Computational Neuroscience

A Comprehensive Approach

CHAPMAN & HALL/CRC
Mathematical Biology and Medicine Series

Aims and scope:
This series aims to capture new developments and summarize what is known over the whole spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical and computational methods into biology by publishing a broad range of textbooks, reference works and handbooks. The titles included in the series are meant to appeal to students, researchers and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press UK
23 Blades Court
Deodar Road
London SW15 2NU
UK

Chapman & Hall/CRC Mathematical Biology and Medicine Series

COMPUTATIONAL NEUROSCIENCE

A COMPREHENSIVE APPROACH

EDITED BY
JIANFENG FENG

CHAPMAN & HALL/CRC
A CRC Press Company
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Computational neuroscience : comprehensive approach / edited by Jianfeng Feng.
p. cm.—(Chapman & Hall/CRC mathematical biology and medicine series)
Includes bibliographical references and index.
ISBN 1-58488-362-6 (alk. paper)
1. Computational neuroscience. I. Feng, Jianfeng. II. Series.

QP357.5.C633 2003
573.81—dc21 2003051578

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com

International Standard Book Number 1-58488-362-6
Library of Congress Card Number 2003051578
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Contents

Preface

The main purpose of this book is to provide a fairly unified and self-contained treatment of the computational (mathematical) theory of the nervous system, starting from atomic level modelling of single channels, finishing with modelling attention and *en route* using many concrete biological systems such as the Hypothalamo-Hypophysial system, weakly electric Fish, the barrel cortex, fly motion vision etc. as examples. The emphasis is on models that are closely tied with experimental observations and experimental data. This is reflected by many chapters where an in-depth mathematical theory intermingles with up-to-date experimental result.

After a century of research, our knowledge of the phenomenology of neural function is staggering. Hundreds of different brain areas have been mapped out in various species. Neurons in these regions have been classified, sub-classified, and reclassified based on anatomical details, connectivity, response properties, and the channels, neuropeptides, and other markers they express. Hundreds of channels have been quantitatively characterized, and the regulation and gating mechanisms are beginning to be understood. Multi-electrode recordings reveal how hundreds of neurons in various brain areas respond in concert to stimuli. Despite this wealth of descriptive data, we still do not have a grasp on exactly how these thousands of proteins are supposed to accomplish computation.

A vast majority of neurons respond to sensory or synaptic inputs by generating a train of stereotypical responses called action potentials or spikes. Deciphering the encoding process which transforms continuous, analog signals (photon fluxes, acoustic vibrations, chemical concentrations and so on) or outputs from other neurons into discrete, fixed-amplitude spike trains is essential to understand neural information processing and computation, since often the nature of representation determines the nature of computation possible. Researchers, however, remain divided on the issue of the neural code used by neurons to represent and transmit information.

On the one hand, it is traditionally assumed that the mean firing rate of a neuron, defined as its averaged response to stimulus presentation, is the primary variable relating neuronal response to a sensory stimulus. This belief is supported by the existence of a quantitative causal relationship between the average firing rate of single cortical neurons and psychophysical judgements made by animals trained to perform specific tasks. The highly variable temporal structure of neural spike trains observed *in vivo* further strengths the view that any statistic other than the averaged response is too random to convey information. Recent findings have indicated that spike timing can be precise and have demonstrated that the fine structure of spike intervals can potentially convey more information than a firing rate code, providing evidence for temporal coding. The precise relative temporal relationships between the outputs of different neurons also appears to be relevant for coding in certain cases.

Presently, it is unclear which, if any, is the universal coding strategy used in the brain. In summary we simply lack a coherent view of how a neuron processes incoming signal, emits spikes, and then controls our daily activity, despite a century long of research. In this book, we attempt to present our current views on the issue.

In contrast to other fields in biology, mathematical thinking and methodology have become entrenched in neuroscience since its very beginning, as is witnessed by the classical work of Hodgkin and Huxley. Indeed, important developments in mathematics, and particularly in statistics (for example, point processes theory), have their roots in this field. Therefore, it will not be surprising that the mathematical analysis of the models studied in this book is often not simple, although to a large extent only calculus, linear algebra, and some basic probability theory are needed. The book is aimed at the level of the final year of undergraduate, graduate students or researchers. In general, we had a reader in mind, who is interested in the theoretical aspects of neuroscience and has a fair knowledge of basic mathematics and a certain fluency in algebraic manipulations.

The book opens with a chapter that introduces the basic mathematical concepts used in computational neuroscience. This chapter may serve as an introduction to the field for students from mathematical or physical sciences, and for biology or neuroscience students who are already acquainted with neuroscience, but feel a need for a theoretically oriented and rigorous introduction.

Chapter 2 deals with computer simulations of ion channel proteins. It reviews briefly the key properties of ion channels for which there is structural information, and discusses different simulation approaches. The next two chapters are devoted to modelling calcium and an important neuronal transmitter: nitric oxide. Calcium has been called the ubiquitous second messenger due to its widespread use in cellular signaling pathways. Calcium plays a role in many different cells for processes such as the release of hormones or neurotransmitters, cell motility or contraction, and the control of gene expression and development. Neurotransmitter release in nerve terminals is mediated by calcium. Calcium is also thought to play a role in synaptic plasticity and learning. Chapter 4 presents a general account of models of nitric oxide (NO). NO is a kind of neurotransmitter that acts, through diffusion, over volumes that potentially contain many neurons and can facilitate signalling between neurons that are not synaptically connected.

In Chapter 5, a general theory based upon the Markov chain for analyzing single channel activity is presented. The information in a single channel record is contained in the amplitudes of the openings, the durations of the open and shut periods, and correlations between them. We use this information to try to identify the nature of the kinetic mechanism of the channel and to estimate rate parameters governing transitions between the various states of such a mechanism. In Chapter 6, we look at the biophysical basis of random firing observed in the nervous system. The chapter intends to bridge the gap between stochastic single channel activity and single neuron activity.

Whereas the first six chapters are focused on spikes and subcellular activity, the emphasis in Chapter 7 and Chapter 8 is more on analyzing bursting activity. Chapter 7 is concentrated on the Hypothalamo-Hypophysial system where a relatively simple output signal is observed: the strength of hormone release. The importance of bursting activity is presented in another example, the electrosensory system of South American weakly electric fish which has proven to be extremely well suited for combined neuroethological and computational studies of information processing from systems neuroscience to the characteristics of ion channels. It first gives a brief introduction to the electrosensory system, describes in detail the *in vivo* firing properties of electrosensory pyramidal cells in the hindbrain of these fish, and reports on the potential behavioral role of bursts. Next, it presents results of *in vitro* studies that have elucidated some of the cellular mechanisms underlying burst generation in pyramidal cells. This is followed by a discussion of detailed compartmental models that successfully reproduce *in vitro* bursting and reduced models offering a dynamical systems perspective on burst mechanisms. It is concluded by comparing burst firing in weakly electric fish to other systems.

As we have mentioned before, a typical feature of a nervous system is the variability observed in single channels, in spikes, in local field potentials and in reaction times. How to tackle and make sense of the experimental data is a challenging task for every theoretician. In Chapter 9, statistical methods of point processes are introduced to handle experimental data (spike trains). It is shown how any point process model may be characterized in terms of its conditional intensity function. The authors then apply the likelihood methods in two actual data analyses: a spike train time-series from a retinal ganglion neuron and the spatial receptive fields of a hippocampal neuron recorded while a rat executes a behavioural task on a linear track.

Commencing from Chapter 10, spiking neuronal networks are employed as tools to explore various function of the nervous system. In Chapter 10, how to build up a biologically realistic model is demonstrated. The requirements for biologically-detailed, realistic network modelling is discussed. The requirements are divided into those for the neurone models, the synaptic models, the pattern of connections between cells, the network inputs. To illustrate these requirements a model of the granule cell layer in the cerebellum and a model of olfactory bulb are are presented.

Chapter 11 and Chapter 12 deal with two important aspects: temporal and spatial dependence of neuronal activity. Spike-timing-dependent plasticity (STDP) is a plasticity rule based on the timing of pre- and postsynaptic spikes. Recent experiments provide ample biological support for this plasticity rule. STDP gives detailed predictions for naturalistic inputs and makes it feasible for the first time to directly compare plasticity rules for naturalistic inputs with experimental data. Therefore it is important to develop a theory to establish the fundamental properties of this plasticity rule and the favorable interplay between theory and experiments will likely make STDP an important area of study. Chapter 11 seeks to summarize recent results in these directions and place them in a coherent framework in comparison to Hebbian rules based on rates. Chapter 12 reviews work related to two perspectives on correlated activity: the high-level approach at which function serves to guide the analysis, and the low-level approach that is bound to the biophysics of single neurons.

Information theory has been widely used and played an important role in neuroscience. Chapter 13 introduces the *series expansion* method of estimating mutual information, which is specifically tailored to the case of sparsely responding neurons. It is applied to a popular model system - the somatosensory cortex of the rat - where the number of evoked spikes per stimulus is also small, and has thereby been able to study issues of spike timing and population coding in a rigorous manner.

Although our sensory systems are composed of auditory, somatosensory, taste and olfaction, and vision, the most studied and perhaps the best understood one is vision. Chapter 14 explores the amazing fly motion vision system. It summarizes our current understanding of fly motion vision with an emphasis on modelling rather than on the large set of available experimental data. After giving an overview of the fly motion vision system, the next part of the chapter introduces the correlation-type of motion detector, a model for local motion detection that has been successfully applied to explain many features of motion vision, not only in flies but also in higher vertebrates including man. This is followed by an outline of how local motion signals become spatially processed by large-field neurons of the lobula plate in order to extract meaningful signals for visual course control. In a final section, the chapter will discuss in what directions current research efforts are pointing to fill in the missing pieces.

Recent advances in neurophysiology have permitted the development of biophysically-realistic models of cortical recurrent local networks. Those models have proved to be valuable for investigating the cellular and network mechanisms of information processing and memory in cortical microcircuits. Recurrent network models of spiking neurons are highly non-linear and display a variety of interesting dynamical behaviors, such as spatially organized firing patterns, synchronous oscillations or coexistence of multiple attractor states. It is notoriously hard to predict, solely by intuition, the behaviour of such networks. Moreover, these large-scale network models routinely consist of many thousands of spiking neurons, so their computer simulations are time-consuming. Therefore, it is highly desirable to develop and mathematically analyze simplified yet faithful *mean-field* theories that are derived from these biophysically-based models. In Chapter 15, the authors review the mean-field theory of recurrent cortical network models in which neurons discharge spikes stochastically with high variability and interact with each other via biologically realistic synapses. Using the theory, models of working memory (active short-term memory) are discussed.

Chapter 16 describes memory systems in the brain based on closely linked neurobiological and computational approaches. The neurobiological approaches include evidence from brain lesions which shows the type of memory for which each of the brain systems considered is necessary; and analysis of neuronal activity in each of these systems to show what information is represented in them, and the changes that take place during learning. The computational approaches are essential in order to understand how the circuitry could retrieve as well as store memories, the capacity of each memory system in the brain, the interactions between memory and perceptual systems, and the speed of operation of the memory systems in the brain.

All chapters before tried to describe, explore and reveal mechanisms behind the nervous systems. In Chapter 17, the authors intend to model motor control and thus close the gap between sensory inputs and motor output.

A key challenge for neural modelling is to explain how a continuous stream of multi-modal input from a rapidly changing environment can be processed by neural microcircuits (columns, minicolumns, etc.) in the cerebral cortex whose anatomical and physiological structure is quite similar in many brain areas and species. However a model that could explain the potentially universal computational capabilities of such microcircuits has been missing. In Chapter 18, the authors propose a computational model that does not require a task-dependent construction of neural circuits. Instead it is based on principles of high dimensional dynamical systems in combination with statistical learning theory, and can be implemented on generic evolved or found recurrent circuitry. This new approach towards understanding neural computation on the micro-level also suggests new ways of modeling cognitive processing in larger neural systems. In particular it questions traditional ways of thinking about neural coding.

Selective visual attention is the mechanism by which we can rapidly direct our gaze towards objects of interest in our visual environment. From an evolutionary viewpoint, this rapid orienting capability is critical in allowing living systems to quickly become aware of possible prey, mates or predators in their cluttered visual world. It has become clear that attention guides where to look next based on both bottom-up (image-based) and top-down (task-dependent) cues. As such, attention implements an information processing bottleneck, only allowing a small part of the incoming sensory information to reach short-term memory and visual awareness. That is, instead of attempting to fully process the massive sensory input in parallel, nature has devised a serial strategy to achieve near real-time performance despite limited computational capacity: Attention allows us to break down the problem of scene understanding into a rapid series of computationally less demanding, localized visual analysis problems. Chapter 19 addresses the issue of how to model attention mechanisms.

Finally, I would like to express my sincere thanks to Seongsook Choi and the people from CRC Press for their help and support during the arduous and lengthy process of finalizing this project. The work was supported by the UK Science and Engineering Research Council.

CONTRIBUTORS

Riccardo Barbieri

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital (U.S.)
Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology (U.S.)

Alexander Borst

Max-Planck-Institute of Neurobiology, Department of Systems and Computational Neurobiology, Am Klopferspitz 18a D-82152 Martinsried (Germany)
borst@neuro.mpg.de,
http://www.neuro.mpg.de/research/scn/

Emery N. Brown

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital (U.S.) Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology (U.S.)
brown@neurostat.mgh.harvard.edu,
http://neurostat.mgh.harvard.edu/brown/

Nicolas Brunel

CNRS, Neurophysique et Physiologie du Système Moteur, Université Paris René Descartes
45 rue des Saints Pères, 75270 Paris Cedex 06, France

Andrew Davison

Yale University School of Medicine, Section of Neurobiology, P.O. Box 208001, New Haven CT 06520-8001 (U.S.)
andrew.davison@yale.edu
http://info.med.yale.edu/neurobio/shepherd/lab/

Uri T. Eden

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital (U.S.), Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology (U.S.)

Jianfeng Feng

Department of Informatics, University of Sussex, Brighton BN1 9QH (U.K.)

Loren M. Frank

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, (U.S.)
Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology (U.S.)

Fabrizio Gabbiani

Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 (U.S.)
babbiani@bcm.tmc.edu,
http://glab.bcm.tmc.edu/

Alan G. Hawkes

European Business Management School, University of Wales Swansea, Swansea SA2 8PP (U.K.)
alan.hawkes@ntlworld.com

Phil Husbands

Centre for Computational Neuroscience and Robotics (CCNR), Department of Informatics, University of Sussex, Brighton (U.K.)

Laurent Itti

University of Southern California, Hedco Neuroscience Building HNB-30A, Los Angeles, CA 90089-2520 (U.S.)
itti@usc.edu,
http://ilab.usc.edu/

Saleet M. Jafri

Program in Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, 10900 University Blvd., Manassas, VA 20110 (U.S.)
sjafri@gmu.edu,
http://www.binf.gmu.edu/jafri/

Rüdiger Krahe

Beckman Institute for Advanced Science and Technology, Department of Molecular and Integrative Physiology, University of Illinois at Urbana/Champaign, 405 N. Mathews Ave. Urbana, IL 61801 (U.S.)

Gareth Leng

Division of Biomedical Sciences, College of Medical and Veterinary Sciences, University of Edinburgh, Hugh Robson Bldg., George Square, Edinburgh EH9 8 XD (U.K.)
gareth.leng@ed.ac.uk

Mike Ludwig

College of Medical and Veterinary Sciences, University of Edinburgh EH9 8XD (U.K.)

Wolfgang Maass

Institute for Theoretical Computer Science, Technische Universitaet Graz, A-8010 Graz (Austria)
maass@igi.tu-graz.ac.at,
http://www.cis.tugraz.at/igi/maass/

Henry Markram

Brain Mind Institute, EPFL, Lausanne, Switzerland

Pietro G. Morasso

University of Genova, DIST (Department of Informatics Systems and Telecommunications), Via Opera Pia 13, I-16145 Genova (Italy)
morasso@dist.unige.it,
http://www.laboratorium.dist.unige.it/ piero/

Thomas Natschlaeger

Institute for Theoretical Computer Science, Technische Universitaet Graz, A-8010 Graz, Austria

Michael O'Shea

Centre for Computational Neuroscience and Robotics (CCNR), Department of Biology, University of Sussex, Brighton (U.K.)

Stefano Panzeri

UMIST, Department of Optometry and Neuroscience, Manchester M6O 1QD (U.K.)

Rasmus S. Petersen

Cognitive Neuroscience Sector, International School for Advanced Studies, Via Beirut 2/4, 34014 Trieste (Italy)
petersen@sissa.it,
http://www.sissa.it/cns/peopletmp/petersen.html

Andrew Philippides

Centre for Computational Neuroscience and Robotics (CCNR), Department of Informatics, University of Sussex, Brighton (U.K.)
andrewop@cogs.susx.ac.uk,
http://www.cogs.susx.ac.uk/users/andrewop/

Arleta Reiff-Marganiec

College of Medical and Veterinary Sciences, University of Edinburgh. EH9 8XD (U.K.)

Alfonso Renart

Volen Center for Complex Systems, Brandeis University, Waltham, MA 02254 (U.S.)

Hugh P.C. Robinson

Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG (U.K.)
hpcr@cus.cam.ac.uk,
http://www.physiol.cam.ac.uk/STAFF/ROBINSON/Index.htm

Edmund T. Rolls

University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, (England)
Edmund.Rolls@psy.ox.ac.uk,
http://www.cns.ox.ac.uk

Nancy Sabatier

College of Medical and Veterinary Sciences, University of Edinburgh EH9 8XD (U.K.)

Emilio Salinas

Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1010 (U.S.)
esalinas@wfubmc.edu,
http://www.wfubmc.edu/nba/faculty/salinas/salinas.html

Vittorio Sanguineti

University of Genova, DIST (Department of Informatics Systems and Telecommunications), Via Opera Pia 13, I-16145 Genova, Italy

Terrence J. Sejnowski

Computational Neurobiology Laboratory, Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla CA 92037, U.S., Department of Biology, University of California at San Diego, La Jolla, CA 92093 (U.S.)

Song Sen

Freeman Building, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY 22734 (U.S.)
songs@cshl.edu,
http://people.brandeis.edu/ flamingo/

Tom Smith

Centre for Computational Neuroscience and Robotics (CCNR), Department of Informatics, University of Sussex, Brighton (U.K.)

Peter D. Tieleman

Dept. of Biological Sciences, University of Calgary, Alberta Canada T2N 1N4
tieleman@ucalgary.ca,
http://moose.bio.ucalgary.ca

Henry C Tuckwell

Epidemiology and Information Science, Faculty of Medicine St Antoine, University of Paris 6, 27 rue Chaligny, 75012 Paris, France
tuckwell@u444.jussieu.fr

Xiao-Jing Wang

Volen Center for Complex Systems, Brandeis University, 415 South Street, Waltham, MA 02254-9110 (U.S.)
xjwang@brandeis.edu,
http://www.bio.brandeis.edu/faculty01/wang.html

Keun-Hang Yang

Department of Neurology, The Johns Hopkins University, School of Medicine, 600 North Wolfe Street, Meyer 2-147 Baltimore, MD 21287 (U.S.)

Chapter 1

A Theoretical Overview

Henry C Tuckwell[1], **and Jianfeng Feng**[2]

[1]*Epidemiology and Information Science, Faculty of Medicine St Antoine, University of Paris 6, 27 rue Chaligny, 75012 Paris, France,* [2]*Department of Informatics, University of Sussex, Brighton BN1 9QH, U.K.*

CONTENTS

1.1 Introduction

This chapter contains a mathematical foundation for the chapters follow. Computational neuroscience is a highly interdisciplinary subject and various mathematical languages are used to deal with the problems which arise. We sketch a mathematical foundation and provide references which can be consulted for further details. We also demonstrate how to mathematically and rigorously formulate certain neuroscience problems. The basic topics we intend to cover are deterministic dynamical system theory [53, 85], stochastic dynamical system theory [32, 38] and a few extra topics related to information theory [4, 11] and optimal control [26, 52].

Excellent books or recent reviews containing an introduction to Neuroscience are [3, 36, 66]. One of the central topics in *Computational Neuroscience* is the coding problem [1, 50, 51]: how does the nervous system encode and then decode information? Early theories of coding in the nervous system focused on rate coding, as in conventional neural network theory, in which the average frequency of action potentials carries information about the stimulus. This type of coding, which was established in the firing responses of such cells as motoneurons and sensory receptors have become associated with the integrate-and-fire mode of operation of a nerve cell where timing is not a key factor. This corresponds to neuron models with stochastic activity, as discussed in Section 1.3. The idea of coding has also been developed in the past few decades that time intervals between impulses in patterns of action potentials play a key role; this manner of information delivery has become known as temporal coding and is suitably described by a deterministic dynamical system (Section 1.2).

1.2 Deterministic dynamical systems

1.2.1 Basic notation and techniques

Let us introduce some basic terms such as attractor, attractive basin, bifurcation, Lyapunov exponent, Lyapunov function; and some basic techniques such as Laplace transformation.

Suppose that $\mathbf{X}(t)$ is a dynamical variable, for example, the membrane potential of a neuron at time t. An *attractor* A of $\mathbf{X}(t)$ is a set of states such that $\mathbf{X}(t) \in A, t > 0$ if $\mathbf{X}(0) \in A$. Examples are fixed point attractors and limit cycle attractors. The resting potential of a neuron is usually a fixed point attractor, and spikes are limit cycle attractors. In most cases, we need to know not only the attractors, but also their *attractive basins*. If we can think of a (fixed point) attractor as the bottom of a bowl, then its attractive basin will be the bowl itself. A ball starting from anywhere inside the bowl (attractive basin) rolls down to its bottom (attractor) and stays there. Hence the attractive basin of an attractor A is the set of initial values from which $\mathbf{X}(t)$ will finally converge to A. For example, when a neuron receives subthreshold stimuli, the membrane potential may finally settle down to a certain value, say A (fixed point attractor). In other words subthreshold stimuli are in the attractive basin of the attractor A. A suprathreshold current stimulus is in the attractive basin of spikes.

The behaviour of a dynamical system $\mathbf{X}(t)$ depends on a parameter λ and might change substantially when λ passes through a critical point λ_0. For example, when $\lambda < \lambda_0$, $\mathbf{X}(t)$ may have only one attractor; but when $\lambda \geq \lambda_0$, there might be two attractors for $\mathbf{X}(t)$. Such a point is called *bifurcation point*. For example, λ could be the (constant) current input to a neuron. When $\lambda < \lambda_0$, the neuron is silent (without spiking) and the only attractor is a fixed point attractor. When the stimulus is strong enough $\lambda > \lambda_0$, the neuron emits spikes, and the attractor is a limit cycle. The various properties at the bifurcation point can be further classified into different categories (see [85] for more details).

Another very useful quantity to characterize a dynamical system is a *Lyapunov exponent*. Consider a one-dimension dynamical system

$$dX(t)/dt = \lambda X(t).$$

$X(t)$ exhibits two distinct behaviours. If λ is larger than zero, $X(t)$ will become infinity; otherwise it will remain finite. In particular, when it is less than 0, $X(t)$ will converge to zero. For a high order dimensional dynamical system satisfying

$$d\mathbf{X}(t)/dt = \Sigma\mathbf{X}(t)$$

Quantities corresponding to λ are the eigenvalues of the matrix Σ. When all eigenvalues are real, we can assert that $\mathbf{X}(t)$ converges if the largest eigenvalue is negative, and diverges if the largest eigenvalue is positive. For a nonlinear dynamical system, we can have a similar picture if we expand it locally at an attractor. Hence from the sign of the largest Lyapunov exponent we can tell whether the dynamical system is stable or not. An attractor of a dynamical system is called a *strange attractor* if its largest Lyapunov exponent is positive (page 186, [85]).

To check whether a given dynamical system is stable or not is not a trivial issue. One of the most widely used quantities is the *Lyapunov function*. A Lyapunov function is such that the dynamical system will move downwards along the trajectory of the system, i.e. $d(\mathbf{X}(t))/dt < 0$. For a given system, however, how to construct a Lyapunov function is not always straightforward (see for example [16]).

Many neuronal models are *relaxation oscillators* which may be roughly described as follows. Consider a seesaw as shown with a container on one side (A) in which electricity (water) can be held. If the container is empty, the other side (B) of the seesaw touches the ground. From a tap, water is dripping into the container and at a certain water level, point B rises and point A will touch the ground. At this moment the container empties itself, the seesaw bounces quickly back to its original position and the process starts again. Such an oscillation is characterized by intervals of time in which very little happens, followed by short intervals of time in which notable changes take place.

For a function $f(t)$ defined on $[0,\infty)$, the *Laplace transformation* of $f(t)$ is the function $F(s)$ defined by the integral

$$F(s) = \int_0^\infty \exp(-st)f(t)dt$$

We denote the Laplace transformation of $f(t)$ as $\mathscr{L}(f)(s)$. The idea behind Laplace transforms is to *simplify the problem*. After applying Laplace transformation to a linear differential equation, we have to apply the inverse Laplace transformation to obtain the solution.

1.2.2 Single neuron modelling

We introduce point model, cable model and multi-compartment model. The dynamics of individual neurons are governed by a multiplicity of state variables, including membrane voltage, channel activation parameters, intracellular ion concentrations, and cell morphology etc.

Point model

- Simplified models. The simplest model of a neuron is the integrate-and-fire model, an equivalent description of the seesaw model in the previous section. If $V(t)$ is the membrane potential and $V(t) \leq V_{thre}$ (threshold), then

  $$dV(t)/dt = -\frac{V(t) - V_{rest}}{\gamma} + I \tag{1.1}$$

  where I is input current and γ is the membrane time constant. When $V(t) = V_{thre}$ we have $V(t+) = V_{rest}$, the resting potential. A refractory period T_{ref} can be introduced if we define $V(t + T_{ref}) = V_{rest}$.

  Another widely used model is called the *FitzHugh-Nagumo (FHN) model.* It has two variables which satisfy

  $$\begin{cases} dv/dt = K[-v(v-\alpha)(v-1) - w] + I \\ dw/dt = b[v - cw]. \end{cases}$$

  Here v is the membrane potential and w is the recovery variable, α, c, K, b are all constants. Typical parameters are $\alpha = 0.2, c = 2.5, K = 100, b = 0.25$.

  When $dw/dt = 0$, i.e. $w = bv/c$, the model is usually called the reduced FHN model, or the Landau model in physical literature. Introducing a threshold to the reduced FHN model as in the integrate-and-fire model, the model is called the *IF-FHN model* [18]. Comparing with the integrate-and-fire model, the IF-FHN model has a nonlinear leakage term $K(v-1)(v-\alpha) + b/c$. Due to the nonlinear leakage term, some interesting properties including more sensitive to correlated inputs than the IF model have been reported [6, 23].

- Biophysical models. Consider the *Hodgkin-Huxley model*, which is the basis of most other biophysical models.

  The Hodgkin-Huxley model is written by

  $$\begin{aligned} CdV/dt = & -g_{Na}m^3h(V - V_{Na})dt - g_K n^4(V - V_K) \\ & -g_L(V - V_L) + I. \end{aligned} \tag{1.2}$$

  The term m^3h describes the sodium channel activity and n^4 reflects the potassium channel activity. Equations about m, n, h and parameters g_{Na}, g_k, g_L, V_{Na}, V_K, V_L used in the Hodgkin-Huxley model are in Appendix A. Other channels may be included in the Hodgkin-Huxley model, the most commonly being the calcium channel. In Chapter 3, the authors present a detailed review on modelling calcium.

The model exhibits different behaviours in response to external current inputs. A response may be described by an $F-I$ curve, where F is the output frequency and I is the input current. When $F-I$ is continuous, it is termed *Type I* neuron (the integrate-and-fire model is of this type); when $F-I$ is not continuous, it is *Type II* neuron (the Hodgkin-Huxley is of this type). Mathematically, the Type I neuron is usually due to a saddle-node bifurcation, whereas the Type II neuron is due to a Hopf bifurcation.

Cable model In point models, we ignore the geometric properties of cells. One way to include cell morphology in modelling is to treat cell segments as cylinders, as described in the cable Equation (1.11).

Multi-compartment models Another way to include cell morphology is with a multi-compartment model. According to actual neuronal anatomy, a biophysical model of a few hundred compartments may be appropriate and it is a formidable task to analyze such a model. Models with only a few compartments have been investigated and it is surprising that such simplified models usually fit more complicated models well [12, 56]. Here we consider two-compartment models, composed of a somatic and a dendritic compartment.

- A two-compartment abstract model (see the integrate-and-fire model before). When the somatic membrane potential $V_s(t)$ is below the threshold V_{thre},

$$\begin{cases} dV_s(t)/dt = -\frac{1}{\gamma}(V_s(t) - V_{rest}) + g_c \frac{V_d(t) - V_s(t)}{p} \\ dV_d(t)/dt = -\frac{1}{\gamma}(V_d(t) - V_{rest}) + g_c \frac{V_s(t) - V_d(t)}{1-p} + \frac{I}{1-p} \end{cases} \tag{1.3}$$

  where $1/\gamma$ is the decay rate, p is the ratio between the membrane area of the somatic compartment and the whole cell membrane area, V_d is the membrane potential of the dendritic compartment, $g_c > 0$ is a constant.

  The properties of the two-compartment model above have been examined by a few authors, see for example [62]. We have reported that a two-compartment model is naturally a slope detector [21, 34, 46].

- A two-compartment biophysical model. A simplified, two-compartment biophysical model, proposed by Pinsky and Rinzel [56] is described here. They have demonstrated that the model mimics a full, very detailed model of pyramidal cells quite well.

The Pinsky-Rinzel model is defined by

$$\begin{cases} C_m dV_s(t)/dt = -I_{Leak}(V_s) - I_{Na}(V_s,h) - I_{K-DR}(V_s,n) \\ \qquad\qquad\qquad + g_c \dfrac{V_d(t)-V_s(t)}{p} \\ C_m dV_d(t)/dt = -I_{Leak}(V_d) - I_{Ca}(V_d,s) - I_{K-AHP}(V_d,q) \\ \qquad\qquad\qquad -I_{K-C}(V_d,Ca,c) + \dfrac{I}{1-p} + g_c \dfrac{V_s(t)-V_d(t)}{1-p} \\ \qquad [Ca]' = -0.002 I_{Ca} - 0.0125[Ca] \end{cases} \tag{1.4}$$

All parameters and other equations of ionic channels can be found in [56], $[Ca]$ is the calcium concentration (see Chapter 3 for a detailed account of modelling calcium activity).

1.2.3 Phase model

It is relatively easy to characterize deterministic neuronal models, in comparison with models with stochastic behaviour which are described in the next section. One of the popular ways to carry out such an analysis is by means of the so-called *phase model* [15, 42, 69].

Models When the spiking output of a cell is approximately periodic, the underlying dynamics may be described by a single variable known as the phase, usually denoted by $\theta(t) \in [0, 2\pi]$. As θ changes from 0 to 2π, the neuronal oscillator progresses from rest to depolarization to spike generation to repolarization and around again over the course of one period.

Within the phase description framework, sometimes we are able to calculate and understand how the detailed description of the synaptic interactions among neurons can effect their spiking timing and, thus, lead to the formation of spatially and temporally patterned electrical output. More specifically,

$$\frac{d\theta_i}{dt} = \omega_i + \sum_{j=1}^{N} \Gamma_{ij}(\theta_j - \theta_i), i = 1, \cdots, N. \tag{1.5}$$

where θ_i is the phase of the *ith* neuron, ω_i is the initial phase and Γ is the interaction between neurons. Even though the reduction to a phase model represents a great simplification, these equations are still too difficult to analyze mathematically, since the interaction functions could have arbitrarily many Fourier harmonics, which is always the case in an actual situation, and the connection topology is unspecified and largely unknown in biology.

The *Kuramoto model* corresponds to the simplest possible case of equally weighted, all-to-all, purely sinusoidal coupling

$$\Gamma_{ij}(\theta_j - \theta_i) = \frac{K}{N}\sin(\theta_j - \theta_i). \tag{1.6}$$

Many theoretical results are known for the Kuramoto model, as briefly described below.

Synchronization The properties of the Kuramoto model have been intensively studied in the literature [69]. Let us first introduce two parameters which characterize the synchronization of a group of oscillators

$$r(t)\exp(i\psi) = \frac{1}{N}\sum_{i=1}^{N}\exp(i\theta_i) \tag{1.7}$$

Geometrically $r(t)$ is the order parameter describing the synchronization among neurons. Numerically, it is shown that when the coupling strength K between neurons is smaller than a constant K_c, the neurons act as if they were uncoupled. $r(t)$ decays to a tiny jitter of size of order $1/\sqrt{N}$. But when K exceeds K_c, the incoherent state becomes unstable and $r(t)$ grows exponentially, reflecting the nucleation of a small cluster of neurons that are mutually synchronized, thereby generating a collective oscillation. Eventually $r(t)$ saturates at some level being smaller than 1. For some most recent results on the Kuramoto model, we refer the reader to [54].

A detailed computation on how two neurons synchronize their activity has also been carried out in [86]. They found that inhibitory rather than excitatory synaptic interactions can synchronize neuron activity.

1.3 Stochastic dynamical systems

1.3.1 Jump processes

The observed electrical potentials of neurons as determined either by extracellular or intracellular recording are never constant. The same is true for grossly recorded field potentials and brain recordings such as the electroencephalogram. Often such recordings of potential exhibit quite sudden changes or jumps. If the sample paths of a continuous time random process have *discontinuities* then it is called a *jump process*. A process may be a pure jump process, like a *Poisson process* or a random walk, or there may be drift and or diffusion between the jumps.

Motivation for using jump processes in neurobiological modelling sprang primarily from observations on excitatory and inhibitory synaptic potentials (EPSPs and IPSPs). Examination of, for example, motoneuron or pyramidal cell somatically recorded EPSPs may show a rapid depolarization of several millivolts relative to resting potential, followed by an exponential decay with a characteristic time constant [70]. A complete understanding of these events requires the use of complex *spatial models*, (see below) but in the majority of studies attempting to model neuronal electrophysiological properties in the last several decades, spatial extent has, regrettably, been ignored, probably because of the unwillingness of but a few theorists to confront partial differential equations rather than ordinary ones.

Putting aside the matter of spatial versus *point models*, if $N = \{N(t), t \geq 0\}$ is a simple standard (unit jumps) Poisson process, with rate parameter λ, then a simple one-dimensional model for the subthreshold (less than $V_{thre} \approx$10-20mV) depolarization of a single neuron can be written as the stochastic differential equation (SDE)

$$dV = -\frac{V - V_{rest}}{\gamma} dt + a_E dN, \qquad V < V_{thre}, \qquad V(0) = V_0, \tag{1.8}$$

where γ is the time constant of decay and $a_E > 0$ is the magnitude of an EPSP (c.f. Equation (1.1)). In the very small time interval $(t, t + \Delta t]$, either N and hence V jumps, with probability $\lambda \Delta t$ or doesn't; if it doesn't then V decreases according to $dV/dt = -V/\gamma$.

A characteristic of a Poisson process is that the (random) time T between jumps or events has an *exponential distribution* with mean $1/\lambda$:

$$Pr\{T \leq t\} = 1 - \exp(-\lambda t), t \geq 0.$$

We may suppose that jumps still occur at rate λ, but that the jump size is random with probability density $\phi(u)$ or a distribution function $\Phi(u)$. This gives a *compound Poisson process* X. This means that jumps with amplitudes in $(u, u + du]$ occur with a relative rate of $\phi(u)du$ or an absolute rate of $\lambda\phi(u)du$. We let $N(du,t)$ count the number of such events in $(0,t]$ and the total contribution of these events will be $uN(du,t)$, being jump amplitude multiplied by the number of jumps with this amplitude. The whole compound Poisson process will be obtained by integrating over all possible amplitudes

$$X(t) = \int_{-\infty}^{-\infty} uN(du,t),$$

where the total jump or event rate is of course $\int \lambda\phi(u)du = \lambda$.

The above *leaky integrate and fire* model (1.8) may thus be extended to include an arbitrary distribution of postsynaptic potential amplitudes:

$$dV = -\frac{V}{\gamma} dt + \int_{-\infty}^{-\infty} uN(du, dt), \qquad V < V_{thre}, \qquad V(0) = V_0. \tag{1.9}$$

Since the positions of inputs are not distinguished in such point models, the input here could arise from many separate excitatory and inhibitory synapses. The representation [79]

$$V(t) = V(0)\exp(-\frac{t}{\gamma}) + \int_0^t \exp\left(-\frac{t-s}{\gamma}\right) \int uN(du, ds),$$

enables us to find the mean of the unrestricted potential at time t

$$E[V(t)] = E[V(0)]\exp\left(-\frac{t}{\gamma}\right) + \mu_1(1 - \exp\left(-\frac{t}{\gamma}\right))$$

where $\mu_1 = \int u\phi(u)$ is the mean postsynaptic potential amplitude. Similarly, the variance is found to be

$$\mathrm{Var}[V(t)] = \frac{\mu_2}{2}(1 - \exp\left(-\frac{2t}{\gamma}\right)),$$

where μ_2 is the second moment of ϕ.

For the model (1.8) analytical solutions for the firing time (ISI) are difficult to obtain because they involve differential-difference equations. These were first solved in [72] and later using analytical and numerical methods for excitation and inhibition in [10]. Such discontinuous processes had been neglected because it was easier to deal with differential equations, which arise in the theory and properties of diffusion processes (see below). However, with the great power of the desktop computers now available, it is a simple task to quickly estimate the interspike time distribution generated by a model such as (1.8) or (1.9) using simulation.

We may add physiological realism to (1.9) by including synaptic *reversal potentials*. These make postsynaptic potential amplitudes smaller when the equilibrium potentials for transmitter-induced conductance changes are approached [73]. This gives,

$$\begin{aligned} dV &= -\frac{V}{\gamma}dt + (V - V_E)\int_{-\infty}^{-\infty} uN_E(du,dt) + \\ &\quad +(V - V_I)\int_{-\infty}^{-\infty} uN_I(du,dt), \qquad V < V_{thre}, \qquad V(0) = V_0, \end{aligned} \tag{1.10}$$

where V_E and V_I are the reversal potentials for excitation and inhibition and N_E and N_I give the input frequency and amplitude distributions.

A *spatial model* may similarly be constructed with jump processes [77] using the linear cable equation on (a,b)

$$\begin{aligned} \frac{\partial V}{\partial t} &= -V + \frac{\partial^2 V}{\partial x^2} + (V - V_E)\sum_{i=1}^{n_E} \delta(x - x_{E,i})a_{E,i}\frac{dN_{E,i}}{dt} \\ &\quad +(V - V_I)\sum_{j=1}^{n_I} \delta(x - x_{I,j})a_{I,j}\frac{dN_{I,j}}{dt}. \end{aligned} \tag{1.11}$$

This gives a better representation of a neuron than either (1.9) or (1.10) because now the postsynaptic potentials have finite rise times according to experimental data and separate spatial locations for inputs are distinguished. Here there are n_E excitatory synaptic inputs at positions $x_{E,i}, i = 1,...,n_E$ with amplitudes $a_{E,i}$ and n_I inhibitory synaptic inputs at positions $x_{I,j}, j = 1,...,n_I$ with amplitudes $a_{I,j}$ and it is assumed that all excitatory (inhibitory) inputs have the same reversal potential, which is often the case. A region of low threshold is chosen as a trigger zone, at say $x = 0$, and boundary conditions must be imposed at $x = a$ and $x = b$. A realistic boundary condition at the soma could be a lumped-soma, being a capacitance and resistance in parallel to represent somatic membrane.

General jump process point model with drift

Multidimensional continuous time Markov processes are well suited for describing much of the electrophysiological and biochemical behavior of neurons and networks of neurons, especially when neuronal spatial extent is ignored. Thanks mainly to Itô [35], Feller and Kolmogorov, the analytical theory of such processes and their representations by *stochastic integrals* (or equivalently stochastic differential equations) was rigorously developed in the middle of the 20th century. Consider a deterministic differential equation for the n-vector $\mathbf{X}(\mathbf{t})$ of the form

$$\frac{d\mathbf{X}(\mathbf{t})}{dt} = \mathbf{f}(\mathbf{X}(\mathbf{t}),t),$$

with initial value $\mathbf{X}(\mathbf{0})$. Interrupting the deterministic trajectories there may be superimposed jumps of various amplitudes and frequencies, both possibly time-dependent or depending on the values of the components of $\mathbf{X}(\mathbf{t})$, representing synaptic inputs. The function $\mathbf{f}$ is called the *drift* and with the discontinuous components representing synaptic inputs of various amplitudes and frequencies we have

$$d\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t),t)dt + \int \mathbf{h}(\mathbf{X}(t),t,\mathbf{u})N(d\mathbf{u},dt),$$

N being a Poisson random measure defined on subsets of $R^n \times [0,\infty)$. Such a general system covers all non-spatial conductance-based models such as Hodgkin-Huxley or approximations like Fitzhugh-Nagumo with almost every possible pattern of synaptic input. For more details in relation to neuronal modelling and extensions to the spatially distributed case see [77, 79].

1.3.2 Diffusion processes

Diffusion processes are an abstract approximation to empirical processes which have the advantage of being less cumbersome to analyze than processes with jumps. Providing the postsynaptic potentials as seen at the soma are not very large and are fairly frequent, a diffusion model should perform reasonably well.

The simplest diffusion model

The simplest diffusion process employed for modelling a neuron is based on the unrealistic *perfect integrator* model and hence is only of historical interest. Unfortunately it is the only diffusion model which can be solved exactly for all parameter values. It consists of a *Wiener process* (Brownian motion) with drift and was introduced by Gerstein and Mandelbrot [28]. Consider a random walk consisting of the difference of two Poisson processes N_E and N_I, corresponding to excitatory and inhibitory input respectively,

$$dX = a_E dN_E - a_I dN_I,$$

where $a_E \geq 0$ and $a_I \geq 0$ are the magnitudes of steps up or down. Since, using the properties of Poisson random variables, the mean of $X(t)$ is $E[X(t)] = (\lambda_E a_E - \lambda_I a_I)t$ and its variance is $Var[X(t)] = (\lambda_E a_E^2 + \lambda_I a_I^2)t$, a diffusion approximation V to X is given by

$$dV = (\lambda_E a_E - \lambda_I a_I)dt + \sqrt{(\lambda_E a_E^2 + \lambda_I a_I^2)}dW, \qquad V < V_{thre}, \qquad V(0) = V_0.$$

(Note that W, a standard Wiener process, is equivalent to a standard Brownian motion B).

Putting $\mu = \lambda_E a_E - \lambda_I a_I$ and $\sigma^2 = \lambda_E a_E^2 + \lambda_I a_I^2$, it can be shown that V will reach $V_{thre} > V(0)$ with probability one if and only if $\mu \geq 0$; that is, the net excitatory drive is greater or equal to the net inhibitory drive. This is not true for more realistic models. When $\mu \geq 0$ the probability density of the time for V to get from a value $V(0) < V_{thre}$ to threshold is the *inverse Gaussian*

$$f(t) = \frac{(V_{thre} - V(0))}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(V_{thre} - V(0) - \mu t)^2}{2\sigma^2 t}\right], t > 0.$$

Gluss model - Ornstein-Uhlenbeck process (OUP)

The jump process model with exponential decay given by (1.9) can be similarly approximated by a diffusion model which is, for subthreshold V

$$dV = \left(-\frac{V}{\gamma} + \mu\right) dt + \sigma dW. \tag{1.12}$$

This defines an *Ornstein-Uhlenbeck process*, which as a neuronal model was first derived and analyzed by Gluss in 1967 [31]. Here V has continuous paths and, if unrestricted, the same first and second moments as the jump process. Both processes may get to the same values, such as a threshold for an action potential, at about the same time, but in many cases this will be far from the truth, depending on the values of the four parameters $\lambda_E, a_E, \lambda_I$ and a_I – see [81] for a complete discussion. Thus *extreme caution* must be exercised if using a diffusion model such as (1.12) to obtain input-output relations for neurons. For example, it was found in [27] that an inhibitory input can drive the model to fire faster, with a fixed excitatory input, but this is simply due to the error introduced by the diffusion approximation. Roy and Smith [63] first solved the difficult problem of obtaining an exact expression for the mean firing time in the case of a constant threshold. Even recently this model has attracted much attention [18, 29, 44]. The OUP has also been shown to be a suitable approximation for *channel noise*, see for example [75], Chapter 5 and Chapter 6.

The general theory of diffusion processes is broad and often quite abstract. Such fundamental matters as (appropriate) definition of stochastic integral and boundary classifications are important but generally outside the domain of most computational neuroscientists [38]. Fortunately such matters can be sidestepped as most modelling is pragmatic and will involve trial and error *simulation methods* using software packages. However, it is useful to realize that diffusion processes, whether of one or several dimensions, have an associated linear partial differential equation satisfied by the transition probability function or its density. This is also true for Markov jump processes, but the corresponding equations are more complicated and far less studied.

Analytical theory for diffusion processes

Letting $\mathbf{X(t)}$ be a vector with n components, all neuronal ordinary differential equation models in which the input current is approximated by white noise have the general form

$$d\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t), t)dt + \mathbf{g}(\mathbf{X}(t), t)d\mathbf{W}(t),$$

where $\mathbf{f}$ also has n components, $\mathbf{g}$ is an $n \times m$ matrix and $\mathbf{W}$ is an m-vector of (possibly) independent standard Wiener processes. This form covers nonlinear models such as Hodgkin-Huxley, Fitzhugh-Nagumo etc. and certain network approximations. Let $p(\mathbf{y},t|\mathbf{x},s)$ be the transition probability density function of the process $\mathbf{X}$, defined with $s < t$ through, $p(\mathbf{y},t|\mathbf{x},s)d\mathbf{y}) = \Pr\{\mathbf{X}(t) \in (\mathbf{y},\mathbf{y}+d\mathbf{y})|\mathbf{X}(s) = \mathbf{x}\}$. Then p satisfies two partial differential equations. Firstly, the *forward Kolmogorov equation* (sometimes called a Fokker-Planck equation)

$$\frac{\partial p}{\partial t} = -\sum_{k=1}^{n} \frac{\partial}{\partial y_k}[f_k(\mathbf{y},t)p] + \frac{1}{2}\sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial^2}{\partial y_k \partial y_l}\left[(\mathbf{g}(\mathbf{y},t)\mathbf{g}^T(\mathbf{y},t))_{kl}p\right],$$

where superscript T denotes transpose. Secondly, holding forward variables fixed gives the *backward Kolmogorov equation*

$$\frac{\partial p}{\partial s} = -\sum_{k=1}^{n}[f_k(\mathbf{x},s)]\frac{\partial p}{\partial x_k} - \frac{1}{2}\sum_{l=1}^{n}\sum_{k=1}^{n}(\mathbf{g}(\mathbf{x},s)\mathbf{g}^T(\mathbf{x},s))_{kl}\frac{\partial^2 p}{\partial x_k \partial x_l}.$$

We may write this as

$$\frac{\partial p}{\partial s} + L_{\mathbf{x}}p = 0,$$

to define the operator $L_{\mathbf{x}}$ which is useful in finding first passage times such as the time to reach a specified electrophysiological state – see [79] for details.

An example - analytical results for the OUP

For the model (1.12), the analytical theory is simple and the resulting differential equations are easily solved. Putting the time unit as the membrane time constant, the forward equation is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}[(-y+\mu_1)p] + \frac{\mu_2}{2}\frac{\partial^2 p}{\partial y^2}.$$

In Feller's terminology, the points at $y = \pm\infty$ are *natural* boundaries and p must vanish there. All other points are *regular* so they are visited with probability one in a finite time. The *unrestricted process* is Gaussian and its distribution is easily found using the mean and variance for (1.3.2). If a *threshold* is put at $y = V_{thre}$, an absorbing condition $p(V_{thre},t|x,s) = 0$ gives a solution from which the neuronal firing time distribution can be found. Alternatively one may solve on $(-b, V_{thre})$ with $V_{thre}, b > 0$ the equation for the mean exit time from $M(x)$, starting at x, from $(-b, V_{thre})$

$$L_x M(x) = \frac{\mu_2}{2}\frac{d^2 M}{dx^2} + (\mu_1 - x)\frac{dM}{dx} = -1$$

with boundary conditions $M(-b) = M(V_{thre}) = 0$. Letting $b \to \infty$ gives the mean time for the neuronal depolarization from rest to get to the threshold for an action potential. The solution (using the Laplace transformation introduced in the previous section) is in a series of parabolic cylinder functions [63].

Spatial diffusion process models - SPDEs

If the postsynaptic potentials are not too large and fairly frequent, a diffusion approximation for a spatial model such as (1.11) may be employed. Linear models of this kind may involve distributed one-parameter white noises representing each synaptic input or group of synaptic inputs

$$\frac{\partial V}{\partial t} = -V + \frac{\partial^2 V}{\partial x^2} + (V - V_E)\sum_{i=1}^{n_E} \delta(x - x_{E,i})\big(a_{E,i}\lambda_{E,i} + |a_{E,i}|\sqrt{\lambda_{E,i}}\frac{dW_{E,i}}{dt}\big)$$
$$+(V - V_I)\textstyle\sum_{j=1}^{n_I} \delta(x - x_{I,j})\big(a_{I,j}\lambda_{I,j} + |a_{I,j}|\sqrt{\lambda_{I,j}}\frac{dW_{I,j}}{dt}\big).$$

Simplified versions of this and similar models were analyzed in [83, 84]. Alternatively, if the synapses are very densely distributed, a two-parameter white noise may be employed as an approximation:

$$\frac{\partial V}{\partial t} = -V + \frac{\partial^2 V}{\partial x^2} + f(x,t) + g(x,t)\frac{\partial^2 W}{\partial t \partial x},$$

where $W(t,x)$ is a standard two-parameter Wiener process. For details see [84].

1.3.3 Jump-diffusion models

It is possible that some inputs to a neuron, including channel noise are frequent and of small amplitudes whereas others are less frequent and large amplitude, such as occur at certain large and critically placed synapses or groups of synapses of the same type. Such a model was introduced in [74] and in its simplest form has the stochastic equation

$$dV = -Vdt + a_E dN_E + a_I dN_I + \sigma dW, \tag{1.13}$$

where the unit of time is the time constant. The corresponding equations for the n-th moments of the firing time for an initial potential x can be obtained by solving:

$$\frac{\sigma^2}{2}\frac{d^2 M_n}{dx^2} - x\frac{dM_n}{dx} + \lambda_E M_n(x + a_E) + \lambda_I M_n(x - a_I) - (\lambda_E + \lambda_I)M_n(x)$$
$$= -nM_{n-1}(x),$$

$n = 1, \ldots,$ with $M_0 = 1$. Here λ_E, λ_I are the mean frequencies of excitation and inhibition, respectively. However, for particular parameter values, solutions can be readily obtained by simulating the Poisson and white noise inputs. Neuron models with correlated inputs can be exactly written as Equation (1.13) [19, 68, 90].

1.3.4 Perturbation of deterministic dynamical systems

One of the key topics addressed in the theory of differential equations or dynamical systems is the asymptotic (large time) effect of a small disturbance or perturbation on a reference solution or orbit, such as an equilibrium point or limit cycle. The fundamental methods employed for deterministic systems are called Lyapunov's first and second methods. For stochastic dynamical systems, which have an extra dimension, results on stability are naturally more difficult to obtain [5]. Here we consider the effects (or methods of determining them) on some neuronal systems of perturbations with small Gaussian white noise.

Firing time of a model neuron with small white noise

Consider an OUP model with threshold V_{thre} and stochastic equation

$$dV = (-V + \mu)dt + \sigma dW.$$

It should be noted that in the absence of noise and in the absence of a threshold, the steady state potential is μ. If $\mu \leq V_{thre}$ the deterministic neuron never fires whereas if $\mu > V_{thre}$ the firing time is

$$T = T_R + \ln\left(\frac{\alpha}{\alpha - 1}\right),$$

where $\alpha = \mu/V_{thre}$ and T_R is the refractory period. If we define the small noise parameter $\varepsilon^2 = \sigma/V_{thre}$ then using perturbation techniques ([87] the mean and variance of the firing time can be found to order ε^2 as follows.

Steady state well above threshold.

When $\alpha >> \varepsilon + 1$,

$$E[T] \approx T_R + \ln\left(\frac{\alpha}{\alpha - 1}\right) - \frac{\varepsilon^2}{4}\left[\frac{1}{(\alpha-1)^2} - \frac{1}{\alpha^2}\right],$$

$$Var[T] \approx \frac{\varepsilon^2}{2}\left[\frac{1}{(\alpha-1)^2} - \frac{1}{\alpha^2}\right].$$

These results show clearly how small noise reduces the mean interspike interval.

Steady state well below threshold.

When $\alpha << 1 - \varepsilon$, the expectation of the interspike interval is

$$E[T] \approx T_R + \frac{\varepsilon\sqrt{\pi}}{1-\alpha} exp\left[\frac{(1-\alpha)^2}{\varepsilon^2}\right],$$

and the variance is

$$Var[T] \approx \frac{\varepsilon^2 \pi}{(1-\alpha)^2} exp\left[\frac{2(1-\alpha)^2}{\varepsilon^2}\right],$$

Many other results are given in the aforementioned reference, which includes an exhaustive study of the dependence of the *coefficient of variation* of T on the input parameters.

Differential equations for moments under Gaussian white noise perturbations

Ordinary differential equations have been derived for the asymptotic moments of the dynamical variables in a general system of coupled (nonlinear) stochastic differential equations with white noise perturbations of the form

$$dX_j = f_j(\mathbf{X}, t)dt + \sum_{k=1}^{m} g_{jk}(\mathbf{X}, t)dW_k$$

where the W_k are standard Wiener processes – see [61]. For example, consider the Fitzhugh-Nagumo system

$$dX = [f(X) - Y + I]dt + \beta dW$$

$$dY = b[X - cY]dt,$$

where $f(X) = kX(X-a)(1-X)$. The means of X and Y, denoted by m_1 and m_2 respectively, satisfy the equations

$$dm_1/dt = f(m_1) - m_2 + f''(m_1)S_1/2 + I(t)$$

$$dm_2/dt = b(m_1 - cm_2)$$

where S_1 is the variance of X. Denoting the variance of Y by S_2 and the covariance of X and Y by C_{12} we also have

$$dS_1/dt = 2f'(m_1)S_1 - 2C_{12} + \beta^2$$

$$dS_2/dt = 2b(C_{12} - cS_2)$$

and

$$dC_{12}/dt = bS_1 - S_2 + C_{12}[f'(m_1) - cb].$$

This system of five ordinary differential equations may be easily solved and for small β gives good agreement with moments from simulations (see [82]). The method can also be used for small biological neuronal networks.

White noise perturbation of spatial nonlinear neuronal models

The analysis of spatial neuronal nonlinear model equations under the effects of white noise perturbations has been performed for both scalar and vector forms of the Fitzhugh-Nagumo model. In all cases a perturbation expansion was used to obtain the moments of the dynamical variables. As a simple example consider the Fitzhugh-Nagumo system without recovery driven by white noise of small amplitude:

$$u_t = u_{xx} + f(u) + \varepsilon(\alpha + \beta W_{xt})$$

where W is a two-parameter Wiener process. An expansion in powers of ε

$$u = u_0 + \sum_{k=1}^{\infty} \varepsilon^k u_k$$

yields a recursive system of linear stochastic partial differential equations for the u_k. Solving the system recursively yields series expressions for the moments and spectrum of the potential. These results, results on the full Fitzhugh-Nagumo system of SPDEs and a general result on perturbation of a nonlinear PDE with white noise are derived in [76, 78, 80].

A general approach to analyze dynamical systems with small perturbations has been developed in recent years called *large deviation theory*. Basically, it is a generalization of the well-known Kramer's formula [60]. A general description is contained in [27], see also [2, 14].

Finally a few remarks on the relationship between deterministic and stochastic dynamical systems. *A deterministic neuron model is a special case of a stochastic neuron model.* When the noise term vanishes, a stochastic neuron model automatically becomes deterministic. Usually there is a correspondence between the notations of stochastic and deterministic dynamical systems. For example, for the Lyapunov exponent introduced in the previous section for the deterministic system, we can introduce analogous notation for a stochastic dynamical system (see for example [55]). With the help of the Lyapunov exponent, we can understand some phenomena such as how stochastic but not deterministic currents can synchronize neurons with different initial states [20, 49].

1.4 Information theory

The nervous system is clearly a stochastic system [65], so we give a brief introduction of information theory [67]. Since neurons emit spikes randomly, we may ask how to characterize their input-output relationships. The simplest quantity is the correlation between input and output. However, information theory, with its roots in communication theory, has its own advantage.

1.4.1 Shannon information

Intuitively, information is closely related to the element of surprise. Hence for an event A, we define

$$S(A) = -\log_2(P(A))$$

as the (Shannon) information of the event A, so that the information in a certain event is zero.

For a discrete random variable X with $P(X = j) = p_j$, its entropy is the mean of its information, i.e.,

$$H(X) = -\sum_{j=1}^{n} p_j \log_2(p_j)$$

When X is a continuous random variable its entropy is thus given by

$$H(X) = -\int p(x)\log_2 p(x)dx$$

where $p(x)$ is the density of X. The notation of entropy in information theory was first introduced by Claude Shannon, after the suggestion of John von Neumann. “You should call it *Entropy* and for two reasons: first, the function is already in use in thermodynamics under that name; second, and more importantly, most people don’t know what entropy really is, and if you use the word *entropy* in an argument you will win every time!”.

1.4.2 Mutual information

For a random vector $\mathbf{X}$, let $f_{\mathbf{X}}(\mathbf{x})$ be its probability density. For two random vectors $\mathbf{X}, \mathbf{Y}$, denote $H_{\mathbf{X}}(\mathbf{Y})$ as a measure of the information content of $\mathbf{Y}$ which is not contained in $\mathbf{X}$. In mathematical terms it is

$$H_{\mathbf{X}}(\mathbf{Y}) = -\int p(\mathbf{y}|\mathbf{x}) \log p(\mathbf{y}|\mathbf{x}) d\mathbf{y}$$

where $p(\mathbf{y}|\mathbf{x})$ is the conditional density of $\mathbf{Y}$, given $\mathbf{X}$. The mutual information between $\mathbf{X}$ and $\mathbf{Y}$ is

$$I(\mathbf{X},\mathbf{Y}) = H(\mathbf{Y}) - H_{\mathbf{X}}(\mathbf{Y}) = \int\int f_{(\mathbf{X},\mathbf{Y})}(\mathbf{x},\mathbf{y}) \log \frac{f_{(\mathbf{X},\mathbf{Y})}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})} d\mathbf{x} d\mathbf{y}$$

where the information content of $\mathbf{Y}$ which is also contained in $\mathbf{X}$. In other words, the mutual information is the Kullback-Leibler distance (relative entropy): the distance between $(\mathbf{X},\mathbf{Y})$ and $\mathbf{X},\mathbf{Y}$, where $\mathbf{X},\mathbf{Y}$ are treated as independent variables. The mutual information measures the distance between possibly correlated random vectors $(\mathbf{X},\mathbf{Y})$ and independent random vectors $\mathbf{X},\mathbf{Y}$.

From the definition of mutual information, we would expect that there is a close relationship between mutual information and correlation. In fact we have the following conclusions. If X and Y are normally distributed random variables, then

$$I(X,Y) = -\frac{1}{2} \log(1-\rho^2)$$

where ρ is the correlation coefficient between X and Y.

From recorded neuronal data, to calculate the mutual information between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is usually not an easy task when one of them is a random vector in a high dimensional space. To estimate the joint distribution of $\mathbf{X},\mathbf{Y}$ from data is already a formidable task in a high dimensional space. See Chapter 13 for a detailed account on how how to overcome the difficulties.

1.4.3 Fisher information

The Fisher information is introduced from an angle totally different from the Shannon information. For a random variable with distribution density $p(x;\theta)$, the Fisher information is

$$I(\theta) = \int \left(\frac{\partial p(\mathbf{x};\theta)/\partial\theta}{p(\mathbf{x};\theta)} \right)^2 p(\mathbf{x};\theta) d\mathbf{x}$$

where θ is the parameter which could be multi-dimensional.

Example 1 Let us assume that

$$T \sim E[T]\exp(-t/E[T]), \qquad t \geq 0$$

where T is the interspike interval and $E[T]$ is the expectation of T. Suppose that $E[T]$ depends on a parameter λ. The Fisher information with respect to λ [45] is defined by

$$\begin{aligned} I(\lambda) &= \frac{1}{E[T]}\int_0^\infty \left(\frac{\partial \log p}{\partial \lambda}\right)^2 \exp\left(-\frac{t}{E[T]}\right) dt \\ &= \frac{1}{E[T]}\int \left(\frac{(E[T])'}{E[T]} - \frac{(E[T])'t}{(E[T])^2}\right)^2 \exp\left(-\frac{t}{E[T]}\right) dt \\ &= \frac{[(E[T])']^2}{[E[T]]^2} \end{aligned} \tag{1.14}$$

where $(E[T])'$ is the derivative with respect to the parameter λ.

For a Poisson process with $E[T] = 1/\lambda$, we have $I = 1/\lambda^2 = (E[T])^2$. The larger $E[T]$ is, the larger the Fisher information. Of course, when $E[T]$ is a nonlinear function λ, we see that $I(\lambda) = 0$ whenever λ satisfies $(E[T])' = 0$.

The Fisher information is useful since it is related to the variance of an estimate δ of $g(\lambda)$. Assume that we have

$$E[\delta] = g(\lambda) + B(\lambda)$$

where $B(\lambda)$ is the bias of the estimate δ. We then have the following information inequality

$$\mathrm{Var}(\delta) \geq \frac{[g'(\lambda) + B'(\lambda)]^2}{I(\lambda)} \tag{1.15}$$

The information inequality is often called Cramer-Rao inequality. However, it seems the inequality was first discovered by Frechet (1943), and then rediscovered or extended by Darmois (1945), Rao(1945) and Cramer (1946) ([45], page 143). The information inequality gives us the lowest bound for an estimate δ can attain.

Coming back to the example above, from Equation (1.14) we see that the Fisher information is zero whenever $(E[T])'$ is zero. In other words, when $E[T]$ reaches its maximum or minimum points, depending on the parameter λ, the Fisher information vanishes. This leaves us a question of how to estimate the values of λ with $(E[T])' = 0$. In fact, these values could be most interesting since neurons fire with their highest or lowest rate. In [17], we have discussed in details on how to estimate (decode) the input information when its Fisher information vanishes.

In the framework of maximum likelihood estimate, the Fisher information asymptotically (when the sample sizes are large enough) gives the confidence interval of an estimate. In Chapter 9 the authors explore the applications of the maximum likelihood estimates to neuronal spike train data.

1.4.4 Relationship between the various measurements of information

After introducing various measures of information, we are interested in knowing the relationship between them [37, 45].

Let $X_1, X_2, \cdots, X_n$ be an identically and independently distributed samples from a density $f(x, \theta)$ and let $S_n(\pi)$ be the Shannon information of the sample, i.e.,

$$S_n(\pi) = E\left[\int \log \frac{\pi(\theta|X_1, \cdots, X_n)}{\pi(\theta)} d\pi(\theta|X_1, \cdots, X_n)\right],$$

where $\pi(\theta)$ is a prior distribution. Then, as n is large enough, we have

$$S_n(\pi) \sim \frac{k}{2}\log\frac{n}{2\pi e} + \int \pi(\theta)\log\frac{|I(\theta)|^{1/2}}{\pi(\theta)}d\theta$$

Hence the optimal prior is the Jeffreys prior which is proportional to $\sqrt{I(\theta)}$.

1.5 Optimal control

Cortical activity related to some simple motor movements might be relatively easier to characterize than high brain functions such as memory and attention etc. [25, 39, 40, 58]. To understand biological movement control is of great potential applications in robot control, as reviewed in Chapter 17. Here we present some examples to illustrate optimal control theory and refer the reader to Chapter 17 for more details.

In theory, (stochastic) optimal control is a well developed area, with wide and successful applications in finance. In general, to find an optimal control signal is reduced to solve the Hamilton-Jacobi-Bellman (HJB) equation [52]. However, the HJB equation is usually difficult to solve, even numerically [43]. In the simplest case, i.e., when the control problem is an open loop control, we can analytically obtain the solution of the control problem (see [71] for some recent results with a feedback control).

1.5.1 Optimal control of movement

The Model

We consider a simple model of saccadic movement. Let $x_1(t)$ be the position of eye (in degrees) and $x_2(t)$ be its velocity (degree/sec) [24]. We then have

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\dfrac{1}{\tau_1\tau_2}x_1 - \dfrac{\tau_1+\tau_2}{\tau_1\tau_2}x_2 + \dfrac{1}{\tau_1\tau_2}\bar{u} \end{cases} \tag{1.16}$$

where τ_1, τ_2 are parameters and $\bar{u}$ is the input signal as defined below. However, we are more interested in general principles rather than numerically fitting of experimental data. From now on, we assume that all parameters are in arbitrary units, although a fitting to biological data would be straightforward. In matrix term we have

$$d\mathbf{X} = A\mathbf{X}dt + d\mathbf{U} \tag{1.17}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -\dfrac{1}{\tau_1\tau_2} & -\dfrac{\tau_1+\tau_2}{\tau_1\tau_2} \end{pmatrix} \tag{1.18}$$

and

$$d\mathbf{U} = \left(\begin{array}{c} 0 \\ \dfrac{\lambda(t)dt + \lambda(t)^{\alpha} dW_t}{\tau_1 \tau_2} \end{array} \right)$$

with W_t being Brownian motion, $\lambda(t)$ is the control signal, and $\alpha > 0$ (see Section 1.3). Basically, under the rate coding assumption, $\alpha = 1/2$ corresponds to Poisson inputs [57, 77] (also see Section 1.3). We call $\alpha < 1/2$ *sub-Poisson inputs*, and $\alpha > 1/2$ *supra-Poisson inputs*.

Here is the control problem.

- For a fixed, deterministic time T, find $\lambda(s) \in \mathscr{L}^{2\alpha}[0, T+R]$ which minimizes

$$\int_T^{T+R} \text{var}(x_1(t))dt \tag{1.19}$$

subject to the constraint

$$E[x_1(T)] = D \text{ for } t \in [T, T+R] \tag{1.20}$$

The meaning is clear. Equation (1.20) ensures that at time T, the saccadic movement stops at the position D and keeps there for a while, i.e., from time T to $T+R$. During the time interval $[T, T+R]$, it is required that the fluctuation is as small as possible (Equation (1.19)).

When $\alpha > 1/2$, define

$$\begin{cases} b_{12}(t-s) = \dfrac{\tau_1\tau_2}{\tau_2 - \tau_1}\left[\exp\left(-\dfrac{1}{\tau_2}(t-s)\right) - \exp\left(-\dfrac{1}{\tau_1}(t-s)\right)\right] \\ b_{22}(t-s) = \dfrac{\tau_1\tau_2}{\tau_2 - \tau_1}\left[\dfrac{1}{\tau_1}\exp\left(-\dfrac{1}{\tau_1}(t-s)\right) - \dfrac{1}{\tau_2}\exp\left(-\dfrac{1}{\tau_2}(t-s)\right)\right], \end{cases}$$

so the solution of the optimal control problem is

$$\lambda(s) = \frac{\left|\mu_1 \exp\left(\dfrac{s}{\tau_1}\right) + \mu_2 \exp\left(\dfrac{s}{\tau_2}\right)\right|^{1/(2\alpha-1)} \text{sgn}\left[\mu_1 \exp\left(\dfrac{s}{\tau_1}\right) + \mu_2 \exp\left(\dfrac{s}{\tau_2}\right)\right]}{\left(\displaystyle\int_0^{T+R} b_{12}^2(t-s)dt\right)^{1/(2\alpha-1)}} \tag{1.21}$$

with μ_1, μ_2 being given by

$$\begin{cases} D\tau_2 \exp\left(\dfrac{T}{\tau_2}\right) = \displaystyle\int_0^T \exp\left(\dfrac{s}{\tau_2}\right) \cdot \dfrac{A(s)}{B(s)} ds \\ D\tau_1 \exp\left(\dfrac{T}{\tau_1}\right) = \displaystyle\int_0^T \exp\left(\dfrac{s}{\tau_1}\right) \cdot \dfrac{A(s)}{B(s)} ds. \end{cases} \tag{1.22}$$

with

$$A(s) = \left|\mu_1 \exp\left(\frac{s}{\tau_1}\right) + \mu_2 \exp\left(\frac{s}{\tau_2}\right)\right|^{1/(2\alpha-1)} \text{sgn}\left[\mu_1 \exp\left(\frac{s}{\tau_1}\right) + \mu_2 \exp\left(\frac{s}{\tau_2}\right)\right]$$

and

$$B(s) = \left(\int_0^{T+R} b_{12}^2(t-s)dt \right)^{1/(2\alpha-1)}.$$

The control problem with $\alpha = 1$ discussed in this subsection is first proposed in [33], see also [64]. Their numerical results show an excellent agreement with experimental data. The basic principle: *task optimization in the presence of signal-dependent noise*, is intriguing and possibly functional meaningful in motor control. Unfortunately, the results are not degenerate only when the inputs are supra-Poisson. To illustrate the point, we next consider a simpler control problem.

1.5.2 Optimal control of single neuron

Controlling neuron activity is another interesting area of research, with possible applications in medicine such as in the treatment of Parkinson's disease [8].

The Model

The neuron model we use here is the classical integrate-and-fire model [9, 77], as in Section 1.3. When the membrane potential V is below the threshold V_{thre}, it is given by

$$dV = -\frac{V_t - V_{rest}}{\gamma}dt + dI_{syn}(t) \tag{1.23}$$

where the synaptic input is

$$I_{syn}(t) = a_E \sum_{i=1}^{n_E} E_i(t) - a_I \sum_{j=1}^{n_I} I_j(t)$$

with $E_i(t), I_i(t)$ as point processes (Poisson process being a special case), $a_E > 0, a_I > 0$ are magnitude of each EPSP and IPSP, n_E and n_I are the total number of active excitatory and inhibitory synapses. Once V crosses V_{thre} from below a spike is generated and V is reset to V_{rest}, the resting potential. The interspike interval of efferent spikes is

$$T = \inf\{t : V(t) \geq V_{thre}\}$$

In the following developments, we further assume that $V_{rest} = 0, a_I = 0$ and use diffusion approximations to approximate synaptic inputs [77], as in Section 1.3

$$di_{syn}(t) = a_E \lambda(t)dt + a_E \lambda^{\alpha}(t) \cdot dW_t$$

where W_t is a standard Brownian motion and $\alpha > 0$. When $\alpha = 1/2$, the input is a Poisson process. The larger the α is, the more randomness the synaptic inputs are.

For a fixed time T_f, let us define

$$I(\lambda) = \text{var}(V(T_f)),$$

i.e., $I(\lambda)$ is the variance at the end-point T_f of the membrane potential with the input signal $\lambda(t)$. Then we have the following.

- Control problem:

  To find a synaptic input $\lambda(s)$ satisfying

$$E[V(T_f)] = V_{thre}, \tag{1.24}$$

$$I(\lambda^*) = \min_{\lambda} I(\lambda). \tag{1.25}$$

The meaning of the optimal control problem is as follows. Suppose that we intend to drive a neuron to fire with a fixed frequency, say $1000/T_f$ Hz so that we can fix the time T_f. Equation (1.24) satisfies the requirement. The second requirement Equation (1.25) indicates the we intend to determine an optimal (control) signal so that the variance of the membrane potential at time T_f attains its minimum value, among all possible control signals. Here all possible control signals $(\lambda(t))$ mean all possible positive function of time t. The more difficult mathematically and indeed more realistic problem is to insist that with a stochastic input

$$E[T] = T_f$$

and seek to minimize $I_1(\lambda) = \text{Var}[T]$. Although minimizing the variance of the membrane potential is not the same as minimizing the variance of the interspike interval, it is likely for most physiologically realistic inputs that the relationship between them is monotonic. Hence we proceed on the reasonable assumption that when the variance of the membrane potential reaches its minimum value, the corresponding variance of interspike intervals attains its minimum as well.

Optimal control

We have the following conclusions

- For $\alpha > 1/2$, the unique optimal control signal $\lambda(s)$ is

$$\lambda(s) = \frac{(2\alpha-2)V_{thre}}{(2\alpha-1)a\gamma\left[1-\exp\left(-\frac{T_f(2\alpha-2)}{\gamma(2\alpha-1)}\right)\right]} \cdot \exp\left(\frac{T_f - s}{(2\alpha-1)\gamma}\right) \tag{1.26}$$

  for $\quad 0 \le s \le T_f$. In particular, when $\alpha = 1$, we have

$$\lambda(s) = \frac{V_{thre}}{aT_f}\exp\left(\frac{T_f - s}{\gamma}\right). \tag{1.27}$$

- For $\alpha = 1/2$, the unique optimal control signal $\lambda(s) = \delta_0(s)$, the delta function at time zero.

- For $\alpha < 1/2$, the optimal control signal $\lambda(s) = \delta_y(s)$, the delta function at $y \in [0, T_f]$. Hence the solution is not unique.

As a direct consequence of the results above, we have

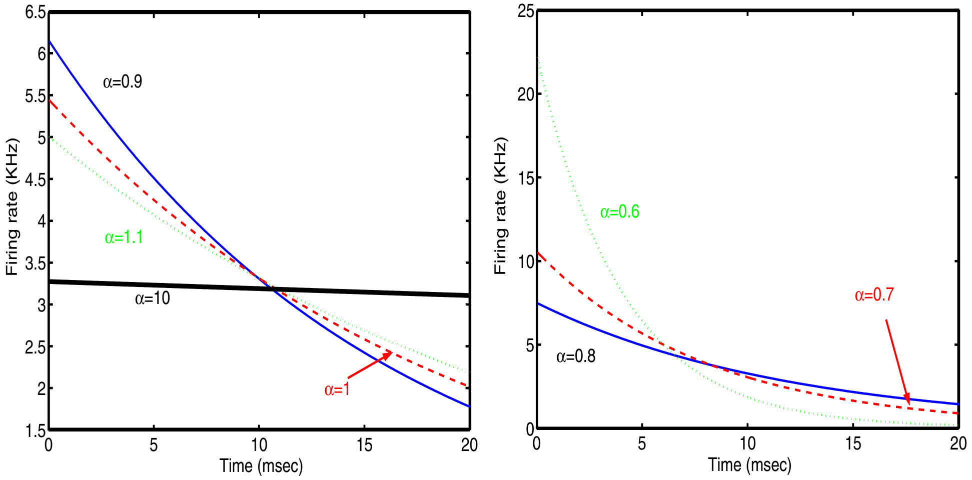

Figure 1.1

Optimal control signals ($\lambda^*(t)$) with $\alpha = 0.6, 0.7, \cdots, 1.2, 10$. $T_f = 20$ msec.

- For $\alpha > 1/2$,

$$I(\lambda^*(s)) = a^{2-2\alpha} V_{thre}^{2\alpha} \left| \frac{2\alpha - 2}{(2\alpha - 1)\gamma} \right|^{2\alpha - 1} \cdot \left| \left(1 - \exp\left(-\frac{T_f(2\alpha - 2)}{\gamma(2\alpha - 1)} \right) \right) \right|^{1-2\alpha}. \tag{1.28}$$

In particular, when $\alpha = 1$, we have

$$I(\lambda^*(s)) = \frac{V_{thre}^2}{T_f}. \tag{1.29}$$

- For $\alpha = 1/2$,

$$I(\lambda^*) = aV_{thre} \exp\left(-\frac{T_f}{\gamma} \right). \tag{1.30}$$

- For $\alpha < 1/2$,

$$I(\lambda^*) = 0. \tag{1.31}$$

Having found the optimal control signals, we can further discuss how to implement them on a neuronal basis [22].

In conclusion, we find that *task optimization in the presence of signal-dependent noise* is promising in applications to motor and neuron control problems [7]. However, the optimal problem turns out to be degenerate with Poisson and sub-Poisson inputs. With a biologically reasonable feedback (for example, the reversal potentials are the natural form of feedback), we would expect that the optimal problem becomes not degenerate.

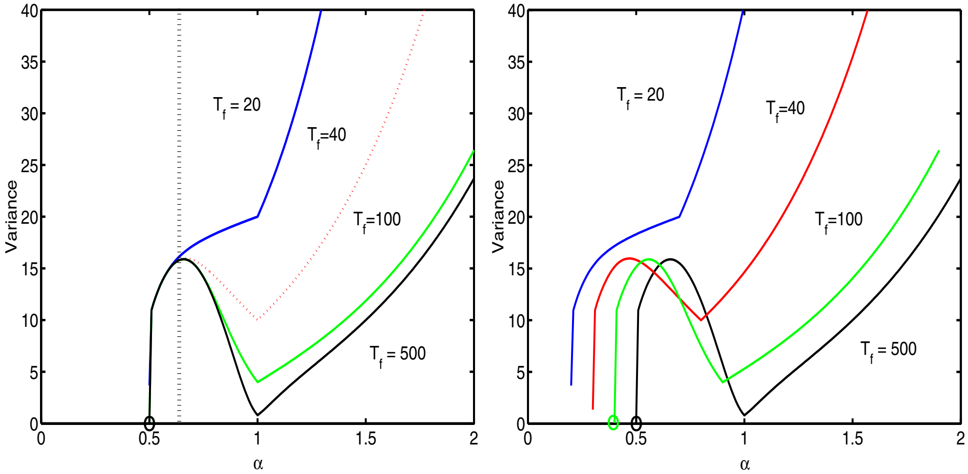

Figure 1.2

Optimal variance $I(\lambda^*)$ against α for $T_f = 20, 40, 100, 500$. Right is the same as left, but $T_f = 20$ is shifted towards left with 0.3 units, $T_f = 40$ with 0.2 units, and $T_f = 100$ with 0.1 units.

Acknowledgments: JF was partially supported by grants from EPSRC (GR/ R54569), (GR/ S30443), (GR/ S20574) and (GR/ S63830) (U.K.), a grant of the Royal Society and an exchange grant between U.K. and China of the Royal Society.

References

[1] Abeles, M. (1990). *Corticonics*, Cambridge Univ. Press: Cambridge, UK.

[2] Albeverio, S., Feng, J. F., and Qian, M. (1995). Role of noise in neural networks. *Phys. Rev. E.* **52**: 6593-6606.

[3] Albright, T.D., Jessell, T.M., Kandel, E.R., and Posner, M.I. (2000). Neural science: a century of progress and the mysteries that remain. *Cell* **100**: s1-s55.

[4] Applebaum, D. (1996). *Probability and Information*. Cambridge University Press: Cambridge, UK.

[5] Arnold, L. (1998). *Random dynamical systems.* Springer-Verlag, Berlin.

[6] Azouzl, R., and Gray, C.M. (2003). Adaptive coincidence detection and dynamic gain control in visual cortical neurons *in vivo*. *Neuron*, in press.

[7] van Beers, R. J., Baraduc, P., and Wolpert, D.M.(2002). Role of uncertainty in

sensorimotor control. *Philos. T. Roy. Soc.* **357**: 1137-1145.

[8] Benabid A.L., Pollak, P., Gervason, C., Hoffmann, D., Gao, D.M., Hommel, M., Perret, J.E., and Derougemont, J. (1991). Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. *Lancet* **337**: 403-406.

[9] Brown, D., Feng, J.F., and Feerick, S.(1999). Variability of firing of Hodgkin-Huxley and FitzHugh-Nagumo neurons with stochastic synaptic input. *Phys. Rev. Lett.* **82**: 4731-4734.

[10] Cope, D.K., and Tuckwell, H.C. (1979). Firing rates of neurons with random excitation and inhibition. *J. Theor. Biol.* **80**: 1-14.

[11] Cover, T.M., and Tomas, J.A. (1991). *Elements of Information Theory*. New York: Wiley.

[12] Davison, A., Feng, J. F., and Brown, D. (2000). A reduced compartmental model of the mitral cell for use in network models of the olfactory bulb. *Brain Research Bulletin* **51**: 393-399.

[13] Dayan, P., and Abbott, L. (2002). *Theoretical Neuroscience*. MIT press: Cambridge, Massachusetts.

[14] E, W.N., Ren, W.Q., Vanden-Eijnden, E. (2002). String method for the study of rare events. *Phys. Rev. B* **66**: 052301.

[15] Ermentrout, G.B., and Kleinfeld, D. (2001). Travelling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron* **29**: 33-44.

[16] Feng, J.F. (1997). Lyapunov functions for neural nets with nondifferentiable input-output characteristics. *Neural Computation* **9**: 45-51.

[17] Feng, J.F. (2001). Optimally decoding the input rate from an observation of the interspike intervals. *J. Phys. A.* **34**: 7475-7492.

[18] Feng, J.F. (2001). Is the integrate-and-fire model good enough? – a review. *Neural Networks* **14**: 955-975.

[19] Feng, J.F. (2003). Effects of correlated and synchronized stochastic inputs to leaky integrator neuronal model. *J. Theore. Biol*: in press.

[20] Feng, J.F., Brown, D., and Li, G. (2000). Synchronization due to common pulsed input in Stein's model. *Phys. Rev. E*, **61**: 2987-2995.

[21] Feng, J.F., and Li, G. (2002). Impact of geometrical structures on the output of neuronal models – a theoretical and numerical analysis. *Neural Computation* **14**: 621-640.

[22] Feng, J.F., and Tuckwell, H.C. (2003). Optimal control of neuronal activity *Phys. Rev. Letts.* (in press).

[23] Feng, J.F., and Zhang, P. (2001). The behaviour of integrate-and-fire and

Hodgkin-Huxley models with correlated input. *Phys. Rev. E.* **63**: 051902.

[24] Feng, J.F., Zhang, K. W., and Wei, G. (2002). Towards a mathematical foundation of minimum-variance theory. *J. Phs. A* **35**, 7287-7304.

[25] Flash, T., and Sejnowski, T.J. (2001). Computational approaches to motor control. *Curr. Opin. Neurobiol.* **11** : 655-662.

[26] Fleming, W.H., and Rishel, R.W. (1975). *Deterministic and Stochastic Optimal Control* Springer-Verlag.

[27] Freidlin, M.I., and Wentzell, A.D. (1984) *Random Perturbations of Dynamical Systems*, Springer-Verlag: New York.

[28] Gerstein, G.l. and Mandelbrot, B. (1964). Random walk models for the spike activity of single neurons. *Biophys. J.* **4**, 41-68.

[29] Gerstner, W., and Kistler, W. (2002) *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press: Cambridge.

[30] Gihman, I.I., and Skorohod, A.V. (1972). *Stochastic Differential Equations.* Springer-Verlag, Berlin.

[31] Gluss, B. (1967). A model for neuron firing with exponential decay of potential resulting in diffusion equations for probability density. *Bull. Math. Biophys.* **29**: 233-243.

[32] Grimmett, G., and Stirzaker, D. (2001) *Probability and Random Processes*, Third Edition, Oxford University Press: Oxford, UK.

[33] Harris, C.M., and Wolpert, D.M. (1998). Signal-dependent noise determines motor planning. *Nature* **394**: 780-784.

[34] Metzner, W., Koch, C., Wessel, R., and Gabbiani, F. (1998). Feature extraction by burst-like spike patterns in multiple sensory maps. *J. Neuroscience* **18** 2283-2300.

[35] Itô, K. (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.*, Volume **4**.

[36] Kandel, E.R., Schwartz, J.J., and Jessell, T.M. (1991). *Principles of Neural Science*, 3rd Edition, Prentice-Hall International Inc.

[37] Kang, K., and Sompolinsky, H. (2001). Mutual information of population codes and distance measures in probability space. *Phys. Rev. Letts.* **86**: 4958-4961.

[38] Karlin, S., and Taylor, H.M. (1982) *A Second Course in Stochastic Processes* Academic Press, New York.

[39] Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology* **9**: 718-727.

[40] Kitazawa, S. (2002). Optimization of goal-directed movements in the cerebel-

lum: a random walk hypothesis. *Neurosci. Res.* **43**: 289-294.

[41] Koch, C. (1999) *Biophysics of Computation*, Oxford University Press.

[42] Kuramoto, Y. (1984) *Chemical Oscillations, Waves and Turbulence* Springer Verlag: New York.

[43] Kushner, H.J. (1999). Consistency issues for numerical methods for variance control, with applications to optimization in finance. *IEEE T. Automat. Contr.* **44**: 2283-2296.

[44] Lansky, P., and Sacerdote, L. (2001). The Ornstein-Uhlenbeck neuronal model with signal-dependent noise. *Phys. Lett. A* **285**: 132-140.

[45] Lehmann, E. and Casella G. (1999) *Theory of Point Estimation,* Springer-Verlag: New York, Berlin etc.

[46] Lisman, J.E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neuroscience* **20**: 38-43.

[47] Liu, R.C., and Brown, L.D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Ann. Statis.* **21**: 1-13.

[48] Mainen, Z.F., and Sejnowski, T.J. (1996). Influence of dendritic structure on firing pattern in model neocortical neuron. *Nature* **382**: 363-366.

[49] Mainen, Z.F., and Sejnowski, T.J. (1995). Reliability of spike timing in neocortical neurons. *Science* **268**: 1503-1506.

[50] Mazurek, M.E., and Shadlen, M.N. (2002). Limits to the temporal fidelity of cortical spike rate signals. *Nat. Neurosci.* **5**: 463-471.

[51] Mehta, M.R., Lee, A.K., and Wilson, M.A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, **417**: 741-746.

[52] Oksendal, B. (1989) *Stochastic Differential Equations* (second edition), Springer-Verlag: Berlin.

[53] Ott, E. (2002) *Chaos in Dynamical Systems* 2nd Edition. Cambridge University Press: Cambridge, UK.

[54] Ott, E., So, P., Barreto, E., et al. (2002). The onset of synchronization in systems of globally coupled chaotic and periodic oscillators. *Physica D***173**: 29-51.

[55] Pikovsky, A.S. (1992). Statistics of trajectory separation in noisy dynamic-systems. *Phys Lett A.* **165**: 33-36.

[56] Pinsky, P.F., and Rinzel, J. (1994). Intrinsic and network rhythmogensis in a reduced Traub model for CA3 neurons. *J. Computational Neuroscience* **1**: 39-60.

[57] Ricciardi, L.M., and Sato, S. (1990), Diffusion process and first-passage-times

problems. *Lectures in Applied Mathematics and Informatics,* ed. Ricciardi, L.M., Manchester: Manchester University Press.

[58] Richardson, M.J.E., and Flash, T. (2002). Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *J. Neurosci.* **22**: 8201-8211.

[59] Rieke, F., Warland, D., de Ruyter van Steveninch, R., and Bialek, W. (1997). *Spikes: Exploring The Neural Code*. The MIT Press, Cambridge, Massachusetts.

[60] Risken, S. (1989). *The Fokker-Planck Equation*. Springer-Verlag: Berlin.

[61] Rodriguez, R. and Tuckwell, H.C. (1996). Statistical properties of stochastic nonlinear dynamical models of single neurons and neural networks. *Phys. Rev. E.* **54**: 5585-5590.

[62] Rodriguez, R., and Lansky, P. (2000). Effect of spatial extension on noise-enhanced phase locking in a leaky integrate-and-fire model of a neuron. *Phys Rev E* **62**: 8427-8437.

[63] Roy, B.K. and Smith, D.R. (1969). Analysis of the exponential decay model of the neuron showing frequency threshold effects. *Bull. Math. Biophys.* **31**: 341-357.

[64] Sejnowski, T.J. (1998). Making smooth moves, *Nature*, **394**: 725-726.

[65] Shadlen, M.N., and Newsome, W.T. (1994). Noise, neural codes and cortical organization, *Curr. Opin. Neurobiol.* **4**: 569-579.

[66] Shepherd, G. (1994) *Neurobiology* Third Ed., Oxford University Press: Oxford, UK.

[67] van Steveninck, R.R.D., and Laughlin, S.B. (1996). The rate of information transfer at graded-potential synapses. *Nature*, **379**: 642-645.

[68] Stevens, C.F., and Zador, A.M. (1998). Input synchrony and the irregular firing of cortical neurons. *Nat. Neurosci.* **1**: 210-217.

[69] Strogatz, S.H. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* **143**: 1-20.

[70] Thompson, A.M. (1997). Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons *in vitro*. *J. Physiol.* **502**: 131-147.

[71] Todorov, E. (2002). Cosine tuning minimizes motor errors. *Neural Computation* **14**: 1233-1260.

[72] Tuckwell, H.C. (1975). Determination of the interspike times of neurons receiving randomly arriving postsynaptic potentials. *Biol. Cybernetics* **18**: 225-237.

[73] Tuckwell, H.C. (1979). Synaptic transmission in a model for stochastic neural activity. *J. Theor. Biol.* **77**: 65-81.

[74] Tuckwell, H.C. (1981). Poisson processes in biology. *Stochastic Nonlinear Systems*, Springer, Berlin, pp 162-171.

[75] Tuckwell, H.C. (1987a). Diffusion approximations to channel noise. *J. Theor. Biol.* **127**: 427-438.

[76] Tuckwell, H.C. (1987b). Statistical properties of perturbative nonlinear random diffusion from stochastic integral representations. *Phys. Lett. A* **122**: 117-120.

[77] Tuckwell, H.C. (1988a). *Introduction to Theoretical Neurobiology, Vol 2: Stochastic and Nonlinear Theories*. Cambridge University Press, New York.

[78] Tuckwell, H.C. (1988b). Perturbative analysis of random nonlinear reaction-diffusion systems. *Physica Scripta* **37**: 321-322.

[79] Tuckwell, H.C. (1989). *Stochastic Processes in the Neurosciences*. SIAM, Philadelphia.

[80] Tuckwell, H.C. (1992), Random fluctuations at an equilibrium of a nonlinear reaction-diffusion equation. *Applied Math. Letters* **6**: 79-81.

[81] Tuckwell, H.C., and Cope, D.K. (1980), The accuracy of neuronal interspike times calculated from a diffusion approximation. *J. Theor. Biol.* **83**: 377-387.

[82] Tuckwell, H.C., and Rodriguez, R. (1998). Analytical and simulation results for stochastic Fitzhugh-Nagumo neurons and neural networks. *J. Computational Neuroscience* **5:** 91-113.

[83] Tuckwell, H.C., Wan, F.Y.M., and Rospars, J-P. (2002). A spatial stochastic neuronal model with Ornstein-Uhlenbeck input current. *Biol. Cybernetics* **86**: 137-145.

[84] Tuckwell, H.C., Wan, F.Y.M., and Wong, Y.S. (1984). The interspike interval of a cable model neuron with white noise input. *Biol. Cybern.* **155**: 155-167.

[85] Verhulst, F. (1990) *Nonlinear Differential Equations and Dynamical Systems*, Springer-Verlag: Berlin, Heidelberg et al.

[86] van Vreeswijk, C., Abbott, L.F., and Ermentrout, G.B. (1994). When inhibition not excitation synchronizes neural firing. *J. Comp. Neurosci.* **4**: 313-321.

[87] Wan, F.Y.M, and Tuckwell, H.C. (1982). Neuronal firing and input variability. *J. Theoret. Neurobiol.* **1**: 197-218.

[88] Williams, S.R., Toth, T.I., Turner, J.P., Hughes, S.W., and Crunelli, V. (1997). The 'window' component of the low threshold Ca2+ current produces input signal amplification and bistability in cat and rat thalamocortical neurones. *J. Physio.-London* **505**: 689-705.

[89] Wolpert, D.M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neuroscience* **3**:1212-1217

[90] Zohary, E., Shadlen, M.N., and Newsome, W.T. (1994). Correlated neuronal

discharge rate and its implications for psychophysical performance. *Nature* **370**: 140-143.

Appendix A: The Hodgkin-Huxley Model. The remaining equations in the Hodgkin-Huxley model are as following (see [9, 13]).

$$\frac{dn}{dt} = \frac{n_\infty - n}{\tau_n}, \qquad \frac{dm}{dt} = \frac{m_\infty - m}{\tau_m}, \qquad \frac{dh}{dt} = \frac{h_\infty - h}{\tau_h}$$

and

$$\begin{aligned} n_\infty &= \frac{\alpha_n}{\alpha_n + \beta_n}, & m_\infty &= \frac{\alpha_m}{\alpha_m + \beta_m}, & h_\infty &= \frac{\alpha_h}{\alpha_h + \beta_h} \\ \tau_n &= \frac{1}{\alpha_n + \beta_n}, & \tau_m &= \frac{1}{\alpha_m + \beta_m}, & \tau_h &= \frac{1}{\alpha_h + \beta_h} \end{aligned}$$

with

$$\begin{aligned} \alpha_n &= \frac{0.01(V+55)}{1 - \exp(-\frac{V+55}{10})} & \beta_n &= 0.125 \exp(-\frac{V+65}{80}) \\ \alpha_m &= \frac{0.1(V+40)}{1 - \exp(-\frac{V+40}{10})} & \beta_m &= 4 \exp(-\frac{V+65}{18}) \\ \alpha_h &= 0.07 \exp(-\frac{V+65}{20}) & \beta_h &= \frac{1}{\exp(-\frac{V+35}{10}) + 1} \end{aligned}$$

The parameters used in Equation (1.2) are $C = 1, g_{Na} = 120, g_K = 36, g_L = 0.3, V_k = -77, V_{Na} = 50, V_L = -54.4$.

Chapter 2

Atomistic Simulations of Ion Channels

Peter D. Tieleman
Dept. of Biological Sciences, University of Calgary, Alberta, Canada T2N 1N4

CONTENTS

Abbreviations: Alm, alamethicin; BD, Brownian dynamics; gA, gramicidin A; KcsA, K channel from Streptomyces lividans; M2δ, synthetic M2 peptide from nAChR M2 helix of δ-subunit; MC, Monte Carlo; MD, molecular dynamics; nAChR, nicotinic acetylcholine receptor; PB, Poisson-Boltzmann; PNP, Poisson-Nernst-Planck

2.1 Introduction

2.1.1 Scope of this chapter

This chapter deals with computer simulations of ion channel proteins. I review briefly the key properties of ion channels for which there is structural information, and discuss different simulation approaches. In my opinion, the general goal of simulations of ion channels could be stated as follows:

The goal of simulation studies is to link the structure of ion channels in atomic detail to ion currents, including a fundamental understanding of how ion channels are controlled by external factors such as voltage, ligand binding, ion gradients, pH, and interactions with toxins and other blockers.

Simulations can be used to study ion channels at different levels of detail. At the highest level of detail, all individual atoms are included in the simulation, but a connection to macroscopic observables like voltage-current relationships is hard to make. At intermediate levels of detail, protein and ions are treated as individual particles but solvent is not. At this level a direct link between simulation and experiment is now feasible. Finally, at the lowest level of detail, both ions and solvent are treated implicitly, as mean fields, and the channel itself may be either atomic detail or simplified as well. At this level of approximation it is computationally easy to calculate macroscopic properties, but it is challenging to obtain accurate results. Recently, simulations have also begun to study interactions of toxins and other small molecules with ion channels, providing an additional possible link with experimental work. I conclude with a review of selected recent applications to a number of model ion channels and give a brief outlook of how simulations might contribute in the future to a better understanding of how ion channels work.

2.1.2 Ion channels

Ion channels are found in all organisms, in a staggering variety. In higher organisms, they play an important role in e.g. cell excitability [58], maintaining and regulating osmotic balance, and signal transduction. Ion channels in excitable cells occur in a wide variety and are usually grouped based on the ions they primarily conduct: e.g., sodium channels, potassium channels, or calcium channels. An alternative classification considers their control mechanism: voltage-gated, ligand-gated, cyclic-nucleotide gated etc. Many ion channels are involved in diseases termed 'channelopathies', in which ion channels are defective in some way [8]. Because of the central role of many ion channels in cell physiology, ion channels are also major targets for drugs and drug development.

The basic function of ion channels is simple: they conduct ions. Broadly speaking, ion channels have three fundamental properties:

1. *Conductance*, or current-voltage relationships at specific conditions. These relationships can be almost linear, but they can also show features such as saturation and rectification (preferential current in one direction), and they can show very complex behaviour in the presence of multiple types of ions. Many ion channels conduct ions close to the maximum rate allowed by diffusion.

2. *Selectivity*. Most ion channels preferentially conduct one type of ion over another. Selectivity is an intriguing property. The difference between potassium

and sodium in radius and electronic properties is small, but certain potassium channels are more than a 1000-fold selective for potassium over sodium. 'Valence selectivity' refers to selectivity for e.g., divalent ions such as calcium over sodium or potassium, and 'charge selectivity' is selectivity for cations over anions or vice versa. Most specific cation channels do not conduct chloride, whereas chloride channels do not conduct cations. Others, such as the nicotinic acetylcholine receptor, are only somewhat selective.

3. *Gating*. The ion channels of excitable membranes are controlled by their environment, including voltage, ion concentrations, and ligand-binding. The precise mechanism of gating is not yet known, although there is much evidence for voltage-gated channels that points to specific parts of the sequence, and the crystal structure of a large part (although not the voltage-sensitive part) of a voltage-gated potassium channel is known [54].

Bacteria contain homologues of many of the channels from higher organisms. Ion channel activity (conductance of ions) is also exhibited by a wide variety of other proteins that are not normally considered ion channels, such as bacterial and mitochondrial porins, gap junctions, bacterial toxins such as alpha-haemolysin, and ion channels formed by the aggregation of peptides such as gramicidin, alamethicin, mellitin, and several designed peptides. In addition to these 'natural' channels, many organic and inorganic compounds have been designed that form supra-molecular aggregates with clear ion channel activity. Although neither these artificial ion channels nor pore forming toxins etc. are directly relevant for ion channels in the nervous system, they form important model systems for theoretical understanding of the molecular basis of the properties of these physiologically more relevant ion channels, primarily because they are simpler.

Until recently, the molecular basis of selectivity, the high conductance of some channels, and of gating was inferred indirectly from electrophysiology and other experiments. Experimentally, the channels can be studied at the single-molecule level by the techniques of electrophysiology, often combined with site-directed mutagenesis and the use of specific blockers [8, 58]. For the purpose of simulations, more detailed structural information is usually required. In recent years, our understanding of the structural reasons underlying ion channel properties has increased significantly. Currently the high-resolution crystal structure of two potassium channels, a chloride channel, and two mechano-sensitive channels are known. An overview of some of the solved structures is given in Figure 2.1.

Potassium channels vary in complexity but are generally tetrameric with 2-6 transmembrane helices per monomer. The known structures are of bacterial homologues. KcsA is a tetrameric channel with 2 transmembrane helices per unit. Figure 2.2 shows the topology of KcsA. The pore-lining segment of the helices M1 and M2 (called S5 and S6 in more complex channels) and the 'pore helix' P is conserved across potassium channels as well as most likely in sodium and calcium channels. In the crystal structures of KcsA, 20-30 residues of each of the terminal sequences are missing [40].

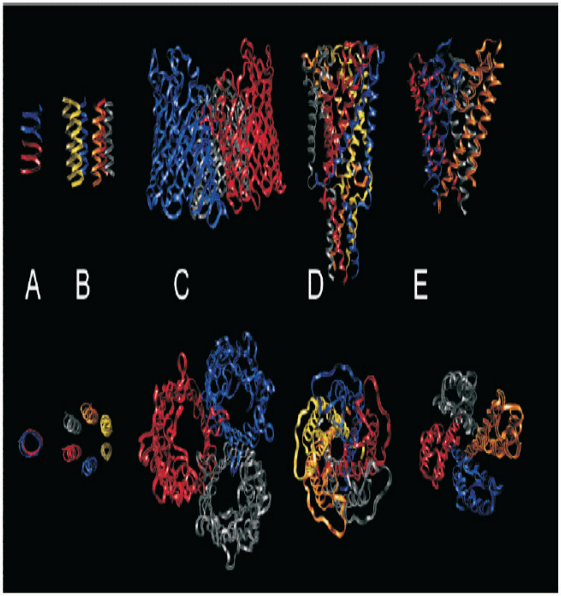

Figure 2.1

The structure of a number of ion channels: A. Gramicidin A; B. Model of an alamethicin channel with six helices; C. OmpF porin; D. The large conductance mechanosensitive channel MscL; E. The potassium channel KcsA. In all cases, different monomers have different colors. Figures 2.1, 2.2, 2.4, 2.6, 2.8 were made with the program VMD.

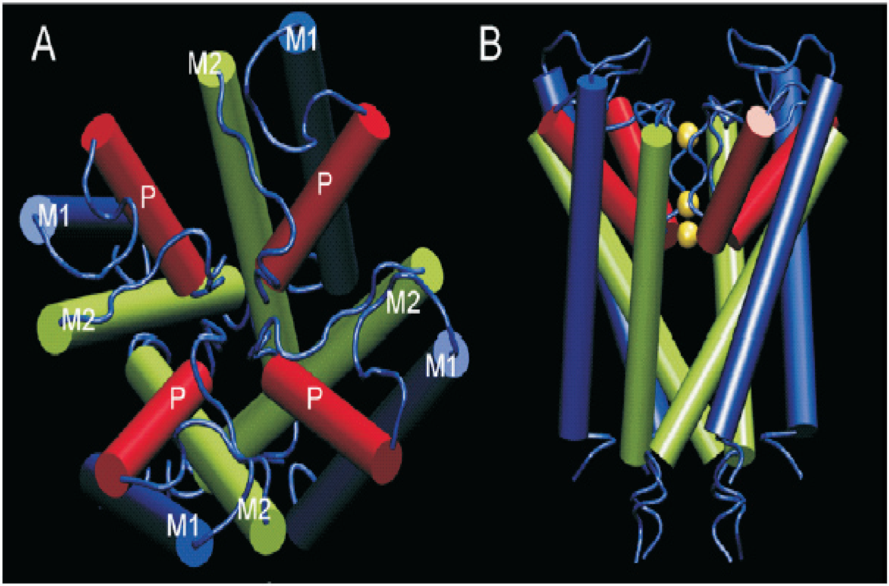

Figure 2.2

Schematic structure of the KcsA channel, with the pore lining M2 helices, the outer M1 helices, and the pore-helices P. A. View from the extracellular side; B. View from the side. The spheres indicate some of the positions where ion density is found in the crystal structure.

More complex potassium channels have large extra-cellular domains instead of these relatively short sequences. The structure of some of these domains has been solved separately, for example the T1 and beta domains of a voltage gated potassium channels [54]. A key missing piece at the moment is a structure of the other transmembrane helices of voltage-gated channels with 6 helices per subunit. The structures of KcsA currently known are all in the closed state. At the moment one structure of an open potassium channel is known: MthK is a calcium-gated potassium channel that was trapped in the open state by the presence of calcium [65].

Other channel structures that are known experimentally are two chloride channels [41] and two mechanosensitive channels [9, 29]. Several modelling and simulation studies have considered the large conductance mechanosensitive channel MscL [17, 45, 55, 106], but to date no simulation studies of the chloride channels have appeared. All of these channels are from bacteria, but the potassium and chloride channels have significant homology to eukaryotic channels. Homology modelling techniques may be used to build molecular models of at least parts of these channels. For example, as mentioned above two out of the six helices that make up the transmembrane domain of the voltage-gated Kv channels, share homology with KcsA, and models of the inner two helices of such channels have been build (recently reviewed by [25]). Similarly, the pore parts of voltage-gated sodium channels and calcium channels are also likely to be homologous to KcsA. If we try to link conductance properties to such models an additional degree of uncertainty is introduced by the use of homology models rather than ‘real’ structures. For this reason I do not consider homology models in this chapter, although clearly the proteins modeled are physiologically very important, and often are important drug targets. As more template structures become available and computational procedures improve, this type of modelling is expected to become increasingly important and useful.

A final important class of channel proteins for which structural information is available but which is not related to the potassium/sodium/calcium channels or the chloride channels is the class of ligand-gated channels. These include neurotransmitter gated channels such as the nicotinic acetylcholine receptor, GABA receptors, and glycine and serotonin receptors. A lower resolution structure of the nicotinic acetylcholine receptor has been obtained from electron crystallography studies [117]. This protein is a hetero-pentamer consisting of a mixture of homologous subunits. It probably has 4 transmembrane helices per protein, and has large domains outside the membrane. The extracellular domain is highly homologous to a recently discovered water-soluble acetylcholine binding protein, the structure of which was solved by crystallography [22]. The pore-lining helices by themselves aggregate into a channel with conductance properties reminiscent of those of the full channel [84]. Several models of this peptide channels have given an idea of what it looks like [71, 80]. Combined with the low-resolution structure of the full protein and the high-resolution structure of the acetylcholine-binding domain, models/structures of the full channel as well as homologous proteins are likely to appear soon.

Most simulation and modelling systems have used a set of well-characterized model systems, ranging in complexity from an infinitely long featureless cylinder to the actual potassium channel KcsA. The literature on these channels is quite extensive, and the references given here are mostly just examples of recent papers or reviews. Popular model systems include simple geometric shapes; gramicidin A, a very well studied peptide antibiotic [92]; alamethicin, another channel forming peptide [114]; the leucine-serine peptides LS2/LS3 [72], designed peptides that form channels; channels formed by the pore-lining helices of the nicotinic acetylcholine receptor (M2δ) [71, 80]; the transmembrane segment of Influenza A M2, a four-helix proton channel [47]; and OmpF, a large and stable bacterial porin [61, 62, 89]. The structures of OmpF and gramicidin A are known experimentally, and reasonable models based on experimental constraints can be constructed for the others. Below I focus mainly on gramicidin A, alamethicin, OmpF porin, and KcsA.

2.2 Simulation methods

Computer simulations make it possible to explore dynamic aspects of ion channels that cannot be addressed directly by the methods of structural biology. For example, X-ray diffraction provides a time- and space-averaged structure of a membrane protein in a specific crystal environment, whereas with simulations we can explore the structural dynamics of a single channel protein molecule embedded in a realistic model of the bilayer environment. In this section, different simulation methods that are currently applied to ion channels are reviewed briefly. Several recent reviews have described various aspects of channel permeation calculations in more detail, and contain references to older reviews [32, 70, 91, 93, 111].

Ion channels can be described at many different levels of detail, and a major consideration of any simulation study is its choice of level of detail. In principle one could treat the entire system (channel, water, ions, membrane) quantum mechanically, taking the distribution of all electrons in the system into account. This is currently not possible in practice, and in general not desirable even if it would be possible because it seems unlikely electronic detail everywhere in the model is relevant for the process (ion conduction) that we would like to study. Mixed quantum mechanics and molecular mechanics (QM/MM) simulations are becoming increasingly feasible for proteins, such as enzymes in which the active site is modelled by quantum mechanics but the environment is modelled classically [27], or rhodopsin where the chromophore and its direct environment are treated by quantum mechanics [90]. This approach has not been widely applied to ion channels [51] and, indeed, it remains unclear whether this level of detail will prove necessary. The majority of channel simulations could be conveniently divided into two categories: atomistic and coarse–grained. In atomistic simulations all (or most) of the atoms in a simulation system are treated explicitly. For a system made up of a channel molecule embedded in a small (ca. 100 Åx 100 Å) patch of lipid bilayer with water and ions on either side this amounts to ca. 50,000 atoms or more. Such simulations have become possible in recent years following developments in accurate simulations of pure lipid bilayers

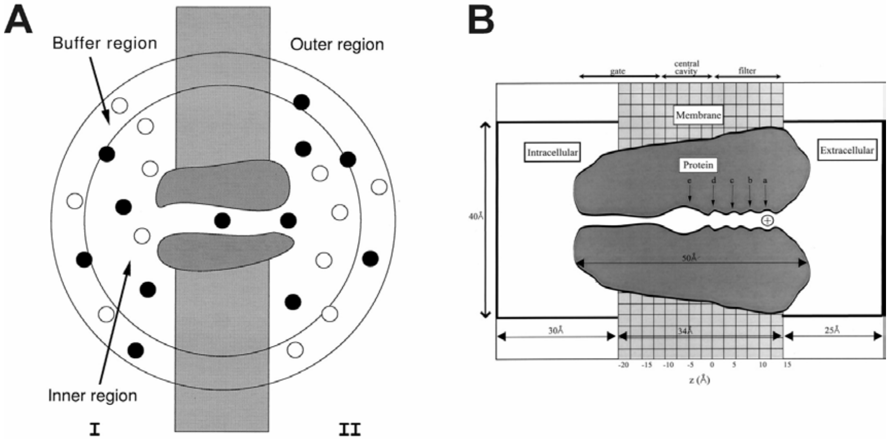

Figure 2.3

Different schemes to separate ***interesting*** from ***less interesting*** regions in simulations. A. The partitioning of a system for BD calculations proposed by Im and Roux. Reproduced with permission from [63]. B. The partitioning of the system used by Burykin et al. The area within the black square is explicitly represented in the simulation. Reproduced with permission from [23]. Simulation methods used in both studies are different, but have in common that the interior of the protein is treated in most detail, whereas remote regions of the solvent and membrane are treated in very little detail.

[96, 102, 116], faster computers, and more efficient algorithms. With current computational resources system sizes are limited to about 100,000 – 200,000 atoms and simulation times of up to 10–100 ns, although this of course depends critically on available computers and software. For some very wide or simplified channels, this means a conductance can be calculated from an MD simulation [35], but in general this time scale is insufficient to get accurate average currents, even when ignoring all other sources of errors inherent in MD simulations.

An alternative approach is to use a coarse grained simulation in which the simulation system is considerably simplified. Often, but not always, the protein is treated in explicit atomic detail, whereas the surrounding environment is treated as a continuum. In these approximations, simulation techniques are usually either Monte Carlo or Brownian dynamics, combined with continuum electrostatics theories to incorporate the effects of the atomic detail that is ignored in these calculations. Various mean field approaches allow simulation times to be comparable to experimental times for ion permeation. Brownian dynamics in particular simulates the dynamics of ions and can reach microseconds or longer, long enough to accurately calculate a current. However, as detailed below, significant difficulties remain regarding the choice of the best mean field model and the connection between the results of short timescale atomistic simulations and long timescale mean field simulations. In some of the most interesting recent methodological studies, different levels of detail are used for different parts of the system; for example, the inside of a channel is described in as much detail as possible, whereas bulk water and membrane away from the channel are highly simplified. Two examples are shown in Figure 2.3.

2.2.1 Molecular dynamics

The most common atomistic simulation technique as applied to ion channels is molecular dynamics (MD) [56]. In MD simulations the interactions between all atoms in the system are described by empirical potentials. An example of a common potential function is:

$$\begin{aligned} V(\mathbf{r}^N) = & \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 \\ & + \sum_{\text{torsions}} \frac{V_n}{2}(1+\cos(n\omega-\gamma)) + \\ & \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \end{aligned} \tag{2.1}$$

This potential function contains harmonic terms for bonds and angles, a cosine expansion for torsion angles, and Lennard-Jones and Coulomb interactions for nonbonded interactions. The constants k_i are harmonic force constants, l_i, is the current, $l_{i,0}$ the reference bond length, θ_i the current, $\theta_{i,0}$ the reference angle, V_n, n and γ describe dihedral angles (rotations around a central bond), ε and σ are Lennard-Jones parameters (a different pair for each possible combination of two different

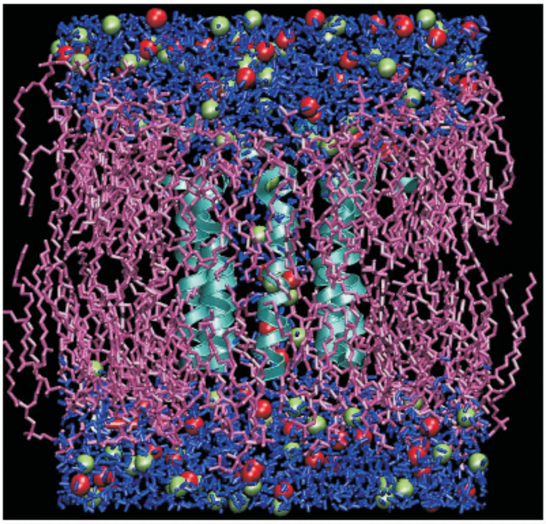

Figure 2.4
Snapshot of an ion channel from an MD simulation of alm K18 (see Section 2.3.3). The 8 helices are shown as blue ribbons; the lipids are shown in purple, water in blue. Chloride ions are green spheres, potassium ions red spheres. (See color insert.)

atom types), q_i and q_j are (partial) atomic charges, and r_{ij} is the distance between two atoms. Using this potential function the forces (the derivative of the potential with respect to position) on all atoms in the system of interest are calculated and used to solve classical equations of motions to generate a trajectory of all atoms in time. An example of a simulation system is given in Figure 2.4. This system contains a model ion channel, an explicit lipid bilayer, water, and salt.

This trajectory is the primary result of the simulation, from which specific details of the system can be analyzed. This is an exciting idea, because atoms can be followed as they move in real time on a timescale of up to ca. 100 ns, although longer simulations have also been reported. In principle, any properties that depend on coordinates, velocities or forces can be calculated, given sufficient simulation time. No assumptions are required about the nature of the solvent, there is no need to choose dielectric boundaries because all atoms are explicitly present, and in principle all interactions (water-ions, water-protein, water-lipid, lipid-protein etc.) are incorporated. This method automatically includes the dynamics of the ion channel protein itself as well as any dynamic effects of the lipids on the ion channel.

Molecular dynamics simulations have been applied by many groups, to a large selection of ion channels as well as to homology models of complex ion channels [91, 111]. Although atomistic simulations have the significant advantage of providing a detailed view of ion, water and protein dynamics, it remains challenging to link such simulations with macroscopic observed properties. Molecular dynamics simulations so far have been useful for a number of problems:

1. MD simulations have shown how the properties of water and ions in narrow channels change significantly compared to bulk. In particular, in many cases water molecules are strongly oriented due to local electric fields from the protein. Examples of this may be found in porin [112] and channels formed by parallel helix bundles [111]. In addition, water and ion diffusion coefficients are significantly lower than in bulk, which is relevant for coarse-grained simulations.

2. MD simulations have given insight into the actual process of ion motion in potassium channels, as well as into local structural changes that may explain the experimentally observed differences between e.g. sodium and potassium in the potassium channel, or different types of ions in gramicidin A (see below).

3. MD simulations have been useful to construct models of channels for which the structure is not known, when such simulations are combined with other modelling techniques and experimental data (see [25] for a review).

4. MD simulations have begun to give detailed insight into the interactions of small molecules and toxins with ion channels (e.g. [36, 46]).

5. MD simulations can be used to make models of states of ion channels that are not present in crystal structures. The open channel models for KcsA of Biggin and Sansom are a good example of this [15].

6. MD simulations give insight into the effect of the environment (lipids) on the channel protein and vice versa (e.g. [85, 113]). This is an important aspect because there are few other techniques available to study this directly.

7. MD simulations can be used to provide parameters and other information for more course-grained simulations. This is potentially a very powerful use of molecular dynamics simulations that has been applied in a number of cases. Two recent examples can be found in interesting studies of OmpF [61] and KcsA [23].

Nonetheless, there are several important caveats and limitations that have to be taken into account. An obvious limitation is the combination of system size and simulation length, which is mainly determined by available computer power and software efficiency. In particular, the maximum time scale of ca. 100 ns is not enough to accurately determine the average number of ions passing through a channel, except for very wide channels such as porins or simplified geometrical models. This means that by timescale alone one of our primary objectives, connecting atomic models with current-voltage curves, is still mostly out of reach of MD simulations. A second limitation is inherent in the specific choice of algorithms used. For example, it is now quite clear that electrostatic interactions must be accurately calculated, but due to their $1/r$ dependence (long-ranged compared to the system size) this entails a certain degree of approximation. Several methods have been proposed (e.g. [114, 23]

for references), the most popular of which currently is Particle Mesh Ewald [95], although in my opinion this may not be the final answer for membrane systems due to its artificial symmetry. More fundamentally, the simple potential functions used might not be accurate enough for important details of e.g. ion-protein interactions across a range of ions [66]. Although any potential energy function could be used in principle, including much more complex versions than the one shown in Eq. 2.1, parameterizing more complex functions is a daunting task and there might not be sufficient experimental data to test the parameters.

Certain crucial aspects of ion channel function are hard to incorporate in MD simulations of periodic systems with tens of thousands of atoms. One problem is that incorporating transmembrane potential differences is not straightforward, although a reasonable and promising approximation has been developed [35]. Clearly, this is crucial if we want to calculate current-voltage curves. A second problem is that ionic concentrations are difficult to model. Even uniform low salt concentrations are not straightforward to represent in a simulation, because it is difficult to sample the motion of the 27 K^+ and 27 Cl^- ions that would make up a 0.15 M KCl solution in a simulation with 10,000 water molecules. Such a simulation would also ignore the effect that the lipids have on the local salt concentration near the bilayer [28], which differs significantly from the bulk concentration. Biologically relevant concentrations of calcium or protons are even more problematic: in a typical simulation system a physiological calcium concentration in the micromolar range would correspond to much fewer than 1 ion. Modelling the effect of pH has similar problems with concentration and the additional problem that it is hard to make protons hop between different groups in a classical potential. Usually, pH is incorporated by calculating the pKa of ionisable residues and adjusting the protonation state of ionisable amino acids according to the desired pH.

Finally, the starting models used for simulations are rather critical at the moment. Most simulations of ion channels have been carried out on a handful of crystal structures, including gramicidin A, OmpF, a mechanosensitive channel, and the potassium channel KcsA. From an ion channel perspective KcsA is by far the most interesting of these, but the crystal structure initially had a fairly low resolution, which caused some uncertainty in the starting structures for the simulations. Simulations have also been done on homology models of various channels, in which case it becomes even more important to carefully consider the sensitivity of the results obtained to changes in the model [26].

Before examining continuum models, I would like to mention a class of simulations based on semi-microscopic models that combine fully atomistic detail in parts of the system with long-range electrostatic corrections based on a series of models that treat the environment as a lattice of rotating dipoles [23, 78, 6]. This method appears quite accurate and flexible, but is not implemented in most of the common software packages for molecular dynamics simulations.

2.2.2 Continuum electrostatics

Theories based on a continuum or field description of ion channels operate closer to the level of measurements of channel current-voltage (I-V) curves, and indeed their main advantage is that it is possible to calculate actual IV curves, and from the IV curves properties like selectivity with only modest computational effort. In recent years, a number of papers have applied electrostatic and electrodynamic theory to ion channels. Much of this work has been reviewed, e.g. by [70, 33, 43, 64]. The relative merits of different approaches, most notably those based on kinetic models, transport equations and on Brownian dynamics, have been hotly debated in a discussion in the *Journal of General Physiology* (1999, vol. 113). Rate models, which describe ion permeation in terms of movement across barriers between binding sites, have a long history in ion channels [58]. They might be most useful in helping to infer details of channel structure from experimental observations, rather than vice versa, and are outside the scope of this chapter. In general, continuum theories for ion channels are based on similar theories from physical chemistry, developed for macroscopic systems such as electrolyte solutions. I will outline the Poisson-Nernst-Planck theory and then focus on Brownian dynamics.

Poisson-Nernst-Planck

The Nernst-Planck equation describes flux of ions driven by an electrochemical potential gradient across the ion channel [58]. The flux $\mathbf{J}_i$ of particles (i.e. ions) of type i is given by:

$$\mathbf{J}_i(\mathbf{r},t) = -D_i(\mathbf{r})\left[\triangle n_i(\mathbf{r},t) + \frac{n_i(\mathbf{r},t)}{kT}\triangle\mu_i(\mathbf{r})\right] \tag{2.2}$$

where $D_i(\mathbf{r})$ is the spatially dependent diffusion coefficient, n_i the number density and $\mu_i(\mathbf{r})$ the external potential acting on the particles. Particles move under the influence of a chemical potential gradient. This general formulation enables incorporation of arbitrary factors that influence effect ion permeation as long as they can be expressed as a chemical potential. Such factors might include e.g. interactions between the walls and ions, or interactions between ions at short distances [74]. When applied to ion channels in the steady state limit, the arbitrary chemical potential $\mu_i(\mathbf{r})$ gradient usually has been replaced by the electrostatic potential gradient. Perhaps this could be exploited in improvements to this theory as applied to ion channels, as numerical methods and modern computers should allow quite complex potentials of mean force. When the electrostatic potential gradient is used in stead of the chemical potential, the resulting equation is the Nernst-Planck equation for ions of type I (simplified to flux in one dimension, z)

$$J_i(z) = -D_i\left(\triangle n_i(z) + \frac{q_i n_i(z)}{kT}\triangle\phi(z)\right) \tag{2.3}$$

where D_i is the diffusion coefficient of species i, n_i is the position dependent number density, q_i the charge, and ϕ the electrostatic potential. In this form, the driving forces for ion permeation are a concentration gradient and an electrostatic potential gradient. The flux J_i is related to the current I_i carried by ion type i by $I_i = q_i FSJ_i$, where q_i is the charge of ion type i, F is Faraday's constant, S is the channel cross sectional area (as the flux through ion channels is calculated per area, not per volume element), and J_i is the flux in the z-direction of ion i, assuming the channel axis is parallel to the z-axis. (Note that this leaves us with a problem of how to define S for a channel of non-uniform geometry). The Nernst-Planck equation for current can be rewritten in integral form as:

$$I_i = z_i F \frac{c_{out,i} - c_{in,i} \exp(z_i F \triangle V/(RT))}{\displaystyle\int_0^{\delta} \frac{\exp(z_i F \phi(z)/(RT))}{D_i} dz} \tag{2.4}$$

where $c_{out,i}$ is the extracellular concentration of species i, $c_{in,i}$ is the intracellular concentration of species i, $\triangle V$ is the transmembrane potential difference, ϕ is the electrostatic potential, δ is the thickness of the membrane, and the extracellular membrane face is at z = 0. In principle, this equation relates the measured current (I) to the transmembrane potential (V), with as input parameters the ion concentrations outside the channel and the electrostatic potential profile $\phi(z)$. However, this potential profile is generally not known and difficult to calculate.

As a first approximation, this potential can be assumed to be linear (i.e. the field everywhere in the channel is the same), leading to the classical Goldman-Hodgkin-Katz solution of the Nernst-Planck Equation [58]. This is a strong approximation. An improvement is to calculate the electrostatic potential in the ion channel from the Poisson equation and the (partial) charges on all atoms of the ion channel plus the induced charges due to the different dielectric constants of the protein, membrane and solvent. The assumption then becomes that permeating ions do not significantly change the local electrostatic potential. For many channels this is unrealistic. By using the Poisson-Boltzman equation instead of the Poisson equation, the shielding due to the presence of salt can be taken into account. However, the Poisson-Boltzman theory is only valid in equilibrium, which is not normally the case in ion channels [43], and at low concentrations, whereas a typical charge density in an ion channel is in the molar range. A second approximation is that the Boltzmann factor is not determined by the mean electrostatic potential, as assumed in the PB theory, but rather by the potential of mean force between ions. In particular, this means that at short distances from ion i, the potential of mean force does not vary smoothly as $z_j \phi(r_j)$ for all distances larger than the ion diameter, but has a more complicated behaviour. An example of PB calculations on a channel is given in Figure 2.7 in Section 2.3.3, where the effect of ionic strength and one of the input parameters for PB calculations (the Stern exclusion radius) is investigated for a model ion channel.

The Nernst-Planck and Poisson equations can be solved simultaneously, with appropriate boundary conditions to take into account the transmembrane difference in electrochemical potential. This means that the local electrostatic potential in the channel depends on the fixed charges of the protein, the permeating ions, the induced charges due to the different dielectric constants of the membrane and the solvent, and the boundary conditions. The resulting equations are often referred to as Poisson-Nernst-Planck (PNP) equations in the literature on ion channels. They consist of

Poisson's equation for the ion channel system:

$$\triangle.[\varepsilon(\mathbf{r})\triangle\phi(\mathbf{r}) = -4\pi\left[\rho(\mathbf{r}) + \sum_{i=1}^{N} z_i e n_i(\mathbf{r})\right] \tag{2.5}$$

(where the first term on the right-hand side is the charge density of the fixed charges in the channel and the membrane, and the second term is the average charge density of the mobile charges) combined with the steady-state equation for drift-diffusion to accommodate the fluxes of the mobile ions:

$$0 = \triangle.\mathbf{J}_i = \triangle.\left[-D_i(\triangle n_i(\mathbf{n}_i(\mathbf{r}) + \frac{n_i(\mathbf{r})}{kT}\triangle\mu_i(\mathbf{r}))\right]; \qquad i = 1,\cdots,N \tag{2.6}$$

Here $\mu_i(\mathbf{r})$ is the chemical potential. In its simplest form, this could just be $z_i e\phi(z)$, which assumes the chemical potential can be approximated by the electrostatic potential and only depend on z, the depth inside the channel and membrane. In this case, ions interact through the average potential $\phi(z)$. However, like before $\mu_i(\mathbf{r})$ can also include other interactions, providing a way to improve the theory. In these equations, $c_i, \mathbf{J}_i, z_i$, and D_i are respectively the concentrations, fluxes, valences, and diffusion constants of the ion species i. These two equations are coupled, because the flux changes the potential due to the mobile charges, and the potential changes the flux. In practice, they are solved simultaneously to self-consistency using numerical methods. When all the fluxes J_i are zero, and $n_i \sim \exp(-z_i e\phi/(kT))$, again with ϕ the average potential, these equations reduce to the normal Poisson-Boltzmann equation. Thus, the PNP equations are an extension of the PB equation, and the same assumptions as in PB theory underly PNP. To single out one assumption: ions interact with each other only through the average charge density. This may be problematic, as in ion channels these interactions are discrete: a binding site with an average occupancy of 0.25 will enter the average charge density as a charge of 0.25, but this does not reflect the real situation. It may be argued that this may not be too serious a problem in that the ion channel walls have such a high charge density that ion-ion interactions are less important, and the average charge density is good enough. However, this remains to be determined in specific cases.

2.2.3 Brownian dynamics

A different approach, that maintains some of the benefits of atomistic simulation, is provided by Brownian dynamics (BD). In BD, typically ions and the ion channel are represented explicitly whereas solvent and lipids are represented implicitly (see also Figure 2.3). In these simulations, ions move stochastically in a potential that is a combination of ion-ion interactions, ion-protein interactions, and a mean field. These three components can be treated at different levels of complexity, analogous to the calculation of the electrostatic potential for use in the Nernst-Planck equation. In Brownian dynamics simulations the trajectories of individual ions are calculated using the Langevin equation:

$$m_i \frac{d\mathbf{v}_i}{dt} = -\gamma_i \mathbf{v}_i + \mathbf{F}_R + q_i \mathbf{E}_i + \mathbf{F}_s \tag{2.7}$$

where $m_i, q_i, \mathbf{v}_i$ are the mass, charge and velocity of ion i. Water molecules are not included explicitly, but are present implicitly in the form of a friction coefficient $m_i \gamma_i = kT/D_i$ and a stochastic force $\mathbf{F}_R$ arising from random collisions of water molecules with ions, obeying the fluctuation-dissipation theorem. The term $q_i\mathbf{E}_i$ is the force on a particle with charge q_i due to the electric field $\mathbf{E}_i$ at the position of particle i. In a first approximation, this field is due to the partial charges plus an applied external field arising from the transmembrane potential. However, this term should also include the effect of multiple ions and reaction field terms (image charge effects) due to moving ions near regions with a low dielectric constant. Finally, $\mathbf{F}_S$ is a short-range repulsive force between ions and possibly between ions and protein. This short-range force could be modelled as a hard sphere potential or as the repulsive part of a Lennard-Jones potential. When the friction is large and the motions are overdamped, the inertial term $m_i d\mathbf{v}/dt$ may be neglected, leading to the simplified form

$$\mathbf{v}_i = \frac{D_i}{kT}(q_i\mathbf{E_i} + \mathbf{F}_s) + \mathbf{F}_R \tag{2.8}$$

This is the approximation made in Brownian dynamics. This form has been used in several ion channel simulations. When the free energy profile changes rapidly on the scale of the mean free path of an ion, the full Langevin equation including inertial effects must be used. BD simulations require only a few input parameters: in its simplest form the diffusion coefficients of the different species of ions and the charge on the ions. However, the model can be refined. The ion channel is present as a set of partial charges, and some form of interaction potential between the mobile ions and the protein must be specified (see above). The result of BD simulations is a large set of trajectories for ions, from which macroscopic properties such as conductance and ion selectivity can be calculated by counting ions crossing the channel. In addition, the simulations yield molecular details of the permeation paths for different types of ions. Such simulations have been performed of a series of different systems, including simplified ion channel models [11, 31, 34, 75, 82], gramicidin A [42], KcsA [23, 30] and OmpF porin and mutants [61, 86].

Although Brownian dynamics simulations are conceptually simple, in practice they can give rise to a number of problems for which there may not be an obvious solution. Representing all solvent effects by a random force and a diffusion coefficient is a drastic simplification, particularly for narrow regions where ions interact very strongly with one or two highly oriented water molecules. A recent study on gramicidin suggested continuum electrostatics calculations cannot provide a potential for Brownian dynamics that is accurate enough to reproduce the experimental data on gramicidin A [42]. Such calculations require several parameters such as dielectric constants whose values cannot be derived from basic principles [23]. In the context of pKa calculations for titratable amino acids by continuum calculations this has been addressed in great detail, and it has been shown that the dielectric constant for a protein depends on the approximations made; it might best be chosen differently for different types of interactions [101]. Brownian dynamics simulations so far do not take protein flexibility into account, but there is evidence from MD simulations that this can be quite critical for e.g., potassium channels and gramicidin A. Finally, even if continuum electrostatics theory provides the field inside the channel with sufficient accuracy to get realistic currents, calculating the field is computationally challenging. Only recently was the important reaction field contribution (due to image forces caused by moving permeating ions from a high to a lower dielectric environment) incorporated in simulations of realistic 3D ion channel models [23, 32, 61].

2.2.4 Other methods

One of the biggest problems of PNP and its equilibrium sibling PB is probably that the short-range potential of mean force is not correct, which leads to incorrect interactions between ions and protein and between ions and other ions. Similarly, in BD the short-range interactions between ions and ion and protein are not correct because of the continuum representation of the solvent. However, there are other theories, taken from the physical chemistry of ionic solutions, that go beyond PB that improve on this aspect of the simulations. In principle, one would like to use a theory that included e.g. the finite size of ions and single filing of ions and water, using techniques from statistical mechanics of electrolytes. A number of interesting recent papers have started to explore application of more advanced statistical mechanical approaches to channels and channel-like systems. These methods include Monte Carlo [50], density functional theory [49], and calculations using the mean spherical approximation for electrolytes [83]. Some applications of these methods are reviewed below.

2.3 Selected applications

2.3.1 Simplified systems

Clearly, the availability of high-resolution ion channel structures has been the key factor spurring rapidly growing interest in simulations and theory of ion channels. Perhaps one of the most significant other developments in the rapidly growing interest in the theory of ion channels is the influx of methods from other areas, such as the physics and physical chemistry of fluids in confined geometries and of electrolyte solutions. Initial efforts to combine these theories with biological problems in the area of channels have already given very interesting results, and I think we can expect much progress from continued work in this area. Simulations of simplified pore models, ranging for inifinitely extended cylinders to artificial pores with atomic detail, such as a carbon nanotubes, have contributed greatly to a better understanding of basic physical principles that affect selectivity, diffusion, collective behavior of ions and water in narrow pores and similar phenomena. Without trying to be exhaustive, it is interesting to consider a few recent studies on highly idealized systems that consider different aspects of simulations of channels.

Hydrophobic pores. The nicotinic acetylcholine pore has a significant hydrophobic stretch. Recent structures of aquaporin and glyceroporin also show that part of the water/glycerol permeation pathway is hydrophobic. Why might this be? One possibility is that a hydrophobic pore allows control of gating by reversibly filling and emptying. Beckstein et al. performed simulations of an artificial channel in a membrane made of 'methane-like' atoms, basically hydrophobic spheres that were harmonically restrained to generate pores of varying lengths and geometry. They found that there is a limiting radius of the pore below which water no longer fills the channel. Only a slight increase in the radius will make the channel fill with water. Introducing dipoles in the channel walls to make the pore lining more hydrophobilic makes it more favourable for water to fill the pore. Interestingly, the pore seems to fill and empty in rapid bursts [10]. Similar behavior was observed in simulations of a carbon nanotube in water. This tube alternated between completely filled and completely empty, with very rapid filling and emptying [60].

These studies might be relevant for the gating of the nicotinic acetylcholine receptor, where small changes in the radius of the hydrophobic part of the channel might switch the channel between open and closed. It is also interesting in the context of water permeation through aquaporins [38, 107]. A third recent study considered water dynamics in a narrow perfectly cylindrical channel in a dielectric slab [3]. This study solved a number of technical problems with combining atomistic detail (water) with continuum electrostatics concepts (low dielectric slab as membrane). It also showed that at a certain (realistic) threshold radius, the channel fills intermittently with water. The probability of this happening increased strongly when an ion was present, which might also be related to gating mechanisms.

These studies all used molecular dynamics simulations. Simplified models have also been used to study the effect of geometry of narrow channels on the distribution of ions, using a number of other statistical mechanics methods, including Monte Carlo simulations and density functional theory. Without going into further details, one study showed that by size alone sodium/potassium selectivity could arise, even without electrostatic effects, given certain radii for the channel [49]. A caveat here might be that this is a highly idealized channel. Other simulations considered the selectivity mechanism of calcium channels, for which the structure is not known [18, 48, 83]. It appears that a simple combination of different charge and different size for calcium and sodium ions can account for much of the observed selectivity effects in calcium channels, without invoking any specific details of the channel structure. The main detail used in these calculations was the knowledge that there are 4 highly conserved glutamate residues (the signature of calcium channels) close to each other in a narrow volume.

Comparison of continuum methods on model systems. Another useful application of simplified geometrical models is to compare different methods. Chung, Kuyucak and coworkers in particular have compared several methods to simulate ion channels, both in idealized geometries and realistic channels like gramicidin A and KcsA. Moy et al. [82] compared the results from Poisson-Boltzmann theory with Brownian dynamics in different geometries, including a catenary shape similar to that of the acetylcholine receptor. In these calculations, a spherical, cylindrical or catenary-shaped region of high dielectric is embedded in a region of low dielectric. The force on a test ion at different locations in the models is calculated using both methods, which provides a sensitive test with instructive results. The force on an ion in the center of a spherical cavity with a radius of 20 Å in 0.5 M NaCl solution (Debye length 4.3 Å) is nearly 0, both in BD and PB. As the ion is moved closer to the wall, the repulsive force from BD persists up to about 2 Debye lengths from the walls, whereas the PB solution shows a force of almost 0. Close to the wall, the BD force increases steeply but the PB force is a factor of 4 less. The problem appears to be that counter ions in the PB case provide shielding, even though there is no physical space for counter ions [82]. In the two more complicated cases (a channel-like cylinder with rounded edges and a catenary shape) similar effects were seen. In wider channels the difference in maximum force between PB and BD decreased. Thus, when force profiles are integrated, PB underestimates the height of the barriers in the electrostatic potential energy profile. In a second paper, PNP theory was compared with BD in the same model cylindrical channels and also in a potassium channel-like geometry [34]. When the channel diameter became less than 2 Debye lengths (n.b. the Debye length is 8 Å in 150 mM KCl), the PNP approach severely overestimated shielding. Overall the results provide clear examples of the dangers of failing to treat ions as discrete entities within confined spaces.

2.3.2 Gramicidin A

Gramicidin A is a peptide ion channel that has long served as a model for theoretical work on ion channels, partly because of its simple structure and partly because it was the only high-resolution structure known (Figure 2.1A). It has an unusual structure with a mixture of L and D amino acids. Because it can be chemically synthesized it has been modified in many ways. Two examples are the replacement of tryptophans with fluorinated equivalents [5] and the incorporation of a photoactive switch [19]. Although the structure of gramicidin is rather unusual, it does provide a pore lined with carbonyl groups, similar to those found in potassium channels. There is a vast literature on theoretical studies of gramicidin A, and I only present a few papers that illustrate important points learned from gramicidin A. Other reviews include [92, 111, 68, 4, 119].

In a pioneering study, Aqvist and Warshel [7] calculated a free energy profile for sodium in the gramicidin channel based on the electrostatic Protein Dipoles Langevin Dipoles (PDLD) model, a semi-microscopic model midway between an all-atom representation and a continuum model, and found reasonable well depths and barriers. They compared the results from this model with free energy perturbation calculations of the solvation of Na^+ in bulk water and inside the channel, yielding similar results when the induced dipoles due to the low dielectric membrane environment were taken into account. An important aspect of this work is that the low dielectric environment from the membrane was taken into account explicitly. This allows an estimate of the significant effect of this term (ca. 10 kcal/mol), which has not always been included in later simulations. Several molecular dynamics simulations have suggested that ion permeation is coupled to motions of the channel. Roux and

Karplus found that there is a 'peristaltic' change in conformation as a cation passes along the channel [94], displacing the carbonyl oxygens of the peptide backbone towards the channel axis. Deformability of the channel seems to play an important role in the dynamics and energetics of permeation of the channel by water and by ions. A reaction path simulation by Elber and coworkers also showed that the motion of a permeating sodium ion is coupled to motions of water and the channel. Their calculations suggest that Na^+ does not take a straight path through the channel, and questions the validity of a potential of mean force calculation for a single ion at different locations in the channel [44].

More recently, gramicidin A has been used as a test system for a variety of electrostatics, Brownian dynamics and molecular dynamics simulations. The first PNP calculations on a realistic channel used gramicidin A [69, 24, 59]. These calculations reproduced some of the properties of the experimental IV-curves but it is not so clear how critical these tests are. A recent detailed study tried to obtain reasonable free energy profiles from continuum electrostatic theory but was unable to find a combination of parameters (mainly dielectric constants for water, channel, and environment) that gave a satisfactory profile that would give a conductance comparable to experiment [42]. De Groot et al. used molecular dynamics simulations to explain why a particular form of gramicidin makes an excellent water channel [39]. Gramicidin A is found in at least two different forms, and there has been some controversy regarding which form is the main ion channel form, although the so-called head-to-head dimer is now almost universally thought to be the ion channel form [4]. The simulations of De Groot et al. suggest that by removing a formyl group from the N-terminus, the double helix form rather than the head-to-head dimer becomes dominant, and is a better channel for water transport. Tang and Xu used gramicidin to test the effect of the anaesthetic halothane on the structure and dynamics of the channel [108]. They found some changes in the dynamics of the channel in the close presence of halothane, which might be a mechanism for general anaesthetics that could modulate ion channels in neuronal membranes in a similar fashion.

Clearly, the well-known structure of gramicidin A and the vast amount of experimental data, including many chemical modifications, means this channel will remain an important system to test theoretical methods on.

2.3.3 Alamethicin

Alamethicin (Alm) is a 20-residue Aib-rich channel-forming peptide, a member of the family of peptaibols. It has been intensively studied by experimental and computational methods (reviewed in [111, 114]). There exist several variants, including covalently linked dimers of alamethicin [118] and peptides in which the Aib residues have been replaced by Leu [97]. Alm forms channels with well-defined conductance levels that are correspond to channels formed by different numbers of peptides. Breed et al. have constructed models of these channels formed by 4 to 8 helices [21]. We have done a number of simulation studies in a palmitoyl-oleoyl-phosphatidylcholine lipid bilayer on these and related channel models [110, 114] to examine the conformational stability of N = 4 to 8 helix bundle models, and to at-

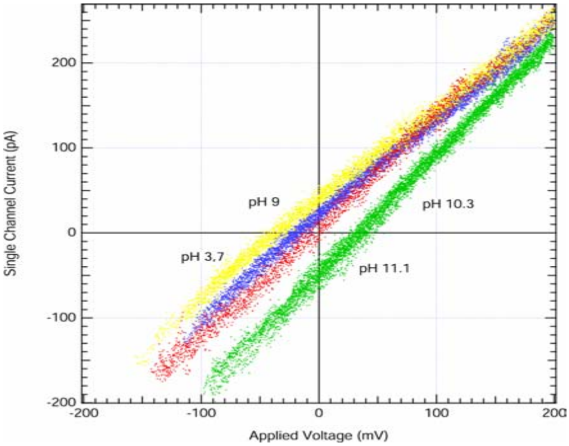

Figure 2.5

Current-voltage curves for alamethicin K18 as function of pH. Many independent measurements have been superimposed. The green data indicate that K18 is cation selective at high pH (positive reversal potential, where the current equals 0), red is not selective (intermediate pH), whereas yellow and blue are anion selective (low pH). See [20] for more details. Figure provided by Dr. Woolley. (See color insert.)

tempt to link the models with experimental conductance data. The simulations are of duration comparable to the mean passage time of a single ion through such a channel (ca. 5 ns is equivalent to an ionic conductance of 250 pS at 125 mV). Although these simulations have suggested strongly which experimental conductance levels correspond to which aggregation number of helices, a more reliable method to connect an atomistic model to conductance levels is highly desirable. Because of its simplicity, alamethicin has also been a useful test system to investigate the effect of different simulation algorithms [114], and for one particular system (N6, a parallel hexameric bundle) a simulation has been extended to 100 ns [109], one the longest simulations on a channel to date. In this fashion we obtained a validated 'best guess' model for at least one conductance level of the Alm channel, which should prove useful as the basis for future more in depth calculations of channel electrostatics and permeation models.

Woolley and coworkers designed an interesting variant of alamethicin, called K18 [105]. In this peptide, the glutamate/glutamine in position 18 has been replaced by a lysine that points into the pore, and two alamethicin peptides have been covalently coupled. The resulting peptide shows preferential stabilization of half the channel levels of normal alamethicin, suggesting pairs of helices insert simultaneously and contribute to the channels. This makes it easier to determine the number of helices in a given measured conductance level. A second interesting property of K18 is that it forms channels with a pH dependent selectivity [20].

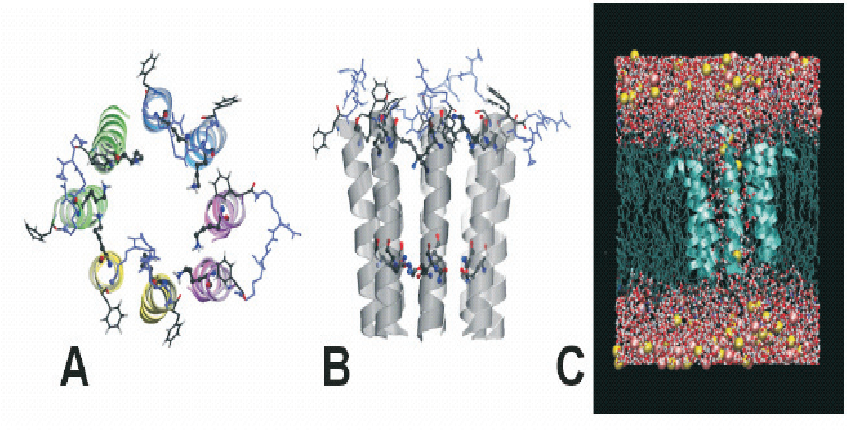

Figure 2.6
Models and a simulation system for alamethicin K18. A, B: starting model; C: full simulation system, including salt, lipids and water. Reproduced with permission from [115].

The maximum anion selectivity of the putative octameric conducting state is obtained at pH 7 or lower. Since no change in selectivity is seen between pH 7 and pH 3, and since protons are expected to be in equilibrium with the open state of the channel during a selectivity measurement, the channel should be fully charged (i.e., all 8 lysines protonated) at pH 7. An example of single channel I-V measurements is shown in Figure 2.5. This work poses several questions that simulations might be able to address. First, what is the structure of these channels? Because of the simple sequence of alamethicin we can be somewhat more confident of models of this channel than of models of complex physiological ion channels. A modeled structure is shown in Figure 2.6.

Second, can we link the models to the measured pH-dependent selectivities? Third, why is the channel not more selective for anions, even with 8 charged lysines pointing into the channel? To address these questions a number of computer simulations of the system has been performed, of 10 ns each of the octameric bundle in a lipid bilayer environment, with either 0, 4, or 8 lysines charged in the absence of salt, and with 8 lysines charged in the presence of 0.5 M or 1 M KCl. Without salt present and with all lysines charged, on average 1.9 Cl^- ions are inside the channel and the channel significantly deforms. With 0.5 M KCl present, 2.9 Cl^- ions are inside the channel. With 1 M KCl present, 4 Cl^- ions are present and the channel maintains a regular structure. Poisson-Boltzmann calculations on the same system showed the effect of ionic strength on the calculated electrostatic potential in the channel (Figure 2.7). The barriers in these graphs can be linked to a conductance through the Nernst-Planck equation. Clearly, the results are rather sensitive to the exact algorithm used

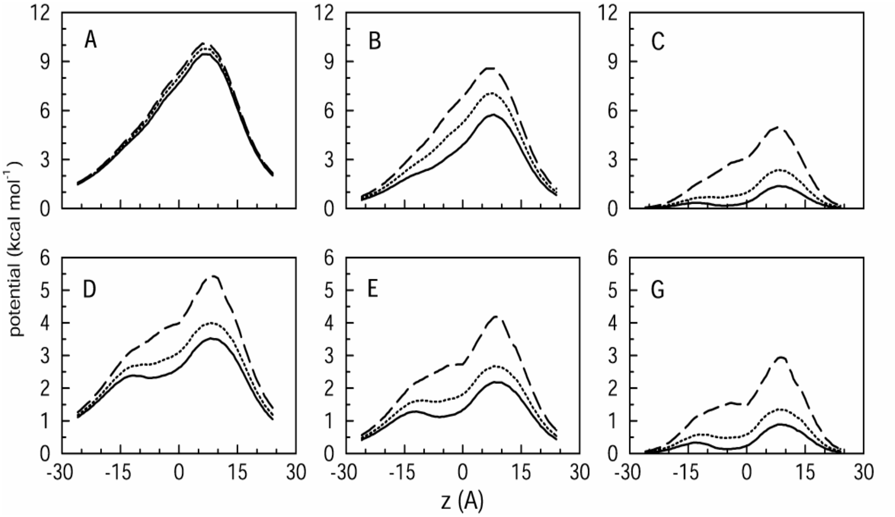

Figure 2.7

Electrostatic potential profiles calculated from the linear Poisson-Boltzmann equation (A-C) and the non-linear Poisson-Boltzmann equation (D-F) for three different values of the ionic strengths and different Stern radius values (solid line 1 Å, dotted 2 Å, long dashed 3 Å). Linear: A) 10 mM; B) 100 mM; C) 1 M; non-linear: D) 10 mM; E) 100 mM; and F) 1M. Reproduced with permission from [115].

(linearized or non-linear PB calculation) and the Stern exclusion radius (a zone in which no ions are assumed to be present). These calculations did not consider different choices of dielectric constants, although these are non-trivial. In the reasonable case in Figures 2.7D, 2.7E, 2.7G PB calculations on models of the octameric channel predict an average of 2 to 4 Cl^- ions near the lysine residues as a function of ionic strength, comparable to the numbers found from MD simulations.

These counterions lower the apparent charge of the channel, which may underlie the decrease in selectivity observed experimentally with increasing salt concentrations. We suggested that to increase the selectivity of Alm K18 channels, positive charges could be engineered in a narrower part of the channel. Because Alm K18 is essentially a designer channel, and artificially synthesized, new versions of this channel can be created, redesigned with the knowledge of the simulations in mind.

2.3.4 OmpF

Porins form large trimeric pores in the outer membrane of Gram-negative bacteria, which passively transport small molecules down their concentration gradients. They can either be general porins or transport specific substrates such as maltose. OmpF is a general diffusion pore from the outer membrane of E. coli that transports

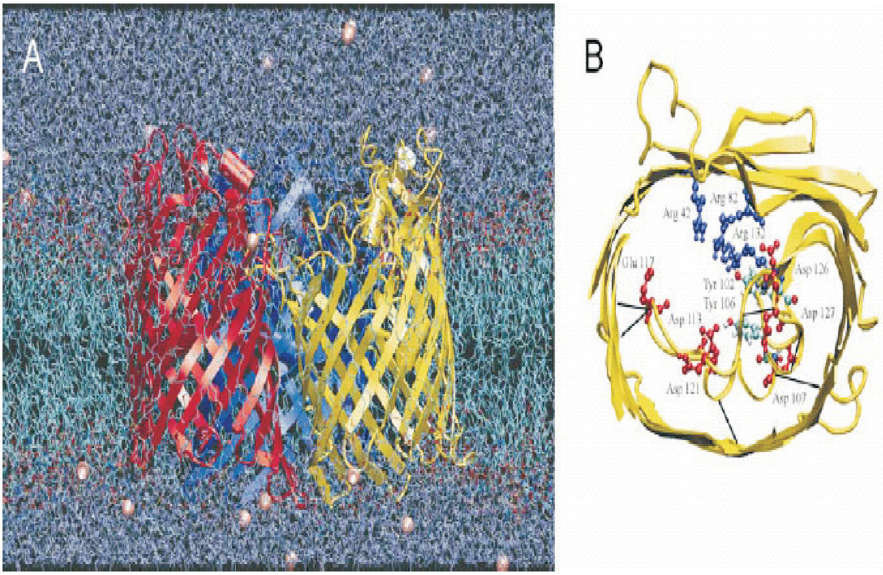

Figure 2.8

A. Simulation model of OmpF porin in a dimyristoyl-phosphatidyl-choline bilayer. The total system contains about 70,000 atoms. B. Close up of the eyelet region of one of the OmpF monomers. There is a significant separation of positive charges (blue) and negative charges (red) across the eyelet, resulting in a large local electric field. (See color insert.)

molecules up to ca. 650 dalton. It shows gating behavior, but the molecular basis and the physiological relevance of this phenomenon are not known [88]. OmpF is weakly cation selective, and its selectivity depends on the ionic strength of the solution. OmpF has been extensively studied by electrophysiology methods, although not all of this data is straightforward to interpret in terms of the properties of a single protein. Nonetheless, OmpF is an attractive model pore for calculations because its high-resolution structure is known, as are structures of a range of mutants with altered electrophysiological properties. This combination of high-resolution structures and electrophysiological data allows systematic testing and calibrating of simulation methods. Experimentally, OmpF is relatively easy to work with because it is present in high concentrations in the outer membrane and it is very stable. Mutations are also comparatively easy to make, which hopefully will facilitate testing of predictions from simulations. Structurally, OmpF is a 16-stranded betabarrel, consisting of three monomers. Porins have relatively long loops on the extracellular side and short turns on the intracellular side. The L3 loops folds back into the pore and forms the so-called eyelet region or constriction zone (Figure 2.8).

The arrangement of oppositely charged residues on opposite walls of the narrowest region of the pore creates a strong transverse electrostatic field, which is expected to have a profound effect on the behaviour of ions, water, and permeating molecules in this region.

OmpF has been the topic of a number of realistic molecular dynamics studies [62, 88, 89, 112] as well as Brownian dynamics [86, 100] and PNP calculations [61]. The crystal structure of OmpF embedded in a simulation model for molecular dynamics simulations, including lipids and solvent is shown in Figure 2.8. The BD and PNP calculations use the same protein structure but replace the membrane and water by regions of different dielectric constants.

An MD simulation of the weakly cation selective porin OmpF in a POPE bilayer [112] was the first simulation of a complex protein channel in a full bilayer environment. The simulations (of 1ns duration) revealed complex, non-bulk properties of water within the transbilayer pore. Within the pore, water diffusion coefficients were reduced by up to 10x relative to bulk. The transverse electrostatic field in the pore resulted in a high degree of alignment of the water dipoles. This would be expected to reduce the dielectric in this region, in agreement with the experimental studies cited above. The local field within the pore reached a maximum of ca. 10^9 V/m. At such a field strength the water will not behave as a linear dielectric medium. This should be taken into account in mean field treatments of such ion channels. In a followup study the orientation of permeating small dipolar molecules was studied. Both alanine and glucose strongly oriented in the narrow part of the pore, but not appreciably outside this part [89]. Im and Roux described in great detail the permeation of cations and anions obtained from atomistic MD simulations of OmpF in a bilayer with 1M KCl [62], based on a 5 ns simulation. They found different permeation paths for cations and anions in most of the channel (Figure 2.9).

This is consistent with a number of Brownian dynamics of ion flow through OmpF that showed that cations and anions follow distinct pathways with little overlap through the pore [86, 100]. It appears that anions probably require cations to permeate efficiently. OmpF is slightly more favourable for cations than for anions, but ion pair formation counteracts this to some extent. Preliminary simulations of the same system with different salt concentrations and an applied electric field have also been described by Robertson and Tieleman [88].

Although OmpF is a wide pore with a very large conductance, the MD calculations so far have not been long enough to calculate reliably a conductance. However, several continuum methods and BD have also been applied to OmpF. A recent study by Im and Roux compared ion distributions obtained from the non-linear Poisson-Boltzmann equation, Brownian dynamics and MD in OmpF. All three methods gave very similar results, with as most conspicuous feature the separation of cations and anions in two distinct sets of pathways through the channel that was observed in the earlier MD study. This is interesting, because it shows that treating OmpF as a rigid protein and the solvent as a dielectric constant seems a reasonable approximation, probably because of the large size of the pore. PNP and BD were used to calculate the conductance and the reversal potential (a measure of the selectivity of the channel). PNP and BD gave similar results, both close to the experiment, for the reversal

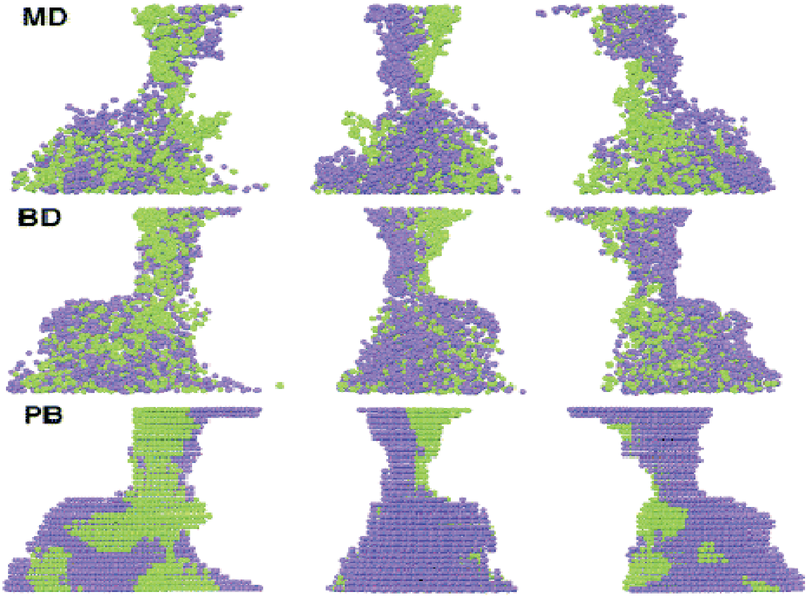

Figure 2.9

Ion distributions in OmpF porin. Two well-separated specific ion pathways with a left-handed screw-like fashion can be distinguished. The potassium ions are magenta and the chloride ions are green. MD. A superimposition of 100 snapshots of ions every 50 ps from the 5 ns trajectory. All the ions in two other pores were superimposed into one pore by rotations; BD. A superimposition of 300 snapshots of ions every 60 ps from the 60 ns trajectory; PB. An ion distribution 3D-grid map. (Left) View from perpendicular to the threefold symmetric axis. (Middle) Left view rotated by 120 degrees. (Right) Left view rotated by 240 degrees. Reproduced with permission from [61].

potential, but PNP overestimated the conductance by about 50% [61].

2.3.5 The potassium channel KcsA

The bacterial potassium channel KcsA was the first structure of a potassium channel to be solved [40] (Figures 2.1E, 2.2) and has been a prime target for simulation and modelling studies. There now is a considerable body of evidence that suggests this bacterial channel shares its main features with eukaryotic potassium channels, including evidence from toxin binding studies [79], conductance measurements [73], and from direct substitution of the KcsA filter in Shaker and inward rectifier potassium channels [77]. This last experiment is especially impressive: replacing the entire pore section of the voltage-gated Shaker channel by the pH-gated KcsA results in a voltage-gated hybrid channel.

The overall shape of KcsA resembles a truncated cone with a central pore running down the centre. The wider end of the cone corresponds to the extracellular mouth of the channel. The transbilayer pore is formed by a bundle of eight TM helices, four M1 and four M2 helices. The selectivity filter with the K channel signature motif TVGYG is located near the extracellular mouth of the channel. This filter contains distinct ion binding sites that are well resolved in the crystal structures [120]. Below the selectivity filter is a central water-filled cavity, which also shows a well-resolved ion-binding site in the high-resolution structure [120]. The pore-lining M2 helices constrict the intracellular mouth to form a putative gate region where the pore radius falls to ca. 1.1 Å (i.e. less than the Pauling radius of a K^+ ion, 1.3 Å). The recent structure of the calcium-gated potassium channel MthK in an open state suggests which considerable conformational changes take place upon gating. This new structure has to my knowledge not yet been exploited in published simulation and modelling studies, but this will only be a matter of time. Several groups have built models of what an open version of KcsA might look like, using a variety of methods including purely theoretical methods [15] and extensive mutagenesis with spin labelling for ESR measurements [76].

There now have been quite a number of MD simulations based on the 1998 structure of KcsA, with varying degrees of approximation of the protein and its environment [91, 111, 98, 99]. A simplified model of KcsA with an atomistic filter and the remainder of the protein treated as a hydrophobic continuum was used by Allen et al. [2]. The whole protein, with restraints on parts of the protein to compensate for the missing membrane environment, with water molecules within the pore at either mouth has been simulated in a number of studies, e.g., [16]. In a next step up in complexity, the unrestrained protein has been simulated embedded in a bilayer-mimetic environment made up of a 'slab' of octane molecules [53] or of *hydrocarbon-like atoms* [6]. Finally, several studies have attempted a more realistic representation of the environment of KcsA, including a fully solvated phospholipid bilayer, e.g., [12, 103] and other more recent studies. Thus, although it remains to be established what level of detail is necessary to get (sufficiently) accurate results, there is a reasonable body of simulation data upon which to draw. Below I consider some of the results obtained.

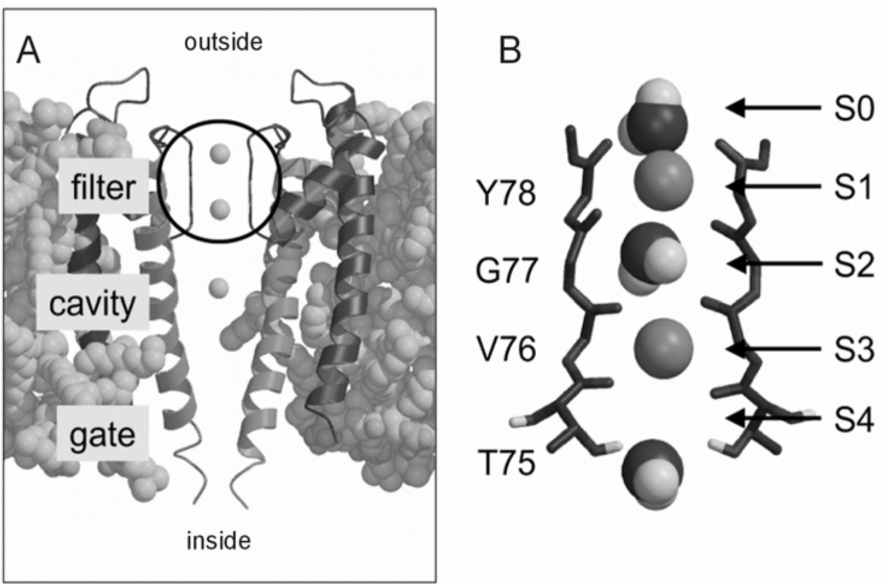

Figure 2.10

Overview of the simulation system (A), including a definition of the four sites (S1 to S4) in the selectivity filter (B). (A) The KcsA channel (shown using two polypeptide chains out of the four) is embedded in a lipid bilayer. The structure of KcsA can be thought of as made up of a selectivity filter formed by the TVGYG-motifs of the P-loop, a central cavity, and an intracellular gate where the cavity-lining M2 helices pack closely together so as to occlude the central pore. (B) The water molecules and K^+ ions in the filter are in the configuration: W(S0)-K1(S1)-W(S2)-K2(S3)-e(S4)-W(C), where S0 is the extracellular mouth, C is the cavity, and e indicates that a site is empty. This corresponds to the initial configuration of simulation KA13C. Reproduced with permission from [104].

2.3.5.1 Diffusion of ions in the channel

Given the presence of multiple K^+ ions within the selectivity filter of KcsA, a number of simulations looked at the spontaneous motions of different configurations of K^+ ions and water molecules in the filter. Several MD simulations of more than 1 ns duration have been carried out, e.g., [12, 14, 103, 104]. It is interesting to compare these simulations to see which sites within the filter are most often occupied by the ions. In the crystal structure the two K^+ ions within the filter occupied S1 and (S3 or S4). The recent 2.0 Å structure actually shows 7 different binding sites, with ion density (at high K^+ concentration) in sites S1-S4 as well as two more sites somewhat outside the filter on the extra cellular side and one ion in the cavity. Comparing the various simulations, the preferred sites when two ions are present in the filter are: (i) S2 and S4 [103]; (ii) S2 and S4 [1] (iii) S1 and S3 [53]; (iv) S2 and S4 [6]; and (v) S2 and S4 (with a 3rd ion at S0) [12]. Interestingly, two independent simulations predicted there was a favourable location for a potassium ion outside the filter, which was confirmed by the recent high-resolution crystal structure [13, 104]. As discussed below, several free energy calculations [6, 13, 23] have suggested that the difference in free energy between K^+ ions at S2 and S4, and at S1 and S3 is quite low. This is consistent with the high permeation rate of potassium ions.

In the multi-nanosecond simulations concerted motions of the K^+ ions in the filter were seen. This is illustrated in Figure 2.11, from which it can be seen that the K-W-K (i.e., ion-water-ion) triplet moves in a concerted fashion [104]. This is direct evidence for concerted single-file motion within a K channel selectivity filter. Clearly this complicates attempts to simulate ion flow through K channels as a diffusion process. It is also significant that in most simulations [1, 12, 103, 104] small (generally ca. 0.5 Å) changes in conformation of the backbone carbonyls occur. In particular, a 'flipping' of the carbonyl of V76 is observed. This is important, as it indicates that the conformation of the selectivity filter is not static, but can undergo dynamic changes on a timescale comparable to that of passage of the ions through the filter. Indeed, at low potassium concentrations (3 mM), ions are seen in the crystal structure mainly at S1 and S4, with some deformation of the filter consistent with observations in MD simulations. This may complicate mean field approaches, which thus far do not take protein flexibility into account, to simulation of ion permeation through KcsA.

2.3.5.2 Energetics of permeation

A number of groups have used atomistic simulations to explore the energetics of permeation of KcsA. Allen et al. have calculated free energy profiles for K^+ and Na^+ ions in a somewhat simplified model of a K channel, based on a channel-shaped hydrophobic pore onto which a model of the KcsA filter is grafted [2]. Their results broadly support the 'rigid filter' model of K channel selectivity (see below). However, the sensitivity of the results to initial assumptions of the rigidity of the filter is a little unclear. In a subsequent paper the same authors [1] using a complete model of the protein (but omitting the surrounding bilayer) found that the free energy differences between K^+ and Na^+ were about half those with the simplified model. Several

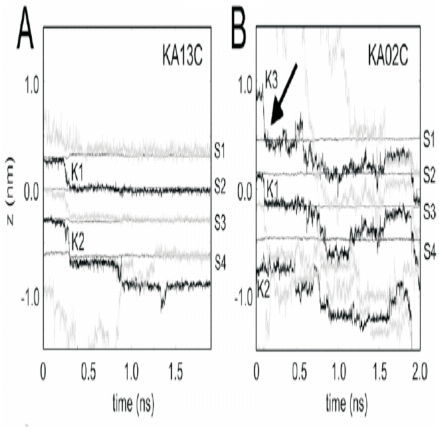

Figure 2.11

Trajectories (along the pore axis) of K^+ ions (thick black lines) and water molecules (gray lines) for two simulations with different starting configurations in sites S1-S4: (A) KA13C; (B) KA02C; Note that, for clarity, not all water molecules within the filter are shown. The locations on z (pore axis) of the four sites (S1 to S4) defined by the geometric center of the 8 oxygen atoms are indicated by the thin black lines. At each point in time, the origin of the coordinate system is defined as the center of gravity of the 16 oxygen atoms that line the selectivity filter. The black arrow in B indicates the time at which a K^+ ion enters the selectivity filter from the extracellular mouth (S0) of the channel. Reproduced with permission from [104] other groups have calculated potentials of mean force for ions in the selectivity filter. Åqvist and Luzhkov [6] showed that occupancy of sites S2 and S4 of the filter (see Figure 2.11) by two K^+ ions was more favourable (by ca. 2 kcal./mol) than occupancy of sites S1 and S3. Other configurations were of higher free energy. Thus, a permeation model based on switching of pairs of K^+ ions between these two configurations was proposed. Berneche and Roux used umbrella sampling to calculate a two-dimensional free energy map describing possible pathways for translocating ions and suggest a plausible mechanism involving correlated motions of at least 3 ions and water on a relatively flat energy landscape [13]. A third study, by Burykin et al., also calculated potentials of mean force using free energy perturbation [6]. As such calculations are becoming increasingly feasible on standard computers, it seems likely there will be significant progress in this area in the near future.

2.3.5.3 Selectivity

Why are potassium channels so selective for potassium over sodium? The key differences between potassium and sodium appear to be only a small difference in radius and in polarizability. On the basis of the X-ray structure of KcsA it has been suggested that a 'rigid' selectivity filter provides stronger cation-oxygen interactions for K^+ ions than for Na^+ ions. Thus, the energetic cost of dehydrating K^+ ions is repaid by ion/protein interactions, while ion/protein interactions are too weak to balance the cost of dehydrating Na^+ ions. Several simulations have tried to address this question, and suggest the picture might be somewhat more complex.

The deciding factor for selectivity in channels is that of the free energy of permeation; i.e., how the free energy of the system varies as different species of ion pass through the channel. The potential energies at various points along the central pore axis, which are much easier to calculate than free energies, are a first approximation. Even with this type of calculation a difference between K^+ and Na^+ ions can be observed [16]. However, for a more quantitative description free energy calculations are needed. Such calculations can yield the difference between two species of ions at a particular location, in addition to the full potential of mean force for moving a particular type of ion (as in the previous section). Allen et al. calculated that the free energy (for a $K^+ \rightarrow Na^+$ transformation) is positive within the filter region [1], which means it is more favourable for a potassium ion to be in the filter than it is for a sodium ion. However, the exact figure arrived at was somewhat sensitive to the nature of the restraints applied to the filter during the simulation. Åqvist and Luzhkov [6, 78] have performed more detailed free energy perturbation calculations. Their results also supported the 'rigid filter' model of K channel selectivity. However, it should be noted that in all three of these simulation studies it is not clear that the filter had time to fully 'relax' around the different species of cation. Longer MD simulations of KcsA with K^+ ions or with Na^+ ions in the filter suggest that the filter may be able to alter its conformation such that Na^+ ions can bind tightly within (and thus block) the filter. The geometry of interaction of Na^+ ions with the filter appears to be different from the geometry of interaction of K^+ ions [104].

Furthermore, long simulations with either K^+ or Na^+ ions at the extracellular mouth of the filter suggest a degree of selectivity in terms of which ions enter the filter [52, 104]. It is clear that very careful simulations are required to obtain the correct balance of ion/water, ion/protein and protein deformation energies. There is experimental data for other cations, e.g., rubidium [81]. In principle these could be simulated too, but they require additional testing of parameters because they are not commonly used in biomolecular simulations.

Clearly, it is becoming possible to carry out detailed numerical studies on potassium channels. The simulation results are sensitive to dynamic structural details and depend on simulation lengths and model accuracy, which might explain some of the differences in results from different labs. The fact that dynamic structural changes appear important will probably cause problems with respect to the use of restrained models (i.e., those omitting a lipid bilayer) to calculate permeation energetics. If such models are to be used, then care must be taken as to the strength and nature of the restraints. It also means simulation lengths need to be carefully checked to ensure sufficient sampling.

2.3.5.4 Interactions with toxins

Although I have not considered simulations of homology models of potassium channels in detail, I would like to emphasize a relatively new direction in simulations of potassium channels for which the full structure is not known. Potassium channels and related channels show strong binding to certain toxins, either small molecules or peptides. The voltage-gated Shaker channel and other eukaryotic voltage-gated channels interact strongly with scorpion toxins such as charybdotoxin and agitoxin. The channel and toxin form very specific complexes with dissociation constants in the nanomolar range. For this reason, these toxins as well as others for different channels have been used extensively to probe the functional properties of ion channels. By combining site-specific mutations in the toxin and in the channel, structural information on the channel (as the structure of many toxins is known) can be inferred from cooperative effects of mutations on the binding constant [57]. The resulting information is a form of low-resolution structural information on the channel as well as on the mode of interaction between the toxin and the channel. Now that there are several high-resolution structures of potassium channels, molecular modelling and simulation studies can be used to understand how these toxins bind and interfere with channel function. Several recent studies have constructed models of voltage-gated channels and their interactions with toxins, and one study used the double mutant data to at the same time refine the model of the ion channel using several molecular-dynamics based techniques.

Cui et al. used Brownian dynamics simulations to dock the scorpion toxin Lq2, a member of the charybdotoxin family, in a model of a voltage-gated potassium channel [37]. Lq2 has the interesting property that it blocks three families of potassium channels (voltage gated, calcium activated and inward rectifying channels), so that it is likely to interact with a common set of amino acids in the ion channels. This study used all 25 NMR structures for the toxin and studied their interactions simply by generating trajectories of the two proteins, without internal degrees of freedom in the proteins, and analyzing the results. The main result is a good suggestion for the mode of docking, given the homology model for the potassium channel. Similar studies have been carried out on related channels and toxins.

Eriksson and Roux recently used the experimental data on agitoxin-Shaker interactions to refine a homology model of Shaker, and at the same time to determine how the toxin binds to the channel [46]. This method is significantly more involved, and uses thermodynamic data from double-mutant cycles to restrain the modes of interactions and the possible models. Their main result is a model of Shaker and a detailed description of how the toxin interacts with Shaker, including an explanation for some ambiguous experimental data. Without going into specific details of the results, this is an interesting development: it opens a range of new experimental data for use in model building and model validation, as well as a range of new conductance data that e.g., BD simulations should be able to reproduce when the effects of the toxin is incorporated in BD simulations for cases where the toxin does not block completely.

These are technical uses, of interest in the context of this review, but of course there are also more practical implications for drug design. Ion channels already are an important target for drugs, or an important target for drugs to avoid (to prevent side effects). Two interesting examples of the use of double mutant cycle analyses, homology modelling and docking, followed by synthesis of new peptides with higher specificity as predicted by the theoretical work can be found in the work of Kalman et al. on voltage gated channels from T-lymphocytes [67] and from Rauer et al. on voltage and calcium gated channels from the same cells [87].

2.4 Outlook

Progress in modelling and simulation of ion channels in the last 5 years has been phenomenal. I think this progress has been inspired by a number of factors, including the availability of crystal structures of physiology relevant ion channels, the obvious relevance of ion channels for biomedical and pharmaceutical research, the (at first sight) comparatively simple function and basic science of ion channels, the development of efficient and sophisticated simulation and modelling software, and the rapid increase in computer power available to an increasing number of researchers. In spite of this progress, we are still short of being able to link microscopic atomistic structures to macroscopic properties of ion channels. Nonetheless, there are several reasons to be optimistic about future work in this direction.

Molecular dynamics simulations include all atomic detail and can deal with protein flexibility and conformational changes. They have been successfully used in a large number of studies to simulate local changes in structure and diffusion of water and ions, as well as to calculate potentials of mean force for ions in channels that can be used for BD simulations or kinetic theories. MD simulations are limited in time scale, system size, and the accuracy of the description of atomic interactions. The time scale that is accessible depends mostly on the speed of computers, software, and algorithmic improvements, all of which combine to allow simulations of several orders of magnitude longer than currently possible. The accuracy of current parameter sets might not be high enough to, for example, distinguish accurately between different cations. However, the potential function in Equation 2.1 is not a fundamental property of molecular dynamics, and much more complex functions could be used, including potential functions that incorporate essential electronic effects. Other improvements might be developed, perhaps based on a combination with semi-microscopic models, to deal more accurately with transmembrane potentials and differences in concentrations.

Brownian dynamics simulation is currently the most feasible way to link an ion channel structure to macroscopic properties, but it requires a description of the free energy profile for ion permeation and does not take protein flexibility into account. The latter might or might not be important for physiologically relevant channels. A description of the free energy profile is not easy to obtain. In most applications so far, only electrostatic interactions (combined with a simple short-range potential) were taken into account, calculated from Poisson or Poisson-Boltzmann equations. Free energy profiles for permeation have been calculated from MD simulations but these do not yet appear to be accurate enough. Nonetheless, these problems should be surmountable.

The limitations and prospects of mean field models depend on what they are being used for. I am not optimistic about the use of mean field models in which both ions and water are represented implicitly but a protein structure is represented in atomic detail, because the transition between mean field and atomic detail in one system is very large, and occurs on very short length scales. In many or maybe most ion channels specific interactions with ions and water appear important. In more simplified models, in which the protein is also simplified, mean field models are very interesting. They can suggest basic mechanisms for properties like selectivity, independent of atomic detail. Solving mean field models computationally only requires a fraction of the computational effort of molecular dynamics or Brownian dynamics simulations.

Combining methods from both atomistic and coarse-grained levels, using information from more detailed methods in less detailed methods that are closer to experimental data seems a promising approach to understanding the properties of ion channels in atomic detail. A start has been made in the last few years, with exciting first results. As methods are developed further and additional experimental information becomes available, simulations should be able to provide detailed insight into ion channel structure-function relationships.

Acknowledgments DPT is a Scholar of the Alberta Heritage Foundation for Medical Research. Research in his group is supported by grants from the Canadian Institutes for Health Research (CIHR), the National Research Council for Science and Engineering (NSERC), and the Protein Engineering Network of Centres of Excellence (PENCE).

I would like to thank Dr. Mark Sansom and colleagues in Oxford for interesting discussions and projects, including a review paper in *Quarterly Reviews of Biophysics* (2001. 34(4): p. 473-561) on which parts of this chapter are based. I would also like to thank Dr. Drew Woolley and Kindal Robertson for their involvement in the alamethicin and OmpF work, respectively.

References

[1] Allen T.W., Bliznyuk A., Rendell A.P., Kuyucak S., and Chung S.H. (2000). The potassium channel: structure, selectivity and diffusion. *J. Chem. Phys.* **112**: 8191-204.

[2] Allen T.W., Kuyucak S., and Chung S.H. (1999). Molecular dynamics study of the KcsA potassium channel. *Biophys. J.* **77**: 2502-16.

[3] Allen R., Melchionna S., and Hansen J.P. (2002). Intermittent permeation of cylindrical nanopores by water. *Physical Review Letters* **89**: art. no.-175502.

[4] Andersen O.S., Apell H.J., Bamberg E., Busath D.D., Koeppe R.E., et al. (1999). Gramicidin channel controversy - the structure in a lipid environment. *Nature Structural Biology* **6**: 609-.

[5] Anderson D.G., Shirts R.B., Cross T.A., and Busath D.D. (2001). Noncontact dipole effects on channel permeation. V. Computed potentials for fluorinated gramicidin. *Biophys. J.* **81**: 1255-64.

[6] Aqvist J., and Luzhkov V. (2000). Ion permeation mechanism of the potassium channel. *Nature* **404**: 881-4.

[7] Aqvist J., and Warshel A. (1989). Energetics of ion permeation through membrane channels. Solvation of Na^+ by gramicidin A. *Biophys. J.* **56**: 171-82.

[8] Ashcroft F.M. (2000). *Ion Channels and Disease*. London: Academic Press.

[9] Bass R.B., Strop P., Barclay M., and Rees D.C. (2002). Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel.*Science* **298**: 1582-7.

[10] Beckstein O., Biggin P.C., and Sansom M.S.P. (2001). A hydrophobic gating mechanism for nanopores. *Journal of Physical Chemistry B* **105**: 12902-5.

[11] Bek S., and Jakobsson E. (1994). Brownian dynamics study of a multiply-occupied cation channel: application to understanding permeation in potassium channels. *Biophys. J.* **66**: 1028-38.

[12] Berneche S., and Roux B. (2000). Molecular dynamics of the KcsA K^+ channel in a bilayer membrane. *Biophys. J.* **78**: 2900-17.

[13] Berneche S., and Roux B. (2001). Energetics of ion conduction through the

K^+ channel. *Nature* **414**: 73-7.

[14] Berneche S, and Roux B. (2001). Mechanism of ions permeation in the KcsA potassium channel. *Biophys. J.* **80**: 674.

[15] Biggin P.C., and Sansom M.S. (2002). Open-state models of a potassium channel. *Biophys. J.* **83**: 1867-76.

[16] Biggin P.C., Smith G.R., Shrivastava I., Choe S., and Sansom M.S. (2001). Potassium and sodium ions in a potassium channel studied by molecular dynamics simulations. *Biochim. Biophys. Acta* **1510**: 1-9.

[17] Bilston L.E., and Mylvaganam K. (2002). Molecular simulations of the large conductance mechanosensitive (MscL) channel under mechanical loading. *FEBS Lett.* **512**: 185-90.

[18] Boda D., Henderson D., and Busath D.D. (2002). Monte Carlo study of the selectivity of calcium channels: improved geometrical model. *Molecular Physics* **100**: 2361-8.

[19] Borisenko V., Burns D.C., Zhang Z.H., and Woolley G.A. (2000). Optical switching of ion-dipole interactions in a gramicidin channel analogue. *J. Am. Chem. Soc.* **122**: 6364-70.

[20] Borisenko V., Sansom M.S., and Woolley G.A. (2000). Protonation of lysine residues inverts cation/anion selectivity in a model channel. *Biophys. J.* **78**: 1335-48.

[21] Breed J., Biggin P.C., Kerr I.D., Smart O.S., and Sansom M.S. (1997). Alamethicin channels - modelling via restrained molecular dynamics simulations. *Biochim. Biophys. Acta* **1325**: 235-49.

[22] Brejc K., van Dijk W.J., Klaassen R.V., Schuurmans M., van Der Oost J, et al. (2001). Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**: 269-76.

[23] Burykin A., Schutz C.N., Villa J., and Warshel A. (2002). Simulations of ion current in realistic models of ion channels: The KcsA potassium channel. *Proteins-Structure Function and Genetics* **47**: 265-80.

[24] Cardenas A.E., Coalson R.D., and Kurnikova M.G. (2000). Three-dimensional Poisson-Nernst-Planck theory studies: influence of membrane electrostatics on gramicidin A channel conductance. *Biophys. J.* **79**: 80-93.

[25] Capener C.E., Kim H.J., Arinaminpathy Y., and Sansom M.S.. (2002). Ion channels: structural bioinformatics and modelling. *Hum. Mol. Genet.* **11**: 2425-33.

[26] Capener C.E., and Sansom M.S. (2002). Molecular dynamics simulations of a K channel model: Sensitivity to changes in ions, waters, and membrane environment. *Journal of Physical Chemistry B* **106**: 4543-51.

[27] Carloni P., Rothlisberger U., and Parrinello M. (2002). The role and perspec-

tive of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* **35**: 455-64.

[28] Cevc G. (1990). Membrane electrostatics. *Biochim Biophys Acta* **1031**: 311-82.

[29] Chang G., Spencer R.H., Lee A.T., Barclay M.T., and Rees D.C. (1998). Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science* **282**: 2220-6.

[30] Chung S.H., Allen T.W., and Kuyucak S. (2002). Conducting-state properties of the KcsA potassium channel from molecular and Brownian dynamics simulations. *Biophys. J.* **82**: 628-45.

[31] Chung S.H., Hoyles M., Allen T., and Kuyucak S. (1998). Study of ionic currents across a model membrane channel using Brownian dynamics. *Biophys. J.* **75**: 793-809.

[32] Chung S.H., and Kuyucak S. (2002). Ion channels: recent progress and prospects. *European Biophysics Journal with Biophysics Letters* **31**: 283-93.

[33] Cooper K., Jakobsson E., and Wolynes P. (1985). The theory of ion transport through membrane channels. *Prog. Biophys. Mol. Biol.* **46**: 51-96.

[34] Corry B., Kuyucak S., and Chung S.H. (2000). Tests of continuum theories as models of ion channels. II. Poisson- Nernst-Planck theory versus brownian dynamics. *Biophys. J.* **78**: 2364-81.

[35] Crozier P.S., Rowley R.L., Holladay N.B., Henderson D., and Busath D.D. (2001). Molecular dynamics simulation of continuous current flow through a model biological membrane channel. *Physical Review Letters* **86**: 2467-70.

[36] Crouzy S., Berneche S., and Roux B. (2001). Extracellular blockade of K+ channels by TEA: Results from molecular dynamics simulations of the KcsA channel. *Journal of General Physiology* **118**: 207-17.

[37] Cui M., Shen J., Briggs J.M., Luo X., Tan X., et al. (2001). Brownian dynamics simulations of interaction between scorpion toxin Lq2 and potassium ion channel. *Biophys. J.* **80**: 1659-69.

[38] de Groot B.L., and Grubmuller H. (2001). Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**: 2353-7.

[39] de Groot B.L., Tieleman D.P., Pohl P., and Grubmuller H. (2002). Water permeation through gramicidin A: desformylation and the double helix: a molecular dynamics study. *Biophys J* **82**: 2934-42.

[40] Doyle D.A., Morais Cabral J., Pfuetzner R.A., Kuo A., Gulbis J.M., et al. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**: 69-77.

[41] Dutzler R., Campbell E.B., Cadene M., Chait B.T., and MacKinnon R. (2002).

X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. *Nature* **415**: 287-94.

[42] Edwards S., Corry B., Kuyucak S., and Chung S.H. (2002). Continuum electrostatics fails to describe ion permeation in the gramicidin channel. *Biophys. J.* **83**: 1348-60.

[43] Eisenberg R.S. (1999). From structure to function in open ionic channels. *J. Membr. Biol.* **171**: 1-24.

[44] Elber R., Chen D.P., Rojewska D., and Eisenberg R. (1995). Sodium in gramicidin: an example of a permion. *Biophys. J.* **68**: 906-24.

[45] Elmore D.E., and Dougherty D.A. (2001). Molecular dynamics simulations of wild-type and mutant forms of the Mycobacterium tuberculosis MscL channel. *Biophys. J.* **81**: 1345-59.

[46] Eriksson M.A., Roux B. (2002). Modeling the structure of agitoxin in complex with the shaker k^+ channel: a computational approach based on experimental distance restraints extracted from thermodynamic mutant cycles. *Biophys. J.* **83**: 2595-609.

[47] Forrest L.R., Kukol A., Arkin I.T., Tieleman D.P., and Sansom M.S. (2000). Exploring models of the influenza A M2 channel: MD simulations in a phospholipid bilayer. *Biophys. J.* **78**: 55-69.

[48] Gillespie D., Nonner W., Henderson D., and Eisenberg R.S. (2002). A physical mechanism for large-ion selectivity of ion channels. *Physical Chemistry Chemical Physics* **4**: 4763-9.

[49] Goulding D., Hansen J.P., and Melchionna S. (2000). Size selectivity of narrow pores. *Phys. Rev. Lett.* **85**: 1132-5.

[50] Graf P., Nitzan A., Kurnikova M.G., and Coalson R.D. (2000). A dynamic lattice Monte Carlo model of ion transport in inhomogeneous dielectric environments: Method and implementation. *Journal of Physical Chemistry B* **104**: 12324-38.

[51] Guidoni L., and Carloni P. (2002). Potassium permeation through the KcsA channel: a density functional study. *Biochimica Et Biophysica Acta-Biomembranes* **1563**: 1-6.

[52] Guidoni L., Torre V., and Carloni P. (1999). Potassium and sodium binding in the outer mouth of the K^+ channel. *Biochem.* **38**: 8599-604.

[53] Guidoni L., Torre V., and Carloni P. (2000). Water and potassium dynamics inside the KcsA K^+ channel. *FEBS Lett.* **477**: 37-42.

[54] Gulbis J.M., Zhou M., Mann S., and MacKinnon R. (2000). Structure of the cytoplasmic beta subunit-T1 assembly of voltage- dependent K^+ channels. *Science* **289:** 123-7.

[55] Gullingsrud J., Kosztin D., and Schulten K. (2001). Structural determinants

of MscL gating studied by molecular dynamics simulations. *Biophys. J.* **80**: 2074-81.

[56] Hansson T., Oostenbrink C., and van Gunsteren W. (2002). Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **12**: 190-6.

[57] Hidalgo P., and MacKinnon R. (1995). Revealing the architecture of a K^+ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**: 307-10.

[58] Hille B. (2001). *Ion channels of excitable membranes.* Sunderland: Sinauer Associates, Inc.

[59] Hollerbach U., Chen D.P., Busath D.D., and Eisenberg B. (2000). Predicting function from structure using the Poisson-Nernst- Planck equations: Sodium current in the gramicidin A channel. *Langmuir* **16**: 5509-14.

[60] Hummer G., Rasaiah J.C., and Noworyta J.P. (2001). Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* **414**: 188-90.

[61] Im W., and Roux B. (2002a). Ion permeation and selectivity of OmpF porin: a theoretical study based on molecular dynamics, Brownian dynamics, and continuum electrodiffusion theory. *J. Mol. Biol.* **322**: 851-69.

[62] Im W., and Roux B. (2002b). Ions and counterions in a biological channel: A molecular dynamics simulation of OmpF porin from Escherichia coli in an explicit membrane with 1 M KCl aqueous salt solution. *Journal of Molecular Biology* **319**: 1177-97.

[63] Im W., Seefeld S., and Roux B. (2000). A Grand Canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels. *Biophys. J.* **79**: 788-801.

[64] Jakobsson E. (1998) Using theory and simulation to understand permeation and selectivity in ion channels. *Methods* **14**:342-351.

[65] Jiang Y., Lee A., Chen J., Cadene M., Chait B.T., and MacKinnon R. (2002). Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**: 515-22.

[66] Jordan P.C. (2002). Trial by ordeal: ionic free energies in gramicidin. *Biophys. J.* **83**: 1235-6.

[67] Kalman K., Pennington M.W., Lanigan M.D., Nguyen A., Rauer H., et al. (1998). ShK-Dap22, a potent Kv1.3-specific immunosuppressive polypeptide. *J. Biol. Chem.* **273**: 32697-707.

[68] Koeppe R.E., and Anderson O.S. (1996). Engineering the gramicidin channel. *Annu. Rev. Biophys. Biomol. Struct.* **25**: 231-58.

[69] Kurnikova M.G., Coalson R.D., Graf P., and Nitzan A. (1999). A lattice relaxation algorithm for three-dimensional Poisson-Nernst- Planck theory with application to ion transport through the gramicidin A channel. *Biophys. J.* **76**:

642-56.

[70] Kuyucak S., Andersen O.S., and Chung S.H. (2001). Models of permeation in ion channels. *Reports on Progress in Physics* **64**: 1427-72.

[71] Law R., Tieleman D.P., and Sansom M.S.P. (2003). Pores formed by the nicotinic receptor M2d peptide: a molecular dynamics study. *Biophys. J.* in press.

[72] Lear J.D., Wasserman Z.R., and DeGrado W.F. (1988). Synthetic amphiphilic peptide models for protein ion channels. *Science* **240**: 1177-81.

[73] LeMasurier M., Heginbotham L., and Miller C. (2001). KcsA: it's a potassium channel. *J. Gen. Physiol.* **118**: 303-14.

[74] Levitt D.G. (1991). General continuum theory for multi-ion channel. 1. Theory. *Biophys. J.* **59**: 271-7.

[75] Li S.C., Hoyles M., Kuyucak S., and Chung S.H. (1998). Brownian dynamics study of ion transport in the vestibule of membrane channels. *Biophys. J.* **74**: 37-47.

[76] Liu Y.S., Sompornpisut P., and Perozo E. (2001). Structure of the KcsA channel intracellular gate in the open state. *Nat. Struct. Biol.* **8**: 883-7.

[77] Lu Z., Klem A.M., and Ramu Y. (2001). Ion conduction pore is conserved among potassium channels. *Nature* **413**: 809-13.

[78] Luzhkov V.B., and Aqvist J. (2001). K(+)/Na(+) selectivity of the KcsA potassium channel from microscopic free energy perturbation calculations. *Biochim. Biophys. Acta* **1548**: 194-202.

[79] MacKinnon R., Cohen S.L., Kuo A., Lee A., and Chait B.T. (1998). Structural conservation in prokaryotic and eukaryotic potassium channels. *Science* **280**: 106-9.

[80] Montal M., and Opella S.J. (2002). The structure of the M2 channel-lining segment from the nicotinic acetylcholine receptor. *Biochim. Biophys. Acta* **1565**: 287-93.

[81] Morais-Cabral J.H., Zhou Y., and MacKinnon R. (2001). Energetic optimization of ion conduction rate by the K^+ selectivity filter. *Nature* **414**: 37-42.

[82] Moy G., Corry B., Kuyucak S., and Chung S.H. (2000). Tests of continuum theories as models of ion channels. I. Poisson- Boltzmann theory versus Brownian dynamics. *Biophys. J.* **78**: 2349-63.

[83] Nonner W., Catacuzzeno L., and Eisenberg B. (2000). Binding and selectivity in L-type calcium channels: A mean spherical approximation. *Biophys. J.* **79**: 1976-92.

[84] Opella S.J., Marassi F.M., Gesell J.J., Valente A.P., Kim Y., et al. (1999). Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. *Nat. Struct. Biol.* **6**: 374-9.

[85] Petrache H.I., Grossfield A., MacKenzie K.R., Engelman D.M., and Woolf T.B. (2000). Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J. Mol. Biol.* **302**: 727-46.

[86] Phale P.S., Philippsen A., Widmer C., Phale V.P., Rosenbusch J.P., and Schirmer T. (2001). Role of charged residues at the OmpF porin channel constriction probed by mutagenesis and simulation. *Biochemistry.* **40**: 6319-25.

[87] Rauer H., Pennington M., Cahalan M., and Chandy K.G. (1999). Structural conservation of the pores of calcium-activated and voltage- gated potassium channels determined by a sea anemone toxin. *J. Biol. Chem.* **274**: 21885-92.

[88] Robertson K.M., and Tieleman D.P. (2002). Molecular basis of voltage gating of OmpF porin. *Biochem. Cell Biol.* **80**: 517-23.

[89] Robertson K., and Tieleman D. (2002). Orientation and interactions of dipolar molecules during transport through OmpF porin. *FEBS Lett.* **528**: 53.

[90] Rohrig U.F., Guidoni L., and Rothlisberger U. (2002). Early steps of the intramolecular signal transduction in rhodopsin explored by molecular dynamics simulations. *Biochemistry* **41**: 10799-809.

[91] Roux B. (2002). Theoretical and computational models of ion channels. *Current Opinion in Structural Biology* **12**: 182-9.

[92] Roux B. (2002). Computational studies of the gramicidin channel. *Accounts of Chemical Research* **35**: 366-75.

[93] Roux B., Berneche S., and Im W. (2000). Ion channels, permeation, and electrostatics: Insight into the function of KcsA. *Biochemistry* **39**: 13295-306.

[94] Roux B, and Karplus M. (1994). Molecular dynamics simulations of the gramicidin channel. *Annu. Rev. Biophys. Biomol. Struct.* **23**: 731-61.

[95] Sagui C., and Darden T.A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **28**: 155-79.

[96] Saiz L., Bandyopadhyay S., Klein M.L. (2002). Towards an understanding of complex biological membranes from atomistic molecular dynamics simulations. *Biosci. Rep.* **22**: 151-73.

[97] Sansom M.S. (1993). Structure and function of channel-forming peptaibols. *Q. Rev. Biophys.* **26**: 365-421.

[98] Sansom M.S., Shrivastava I.H., Ranatunga K.M., and Smith G.R. (2000). Simulations of ion channels–watching ions and water move.*Trends. Biochem. Sci.* **25**: 368-74.

[99] Sansom M.S., Shrivastava I.H., Bright J.N., Tate J., Capener C.E.,and Biggin P.C. (2002). Potassium channels: structures, models, simulations. *Biochim. Biophys. Acta* **1565**: 294-307.

[100] Schirmer T., and Phale P.S. (1999). Brownian dynamics simulation of ion flow through porin channels. *J. Mol. Biol.* **294**: 1159-67.

[101] Schutz C.N., and Warshel A. (2001). What are the dielectric ”constants” of proteins and how to validate electrostatic models? *Proteins.* **44**: 400-17.

[102] Scott H. (2002). Modeling the lipid component of membranes. *Curr. Opin. Struct. Biol.* **12**: 495.

[103] Shrivastava I.H., and Sansom M.S.P. (2000). Simulations of ion permeation through a potassium channel: Molecular dynamics of KcsA in a phospholipid bilayer. *Biophys. J.* **78**: 557-70.

[104] Shrivastava I.H., Tieleman D.P., Biggin P.C., and Sansom M.S.P. (2002). K^+ versus Na^+ ions in a K channel selectivity filter: A simulation study. *Biophys. J.* **83**: 633-45.

[105] Starostin A.V., Butan R., Borisenko V., James D.A., Wenschuh H., et al. (1999). An anion-selective analogue of the channel-forming peptide alamethicin. *Biochemistry* **38**: 6144-50.

[106] Sukharev S., Betanzos M., Chiang C.S., and Guy H.R. (2001). The gating mechanism of the large mechanosensitive channel MscL. *Nature* **409**: 720-4.

[107] Tajkhorshid E., Nollert P., Jensen M.O., Miercke L.J., O’Connell J., et al. (2002). Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**: 525-30.

[108] Tang P., and Xu Y. (2002). Large-scale molecular dynamics simulations of general anesthetic effects on the ion channel in the fully hydrated membrane: The implication of molecular mechanisms of general anesthesia. *Proceedings of the National Academy of Sciences of the United States of America* **99** 16035-16040.

[109] Tieleman DP. (2002). Molecular motions in ion channels: a possible link to noise in single channels. In: S. Bezrukov, editor. *AIP Symposia proceedings 665*; 2002. AIP Press, pp. 298-304.

[110] Tieleman D.P., Berendsen H.J., and Sansom M.S. (1999). A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **76**: 1757-69.

[111] Tieleman D.P., Biggin P.C., Smith G.R., and Sansom M.S.P. (2001). An alamethicin channel in a lipid bilayer: molecular dynamics simulations. *Quarterly Reviews of Biophysics* **34**: 473-561.

[112] Tieleman D.P., and Berendsen H.J. (1998). Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: molecular dynamics simulations. *Biophys. J.* **74**: 2786-801.

[113] Tieleman D.P., Forrest L.R., Sansom M.S., and Berendsen H.J. (1998). *Bio-*

chemistry **37**: 17554-61.

[114] Tieleman D.P., Hess B., and Sansom M.S. (2002). Analysis and evaluation of channel models: simulations of alamethicin. *Biophys. J.* **83**: 2393-407.

[115] Tieleman D.P., Hess B., Sansom M.S.P., and Woolley G.A. (2003). Understanding pH-dependent selectivitiy in alameticin K18 channels by computer simulations. *Biophys J* **84**: 1464-1469.

[116] Tieleman D.P., Marrink S.J., and Berendsen H.J. (1997). A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochim. Biophys. Acta* **1331**: 235-70.

[117] Unwin N. (2000). The Croonian Lecture 2000. Nicotinic acetylcholine receptor and the structural basis of fast synaptic transmission. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**: 1813-29.

[118] Woolley G.A., Biggin P.C., Schultz A., Lien L., Jaikaran D.C., et al. (1997). Intrinsic rectification of ion flux in alamethicin channels: studies with an alamethicin dimer. *Biophys. J.* **73**: 770-8.

[119] Woolley G.A., and Wallace B.A. (1992). Model ion channels: gramicidin and alamethicin. *J. Membr. Biol.* **129**: 109-36.

[120] Zhou Y., Morais-Cabral J.H., Kaufman A., and MacKinnon R. (2001). Chemistry of ion coordination and hydration revealed by a K+ channel- Fab complex at 2.0 A resolution. *Nature* **414**: 43-8.

Chapter 3

Modelling Neuronal Calcium Dynamics

Saleet M. Jafri[1], and Keun-Hang Yang[2]

[1]*Program in Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, 10900 University Blvd., Manassas, VA 20110, U.S.,* [2]*Department of Neurology, The Johns Hopkins University, School of Medicine, 600 North Wolfe Street, Meyer 2-147 Baltimore, MD 21287, U.S.*

CONTENTS

3.1 Introduction

Calcium has been called the ubiquitous second messenger due to its widespread use in cellular signaling pathways. Calcium plays a role in many different cells for processes such as the release of hormones or neurotransmitters, cell motility or contraction, and the control of gene expression and development. In neurons, many of these functions have been observed. Intracellular calcium plays a crucial role in hormone release in pituitary gonadotropes. Neurotransmitter release in nerve terminals is mediated by calcium. Calcium is also thought to play a role in synaptic plasticity and learning. The cooperation between computational models and experimental studies has lead to greater understanding of neuronal calcium dynamics.

Neurons vary significantly in their size, shape, and function. Hence, the role of

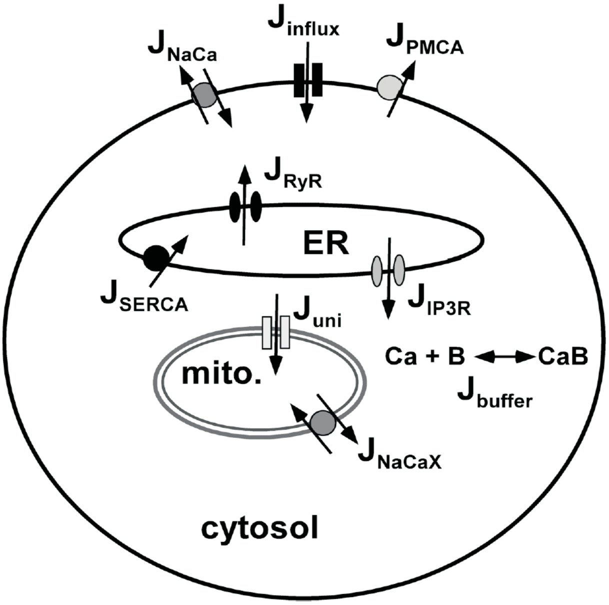

Figure 3.1

Schematic representation of the different fluxes that contribute to neuronal calcium dynamics: calcium influx across the plasma membrane through calcium channels (J_{pm}), calcium efflux across the plasma membrane (J_{efflux}) via calcium pumps (J_{PMCA}), and exchangers (J_{NaCa}), release from intracellular calcium stores via either ryanodine receptors (J_{RyR}) or IP$_3$ receptors (J_{IP3R}), sequestrations of calcium into the ER by calcium pumps (J_{SERCA},) calcium binding proteins (J_{buffer}), and mitochondrial calcium handling (J_{mito}) by the uniporter (J_{uni}) and $Na^+ - Ca^{2+}$ exchange (J_{NaCaX}).

calcium signaling in neurons will vary between different types of neurons. This work will describe general principles used to model neuronal calcium that can be applied not only to a variety of neurons, but to other cells as well. Specific examples will be used when possible to demonstrate the features mentioned.

Calcium is maintained at a very low level in resting cells (∼0.1 μM) compared to extracellular calcium (∼2.0 mM). This 10,000-fold difference results in a large electrochemical gradient from the outside to the inside of the cell. Furthermore, calcium in the internal stores is also much higher than resting cytosolic calcium (∼1.0 mM). Given these two large gradients, calcium can be quickly brought into the cytosol via calcium channels and signal various events, such as contraction, secretion, gene expression, etc. There are several mechanisms used by the cell to maintain low cytosolic calcium at rest and to allow transient increases in calcium that are mentioned in the next paragraph and described below.

There are several common features that play a role in cellular calcium signaling (Figure 3.1). These include calcium influx across the plasma membrane through calcium channels (J_{influx}), calcium efflux across the plasma membrane (J_{efflux}) via calcium pumps (J_{PMCA}), and exchangers (J_{NaCa}), release from intracellular calcium stores (J_{release}) via either ryanodine receptors (J_{RyR}) or IP3 receptors (J_{IP3R}), sequestrations of calcium by internal stores (J_{uptake} or J_{SERCA}), calcium binding proteins (J_{buffer}), mitochondrial calcium handling (J_{mito}) by the uniporter (J_{uni}) and Na^+-Ca^{2+} exchange (J_{NaCaX}), and diffusion of calcium ($J_{\text{diffusion}}$). The change of calcium concentration with respect to time is simply a sum of the fluxes that arise from these features

$$\frac{d[Ca_i]}{dt} = J_{\text{influx}} + J_{\text{efflux}} + J_{\text{release}} + J_{\text{uptake}} + J_{\text{buffer}} + J_{\text{mito}} + J_{\text{diffusion}} \tag{3.1}$$

To correctly describe cellular calcium dynamics, biophysically accurate mathematical representations for the above are needed. This chapter will first describe these common features of calcium signaling and their mathematical representation. Moreover, specific examples of how the cellular microscopic ultrastructure provides a framework in which these components can yield complex behaviors.

3.2 Basic principles

Calcium Buffering

Calcium is maintained at a very low level in resting cells. This is due in part to the presence of calcium buffers that bind approximately 90-99% of the total calcium found in the cytosol. These buffers are typically calcium binding proteins, such as calbindin, calretinin, calmodulin, calsequestrin, calcineurin and parvalbumin, or the negative charges associated with the cellular membranes [25, 47]. These buffers are quite fast so that they have the effect of binding almost all the calcium that enters cytosol. This has the effect of reducing the amplitude of changes in cytosolic calcium concentration.

Another effect of calcium buffers is its effect on calcium mobility. While the majority of calcium buffers are stationary or fixed to an immobile component of the cell, there are also mobile buffers that diffuse throughout the cytoplasm. Calcium indicator dyes are typically mobile buffers (exogenous buffers). Due to the high concentration of calcium buffers, a calcium ion can only diffuse a short distance before being immobilized by binding to a calcium buffer. The net effect is that the effective diffusion constant of calcium in the cytoplasm is 36 μm^2/s while the calcium diffusion constant in cytoplasmic extracts is 223 μm^2/s [2]. This reduction in calcium mobility by calcium buffers has to be included in any realistic model of spatial calcium signaling.

The reaction of calcium (Ca) binding to a buffer (B) can be described as a chemical

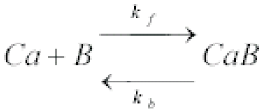

Figure 3.2
Reaction of calcium binding to a buffer.

equation as in Figure 3.2: in which calcium binds to buffer with a forward rate constant k_f and unbinds with a backward rate constant k_b. This buffering flux describing the rate of change for calcium can be represented by the equation

$$J_{\text{buffer}} = -k_f[Ca][B] + k_b[CaB] \tag{3.2}$$

In addition to the [Ca] balance equation, it is necessary to describe [B] and [CaB] by differential equations for their rate of change. If it is assumed that the total amount of buffer ($[B_{\text{total}}]$) remains constant, then we can assume that ($[B_{\text{total}}]$ = [B] + [CaB]), which allows the reduction of the total number of differential equations by one for each buffer. Typically the rate constants k_f and k_b are very fast compared to other cellular processes. This requires for a very small time step to be used when solving the differential equations associated with buffering. Because of the difference in time scale between these reactions and other cellular processes, this will increase the computational time necessary to solve the problem. A solution to this problem is the rapid buffering approximation, suggested by Wagner and Keizer [53], which assumes that the buffering reaction (Eq. 3.2) is in equilibrium. Using a steady state approximation with the differential equations describing buffering, and the chain rule, the equation describing the rate of change for total cytosolic calcium with respect to time can be broken down into a term for the rate of change for the free calcium with respect to time multiplied by the rate of change of the total calcium with respect to the free calcium. This yields a buffering factor β that scales the other fluxes in the calcium balance equation to account for the fraction of other fluxes that is not bound by the buffer.

$$\beta = \left(1 + \frac{[B_{S,\text{total}}]K_{S,eq}}{(K_{S,eq} + [Ca])^2} + \frac{[B_{M,\text{total}}]K_{M,eq}}{(K_{M,eq} + [Ca])^2} + \frac{[B_{E,\text{total}}]K_{E,eq}}{(K_{E,eq} + [Ca])^2}\right)^{-1} \tag{3.3}$$

where $K_{S,eq}, K_{M,eq}$, and $K_{E,eq}$ are the disassociation constants for calcium and the stationary, mobile, and exogenous buffers, respectively. The total concentrations for the stationary, mobile, and exogenous buffers are $[B_{S,\text{total}}]$, $[B_{M,\text{total}}]$, and $[B_{E,\text{total}}]$, respectively. Additional terms can be added to account for different buffers and calcium binding dyes on calcium dynamics. The rapid buffering approximation also accounts for the effect of buffering on diffusion by altering the diffusive flux to

$$J_{\text{diffusion}} = D_{Ca}\frac{\partial^2[Ca]}{\partial x^2} \tag{3.4}$$

to

$$J_{\text{diffusion}} = \beta \left[(D_{Ca} + \gamma_M D_M + \gamma_E D_E) \frac{\partial^2 [Ca]}{\partial x^2} \right. \\ \left. -2 \left\{ \frac{\gamma_M D_M}{K_{M,eq} + [Ca]} + \frac{\gamma_E D_E}{K_{E,eq} + [Ca]} \right\} \Delta[Ca] \cdot \Delta[Ca] \right] \tag{3.5}$$

with

$$\gamma_l = \frac{[B_{l,\text{total}}] K_{l,eq}}{(K_{l,eq} + [Ca])^2}, \qquad \text{with } l = M, E \tag{3.6}$$

where $K_{l,eq}, D_l$ are the calcium disassociation constant and diffusion constant for species $l = M, E$ for mobile and exogenous buffers respectively [23, 53]. The buffering factor scales the flux to account for the calcium binding effect of the buffers. The first term in the square brackets represents the diffusive transport of calcium. The last term in the square brackets represents the uptake of calcium by mobile buffers as free calcium moves down its concentration gradient and is a non-diffusive term. The approach assumes that the sum of the concentrations of bound and unbound buffer at any given point in space remains constant. Other equilibrium approaches for approximating buffering have been suggested by Zhou and Neher [56] and are evaluated for accuracy, i.e., under what conditions they can be used, by Smith [46]. These have not been discussed here for brevity, but the reader should seek the sources mentioned for more details.

The principles described above can be applied to models for neurons. Neher [37] developed an equation for the effective diffusion constant based on these principles that can be used to calculate the diffusion profile of a calcium flux release from a channel. This profile would fall off more sharply in the presence of calcium buffers. With this, the distance between a release site and channel activated by calcium can be calculated by determining the concentration of an exogenous buffer with known calcium dissociation constant, such as BAPTA, at which the activation of the calcium sensitive signal is abolished upon stimulation. For example, in chick dorsal root ganglion cells, calcium enters via voltage gated calcium channels. This method suggests that calcium-activated chloride channels are 50-400 nm distant and the ryanodine receptors are 600 nm distant [55].

While the buffering of the charges of the cellular membranes can be approximated by the rapid buffering approximations above, this fails to account for the effect of the electrical charges on diffusion of charged particles. For this purpose, a model of electrodiffusion based on the Nernst-Planck electrodiffusion equations can be used. This approach has been used to describe diffusion in membrane-restricted spaces such as the cardiac diadic junction by Soeller and Cannell [48].

3.2.1 Intracellular calcium stores

The main intracellular calcium store is the endoplasmic reticulum (ER). This is known as the sarcoplasmic reticulum (SR) in muscle cells. The major calcium handling components in the ER are the calcium release channels, the calcium uptake pumps, and the calcium buffers in the ER lumen.

The ER has two calcium release channels, the ryanodine receptor (RyR) and the inositol 1,4,5-trisphosphate receptor (IP_3R) [43, 54]. The RyR releases calcium from the ER in response to an increase in calcium. This positive feedback of calcium release is termed calcium-induced calcium release. There are three major isoforms of the RyR, RyR1, RyR2 and RyR3, which differ in their biophysical properties [18]. RyR1 is the main isoform in skeletal muscle. RyR2 is the most common isoform in heart and brain. RyR3 has been found in many tissues including brain, lung epithelia, diaphragm, and smooth muscle [29]. All three isoforms have been detected in the nervous system [18].

The IP_3R is a calcium release channel that is activated by both cytosolic inositol 1,4,5-trisphosphate and cytosolic calcium and is inhibited by very high cytosolic calcium levels. The three isoforms of the IP_3R are called Type I, Type II, and Type III. Type I is found in neurons and is dramatically enriched in Purkinje Neurons [45]. Type II is found in glia cells, but not neurons [45]. Type III is found in neurons, but not glia [45].

The modelling of the calcium release channels (IP_3R and RyR) uses the same formalism used for membrane ionic channels. In this case, the flux through the channel is the product of the channel permeability (or conductance) and the driving force. The channel permeability is typically the product of the channel open probability (Popen) and some maximal permeability $\bar{G}$.

$$J_{\text{release}} = \bar{G}P_{\text{open}}([Ca_{ER}] - [Ca_i]) \tag{3.7}$$

The open probability (P_{open}) of a channel can be determined in a number of ways: 1) An empirical expression of its dependence on allosteric regulators such as calcium can be used; 2) A Hodgkin-Huxley type formulation of gating variables can be implemented; or 3) A Markov state model of the channel can be constructed and where Popen would be the fraction of channels in the open state. With the Markov state model either a deterministic or stochastic approach can be implemented.

The SR sequesters cytosolic calcium through the sarcoplasmic and endoplasmic reticulum calcium ATP-ase (SERCA). SERCA consumes ATP and pumps calcium into the ER (or SR) against a concentration gradient. There are different isoforms of SERCA found in different tissues. For example, SERCA2 and SERCA3 are found in Purkinje neurons, SERCA1a is found in skeletal muscle, and SERCA2b is found in heart muscle [6, 35]. The action of SERCA has typically been modeled as a saturating pump with sigmoid kinetics.

$$J_{\text{SERCA}} = \frac{V_{\text{max}}[Ca_i]^2}{K_M^2 + [Ca_i]^2} \tag{3.8}$$

where K_M, the calcium dissociation constant, is the calcium concentration at which the pump rate is half the maximum pump rate (V_{max}). This representation of the pump requires a leak term to ensure calcium homeostasis at resting intracellular calcium. The leak is typically formulated as a passive leak

$$J_{\text{leak}} = k_{\text{leak}}([Ca_{ER}] - [Ca_i]) \tag{3.9}$$

where k_{leak} is the leak rate constant.

More recent studies of the SERCA pump in cardiac cells have suggested that part of the leak out of the SR is partially due to a backward flux through the SERCA pump [44]. This can be represented by

$$J_{\text{SERCA}} = \frac{V_{\text{maxf}}\left(\frac{[Ca_i]}{K_{mf}}\right)^{hf} - V_{\text{maxb}}\left(\frac{[Ca_i]}{K_{mb}}\right)^{hb}}{1+\left(\frac{[Ca_i]}{K_{mf}}\right)^{hf} + V_{\text{maxb}}\left(\frac{[Ca_i]}{K_{mb}}\right)^{hb}} + k_{\text{leak}}([Ca_{ER}]-[Ca_i]) \quad (3.10)$$

The first term in the numerator of Eq. (3.10) describes the forward rate of the pump with maximal rate V_{maxf} , binding constant K_{mf}, and cooperativity k_f. The second term in the numerator describes the backward rate of the pump with maximal rate V_{maxb} , binding constant K_{mb}, and cooperativity k_b. The second term in Eq. (3.10) describes the passive leak out of the SR with k_{leak} being the leak rate constant. Calcium in the SR lumen is buffered by the large amounts of low affinity calcium binding proteins resulting in a large calcium reserve in the SR. Typical SR calcium binding proteins are calsequestrin and calreticulin. They can be modeled in the same way at the cytosolic calcium binding proteins.

3.2.2 Calcium channels

In addition to the calcium channels found in the ER, there are different types of calcium channels that allow calcium entry across the cell membrane, namely, the L-type calcium channel, the N-type calcium channel, P/Q-type calcium channels, the NMDA (N-methyl-D-aspartate) receptor, and store operated channels. The L-, P/Q- and N-type calcium channels are voltage gated calcium channels, i.e. they open when the cell membrane depolarizes. The P/Q-type calcium channels deactivate with time. The N-type and L-type calcium channels not only deactivate with time, but undergo calcium-dependent inactivation. The NMDA receptor is a ligand gated pore, activated by glutamate, in which a voltage dependent block by magnesium is relieved during depolarization [21, 44].

3.2.3 Calcium pumps and exchangers

In order for calcium homeostasis to occur, calcium must be removed from the cell in an amount equal to that which enters. Thus, if calcium entry through channels occurs, there must be calcium extrusion mechanisms such as calcium pumps and exchangers. The main calcium pump in the plasmalemmal is the plasmalemmal calcium ATPase (PMCA). The PMCA hydrolyzes ATP to pump calcium up a concentration gradient. The typical formulation for this flux is

$$J_{\text{PMCA}} = \frac{V_{\text{max}}[Ca_i]^2}{K_M^2+[Ca_i]^2} \quad (3.11)$$

where V_{max} is the maximal pump rate and K_M is the calcium dissociation constant for the pump or the calcium concentration at which the pump is working at half its maximal rate.

Another set of calcium extrusion mechanism is the calcium exchangers. One such exchanger is the $Na^+ - Ca^{2+}$ exchanger. This pump exchanges three sodium ions for one calcium ion making it electrogenic in nature, i.e., it carries a current across the plasma membrane and must be included in any equation for membrane potential. In forward mode, the $Na^+ - Ca^{2+}$ exchanger uses the sodium gradient to bring in three sodium ions and extrude one calcium ion. However, this exchanger has a voltage dependence with the property that under depolarized conditions it can bring in one calcium ion and extrude three sodium ions in what is termed reverse mode. This current has been formulated by Luo and Rudy [28] for cardiac myocytes as

$$\begin{aligned} I_{\text{NaCaX}} = {} & k_{\text{NaCa}} \frac{1}{K_{m,Na}^3 + [Na_0]^3} \frac{1}{K_{m,Ca} + [Ca_i]} \\ & \cdot \left\{ \exp\left(\eta \frac{VF}{RT}\right) [Na_i]^3 [Ca_0] - \exp\left((\eta - 1)\frac{VF}{RT}\right) [Na_0]^3 [Ca_i] \right\} \\ & \cdot \frac{1}{1 + k_{\text{sat}} \exp((\eta - 1)VF/(RT))} \end{aligned} \tag{3.12}$$

where k_{NaCa} is the scaling factor for the current, k_{sat} is the saturation factor of the current at very low potentials, V is membrane potential, F is Faraday's constant, R is the ideal gas constant, T is the absolute temperature, $K_{m,Na}$ is the dissociation constant for external sodium, $K_{m,Ca}$ is the dissociation constant for internal calcium, and η is the position of the energy barrier controlling voltage dependence of the current.

3.2.4 Mitochondrial calcium

The mitochondria also regulate calcium in the cell. They sequester, store, and release calcium and thus are in effect a calcium store (Figure 3.3). A large electrochemical potential is maintained across the inner mitochondrial membrane mainly through the respiration driven proton pumps. The uniporter (J_{uni}) uses this energy gradient to move calcium into the mitochondria. Calcium in the mitochondria is buffered through precipitation when it combines with inorganic phosphate to form calcium phosphate (J_{buffer}). Calcium is extruded from the mitochondrial by sodium-calcium exchange (J_{NaCaX}).

Mitochondrial calcium uptake is complicated by its rapid uptake mode in which calcium is sequestered very quickly if cytosolic calcium is high [9, 49]. This condition might occur at places where the mitochondria are near the ER calcium release channel [36, 38] or calcium influx through voltage-gated calcium channels [50]. Recently, Beutner and co-workers have found a ryanodine receptor in the mitochondria that when blocked with ryanodine, suppresses mitochondrial calcium uptake [8]. It is possible that this channel plays a role in rapid uptake of calcium.

Modelling of mitochondrial calcium dynamics can be the topic of a manuscript in itself. In the limited space here, it suffices to mention a few of the more recent

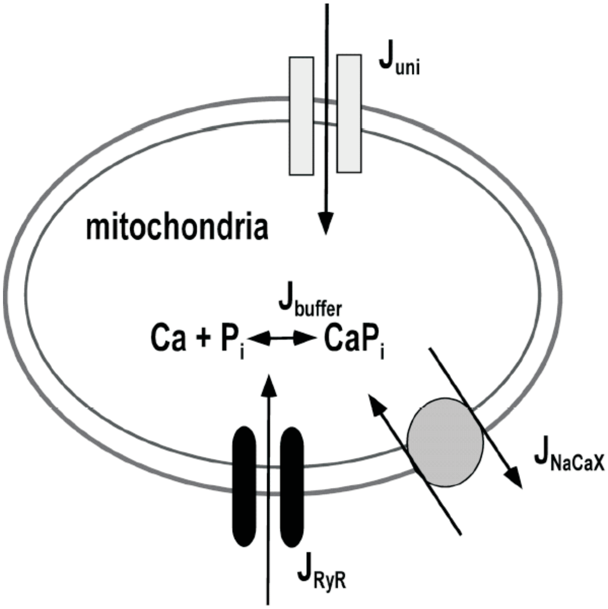

Figure 3.3

Schematic representation of mitochondrial calcium dynamics. Shown are the fluxes for calcium uptake by the uniporter (J_{uni}), calcium extrusion by $Na^+ - Ca^{2+}$ exchange ($J_{\mathbf{NaCaX}}$), and calcium buffering ($J_{\mathbf{buffer}}$) through precipitation with phosphate ($\mathrm{P}_{\mathbf{i}}$).

models in this area. Magnus and Keizer [30] proposed a model for mitochondrial calcium dynamics that included the processes that generated the membrane potential across the inner mitochondrial membrane since this is responsible for powering the calcium uniporter. This is a comprehensive model and has been incorporated into the pancreatic β cell [31, 32]. One of the key predictions of the model is that when cytosolic calcium rises, it increases mitochondrial calcium, which reduces the mitochondrial membrane potential resulting in decreased ATP production. More recently, this model has been incorporated by Fall and Keizer [16] into a model of calcium signaling to show how mitochondrial calcium dynamics affected calcium signaling. Depending on the parameters chosen, the DeYoung-Keizer model [13] can give oscillatory or bi-stable calcium dynamics. The addition of the mitochondrial model to a DeYoung-Keizer model [16] tuned to give bi-stable behavior results in a model with oscillatory calcium dynamics. Furthermore, the model predicts that increasing metabolism slows the frequency of calcium oscillations consistent with experiments [26].

This latter finding was also modeled by Falcke and co-workers [15]. They added a simplified model of uniporter and $Na^{+} - Ca^{2+}$ exchange to the Tang and Othmer model [51] for calcium signaling. They simulated energization of the mitochondria by increasing the maximal calcium uptake rate for the uniporter.

Finally, a recent model for the effects of mitochondrial calcium dynamics on cellular calcium signaling was developed for sympathetic neurons by Colegrove and co-workers [10]. Once again, simple formulations for mitochondrial uptake and release were implemented into a model for neuronal calcium dynamics. The model suggested that the mitochondria will accumulate calcium even under low amplitude fluctuations of cytosolic calcium and that the impact of mitochondrial calcium dynamics on cytosolic calcium is influenced greatly by non-mitochondrial calcium handling mechanisms. Furthermore, the model predicted that the buffering and non-buffering modes of mitochondrial calcium dynamics correspond to two different calcium signaling regimes.

3.3 Special calcium signaling for neurons

Thus far, the mechanisms of calcium signaling described are general and can be applied to many different cell types. Neurons are specialized cells that are varied in their function and morphology. This results in many calcium handling features designed to perform specific functions. Although some of them functions might be specific to neurons, the constructs to describe these are often found in other cells. These features occur in different parts of the neuron, namely, the soma, nerve terminal, dendrites and dendritic spines, and axon. In the next section, three specific calcium signaling mechanisms will be discussed: local calcium signaling, the control of gene expression by calcium, and cross-talk between channels mediated by calcium.

3.3.1 Local domain calcium

Local calcium domains have been demonstrated to be crucial to neuronal function. For example, the secretion of neurotransmitter in nerve terminal is critically dependent on the activation of specific voltage gated calcium channels, but not on release from internal stores [4, 5]. In PC 12 cells (rat pheochromocytoma cells), calcium entry across the plasma membrane through voltage gated channels is essential the secretion of catecholamines [3]. Calcium release from internal stores does not trigger secretion of catecholamines even in the presence of membrane depolarization in calcium free medium. This suggests that the elevation of calcium close to the plasma membrane calcium channels is essential for secretion.

The requirement of calcium entry might also depend on the specific voltage gated calcium channels involved. In mouse or neonatal rat motor nerve terminals, experiments have indicated that neurotransmitter release is activated by the opening of P/Q-type calcium channels but not by either L- or N-type calcium channels [42, 52]. Not only is calcium necessary for synaptic vesicle release, there is also evidence that elevated calcium in the nerve terminals is also necessary for the synaptic vesicle endocytosis also [11].

To model this, one must consider local domains of elevated calcium (Figure 3.4). In the discussion above, the vesicles respond to calcium local to specific voltage gated calcium channels in the plasma membrane. Bulk elevations of calcium do not activate vesicle exocytosis, but the high local calcium that occurs during plasma membrane voltage gated channels does. This suggests voltage gated channels are in close proximity to the vesicles as depicted in the figure as P/Q type channels. Other voltage gated channel types (N-type) or release from internal stores is more distant and does not activate vesicle fusion.

This has been modeled by Bertram and co-workers [7]. In their model, they explored the effect of overlapping calcium microdomains in activating vesicle fusion. They used a deterministic set of reaction-diffusion equations to describe the system. They concluded that calcium current cooperativity increases with the number of channels in the release site. Furthermore, they found that this increase is much less than the increase in the number of channels, giving an upper bound on the increase in cooperativity. Another interesting prediction was that the calcium channel cooperativity was an increasing function of channel distance.

Another model describing transmitter release in the mammalian CNS was proposed by Meinrenken and co-workers [33]. In this work, the effect of the spatial distribution of clusters of voltage-gated calcium channels on vesicle release was explored. In this model, vesicles at different locations are exposed to different calcium concentrations resulting in different release probabilities. The authors suggest that this spatially heterogeneous release probability has functional advantages for synaptic transmission.

A third model of calcium dynamics in the synapse of the frog saccular hair cell has been proposed by Roberts [40, 41]. In this model, an array of calcium channels

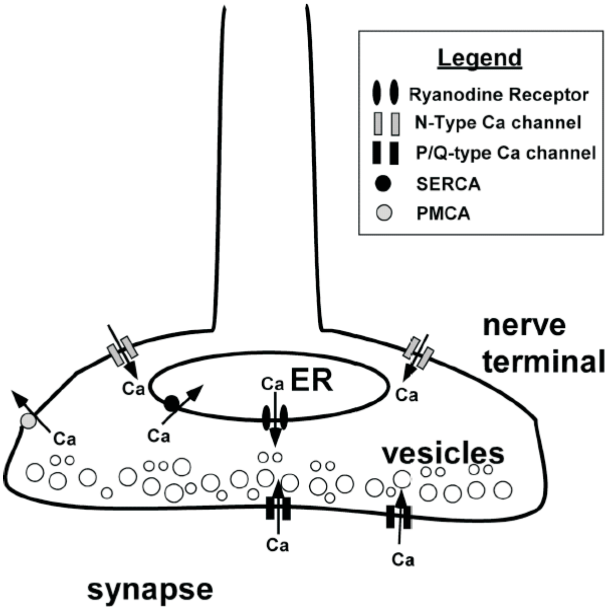

Figure 3.4

Schematic representation of calcium dynamics in the nerve terminal. Synaptic vesicles are found near the synaptic cleft. Voltage gated calcium channels (P/Q-type) are located near the vesicle so that their activation leads to high local calcium near the vesicle initiating vesicle fusion. Other calcium entry (N-type channels) and release from the ER through ryanodine receptors will elevate nerve terminal calcium, but not activate vesicle fusion. Also shown are the pumps that return calcium to resting levels.

and calcium activated potassium channels is modeled along with the endogenous mobile buffer calbindin. Also, the calbindin captures calcium microseconds after it enters the cells and carries it away from the channel mouth. This in effect, causes calcium to quickly reach a steady state level near one or more open channels. Furthermore, it restricts the area in which calcium is elevated, i.e., it causes calcium to fall off more steeply as the distance from the channel increases. It also showed that a calcium binding molecule with a calcium dissociation constant and diffusional properties similar to calbindin is necessary to simulate the experimental results.

3.3.2 Cross-talk between channels

Another area in which local domains of calcium is important is the communication or cross-talk between ion channels (Figure 3.5). In such situations, ion channels are located in close proximity so that calcium entry through one channel causes high local calcium concentrations that can act on other nearby channels of the same or different types. Sometimes this will occur in a subspace bounded by cell membranes and organelles, and other times not. The calcium concentration in the local domains can easily be 100 times that of the bulk cytoplasm. In Figure 3.5, calcium-activated chloride channels are located in close proximity to voltage gated calcium channels. The ryanodine receptors in the ER are located at some slightly further distance. Under highly buffered, low level activation of the voltage gated channels might activate the chloride channels but not the ryanodine receptors as observed by Ward and coworkers [55].

This has been modeled extensively in the area of excitation-contraction (EC) coupling in cardiac cells on the cellular level [24]. In EC coupling, opening of voltage-gated L-type calcium channels in the cell membrane allows calcium entry that triggers release from internal stores via the ryanodine receptor. This process is termed calcium-induced calcium release. In this system, the L-type calcium channel and the ryanodine receptors are situated in a membrane restricted subspace with only 12-15 nm between the membranes containing the two types of channels. In these cellular models, the domain is treated as a small compartment that contains ion channels and buffers, and communicates with the bulk cytoplasm. The group behavior of these ion channels is modeled, which smooths the rapid changes of calcium fluxes due to channel opening. In these models, since the time scale of calcium dynamics is much slower than the kinetics of the buffers, the rapid buffering approximation should be used.

Another approach is to model the details of this compartment in cardiac cells [39, 48]. These models also contain ion channels and buffers, and also communicate with the bulk myoplasm. However, they are generally stochastic to simulate ion channel dynamics and do not include whole cell calcium dynamics. Since a small number of channels is modeled, there are rapid changes in the calcium fluxes. This results in large time dependent changes in calcium requiring a small time step. This necessitates that the buffering equations be solved dynamically rather than with a steady-state approximation.

Calcium induced calcium release has also been observed in neurons [1, 22, 55].

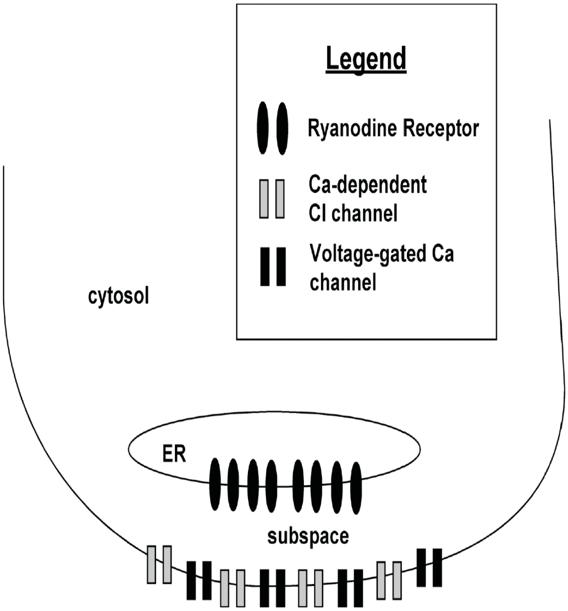

Figure 3.5

Schematic representation of one possible local calcium signaling scenario. Close association of the ER and cell membrane form a subspace where ion channel cross-talk can occur. Here, calcium entry through the voltage gated calcium channels can trigger adjacent calcium-dependent chloride channels or farther ryanodine receptors.

In these systems calcium entry through voltage gated calcium channels activates calcium release from ryanodine receptors located in the ER. A model to describe this phenomenon can use the same principles and formulations used in the work on cardiac cells described above. In fact, the work of Albrecht and co-workers [1] presents a model to explain their experimental results. They construct a simple model that includes two dynamic equations for the calcium concentrations in the ER and cytosol, RyR calcium release, SERCA pumps, and calcium entry and extrusion across the plasma membrane. Their model demonstrates that there can be a low-gain mode of CICR that operates under weak stimulation and a high-gain mode of CICR that operates at high cytosolic calcium.

3.3.3 Control of gene expression

Calcium has long been thought to play a role in controlling gene expression. Early research in this area suggested that calcium elevations and oscillation could control gene expression through frequency encoding [19]. Concrete evidence of frequency encoding of gene expression activation by calcium oscillations was shown in T-lymphocytes concurrently by Dolmetsch and co-workers [14] and Tsien and co-workers [27]. A viable mechanism for this frequency encoding of gene expression has been proposed in a computational model by Fisher and co-workers [17] that involves the activation of calcium dependent transcription factors, NFAT and NF_kB, and their translocation into the nucleus (Figure 3.6). At rest NFAT is located in the cytoplasm and is phosphorylated (Figure 3.6 in oval). The activation of NFAT involves the activation calcineurin (C), a calcium/calmodulin-dependent phosphatase, which binds to NFAT. Calcineurin then dephosporylates NFAT, allowing its translocation into the nucleus. A similar set of reactions can occur in the nucleus. The nuclear form of dephosphorylated NFAT is the transcriptionally active form and is shown in the box in Figure 3.6. The model describes the different steps in these biochemical pathways and derived differential equations using the laws of mass action applied to these pathways. Translocation of the transcription factors across the nuclear membrane is also treated as a biochemical reaction. The rate constants can be determined by experimentally determined equilibrium constants and reaction half times.

Activity-dependent gene expression has also been observed in neurons. Neurons are able to differential between different types of stimulus, i.e., inputs that engage synaptic transmission are much more effective that inputs that do not [34]. Experiments in hippocampal neurons have shown that the calcium-dependent activation of calcineurin (and hence NFAT) is critically dependent on calcium entry through L-type calcium channels [20]. This again involves local calcium signaling in the nerve terminals that can have global effects. This synapse-to-nucleus signaling that leads to gene expression might play a role in synaptic plasticity and memory [12, 20]. A similar approach as presented in the work by Fisher and co-workers as described above would provide a good way to describe this phenomenon.

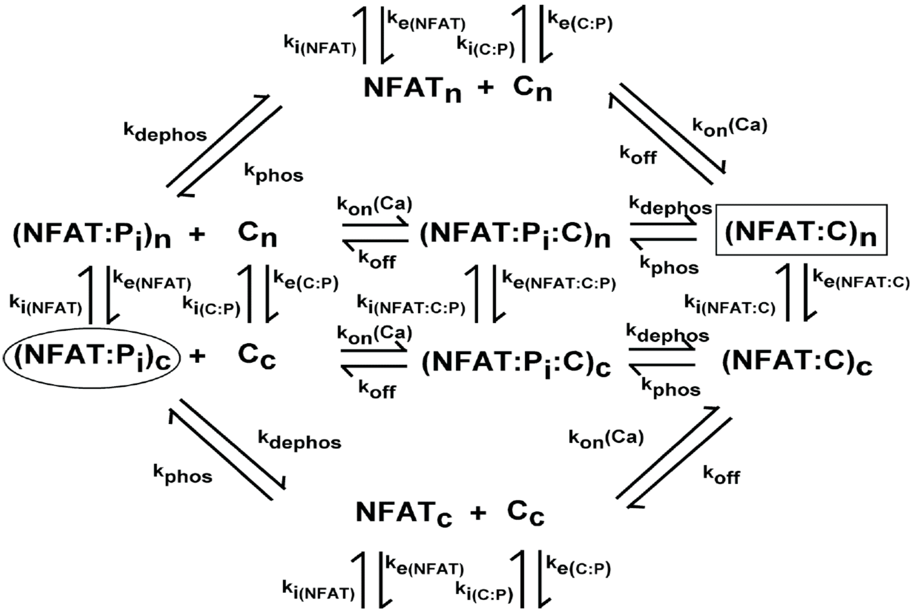

Figure 3.6

The biochemical reaction scheme for activation of the calcium-dependent transcription factor calcineurin. The top half of the figure described nuclear reactions (subscript n) and the bottom half describes cytosolic reactions (subscript c). At rest, NFAT is phosphorylated. Upon activation by calcium, calcineurin (C) binds NFAT and dephsophorylates by allowing its translocation into the nucleus. The transcriptionally active form of NFAT is dephosphorylated and nuclear and is enclosed in a box. The resting form is the phosphorylated cytoplasmic state, which is enclosed in an oval.

3.4 Conclusions

Calcium dynamics is a complex non-linear phenomenon that has benefited greatly from mathematical modelling. The basic mechanisms of calcium signaling, such as buffering, the endoplasmic reticulum, plasma membrane channels and exchangers, and mitochondria can be modeled using very simple or biophysically detailed representations. These can be combined to describe the complex morphologies and structures that give rise to calcium signaling in neurons. A few of these structures have been discussed such as vesicle fusion in the nerve terminal and submembrane spaces. Calcium also can play a role in complex biochemical signaling pathways such as those controlling gene expression.

Modelling efforts such as these have contributed to the fundamental understanding of cellular function. With the models, hypotheses about the mechanisms can be tested making way for new experiments to test predictions of the models. The models can then be further refined to reflect the new experimental data. In this fashion, continued interplay between modelling and experimental science will lead to greater advances in the study of neuronal function.

References

[1] Albrecht M.A., Colegrove S.L., Hongpaisan J., Pivovarova N.B., Andrews S.B., and Friel D.D. (2001). Multiple Modes of Calcium-induced Calcium Release in Sympathetic Neurons I: Attenuation of Endoplasmic Reticulum Ca^{2+} Accumulation at Low $[Ca^{2+}]_i$ during Weak Depolarization. *J. Gen. Physiol.* **118**:83-100.

[2] Allbritton N.L., Meyer T., and Stryer L. (1992). Range of Messenger Action of Calcium Ion and Inositol 1,4,5-trisphosphate. *Science*. **258**:1812-1815.

[3] Ashery T., Weiss C., Sela D., Spira M.E., and Atlas D. (1993). Membrane Depolarization Combined with Release of Calcium from Internal Stores does not Trigger Secretion from PC 12 Cells. *Receptors Channels.* **1**:217-220.

[4] Atlas D. (2001). Functional and Physical Coupling of Voltage-sensitive Calcium Channels with Exocytotic Proteins: Ramifications for the Secretion Mechanism. *J. Neurochem.* **77**:972-985.

[5] Atlas D., Wiser O., and Trus M. (2001). The Voltage-gated Ca^{2+} Channel is the Ca^{2+} Sensor of Fast Neurotransmitter Release. *Cell. Mol. Neurobiol.* **21**:171-731.

[6] Baba-Aissa F., Raeymaekers L., Wuytack F., Callewaert, G., Dode, L., et al. (1996). Purkinje Neurons Express the SERCA3 Isoform of the Organellar Type

Ca^{2+}-transport ATPase. *Brain Res. Mol. Brain Res.* **41**:169-174.

[7] Bertram R., Smith G.D., and Sherman A. (1999). Modeling Study of the Effects of Overlapping Ca^{2+} Microdomains on Neurotransmitter Release. *Biophys. J.* **76**:735-750.

[8] Beutner G., Sharma V.K., Giovannucci D.R., Yule D.I., and Sheu S-S. (2001). Identification of a Ryanodine Receptor in Rat Heart Mitochondria. *J. Biol. Chem.* **276**:21482-21488.

[9] Buntinas L., Gunter K.K., Sparagna G.C., and Gunter T.E. (2001). The Rapid Mode of Calcium Uptake into Heart Mitochondria (RaM): Comparison to RaM in Liver Mitochondria. *Biochim. Biophys. Acta.* **1504**:248-261.

[10] Colegrove S.L., Albrecht M.A., and Friel D.D. (2000). Quantitative Analysis of Mitochondrial Ca^{2+} Uptake and Release Pathways in Sympathetic Neurons. *J. Gen. Physiol.* **115**:371-388.

[11] Cousin M.A. (2000). Synaptic Vesicle Endocytosis: Calcium Works Overtime in the Nerve Terminal. *Mol. Neurobiol.* **22**:115-128.

[12] Deisseroth K., and Tsien R.W. (2002). Dynamic Multiphosphorylation Passwords for Activity-dependent Gene Expression. *Neuron.* **34**:179-182.

[13] DeYoung G.W., and Keizer J. (1992). A Single-pool Inositol 1,4,5-trisphosphate-receptor-based Model for Agonist-stimulated Oscillations in Ca^{2+} Concentration. *Proc. Natl. Acad. Sci., USA.* **89**:9895-9899.

[14] Dolmetsch R.E., Xu K., and Lewis R.S. (1998). Calcium Oscillations Increase the Efficiency and Specificity of Gene Expression. *Nature.* **392**:933-936.

[15] Falcke M., Hudson J.L., Camacho P., and Lechleiter J.D. (1999). Impact of Mitochondrial Ca^{2+} Cycling on Pattern Formation and Stability. *Biophys. J.* **77**:37-44.

[16] Fall C.P., and Keizer J.E. (2001). Mitochondrial Modulation of Intracellular Signaling. *J. Theor. Biol.* **210**:151-165.

[17] Fisher W.G., Medikonduri R.K., and Jafri M.S. (2003). Models for NFAT and NF B Activation in T-lymphocytes. *Biophys. J.* (abstract, 1891).

[18] Giannini G., Conti A., Mammarella S., Scrobogna M., and Sorrentino V. (1995). The Ryanodine Receptor/Calcium Channel Genes are Widely and Differentially Expressed in Murine Brain and Peripheral Tissues. *J. Cell. Biol.* **128**:893-904.

[19] Golbeter A., Dupont G., and Berridge M.J. (1990). Minimal Model for Signal-induced Ca^{2+} Oscillations and for their Frequency Encoding through Protein Phosphorylation. *Proc, Natl. Acad. Sci., USA.* **87**:1461-1465.

[20] Graef I.A., Mermelstein P.G., Stankunas K., Neilson J.R., Deisseroth K., Tsien R.W., et al. (1999). L–type Calcium Channels and GSK-3 Regulate the Activity of NF-ATc4 in Hippocampal Neurons. *Nature.* **401**:703-708.

[21] Hille B. (1992). *Ionic Channels of Excitable Membranes.* 2nd ed. Sunderland, MA: Sinauer Associates, Inc.

[22] Hongpaisan J., Pivovarova N.B., Colegrove S.L., Leapman R.D., Friel D.D., and Andrews S.B. (2001). Multiple Modes of Calcium-induced Calcium Release in Sympathetic Neurons II: A $[Ca^{2+}]_i$- and Location-dependent Transition from Endoplasmic Reticulum Ca Accumulation to Net Ca Release. *J. Gen. Physiol.* **118**:101-112.

[23] Jafri M.S., and Keizer J. (1995). On the Roles of Ca^{2+} Diffusion, Ca^{2+} Buffers, and the Endoplasmic Reticulum in IP3-induced Ca^{2+} Waves. *Biophys. J.* **69**:2139-2153.

[24] Jafri M.S., Rice J.J., and Winslow R.L. (1998). Cardiac Calcium Dynamics: the Role of Ryanodine Receptor Adaptation and Sarcoplasmic Reticulum Ca^{2+} Load. *Biophys. J.* **74**.

[25] Jinno S., and Kosaka T. (2002). Patterns of Expression of Calcium Binding Proteins and Neuronal Nitric Oxide Synthase in Different Populations of Hippocampal BAGAergic Neurons in Mice. *J. Comp. Neurol.* **449**:1-25.

[26] Jouaville L.S., Ichas F., Holmuhamedov E.L., Camacho P., and Lechleiter J.D. (1995). Synchronization of Calcium Waves by Mitochondrial Substrates in Xenopus Oocytes. *Nature.* **377**:438-441.

[27] Li W., Llopis J., Whitney M., Zlokarnik G., and Tsien R.Y. (1998). Cell-permeant Caged IP3 Ester Shows that Ca^{2+} Spike Frequency Can Optimize Gene Expression. *Nature.* **392**:863-866.

[28] Luo C-H, and Rudy Y. (1994). A Dynamic Model of the Cardiac Ventricular Action Potential. *Circ. Res.* **74**:1097-1113.

[29] MacKrill J.J., Challiss R.A.J., O'Connell D.A., Lai F.A., and Nahorski S.R. (1997). Differential Expression and Regulation of Ryanodine Receptor and Myo-inositol 1,4,5-trisphosphate Receptor Ca^{2+} Release Channels in Mammalian Tissues and Cell Lines. *Biochem. J.* **327**:251-258.

[30] Magnus G., and Keizer J. (1997). Minimal Model of Beta-cell Mitochondrial Ca^{2+} Handling. *Am. J. Physiol.* **273**:C717-C732.

[31] Magnus G., and Keizer J. (1998). Model of Beta-cell Mitochondrial Calcium Handling and Electrical Activity II. Mitochondrial Variables. *Am. J. Physiol.* **274**:C1174-C1184.

[32] Magnus G., and Keizer J. (1998). Model of Beta-cell Mitochondrial Calcium Handling and Electrical Activity II. Cytoplasmic Variables. *Am. J. Physiol.* **274**:C1158-C1173.

[33] Meinrenken C.J., Borst J.G., and Sakmann B. (2002). Calcium Secretion Coupling at Calyx of Held Governed by Nonuniform Channel-vesicle Topography. *J. Neurosci.* **22**:1648-1667.

[34] Mermelstein P.G., Bito H., Deisseroth K., and Tsien R.W. (2000). Critical Dependence of cAMP Response Element-Binding Protein Phosphorylation of L-type Calcium Channels Supports a Selective Response to EPSPs in Preference to Action Potentials. *J. Neurosci.* **21**:266-273.

[35] Misquitta C.M., Mack D.P., and Grover A.K. (1999). Sarco/endoplasmic Reticulum Ca^{2+} (SERCA)-pumps: Link to Heart Beats and Calcium Waves. *Cell Calcium.* **25**:277-290.

[36] Merchant J.S., Ramos V., and Parker I. (2002). Structural and Functional Relationships between Ca^{2+} Puffs and Mitochondria in Xenopus Oocytes. *Am. J. Physiol. Cell Physiol.* **282**:C1374-C1386.

[37] Neher E. (1986) Concentration Profiles of Intracellular Calcium in the Presence of a Diffusible Chelator. In: Heinemann U, Klee M, Neher E, Singer W, editors. *Calcium Electrogenesis and Neuronal Functioning.* Berlin: Springer-Verlag; p. 80-96.

[38] Pacher P., Csordas P., Schneider T., and Hajnoczky G. (2000). Quantification of Calcium Signal Transmission from Sarco-endoplasmic Reticulum to the Mitochondria. *J. Physiol.* **529**:553-564.

[39] Rice J.J., Jafri M.S., and Winslow R.L. (1999). Modeling Gain and Gradedness of Ca^{2+} Release in the Functional Unit of the Cardiac Diadic Space. *Biophys. J.* **77**:1871-1884.

[40] Roberts W.M. (1993). Spatial Calcium Buffering in Saccular Hair Cells. *Nature.* **363**:74-76.

[41] Roberts W.M. (1994). Localization of Calcium Signals by a Mobile Calcium Buffer in Frog Saccular Hair Cells. *J. Neurosci.* **14**:3246-3262.

[42] Rosato-Siri M.D., Piriz J., Tropper B.A., and Uchitel O.D. (2002). Differential Ca^{2+}-dependence of Transmitter Release Mediated by P/Q- and N-type Calcium Channels at Neonatal Rat Neuromuscular Junctions. *Eur. J. Neurosci.* **15**:1874-1880.

[43] Ross C.A., Meldosi J., Milner T.A., Satoh T., Suppattapone S., and Snyder S.H. (1989). Inositol 1,4,5-trisphosphate Receptor Localized to Endoplasmic Reticulum in Cerebellar Purkinje Neurons. *Nature.* **339**:468-470.

[44] Shannon T.R., Ginsburg K.S., and Bers D.M. (2000). Potentiation of Fractional Sarcoplasmic Reticulum Calcium Release by Total and Free Intra-sarcoplasmic Retuciulum Calcium Concentration. *Biophys. J.* **78**:334-343.

[45] Sharp A.H., Nucifora F.C.J., Blondel O., Sheppard C.A., Zhang C., Snyder S.H., et al. (1999). Differential Cellular Expression of Isoforms of Inositol 1,4,5-trisphosphate in Neurons and Glia in Brain. *J. Comp. Neurol.* **406**:207-220.

[46] Smith G.D., Dai L., Miura R., and Sherman A. (2001). Asymptotic Analysis of Buffered Ca^{2+} Diffusion Near a Point Source. *SIAM J. Applied Math.*

61:1816-1838.

[47] Slepecky N.B., and Ulfendahl M. (1993). Evidence for Calcium-binding Proteins and Calcium-dependent Regulatory Proteins in Sensory Cells of the Organ of Corti. *Hear. Res.* **70**:73-84.

[48] Soeller C., and Cannell M.B. (1997). Numerical Solution of Local Calcium Movements during L-Type Calcium Channel Gating in the Cardiac Diad. *Biophys. J.* **73**:97-111.

[49] Sparagna G.C., Gunter K.K., Sheu S-S, and Gunter T.E. (1995). Mitochondrial Calcium Uptake from Physiological-type Pulses of Calcium. *J. Biol. Chem.* **270**:27510-27515.

[50] Svichar N., Kostyuk P., and Verkhratsky A. (1997). Mitochondria Buffer Ca^{2+} Entry but Not Intracellular Ca^{2+} Release in Mouse DRG Neurones. *Neuroreport.* **8:**3929-3932.

[51] Tang Y., and Othmer H.G. (1995). Frequency Encoding in Excitable Systems with Applications to Calcium Oscillations. *Proc. Natl. Acad. Sci., USA.* **92**:7869-7873.

[52] Urbano F.J., Depetris R.S., and Uchitel O.D. (2001). Coupling of L-type Calcium Channels to Neurotransmitter Release at Mouse Motor Nerve Terminals. *Pflugers Arch.* **441**:824-831.

[53] Wagner J., and Keizer J. (1994). Effects of Rapid Buffers on Ca^{2+} Diffusion and Ca^{2+} Oscillations. *Biophys. J.* **67**.

[54] Walton P.D., Airey J.A., Sutko J.L., Beck C.F., Mignery G.A., Sudhof T.C/, et al. 1991. Ryanodine and Inositol Trisphosphate Receptors Coexist in Avian Purkinje Neurons. *J. Cell Biol.* **113**:1145-1157.

[55] Ward S.M., and Kenyon J.L. (2000). The Spatial Relationship between Ca^{2+} Channels and Ca^{2+}-activated Channels and the Function of Ca^{2+}-buffering in Avian Sensory Neurons. *Cell Calcium.* **28**:233-246.

[56] Zhou Z., and Neher E. (1993). Mobile and Immobile Calcium Buffers in Bovine Adrenal Chromaffin Cells. *J. Physiol.* **469**:245-273.

Chapter 4

Structure-Based Models of NO Diffusion in the Nervous System

Andrew Philippides[1], Phil Husbands[2], Tom Smith[1], and Michael O'Shea[1]

Centre for Computational Neuroscience and Robotics (CCNR), [1]Department of Biology, [2]Department of Informatics, University of Sussex, Brighton, U.K.

CONTENTS

4.1 Introduction

While the transmission of electrical signals across neuronal networks is a fundamental aspect of the operation of nervous systems, and this feature has traditionally been the main focus of computational neuroscience [10, 23], neurochemistry adds many dimensions to the picture. For instance, it is now recognised that nitric oxide (NO) is a novel kind of neurotransmitter that acts, through diffusion, over volumes that potentially contain many neurons and can facilitate signalling between neurons that are not synaptically connected [14, 46, 47]. This chapter aims to demonstrate that computational and mathematical modelling have an important role to play in trying to understand this particularly interesting mode of signalling.

Traditionally, chemical signaling between nerve cells was thought to be mediated solely by messenger molecules or neurotransmitters which are released by neurons at synapses [22] and flow from the presynaptic to postsynaptic neuron. Because most neurotransmitters are relatively large and polar molecules (amino acids, amines and peptides), they cannot diffuse through cell membranes and do not spread far from the release site. They are also rapidly inactivated by various reactions. Together these features confine the spread of such neurotransmitters to be very close to the points of release and ensure that the transmitter action is transient. In other words, chemical synaptic transmission of the classical kind operates essentially two-dimensionally (one in space and one in time). This conventional interpretation is coupled to the idea that neurotransmitters cause either an increase or a decrease in the electrical excitability of the target neuron. According to a traditional view of neurotransmission therefore, chemical information transfer is limited to the points of connection between neurons and neurotransmitters can simply be regarded as either excitatory or inhibitory. In recent years a number of important discoveries have necessitated a fundamental revision of this model. It is now clear that many neurotransmitters, perhaps the majority, cannot be simply classified as excitatory or inhibitory [17]. These messenger molecules are best regarded as modulatory because among other things they regulate, or modulate, the actions of conventional transmitters. Modulatory neurotransmitters act in an indirect way by causing medium and long-term changes in the properties of neurons by influencing the rate of synthesis of so-called second messenger molecules. By altering the properties of proteins and even by changing the pattern of gene expression, these second messengers cause complex cascades of events resulting in fundamental changes in the properties of neurons. In this way modulatory transmitters greatly expand the diversity and the duration of actions mediated by the chemicals released by neurons.

However, when coupled with this expanded picture of the nervous system, it is the recent discovery that the gas nitric oxide is a modulatory neurotransmitter that has opened entirely unexpected dimensions in our thinking about neuronal chemical signaling [14, 15, 19]. Because NO is a very small and nonpolar molecule it diffuses isotropically in aqueous and lipid environments, such as the brain, regardless of intervening cellular structures [47]. NO therefore violates some of the key tenets of point-to-point chemical transmission and is the first known member of an entirely new class of transmitter, the gaseous diffusable modulators (CO and H_2S are the other two identified examples (see e.g., [4]. NO is generated in the brain by specialised neurons that contain the neuronal isoform of the calcium activated enzyme, nitric oxide synthase or nNOS [3]. NO synthesis is triggered when the calcium concentration in nNOS-containing neurons is elevated, either by electrical activity or by the action of other modulatory neurotransmitters. NO activates the synthesis of cyclic-GMP, an important second messenger which regulates a wide variety of cellular processes in target neurons, some of which underlie synaptic plasticity [19]. Hence NO is involved in many neuronal functions from visual processing to memory formation and blood flow regulation [16, 19, 46].

The existence of a freely diffusing modulatory transmitter suggests a radically different form of signalling in which the transmitter acts four-dimensionally in space and time, affecting volumes of the brain containing many neurons and synapses. NO cannot be contained by biological membranes, hence its release must be coupled directly to its synthesis. Because the synthetic enzyme nNOS can be distributed throughout the neuron, NO can be generated and released by the whole neuron. NO is therefore best regarded as a 'non-synaptic' transmitter whose actions moreover cannot be confined to neighbouring neurons [18, 33]. So not only can NO operate over a large region, it can also mediate long-lasting changes in the chemical and electrical properties of neurons within that volume [2, 41].

Because nNOS is a soluble enzyme and thus likely to be distributed throughout a neuron's cytoplasm, the whole neuron surface is a potential release site for NO. Thus the morphology of NO sources, as well as the presence of structured sinks (such as blood vessels), will have a major influence on the dynamics of NO spread. Understanding this dynamics is clearly a very important part of a more general understanding of volume signalling processes. However, because the NO molecule is so small and non-polar it is very difficult to gather accurate empirical data in this area. Therefore it is natural to turn to computational modelling to shed light on volume signalling.

Somewhat ironically, many of NO's characteristics that complicate its empirical investigation, make it much easier to model than many conventional neurotransmitters whose large size and polarity make them impermeable to cell membranes. Thus while these molecules also diffuse, their movement is restricted to the extracellular space near their release site and to model their spread would therefore require accurate modelling of the morphology of the extracellular space and any local inhomogeneities. In contrast, because of NO's minute size and non-polarity it can be assumed as a good first approximation to diffuse isotropically through most brain tissue and so the morphology of the synaptic cleft and other surrounding matter need not be modelled. This means that complex factors such as tortuosity and viscosity, which affect the movement of larger molecules, do not need to be included in the governing equations.

This chapter demonstrates how to model NO diffusion from continuous structures of biologically realistic dimensions. The central part of the chapter describes and justifies in some detail the methods used to build such models. It then goes on to show how these models provide insights into a number of salient functional questions that arise in the context of volume signalling. Chief among these is how large a volume can be affected, and for how long, from various NO generating neuronal structures. Finally, work on more abstract computational models of neural networks incorporating functionally active diffusing neuromodulators is introduced. These networks serve as the nervous system of autonomous robots, generating sensorimotor behaviours in these devices, and thus help to give insights into possible functional roles of gaseous diffusing modulators in real nervous systems.

4.2 Methods

This section gives a detailed overview of the methods used to model the diffusion of NO in the CNS.

4.2.1 Equations governing NO diffusion in the brain

> *The rate of change of concentration in a volume element of a membrane, within the diffusional field, is proportional to the rate of change of concentration gradient at that point in the field.* Fick's second law (Fick 1855)

The equations governing diffusive movement can be understood by considering the motion of individual molecules. In a dilute solution, each molecule behaves independently of the others as it rarely meets them, but is constantly undergoing collisions with solvent molecules which move it in random directions. Thus its path can be described as a random walk* resulting in a net transfer of molecules from high to low concentrations at a rate proportional to the concentration gradient. This process is captured by what is commonly known as Fick's first law, that in isotropic substances the rate of transfer of diffusing substance through unit area of a section is equal to the product of the diffusion coefficient, D, and the concentration gradient measured normal to the section [8]. While in some cases D depends on concentration, it can be taken to be constant for dilute solutions [8]. As this is the case for diffusion of NO in the brain [45], we will only consider these situations. Representing the concentration at a point $\mathbf{x}$ and time t as $C(\mathbf{x},t)$, the following equation for diffusion in the brain (Fick's second law) can then be derived from Fick's first law:

$$\frac{\partial C(\mathbf{x},t)}{\partial t} = D\left(\frac{\partial^2 C(\mathbf{x},t)}{\partial x^2} + \frac{\partial^2 C(\mathbf{x},t)}{\partial y^2} + \frac{\partial^2 C(\mathbf{x},t)}{\partial z^2}\right) \tag{4.1}$$

or, more generally:

$$\frac{\partial C(\mathbf{x},t)}{\partial t} = D\,\nabla^2 C(\mathbf{x},t) \tag{4.2}$$

While the above equations govern the diffusive element of NO's spread, they do not take into account its destruction. NO does not have a specific inactivating mechanism, and is lost through reaction with oxygen species and metals, as well as heme containing proteins [25, 44]. This means that the movement of other molecules and

*Diffusion processes are thus amenable to *Monte Carlo* methods, where a (in the case of diffusion) uniform probability distribution, representing the probability of a molecule moving in a given direction, together with a random number generator are used to calculate the path of each molecule. However, the relatively long running times to achieve a good approximation render this method inappropriate for our needs [1].

receptors and their interactions need not be modelled and instead a more general loss function can be used. Thus we have:

$$\frac{\partial C(\mathbf{x},t)}{\partial t} - D\,\nabla^2 C(\mathbf{x},t) = -L(C,\mathbf{x},t) \tag{4.3}$$

where the term on the right-hand side is the inactivation function. This function will be composed of a global component for general, background NO reactions, $L_1(C)$, and spatially localised components, $L_2(C,\mathbf{x},t)$, representing structures which act as local NO sinks such as blood vessels. The kinetics of these reactions are not understood perfectly [47], but empirical data indicates either first or second order decay [25, 27, 28, 43, 45], as represented by:

$$L_i(C,\mathbf{x},t) = k_i(\mathbf{x}) \times C(\mathbf{x},t)^n \tag{4.4}$$

where n is the order of the reaction and is equal to either 1 or 2, referred to as first or second decay order respectively, and k_i is the reaction rate, commonly given in terms of the half-life $t_{1/2} = ln(2)/k_i$. The reaction may also depend on the concentration of the oxidative substance (usually oxygen), but, apart from very special cases (as in [7], for example), this can be assumed to be constant and is left out of the equation as it is subsumed by the reaction rate constant.

Values for the half-life of NO have been determined empirically and are dependent on the chemical composition of the solvent within which NO is diffusing. For instance, the half-life of NO in the presence of haemoglobin (Hb) is reported as being between about $1ms$ and $1\mu s$, depending on the Hb concentration [23, 26, 28, 44]. In contrast, half-life values used for extravascular tissue, normally associated with background NO consumption, are more than 1000 times longer, ranging from 1 to $> 5s$ [29, 31, 43, 47].

The order of the reaction is also dependent on the nature of the diffusive environment. In environments where there is a high concentration of Hb, as in NO sinks, recent work by Liu et al. [28] has shown that the reaction of NO with intact red blood cells exhibits first order kinetics, which is in agreement with earlier measurements [23, 26, 44]. Similarly, for modelling the global part of the loss function, as would be seen in most extravascular regions of the brain, measurements of NO loss are also consistent with first order decay [25, 28, 43]. Although second order decay has been used for extravascular NO comsumption [27, 45], in these models NO is diffusing *in vitro* in an air-saturated aqueous solution which is molecular-oxygen rich (unlike intact extravascular brain tissue). Thus the dynamics of decay are taken from empirical data in molecular-oxygen rich environs and are unlike those in the intact brain. As we are concerned with modelling NO diffusion in the brain *in vivo*, we have therefore used first order decay to model global NO loss in extravascular tissue as well as in localised sinks. This gives us the following widely used equation [24, 25, 26, 43, 44, 47] for diffusion of NO in the brain:

$$\frac{\partial C(\mathbf{x},t)}{\partial t} - D\,\nabla^2 C(\mathbf{x},t) = -k(\mathbf{x})\; C(\mathbf{x},t) \tag{4.5}$$

referred to as the *modified diffusion equation*. A production term can also be added to the right-hand side of Equation 4.5 though this is often factored into the solution later via the initial conditions (see Section 4.2.2).

Under certain conditions and for some source morphologies, Equation 4.5 can be solved analytically (the analytical solution), although this usually involves some numerical integration. Since the more complex a system is, the more numerical integration is required, this approach is often impractical and, in general, radial symmetry is required for tractability. If the analytical solution cannot be derived, a numerical approximation method must be used [26]. That is not to say that the numerical solutions are somehow ‘worse’ than the analytic ones or that they are simply crude approximations to the true solution [39]. Rather, they can usually be made as accurate as desired, or as accurate as the situation warrants given the unavoidable errors in empirical measurements of diffusion parameters. Indeed, as all the analytic solutions presented here required numerical integration they are also approximate and all results have been derived to the same degree of accuracy. However, a numerical approximation is always an approximation to the analytical solution and so it seems sensible to use the latter if its calculation is tractable. A more practical reason for doing so is that when it is available, evaluating the analytical solution normally requires much less computational power. Our approach therefore is to use the analytical solution whenever possible and, when not, to employ finite difference methods. In the next two sections, we discuss these techniques.

4.2.2 Analytic solutions to the diffusion equation

> *God does not care about our mathematical difficulties. He integrates empirically.* Albert Einstein.

In this section we discuss how analytic solutions to the diffusion equation are generated. The solution for a point-source is given first and we then show how solutions for other simple structures are derived from this. We next state the solutions thus obtained for hollow spherical and tubular sources and finally give some details of the numerical integration techniques used to calculate these solutions.

4.2.2.1 Modelling NO diffusion from a point-source

As stated earlier (Section 4.2.1), the dynamics of diffusion are governed by the modified diffusion equation. Assuming that there are no local NO sinks present and only global decay is acting, this equation becomes:

$$\frac{\partial C}{\partial t} - D\nabla^2 C = -\lambda C \tag{4.6}$$

where C is concentration, D is the diffusion coefficient and λ the decay-rate [8]. We first generate the *instantaneous solution*, that is, the solution for an instantaneous burst of synthesis from a point source positioned at the origin of some co-ordinate system. To do this we envision an amount S_0 of NO being deposited instantaneously at the origin at time $t = 0$. We then solve the diffusion equation under this initial condition, which gives us the following equation describing the evolution of the concentration of NO from a point [8]:

$$C_P(r,t) = \frac{S_0}{8(\pi Dt)^{3/2}} exp\left(\frac{-r^2}{4Dt}\right) e^{-\lambda t} \tag{4.7}$$

where $C_P(r,t)$ is the concentration of NO at time t at a point (r,θ,ϕ), defined in a spherical polar coordinate system. Note, however, that C_P depends only on time and the distance from the point source, r, as the system is radially symmetric.

The solution for a point-source which emits NO continuously, the *continuous solution*, is derived from the instantaneous solution for a point source, described earlier, in the natural way, via the principle of superposition of linear solutions [8]. First we define the 'strength' of a source to be its rate of NO production. Next we define the concentration at time t' and distance r from the origin, due to an instantaneous source of unit strength to be $f(r,t')$. Thus, if a source emits NO continuously at a rate governed by $S(t)$, we have:

$$C(r,t) = \int_0^t S(t-t') f\left(r,t'\right) dt' \tag{4.8}$$

This can be understood by seeing that the contribution at time $t' \leq t$ is due to an instantaneous pulse of NO t' seconds previously, with $S(t-t')$ the amount of NO per second produced at time $t-t'$ that is, t' seconds earlier. Thus, in Equation 4.8, the most recent pulses of NO are responsible for the lower limit of the integration, whilst the oldest pulses account for the upper limit. Similarly, we can derive the solution for times after a source which emitted NO continuously has stopped synthesising. If the source synthesises for T seconds and, as before, the instantaneous solution is $f(r,t')$, then the concentration at a distance r from the source, t_1 seconds after it has stopped synthesising is:

$$C(r,t_1+T) = \int_{t_1}^{t_1+T} S(t_1+T-t') f\left(r,t'\right) dt' \tag{4.9}$$

where $t_1 > 0$. This approach is valid since the diffusion equation is linear and the principle of superposition of linear solutions therefore applies.

4.2.2.2 Modelling NO diffusion from a symmetrical 3D structure

To model the spread of an amount of NO produced instantaneously throughout a continuous structure, we use methods developed in the field of thermodynamics which are readily applicable to modelling diffusion [6]. The main technique is to build up solutions for complicated structures from summation of contributions from point sources distributed throughout the structure. Of course, we are not implying that there are an infinite number of NO sources in the structure, but they are small enough that we are justified in imagining that they are uniformly distributed throughout the source with some density ρ (see Section 4.2.4.3). Hence for a spherical source, M, of radius a, the method is to sum the contributions to the concentration at a point in space $\mathbf{y}$, from all the points within the sphere as described below.

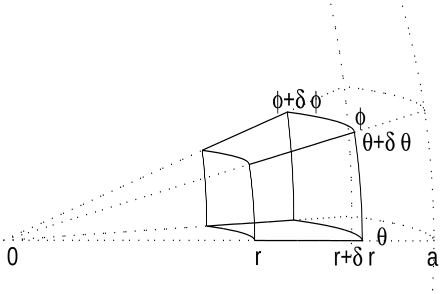

Figure 4.1

The element X (see Equation 4.10) of a sphere of radius a containing the points (r', θ', ϕ') where: $r \leq r' \leq r + \delta r$; $\theta \leq \theta' \leq \theta + \delta\theta$ and $\phi \leq \phi' \leq \phi + \delta\phi$. X is outlined by solid lines with dashed lines denoting radii and surface of the sphere.

Take a volume X within the sphere containing the points (r', θ', ϕ') where:

$$r \leq r' \leq r + \delta r; \qquad \theta \leq \theta' \leq \theta + \delta\theta; \qquad \phi \leq \phi' \leq \phi + \delta\phi. \tag{4.10}$$

as shown in Figure 4.1. If this element, X, is relatively small (i.e., if δr, $\delta\theta$ and $\delta\phi$ are sufficiently small), we can approximate its volume with:

$$V_X \approx r^2 \sin\theta\, \delta\theta\, \delta\phi\, \delta r \tag{4.11}$$

with the error in the approximation getting smaller as the dimensions of the element (δr, $\delta\theta$ and $\delta\phi$) are reduced and the error becoming zero in the limit of the dimensions becoming vanishingly small. Now, the amount of NO produced per second in a volume V of NO-producing tissue is:

$$S_V = Q \times N_V \tag{4.12}$$

where Q is the amount of NO produced per second from a single NO producing unit and N_V is the number of these units within V. This number is simply the product of the volume of V and ρ, the density of units in V. Hence for the element, X (Figure 4.1), we have a strength/second term, S_X, of:

$$S_X = Q \times N_X = Q \times \rho V_X \approx Q\rho\, r^2 \sin\theta\, \delta\theta\, \delta\phi\, \delta r \tag{4.13}$$

In this equation, r, θ and ϕ are variables whilst the product $Q\rho$, the concentration of NO produced per second, is independent of the particular shape of the structure being studied and so can be determined by empirical experiments as in [47].

Making δr, $\delta\theta$ and $\delta\phi$ vanishingly small makes the approximation in Equations 4.11 and 4.13 exact and we can assume that the contribution to the concentration at a point, $\mathbf{y} = (\tilde{r}, 0, 0)$, from the volume X is as if from a point source at the point: $\mathbf{x} = (r, \theta, \phi)$. Inspecting Equation 4.7 we see that the concentration depends on the time after synthesis and the distance between $\mathbf{x}$ and $\mathbf{y}$, $\parallel \mathbf{x} - \mathbf{y} \parallel$. Substituting this and our strength/second term, S_X, from Equation 4.13 into Equation 4.7 we obtain the concentration of NO at $\mathbf{y}$ due to $\mathbf{x}$ at a time t after synthesis:

$$C_P(\parallel \mathbf{x} - \mathbf{y} \parallel, t) = \frac{Q\rho\, r^2 \sin\theta\, dr\, d\theta\, d\phi}{8(\pi D t)^{3/2}}\, exp\left(\frac{-\parallel \mathbf{x} - \mathbf{y} \parallel^2}{4Dt}\right) e^{-\lambda t} \tag{4.14}$$

Now, to get the concentration at $\mathbf{y}$ due to the whole sphere we must sum up the contributions from all the points $\mathbf{x} = (r, \theta, \phi)$ inside the sphere, M, (i.e., $0 \leq r \leq a$; $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$) as shown below:

$$\begin{aligned} C_S(a, \tilde{r}, t) &= \sum_{\mathbf{x} \in M} C_P(\parallel \mathbf{x} - \mathbf{y} \parallel, t) && (4.15) \\ &= \int_0^a \int_0^\pi \int_0^{2\pi} \frac{Q\rho\, r^2 \sin\theta}{8(\pi D t)^{3/2}} \exp\left(\frac{-\parallel \mathbf{x} - \mathbf{y} \parallel^2}{4Dt}\right) && (4.16) \\ &\qquad \cdot e^{-\lambda t}\, dr\, d\theta\, d\phi && (4.17) \end{aligned}$$

Using:

$$\parallel \mathbf{x} - \mathbf{y} \parallel^2 = \tilde{r}^2 + r^2 - 2\tilde{r}r \cos\theta \tag{4.18}$$

and noting that there is radial symmetry so that the concentration at any point $\mathbf{z} = (\tilde{r}, \theta, \phi)$ at a distance of $\tilde{r}$ from the origin is equal to the concentration at $\mathbf{y} = (\tilde{r}, 0, 0)$, we therefore obtain:

$$\begin{aligned} C_S(a, \tilde{r}, t) = Q\rho\, e^{-\lambda t} \Bigg[& \frac{1}{2}\left(\operatorname{erf}\left(\frac{a + \tilde{r}}{2\sqrt{Dt}}\right) + \operatorname{erf}\left(\frac{a - \tilde{r}}{2\sqrt{Dt}}\right)\right) \\ & - \frac{1}{r'}\sqrt{\frac{Dt}{\pi}}\left(\exp\left(\frac{(a - \tilde{r})^2}{4Dt}\right) - \exp\left(\frac{(a + \tilde{r})^2}{4Dt}\right)\right)\Bigg] \end{aligned} \tag{4.19}$$

where:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-u^2\right) du \tag{4.20}$$

for the concentration at a distance $\tilde{r}$ from the centre of a solid sphere of radius a at a time t after synthesis. This leads naturally to the solution for a hollow sphere of inner radius a and outer radius b:

$$C_H(a, b, \tilde{r}, t) = C_S(b, \tilde{r}, t) - C_S(a, \tilde{r}, t) \tag{4.21}$$

The analytical method outlined also yields the concentration at distance $\tilde{r}$ from the centre of an annulus of inner radius R_1 and outer radius R_2 at a time t after synthesis:

$$C_A(R_1,R_2,\tilde{r},t) = \frac{Q\rho}{2Dt} e^{-\lambda t} \exp\left(\frac{-\tilde{r}^2}{4Dt}\right) \int_{R_1}^{R_2} \exp\left(\frac{-r^2}{4Dt}\right) \mathrm{I}_0\left(\frac{r\tilde{r}}{2Dt}\right) dr \qquad (4.22)$$

where $\mathrm{I}_0(x)$ is the modified bessel function of order zero [6].

These 'instantaneous' solutions can then be integrated over the appropriate time intervals to get the solutions for the evolution of concentration of NO synthesised for a finite time interval, in the same way as for the point source (as in Equations 4.8 and 4.9). In these cases, however, as the volume term has already been implicitly factored into the integrals, we replace $Q\rho$, the concentration/second at each instant, with $\tilde{S}(t)$, a function which, for each time t, gives the value of $Q\rho$ t seconds after the start of synthesis. Traditionally, instantaneous switch-on and off of synthesis has been assumed meaning that $\tilde{S}(t)$ will be a square wave with maximum value of $Q\rho$. This has the advantage of simplicity since $\tilde{S}(t)$ is now constant and can be moved outside the integral in Equations 4.8 and 4.9. Obviously, such a mechanism of release is not strictly biologically plausible but, in the absence of experimental data on the kinetics of nNOS activation *in vivo* and given the insensitivity of the diffusion process to small scale heterogeneity [35], this is a reasonable approximation. However, if numerical integration techniques are being used, more complicated strength functions can be used to model the time-course of NO synthesis. For instance, in [34] we used a strength function relating the amount of NO released to the amount of depolarisation caused by an action potential.

4.2.2.3 Numerical integration of analytical solutions

As mentioned previously, numerical integration is generally required to generate the analytic solutions especially if any reasonable inactivation is included. We will now describe the methods used for the results detailed below, but a full review of numerical integration techniques can be found in, for example, [9].

Equation 4.19 requires numerical integration over time. This was performed by the 'quad8' function in the programming language Matlab, which uses an adaptive recursive Newton Cotes 8 panel rule [9] to a relative accuracy of 0.1%. This is an extension of the *extended trapezoidal rule* in which the integral of $y(x)$ between x_A and x_B is estimated by dividing the range into N sections of width h, and approximating the area under the curve in each segment by the area of a trapezium, giving:

$$\int_{x_A}^{x_B} y(x) = h\left[\frac{1}{2}y_0 + y_2 + y_3 + \ldots + y_{N-1} + \frac{1}{2}y_N\right] + O\left(\frac{(x_B - x_A)^3 y''}{N^2}\right) \qquad (4.23)$$

where $y_i = y(x_A + ih)$. While this method is robust for functions that are not very smooth, it is relatively slow and the 'quad8' function achieves much faster convergence by adaptively changing the positions and weightings of the estimates y_i in Equation 4.23. However the extra speed comes at the cost of a loss of accuracy over the less smooth areas of the integrand near the temporal origin. This means that the integration procedure must be modified for cases where the lower limit of integration in Equation 4.19 is less than $1ms$ to include the following analytical approximation for the second part of the integral:

$$\int_0^{\varepsilon} Q\rho e^{-\lambda t}\frac{1}{r}\sqrt{\frac{Dt}{\pi}}\left[\exp\left(\frac{(a-r)^2}{4Dt}\right)-\exp\left(\frac{(a+r)^2}{4Dt}\right)\right]dt \tag{4.24}$$

$$\simeq Q\rho\frac{\left(1+e^{-\lambda\varepsilon}\right)}{2}\left[h\left(\frac{(r-a)^2}{4D},r,\varepsilon\right)-h\left(\frac{(r+a)^2}{4D},r,\varepsilon\right)\right] \tag{4.25}$$

where:

$$\begin{aligned} h(k,r,\varepsilon) &= \textstyle\int_0^{\varepsilon}\frac{1}{r}\sqrt{\frac{Dt}{\pi}}e^{-\frac{k}{t}}\,dt \\ &= \textstyle\frac{2}{3r}\sqrt{\frac{D}{\pi}}\left(e^{-\frac{k}{\varepsilon}}\sqrt{\varepsilon}\,(\varepsilon-2k)+2\sqrt{\pi k^3}\,\mathrm{erfc}\sqrt{\frac{k}{\varepsilon}}\right) \end{aligned} \tag{4.26}$$

$$\mathrm{erfc}(x) = 1-\mathrm{erf}(x) \tag{4.27}$$

and noting that at $t=0$ the instantaneous solution is:

$$C_S(a,r,t) = \begin{cases} Q\rho & \text{for } r<a \\ Q\rho/2 & \text{for } r=a \\ 0 & \text{else} \end{cases} \tag{4.28}$$

In the above a is the radius of the sphere, r is the distance from its centre and $\varepsilon \leq 1ms$. Solutions for the hollow sphere and when the lower limit of integration is greater than zero can be derived from the above equations.

The approximation in Equation 4.24 is based on the principle that if $fmin_{[0,\varepsilon]}$ is the minimum value attained by a function $f(t)$ over the range $[0,\varepsilon]$ and $fmax_{[0,\varepsilon]}$ is the maximum of $f(t)$ over the same range then:

$$\int_0^{\varepsilon} f(t)g(t)\,dt \simeq \frac{fmin_{[0,\varepsilon]}+fmax_{[0,\varepsilon]}}{2}\int_0^{\varepsilon} g(t)\,dt \tag{4.29}$$

which has a maximum error of:

$$\frac{fmax_{[0,\varepsilon]}-fmin_{[0,\varepsilon]}}{2}\int_0^{\varepsilon} g(t)\,dt \tag{4.30}$$

Thus, the actual value of the error is dependent on the parameter values used but, for the parameters used here, the errors are small enough to keep the solutions to within a relative accuracy of 0.1%.

The continuous solution for the tubular source (Equation 4.22) has to be integrated over both space and time, necessitating a slightly different approach since multi-dimensional integration is significantly more time-consuming and can magnify errors and instabilities in the methods used [37]. While Monte-Carlo integration can be used, it is inappropriate for our needs due to the asymptotically slow convergence. Due to the accuracy requirements and relative smoothness of the function, the approach we have taken is to use successive applications of one-dimensional integration. In this method, to evaluate $y(x,t_i)$ for each abscissa, t_i, of an iteration of the outer integration of $\int\int y(x,t)dxdt$, we must perform one whole numerical integration over x using that value of t_i to evaluate $y(x_j,t_i)$ at each x_j. This means that if it takes N function evaluations to get a sufficiently accurate estimate for the one-dimensional integral, we will need around N^2 evaluations to achieve the same accuracy for the two-dimensional integral. Moreover, since it is good practice for accuracy reasons to use simple numerical integration routines, the exponential growth in the number of operations needed is exacerbated by the slow convergence of these methods. Given these considerations, solutions for reasonably complicated functions requiring greater than double integration is probably best handled with one of the numerical methods discussed in the next section.

For the results detailed here, the extended trapezoidal rule given in Equation 4.23 [37] was used for the outer integration (over time) with the inner integration (over radial distance) performed by the 'quad8' function to speed up convergence. However, to ensure accuracy, we checked the solutions by performing both inner and outer integrations using the extended trapezoidal rule. In evaluating the continuous solution for the tubular source, it should be noted that at $t = 0$ the instantaneous solution is:

$$C_A(R_1,R_2,r,t) = \begin{cases} Q\rho & \text{for } R_1 < r < R_2 \\ Q\rho/2 & \text{for } r = R_1 \text{ , } r = R_2 \\ 0 & \text{else} \end{cases} \tag{4.31}$$

Solutions were accurate to a relative accuracy of 0.5%. Accuracy of solutions for the tubular and spherical sources were further checked using the numerical integration package in the programming language Maple which is very accurate and accounts for improper integrals correctly, but is too slow for general use.

4.2.3 Modelling diffusion of NO from an irregular 3D structure

4.2.3.1 Finite difference methods for diffusive problems

From the previous section it is clear that the analytical method is not tractable for many situations which we might want to investigate. In particular, modelling irregularly shaped sources and sinks is inappropriate and other numerical techniques for solving the partial differential equations (PDEs) governing the spread of NO must be used. For diffusive problems evolving over the short time-scales associated with NO diffusion in the brain, one recommended approach is to use finite differences [1, 37].

These methods proceed by approximating the continuous derivatives at a point by difference quotients over a small interval, for instance replacing $\frac{\partial x}{\partial t}$ by:

$$\frac{\delta x}{\delta t} = \frac{x(t+\Delta t) - x(t)}{\Delta t} \tag{4.32}$$

In this way, given some initial conditions for x at $t = t_0$, we can define a recurrence relation:

$$x_0 = a, x_{n+1} = x_n + \Delta t f(x_n, t_n), \text{where: } x_n = x(t_n), t_n = t_0 + n\Delta t \tag{4.33}$$

which can be solved iteratively. The error in the approximation is dependent on the size of Δt, the step-size, with schemes being said to be nth order accurate in a given variable (or variables), meaning that the error is essentially a constant multiplied by the step-size raised to the *nth* power [1]. As well as governing this truncation error, one must also ensure that the spatial and temporal step-sizes used do not make the set of equations unstable, resulting in erroneous answers. For instance, explicit difference equations (DEs), where the values at time-step $n+1$ are calculated using only values known at time n as in Equation 4.33 above, tend to have stability problems. Thus, compartmental models, a common finite difference method used to solve the diffusion equation (Equation 4.2) [14, 24, 26], are hampered by the fact that for stability:

$$\frac{D\Delta t}{(\Delta x)^2} \leq \frac{1}{2^n} \tag{4.34}$$

where D is the diffusion coefficient, n is the spatial dimension and Δx and Δt are the spatial and temporal step-sizes respectively [37]. This puts a limitation on the size of the time-step to be used which, in less abstract terms, means that, in one space dimension, it must be less than the diffusion time across a cell of width Δx.

However, different schemes have different stability properties and so the restrictive bounds of the compartmental model can be avoided. For instance, implicit DEs, where values at time $n+1$ are defined in terms of each other, are often stable for all step-sizes. However, while explicit DEs are inherently easy to solve as the solution is simply propagated forward in time, implicit DEs require the solution of a set of simultaneous Equations [30]. In order to avoid computationally intensive routines, it is therefore important that the DE is designed so that the resulting system of equations is tridiagonal.[†] One such equation, known as the Crank-Nicholson scheme, is recommended for diffusive problems in one space dimension [37]. Applied to the one-dimensional version of Equation 4.2 we have, using the notation of Equation 4.34:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = D\left[\frac{\left(u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}\right) + \left(u_{i+1}^n - 2u_i^n + u_{i-1}^n\right)}{(\Delta x)^2}\right] \tag{4.35}$$

[†] A tridiagonal system of equations $Ax = b$ is one where the matrix A is tridiagonal, that is, where the elements of A, a_{ij}, equal 0 if $i > j+1$ or $j > i+1$. In other words, if the row and column number differ by more than one, the entry must be zero. The entries are otherwise unrestricted. In non-mathematical terms this means that the resulting equations can be solved quite straightforwardly in O(n) operations where n is the number of equations, which in the context of difference equations equates to the number of spatial points at which the equation is to be evaluated.

where u_i^n is the concentration at time $n\Delta t$ and position $i\Delta x$. This scheme is second order accurate in space and time while maintaining stability for all choices of Δt.

This equation can be generalised for higher spatial dimensions quite easily but the resulting systems of equations are no longer tridiagonal and so much more computationally expensive to solve. As the number of operations required to solve multi-dimensional DEs increases exponentially with the dimension (as in numerical integration), this method is impractical [1]. Generalising explicit schemes in this way is feasible due to the speed with which they can be solved. However the limit on the step-sizes that can be used while ensuring stability (Equation 4.34) becomes even more restrictive obviating this approach as well.

To get around these problems one can use a second class of techniques, known as *alternating-direction implicit* (ADI) methods. The basic idea in these schemes is to split up a single time-step into n sub-steps, one for each spatial dimension. For each one of these sub-steps we evaluate only one spatial derivative at the advanced time step which ensures that the resultant sub-system is tridiagonal, and thus solving for each coordinate direction in turn. To see the general principle consider the following generalisation of Equation 4.35 the two-dimensional diffusion equation:

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) \tag{4.36}$$

Defining:

$$u_n \equiv u_{i,j,n} \equiv u(i\Delta x, j\Delta y, n\Delta t) \tag{4.37}$$

$$\delta_x^2 u_n \equiv \delta_x^2 u_{i,j,n} \equiv u_{i+1,j,n} - 2u_{i,j,n} + u_{i-1,j,n} \tag{4.38}$$

$$\delta_y^2 u_n \equiv \delta_y^2 u_{i,j,n} \equiv u_{i,j+1,n} - 2u_{i,j,n} + u_{i,j-1,n} \tag{4.39}$$

where $\Delta x, \Delta y$ and Δt are the spatial and temporal step-sizes respectively, we have:

$$\frac{u_{n+1/2} - u_n}{\frac{1}{2}\Delta t} = D\left(\delta_x^2 u_{n+1/2} + \delta_y^2 u_n\right) \tag{4.40}$$

which is tridiagonal. This is solved everywhere and the solution for the half time-step is then used in the following difference:

$$\frac{u_{n+1} - u_{n+1/2}}{\frac{1}{2}\Delta t} = D\left(\delta_x^2 u_{n+1/2} + \delta_y^2 u_{n+1}\right) \tag{4.41}$$

to solve for the full time-step. This results in a scheme which is stable, second order accurate in space and time and only requires solution of tridiagonal systems. Such methods, while certainly requiring a significant amount of computation, are at least practical and have been used extensively for diffusive IVPs [1, 37].

Before detailing the specific difference schemes used, it should be noted that the major issue in using difference equations for multi-dimensional diffusive DEs is that of computational power. As well as the number of operations required scaling exponentially with the number of dimensions, so do the memory (RAM) requirements.

One could attempt to alleviate this by reducing the problem's size by using large spatial and temporal scales and offsetting the loss of accuracy incurred by employing higher-order (in terms of accuracy) methods. However, unless the DE is extremely stable, using too high an order can introduce spurious solutions and for second order initial value problems, such as the diffusion equation, Press et al. [37] recommend that one should go no higher than second order in space and time. Thus for multi-dimensional DEs one must resign oneself to long running times, high memory requirements and a certain loss of accuracy. If this is not practical a lower dimensional model can in some circumstances be used to approximate higher dimensions [45].

4.2.3.2 Finite difference schemes used

For the numerical solutions given here, we used the alternating direction implicit (ADI) method in two and three space dimensions [1]. This is recommended for diffusive problems as it is fast, second-order accurate in space and time, unconditionally stable and unlike simpler schemes, allows for examination of the solution at all time-steps [1, 30, 37]. The equation to be approximated is:

$$\frac{\partial C}{\partial t} - D\nabla^2 C = P(\vec{x},t) - S(\vec{x})C - \lambda C \tag{4.42}$$

where:

$$P(\vec{x},t) = \begin{cases} Q\rho & \text{for points inside the source during synthesis} \\ 0 & \text{else} \end{cases} \tag{4.43}$$

and:

$$S(\vec{x}) = \begin{cases} \eta & \text{for points inside sinks} \\ 0 & \text{else} \end{cases} \tag{4.44}$$

where a sink is a local high concentration of an NO-binding moiety such as a heme-protein. Thus in two dimensions we have [1]:

$$\begin{aligned} \tfrac{u_{n+1/2}-u_n}{\frac{1}{2}\Delta t} &= D\left(\delta_x^2 u_{n+1/2} + \delta_y^2 u_n\right) + P(i,j,n) \\ &\quad - \left(\tfrac{\lambda+S(i,j)}{2}\right)\left(u_{n+1/2}+u_n\right) \end{aligned} \tag{4.45}$$

$$\begin{aligned} \tfrac{u_{n+1}-u_{n+1/2}}{\frac{1}{2}\Delta t} &= D\left(\delta_x^2 u_{n+1/2} + \delta_y^2 u_{n+1}\right) + P\left(i,j,n+\tfrac{1}{2}\right) \\ &\quad - \left(\tfrac{\lambda+S(i,j)}{2}\right)\left(u_{n+1}+u_{n+1/2}\right) \end{aligned} \tag{4.46}$$

Extending these equations to three space variables leads to a method that is unstable for any useful spatial and temporal scales [1] and so the following variant is used. Instead of taking three third-steps one generates three subsequent approximations for the solution at the advanced time-step, the third one being used as the actual solution. We obtain the first approximation u^*_{n+1} at time-step $n+1$ in the following way [1]:

$$\begin{aligned} \tfrac{u^*_{n+1}-u_n}{\Delta t} &= D\left[\tfrac{1}{2}\delta_x^2\left(u^*_{n+1}+u_n\right) + \delta_y^2 u_n + \delta_z^2 u_n\right] \\ &\quad + P(n) - \tfrac{\lambda}{2}\left(u^*_{n+1}+u_n\right) \end{aligned} \tag{4.47}$$

where Equations 4.37-4.39 have been extended in the obvious way so that, for instance:

$$u_n \equiv u_{i,j,k,n} \equiv u(i\Delta x, j\Delta y, k\Delta z, n\Delta t) \tag{4.48}$$

$$\delta_z^2 u_n \equiv \delta_z^2 u_{i,j,k,n} \equiv u_{i,j,k+1,n} - 2u_{i,j,k,n} + u_{i,j,k-1,n} \tag{4.49}$$

The second approximation u^{**}_{n+1} is then calculated using the first via:

$$\begin{aligned} \frac{u^{**}_{n+1}-u_n}{\Delta t} = {} & D\left[\tfrac{1}{2}\delta_x^2\left(u^{*}_{n+1}+u_n\right)+\tfrac{1}{2}\delta_y^2\left(u^{**}_{n+1}+u_n\right)+\delta_z^2 u_n\right] \\ & +P(n)-\tfrac{\lambda}{2}\left(u^{**}_{n+1}+u_n\right) \end{aligned} \tag{4.50}$$

and the final solution u_{n+1} with:

$$\begin{aligned} \frac{u_{n+1}-u_n}{\Delta t} = {} & \tfrac{D}{2}\left[\delta_x^2\left(u^{*}_{n+1}+u_n\right)+\delta_y^2\left(u^{**}_{n+1}+u_n\right)+\delta_z^2\left(u_{n+1}+u_n\right)\right] \\ & +P(n)-\tfrac{\lambda}{2}\left(u_{n+1}+u_n\right) \end{aligned} \tag{4.51}$$

In the above the reaction term for a sink has been dropped as this was not used with a three-dimensional model, though its inclusion is straightforward.

The two-dimensional ADI equation was implemented with spatial scale of $1\mu m$, on a square grid of size 1000×1000 and time step $1ms$. The three-dimensional version also used a space-step of $1\mu m$, but on a cubic $300 \times 300 \times 300$ grid with a time-step of $4ms$. The effects of the step-sizes were checked by running the equations with smaller scales and were found to be negligible ($< 0.5\%$ relative error). The equations were run using Neumann boundary conditions with the gradient at the edge of the grid set to be constant. However, to ensure that the size of the grid and boundary condition did not affect the results significantly, we checked the simulations by rerunning them with a flat gradient at the boundary. The equations were written in C. For full details of the implementation see [35].

4.2.4 Parameter values

The values of the main parameters used here, the diffusion coefficient, D, the decay rates, λ and η, and the concentration rate, $Q\rho$, warrant some discussion. We also discuss the choice of an NO threshold and the localisation of nNOS.

4.2.4.1 Diffusion coefficient and decay rate

The value of D in an aqueous salt solution has been measured as $3300\mu m^2 s^{-1}$ [29]. This value has been used widely [24, 44, 47] and has also been derived with reference to a model [45]. It is reasonable to assume that it will not be significantly affected in a lipid or protein aqueous medium due to the very small molecular dimension and nonpolarity of NO. In addition, because NO is dilute, D is assumed to be independent of NO concentration and constant [45] and so we use this value throughout.

The value of the decay rate used gives a half-life of $5s$, which is that recorded for dissolved NO perfused over living tissues in oxygenated saline solution (Moncada et al. 1989). Whilst other rate constants can be used, these are basically dependent on the oxidising environment in which NO is diffusing. If this is other than a simple environment, with a $t_{1/2} \ll 5s$, it should be treated more carefully [7, 45], whilst anything longer has hardly any effect over the spatial and temporal scales examined here [47]. We have thus made the simplifying assumption that the background half-life is $5s$. For strong NO sinks, η, has a value of $693.15s^{-1}$ equivalent to a half-life of $1ms$ which was chosen as a conservative value based on the rate of NO uptake by a nearby haemoglobin containing structure such as a blood vessel [23, 28].

4.2.4.2 NO production rate

The value of the synthesis rate $Q\rho$ is a more open question and several values have been determined via different models. Before these are discussed, however, it should be noted that the effect of this parameter is purely one of scale as it is a constant which simply multiplies the concentrations. Thus, whatever the actual value of this parameter, the qualitative nature of the results is unchanged and it is easy to see what effect a different value would have simply by rescaling.

There are two determinations of $Q\rho$ that have underpinned NO diffusion modelling to date, both of which are based on the experimental findings of Malinski et al. [29]. Both results are measurements of NO from endothelial cells of a rabbit aorta, one *in vivo* and the other *in vitro*. The first measurement is the concentration of NO produced from stimulated cells of the aorta $100\mu m$ away after diffusion through muscle cells. The second measurement is taken at the surface of a single endothelial cell in culture stimulated to produce NO.

Vaughn et al. [45] chose to use the *in vivo* determination. This is a very complex situation since the reaction with smooth muscle has to be taken into consideration and the size of the synthesising region is unknown and to complete the calculation many simplifying assumptions had to be made. One of these was that NO was produced at the surface of the endothelial cells only, which could seriously alter the results and renders the production rate gained unusable in the models we employed.

We, like Wood and Garthwaite [47], base our model on the *in vitro* determination. However, unlike Wood and Garthwaite [47], who used a point-source model to represent a spherical neuron of diameter $1\mu m$, we employed a structure-based analysis. For this task, we used a hollow sphere of inner radius 6, outer radius 10, with the result that a value for $Q\rho$ of $1.32 \times 10^{-4} mol\mu m^{-3}s^{-1}$ is needed to generate a maximum concentration of $1\mu M$ on the surface of the sphere. These dimensions are chosen to approximate an average endothelial cell but are not incredibly significant [35]. Also, the endothelial cell will not be spherical but again, due to the insensitivity of the results to changes of dimension and given that there is no other data available, this approximation was deemed sufficient. Significantly, the resultant value for $Q\rho$ is about 300 times less than that used previously. Moreover, the peak concentration is attained after about 14 seconds - a result which agrees closely with the empirical data of Malinski et al. [29] but which was unexplained when the point-source model was utilised [26, 35]. The determination of the production rate is discussed in more detail in [35].

4.2.4.3 Distribution of nNOS in neurons

When modelling NO formation one must know where the NOS is located and this differs depending on the type of NOS examined. Here we are mainly concerned with nNOS and so we should say a few words on its distribution within neurons. In the literature nNOS is often referred to as being located on the membrane, as it is frequently shown to be associated with postsynaptic density protein-95 (PSD-95) which in turn is linked to the NMDA-receptor which is membrane-associated (see, for instance [41]. This is consistent with NO acting as a retrograde messenger to induce LTP or LTD in the pre-synaptic neuron. Thus it is natural that nNOS's position near the NMDA receptor is emphasised but this should not be taken to mean that there is no nNOS elsewhere. Indeed, nNOS is a soluble enzyme and will be dispersed throughout the cytoplasm, as demonstrated by NADPH-diaphorase staining [32]. A similar story is true of eNOS which, while it does have an affinity for the membrane, will also be found at positions throughout the endothelial cell [11]. Thus in our model we have made the assumption that nNOS is spread evenly within the synthesising region with a uniform source density which we have denoted as ρ.

4.2.4.4 The NO threshold concentration

To quantify a threshold concentration for effective NO signalling, one first has to specify a particular molecular signalling pathway. Here we follow the thinking of Vaughn et al. [44] who chose the soluble guanylyl cyclase-cyclic GMP (sGC-cGMP) signalling pathway, the major signalling pathway for NO in the brain [16, 38]. The *equilibrium dissociation constant* [42] for NO for sGC is $0.25\mu M$ and this value defines a threshold concentration for NO.

4.3 Results

In this section we apply the methods detailed above to investigate the properties of an NO signal produced by neuron-like morphologies. In so doing, we examine a number of salient functional questions that arise in the context of volume signalling. In particular, we highlight the importance of the morphology of the source in determining the spatial and temporal extent of an NO volume signal. While a number of previous models of NO diffusion in the brain have been published, they are broadly one of two types: point-source models (see for instance [24, 47]) or compartmental models [14, 25, 26], neither of which address the impact of the source structure on the diffusional process. The shortcomings of these approaches are discussed in detail in [35, 36], but we will summarise the main points here.

In a point-source model, as the name suggests, one models NO diffusing from a source as if it were being produced at a dimensionless point at its centre. It is not difficult to see intuitively that problems might arise from the fact that a point source is therefore by definition singular. The singular nature of the solution represents a fundamental problem to modelling sources with morphology which can be appreciated by examining the steady-state solution for the 3D point source used by Wood and Garthwaite [47]:

$$C(r) = \frac{S}{r}\exp\left(-r\sqrt{\frac{\ln 2}{Dt_{1/2}}}\right) \tag{4.52}$$

where r is the distance from the source and S is a constant determined by the production rate of the source. The first thing to note is that the concentration at the source $r = 0$ is infinite. Although the central concentration itself could be ignored, the ramifications of having a singularity at the heart of the solution causes many complications and unrealistic results [34]. Firstly one must decide at what distance from the centre the model is deemed to be 'correct', necessarily a somewhat arbitrary choice. The approach taken by Wood and Garthwaite [47] was to use the surface of the neuron and only consider points outside the cell. This highlights a second problem, namely that the internal concentration is indeterminate which in turn means that obtaining a meaningful solution for hollow structures is impossible [34]. Finally, as the concentration in Equation 4.52 is dependent on the distance from the source only, this model cannot be used to address the impact of different source morphologies as sources with the same value of S but different shapes and sizes will yield identical results.

In compartmental models, on the other hand, one can include some notion of the source morphology. However, while such models do give valid insights into the overall role of a diffusing messenger they are a form of explicit finite difference model [1] and are thus hampered by the limit on the duration of the time step employed given in Equation 4.34 [30]. This limitation necessitates the use of relatively large compartments leading to gross approximations. In view of this, we believe a more sophisticated form of numerical approximation, such as the one presented here, should be employed when the complexity of the morphology makes an analytical solution impractical.

4.3.1 Diffusion from a typical neuron

We first examine the solution for a simple symmetrical structure representing, for example, a neuronal cell body in which NO is synthesized in the cytoplasm but not in the nucleus. We have therefore examined the solution for a hollow spherical source of inner radius $50\mu m$ (the nucleus) and outer radius $100\mu m$ (cell body). These dimensions, though large for many neurons especially in vertebrates, do correspond to the dimensions for some identified giant molluscan neurons whose cell bodies synthesise NO and have been shown to mediate volume signaling [33].

Of course we are not suggesting that neurons are perfectly spherical but rather that hollow spheres are a useful approximation for neurons. They can, for example, tell us about the importance of morphological irregularities. For instance, if one had a cell which was mainly spherical but had a lot of small-scale variability

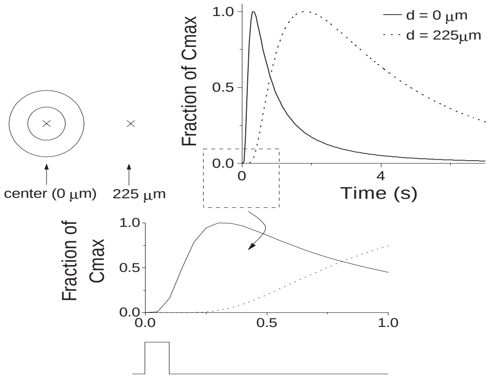

Figure 4.2

Concentration of NO plotted against time after synthesis for a hollow spherical source of inner radius $50\mu m$ and outer radius $100\mu m$ for a $100ms$ burst of synthesis. Here the solid line depicts the concentration at the centre of the cell ($0\mu m$), whilst the dotted line shows the concentration at $225\mu m$ from the centre. Because the absolute values attained at the two positions differ from one another markedly, the concentration is given as a fraction of the peak concentration attained. These peak values are $7.25\mu M$ (centre) and $0.25\mu M$ at $225\mu m$. The cell and the points at which the concentration is measured are depicted to the left of the main figure. Note the high central concentration, which persists for a long time (above $1\mu M$ for about $2s$. Also, there is a significant delay to a rise in concentration at distant points which is more clearly illustrated in the expanded inset figure. The square-wave shown beneath the inset figure represents the strength function.

in its outer structure, we could use two ideal models, one with the outer radius set to the minimum radius and the other with outer radius set to the maximum. In this way analytical solutions can be employed to see whether or not the irregularity has a significant effect. In fact we have seen that due to the speed of diffusion of NO, small-scale irregularities (±2.5% of source size) have a negligible effect [35]. Using such an approach we can also investigate the sensitivity of the diffusional process to other parameters such as boundary conditions whose complexity make the analytical solution intractable. Thus, if we have to make simplifications to a model to render derivation of the analytical solution tractable, we can tell whether or not these simplifications generate gross inaccuracies.

The solution for the hollow sphere was examined for a burst of synthesis of duration $100ms$, with results shown in Figures 4.2 and 4.3. There are two points of note, namely the length of time for which the concentration in the centre of the sphere remains high and the significant delay between the start of synthesis and the rise of concentration for points distant from the source (Figure 4.2). The cause of these phenomena can be seen on examination of Figure 4.3. During the synthesis phase, the concentration outside the cell rises very slowly. In the nucleus, however, a ‘reservoir’ of NO starts to build up (Figure 4.3A), albeit relatively slowly when compared to the

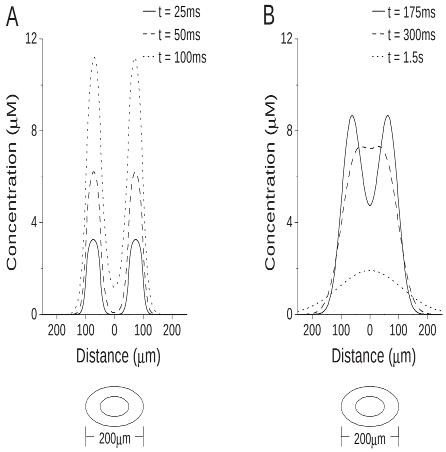

Figure 4.3

Concentration of NO plotted against distance from the centre of a hollow spherical source of inner radius 50μm and outer radius 100μm for a 100*ms* burst of synthesis starting at time $t = 0$. The graphics underneath each plot depict the structure. A. Concentration of NO at times $t = 25$, 50 and 100*ms*, two time points during and one at the end of synthesis. B. Concentration of NO after synthesis at times $t = 175$, 300 and 1.5*s*. The reservoir effect following the end of synthesis is clearly seen as the centrally accumulated NO is trapped by the higher surrounding concentrations.

rise in the synthesising area (the cytoplasm). After the end of synthesis, this reservoir continues to fill up for about 200*ms* as the NO in the cytoplasm diffuses away from its point of origin to points of lower concentration in the nucleus. However, the concentration outside the cell still rises slowly as the NO is dissipated over a larger volume. Later, the situation changes somewhat, as we are now in the position where the concentration in the nucleus is roughly equal to the concentration in the cytoplasm, giving a wide flat peak to the concentration profile. Until this point, the NO which had diffused into the centre had been ‘trapped’ and could not be dissipated due to the higher concentration present in the surrounding cytoplasm. Now though, we see this reservoir spreading away from the cell in a wave of high concentration which starts to raise the distal concentrations to significant levels. However, the concentration at the centre remains high and does not spread outwards very quickly since the concentration gradient is virtually flat, meaning there is very little diffusive pressure on the NO in this area. It is this effect that produces the unexpected time delay at distant points.

Examination of the concentration at 225μm from the centre of the cell (Figure 4.2), shows that it remains low until about 400*ms* after synthesis has stopped. It peaks shortly afterwards and stays relatively high for a relatively long period. This has implications for the temporal dynamics of NO-signalling in a neurobiological context. For example, suppose there was an NO-responsive neuron at a distance of $225\mu m$ from the centre of the source neuron. Assuming a threshold concentration of $0.1\mu M$ this neuron would not be affected until $600ms$ after the end of synthesis and would continue to be affected for a period of about $4s$. Such a process could be used to introduce a time delay in NO-mediated neural signalling. The high central concentration also has implications for neural signalling as the effect of the NO synthesising event remains long after this event has passed.

There is another interesting factor seen in these results, namely the temporal dynamics of the solution in the cytoplasm during synthesis. Here it is enough to note that the concentration continues to rise for a very long time of continuous synthesis before a steady-state is approached. Thus, though much of the work using point source models has considered solutions at steady-state such considerations may be inappropriate in the context of real structures.

4.3.2 Effect of neuron size

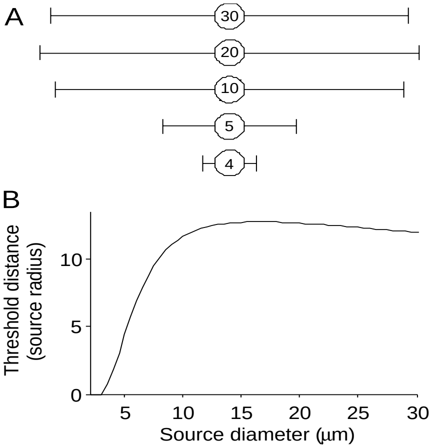

Figure 4.4

Threshold distances for NO generated by single fibres of different diameters. The threshold distance is defined as the distance at which concentration of NO drops below $0.25\mu M$, the dissociation constant for soluble guanylyl cyclase [42]. A. Illustration of threshold distances for several fibres. Numbers inside circles represent fibre diameters in microns. For comparison, the fibres have all been drawn the same size with the threshold distance shown in multiples of the fibre diameter. B. Threshold distance as a function of fibre diameter plotted against fibre diameter. Note the sharp reduction in threshold distance for fibres of diameter less than $10\mu m$.

The effects reported above are generated by the relatively large dimensions typical of some molluscan neurons. Questions therefore arise as to whether similar phenomena are present for smaller neurons such as those found in mammalian brains. In [34] we showed that reducing the radius of the cell can have a significant and somewhat counter-intuitive effect on the signalling capacities of neurons. In particular, for hollow spherical cells whose inner radius is half the outer radius and which synthesise NO for 100*ms*, the signalling capacity slowly increases as the cell size is reduced, peaking for $30 - 40\mu m$ diameter cells. Below this size, however, there is a steep almost linear decline in the threshold distance and sources of $10\mu m$ diameter or less have a very limited signalling capacity.

To further examine this effect, we studied the maximum region around a source which could be affected via the NO-cGMP signalling pathway during and after the generation of NO from tubular sources of various sizes. The affected region is defined by the volume within which the concentration is above the threshold concentration, as discussed in Section 4.2.4.4, and is given in multiples of fibre diameter. This is because the important comparator is the number of potential neuronal targets within the affected region and this depends on their size. A similar phenomenon to that seen in spherical sources is seen for these NO-expressing fibres (Figure 4.4): a slow rise in the affected region as the diameter is reduced, peaking at a fibre diameter of $20\mu m$, followed by a steep decline thereafter indicating that small fibres ($4\mu m$ diameter or less) are unable to generate an effective NO signal. Moreover, for such small sources, the affected region is not increased significantly by increasing the duration of NO synthesis, since a steady-state situation is rapidly approached. Thus for a source of diameter $3\mu m$ or less, a threshold concentration will not be achieved anywhere for no matter how long NO is synthesised (Figure 4.5).

4.3.3 Small sources

Sources which are too small to generate a threshold concentration individually may affect larger regions if they behave as if they were a single larger source with the attendant temporal and spatial phenomena associated with a source of the combined shape and size. An example of this is provided by endothelial cells which act as a multicellular complex of many very small cells (Vaughn et al. 98a). In many other locations in the brain, though, there are many instances in the brain of well-separated NO sources below the critical size required for a volume signal. It would seem therefore that NO from these sources cannot have a functional role. In determining the range of influence of a source, however, it was assumed that each source is acting on its own. What if instead, NO is derived from many small separated sources acting in concert?

In this section, we study this situation, analysing the dynamics of the NO cloud such sources produce. In particular, we examine networks of axonal fibres with diameters of a few microns or less and the functional extent of the volume signal they generate. A knowledge of the characteristics of such a signal could be crucial in helping to understand NO's neuromodulatory role, since this type of source is found in many places in the brain. For instance, while NO producing neurons account for only about 1% of cell bodies in the cerebral cortex, their processes spread so extensively,

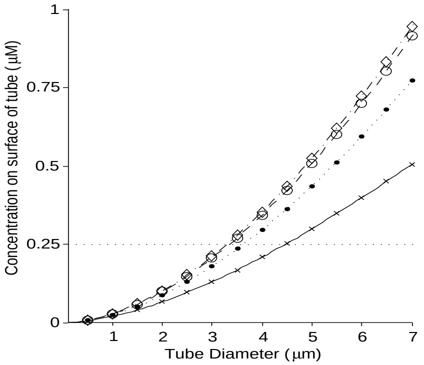

Figure 4.5

Maximum concentration reached on the surface of fibres of different diameters for synthesis bursts of length $0.1, 1, 5$ and $10s$. The dotted line shows the threshold concentration. The surface concentration is the maximum concentration achieved outside the fibre and so if this is less than threshold, these cells cannot affect any others via the NO-cGMP signalling pathway. This is the case for fibres of diameter less than $3.5\mu m$. Note that for cells of this size steady-state is approached after about $5s$ of synthesis, and that the smaller the fibre, the sooner this situation occurs.

that almost every neuron in the cortex is exposed to these small fibres the vast majority of which have diameters of a micron or less [2, 11]. Another example is found in the locust optic lobe where there are highly ordered sets of nNOS-expressing $2\mu m$ diameter fibres [12].

An illustration of how co-operation can occur in a model of fibres in the locust optic lobe is provided by Figure 4.6 which shows the spatial extent of the NO signal generated by a single fibre of $2\mu m$ diameter and by ordered arrays of four, nine and sixteen identical sources separated by $10\mu m$. As can be seen in the figure, an ordered array of N^2 fibres is a 2D arrangement of $N \times N$ fibres. As the fibres are parallel the solution is symmetric along the z-axis (out of the page) and so we give results only for cross-sectional slices in a plane perpendicular to the direction of the fibres. The single fibre does not achieve an above threshold signal principally because the great speed of NO diffusion means that NO will spread rapidly over a large volume. So while NO does not reach threshold anywhere, the volume occupied by NO at a significant fraction of threshold is large relative to the source size. Thus NO derived from small and well-separated individual sources can summate to produce an effective NO cloud. But what are the characteristics of such a signal and what do they imply for the way NO functions? For instance, do we still see the reservoir and delay effects characteristic of signals from single larger sources?

The first thing that one notices from Figure 4.6 is that while the centre effect is still present with NO accumulating in the centre of mass of the sources, the concentration profile appears to be flatter. This can be seen more clearly in Figure 4.7 where an

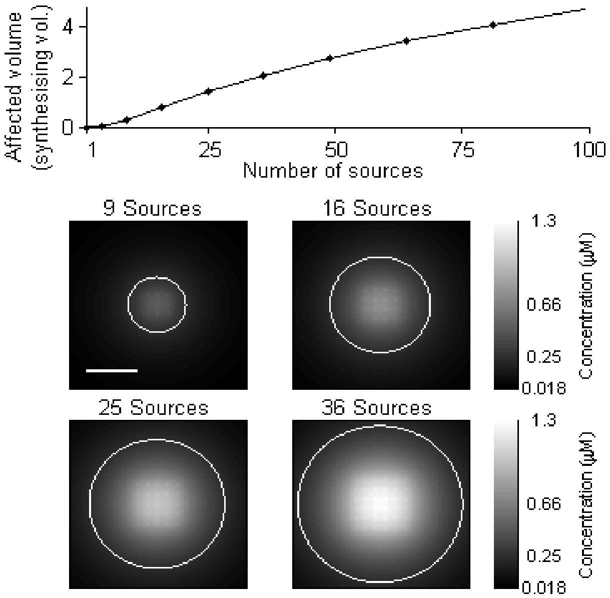

Figure 4.6

Volume over threshold (0.25μM) generated by NO synthesising fibres of 2μm diameter organised in ordered arrays separated by 10μm after 1 second of synthesis. The fibre dimensions and spacing have been chosen so as to approximate the arrangement of the nNOS-expressing fibres in the optic lobe of the locust [12]. The upper graph shows the volume over threshold per unit length of the fibres. The lower four graphs show the concentrations of NO (dark = low, light = high) in a two-dimensional slice through the fibres which project out of the page. Here we see how NO from several sources can combine to produce above threshold concentrations (areas inside the white boundaries) which extend away from the synthesising region. Scale bar is 50μm.

ordered 10 × 10 array of 2μm fibres separated by 36μm. The effect of the spacing between the sources is further illustrated in Figure 4.8, where the area over threshold generated by 100 fibres is shown as a function of the total cross-sectional area of source. What is immediately obvious from Figure 4.8 is that if one's goal is to reach the largest number of potential targets with a given volume of source fibres, then one should use a dispersed source rather than one single source. While the optimal spacing that should be used is dependent on the length of synthesis (and the number of sources [35], results for fibres arranged contiguously so that 100 fibres act as one 20 × 20μm source (the first points on the x-axis in Figures 4.8A-B), are *always* lower than those for dispersed sources (unless the sources are dispersed so widely that they fail to interact). Indeed can affect a volume over twice the size of a solid source simply by dispersing them correctly. Thus, in terms of the extent of the NO signal,

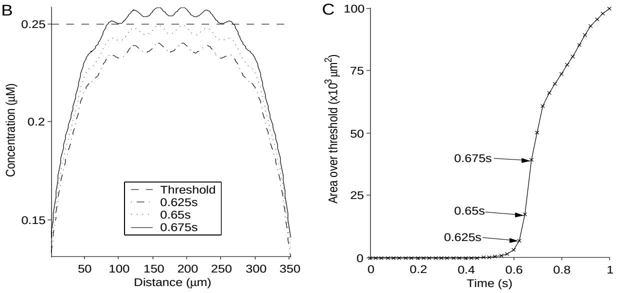

Figure 4.7

Concentrations of NO at several time points during NO synthesis in a line through the centre of ordered arrays of NO synthesising fibres of diameter $2\mu m$. B. NO concentrations generated by 100 fibres separated by $36\mu m$ after $0.625, 0.65$ and $0.675s$ of NO synthesis. The dashed line shows the threshold concentration. C. Area over threshold due to 100 fibres separated by $36\mu m$ for NO synthesis of length $1s$ plotted against time after synthesis. Here even though there has been $600ms$ of synthesis, just $100ms$ more extends the affected region from virtually nothing to over $50000\mu m^2$.

there is a big advantage in using separated sources.

What about the temporal dynamics of the NO signal? Examining the time-course of the NO signal generated by an array of 100 fibres spaced $36\mu m$ apart, we see a delay until areas reach an above threshold concentration as we did for single sources (Figure 4.7). This time, however, rather than the delay being for points outside the source only, here there is a delay until *any* point is affected by NO, after which there is a very steep rise in the volume affected. This is a common feature of signalling from dispersed sources because the summation of NO from several separated fibres means that the concentration in and around them is, in a sense, averaged and hence, smoothed. Thus due to the dynamics of diffusion one tends to get a relatively even concentration within the synthesising region with small peaks around the fibres themselves (Figure 4.7). In conjunction with the use of a threshold concentration, this means that there will come a point when the concentration in a region around the fibres is just sub-threshold and a small increase in the general level of NO will result in large areas rising above threshold, as shown in Figure 4.7. We refer to this feature as the *interaction* effect.

Again, the impact of the interaction effect will vary depending on the spacing used and a large range of temporal dynamics can be seen (Figure 4.8B). In particular, the delay before the start of interaction can vary from nothing to more than a second, with the delay growing as the spacing is increased. This means that a system with optimal spacing, in terms of extent of the affected region, will experience a considerable delay before the region begins to be affected, with the total area affected suddenly rising sharply at the end of the delay. This raises the intriguing functional possibility of a system which is completely unaffected by NO for a given length of time (a

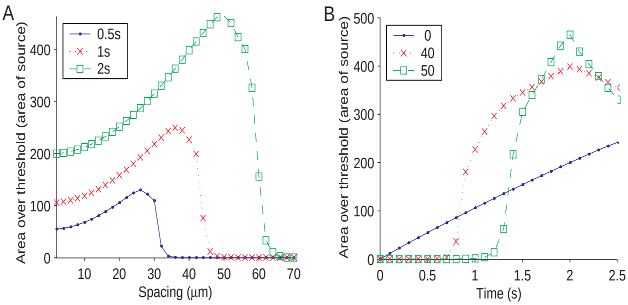

Figure 4.8

Area over threshold as a function of the total cross-sectional area of source for different numbers of evenly spaced fibres of diameter $2\mu m$. A. Affected area against spacing for 100 fibres for NO synthesis of length 0.5, 1 and 2 seconds. B. Affected area over time due to 100 fibres arranged as a single source (spacing = 2) or separated by 40 or $50\mu m$ for 2 seconds of NO synthesis. Note the delay till effective co-operation of the separated sources.

period which would be tunable by changing the spacing), but once past this point, large regions are rapidly 'turned-on' by the cloud of NO.[‡]

Another factor which affects the length of the delay is the thickness of the fibres used. Examining the delay from plexuses composed entirely of fibres of various fixed diameters, we see that the delay is longest for the thinnest fibres (Figure 4.9). This is because the fibres are now too small to achieve an above threshold signal singly and so must cooperate, although the subsequent rise in the affected area is not as steep as for ordered arrays. This is expected since the random nature of the plexus means that the distribution of concentrations is less uniform and the interaction effect is less pronounced. Examining the delay for plexuses of other diameters we see that it rises steeply to that seen for the $1\mu m$ plexuses showing that the interaction needed is much greater for the smaller fibres (Figure 4.9A). As a property of a signal, the obvious role for such a delay is as a low pass filter since there has to be nearly $200ms$ of synthesis before the thin plexus will respond. In the case of the signal mediating increased blood flow, this means that there would need to be significant sustained activity before blood flow increased, whereas for a thick plexus blood flow would react to every short burst of neuronal activity. Other features seen to vary with fibre thickness are maximum concentration, how centred a cloud is on a target region, and the variability of concentrations over a region; all of which have sensible interpretations in terms of neuronal signalling to blood vessels.

[‡] A form of signalling which might also be useful in an artificial neural network (see Section 4.4).

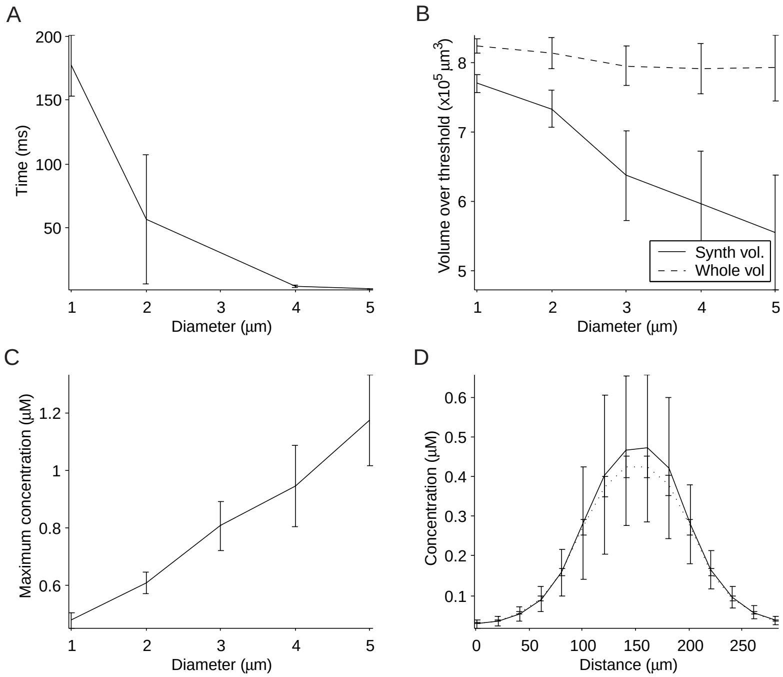

Figure 4.9

Mean results for plexuses composed entirely of fibres of diameters $d = 1,2,3,4$ or $5\mu m$ in a $100 \times 100 \times 100\mu m^3$ synthesising volume for 1 second of NO synthesis. Results averaged over 30 random plexuses grown generated using the growth algorithm detailed in [35]. A. Mean delay until interaction against diameter of plexus fibres. B. Mean volumes over threshold both in total and inside the synthesising volume against diameter of plexus fibres. C. Mean maximum concentration obtained against diameter of plexus fibres. D. Mean concentrations seen in a 1d line through the centre of either thin ($1\mu m$) or thick ($5\mu m$) plexuses.

4.4 Exploring functional roles with more abstract models

The computational models presented above require huge numbers of iterated calculations and inevitably place heavy demands on processing resources. Hence it is not yet feasible to build models of whole neuronal networks at that level of detail and run them in real time, or anything even vaguely approaching it. Therefore, in parallel with the detailed modelling work, we have developed a class of computationally fast artificial neural networks (ANNs) that incorporate a more abstract model of signalling by diffusing neuromodulators. Such networks have been used as artificial nervous systems in autonomous mobile robots and have allowed us to start exploring the properties and potential functional roles of this kind of signalling in the generation of behaviour [20, 21]. We have named this class of ANNs GasNets. Our work with these networks is briefly introduced in this section.

Since as yet we have no deep formal theory for such systems, we have found the use of stochastic search methods (such as evolutionary algorithms) to be a very helpful tool in this exploration. We have used the methods of evolutionary robotics to explore the suitability of this class of networks for generating a range of behaviours in a variety of autonomous robots [20, 21].

4.4.1 The GasNet model

In this strand of our work we have attempted to incorporate into ANNs, in an abstracted form, some of the richness and complexity that characterises the temporal and spatial dynamics of real neuronal signalling, especially chemical signalling by gaseous transmitters. As these systems operate on different temporal and spatial scales to electrical signalling, we have developed models in which electrical and indirect chemical signalling are controlled in ANNs by separate processes. Thus, we developed the GasNet, a standard ANN augmented by a diffusing virtual gas which can modulate the response of other neurons.

The 'electrical' network underlying the GasNet model is a discrete time-step, recurrent neural network with a variable number of nodes. These nodes are connected by either excitatory (with a weight of +1) or inhibitory (with a weight of -1) links with the output, O_i^n, of node i at time-step n described by the following equation:

$$O_i^n = \tanh\left[k_i^n\left(\sum_{j\in C_i} w_{ji}O_j^{n-1} + I_i^n\right) + b_i\right] \tag{4.53}$$

where C_i is the set of nodes with connections to node i and $w_{ji} = \pm 1$ is a connection weight. I_i^n is the external (sensory) input to node i at time n, and b_i is a genetically set bias. Each node has a genetically set default transfer function parameter, k_i^0, which can be altered at each time-step according to the concentration of the diffusing 'gas' at node i to give k_i^n (as described later in the section on modulation).

4.4.2 Gas diffusion in the networks

In addition to this underlying network in which positive and negative 'signals' flow between units, an abstract process loosely analogous to the diffusion of gaseous modulators is at play. Some units can emit virtual 'gases' which diffuse and are capable of modulating the behaviour of other units. The networks occupy a 2D space; the diffusion processes mean that the relative positioning of nodes is crucial to the functioning of the network. The original GasNet diffusion model is controlled by two genetically specified parameters, namely the radius of influence r and the rate of build up and decay s. Spatially, the gas concentration varies as an inverse exponential of the distance from the emitting node with a spread governed by r, with the concentration set to zero for all distances greater than *r* (Equation 4.54). The maximum concentration at the emitting node is 1.0 and the concentration builds up and decays from this value linearly as defined by Equations (4.55 and 4.56) at a rate determined by *s*.

$$C(d,t) = \begin{cases} e^{-2d/r} \times T(t) & d < r \\ 0 & \text{else} \end{cases} \tag{4.54}$$

$$T(t) = \begin{cases} H\left(\frac{t-t_e}{s}\right) & \text{emitting} \\ H\left[H\left(\frac{t_s-t_e}{s}\right) - H\left(\frac{t-t_s}{s}\right)\right] & \text{not emitting} \end{cases} \tag{4.55}$$

$$H(x) = \begin{cases} 0 \; x \leq 0 \\ x \; 0 < x < 1 \\ 1 \text{ else} \end{cases} \tag{4.56}$$

where C(d,t) is the concentration at a distance *d* from the emitting node at time *t*. t_e is the time at which emission was last turned on, t_s is the time at which emission was last turned off, and *s* (controlling the slope of the function *T*) is genetically determined for each node. The total concentration at a node is then determined by summing the contributions from all other emitting nodes (nodes are not affected by their own concentration, to avoid runaway positive feedback).

A variant on this diffusion model, based on the cortical plexus diffusion described earlier in this chapter, involves diffusion of a uniform concentration 'cloud' centred on some genetically specified site distant from the emitting node. This reflects the spatial separation of the main plexus and the body of the controlling neurons. The cloud suddenly turns 'on' or 'off', depending on the state of the controlling neuron, in keeping with the plexus mode of signalling described earlier.

4.4.3 Modulation

In a typical GasNet model [36], each node in the network can have one of three discrete quantities (zero, medium, maximum) of N possible receptors. Each diffusing-neurotransmitter/receptor pairing gives rise to a separate modulation to the properties of the node. The strength of a modulation at node *i* at time *n*, ΔM_j^n, is proportional to the product of the gas concentration at the node, $C_i^n$and the relevant receptor quantity, R_j as described by Equation 4.57. Each modulation makes some change to one or more function parameters of the node. All the variables controlling the process are again set for each node by an evolutionary search algorithm.

$$\Delta M_j^n = \rho_i C_i^n R_j \tag{4.57}$$

A number of different receptor linked modulations have been experimented with, including:

- Action of receptor 1 : increase gain of node transfer function
- Action of receptor 2: decrease gain of node transfer function

- Action of receptor 3: increase proportion of retained node activation from last time step
- Action of receptor 4: if above a threshold switch transfer function of node for sustained period

Most GasNet variants were found to be highly evolvable and capable of the robust generation of sensorimotor behaviours (often visually guided) in noisy environments [20, 21]. Indeed, it has been shown that most forms of GasNet are significantly more evolvable than all other forms of ANN tested over a wide range of continuous sensorimotor tasks. The operation of the networks was often subtle, making use of intricate dynamics involving continuously shifting modulation patterns [40]. Particular functional modules such as oscillators and low-pass 'noise' filters were frequently discovered and used by the evolutionary process. The inherently flexible, and generally loose, coupling between two processes with distinct spatial and temporal properties (the chemical and the electrical) makes these systems highly evolvable and endowed with a powerful kind of plasticity [36]. Future work will increase the biological veracity of the networks while maintaining their abstract computationally tractable nature. They can then be used to illuminate more specific biological questions than they have to date.

4.5 Conclusions

This chapter has concentrated on the details of how to build computational models of NO diffusion in the nervous system and has demonstrated how such models can give important insights into the phenomenon of volume signalling. This kind of signalling cannot be dealt with in a simple point-to-point connectionist framework; our tools and concepts for understanding the operation of neuronal circuitry making use of such signalling must be expanded. The work presented here is intended as a contribution to that expansion.

References

[1] W. Ames (1992). *Numerical Methods for Partial Differential Equations* (third ed). Academic Press.

[2] D. Baranano, C. Ferris and S, Snyder (2001). Atypical neural messengers. *Trends in Neuroscience* **24**: 99-106.

[3] D. Bredt and S. Snyder (1990). Isolation of nitric oxide synthetase, a calmodulin-requiring enzyme. *Proc. Natl. Acad. Sci. USA* **87**: 682-685.

[4] L. Cao, T. Blute and W. Eldred (2000). Localisation of heme oxygenase-2 and modulation of cGMP levels by carbon monoxide and/or nitric oxide. *Visual Neuroscience,* **17**: 319-379.

[5] E. Carlsen and J. Comroe (1958). The rate of uptake of carbon monoxide and nitric oxide by normal humanerythrocytes and experimentally produced spherocytes. *J. Gen. Physiol.,* **42**: 83-107.

[6] H. Carslaw and J. Jaeger (1959). *Conduction of Heat in Solids.* Oxford University Press.

[7] B. Chen, M. Keshive and W. Deen (1998). Diffusion and reaction of nitric oxide in suspension cell cultures. *Biophysical Journal*, **75**: 745-754.

[8] J. Crank (1980). *The Mathematics of Diffusion.* Oxford University Press, Oxford, U.K..

[9] P. Davis and P. Rabinowitz (1984). *Methods of Numerical Integration*, 2nd ed. Academic Press, Orlando, Florida.

[10] P. Dayan and L.F. Abbott (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems,* MIT Press, Cambridge Massachusetts, U.S.

[11] J. DeFilipe (1993). A study of NADPH diaphorase-positive axonal plexuses in the human temporal cortex. *Brain Research*, **615**: 342-346.

[12] M. Elphick and L. Williams and M. O'Shea (1996). New features of the locust optic lobe: evidence of a role for nitric oxide in insect vision. *J. Exp. Biol.* **199:** 2395-2407.

[13] A. Fick (1855). *Ann. Phys.,* **170**: 59.

[14] J.A. Gally, P.R. Montague, G.G. Reeke Jr. and G.M. Edelman (1990) The NO hypothesis: possible effects of a short-lived, rapidly diffusible signal in the development and function of the nervous system. *Proc. Natl. Acad. Sci. USA,* **87:** 3547-3551.

[15] J. Garthwaite, S. Charles and R. Chess-Williams (1988) Endothelium-derived relaxing factor release on activation of NMDA receptors suggests role as intracellular messenger in the brain. *Nature* **336**: 385-388.

[16] J. Garthwaite and C. Boulton (1995). Nitric oxide signaling in the central nervous system. *Ann. Rev. Physiol.* **57:** 683-706.

[17] Z.W. Hall (1992) *An Introduction to Molecular Neurobiology.* Sinauer Associates Inc., Sunderland, Massachusetts.

[18] N.A. Hartell (1996) Strong activation of parallel fibres produces localized calcium transients and a form of LTD that spreads to distant synapses. *Neurons* **16**: 601-610.

[19] C. Holscher (1997) Nitric oxide, the enigmatic neuronal messenger: its role in

synaptic plasticity. *Trends Neurosci.* **20:** 298-303.

[20] P. Husbands and T. Smith and N. Jakobi and M. O'Shea (1998). Better living through chemistry: evolving GasNets for robot control, *Connection Science,* **10**(3&4): 185-210.

[21] P. Husbands, A. Philippides, T. Smith and M. O'Shea, M (2001). Volume signalling in real and robot nervous systems. *Theory in Biosciences* **120:** 251-268.

[22] B. Katz (1969) *The Release of Neural Transmitter Substances.* Liverpool University Press.

[23] C. Koch and I. Segev (1998). *Methods in Neuronal Modeling: From Ions to Networks,* 2nd edition, MIT Press, Cambridge Massachusetts, U.S.

[24] J. Lancaster (1994). Simulation of the diffusion and reaction of endogenously produced nitric oxide. *Proc. Natl. Acad. Sci. USA,* **91**: 8137-8141.

[25] J. Lancaster (1996). Diffusion of free nitric oxide. *Methods in Enzymology,* **268**: 31-50.

[26] J. Lancaster (1997). A tutorial on the diffusibility and reactivity of free nitric oxide. *Nitric Oxide,* **1:** 18-30.

[27] M. Laurent, M. Lepoivre and J.-P. Tenu (1996). Kinetic modelling of the nitric oxide gradient generated *in vitro* by adherent cells expressing inducible nitric oxide synthase. *Biochem. J.,* **314:** 109-113.

[28] X. Liu, M. Miller, M. Joshi, H. Sadowska-Krowicka, D. Clark and J. Lancaster (1998). Diffusion-limited reaction of free nitric with erythrocates. *Journal of biological chemistry,* **273**(30): 18709-18713.

[29] T. Malinski, Z. Taha, S. Grunfeld, S. Patton, M. Kapturczak and P. Tomboutliant (1993). Diffusion of nitric oxide in the aorta wall monitored in situ by porphyrinic microsensors. *Biochem. Biophys. Res. Commun.,* **193:** 1076-1082.

[30] M. Mascagni (1989). Numerical methods for neuronal modeling. In *Methods in Neuronal Modeling: From Ions to Networks*, 2nd edition, C. Koch and I. Segev (1998), MIT Press, Cambridge, Massachusetts, U.S.

[31] S. Moncada, R. Palmer and E. Higgs (1989). Biosynthesis of nitric oxide from l-arginine. *Biochem. Pharm.,* **38**: 1709-1715.

[32] M. O'Shea, R. Colbert, L. Williams and S.Dunn (1998). Nitric oxide compartments in the mushroom bodies of the locus brain. *NeuroReport* **3**: 333-336.

[33] J. Park, V. Straub and M. O'Shea (1998). Anterograde signaling by nitric oxide: characterization and *in vitro* reconstitution of an identified nitrergic synapse. *J. Neurosci.* **18**: 5463-5476.

[34] A. Philippides, P. Husbands and M. O'Shea (2000). Four dimensional neuronal

signaling by nitric oxide: a computational analysis. *Journal of Neuroscience* **20**(3): 1199–1207.

[35] A. Philippides (2001). *Modelling the Diffusion of Nitric Oxide in Brains.* DPhil thesis, University of Sussex.

[36] A. Philippides, P. Husbands, T. Smith and M. O'Shea (2002). Fast and loose: biologically inspired couplings. *Artificial Life VIII*, MIT Press.

[37] W. Press, S.Teukolsky, W. Vetterling, B. Flannery (1971). *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge University Press.

[38] H. Schmidt and U. Walter (1994). NO at work. *Cell* **78:** 919-925.

[39] G. Smith (1985). *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd ed., Clarendon Press, Oxford.

[40] T. Smith, P. Husbands, A. Philippides and M. O'Shea (2003). Neuronal plasticity and temperal adaptivity: GasNet robot control networks. *Adaptive Behaviour.* (in press).

[41] S. Snyder and C. Ferris (2001). Novel neurotransmitters and their neuropsychiatric relevance. *American Journal of Psychiatry* **157:** 1738-1751.

[42] J. Stone and M. Marletta (1996). Spectral and kinetic studies on the activation of soluble guanyly cyclase by nitric oxide. *Biochemistry* **35:** 1093-1099.

[43] D. Thomas, X. Liu, S. Kantrow and J. Lancaster (2001). The biological lifetime of nitric oxide: implications for the perivascular dynamics of NO and O2. *PNAS*, **98(1)**: 355-360.

[44] M. Vaughn, L. Kuo and J. Laio (1998a). Effective diffusion distance of nitric oxide in the microcirculation. *Am. J. Physiol.* **274**: 1705-1714.

[45] M. Vaughn, L. Kuo and J. Laio (1998b). Estimation of nitric oxide production and reaction rates in tissue by use of mathematical model. *Am. J. Physiol.* **274**: 2163-2176.

[46] S. Vincent (1994). Nitric oxide: a radical neurotransmitter in the central nervous systems. *Progress in Neurobiology,* **42:** 129-160.

[47] J. Wood and J. Garthwaite (1994). Model of the diffusional spread of nitric oxide - implications for neural nitric oxide signaling and its pharmacological properties. *Neuropharmacology* **33:** 1235-1244.

Chapter 5

Stochastic Modelling of Single Ion Channels

Alan G. Hawkes
European Business Management School, University of Wales, Swansea, SA2 8PP, U.K.

CONTENTS

5.1 Introduction

Biological cells are enclosed by a phospholipid bilayer, a few nanometres thick, that is almost impermeable to water and water soluble molecules. Ion channels are proteins embedded in the membrane that control the movement of ions across the membrane. The channel protein, also a few nanometres wide, is arranged in a roughly circular array that spans the membrane with a central aqueous pore that can open under certain conditions, allowing electrically charged ions to pass through the pore under the influence of a small electrical potential difference between the intracellular and extracellular sides of the membrane. Typically, different kinds of channels allow the passage of different ions, such as Na^+, K^+, Ca^{2+} or Cl^-. The flow of these charged ions constitutes a flow of electrical current.

The opening and closing of ion channels is called gating. The major types of gating mechanism are *voltage gated* (where channels respond to changes in the membrane potential) and *ligand activated* (where channels are activated by binding with molecules of certain chemicals): other types of channel may respond to changes in temperature or stretching by a mechanical force.

All electrical activity in the nervous system appears to be regulated by ion channel gating, see for example [2]. Channels play a role in many diverse activities including thought processes; transmission of nerve signals and their conversion into muscular contraction; controlling the release of insulin, so regulating the blood glucose level. Understanding their behaviour increases our understanding of normal physiology and the effect of drugs and toxins on an organism, especially the human body. It is therefore an important step towards developing treatments for a wide variety of medical conditions such as epilepsy, cystic fibrosis, and diabetes, to name but a few.

Measurements of electrical currents that are the superposition of currents through very large numbers of channels are called *macroscopic measurements*. For example, in the decay of a miniature endplate current at the neuromuscular junction several thousand channels are involved, a large enough number to produce a smooth curve in which the contribution of individual channels is impossible to see. Early examples include [1, 64, 71]. In this case the time course of the current is often a simple exponential. Forms of macroscopic measurements include (a) relaxation of the current following a sudden change in membrane potential (*voltage jump*), and (b) relaxation of the current following a sudden change in ligand concentration (*concentration jump*). In these cases too, it is common to observe that the time course of the current following the jump can be fitted by an exponential curve, or by the sum of several exponential curves with different time constants. An example is shown in Figure 5.1.

If, on the other hand, we record from a fairly small number of ion channels, the fluctuations about the average behaviour become large enough to measure. Suppose, for example, that there are 400 channels open on average: then the number of channels open will vary by random fluctuations between about 340 and 460. In [52, 53, 72] they showed how these fluctuations (or ‘noise’) could be interpreted in terms of the ion channel mechanism. An elementary discussion is given in [27].

Since the pioneering patch-clamp experiments of Neher and Sakmann [65], the techniques being further refined by [45], it has become routinely possible to observe electric currents of a few picoamperes flowing through a single ion channel in a biological membrane. A great deal of information about how this remarkable feat is achieved is given in [70]. Apart from noise and some inertia in the recording system, it soon becomes clear that we are essentially observing the opening and closing of a pore in the macromolecule that forms the channel. When the channel is open there is a current of approximately constant amplitude; when the channel closes the

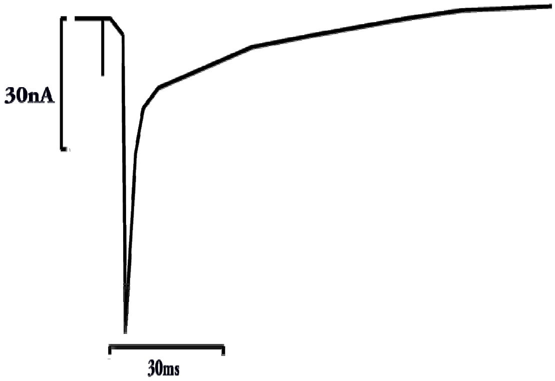

Figure 5.1
Relaxation after a jump in conditions: a mixture of two or three exponentials.

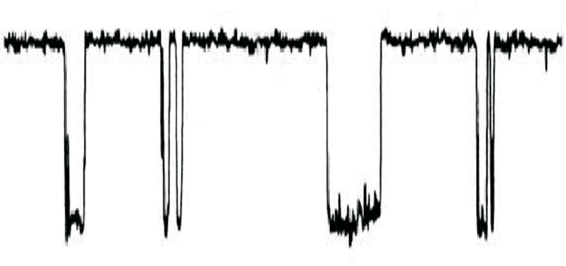

Figure 5.2
An example of current through a single ion channel

current stops; an example is given in Figure 5.2. Methods of extracting the idealised open and closed intervals from the noise are discussed in [35]; see also Section 5.6. There are channels in which there are identifiable sublevels of conductance but in this chapter we will deal almost exclusively with the common case of just two observable current levels, corresponding to channel open and channel closed.

The information in a single channel record is contained in the amplitudes of the openings, the durations of the open and shut periods, and correlations between them. We use this information to try to identify the nature of the kinetic mechanism of the channel and to estimate rate parameters governing transitions between the various states of such a mechanism. In order to do this we need to develop a mathematical model to describe the operation of any postulated mechanism and then use mathematical techniques to predict observable behaviour that the mechanism would exhibit. In this way we can confront theoretical mechanisms with observed data.

In this chapter we give a brief introduction to such models and what can be predicted from them. Inevitably, we shall need to introduce some mathematics but we will try to keep this as simple as possible and not go into too much detail.

5.2 Some basic probability

Activity at the level of individual molecules can be described by probabilistic laws. We need, therefore, to begin with a few basic ideas about probability. The general notation $Prob(A|B)$ denotes the probability of the event A conditional on the event B having occurred; we talk about the probability of the event A given B. We shall need a couple of simple results from probability theory.

First, the multiplication rule of probability says that for any two events A, B the joint probability

$$Prob(\text{A and B}) = Prob(A)Prob(B|A),$$

i.e., the probability of one event multiplied by the conditional probability of the other given the first.

An extension of this is the *Total Probability Theorem*, which says that given a set of *mutually exclusive events*, A_i (one of which **must** occur, but only one), the probability of any other event B can be written as

$$Prob(B) = \sum_i Prob(A_i)Prob(B|A_i)$$

These simple results can be used to solve a vast number of problems in probability theory, including most things we need to describe channel behaviour.

5.3 Single channel models

All of the mechanisms that we shall consider suppose that an ion channel consists of a single macromolecule that can exist in a number of different chemical states, either by itself or in association with molecules of some specific chemicals. For example, one or two molecules of an agonist may be attached to a receptor on the channel or a molecule of a channel blocker, such as a local anaesthetic, may block the pore of an open channel and prevent the passage of ionic current.

The transition rate between two chemical states always has the dimensions of a rate or frequency, viz s^{-1}. For a reaction that involves only a single molecule (e.g., a conformation change) the transition rate is simply the *reaction rate constant* defined by the law of mass action. The same is true of dissociation (unbinding) of a single molecule of a ligand that is bound to a receptor. For a reaction in which a free ligand binds to a receptor, the law of mass action states that the transition rate in this case is the product of the *rate constant* and the free ligand concentration.

The assumption that the transition rates are constant, i.e., they do not change with time, involves the assumption that the free concentration does not change with time; this is usually not true in daily life but is often approximately true in well-controlled

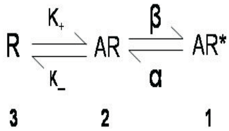

Figure 5.3
CK mechanism.

experiments. Similarly, the channel-shutting rate constant is known to be dependent on membrane potential (for muscle-type nicotinic receptors), so it will stay constant only if the membrane potential stays constant (i.e., only as long as we have an effective voltage clamp).

5.3.1 A three-state mechanism

If a ligand must be bound before the ion channel can open, at least three discrete states are needed to describe the channel mechanism. The mechanism of Castillo and Katz [22], the CK mechanism, has two shut states and one open state; this is usually represented as in Figure 5.3 where R represents a shut channel, R^* an open channel, and A represents the agonist molecule. The states have been numbered to facilitate later mathematical representation. State 1 is the open state in which an agonist molecule is bound to a receptor on the channel; in state 2 a molecule is bound but the channel is shut; in state 3 the channel is shut and its receptor is unoccupied. The rate constants are shown next to each possible transition.

A single channel makes transitions between its states in a random fashion and the transition rates will determine the probability distributions that describe the occupancy times of the various states and the states into which the transitions take place. In the CK mechanism, for example, the shutting rate, α, of an open channel must be interpreted in a probabilistic way: roughly, we can say that the probability of an open channel shutting in the next small interval of time Δt, is approximately $\alpha\Delta t$. More precisely, we can interpret the transition rate as

$$\alpha = \lim_{\Delta t \to 0} Prob(\text{channel shut at } t+\Delta t|\text{ channel open at } t)/\Delta t$$

Thus the transition rate is thought of as the limit of a conditional probability over a small time interval. Notice that this is supposed to be the same at whatever time t we start timing our interval, and also to be independent of what has happened earlier, i.e., it depends only on the present (time t) state of the channel. This is a fundamental characteristic of our type of random process (a homogeneous Markov process).

More generally, we can define any transition rate in this way. Denote by q_{ij} the transition rate from state i to state j. Then, for j not equal to i,

$$q_{ij} = \lim_{\Delta t \to 0} Prob(\text{ channel in state } j \text{ at time } t+\Delta t|\text{ channel in state } i \text{ at time } t)/\Delta t$$

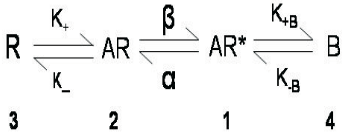

Figure 5.4
CK mechanism with four states.

Thus, for small Δt, the conditional probability is given approximately by

$$Prob(\text{ channel in state } j \text{ at time } t+\Delta t|\text{ channel in state } i \text{ at time } t)/\Delta t \simeq q_{ij}\Delta t$$

It is convenient to define q_{ii} as minus the sum of the transition rates away from state i: that is $q_{ii} = -\sum_{k\neq i} q_{ik}$. Then the probability of remaining in the same state is

$$\begin{aligned}
&Prob(\text{ channel in state } i \text{ at time } t+\Delta t|\text{ channel in state } i \text{ at time } t)\\
&= 1 - Prob(\text{moving to some other state})\\
&\simeq 1 - \sum_{k\neq i} q_{ik}\Delta t = 1 + q_{ii}\Delta t
\end{aligned}$$

If there are m states, the $m \times m$ square matrix $\mathbf{Q}$, whose elements are q_{ij}, is called the transition rate matrix or Q-matrix. The elements in each of its rows sum to zero. For example, the Qf-matrix for the CK model is

$$\mathbf{Q} = \begin{pmatrix} -\alpha & \alpha & 0 \\ \beta & -(k_{-1}+\beta) & k_{-1} \\ 0 & k_{+1}x_A & -k_{+1}x_A \end{pmatrix} \tag{5.1}$$

Note that, as discussed above, the transition rate for binding is the product of the rate constant k_{+1} and the ligand concentration x_A. Each of the other transition rates is given simply by the appropriate rate constant.

When the channel leaves state i it moves into state j with probability $-q_{ij}/q_{ii}$. Thus, for example, when the channel leaves state 2 (shut with a bound molecule) it opens, with probability $\beta/(k_{-1}+\beta)$, or the bound molecule dissociates, with probability $k_{-1}/(k_{-1}+\beta)$.

5.3.2 A simple channel-block mechanism

Suppose that the open channel in the CK model can be blocked by a molecule of a blocker substance, whose concentration is x_B. The model that had two shut states and one open state now has an additional shut state, so it can be represented as in Figure 5.4

The Q-matrix now becomes

$$\mathbf{Q} = \begin{pmatrix} -(\alpha + k_{+B}x_B) & \alpha & 0 & k_{+B}x_B \\ \beta & -(\beta + k_{-1}) & k_{-1} & 0 \\ 0 & k_{+1}x_A & -k_{+1}x_A & 0 \\ k_{-B} & 0 & 0 & -k_{-B} \end{pmatrix} \tag{5.2}$$

Note that the transition rate from open to blocked, like the binding transition, is given by the product of the rate constant and the drug concentration, this time the blocker concentration x_B. Now when the channel is observed to be shut we cannot tell whether it is blocked or closed, although we do know that the proportion of shut times that are blocked is $k_{+B}x_B/(\alpha + k_{+B}x_B)$, the probability of leaving state 1 for state 4 rather than state 2. Moreover, depending on the relative values of some of the transition rates (for example if k_{-B} is somewhat greater than β), it may be that short shut times are more likely to be the result of blocking and long shut times more likely to be the result of shutting, i.e., a sojourn among the pair of shut states $(2,3)$.

5.3.3 A five-state model

We now consider a model, introduced by [29], that has been used by several authors to describe the nicotinic acetylcholine receptor. In this mechanism there may be one agonist molecule (A) or two molecules (A_2) bound to the shut receptor (R) or the open receptor (R^*). In the following diagram that represents this model three shut states $(3,4,5)$ are shown on the bottom row and two open states $(1,2)$ on the top; on the right two agonist molecules are bound, one in the middle and none on the left. If it were possible for the channel to open in the absence of bound agonist then there would be another open state at the top left of the diagram.

Note that the rate constant for binding one molecule when the channel is free is written as $2k_{+1}$ because there are two free receptor sites; similarly, the dissociation rate constants for the unbinding of one of two occupied receptor sites are written as $2k_{-2}$ (for the closed channel) and $2k^*_{-2}$ (for the open channel) (see Figure 5.5).

The transition rate matrix is

$$\mathbf{Q} = \begin{pmatrix} (\alpha_1 + k^*_{+2}x_A) & k^*_{+2}x_A & 0 & \alpha_1 & 0 \\ 2k^*_{-2} & -(\alpha_2 + 2k^*_{-2}) & \alpha_2 & 0 & 0 \\ 0 & \beta_2 & -(\beta_2 + 2k_{-2}) & 2k_{-2} & 0 \\ \beta_1 & 0 & k_{+2}x_A & -(\beta_1 + k_{+2}x_A + k_{-1}) & k_{-1} \\ 0 & 0 & 0 & 2k_{+1}x_A & -2k_{+1}x_A \end{pmatrix} \tag{5.3}$$

In particular, suppose that $k_{+1} = 5 \times 10^7 M^{-1}s^{-1}; k_{+2} = k^*_{+2} = 10k_{+1}$, so that when one agonist molecule is bound the second receptor site is more likely to bind another agonist molecule; dissociation rates for the shut conformation $k_{-1} = k_{-2} = 2000s^{-1}$. Suppose the opening and shutting rates of the singly occupied state are $\beta_1 = 15s^{-1}$, $\alpha_1 = 3000s^{-1}$ while those for the doubly occupied state are $\beta_2 = 15000s^{-1}$, $\alpha_2 = 500s^{-1}$: thus the singly occupied state is slow to open and quick to shut, while

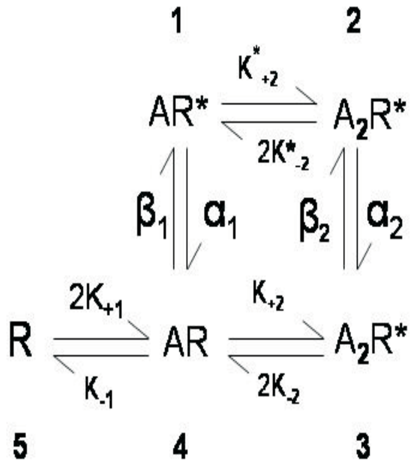

Figure 5.5
A five-state model.

the doubly occupied state opens very much more quickly and closes somewhat more slowly than the singly occupied state. Take the agonist concentration as $x_A = 100nM$.

There is one more rate constant to fix, the dissociation rate k^*_{-2} from the doubly occupied open state. To obtain this we appeal to the principle of *microscopic reversibility*. This states that, in the absence of a source of energy, a system will move to thermodynamic equilibrium in which each individual reaction will proceed, on average, at the same rate in each direction. In particular, if there is a cycle in the reaction, there can be no tendency to move round the cycle in one particular direction. In the model under consideration the states $1, 2, 3, 4$ form a cycle and the assumption of microscopic reversibility implies that the product of transition rates around the cycle are the same in both directions, i.e., $k^*_{+2}x_A\alpha_2 2k_{-2}\beta_1 = \alpha_1 k_{+2}x_A\beta_2 2k^*_{-2}$. Then, with the constants previously defined, $k^*_{-2} = k^*_{+2}(\alpha_2/\beta_2)(k_{-2}/k_{+2})(\beta_1/\alpha_1) = 1/3$.

With these values the transition rate matrix becomes

$$\mathbf{Q} = \left(\begin{array}{cc:ccc} -3050 & 50 & 0 & 3000 & 0 \\ 0.667 & -500.667 & 500 & 0 & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & 15000 & -19000 & 4000 & 0 \\ 15 & 0 & 50 & -2065 & 2000 \\ 0 & 0 & 0 & 10 & -10 \end{array} \right) \tag{5.4}$$

The reason for partitioning of the above matrix will be explained later. As mentioned before, when the channel leaves state i it moves into state j with probability $-q_{ij}/q - ii$. Thus, if we divide the non-diagonal elements of each row of the matrix **Q** by minus the diagonal element on that row, we get a new matrix whose rows represent conditional probability distributions of the next state to be entered. The diagonal elements must be zero.

In this model we get the matrix

$$\begin{pmatrix} 0 & 0.0164 & 0 & 0.9836 & 0 \\ 0.0013 & 0 & 0.9987 & 0 & 0 \\ 0 & 0.7895 & 0 & 0.2015 & 0 \\ 0.0073 & 0 & 0.0242 & 0 & 0.9685 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{5.5}$$

Thus, for example, a channel in state 4 (AR) has a probability of 0.0073 of opening (move to state1), probability 0.0242 of binding a second agonist molecule but has a very high probability of 0.9585 of losing its agonist molecule (move to state 5). In contrast, a doubly occupied channel A_2R (state 3) has a probability of 0.7895 of opening (move to A_2R^*, state 2) rather than losing one of its agonist molecules.

There are many similar models that can be constructed in this way, but the above will suffice for this chapter. Some models can get quite large: Rothberg and Magleby [68] have considered a 50-state model for a calcium-activated potassium channel. Ball [20] has studied a model based on molecular structure that has 128 states of which 4 are open: by exploiting assumptions of symmetry this effectively reduces to a model with 3 open states and 69 closed states – still quite big!

Although we shall see that it is possible to eliminate some models for a particular mechanism on the basis of observable characteristics, a certain amount of indeterminacy arises because we cannot see everything that a channel does (we cannot see which individual state the channel is in, only if it is open or closed). It is possible that two or more distinct models may give rise to the same observable features under fixed conditions, see for example [38, 40, 55]. However, further discrimination between models is possible by observing the same channel under different conditions, changing voltages or agonist concentration.

5.4 Transition probabilities, macroscopic currents and noise

In order to predict macroscopic currents and the behaviour of noise measurements, we first need to study some transition probabilities for a single channel.

5.4.1 Transition probabilities

Assuming that we know the current state of the system, it is useful to predict the state that the system might be in at some time t later. Let $X(t)$ denote the state occupied by the mechanism at time t then, for $i, j = 1$ to m and $t > 0$, we define transition probabilities $p_{ij}(t)$ by

$$p_{ij}(t) = P(X(t) = j|X(0) = i)$$

Then by the total probability theorem, if we consider the situation at times t and $t + \Delta t$,

$$p_{ij}(t+\Delta t) = \sum_k p_{ik}(t) Prob(X(t+\Delta t) = j|X(0) = i \text{ and } X(t) = k)$$

But the Markov property implies that the condition in the second factor of this expression can be replaced by the condition $X(t) = k$ only (the most recent thing known) so that we have, approximately,

$$p_{ij}(t+\Delta t) = \sum_{k \neq j} p_{ik}(t) q_{kj} \Delta t + p_{ij}(t)\{1 + q_{jj}\Delta t\}$$

Then the derivative

$$p'_{ij}(t) = \lim_{\Delta t \to 0} \{p_{ij}(t+\Delta t) - p_{ij}(t)\}/\Delta t = \sum_k p_{ik}(t) q_{kj}$$

or, if the square matrix $\mathbf{P}(t)$ has elements $p_{ij}(t)$ we have the matrix differential equation

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q} \tag{5.6}$$

which has formal solution

$$\mathbf{P}(t) = \exp(\mathbf{Q}t), \qquad t > 0 \tag{5.7}$$

The initial value is $\mathbf{P}(0) = \mathbf{I}$, an identity matrix, because $p_{ii}(0) = 1$ as the system cannot move anywhere in zero time.

So what is the exponential of a matrix? We can define it by a matrix version of the usual series expansion

$$\exp(\mathbf{Q}t) = \mathbf{I} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{Q}^n$$

A 'cookbook' approach to programming the calculations of this matrix function is provided in [32] but, theoretically, the nicest result arises from the so-called spectral expansion of the matrix $\mathbf{Q} = -\sum_{i=1}^{m} \lambda_i \mathbf{A}_i$, where λ_i the are the eigenvalues of the matrix $-\mathbf{Q}$. The spectral matrices $\mathbf{A}_i$ may be calculated from the eigenvectors of $\mathbf{Q}$. The details need not concern us here but the important thing to note is the property that

$$\mathbf{A}_i \mathbf{A}_j = 0, \;\; i \neq j; \qquad \mathbf{A}_i^2 = \mathbf{A}_i$$

If we substitute the spectral form into the series expansion for the exponential, we see that

$$\exp(\mathbf{Q}t) = \sum_{i=1}^{m} \exp(-\lambda_i t)\mathbf{A}_i \tag{5.8}$$

5.4.2 Macroscopic currents and noise

Now let $\pi_i(t) = Prob(X(t) = i)$ be the probability that a channel is in state i at time t. Then, using once again the total probability theorem and conditioning on the state occupied at time zero,

$$\pi_j(t) = \sum_i \pi_i(0) p_{ij}(t)$$

which can be expressed in matrix terms as

$$\pi(t) = \pi(0)\mathbf{P}(t) \tag{5.9}$$

where $\pi(t) = (\pi_1(t), \pi_2(t), \cdots, \pi_m(t))$ is a row vector of occupancy probabilities. In equilibrium conditions the occupancy probabilities should be constant (independent of time). To find this, set the derivative of $\pi(t)$ to zero to get

$$0 = \pi'(t) = \pi(0)\exp(\mathbf{Q}t)\mathbf{Q} = \pi(t)\mathbf{Q}$$

If we omit t, because we now suppose it is independent of t, the row vector, π, of equilibrium occupancies satisfies

$$0 = \pi\mathbf{Q} \text{ together with } \sum_i \pi_i = 1 \tag{5.10}$$

because the probabilities must sum to one.

Now suppose that the current flowing through a single channel is given by $gamma_i$ when the channel is in state i; let the row vector $\gamma^T = (\gamma_1, \gamma_2, \cdots, \gamma_m)$. If there is a large number, N, of similar channels, the macroscopic current can be predicted as N times the expected current through a single channel: thus, using the spectral expansion (5.8),

$$\begin{aligned} I(t) &= N\sum_i \pi_i(t)\gamma_i = N\pi(t)\gamma = N\pi(0)\mathbf{P}(t)\gamma \\ &= N\pi(0)\exp(\mathbf{Q}t)\gamma = \textstyle\sum_{i=1}^{m} w_i \exp(-\lambda_i t) \end{aligned} \tag{5.11}$$

where the scalar weight $w_i = N\pi(0)\mathbf{A}_i\gamma$.

Because the rows of the matrix $\mathbf{Q}$ all sum to zero, it follows that one of the eigenvalues, say λ_1, is zero and so the first term in the above expression will be the constant w_1. Also, at least for reversible processes, it can be shown that all the other eigenvalues are real positive numbers.

Some of the transition rates in the matrix $\mathbf{Q}$ depend on the agonist concentration and/or the voltage difference across the membrane; they do so in an explicit manner for the effect of agonist in agonist gated channels (see, for example, the Q-matrices for the three models introduced in Section 5.3). If we observe the macroscopic current following a jump in agonist concentration or voltage applied in such a way that these are held constant at the new values after the jump, then the current relaxes as a mixture of $(m-1)$ exponential components towards the constant value w_1.

The time constants of these components, the eigenvalues of the matrix $-\mathbf{Q}$, depend on the values of all of the rate constants that pertain in the conditions following the jump. They have, in general, no simple physical significance, although in particular cases they may approximate some physical quantity such as mean open lifetime. Observation of the current therefore tells us something about the number of states the channel may occupy: thus, for example, we expect to see 2 components in the CK model, 3 in the simple channel blocker model and 4 in the 5-state model. However, it often happens that some components of the mixture correspond to large values of λ and small values of w (i.e., very short lived components with very small weight) so that they may be very difficult to detect in practice. The number of states may therefore be greater than 1 plus the apparent number of components in the relaxation observed in experiments.

When we consider the analysis of noise experiments in equilibrium conditions, similar arguments lead to an expression for the autocovariance function of the current fluctuations in the form

$$C(t) = Cov(I(s), I(s+t)) = \sum_{i=2}^{m} \alpha_i \exp(-\lambda_i t)$$

which is a mixture of exponentials with the same $(m-1)$ time constants as for the relaxation equations. Details may be found in [28].

5.5 Behaviour of single channels under equilibrium conditions

In this section we consider probability distributions that describe the behaviour of a single channel under equilibrium conditions of constant ligand concentration and voltage difference.

5.5.1 The duration of stay in an individual state

We are interested in the length of time for which the system stays in a particular state, for example the single open state in the CK model. Each time the channel opens the duration of its stay in the open state varies. These durations are random variables, and we wish to find the probability density function (pdf) that describes this variability. This is a function $f(t)$, defined so that the area under the curve up to a particular time t represents the probability that the duration (or lifetime) is equal to or less than t. Thus, if we denote a random lifetime by T, the cumulative distribution (or *distribution function*), which is usually denoted $F(t)$, is the probability that a lifetime does not exceed t: it is given by

$$F(t) = Prob(T \leq t) = \int_{-\infty}^{t} f(s)ds$$

Conversely, the pdf can be found by differentiating $F(t)$. Thus,

$$\begin{aligned} f(t) &= \frac{dF}{dt} = \lim_{\Delta t \to 0} [F(t+\Delta t) - F(t)]/\Delta t \\ &= \lim_{\Delta t \to 0} [Prob(\text{ lifetime is between } t \text{ and } t+\Delta t\ ]/\Delta t \end{aligned}$$

In order to derive this function it is convenient to introduce the function

$$R(t) = 1 - F(t)$$

This is just the probability that a lifetime is greater than t, and so it is often called the *survivor function* or the *reliability function.*

Now, a channel that is open at time $t = 0$ will remain open throughout the interval from 0 to $t + \Delta t$ if it remains open from 0 to t and then remains open for a further time Δt. Thus

$$\begin{aligned} R(t+\Delta t) &= Prob(T > t + \Delta t) \\ &= Prob(\text{ open throughout 0 to } t) Prob(\text{ open at } t+\Delta t | \text{open at } t) \\ &= R(t)(1 - Prob(\text{ shut at } t+\Delta t | \text{open at } t)) \end{aligned}$$

This is an example of the general multiplication rule of probability and also uses the crucial Markov assumption, discussed earlier, that the conditional probability used here depends only on the channel being open at time t and not on the behaviour prior to that time: that is

$$Prob(\text{ open at } t+\Delta | \text{open throughout 0 to t}) = Prob(\text{ open at } t+\Delta t | \text{open at } t)$$

Then

$$\begin{aligned} \frac{dR}{dt} &= \lim_{\Delta t \to 0} \frac{R(t+\Delta t) - R(t)}{\Delta t} \\ &= -\lim_{\Delta t \to 0} R(t) Prob(\text{shut at } t+\Delta t | \text{open at } t)/\Delta t = -\alpha R(t) \end{aligned}$$

As long as α is a constant (not time dependent), the solution of this equation is

$$R(t) = \exp(-\alpha t) \tag{5.12}$$

because $R(0) = 1$ (i.e., channel cannot move out of the open state in zero time). Then the cumulative distribution function is $F(t) = 1 - R(t) = 1 - \exp(-\alpha t)$.

The required pdf for the open-channel lifetime is the first derivative of this, i.e.,

$$f(t) = \frac{dF}{dt} = -\frac{dR}{dt} = \alpha \exp(-\alpha t) \qquad \text{for } t > 0 \tag{5.13}$$

The density is, of course zero for $t < 0$.

state	1	2	3	4	5
mean	0.328ms	1.997ms	52.6μs	0.484 ms	100ms

Table 5.1 Mean lifetimes of sojourns in individual states.

This pdf is described as an *exponential distribution*, or exponential density, with mean $1/\alpha$. It is a simple exponentially decaying curve. The exponential distribution has a central role in stochastic processes like the Gaussian distribution has in large areas of classical statistics.

For any pdf the mean is given by

$$\text{mean} = \int_{-\infty}^{\infty} t f(t) dt$$

or where, as here, $f(t) = 0$ for $t < 0$ the lower limit may be taken as zero. Then

$$\text{mean} = \int_{0}^{\infty} t f(t) dt \tag{5.14}$$

which is $1/a$ in this case.

The above argument concerned the single open state in the CK mechanism. A similar argument may be used in general, so lifetimes in the *ith* state of any mechanism have an exponential distribution

$$\text{pdf } f(t) = -q_{ii} \exp(q_{ii}t); \text{ for } t > 0$$

and

$$\text{mean } = -1/q_{ii} \tag{5.15}$$

remember that $-q_{ii}$ is the sum of the transition rates away from state i.

In particular, from Equation (5.1), lifetimes of the unbound shut state (state 3) in the CK mechanism have an exponential distribution with mean $1/k_{+1}x_A$ and lifetimes in the occupied shut state (state 2) have an exponential distribution with mean $1/(k_{-1}+\beta)$. In the latter case each lifetime will end with a transition into state 3 (by dissociation of the bound ligand from the receptor) with probability $k_{-1}/(k_{-1}+\beta)$ or a transition into state 1 (by a conformation change that results in the channel opening) with probability $\beta/(k_{-1}+\beta)$.

Similarly, for the 5-state model, the reciprocals of (minus) the diagonal elements of (5.4) give the mean lifetimes of sojourns in individual states as in Table 5.1.

5.5.2 The distribution of open times and shut times

If there is only one open state, as in the CK model or the simple channel-block model, the distribution of the duration of open times is just the distribution of sojourns in a single state. If there is more than one open state then an open time starts when the channel leaves a shut state for an open state, takes a tour round various open states and then enters a shut state. For example, in the 5-state model an open time might start with a transition from state 4 to state 1, make several transitions back and forth between states 1 and 2 then end with a transition from state 1 to state 4 or from state 2 to state 3; alternatively, an open time might start with a transition from state 3 to state 2 then continue in a similar manner to that previously described.

To solve this problem it is convenient to arrange that all the open states be labelled as states $1, 2, \cdots\ m_o$ where m_o is the number of open states; the shut (or closed) states having the highest numbered labels $m_o + 1$ to $m_o + m_c = m.$ Then the Q-matrix can be partitioned in the form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{oo} & \mathbf{Q}_{oc} \\ \mathbf{Q}_{co} & \mathbf{Q}_{cc} \end{pmatrix} \tag{5.16}$$

where $\mathbf{Q}_{oo}$, a square matrix of dimension $m_o \times m_o$, contains all the transition rates between open states; $\mathbf{Q}_{cc}$ contains all the transition rates between closed states; $\mathbf{Q}_{oc}$, $\mathbf{Q}_{co}$ contain, respectively, the transition rates from open to closed states and from closed to open states. These partitions for the 5-state model are shown in Equation (5.4).

Now let $\mathbf{R}_o(t)$ be a matrix function whose *ijth* element, where i and j are open states, is

$$Prob(X(t) = j \text{ and channel open throughout time 0 to t} | X(t) = i).$$

Then, by an argument similar both to that used in deriving the matrix function $\mathbf{P}(t)$ in Equation (5.7) and that used in deriving the reliability $R(t)$ in Equation (5.12), we can show that

$$\mathbf{R}_o(t) = \exp(\mathbf{Q}_{oo}t) \ \ t > 0 \tag{5.17}$$

To get the reliability function of open times we have to sum over the possible open states that the channel might be in at time t and also, with a suitable weight function, sum over the states that an open time might start in: so we get

$$R_o(t) = Prob(\text{ open time } > t) = \Phi_o \exp(\mathbf{Q}_{oo}t)\mathbf{u}_o \tag{5.18}$$

In this equation $\mathbf{u}_o$ is a column vector of m_o 1's and Φ_o, the row vector of probabilities for the initial state of an open time, is given (in a different notation) by Colquhoun and Hawkes (in [29], Equation 5.4) as

$$\Phi_o = \pi_c \mathbf{Q}_{co} / \pi_c \mathbf{Q}_{co} \mathbf{u}_o \tag{5.19}$$

where π_c is a row vector corresponding to the part of the equilibrium occupancy vector that deals with the closed states only, i.e., the equilibrium occupancy vector see Equation (5.10) is partitioned as $\pi = (\pi_o, \pi_c)$.

We get the probability density function of open times by differentiating the reliability function

$$f_o(t) = -R_0'(t) = -\Phi_o \exp(\mathbf{Q}_{oo}t)\mathbf{Q}_{oo}\mathbf{u}_o \tag{5.20}$$

Then mean open time is given by

$$\mu_o = -\Phi_o \mathbf{Q}_{oo}^{-1} \mathbf{u}_o \tag{5.21}$$

where, as usual, the inverse of a matrix is denoted by raising it to the power -1. The results for the duration of a sojourn in a single state, Equation (5.15), are obviously obtained from this when $\mathbf{Q}_{oo}$ is replaced by a scalar q_{ii}. A very important thing to notice is that, if we carry out a spectral resolution of the matrix $-\mathbf{Q}_{oo}$, we see that, like the derivation of the relaxation equation 5.11, the probability density function $f_o(t)$ can be expressed as a mixture of exponential components with time constants given by the m_o eigenvalues of $-\Phi_{oo}$. Note that, unlike the matrix $-\mathbf{Q}$, it will not normally have a zero eigenvalue.

Similarly, just by interchanging o and c, we get the probability density of shut times as

$$f_c(t) = -\Phi_c \exp(\mathbf{Q}_{cc}t)\mathbf{Q}_{cc}\mathbf{u}_c$$

with mean shut time

$$\mu_c = -\Phi_c \mathbf{Q}_{cc}^{-1}\mathbf{u}_c$$

$\mathbf{u}_c$ is a column vector of m_c 1's and Φ_c a row vector of probabilities for the initial state of the shut time calculated by an expression like Equation (5.17) with the o and c interchanged. This can be expressed as a mixture of exponential components with time constants given by the m_c eigenvalues of the matrix $-\mathbf{Q}_{cc}$.

So the distributions of open times and shut times tell us something about the numbers of open states and shut states, again with the caveat that we might not be able to distinguish all components from an experimental record.

5.5.3 Joint distributions

Useful information about the structure of a channel mechanism may be obtained by looking at joint distributions of adjacent intervals: for example, does a long open interval tend to be followed by a short shut interval or vice versa? The problem can be approached by defining a transition density matrix $\mathbf{G}_{oc}(t_o)$ which has dimension $m_o \times m_c$ and whose *ijth* element is a joint probability/probability density for the duration of an open time, T_o, and the shut state that the system moves to when the open time ends, all conditional on starting in the *ith* open state: i.e.,

$$g_{ij}(t_o) = \lim_{\Delta t_o \to 0} Prob(t_o \leq T_o \leq t_o + \Delta t_o \text{ and enter } jth \text{ closed state} \\ \mid \text{starting in } ith \text{ open state})$$

This is given simply by

$$\mathbf{G}_{oc}(t_o) = \mathbf{R}_o(t_o)\mathbf{Q}_{oc} = \exp(\mathbf{Q}_{oo}t)\mathbf{Q}_{oc} \qquad (5.22)$$

If we are only interested in the shut state that is entered at the end of the open time, and not the duration of the open time, we can integrate the above expression with respect to t_o to obtain a transition probability matrix

$$\mathbf{G}_{oc} = \int_0^\infty \mathbf{G}_{oc}(t_o)dt_o = -\mathbf{Q}_{oo}^{-1}\mathbf{Q}_{oc} \qquad (5.23)$$

The joint probability density of an open time T_o and the immediately following shut time T_c is given by

$$\Phi_o \mathbf{G}_{oc}(t_o)\mathbf{G}_{co}(t_c)\mathbf{u}_c \tag{5.24}$$

Similar results may be obtained for any pair of intervals, for example two successive open times (separated by a single shut time) have joint probability density

$$\Phi_o \mathbf{G}_{oc}(t_{o1})\mathbf{G}_{co}\mathbf{G}_{oc}(t_{o2})\mathbf{u}_c$$

See [34] for examples of detailed study of such joint distributions.

The value of looking at joint distributions to distinguish between mechanisms was demonstrated practically by McManus et al. [21, 58, 59, 60].

We can also use these matrices to build up a likelihood for a complete sequence of open times and shut times: if there are M pairs of open and following shut times (t_j, s_j) this takes the form

$$\Phi_o \prod_{j=1}^{M} [\exp(\mathbf{Q}_{oo} t_j)\mathbf{Q}_{oc} \exp(\mathbf{Q}_{cc} s_j)\mathbf{Q}_{co}]\mathbf{u} \tag{5.25}$$

This can be maximised to estimate the parameters of a given model and to test the fit of a model to data – see [17, 38, 49] for details. Using all the information in this way removes some identifiability problems that can occur when using marginal distributions only, see [37], although some identifiability problems may remain.

5.5.4 Correlations between intervals

One way of studying these joint distributions is to look at correlations between the durations of adjacent intervals. The Markov assumption implies that if the system is in a specified state at time t, the future evolution of the system is independent of what happened before time t. The lifetimes of sojourns in individual states are therefore independent of each other. Correlations between open times or shut times can occur, however, if there are at least two open states and two shut states. It can be shown that the correlation between the duration of an open time and the nth subsequent open time has the form

$$\rho_n = \sum w_i \lambda_i^n$$

where the number of terms in the summation is $V - 1$. The λ_i are those eigenvalues of $\mathbf{G}_{oc}\mathbf{G}_{co}$ that are neither zero nor one. Similar results hold for correlations between the durations of shut times and between shut times and open times.

V is known as the (vertex) connectivity of the mechanism and is a measure of the extent to which the set of open states and the set of shut states are connected to each other: it is defined as the smallest number of states that need to be removed (together with any links that they have) in order to separate the set of open states and the set of shut states. We can do this in both the CK model and the simple channel block by removing just one state (either state 1 or state 2), so $V = 1$ and all durations of open times and shut times are uncorrelated, indeed they are mutually independent.

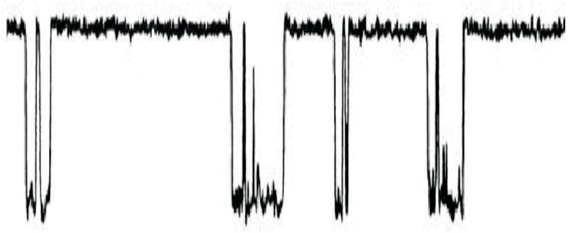

Figure 5.6
An example showing four bursts of openings containing brief shuttings and separated by longer gaps between bursts.

In the 5-state model we need to remove at least two states to separate the open and shut states: we could remove any of the pairs $(1,2),(3,4),(1,3)$ or $(2,4)$. So $V = 2$ and correlations would die away with lag n as a single geometric term.

Correlations between intervals are useful in telling us something about the connectivity between open and shut states.

The results on correlations were given by Fredkin et al. [38] and were extended by Colquhoun and Hawkes [30], Ball and Sansom [16], Ball et al. [7], Ball and Rice [13] among others.

Colquhoun and Hawkes [30] also studied the distribution of openings and shuttings after a jump in agonist concentration or voltage. The first latency (time to first opening) has a different distribution to subsequent shut time durations. If $V > 1$ there continue to be differences among the subsequent durations of both open and shut times; if $V = 1$, however, all open time durations have the same distribution (no matter if it is the first, second etc. opening after the jump) while all shut times apart from the first latency have the same distribution. This happens because of the independence of the various intervals.

Results on the distribution of openings and shuttings as a result of a finite pulse, rather than a single jump, were discussed in [33].

5.5.5 Bursting behaviour

It is usually the case that openings seem to occur in bursts of activity: a sequence of openings will be interspersed with brief shuttings and then there will be a long shut period before the activity starts again. This behaviour can often be largely explained by dividing the shut states into two categories: short-lived shut states and long-lived shut states. For example, in the simple channel blocker, if the rate constant k_{-B} is large then the duration of a stay in the blocked state will be very short and a burst of openings will most likely consist of oscillations between the open and blocked states. An example of bursting behaviour is shown in Figure 5.6.

Considering the 5-state model as another example, we see that the mean duration of a stay in state 5, the unbound state, is very much longer than that in any other state (see Table 5.1). Then shut times (gaps) within bursts of openings are almost certainly sojourns within the pair of shut states (3,4); gaps between bursts will consist of any sojourn within the shut states that includes at least one visit to state 5. Colquhoun and Hawkes [29] provide a detailed treatment of bursting behaviour in general and the 5-state model in particular. For this particular case the distribution of gaps within bursts is a mixture of two exponentials (because there are two short-lived shut states) with a mean of 57.6 μs whereas the distribution of gaps between bursts is more complex and has a mean duration of 3790ms.

Actually the information concerning gaps between bursts is rather dubious because it is possible that there is more than one channel in the recording environment so that, while the activity within a burst of openings almost certainly arises from one channel, different bursts may arise from different channels, making the mean inter-burst gap shorter than it should be. This is one reason why it is of interest to study the behaviour of bursts, as the information arising from within a burst should be fairly reliable. In addition, burst behaviour provides evidence about the finer structure of the underlying process.

Other distributions derived include those for:

- The number of openings per burst (a mixture of 2 geometric distributions with mean 3.82);
- The duration of a burst (a mixture of 4 exponentials with mean of 7.33ms);
- The total open time per burst (a mixture of 2 exponentials with a mean of 7.17ms);
- The distribution of individual openings within a burst (the first, second etc.).

Some results for the 5-state model are shown in parentheses.

If we have data from a channel that really comes from a 5-state model it should be possible to infer from the empirical distributions discussed above that there are at least two open states, at least two short-lived shut states and at least one long-lived shut state.

Methods for studying empirically whether bursting activity takes place are given by Ball and Sansom [14], Ball and Davies [6] and Ball [3].

5.6 Time interval omission

One big problem in observing single channel records is that, because of noise and inertia in the recording system, very short events, openings or shuttings, are likely to be missed. One can see that this is to be expected by looking at the examples in Figure 5.6 and Figure 5.7. The results then get distorted because, for example, what appears to be one long open time may actually be two or three open times separated by shut times that are too short to be distinguished. This is known as the problem of time interval omission (TIO). One way of coping with this problem is to study the total burst length or the total open time per burst, as these should not be very sensitive to missing short shut times.

In order to take account of missed events when dealing with the individual openings and shuttings it has been the custom to assume that there is a critical constant dead-time, ξ, such that all open times or shut times greater than this are observed accurately but times less than this are missed (note that a safe ξ value can be imposed retrospectively on recorded data). We then work with *apparent open times* defined as periods that start with an open time of duration greater than ξ which may then be extended by a number of openings separated by shut periods each of duration less than ξ; they are terminated at the start of a shut period of duration greater than ξ. Apparent shut times may be similarly defined.

A number of people presented approximate solutions for the distributions of apparent open times and apparent shut times before Ball and Sansom [15, 16] obtained exact results in the form of Laplace Transforms and also considered the effect of TIO on correlations between interval durations. Exact expressions for the probability density function of apparent open times and shut times were found by Hawkes et al. [46]. These are fine for small to moderate values of time t, but when trying to compute them they tend to be numerically unstable for large t. These results were also studied by Ball, Milne and Yeo [9, 10] in a general semi-Markov framework. In a series of papers, Hawkes et al. [46] and Jalali and Hawkes [50, 51] obtained asymptotic approximations that are extremely accurate for values of t from very large right down to fairly small; for small t the exact results are readily obtainable, so that the distributions are obtained over the whole range. Ball [3] studied these approximations further and showed mathematically why they are so very good. It is interesting to note that, if the true distribution is a mixture of k exponentials, then the approximation to the distribution of apparent times allowing for TIO is also a mixture of k exponentials: the time constants are, however, different.

These methods were subsequently applied by Colquhoun et al. [34] to study the effect of TIO on joint distributions of apparent open and shut times. They also used it to calculate the likelihood of a complete series of intervals and used this to estimate the parameters of any postulated mechanism - thus generalising the work of Ball and Sansom [17], who used similar methods for the ideal non-TIO case. For a given model, TIO can induce some indeterminacy in the process of estimating parameters. For data recorded under fixed conditions there can be two sets of parameters that seem to fit the data about equally well: typically a *fast solution* and a *slow solution*. A simple example is given by Colquhoun and Sigworth [35]. These can, however, be discriminated by observing the same channel under different conditions of voltage and/or ligand concentration, see [5, 20, 75].

Colquhoun et al. [33] and Merlushkin and Hawkes [62] studied the TIO problem in the context of recording apparent open and shut intervals elicited by a pulse of agonist concentration or voltage change.

5.7 Some miscellaneous topics

We conclude with a few topics that should be mentioned briefly but which are too advanced to be considered in detail in this introductory chapter.

5.7.1 Multiple levels

Almost all of the foregoing refers only to channels that are open or closed. Quite a few channels, however, show two or more different levels of current, presumably corresponding to different states of the mechanism. Examples are given in [44] and [73], the latter showing a channel with four conductance levels in addition to the shut states.

It is perfectly possible to ignore this and just analyse the system as open or closed, but this loses information. Li et al. [57] obtained some theoretical results for a chloride channel with two conductance levels (corresponding to one state each) and four shut states, of which one was short-lived. Ball, Milne and Yeo [12] give a general treatment of a multilevel system in which they derive burst properties including distributions of total charge transfer, and total sojourn time and number of visits to each conductance level during a burst. Building on unpublished work of Jalali, Merlushkin [61] has studied various apparent sojourn distributions allowing for TIO in the multilevel case.

Multiple levels sometimes arise from the presence of more than one channel: if so, they are usually treated as acting independently. However, Ball et al. [8], Ball and Yeo [19] and Ball, Milne and Yeo [11] introduced models for systems of interacting channels.

5.7.2 Hidden Markov Methods of analysis

A number of authors have applied signal processing methods, met in other areas such as speech processing, to the original noisy signals obtained from patch clamp experiments. These Bayesian or Hidden Markov methods have been used to extract the ideal step-function signals (representing opening and shutting) from the noise: in this task they work reasonably well at low signal to noise ratios where threshold methods work poorly. They are also used to estimate parameters in the models directly without identifying the individual open times and shut times. These techniques can cope with multilevel records as well as simply open/shut as illustrated in Figure 5.7.

Notable contributions in this area include [22, 25, 26, 36, 41, 42, 43, 54, 56, 63, 66, 67]. Venkataramanan and Sigworth [74] introduced a method of dealing with the problem of baseline drift that can badly affect application of this method. Markov Chain Monte Carlo (MCMC) methods of Bayesian analysis were applied by Ball et al. [4] and Hodgson [48].

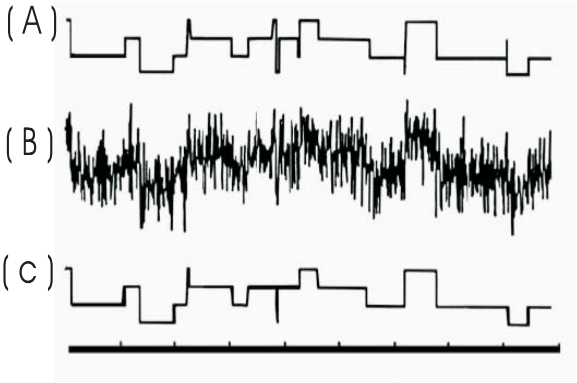

Figure 5.7

(A) shows a simulation of an ideal record from a mechanism with 4 conductance levels (B) shows the record with added noise (C) shows the signal recovered from the noise by HMM techniques - compared with the original it is remarkably good, missing just 4 very brief events.

References

[1] Adams, P. R. (1975). Kinetics of agonist conductance changes during hyperpolarization at frog endplates. *Br. J. Pharmac. Chemother.* **53**, 308-310.

[2] Aidley, D. J. and Stanfield, P. R. (1996). *Ion Channels: Molecules in Action.* Cambridge University Press, New York.

[3] Ball, F. G. (1997). Empirical clustering of bursts of openings in Markov and Semi-Markov models of single channel gating incorporating time interval omission. *Adv. Appl. Prob.* **29**, 909-946.

[4] Ball, F. G., Cai, Y., Kadane, J. B. and O'Hagan, A. (1999). Bayesian inference for ion-channel gating mechanisms directly from single-channel recordings, using Markov chain Monte Carlo. Proc. *R. Soc. Lond. A* **455**, 2879-2932.

[5] Ball, F. G. and Davies, S. S. (1995). Statistical inference for a two-state Markov model of a single ion channel, incorporating time interval omission. *J. R. Statist. Soc. B* **57**, 269-287.

[6] Ball, F. G. and Davies, S. S. (1997). Clustering of bursts of openings in Markov and semi -Markov models of single channel gating. *Adv. Appl. Prob.* **29**, 92-113.

[7] Ball, F. G., Kerry, C. J., Ramsey, R. L., Sansom, M. S. P. and Usherwood, P. N. R. (1988). The use of dwell time cross-correlation functions to study single ion channel gating kinetics. *Biophys. J.* **54**, 309-320.

[8] Ball, F. G., Milne, R. K., Tame, I. D. and Yeo, G. F. (1997). Superposition of interacting aggregated continuous-time Markov chains. *Adv. Appl. Prob.* **29**, 56-91.

[9] Ball, F., Milne, R. K., and Yeo, G. F. (1991). Aggregated semi-Markov processes incorporating time interval omission. *Adv. Appl. Prob.* **23**, 772-97.

[10] Ball, F. G., Milne, R. K. and Yeo, G. F. (1993). On the exact distribution of observed open times in single ion channel models. *J. Appl. Prob.* **30**, 529-537.

[11] Ball F. G, Milne R. K. and Yeo, G. F. (2000). Stochastic models for systems of interacting ion channels. *IMA J. Math. Appl. Med. Biol.* **17**, 263-293.

[12] Ball, R. K., Milne R. K. and Yeo, G. F. (2002). Multivariate semi-Markov analysis of burst properties of multiconductance single ion channels. *J. Appl. Prob.* **39**, 179-196.

[13] Ball, F. G. and Rice, J. A. (1989). A note on single-channel autocorrelation functions. *Math. Biosci.* **97**, 17-26.

[14] Ball, F. G. and Sansom, M. S. P. (1987). Temporal clustering of ion channel openings incorporating time interval omission. *IMA J. Math. Appl. Med. and Biol.* **4**, 333-361.

[15] Ball, F., and Sansom, M. S. P. (1988a). Aggregated Markov processes incorporating time interval omission. *Adv. Appl. Prob.* **20**, 546-72.

[16] Ball, F. G. and Sansom, M. S. P. (1988b). Single channel autocorrelation functions-the effects of time interval omission. *Biophys. J.* **53**, 819-832.

[17] Ball, F. G. and Sansom, M. S. P. (1989). Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. R. Soc. Lond. B* **236**, 385-416.

[18] Ball, F. G. and Yeo, G. F. (1994). Numerical evaluation of observed sojourn time distributions for a single ion channel incorporating time interval omission. *Statist. Comp.* **4**, 1-12.

[19] Ball, F. G. and Yeo, G. F. (1999). Superposition of spatially interacting aggregated continuous time Markov chains. *Meth. Comp. Appl. Prob.* **2**, 93-115.

[20] Ball, S.S. (2000). *Stochastic models of ion channels.* Ph.D. thesis, University of Nottingham.

[21] Blatz, A. L. and Magleby, K. L. (1989). Adjacent interval analysis distinguishes among gating mechanisms for the fast chloride channel from rat skeletal muscle. *J. Physiol. Lond.* **410**, 561-585.

[22] Castillio, J. del, and Katz, B. (1957). Interaction at end-plate receptors between different choline derivatives. *Proc. R. Soc. Lond. B* **146**, 369-381.

[23] Chung, S. H. and Gage, P. W. (1998). Signal processing techniques for channel current analysis based on hidden Markov models. *Methods in Enzymology*

293, 420-437.

[24] Chung, S. H. and Kennedy, R. A. (1996). Coupled Markov chain model: characterisation of membrane channel currents with multiple conductance sublevels as partially coupled elementary pores. *Math. Biosci.* **133**, 111-137.

[25] Chung, S. H., Moore, J. B., Xia, L., Premkumar, L. S. and Gage, P. W. (1990). Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. R. Soc. Lond. B* **329**, 265-285.

[26] Chung, S. H., Krishnamurthy, V. and Moore, J. B. (1991). Adaptive processing techniques based on hidden Markov models for characterising very small channel currents buried in noise and determinstic interference. *Phil. Trans. R. Soc. Lond. B* **334**, 357-384.

[27] Colquhoun, D. (1981). How fast do drugs work? *Trends. Pharmacol. Sci.* **2**, 212-217.

[28] Colquhoun, D., and Hawkes, A. G. (1977). Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Proc. R. Soc. Lond. B* **199**, 231-262.

[29] Colquhoun, D., and Hawkes, A. G. (1982). On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B* **300**, 1-59.

[30] Colquhoun, D., and Hawkes, A. G. (1987). A note on correlation in single ion channel records. *Proc. R. Soc. Lond. B* **230**, 15-52.

[31] Colquhoun, D., and Hawkes, A. G. (1995a). The principles of the stochastic interpretation of ion-channel mechanisms. In *Single-Channel Recording* (2nd edn), Chapter 18 (B. Sakmann and E. Neher, eds.), pp. 397-482. New York: Plenum Press.

[32] Colquhoun, D., and Hawkes, A. G. (1995b). A Q-matrix cookbook. In *Single-Channel Recording* (2nd ed.), Chapter 20 (B. Sakmann and E. Neher, eds.), pp. 589-633. New York: Plenum Press.

[33] Colquhoun, D., Hawkes, A. G., Merlushkin, A. and Edmonds, B. (1997). Properties of single ion channel currents elicited by a pulse of agonist concentration or voltage. *Phil. Trans. R. Soc. Lond. A* **355**, 1743-1786.

[34] Colquhoun, D., Hawkes, A. G., Srodsinski, K. (1996). Joint distributions of apparent open times and shut times of single ion channels and the maximum likelihood fitting of mechanisms. *Phil. Trans. R. Soc. Lond. A* **354**, 2555-2590.

[35] Colqhoun, D. and Sigworth, F. J. (1995). Fitting and statistical analysis of single-channel records. In *Single-Channel Recording* (Second Edition) (B. Sakmann and E. Neher, eds.), pp. 483-587. New York: Plenum.

[36] De Gunst, M. C. M., Huensch, H. R. and Schouten, J. G. (2001). Statistical analysis of ion channel data using hidden Markov models with correlated state-dependent noise and filtering. *J. Amer. Statist. Assoc.* **96**, 805-815.

[37] Edeson, R. O., Ball, F. G., Yeo, G. F., Milne, R. K. and Davies, S. S. (1994). Model properties underlying non-identifiability in single channel inference. *Proc. R. Soc. Lond. B* **255**, 21-29.

[38] Fredkin, D. R., Montal, M. and Rice, J. A. (1985). Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proceedings of the Berkeley Conference in Honour of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Ohlsen, eds.), pp 269-289, Belmont, Wadsworth.

[39] Fredkin, D. R. and Rice, J. A. (1986). On aggregated Markov processes. *J. Appl. Prob.* **23**, 208-214.

[40] Fredkin, D. R. and Rice, J. A. (1987). Correlation functions of a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci.* **87**, 161-172.

[41] Fredkin, D. R. and Rice, J. A. (1992a). Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. R. Soc. Lond. B* **249**, 125-132.

[42] Fredkin, B. R. and Rice, J. A. (1992b). Bayesian restoration of single-channel patch clamp recordings. *Biometrics* **48**, 427-448.

[43] Fredkin, B. R. and Rice, J. A. (2001). Fast evaluation of the likelihood of an HMM: ion channel currents with filtering and coloured noise. *IEEE Trans. Signal Processing* **49**, 625-633.

[44] Hamill, O. P. and Martinac, B. (2001). Molecular basis of mechanotransduction in living cells. *Physiol Rev.* **81**, 685-740.

[45] Hamill, O. P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F. (1981). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Pflügers Archiv. Eur. J. Physiol.* **391**, 85-100.

[46] Hawkes, A. G., Jalali, A., and Colquhoun, D. (1990). The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Phil. Trans. R. Soc. Lond. A* **332**, 511-38.

[47] Hawkes, A. G., Jalali, A., and Colquhoun, D. (1992). Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond. B* **337**, 383-404.

[48] Hodgson, M. E. A. (1999). A Bayesian restoration of an ion channel signal. *J.R. Statist. Soc. B* **61**, 95-114.

[49] Horn, R. and Lange, K. (1983). Estimating kinetic constants from single channel data. *Biophys. J.* **43**, 207-233.

[50] Jalali, A., and Hawkes, A. G. (1992a). The distribution of apparent occupancy times in a two-state Markov process in which brief events cannot be detected, *Adv. Appl. Prob.* **24**, 288-301.

[51] Jalali, A., and Hawkes, A. G. (1992b). Generalised eigenproblems arising in aggregated Markov processes allowing for time interval omission. *Adv. Appl. Prob.* **24**, 302-21.

[52] Katz, B. and Miledi, R.(1970). Membrane noise produced by acetylcholine, *Nature* **226**, 962-963.

[53] Katz, B. and Miledi, R. (1972). The statistical nature of the acetylcholine potential and its molecular components, *J. Physiol.* **224**, 665-699.

[54] Khan, R. N. (2002). *Statistical modelling an analysis of ion channel data based on hidden Markov models and the EM algorithm.* Ph.D. Thesis, University of Western Australia.

[55] Kienker, P. (1989). Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B* **236**, 269-309.

[56] Klein, S., Timmer, J. and Honerkamp, J. (1997). Analysis of multichannel patch clamp recordings by hidden Markov models. *Biometrics* **53**, 870-883.

[57] Li, Y., Yeo, G. F., Milne, R. K., Madsen, B. W. and Edeson, R. O. (2000). Burst properties of a supergated double-barrelled chloride ion channel. *Math. Biosci.* **166**, 23-44.

[58] McManus, O. B., Blatz, A. L. and Magleby, K. L. (1985). Inverse relationship of the duration of adjacent open and shut intervals for Cl and K channels. *Nature* **317**, 625-628.

[59] Magleby, K. L. and Weiss, D. S. (1990). Identifying kinetic gating mechanisms for ion channels by using two-dimensional distributions of simulated dwell times. *Proc. R. Soc. Lond. B* **241**, 220-228.

[60] Magleby, K. L. and Song, L. (1992). Dependency plots suggest the kinetic structure of ion channels. *Proc. R. Soc. Lond. B* **249**, 133-142.

[61] Merlushkin, A. I. (1996). *Some Problems Arising in Stochastic Modelling of Ion Channels due to Time Interval Omission.* Ph.D. thesis, University of Wales.

[62] Merlushkin, A. I. and Hawkes, A. G. (1997). Stochastic behaviour of ion channels in varying conditions. *IMA J. Math. Appl. in Med. and Biol.* **14**, 125-149.

[63] Michalek, S., Lerche, H., Wagner, M., Mitrovic, N., Schiebe, M., Lehmann-Horn, F. and Timmer, J. (1999). On identification of Na+ channel gating schemes using moving average filtered hidden Markov models. *Eur. Biophys J.* **28**, 605-609.

[64] Neher, E. and Sakmann, B. (1975). Voltage-dependence of drug-induced conductance in frog neuromuscular junction. *Proc. Nat. Acad. Sci. U.S.A.* **72**, 2140-2144.

[65] Neher, E., and Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260**, 799-802.

[66] Qin, F., Auerbach, A. and Sachs, F. (2000a). A direct optimisation approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* **79**, 1915-1927.

[67] Qin, F., Auerbach, A. and Sachs, F. (2000b). Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.* **79**, 1928-1944.

[68] Rothberg, B. S. and Magleby, K. L. (1999). Gating kinetics of single large-conductance Ca2+-activated K+ channels in high Ca2+ suggest a two-tiered allosteric gating mechanism. *J. Gen. Physiol.* **114**, 93-124.

[69] Roux, B., and Sauve, R. (1985). A general solution to the time interval omission problem applied to single channel analysis. *Biophys. J.* **48**, 149-58.

[70] Sakmann, B. and Neher, E. (editors) (1995). *Single Channel Recording.* Plenum Press, New York, Second edition.

[71] Sheridan, R. E. and Lester, H. A. (1975). Relaxation measurements on the acetylcholine receptor. *Proc. Natn. Acad. Sci. U.S.A.* **72**, 3496-3500.

[72] Stevens, C. F. (1972). Inferences about membrane properties from electrical noise measurements. *Biophys. J.* **12**, 1028-1047.

[73] Sukharev, S. I., Sigurdson, W. J., Kung, C. AND Sachs, F. (1999). Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. *J. Gen. Physiol.* **113**, 525-539.

[74] Venkataramanan, L. and Sigworth, F. J. (2002). Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.* **82**, 1930-1942.

[75] Yeo, G. F., Milne, R. K., Edeson, R. O. and Madsen, B. W. (1988). Statistical inference from single channel records: two-state Markov model with limited time resolution. *Proc. R. Soc. Lond. B* **235**, 63-94.

Chapter 6

The Biophysical Basis of Firing Variability in Cortical Neurons

Hugh P.C. Robinson

Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, U.K.

CONTENTS

6.1 Introduction

The trajectory of the membrane potential in a cortical neuron is very noisy. The sequence of events which generate this variability is summarized in Figure 6.1. Unlike the input of a sensory receptor, that of a cortical cell is complex and not completely controllable, or even knowable, in experiments. The process of generating the membrane potential signal begins with a set of effectively stochastic presynaptic action potential trains converging on some 10^4 synaptic terminals, distributed over the dendritic tree of the neuron. Each of these spike trains drives a highly unreliable and stochastic process of transmitter release at each terminal. The released transmitter then opens ion channels whose opening and closing behaves exactly like a stochastic (Markov) process, as described in detail in the chapter by Alan Hawkes. As the membrane potential then changes, other Markovian ion channels whose transition

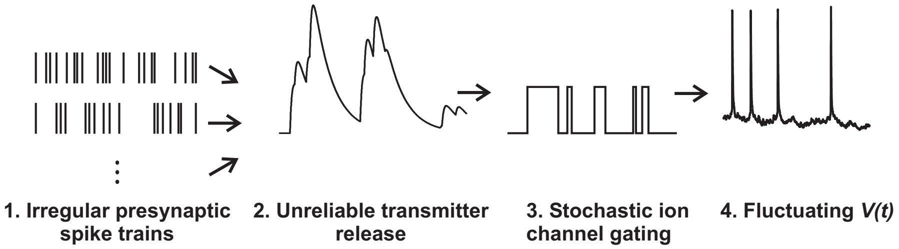
1. Irregular presynaptic spike trains
2. Unreliable transmitter release
3. Stochastic ion channel gating
4. Fluctuating *V(t)*

Figure 6.1

A summary of the noise sources contributing to the fluctuating membrane potential of cortical neurons.

rates are highly voltage-dependent, open and close, generating postsynaptic action potentials. The dynamics of synaptic integration are thus nonlinear, because of this voltage-dependence, and are permeated with noise at each stage. In this chapter, I will focus on the steps in the production of the noisy membrane potential which occur at the level of the single neuron, i.e., steps 2 through 4 in Figure 6.1. For a review of ideas about step 1, see the chapter by Salinas and Sejnowski in this volume.

The biophysical mechanisms involved are central to understanding the reliability of synaptic integration, and hence the strategies used to transmit and transform neural information. What is encoded by the times at which spikes occur? The precision or reliability of responses of individual cells is responsible for the degree of *synchrony* in a connected population of neurons. How precise and how stable can coherent firing amongst cells be? Does dynamical behaviour resulting from the interaction of noise and nonlinearity, such as stochastic or coherence resonance [21], play a role in cortical information processing? Being able to answer such questions will depend on an understanding of the biophysics of firing variability.

6.2 Typical input is correlated and irregular

Because of the difficulty of recording from large numbers of neurons simultaneously across the cortex, much of what we know about the synaptic input to cortical cells is inferred from the firing of single cells. Firing patterns in the functioning cortex are themselves highly variable. In some situations, firing resembles a Poisson process, with an exponential distribution of interspike intervals [8, 12, 31]. Burst firing is evident [39], and there is evidence for weakly periodic firing during certain states of consciousness or sensory stimulation [54, 58]. Overall firing variability is characterised by measures such as the coefficient of variation of interspike intervals, CV(ISI), the ratio between the standard deviation and the mean of interspike intervals. CV(ISI) can be high – higher than the value (1) expected for a Poisson point process, a completely random point process of uniform rate.

The input collected from individual presynaptic neurons is small, with unitary excitatory postsynaptic potentials (EPSPs) typically less than a millivolt at the soma. Assuming that such inputs converge as *independent* stochastic point processes, modelling has suggested that with the number required to fire the postsynaptic cell, the total input would actually be smooth enough that the firing of the postsynaptic cell would be far more regular than is actually observed [56]. Some increase in firing variability might be generated by balanced independent excitation and inhibition [52]. However, it appears (as one might expect from the high local connectivity in the cortex) that a major factor in high firing variability is the quite synchronous, correlated firing of local presynaptic neurons. It is not correct to assume that inputs are independent. Large transients of local population activity i.e., synchronous network firing events, are observed during sensory responses ([2, 5, 7, 64]). Furthermore, experiments in which single cells are stimulated with current [60] or more naturally, conductance stimuli [1, 27] constructed from elementary synaptic events, show that it is necessary to group these in correlated bursts to generate the required level of postsynaptic firing variability. This conclusion is supported by modelling studies, e.g., [17]. A full discussion of the structure of local population activity in the cortical network is outside the scope of this chapter. For the present purpose, it is enough to observe that cells *typically* fire in response to strong fluctuations in input, produced by correlated activity in the surrounding network. The high variability of interspike intervals results from the irregularity of the times of these synchronous firing events, and the contrast between short *intra*burst and long *inter*burst spike intervals. The onset of a strong input fluctuation essentially resets the coherence of spike timing. Later in this chapter, therefore, I will focus on the transient response to a single input fluctuation, as a distinct unit of activity.

6.3 Synaptic unreliability

Cortical excitatory synapses are highly noisy. In the experiment shown in Figure 6.2, five closely-spaced action potentials followed by a single delayed one, are delivered to a presynaptic pyramidal cell, and the EPSP responses are recorded in a postsynaptic cell. Although a depressing trend is evident in the ensemble average (bottom trace), the responses from spike to spike, and from trial to trial, are seen to be highly variable in amplitude. There are numerous failures to release transmitter in response to presynaptic action potentials, although estimates of the overall probability of release vary greatly [13, 23, 26].

The distribution of amplitudes of response at a given synapse is wide – when responses to single presynaptic APs, separated by at least several seconds, are measured, there is evidence for a quantal or multimodal distribution of amplitudes [16]. However, whatever this distribution of amplitudes is, it is not uniform in time during a sequence of synaptic responses. For example, immediately following a release,

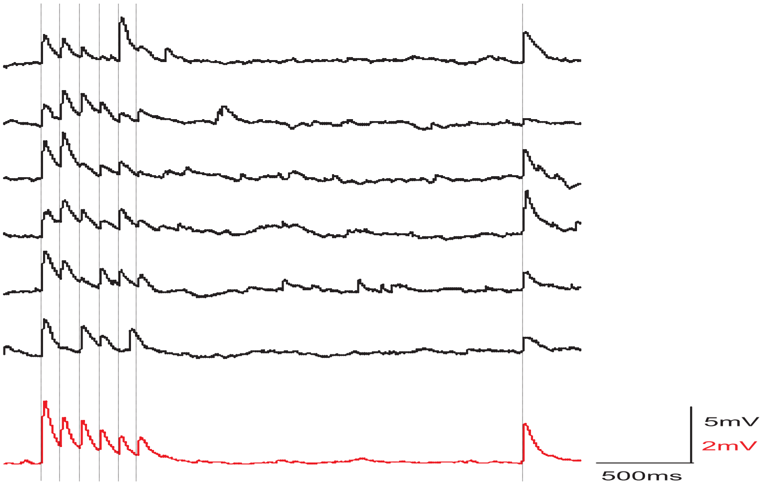

Figure 6.2

An ensemble measurement of EPSPs in a layer 2/3 pyramidal neuron. Top 6 traces: individual trials. Red trace, ensemble average. Dotted lines indicate times at which presynaptic neuron is stimulated. Recording by Ingo Kleppe, Dept. of Physiology, University of Cambridge. (See color insert.)

there is an increased probability of failure, and vice versa [59]. In other words, there are significant correlations over time in the *variability* at individual synapses, as indeed there are in the mean synaptic response. These have been shown to extend over surprisingly long time scales, of several minutes or more during natural-like spike trains [35]. Further understanding of the variability or reliability of individual synapses over time during complex, natural input, is an important goal. One might think that the effect of synaptic fluctuations will be *averaged away* because there are large numbers of synapses per cell. However even if many inputs are active, only one completes the job of taking the voltage over threshold. As will be discussed further below, the nonlinear threshold behaviour of neurons means that the properties of essentially *any* input fluctuations, however small, can determine the firing pattern of the cell – it all depends on the proximity of the spike generation mechanism to its threshold.

6.4 Postsynaptic ion channel noise

The variability introduced by ion channel gating in the postsynaptic cell is rather better understood than synaptic unreliability (see chapter by Hawkes). Single-channel recording has shown in great detail how single ion channels operate as probabilistic machines, or Markov processes with a reasonably small number of discrete states, corresponding to distinct conformations of the ion channel protein [29]. Single channel properties of ion channels – their conductance and average opening duration – were first estimated, before the advent of the gigaohm-seal patch-clamp, by analysing the current or voltage noise produced by channel gating. The properties of this noise have therefore been well characterised and extensively measured [11]. The effects of ion channel gating noise in neurons have recently been reviewed in [66].

If channels in an identical population of size N are independent, and have a single conducting current level i then at steady-state, then the population current variance $\sigma^2 = iI - I^2/N$, where I is the mean population or macroscopic current. The size of fluctuations therefore scales with the square root of the size of the population current, for small open channel probability. For large numbers of channels, the noise amplitude is distributed normally. How big are the populations of channels? This question is complicated by the fact that they are distributed over large electrical distances within the cell. It is the *local* population of channels, in particular around any site of AP initiation, which is relevant. The density of sodium channels in layer 5 pyramidal neurons appears to be fairly uniform in the soma and dendrites, at around several channels per square micrometre, but is probably much higher in the proximal axon, where it is thought that many action potentials initiate [61]. Both calcium and sodium action potentials can, however, be initiated in the remote dendrites. There are several types of potassium channels, the density of which tapers off with distance from the soma in pyramidal neurons [36], but is probably in the region of 0.1 to 1 channels per square micrometer. Less is known about the density of functional calcium channels in cortical pyramidal neurons, but in hippocampal pyramidal neurons densities of 1 to 10 channels per square micrometer of high and low voltage activated calcium channels [40]. The noise from persistent Na channels, because of their maintained activation around threshold, appears to be particularly important in controlling firing patterns in entorhinal cortical cells [65]. However, it is important to realise that the population of open channels is in general much smaller than the total. Channel noise becomes most powerful in its effects when only a very small proportion of channels are open [51]. This is because the size of fluctuations relative to the mean conductance is highest under these conditions.

Another important quantity in determining the variance of channel noise is the size of the single channel conductance. Amongst voltage-gated channels, this varies from a few pS for a low voltage-activated calcium channel to about 200 pS for a maxi calcium-activated potassium channel. At the excitatory synapse, the AMPA receptor has a single channel conductance of around 10 pS [63], while the NMDA receptor has a conductance of 35 to 50 pS [47].

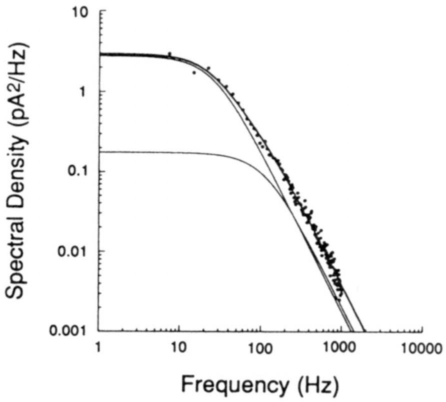

Figure 6.3
Spectrum of NMDA receptor current noise at a single synapse. The spectrum follows a sum of two Lorentzian components, with corner frequencies of 25 Hz and 110 Hz. From [47] with permission.

Equally important, though, is the *timing*, or frequency distribution of the fluctuations. The membrane time "constant" (which is actually dynamically-varying) determines the low-pass filtering of the input signal. Low frequencies in the noise e.g. below 100 Hz are much more effective in distorting the membrane potential than higher frequency fluctuations. Under stationary conditions, channel noise is essentially a linear stochastic process: its autocorrelation function (or equivalently power spectrum) contains all the information available to predict its time course. The autocorrelation function is a sum of exponential components whose rates are given by the eigenvalues of the kinetic matrix (see chapter by Hawkes). Correspondingly, the power spectrum is a sum of Lorentzian components. Almost always, though, for actual channels, the power of one or two of these components is dominant. For example, Figure 6.3 shows the noise through NMDA receptor channels at a single synapse [47]. The power of the lower frequency component is more than ten times that of the higher frequency component.

Roughly speaking, this first component corresponds to the correlation due to the mean open lifetime of channels or predominant burst lifetime. For purposes of modelling the function of this noise, therefore, it is often enough to consider a single exponentially-correlated process. A continuous stochastic process which has an exponential autocorrelation, with amplitude σ and correlation time constant τ, is the Ornstein-Uhlenbeck (OU) process [18, 24], $\xi(t)$, which can be generated by numerically solving:

$$\tau \frac{d\xi}{dt} = \frac{-\xi}{\tau} + c^{1/2} g_w(t) \tag{6.1}$$

where $g_w(t)$ is Gaussian white noise, and the standard deviation of ξ is $\sqrt{c\tau/2}$, and c is a constant.

This generates stationary noise. However, real channel noise is usually not stationary: for example, transmitter-activated channels have transition rates which depend on the changing transmitter concentration, and voltage-activated channels have voltage-dependent transition rates. To account for such nonstationarity with high accuracy, it is necessary to use stochastic simulations of populations of channels, modelling the state transitions of all channels in each population [9, 55]. One may also use an OU process which is nonstationary in σ and τ to approximate the stochastic Hodgkin-Huxley system [19]. However, since as described later, a great deal of spike-time variance is generated by noise in a limited band of membrane potential – around threshold – then a stationary OU noise source or sources may be a reasonably accurate yet simple model for predicting firing variability. An OU process is also a good model for synaptic noise, composed of large numbers of small, identical EPSPs or IPSPs which have a fast onset and decay exponentially and whose arrival times are a Poisson process [1].

6.5 Integration of a transient input by cortical neurons

Cortical neurons fire in response to fast-fluctuating stimuli with much more precision (i.e., with reproducibly-timed action potentials in an ensemble of identical trials), than in response to constant stimuli [41]. In cortical neurons, precision has been found to improve as the frequency of sinusoidal stimulation is increased up to about 25 Hz [44]. In *Aplysia* neurons, it has been demonstrated that the precision depends on the presence of frequencies close to the preferred firing frequency of the cell at the mean level of current [32], an effect which is probably general to many neurons. It is clear that in an ensemble of responses to a complex fluctuating stimulus, a *strong* fluctuation forces coherence across trials by compelling the cell to spike, putting it in a particular dynamical state within a tightly-delimited interval of time. Most of the influence of the preceding history is lost. This is because voltage-gated channels are forced into an activated dynamical state, and the temporarily high conductance of the neuronal membrane allows a high rate of dissipation, or leak of charge stored on the membrane capacitance. In this section, I will discuss the variability of response of a cortical neuron during a single large input fluctuation, from the moment of complete coherence at the beginning of the fluctuation, until the input has decayed sufficiently that the cell is silent. This discussion is based on [48].

Figure 6.4 shows the response of a pyramidal cell to a burst of excitatory synaptic conductance, an exponentially-decaying transient in the rate of arrival of excitatory synaptic conductances. The conductance stimulus is delivered using the technique of conductance injection, or dynamic clamp [49, 53]. This technique is ideal for inves-

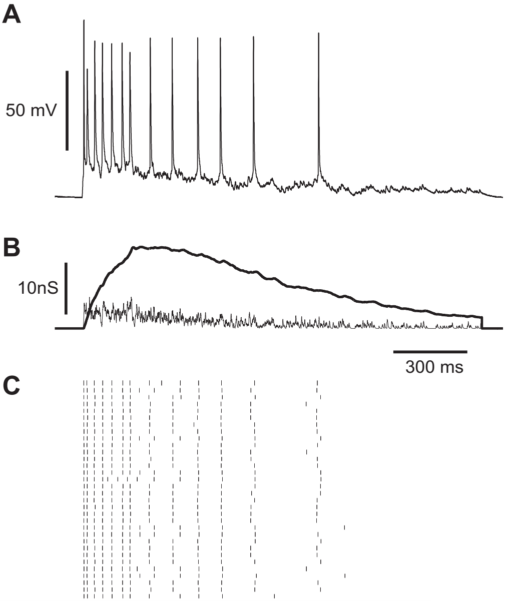

Figure 6.4

Spike time variability in responses of a layer V pyramidal neuron to a natural-like burst conductance input. (A) An example membrane potential response to a burst conductance input at the soma. (B) Conductance input consisting of a train of unitary AMPA+NMDA conductance transients, generated by a nonstationary Poisson process with an exponentially declining rate. Initial peak rate was 2500 Hz, time constant was 500 ms. Total number of unitary input events was 1222. AMPA (thin trace) and NMDA (thick trace) conductance components are shown separately. The NMDA input is subject to a further voltage-dependent block (see Figure 6.8). (C) Raster display of 32 trials with the same stimulus. From [48] with permission.

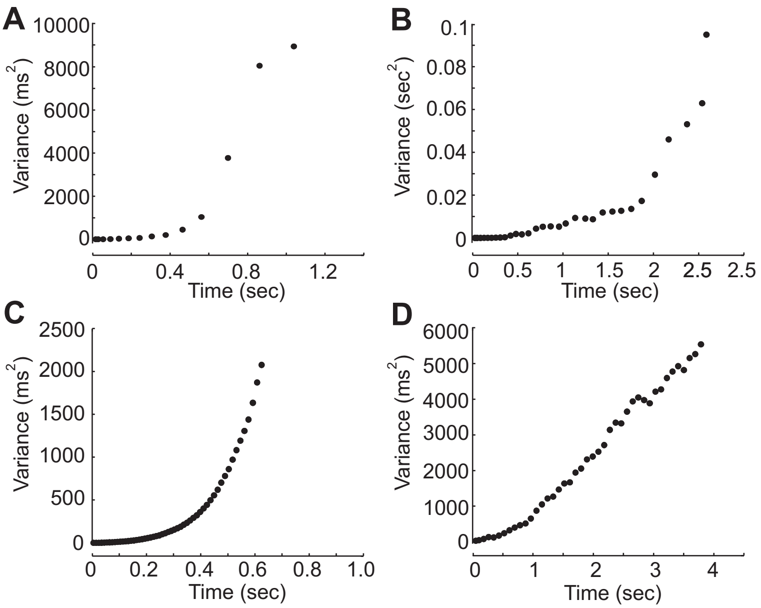

Figure 6.5

Spike time variance plots for 2 types of cortical neuron. (A) a regular-spiking (excitatory, pyramidal) neuron, stimulated by a ramp current decaying from 500 pA to 0 in 3 s. (B) another regular-spiking neuron, stimulated by a ramp current decaying from 500 pA to 0 in 4.5 s. (C) a fast-spiking (inhibitory, basket) neuron, stimulated by a ramp current decaying from 400 pA to 0 in 4 s. (D) a regular-spiking neuron stimulated by a steady current of 100 pA. From [48] with permission.

tigating the variability of neuronal integration, because unlike natural synaptic input, it can be delivered precisely and repeatedly to the neuron, yet is electrically realistic. The excitatory input shown here has two fractions, due to AMPA and NMDA glutamate receptors, each of which are activated during every unitary synaptic event. Because of the intrinsic differences in the receptor kinetics, NMDA receptor conductance outlasts the much more rapid AMPA conductance. However, the amount of NMDA conductance injected is voltage-dependent, and is sharply reduced by depolarization (see Section 6.7). In the experiment in Figure 6.4, the jitter of spike times is at first low, but towards the end of the response rises very rapidly, so that the final spike is scattered over some 100 ms. What is going on here? This will be the focus of the rest of the chapter.

Within a burst response like this, progressive accumulation of spike time variance is opposed by the regrouping effect of the fluctuations in the stimulus. While there is some correlation of each spike time with the previous spike time in the same response, there is also some tendency for spikes to be driven by particular fluctuations.

To remove the latter effect, we can instead deliver smoothly decaying ramps of current. The variance of the time of occurrence of spike i is determined as:

$$v_i = \frac{1}{N_i} \sum_{j, n_j \geq i} (t_{ij} - \overline{t_i})^2 \tag{6.2}$$

where t_{ij} is the time of spike i in response j, N_i is the number of responses with i or more spikes, n_j is the total number of spikes in response j, and $\overline{t_i}$ is the mean time of spike i. The slope of this quantity with time is the rate of generation of spike time variance.

Figure 6.5A-C shows, in several different cells, that the accumulating variance goes through two phases as the stimulus decays, a low-variance stage, and a high-variance stage. The variance in the late stage begins to approach that of a Poisson process - that is the increment in variance per spike approaches the square of the mean spike interval. In contrast, with a steady plateau stimulus, there is, apart from slight initial adaptation, a steady rate of variance generation.

6.6 Noisy spike generation dynamics

What is the biophysical basis of these two stages of variance generation? Although we know that there are many different voltage-gated channels which activate and deactivate over different timescales in cortical neurons [42], many of these are at much slower time scales than the spike. Initiation of axonally-propagated spikes is dominated by populations of fast Na channels and fast K channels, in a restricted site in the neuron, around the soma and initial segment of the axon [57]. The qualitative features of this two-stage behaviour are displayed by a model as simple as an isopotential patch of membrane with stochastically-simulated Na and K channels whose states and transition rates are derived from the deterministic Hodgkin-Huxley model (for details of the model, see [9]). In this model, voltage-dependent probabilistic transitions between channel states are simulated explicitly, and the level of voltage noise increases as the area of membrane (and therefore number of channels) is reduced. At four different membrane areas, the two stages of rising spike time variance are clearly seen in response to the same decaying ramp of current density (Figure 6.6).

An essential difference between the two stages of rising variance is now seen: the gradient of the early stage is highly sensitive to the noise level, increasing in inverse proportion to membrane area, while that of the late, high variance stage is almost independent of membrane area. Mean firing frequency decays only slightly during the burst. The point of transition between low and high variance stages is also sensitive to the noise level. Perhaps surprisingly, the higher the membrane area (i.e. the *lower* the noise level), the *earlier* the transition. This effect leads to a crossover of the relationships at about 270 ms (indicated by an arrow).

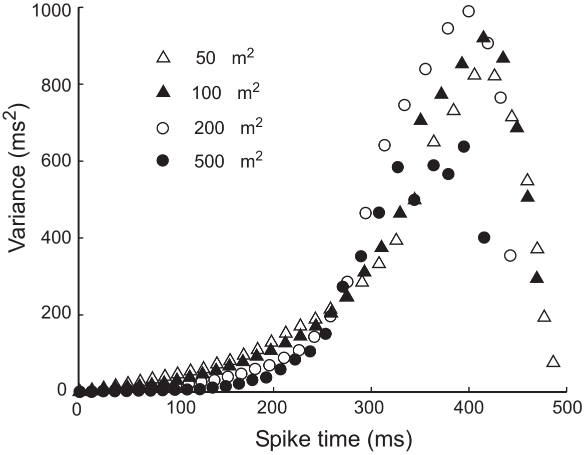

Figure 6.6

Spike time variance plots for stochastically-simulated Hodgkin-Huxley membrane. Areas of membrane are indicated by symbol. The stimulus current density decayed linearly from 50 mA/cm^2 to 0 over 0-500 ms. Channel density 60 Na channels/μm^2, 18 K channels/μm^2, capacitance 1 μF/cm^2. The point of crossover is indicated by an arrow (see text). Modified from [48].

To gain further insight, it is useful to reduce the Hodgkin-Huxley equations from a 4-variable model (m, h, n, and V) to a 2-variable model, and to specify separate terms for the noise and the dynamics, i.e. to use a Langevin equation. Well-known 2-variable spiking models are the FitzHugh-Nagumo (FHN) model and the Morrris-Lecar models (ML1 and ML2). These both have a fast variable (V) reflecting both voltage and sodium channel activation, which vary together, and a slow variable (W) reflecting inactivation and potassium channel activation [22]. O-U noise $\xi(t)$ is added to the derivative of V, which is also a function of the stimulus current I:

$$\dot{V} = F(V,W,I) + \xi(t) \tag{6.3}$$

$$\dot{W} = G(V,W) \tag{6.4}$$

Studying the trajectory of these variables under the influence of noise at different levels of background current reveals the different nature of the two stages of variance (Figure 6.7). At *high stimulus values* (at mean stimulus values well above the threshold or bifurcation point for repetitive firing), i.e. early in the transient response, the vector field defined by the nonlinear function F and the linear or nonlinear function G is such that motion is consistently fast, following a limit cycle, whose *position* in phase space, rather than essential shape, is perturbed by the noise. At low mean stimulus values, well below threshold, spiking still occurs, as a result of the noise, but the quality of the motion is quite different – there is a slow, thoroughly random

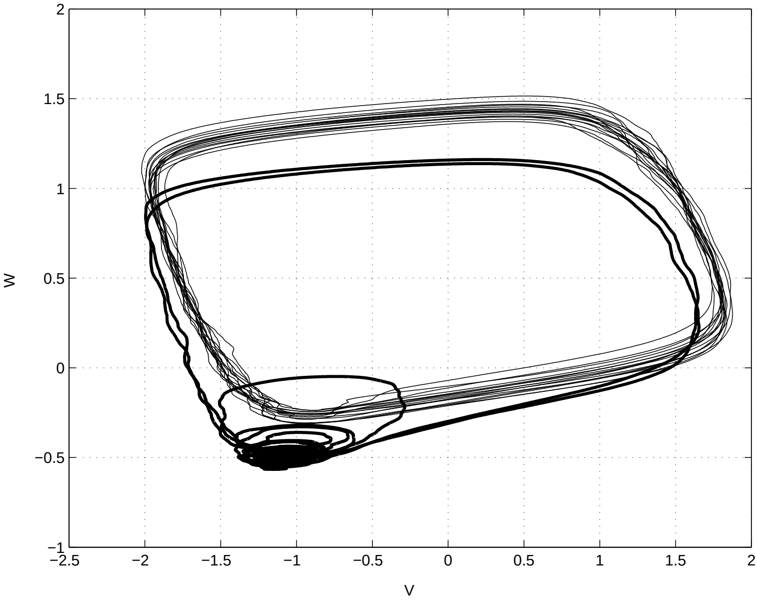

Figure 6.7

Phase trajectory of the variables in the FitzHugh-Nagumo model at two different (steady) stimulus levels: 0.5 (thin, black) and 0.15 (thick, gray) with Ornstein-Uhlenbeck noise, σ=0.075, τ=1. See [48] for details of the FHN model.

walk around the stable fixed point (corresponding to the resting potential) punctuated by occasional excursions around the excited spike trajectory. The generation of spikes under this condition can be described by Kramer's formula for thermal motion escaping from an energy well [9, 38] – because there is essentially a constant escape probability at each instant of time, then spiking occurs as a Poisson process.

Thus, in the response of a cortical neuron to a single decaying transient there is, at a certain point, a switch from low variability to high variability, from fast motion to slow motion, from a uniformly perturbed limit cycle to occasional escapes from an attractive basin. The trajectory of slow motion is much more sensitive to noise than is the fast motion. This difference also explains the phenomenon of coherence resonance, or autonomous stochastic resonance, in which an excitable set of equations such as the FHN model goes through a maximum in minimum CV(ISI) at a certain amplitude of driving noise [45, 46].

6.7 Dynamics of NMDA receptors

Another important factor in firing variability is the nature and timing of the unitary synaptic conductance, in particular of its NMDA receptor-mediated component. As mentioned above, excitatory synapses between cortical cells have a large NMDA receptor-mediated fraction of conductance [28, 34], which has two distinctive features. It is much slower in its time course than the AMPA-receptor-mediated phase (Figure 6.8B), and the conductance is highly voltage-dependent, as a result of block by extracellular magnesium ions [3]. At hyperpolarized potentials, the channel is blocked but it unblocks rapidly as the membrane potential rises. The NMDA receptor conductance is believed to cause a large amount of the variability of spikes within synaptically-driven bursts, because stimulating cortical neurons with conductance injection indicates that activation of the NMDA receptor conductance at the synapse can roughly double the jitter of spikes [27].

Figure 6.8A analyses the effect of NMDA receptor input using a Morris-Lecar class 1 model (see [48] for details), perturbed by stationary OU noise. The relationship between CV(ISI) and firing frequency is plotted when the model is stimulated by a constant level of AMPA or NMDA activation. In this case, the only difference between the two mechanisms lies in the voltage-dependent nonlinearity of the NMDA conductance. This is seen to reduce variability above a firing frequency of about 10 Hz. This seems to happen because the nonlinearity lowers the threshold for fast motion.

What about the effect of the slow kinetics of the NMDA phase in natural synaptic input? Figure 6.8C shows that when stimulating the Morris-Lecar model with a Poisson train of unitary excitatory postsynaptic conductance transients, increasing the NMDA content of unitaries has a very powerful effect on variability, increasing it greatly. This is essentially because the input varies more slowly as its NMDA

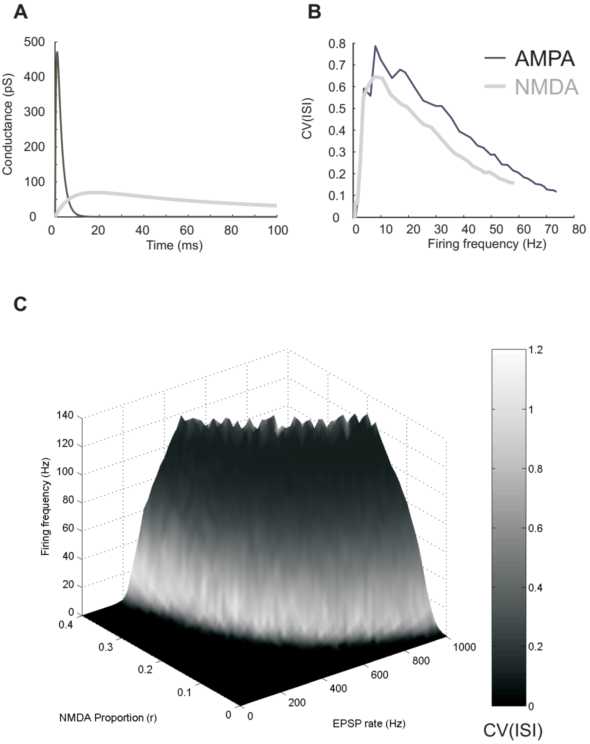

Figure 6.8

Effects of NMDA receptors on firing variability in the Morris-Lecar class 1 (ML1) model. (A) Relationship between CV(ISI) and firing frequency with variation in the level of NMDA or AMPA receptor activation. (B) Dependence of CV(ISI) (indicated by gray level) and firing frequency on EPSP rate and the proportion of NMDA receptor input. Each unitary conductance had two fractions: $g_{AMPA}(t) = 0.1\left(e^{-t/2} - e^{-t/0.5}\right)$, $g_{NMDA}(t) = r[0.062e^{-t/46} + 0.038e^{-t/235} - 0.1e^{-t/7})]/[1 + 0.6e^{-0.06V}]$, where r is the ratio of maximal NMDA conductance to maximal AMPA conductance. Current density is in units of μA/cm^2. V in mV. Conductance in μmS/cm^2. No OU current noise is added.

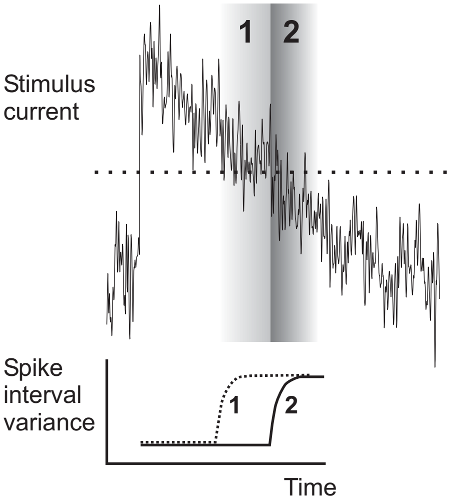

Figure 6.9

Type 1 and type 2 dynamics shift to high variance at different points in a noisy transient stimulus (top). Type 1 neuron switches to high variance when mean stimulus current level is above the threshold level (θ), while type 2 neuron switches when mean stimulus is below threshold.

content is raised, so that more time is spent poised in the late high-variability stage of firing. This figure also illustrates the high variability associated with slow motion just above threshold, discussed above.

NMDA channels are highly regulated by their subunit expression, by phosphorylation, by extracellular glycine levels, and by factors such as intracellular polyamines and pH [50]. Thus it is likely that the precision and reliability of spiking is constantly being tuned by modifying the balance between AMPA and NMDA receptor activation at excitatory synapses.

6.8 Class 1 and class 2 neurons show different noise sensitivities

The manner in which the switch between early and late stage variability happens, depends on the threshold behaviour of the neuron. There are two major classes of simple threshold behaviour in neurons [30]. Class 2 neurons show an abrupt, hard onset of repetitive firing at a high firing frequency – they cannot support regular low-frequency firing. Class 1 neurons show a continuous transition from zero frequency to arbitrarily low frequencies of firing. Squid giant axons, and both the classical Hodgkin-Huxley model and the FitzHugh-Nagumo model, are class 2, crab nerve fibres and gastropod neurons are class 1. The Morris-Lecar model can show both types of behaviour, depending on its parameters. Class 1 behaviour appears with a strong enough nonlinearity in the steady state dependence of W on V. With enough positive curvature, this relationship can result in a saddle-node bifurcation at threshold (the disappearance of two fixed points of the dynamics as the current is increased), creating a situation where the limit cycle just above threshold passes through the extremely slow region of the two-dimensional phase plane where the stable and saddle fixed points have just disappeared [15].

The threshold in a class 2 neuron occurs in a different manner. Here there is only one fixed point, which switches from stable to unstable at (or above) threshold. Oscillations which decay around the fixed point below threshold, blow up above threshold until the trajectory finds a surrounding limit cycle. In a FHN cell, this corresponds to a subcritical Hopf bufurcation. The phase point then migrates outwards to find an already existing limit cycle – so that there is an abrupt onset of a high firing oscillation.

Here, we are concerned with when and how the neuron switches to the late stage of high variability in a transient response. The two classes of neuron behave very differently in this respect (Figure 6.9). Class 1 dynamics begins to enter the late high variability stage when the fluctuations in the stimulus begin to touch the threshold, i.e., on the lower envelope of the noise – trapping in *slow motion* begins at this point. Class 2 dynamics, on the other hand, means that the subthreshold oscillations just below threshold are *fast motion* of roughly the same period as the spiking limit cycle. The late stage does not begin until fluctuations are only occasionally reaching threshold from below, i.e., on the upper envelope of the noise. Thus the two classes show a quite different and, in a sense, opposite sensitivity to the amplitude of noise. Increasing the amplitude of noise accelerates the onset of the late high-variability stage in class 1 neurons, because noise fluctuations touch threshold sooner, while it retards the onset of late stage in class 2, because with larger fluctuations it takes longer for the mean current to decay to the point where fluctuations only occasionally touch threshold from below.

6.9 Cortical cell dynamical classes

How is this discussion of dynamical classes relevant to cortical neurons? Two of the principal cell types of the cortex, regular-spiking, RS (pyramidal, excitatory) cells, and fast-spiking, FS (inhibitory, basket) cells, appear to be of class 1 and class 2, re-

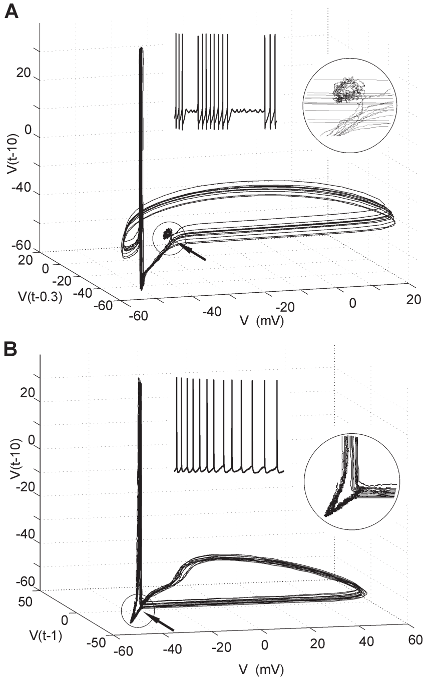

Figure 6.10

Time delay representation of phase space in responses of cortical neuron near threshold. (A) a fast-spiking neuron stimulated by a constant current of 150 pA. (B) a regular-spiking neuron stimulated by a constant current of 300 pA. From [48] with permission.

spectively. It is well known that regular-spiking neurons support extremely low frequency regular firing [10], which is class 1 behaviour. Erisir et al. [14] describe how FS neurons begin repetitive firing "with abrupt onset" for increasing levels of steady stimulus current. We have also observed that fast-spiking neurons have an abrupt onset of regular firing, and when held near the threshold switch between episodes of quite high frequency firing and subthreshold oscillations at a similar frequency (inset of Figure 6.10 [48]). This is typical class 2 behaviour.

To better illustrate the difference between dynamical classes, we examined delay representations [33, 43] of near-threshold responses to steady current injection in RS and FS cells. There is no added noise, only intrinsic noise, essentially ion channel gating noise. Three-dimensional representations are shown in Figure 6.10. Trajectories in this space did not reproducibly self-intersect, suggesting that even with this small number of time delay dimensions, there is a 1:1 transformation of the motion of the principal underlying dynamical variables [62]. Using lags of 1 and 10 ms to unfold movement at fast and slow timescales, FS neurons (Figure 6.10A) showed two patterns of perturbation – uniform perturbation of the spike loop (horizontal limb) and noisy resonant loops (inset) in a basin from which there are intermittent escapes to spike. In RS neurons (Figure 6.10B), uniform perturbation of the spiking loop is also seen, but as for the class 1 model, subthreshold movement lacks fast oscillation structure, and is a very slow drift of random movement (inset). Thus, variability of firing in two major types of cortical neurons, RS and FS, also appear to follow the qualitative patterns shown by class 1 and class 2 models, respectively.

6.10 Implications for synchronous firing

The explosion of variability in the late stage must be partly responsible for breaking up transient synchrony of firing in the local cortical network. As soon as adaptation, inactivation and synaptic depression bring the level of firing down to a critical point, the entry of many cells into the late stage would destroy the *coherence* of firing which is itself essential for maintaining the high level of input to each cell during the transient. Pyramidal RS cells appear to associate inputs from different layers and areas in the cortex, via the back-propagation activated dendritic calcium spike mechanism [37]. Class I dynamics of RS neurons, with its early entry into the late stage, might allow relatively easy *switching* between different tempos in their inputs. It may also promote the generally high level of firing variability in the cortex [25] – probably over 50% of cortical cells are class 1 RS neurons.

On the other hand, class 2 FS neurons, which inhibit each other and other RS neurons locally, are implicated in promoting synchronous firing [6, 20]. They are coupled together by electrical synapses or gap junctions, which helps to synchronise their action potentials precisely. The nature of class 2 dynamics may mean that the phase of rhythmic firing is quite stable even when the mean stimulus goes subthresh-

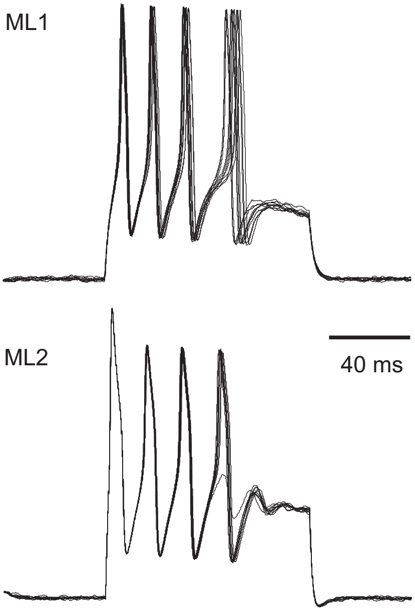

Figure 6.11

Relative precision of spikes in a transient response for a class 2 neuron compared to a class 1 neuron. Morris-Lecar models as in [25]. Noise σ = 0.3, τ=1. ML2 stimulus: ramp from 30 to 22 μA/cm^2 in 100 ms, ML1 stimulus: ramp from 11.5 to 7 μA/cm^2 in 100 ms.

old, because strong subthreshold oscillations could keep the rhythm intact until the stimulus moves above threshold again. For a class 2 neuron, there is more of a tendency for spikes to drop out without a wide dispersion of spike times: this is because the late stage is more restricted and has a harder onset than for class 1 neurons. This is illustrated in Figure 6.11, where we show that changing the class of the Morris-Lecar model from 1 to 2 can greatly reduce the dispersion of the final spikes – class 2 neurons intrinsically prefer to stay coherent or be silent, while class 1 neurons have a smooth transition between the two extremes.

6.11 Conclusions

It is clear that, to understand firing variability, it is very important to consider both the dynamics of spike generation, and the nature and parameters of the noise sources in cortical neurons. Using more sophisticated and more quantitatively exact dynamical models, for example taking account of higher-order patterns spiking such as bursting and slower processes such as dendritic calcium spikes, will uncover a much greater range of excitable behaviour shaped by noise in the cortex. Which phenomena are functionally important will be made clear as we find out more about the cortical network firing patterns of awake, behaving animals. An important principle which should also be considered is the energetic cost of precise firing, and how this has been adapted to [4]. There is still a great deal of theoretical and experimental work to be done in order to arrive at a satisfactory understanding of firing variability in the cortex.

References

[1] J-M. Fellous, A. Destexhe, M. Rudolph and T. J. Sejnowski (2001). Fluctuating synaptic conductances recreate *in-vivo*-like activity in neocortical neurons. *Neuroscience*, **107**:13–24.

[2] A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen (1996). Dynamics of ongoing activity: explanation of the large variability in evoked-cortical responses. *Science*, **273**:1868–1871.

[3] P. Ascher and L. Nowak (1988). The role of divalent cations in the n-methyl-d-aspartate responses of mouse central neurons in culture. *J. Physiol. (Lond.)*, **399**:247–266.

[4] D. Attwell and S. B. Laughlin (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow and Met.*, **21**:1133–1145.

[5] R. Azouz and C.M. Gray (1999). Cellular mechanisms contributing to response variability of cortical neurons *in vivo*. *J. Neurosci.*, **19** :2209–2223.

[6] M. Beierlein, J. R. Gibson, and B. W. Connors (2000). A network of electrically couple interneurons drives synchronized inhibition in neorcortex. *Nature Neurosci.*, **3**:904–910.

[7] T.H. Bullock, M.C. McClune, J.Z. Achimowicz, V.J. Iragui-Madoz, R.B. Duckrow, and S. S. Spencer (1995). Temporal fluctuations in coherence of brain waves. *Proc. Natl. Acad. Sci. USA*, **92**:11568–11572.

[8] B. D. Burns and A. C. Webb (1976). The spontaneous activity of neurones in the cat's cerebral cortex. *Proc. Royal Soc. Lond. B*, **194**:211–233.

[9] C. C. Chow and J. A. White (1996). Spontaneous action potentials due to channel fluctuations. *Biophys. J.*, **71**:3013–3021.

[10] B. W. Connors and M. J. Gutnick (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.*, **13**:99–104.

[11] F. Conti and E. Wanke (1975). Channel noise in nerve membranes and lipid bilayers. *Q. Rev. Biophys.*, **8**:451–506.

[12] A. Dean (1981). The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.*, **44**:437–440.

[13] J. Deuchars, D. C. West, and A. M. Thompson (1994). Relationships between morphology and physiology of pyramid-pyramid single axon connections in rat neocortex in-vitro. *J. Physiol.*, **478**:423–435.

[14] A. Erisir, D. Lau, B. Rudy, and C. S. Leonard (1999). Function of specific k+ channels in sustained high-frequency firing of fast-spiking neocortical interneurons. *J. Neurophysiol.*, **82**:2476–2489.

[15] B. Ermentrout (1996). Type 1 membranes, phase resetting curves, and synchrony. *Neural Computation*, **8**:979–1001.

[16] D. Feldmayer, J. Lubke, R. A. Silver, and B. Sakmann (2002). Synaptic connections between layer 4 spiny neurone-layer 2/3 pyramidal cell pairs in juvenile rat barrel cortex: physiology and anatomy of interlaminar signalling within a cortical column. *J. Physiol.*, **538**:803–822.

[17] J. Feng and D. Brown (1999). Coefficient of variation greater than 0.5. How and when? *Biol. Cybern.*, **80**:291–297.

[18] R. F. Fox (1991). Second-order algorithm for the numerical integration of colored-noise problems. *Phys. Rev. A*, **43**:2649–2654.

[19] R. F. Fox (1997). Stochastic versions of the Hodgkin-Huxley equations. *Biophys. J.*, **72**:2068–2074.

[20] M. Galarreta and S. Hestrin (2001). Spike transmission and synchrony detection in networks of gabaergic interneurons. *Science*, **292**:2295–2299.

[21] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni (1998). Stochastic resonance. *Rev. Mod. Phys.*, **70**(1):223–287.

[22] W. Gerstner and W. Kistler (2002). *Spiking Neuron Models*. Cambridge University Press.

[23] Z. Gil, B. W. Connors, and Y. Amitai (1999). Efficacy of thalamocortical and intracortical synaptic connections: Quanta, innervation, and reliability. *Neuron*, **23**:385–397.

[24] D. T. Gillespie (1996). The mathematics of Brownian motion and Johnson noise. *Am. J. Phys.*, **64**:225–240.

[25] B. S. Gutkin and G. B. Ermentrout (1998). Dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms and cortical spike train statistics. *Neural Computation*, **10**:1047–1065.

[26] N.R. Hardingham and A. U. Larkman (1998). The reliability of excitatory synaptic transmission in slices of rat visual cortex *in vitro* is temperature dependent. *J. Physiol.*, **507**:249–256.

[27] A. Harsch and H. P. C. Robinson (2000). Postsynaptic variability of firing in rat cortical neurons: the roles of input synchronization and synaptic NMDA receptor conductance. *J. Neurosci.*, **20**:6181–6192.

[28] S. Hestrin, P. Sah, and R. A. Nicoll (1990). Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. *Neuron*, **5**:247–253.

[29] B. Hille (2001). *Ion Channels of Excitable Membranes*. Sinauer, Sunderland MA.

[30] A.L. Hodgkin (1948). The local electric changes associated with repetitive action in a non-modulated axon. *J. Physiol. (London)*, **117**:500–544.

[31] G. R. Holt, W. R. Softky, C. Koch, and R. J. Douglas (1996). Comparison of discharge variability *in vitro* and *in vivo* in cat visual cortex neurons. *J. Neurophys.* **75**: 1806-1814.

[32] J. D. Hunter, J. G. Milton, P. J. Thomas, and J. D. Cowan (1998). Resonance effect for neural spike time reliability. *J. Neurophysiol.*, **80**:1427–1438.

[33] H. Kantz and T. Schreiber (1999). *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, UK, second edition.

[34] I. C. Kleppe and H. P. C. Robinson (1999). Determining the activation time course of synaptic ampa receptors from openings of colocalized nmda receptors. *Biophys. J.*, **77**:1418–1427.

[35] I. C. Kleppe and H. P. C. Robinson (2001). Nonlinear prediction of cortical synaptic responses to complex stimuli. *Biophys. J.*, **80**:452.

[36] A. Korngreen and B. Sakmann (2000). Voltage-gated k+ channels in layer 5 neocortical pyramidal neurones from young rats: subtypes and gradients. *J. Physiol (Lond)*, **525**:621–639.

[37] M. E. Larkum, J. J. Zhu, and B. Sakmann (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, **398**:338–341.

[38] H. Lecar and R. Nossal (1971). Theory of threshold fluctuations in nerves. I. relationships between electrical noise and fluctuations in axon firing. *Biophys. J.*, **11**:1048–1067.

[39] J. E. Lisman (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.*, **20**:38–43.

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

[41] Z.F. Mainen and T.J. Sejnowski (1995). Reliability of spike timing in neocortical neurons. *Science*, **268**:1503–1506.

[42] D. A. McCormick (1998). Membrane properties and neurotransmitter actions, in *The Synaptic Organization of the Brain*, pages 37–76. Oxford University Press, fourth ed. edition.

[43] A. Mees, K. Aihara, M. Adachi, K. Judd, T. Ikeguchi, and G. Matsumoto (1992). Deterministic prediction and chaos in squid giant axon response. *Physics Letters A*, **169**:41–45.

[44] L. Nowak, M.V. Sanchez-Vives, and D.A. McCormick (1997). Influence of low and high frequency inputs on spike timing in visual cortical neurons. *Cerebral Cortex*, **7**:487–501.

[45] A. Pikovsky and J. Kurths (1997). Coherence resonance in a noise-driven excitable system. *Phys. Rev. Lett.*, **78**:775–778.

[46] J. R. Pradines, G. V. Osipov, and J. J. Collins (1999). Coherence resonance in excitable and oscillatory systems: the essential role of slow and fast dynamics. *Phys. Rev. E*, **60**:6407–6410.

[47] H. P. C. Robinson, Y. Sahara, and N. Kawai (1991). Nonstationary fluctuation analysis and direct resolution of single channel currents at postsynaptic sites. *Biophys. J.*, **59**:295–304.

[48] H.P.C. Robinson and A. Harsch (2002). Stages of spike time variability during neuronal responses to transient inputs. *Phys. Rev. E*, **66**:061902.

[49] H.P.C. Robinson and N. Kawai (1993). Injection of digitally synthesized synaptic conductance transients to measure the integrative properties of neurons. *J. Neurosci. Meth.*, **49**:157–165.

[50] R. H. Scannevin and R. L Huganir (2000). Postsynaptic organization and regulation of excitatory synapses. *Nature Rev. Neuroscience*, **1**:133–141.

[51] E. Schneidman, B. Freedman, and I. Segev (1998). Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Computation*, **10**:1679–1703.

[52] M. N. Shadlen and W. T. Newsome (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *J. Neurosci.*, **18**:3870–3896.

[53] A. A. Sharp, M. B. O'Neil, L. F. Abbott, and E. Marder (1993). Dynamic clamp – computer-generated conductances in real neurons. *J. Neurophysiol.*, **69**:992–995.

[54] W. Singer and C. M. Gray (1995). Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, **18**:555–586.

[55] E. Skaugen and L. Wallöe (1979). Firing behaviour in a stochastic nerve membrane model based upon the Hodgkin-Huxley equations. *Acta Physiol. Scand.*, **107**:343–363.

[56] W. R. Softky and C. Koch (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *J. Neurosci.*, **13**:334–350.

[57] N. Spruston, G. Stuart, and M. Hausser (1999). Dendritic integration (chapter 10), in *Dendrites*, pages 231–270. Oxford University Press.

[58] M. Steriade, D. A. McCormick, and T. J. Sejnowski (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, **262**:679–685.

[59] C. F. Stevens and Y. Y. Wang (1995). Facilitation and depression at single central synapses. *Neuron*, **14**:795–802.

[60] C. F. Stevens and A. M. Zador (1998). Input synchrony and the irregular firing of cortical neurons. *Nature Neuroscience*, **1**:210–217.

[61] G. J. Stuart and B. Sakmann (1994). Active propagation of somatic action-potentials into neocortical pyramidal cell dendrites. *Nature*, **367**:69–72.

[62] F. Takens (1981). *Detecting Strange Attractors in Turbulence*, volume 898 of *Lecture Notes in Mathematics*. Springer, New York.

[63] S. F. Traynelis, R. A. Silver, and S. G. Cullcandy (1993). Estimated conductance of glutamate-receptor channels activated during epscs at the cerebellar mossy fiber-granule cell synapse. *Neuron*, **11**:279–289.

[64] M. Tsodyks, T. Kenet, A. Grinvald, and A. Arieli (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, **286**:1943–1946.

[65] J. A. White, R. Klink, A. Alonso, and A. R. Kay (1998). Noise from voltage-gated ion channels may influence neuronal dynamics in the entorhinal cortex. *J. Neurophysiol.*, **80**:262–269.

[66] J. A. White, J. T. Rubinstein, and A. R. Kay (2000). Channel noise in neurons. *Trends Neurosci.*, **23**:131–137.

Chapter 7

Generating Quantitatively Accurate, but Computationally Concise, Models of Single Neurons

Gareth Leng, Arleta Reiff-Marganiec, Mike Ludwig, and Nancy Sabatier

College of Medical and Veterinary Sciences, University of Edinburgh, EH9 8XD, U.K.

CONTENTS

7.1 Introduction

Information in the brain is carried by the temporal pattern of action potentials (spikes) generated by neurons. The patterns of spike discharge are determined by intrinsic properties of each neuron and the synaptic inputs it receives; modulation of either of these parameters changes the output of the neurons, and, through this, the behavior or physiology of the organism. Computational models of brain function have principally focussed on how patterns of connectivity contribute to information processing, but most models largely neglect the different intrinsic properties of different neuronal phenotypes.

7.1.1 The scale of the problem

Computational models of single neurons that realistically reflect intrinsic membrane properties can be extremely complex, and hence building large-scale realistic models of neuronal networks is computationally intense. A typical neuron may make 10,000 synaptic contacts with other neurons, and receive a similar number of inputs. Each neuron expresses a large number of channels that contribute to its membrane excitability – including several different classes of Ca^{2+}, K^{+} and Na^{+} channels, and each neuronal phenotype differs in its exact composition of membrane channels. Neurons also differ from each other morphologically, in the distribution of channel types in different cellular compartments, and in intracellular properties that influence channel function. A model of a single neuron incorporating all these factors will have a very large number of parameters that must be estimated with reasonable precision from experimental observations, but many of which must be guessed for particular cell types, as the detailed information is not available. Moreover, experimental observations of biophysical parameters are typically made *in vitro* in conditions that are different from *in vivo* conditions. The relative scarcity of afferent input in *in vitro* preparations must always be taken into account, but beyond this, biophysical measurements often require interventions that fundamentally disturb cell properties. For example, measurements of membrane potential often derive from patch-clamp recordings, which may involve dialysis of the neuronal cytoplasm, altering the composition of the intracellular fluid, changing ion gradients and diluting second messenger systems. Measurements of intracellular Ca^{2+} involve introducing fluorophores into the cell that effectively function as additional Ca^{2+} buffers. Thus measurements of many variables require consideration of the context in which they are measured.

How many neurons must be included in a realistic network model is far from clear. The human brain is commonly estimated to contain about 2×10^{10} neurons, but a rat gets by with perhaps 10^{7} neurons; the major source of this discrepancy is of course in the size of the neocortex. The neocortex, however, is one of the parts of the brain about which we understand least, substantially because the functions that we think it is principally involved in are, in general, not very amenable to reductionist experimental testing at the single cell level. In the rat brain, probably around 10^6 neurons are in the hypothalamus, and this region controls a wide diversity of clearly definable functions that are much more amenable to experimental investigation. Different neuronal groups in the hypothalamus control the release of different hormones from the pituitary gland – oxytocin; vasopressin; prolactin; growth hormone; the gonadotrophic hormones; adrenocorticotrophic hormone (that in turn controls steroid secretion from the adrenal glands); thyroid stimulating hormone (that controls the functions of the thyroid gland); and melanocyte-stimulating hormone. The hypothalamus also controls thirst, feeding behavior (including specific appetites such as sodium appetite), body composition, blood pressure, thermoregulation, and much instinctive or reflex behavior including male and female sexual behavior and maternal behavior. These functions involve highly specialised cells with specific properties; cells for instance that have receptors or intrinsic properties that enable them to respond to glucose concentration, or the osmotic pressure of extracellular fluid, or to detect specific blood-borne signals released from peripheral tissues, such as leptin from fat cells, angiotensin from the kidney, and ghrelin from the stomach. Many of these cells in turn signal to other neurons using distinct chemical messengers: neurotransmitters, neuromodulators and neurohormones but also other types of signalling molecule, that are transduced by specific receptors that can occur in multiple forms even for one given signalling molecule.

Estimating the number of distinguishable neuronal phenotypes in the rat hypothalamus is imprecise, but there seem likely to be up to 1,000, each of which may be represented by about 1,000 to 10,000 individual neurons. This may seem a high estimate of diversity, but let us consider. The ventro-rostral extent of the hypothalamus is bounded by the organum vasculosum of the lamina terminalis (OVLT). This region is highly specialised in lacking a blood-brain barrier; how many cell types it contains we do not know for sure, but they include a highly specialised population of osmoreceptive neurons [1]. Another area lacking a blood-brain barrier marks the dorso-rostral extent of the hypothalamus – this is the subfornical organ and it contains amongst its neurons (there seem to be at least six types) a population of specialised angiotensin- processing neurons. Between these, the preoptic region of the hypothalamus contains several identified nuclei and many different neuronal populations; one small but interesting population comprises about 700 luteinising-hormone releasing hormone (LHRH) neurons; these are remarkable cells, they are born in the nasal placode and migrate into the brain late in development, and are essential for controlling pituitary gonadotrophic secretion and thereby are essential for spermatogenesis in males and ovarian function in females [2]. Though very scattered throughout the preoptic area they nonetheless discharge bursts of electrical activity in synchrony to elicit pulsatile secretion of gonadotrophic hormones from the pituitary. Most of these project to the median eminence, the site of blood vessels that enter the pituitary, but some LHRH neurons project to the OVLT – why, we do not know. The preoptic region also includes a sexually dimorphic nucleus larger in males than in females. In the midline periventricular nucleus are neurosecretory somatostatin neurons that provide inhibitory regulation of growth hormone secretion, alongside growth-hormone releasing-hormone neurons of the arcuate nucleus. Caudal to the periventricular nucleus is the paraventricular nucleus; this contains thyrotropin-releasing hormone neurons that indirectly regulate the thyroid gland; corticotrophin-releasing hormone neurons that indirectly control the adrenal gland, magnocellular vasopressin neurons that control the kidney and magnocellular oxytocin neurons that are responsible for controlling parturition and lactation; in addition, smaller, centrally projecting oxytocin neurons regulate gastric function, and centrally projecting vasopressin neurons that regulate body temperature and blood pressure, some of which project into the spinal cord (as do some oxytocin neurons, a subpopulation that seems to be involved in penile erection). Below this, the suprachiasmatic nucleus is the body's principal circadian pacemaker; one population of neurons here makes vasoactive intestinal peptide, another makes vasopressin; these cells are governed by clock genes that confer 24-h cyclicity on their behavior. Behind the suprachiasmatic nucleus is the arcuate nucleus, that in addition to growth-hormone releasing-hormone neurons contains leptin-sensitive neuropeptide Y neurons that regulate feeding, dopamine neurons that regulate the secretion of prolactin, opioid (β-endorphin) neurons that impact on many neuronal systems through extensive central projections, and a large population of centrally projecting somatostatin neurons of unknown function. Above this, the ventromedial nucleus contains specialised glucoreceptive neurons, and alongside it the lateral hypothalamus contains orexin neurons; orexin is linked to sleep and wakefulness, and orexin knock-out results in narcolepsy.

We have not gone far in the hypothalamus yet, and we have described only some of the best-known populations, and neglected subpopulations of interneurons and many distinctive subnuclei. In addition, the individual cells vary even within a given population: these *homogeneous populations* are far from clones. Moreover, individuals in one population interact to differing extents with individuals of many other populations, and these interactions differ from cell to cell even within a population.

The populations are not fully interconnected but neither are they as separable as we would like. Take for instance the magnocellular oxytocin neurons of the hypothalamus-and we probably know as much or more about these as about any cells in the brain (see [3]). These are simple neurons in many respects; there are about 3,000 of them in the rat brain, and each has a single axon that projects to the neural lobe of the pituitary gland. Oxytocin, released from the nerve endings in the pituitary, controls milk let-down in response to suckling, and it controls the progress of parturition by its actions on the uterus. But in the rat, oxytocin also controls the excretion of sodium at the kidney. Moreover, centrally released oxytocin is involved in maternal behavior, sexual behavior, and affiliative behaviors generally, and stress responsiveness, and the magnocellular oxytocin system is involved in these behaviors through secretion of oxytocin from its dendrites rather than from classical nerve endings. Dendritic secretion unfortunately does not parallel secretion from axonal endings – the mechanisms underlying dendritic secretion differ in important ways from those that govern axonal secretion. The diversity of roles played by oxytocin shows both that the oxytocin neurons receive very functionally diverse inputs, and also that they influence many other neuronal populations, including some to which they are not synaptically connected, even indirectly. It would be dangerous to think that oxytocin neurons are exceptionally complicated, just because we know more about them than other neurons; in fact, from what we do know of other neurons, oxytocin neurons are, if anything, rather simple.

Thus models of any function or part of the brain must take due account of the diversity of neuronal phenotypes. In particular, models that seek to understand information processing must take account of the diversity of electrophysiological phenotypes exhibited within interconnected populations. The number of distinct electrophysiological phenotypes may be less than the number of chemically definable phenotypes alluded to, but the degree of difference between phenotypes can be striking. For example, magnocellular vasopressin neurons discharge spikes in a distinctive *phasic* pattern, alternating between periods of silence and periods of stable spike discharge, and these cells function as true bistable oscillators. Other cells, such as those in the suprachiasmatic nucleus, exhibit highly regular discharge activity, whereas the spontaneous activity of oxytocin neurons appears quasi-random. Other cells display intrinsically oscillatory discharge activity, or display a propensity to discharge in brief rapidly adapting bursts. Each of these radically different electrophysiological phenotypes has significant consequences for information processing within the networks of which they are a part (see [4]).

7.1.2 Strategies for developing computationally concise models

The number of parameters involved in modelling any single neuron to biophysical accuracy is large, at least 100 parameters would seem necessary; to build a realistic network model, these must be estimated for each cell type in the network. Even with the huge computational power available now, the computational task involved in systematically assessing models of such complexity is daunting, if this is to involve a rigorous assessment of the robustness of model performance for variable parameter values. The uncertainties and inaccuracies in estimating individual parameters are so large as to make the utility of the effort questionable. The purpose of any model is to understand a system by simplifying it, revealing the key, important variables. It makes little sense to try to build a model of the brain that is as complex as the brain. Clearly, we need to develop computationally simple models of neurons that preserve essential properties and discard those which have no major impact upon their information processing functions. However, it is not always clear which properties of a neuron are important and must be included in any model, and which, for the moment, can be neglected.

The conventional approach to understanding the role of intrinsic membrane properties in neurons has been to study channel properties in detail through experiments on isolated cells *in vitro*, and then to speculate about how these might contribute to spike patterning or neuronal responses *in vivo*. However, rather than look at membrane properties and speculate about how they might influence firing patterns, it is also possible to look at spontaneous firing patterns to see how they can be modelled most simply, and then look for explanations in terms of the known intrinsic properties [5]. What we describe here is an illustration of this approach. We are not seeking to build a complex model of an oxytocin cell by attempting to incorporate all known features of these cells. Instead, we seek a minimalist, computationally concise representation of its information-processing functions in a way that is both biologically referrable in its parameters and quantitatively accurate in its match to experimental data, while being also founded on explicit assumptions in a manner that gives it true predictive and explanatory power.

7.2 The hypothalamo-hypophysial system

The magnocellular neurosecretory neurons of the hypothalamus are concentrated in the supraoptic and paraventricular nuclei; axons from these cells project to the posterior pituitary gland (also known as the neural lobe, or neurohypophysis) where the hormones that they synthesize, oxytocin and vasopressin, are released into the circulation. These hormones are released in response to spikes that are generated at the cell bodies and conducted along the axons; hormone release from these neurosecretory terminals is analogous to neurotransmitter release from conventional neurons, but unlike transmitter release, hormone release occurs in such large amounts that it can be measured very easily. The soma and axons of the magnocellular neurons are readily identifiable and accessible for experimental manipulation through a wide variety of *in vivo* and *in vitro* experimental approaches. Thus, they are one of the few groups of central neurons in which changes in activity pattern can be related to the physiological stimulus and the precise neuronal response to the stimulus, the state of the organism and the hormonal secretion, respectively (see [5, 6, 7, 8, 9] for reviews).

Through its role as the antidiuretic hormone, vasopressin is primarily concerned with body fluid homeostasis. Vasopressin is released in response to increased plasma osmotic pressure, and in response to reduced plasma volume, and it acts on the kidney to promote conservation of water by concentrating the urine, and to restrict plasma volume by vasopressor actions on blood vessels. The classical roles of oxytocin are in lactation and parturition. At parturition, oxytocin stimulates uterine contractions to promote parturition. During lactation, oxytocin is released in response to suckling in a pulsatile manner, and promotes milk let-down from the mammary gland. Oxytocin is also released in response to increased osmotic pressure, hypovolemia, and gastric distension, reflecting a secondary role at the kidney to stimulate sodium excretion in response to increased sodium intake. Oxytocin and vasopressin also have intriguing behavioral actions. These are intriguing first because oxytocin and vasopressin released into the blood does not re-enter the brain, which is protected by a blood-brain barrier; so these behavioral effects are mediated by central release of vasopressin and oxytocin. The oxytocin and vasopressin cells have few axonal endings within the brain, but they can release very large amounts of these peptides from their dendrites. The behavioral actions seem remarkable apposite to the peripheral roles of the hormones. Oxytocin for instance promotes maternal behavior after parturition, seen in rats as nest building and retrieval of young. These behavioral actions are typical of the effects of central injection of peptides endogenous to the hypothalamus in being complex behaviors expressed over a prolonged period after brief central exposure to the peptide. Interestingly, these behavioral effects appear to be exerted via neuronal receptors expressed in regions where there is little or no innervation by oxytocin-containing fibres, suggesting that neurohormonal-like secretion from dendrites may be the important modulator of behavior.

7.2.1 Firing patterns of vasopressin neurons

When vasopressin cells are activated by a rise in osmotic pressure, their spike discharge activity consists of alternating periods of activity and silence lasting tens of seconds each, so called phasic firing (Figure 7.1A). Vasopressin cells fire phasically when their mean discharge rate exceeds 3 Hz; at lower firing rates no clear pattern is discernible. Phasic firing is important in vasopressin cells, but not because of the apparently obvious temporal patterning. Different vasopressin cells discharge asynchronously, so while the output of individual cells is pulsatile, the net output of the whole population is continuous. Phasic patterning optimises the efficiency of secretion from nerve terminals. In response to trains of electrical stimuli, vasopressin release from nerve endings is facilitated as the frequency of stimulation increases, but fatigues as the duration is extended. The frequency-facilitation of release is ascribed to a facilitation of Ca^{2+} entry at the nerve terminals, resulting in part from a broadening of spike duration at high frequencies of stimulation, in part from depolarisation caused by accumulation of K^{+} in the extracellular clefts, and in part from a progressively more complete invasion of the arborised terminal field of an axon during repetitive stimulation, resulting finally in a greater Ca^{2+} influx through voltage-gated channels to trigger enhanced exocytosis. The facilitation of secretion is transient; stimulation sustained for longer than about 20s results in a steep decline in hormone release, which probably reflects inactivation of Ca^{2+} entry into the terminals. This fatigue is readily reversed, and a 20s rest period will allow a new stimulus again to evoke efficient release. Phasic firing appears to make optimal use of the properties of the terminal membranes to enable hormone release to occur with minimal expenditure of energy on spike generation.

Phasic firing is the result of intrinsic membrane properties. The bursts are not passive responses to a phasically patterned input, nor do they reflect spontaneous oscillations of membrane potential. Instead, the bursts are regenerative, in that the first few spikes of a burst trigger prolonged activity. The bursting depends on intracellular Ca^{2+}; blockade of Ca^{2+} entry or chelation of intracellular Ca^{2+} will block phasic firing. At the start of a burst, a small but long-lasting depolarising after-potential (DAP) follows each spike, and these DAP's summate, bringing the membrane potential close to the spike threshold. After the first few spikes, a depolarising plateau, reflecting a persistent inward current, sustains a burst (Figure 7.1B). The plateau can be viewed as an alternative state of the resting potential. When depolarised by about 10 mV from its normal resting potential of about -70mV, a vasopressin cell will tend to settle at a new, more depolarised resting (plateau) potential, which is sustained by a constant depolarising current, and which is close enough to the spike threshold for EPSP's to frequently trigger spikes. The DAP after each spike brings the mem-

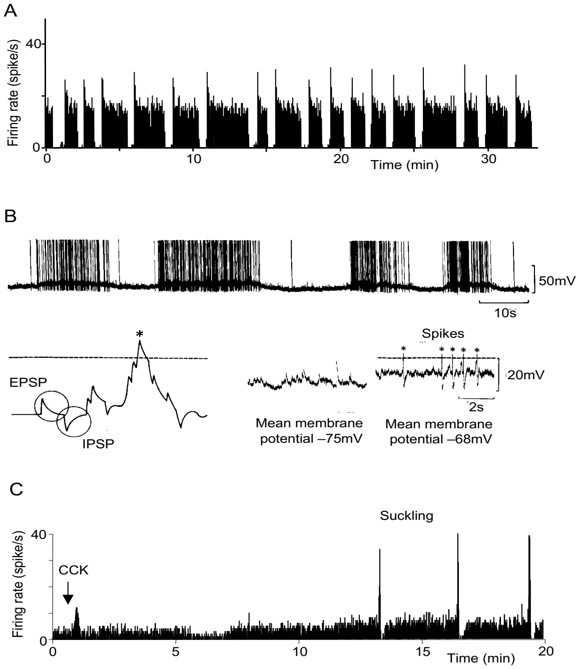

Figure 7.1

(A) Typical example of phasic discharge pattern in a vasopressin neuron characterised by succession of silent periods and active periods. Mean burst duration varies considerably between vasopressin cells, and the bursts occur asynchronously amongst the population of vasopressin cells. (B) Intracellular recording from a vasopressin cell *in vitro*. Spikes arise in these cells when EPSPs summate to bring the membrane potential above spike threshold. Phasic bursts occur because a sustained plateau potential, depolarising the vasopressin cell by typically about 7 mV, produces a sustained increase in excitability. The plateau is, however, itself activity dependent. (C) Typical example of firing pattern in an oxytocin cell, identified by its activation in response to i.v. injection of 20 μg/kg of CCK. During suckling of pups, brief and high frequency bursts of spikes are superimposed upon the slow and irregular background activity. These bursts occur synchronously amongst all oxytocin cells in the hypothalamus.

brane potential back into the range for activating the low threshold, non-inactivating currents, so plateau potentials normally are triggered by, and maintained by, spike activity. The vasopressin cell is bistable, in exhibiting two alternative stable states: the *normal* resting potential, and the plateau potential; it is either active in a burst, or it is silent – and it repeatedly oscillates between these states.

An essential component of phasic activity is activity-dependent inactivation of the plateau, which allows the vasopressin cell to fall silent after a burst until DAP conductances are activable again. This makes the cell a *bistable oscillator*: it will tend not to stay in either stable state indefinitely, but will alternate between the two. Bursts evoked by current injection or by antidromic stimulation are followed by an activity-dependent after-hyperpolarisation resulting from a slow, Ca^{2+}-dependent K^{+}-conductance which functions as a feedback inhibitor of spike activity; this channel type can be identified by use of the toxin apamin; in the presence of apamin, the afterhyperpolarisation (but not the HAP) is blocked; as a result, bursts are more intense but phasic firing is still present, so this after-hyperpolarising mechanism does not alone account for burst termination.

7.2.2 Implications of membrane bistability for responsiveness to afferent input

The membrane properties of vasopressin cells are reflected in distinctive features of their behavior. Electrical stimuli applied to the axons evoke spikes that are propagated antidromically to the cell bodies. Just as spontaneous spikes are followed by a DAP, so are antidromic spikes, and brief trains of antidromic spikes can trigger full bursts of activity in vasopressin cells. Interestingly, trains of antidromic spikes can also *stop* established bursts, through exaggerating the activity-dependent inactivation of the plateau potential. Low-frequency antidromic spikes on the other hand produce the interesting effect that cells appear to compensate for the additional evoked spikes by a matching reduction in spontaneous discharge. Vasopressin cells thus *defend* their firing rate against perturbations, probably via an intrinsic after-hyperpolarisation (AHP), which acts as a feedback inhibitor of spike activity. Antidromic activation mimics particular effects of synaptic excitation – consequences of spike activity *per se* follow whether the spikes are induced by synaptic input or by direct stimulation. Thus excitatory inputs may trigger a burst if a vasopressin cell is silent, or may stop a burst if a cell is active, or may have no effect if it is weak enough to allow the cell to defend its firing rate effectively. Similar paradoxical effects are observed with inhibitory stimuli.

So vasopressin cells fire in bursts, and bursts release vasopressin efficiently. Bursting reflects an intrinsic membrane bistability, and this property causes vasopressin cells to respond to inputs in a complex manner. Vasopressin cells require tonic synaptic input in order to function normally, but are individually relatively insensitive to changes in the level of input, except that small changes can, rather unpredictably, trigger transitions between activity and silence. One role of synaptic input is thus to permit the expression of patterns of activity in vasopressin cells – without synaptic *noise*, this behavior cannot be displayed.

Signals and noise are often thought of as mutually incompatible. However, the reliability of information transfer can, in some systems, be paradoxically enhanced by noise, a phenomenon referred to as stochastic resonance. Background activity in neurons in the absence of an identifiable signal associated with that activity, what we might call neural noise, may not merely reflect activity in neurons poised at the threshold of responsiveness, but may play a role in fashioning the behavior and signal sensitivity of target neurons. One implication of this is that when removal of an input impairs the response of a neuron to a stimulus, we cannot infer, from this observation alone, that the input encodes any information about the stimulus. Some neurons may play an important role even if they carry no identifiable information, and their output is unaffected by physiological stimuli, if their activity provides a level of synaptic noise which is important to support key dynamical behavior, either in neuronal networks or in single cells.

7.2.3 Firing patterns of oxytocin cells

Unlike vasopressin neurons, oxytocin neurons never discharge phasically. Under most circumstances they discharge continuously, at between 0 and 12Hz, but an important exception to this is seen during lactation, when, in response to suckling, in addition to the background activity that resembles that seen in non-lactating animals, oxytocin cells display occasional high frequency discharges of spikes at up to 100Hz for 2 to 4s (Figure 7.1C). In the rat, these *milk-ejection bursts* occur every 5 to 10min, for as long as the suckling continues, and occur synchronously between all oxytocin cells in the hypothalamus. At all other times, no correlation is apparent between the discharge activities of neighbouring oxytocin cells. Thus, normally, oxytocin cells appear to behave autonomously, but in some circumstances they discharge in a manner reflecting positive-feedback interaction amongst the population. At present, it is believed that this interaction reflects a labile dendro-dendritic interaction between oxytocin cells that provides a variable level of weak mutual excitatory interaction. Dendritic release of oxytocin is activity dependent, but it is also modulated by intracellular signalling mechanisms. Importantly, the amount of activity-dependent oxytocin release is determined largely by extrinsic priming factors.

7.2.4 Intrinsic properties

Magnocellular neurons have a resting membrane potential of between 55 and 70mV, an input resistance of between 50 and 250MΩ, and membrane time constants ranging between 9 and 18ms. Single electrode voltage-clamp recordings measured steady-state current-voltage (I-V) relations that were nearly linear between 100 and 60mV when performed from a holding potential near 60mV. Varying the holding potential and the application of various channel blockers, resulting in changes in the I-V relationship, show a variety of Ca^{2+} and K^{+} currents. In particular, whole-cell patch-clamp recordings have revealed at least five different components in the voltage-dependent Ca^{2+} currents in magnocellular neurons: a T-type current with a low threshold of activation (-60mV), rapid inactivation at peak amplitudes ($\sim$40ms), high sensitivity to Ni^{2+}; a low threshold L-type channel with a threshold of around 50mV, slowly inactivating ($\sim$1300ms), and sensitive to nifedipine; a R-type current, threshold of 50mV, inactivation of $\sim$200ms, and insensitive to toxins; a P-type current, non-inactivating, and blocked by ω-agatoxin IVA; and an N-type-current, slowly inactivating ($\sim$1800ms), blocked by ωconotoxin GVIA.

In both oxytocin cells and vasopressin cells, every spike is followed by a hyperpolarising afterpotential (HAP), which lasts between 50 to 100ms, and results from a rapidly activated Ca^{2+}-dependent K^+ conductance, similar to the current termed I_a in other cells. The I_a can be activated when a cell is depolarised following a period of hyperpolarisation, and it serves as a damper to space successive spikes. Accordingly, the HAP sets an upper limit on the maximal firing rate which can be achieved during a depolarising stimulus, and in the case of magnocellular neurons the HAP is large and long lasting, and this upper limit is accordingly quite low. Under most circumstances oxytocin and vasopressin cells will not sustain a discharge rate exceeding 15Hz for more than a few seconds, and the minimum interval between successive spikes is very rarely less than about 30ms. Oxytocin cells adhere to this limit under all circumstances but one: during milk-ejection bursts they dramatically escape this limit. The outward current underlying the HAP is evoked by a depolarising voltage current pulse from a threshold of 75mV, reaches the peak within 7ms and subsequently decays monotonically with a time constant of 30ms. Steady-state inactivation is complete at potentials positive to 55mV, and the inactivation is removed following tens of milliseconds at hyperpolarised voltages. Further, the I_a is strongly dependent on extracellular Ca^{2+}, whereby its influx during a spike may contribute to the repolarisation, as well as to the peak and initial phase of the HAP. Indeed, its amplitude appears to be directly proportional to the external concentration of Ca^{2+}.

In contrast to the HAP, which is evoked by single spike, trains of spikes are followed by a prominent afterhyperpolarisation (AHP). The magnitude of the AHP is proportional to the number of spikes during the preceding spike train, with an exponentially progressing onset and a maximum after the first 15 to 20 spikes, regardless of the frequency at which spikes were evoked. The steady-state amplitude increases logarithmically between 1 and 20Hz. The AHP lasts hundreds of ms, and its duration also depends on the duration and frequency of the spike train. The AHP is associated with a 20 to 60% decrease in input resistance, shows little voltage dependence in the range 70 to 120mV, and is proportional to the extracellular Ca^{2+} concentration. These observations led to the conclusion that the AHP results from the activation of a slow, voltage-independent Ca^{2+}-dependent K^+ conductance, the I_{AHP}. The distinction between the I_{AHP} and the I_{to} (and correspondingly between the post-train AHP and the post-spike HAP) is made since the ionic currents are pharmacologically distinct. The I_{to} is less sensitive to tetraethyl ammonium, but is reduced by 4-aminopyridine and dendrotoxin. In contrast, I_{AHP} is blocked by apamin, with the effect of a threefold increase in the mean firing rate of spontaneously active neurons. Pharmacologically, the AHP appears to have a fast component and a slow component; the fast AHP is blocked by apamin, while the slow AHP is blocked by charybdotoxin, and is affected by low concentrations of tetraethyl ammonium. Apamin generally blocks small-conductance (SK) Ca^{2+}-dependent K^+

channels, while charybdotoxin blocks big-conductance (BK) Ca^{2+} dependent K^+ channels, as well as some other K^+ channels.

7.2.5 Intracellular Ca^{2+} concentration

In response to any spike activity, there is a large Ca^{2+} entry into oxytocin cells and vasopressin cells via several different voltage-gated Ca^{2+} channels. In addition, intracellular Ca^{2+} stores can be mobilised via second messenger pathways to give very large increases in intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$) (see [10, 11, 12, 13]). The dynamics of Ca^{2+} change differ between the cell types, as they differently express Ca^{2+} binding protein calbindin [4]. Oxytocin cells contain more calbindin than vasopressin cells, allowing them a higher Ca^{2+} buffering capacity, which prevents generation of DAPs and therefore phasic firing. DAPs and phasic firing can be evoked in oxytocin cells by neutralising calbindin or by increasing $[Ca^{2+}]_i$. Conversely, phasic firing neurons can be switched to continuous firing by introduction of exogenous calbindin or by chelation of intracellular Ca^{2+}. The amplitude of DAPs depends on Ca^{2+} influx through voltage-dependent Ca^{2+} channels of L- and N-types, but also on Ca^{2+} release from intracellular Ca^{2+} stores, notably thapsigargin-sensitive stores, located in the endoplasmic reticulum.

Both vasopressin and oxytocin cells have thapsigargin-sensitive intracellular Ca^{2+} stores. In oxytocin cells, oxytocin itself induces an increase in $[Ca^{2+}]_i$ by activating IP3 pathway-coupled oxytocin receptors, which results in the release of Ca^{2+} from thapsigargin-sensitive stores. Oxytocin apparently does so without any strong accompanying depolarisation. In vasopressin cells, vasopressin also induces a Ca^{2+} response, but in a more complex way. Vasopressin-induced $[Ca^{2+}]_i$ increase mainly involves an influx of Ca^{2+} via voltage-dependent Ca^{2+} channels of L-, N- and T-types, as it can be reduced by specific blockers of these channel types. In addition to Ca^{2+} coming from the extracellular medium, part of the response to vasopressin is also due to release of Ca^{2+} from thapsigargin-sensitive intracellular stores. The complexity of vasopressin actions probably results from activation of several types of receptors coupled to different intracellular messenger pathways. Vasopressin receptors (described so far) comprise V_{1a} and V_{1b} type receptors, which are coupled to phospholipase C (PLC), and V_2-type receptors, which are coupled to adenylyl cyclase. Agonists of both V_{1a}- and V_2- receptor types can induce a $[Ca^{2+}]_i$ increase in vasopressin cells. In addition, inhibitors of PLC and adenylyl cyclase pathways, by blocking the production of intracellular messengers IP3 and cAMP, decrease the Ca^{2+} response to vasopressin. Vasopressin clearly can depolarise vasopressin cells to induce Ca^{2+} entry via voltage-gated channels, but also induces some liberation of Ca^{2+} from intracellular stores, and can also probably hyperpolarise vasopressin cells via Ca^{2+} -activation of Ca^{2+}-dependent K^+ channels; in practice, it seems that the actions of vasopressin on vasopressin cells are state-dependent; when applied on vasopressin cells *in vivo*, vasopressin tends to excite silent or slow firing cells but it tends to decrease the firing rate in active cells.

7.2.6 Implications

To build a realistic biophysical model of an oxytocin cell or a vasopressin cell from the bottom up, there are in principle a very large number of basic membrane properties to be incorporated. Apart from the Na^+ and K^+ conductances that underlie the generation of the spike, there are a large number of Ca^{2+} conductances and K^+ conductances that appear to play specific, potentially important roles, and since several of the latter are Ca^{2+} -dependent, the intracellular Ca^{2+} dynamics, involving buffering and mobilisation of intracellular Ca^{2+} stores also need to be included. The disposition of these conductances in different cellular compartments is poorly understood, and other conductances, in particular to chloride, and non-specific cation conductances, are also important, as may be the precise cellular topology. To build a network model it would also be necessary to incorporate elements reflecting the nature of stimulus-secretion coupling and the different underlying mechanisms at dendrites and nerve endings; vasopressin cells, for instance, express different populations of Ca^{2+} channels at the soma and nerve terminals.

On the other hand, we might take an approach that is consciously simplistic, incorporating progressively only those features of cells that are essential to explain particular behaviors, and incorporating these in a minimalist way. We need to set a verifiable objective: to develop computational models that mimic cells so closely, that for specific defined attributes, they are essentially indistinguishable in their behavior from real cells. In the example shown hereon, we look at the normal discharge patterning of oxytocin cells and seek a minimalist quantitatively accurate model of this. To be a good model, the spike output of the model cell must be indistinguishable from outputs of real cells by any statistical analysis that is applied.

7.3 Statistical methods to investigate the intrinsic mechanisms underlying spike patterning

7.3.1 Selecting recordings for analysis

Before starting the analysis of the firing activity, we must select suitable recordings. Stationarity is an essential prerequisite for the analysis to be meaningful. A series is stationary if there are no systematic trends or *rhythmic* variations. We started with stable, long recordings (up to 3h) of spontaneous activity from identified rat oxytocin cells *in vivo*, and from these, stationary recordings, or long stationary stretches from recordings that were not stationary throughout, were chosen. Stationarity was checked with the help of *bicubic splines*, a series of smooth cubic curves fitted over short stretches of activity, then joined with the same slope at the joints to form one continuous curve. One way of looking at temporal patterns is by looking at the time between consecutive spikes (*interspike intervals*). Interspike intervals provide information about the relationship between spikes in a way easily accessible for statistical scrutiny. The interspike interval histogram is a graphical representation of the distribution of the occurrence of intervals of a variable length. The distribution allows a first indication about spike patterning in an individual neuron.

7.3.2 Interspike interval distributions

For both oxytocin cells and vasopressin cells, the interspike interval distribution is skewed, with a single mode and a long tail (Figure 7.2). For vasopressin cells, modes are in the range 40 to 60ms, and for oxytocin cells, in the range 30 to 80ms. The tail of each distribution ($>$200ms) can be well fitted by a single exponential, and extrapolation of this exponential shows a deficit of intervals below the curve in the range 0 to 40ms, consistent with the effect of a HAP. For vasopressin cells, there is an excess of intervals above the curve in the range 40 to 100ms (Figure 7.2B), consistent with the effect of a DAP. No such excess is observed in oxytocin cells (Figure 7.2A), indicating that oxytocin cells display little or no DAP when normally active *in vivo*. The good exponential fits for oxytocin cells indicate that, to a first approximation, beyond about 80 ms after any given spike the arrival time of the next spike is essentially random. Thus the activity of oxytocin cells is dominated by factors affecting the probability of spike occurrence that are independent of previous spike history, i.e., the mean resting potential and the rate of synaptic input; and a reduction in excitability following each spike that decays over 40 to 80ms, consistent with the expected effects of a post-spike HAP. This inference is more clearly apparent from the construction of hazard functions (Figure 7.3): these describe cell excitability as a function of time elapsed since a spike, by calculating the probability of cell firing per unit elapsed time from the interspike interval data. For oxytocin cells the hazard function shows a low probability of firing (reflecting the HAP) for about 50ms, and a constant hazard thereafter; vasopressin cells show a sequence of low probability followed by high probability before return to constant hazard.

7.3.3 Modelling

To test this inference, we [15] modelled oxytocin cells and vasopressin cells by a modified *leaky integrate and fire model* [16]. EPSPs and IPSPs generated randomly and independently at mean rates R_E and R_I produce perturbations of membrane potential that decay exponentially. These summate to produce a fluctuating *membrane potential*. When a fluctuation crosses a spike threshold, T, a spike is generated, followed by a HAP, modelled as an abrupt, exponentially decaying increase in $T = T_0(1 + k\exp(-\lambda t)$ where t is the time since the last spike, T_0 is the spike threshold at rest, and k and λ are constants. Intracellular recordings from oxytocin cells reveal EPSPs and IPSPs of 2 to 5mV that last for 5 to 10ms; we assumed that EPSPs and IPSPs at rest were of equal and opposite magnitude (at T_0) with identical half-lives. Oxytocin and vasopressin cells have resting potentials of about -62mV with a spike threshold of about -50mV, and are depolarised in direct response to hyperosmotic stimulation following shrinkage, resulting in inactivation of specialised non-adapting stretch-sensitive K^+ channels [17]. *In vivo*, the peak

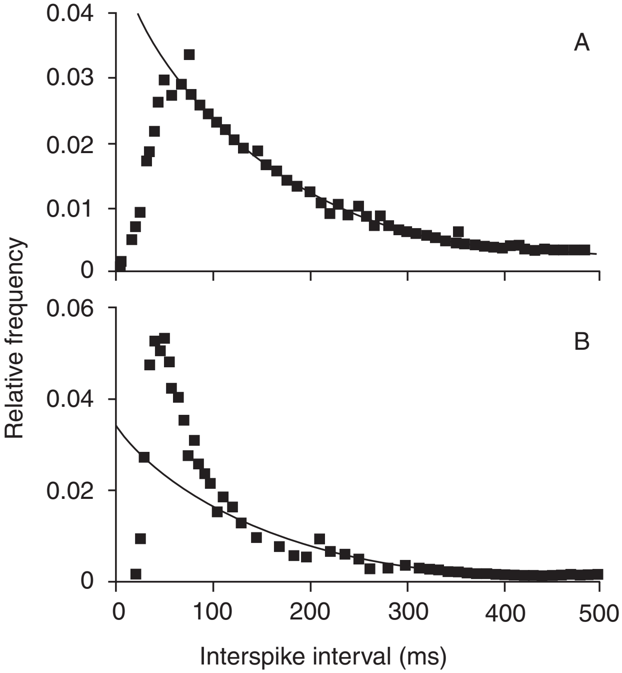

Figure 7.2

Representative inter-spike interval distributions from an oxytocin cell (A) and a vasopressin cell (B) showing exponential curves fitted to the tails of the distributions and extrapolated to lower interval values. Both oxytocin cells and vasopressin cells show a deficit of intervals below the fitted curve in the region 0-30 ms, reflecting the presence in both cells of a strong post-spike hyperpolarisation (the HAP). For vasopressin cells, but not for oxytocin cells, such extrapolations revealed a large excess of intervals above the fitted curve in the range 30 to 150ms, reflecting the presence in vasopressin cells only of a depolarising afterpotential (DAP) following the HAP.

activation (at $\approx$ 12Hz) is attained after osmotic stimulation has raised extracellular [Na$^+$] by $\sim$10mM, producing a direct depolarisation of $\approx$3 to 5mV. The equilibrium value for T, T_0, was thus set at 12mV for the simulations shown, and simulations were conducted over 1mV below to 5mV above this level. We conducted simulations with parameter values systematically spanning the ranges above, restricted to output ranges (0 to 16Hz) consistent with the behavior of oxytocin cells.

7.3.4 Simulating oxytocin cell activity

We found that the inter-spike interval distribution from each oxytocin cell could be closely matched by a model cell with a resting potential (T_0) of 12mV below spike threshold subject to random EPSPs of 4mV amplitude and 7.5ms half-life (Figure

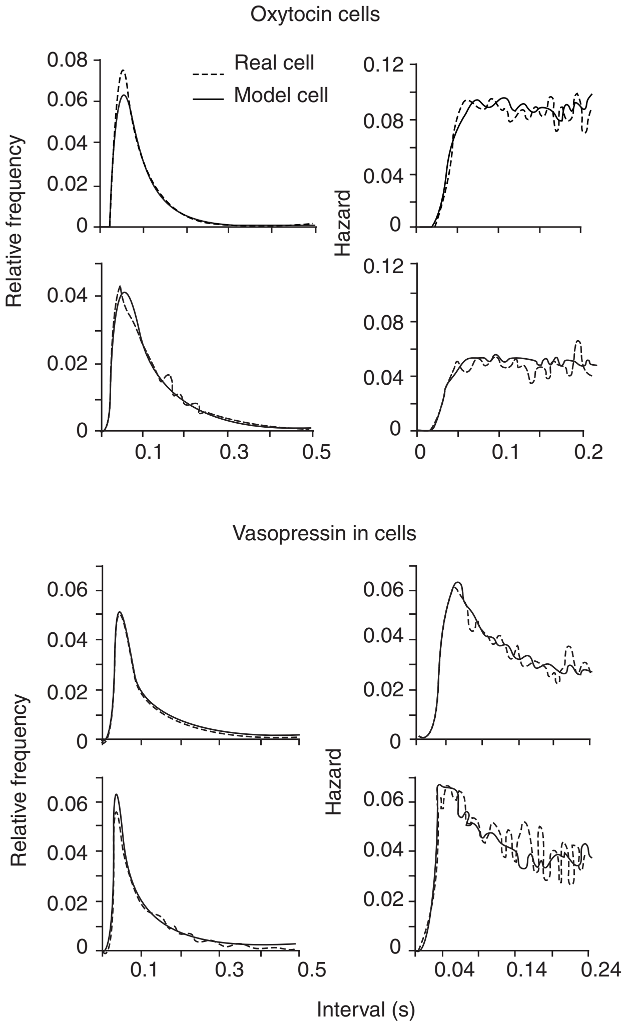

Figure 7.3

Left panels: inter-spike interval distributions from two oxytocin cells (top) and two vasopressin cells (bottom), each of which is modelled by a modified leaky-integrate-and-fire model neuron (see text). The histograms from the model cells are superimposed on the original cell data. Right panels: hazard functions plotted from the same data for each cell and each model neuron superimposed. The interspike interval distributions of oxytocin cells and vasopressin cells can be fit remarkably well with relatively simple models, that in the case of oxytocin cells mimic the effects of a HAP only, and for vasopressin cells that mimic the effects of a HAP and subsequent DAP.

7.3). Each cell could be closely fitted by varying just λ and I_E. The fits were not unique; good fits could be achieved for different values of EPSP size (2 to 5mV) or half-life (3 to 20ms), or for different values for T_0 by changes in R_E, or for different values of k by adjusting λ. Similarly, the shape of the distribution is little affected if the model is challenged not with EPSPs alone but with a mixture of EPSPs and IPSPs. However, for a chosen value of k, every distribution could be characterised by a unique λ, and by T_0, R_E and R_I, which affect the output rate but have little other effect on the shape of the distribution in the relevant range.

For vasopressin cells, inter-spike interval distribution could be well fitted by the output of a similar model cell with the addition of a component to mimic the effect of a slow DAP, modelled as an abrupt, exponentially decaying increase in $T = T_0(1 + k\exp(-\lambda t)$ where t is the time since the last spike (Figure 7.3). Adding this to the oxytocin model produces a sequence of fast HAP followed by a slow DAP. Of course this simple model does not reproduce phasic firing, only the characteristic distribution of inter-spike intervals. Essentially, phasic firing does not occur in this simple model because it incorporates no mechanism for burst termination or burst refractoriness, hence once a burst is triggered it may continue indefinitely, or once ended may be retriggered immediately.

But leaving aside the vasopressin cell model for the present, the good fit of a simple model to oxytocin cell data enabled us to test the hypothesis that the oxytocin cell response to osmotic stimulation arises from an increase in synaptic input combined with a direct depolarisation, with no change in the intrinsic mechanisms that govern post-spike excitability. If so, then it should be possible to fit inter-spike interval histograms from any one oxytocin cell at different levels of activity with a common λ. This proved true for each cell tested (Figure 7.4). We then studied how the firing (output) rate of model cells changes with the synaptic input rate, and with increasing depolarisation.

7.3.5 Experimental testing of the model

In a healthy adult, above a fixed threshold, vasopressin release varies linearly with osmotic pressure over a wide dynamic range. The relationship between the plasma concentration of vasopressin (v) and osmotic pressure (x) fits the equation $v = ax + b$, where b is the threshold osmotic pressure or set point, and a is the slope of the osmoregulatory mechanism. When the behavior of individual vasopressin cells and oxytocin cells is studied, it is strikingly apparent that individual neurons also show a linear increase in firing rate in response to increased osmotic pressure, that this linearity is apparent throughout the normal dynamic range of their spike activity, and that the slope of the response is relatively constant between cells even when their spontaneous firing rates differ markedly (Figure 7.5A).

However, in the absence of IPSPs, an increase in EPSP rate produces a non-linear increase in output (over the physiological output range), and this is true broadly regardless of the parameter values within the ranges as described above. Elaborating the model to incorporate a reversal potential for EPSPs (of 38mV) does not significantly alter this conclusion. The effective dynamic range of oxytocin cells is from

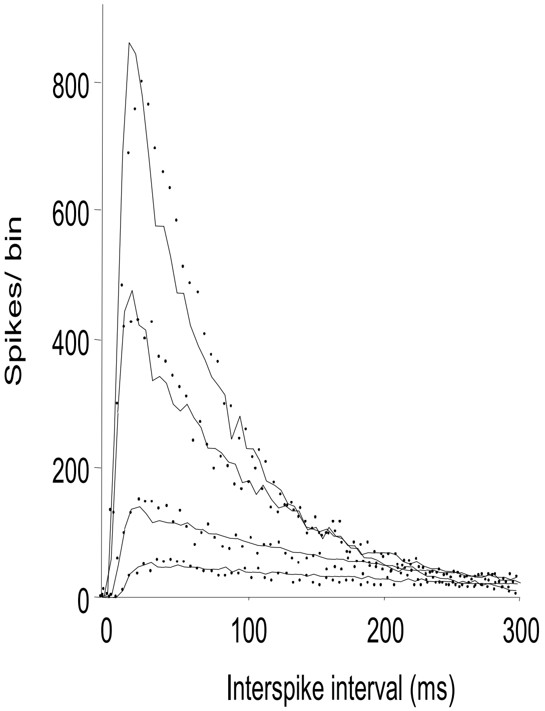

Figure 7.4

Comparison between model inter-spike interval distributions (lines) at different levels of synaptic input, and distributions observed in an oxytocin cell at different times during infusion of hypertonic saline (points). Oxytocin cell inter-spike interval distributions were constructed over 1000s, model cell distributions over a simulated 25000s, normalised for comparison. Oxytocin cell distributions correspond to mean firing rates of 12.8Hz, 9.3Hz 4.8Hz and 2.5Hz. Model cell distributions were constructed for equal average numbers of EPSPs and IPSPs, over a range of PSP frequencies that matched the range in firing rates observed during the period of recording analysed, producing output rates close to the average firing rates of oxytocin cells. A single value of λ (0.08) produces good fits for this cell at all levels of activity.

about 0.5Hz (lower range of spontaneous rates) to about 10Hz (peak sustained rates). This range was spanned in the model cells by a narrow range of R_E typically by a change in R_E from 110/s to 180/s. Osmotic stimulation is accompanied by a direct depolarisation of 3-5mV [17], and an equivalent change in T_0 leads to a compression of the range of R_E needed. This suggests that tonically active cells subject to EPSP input alone will respond strongly to osmotic stimuli as a result of the direct osmotic depolarisation, even with no change in synaptic input. Furthermore, similar changes in R_E from a different initial rate, but accompanied by the same osmotic depolarisation, result in very different amplitudes of responses. This suggests that oxytocin cells that differ in initial firing rate as a result of differing initial EPSP rates will respond in a divergent manner to a subsequent identical stimulus (Figure 7.5B–D).

Although these inferences are broadly independent of assumptions about EPSP size, half-life, and resting potential, they are not consistent with experimentally observed behavior. Osmotic stimulation is accompanied by large increases in the activity of afferent neurons. Moreover, the inference of divergent responsiveness of cells with different initial firing rates is not consistent with the consistency and linearity of the neuronal responses observed *in vivo*.

However, in model cells, the relationship between output rate and input rate becomes shallower as the ratio of IPSPs to EPSPs is increased, and this is true both for models that assume that EPSP and IPSP size are independent of voltage, and for models that incorporate appropriate reversal potentials for both. Comparing simulation results of models with and without reversal potentials, it is apparent that while the latter are less sensitive to synaptic input, it is equally true for both models that a high proportion of IPSPs produces a linearisation of the input-output relationship [18]. We therefore conducted simulations combining a direct depolarisation with an increase in balanced input, comprising equal average numbers of EPSPs and IPSPs (Figure 7.3). Under these conditions, model cells that differ in initial output rate as a result of differing initial input rates respond similarly to a given stimulus.

The simple oxytocin cell model described here indicates that an increase in IPSP frequency that accompanies either an increase in EPSP frequency or a steady depolarising influence, will moderate the rate of increase in firing rate, will linearise the input-output relationship, will extend the effective dynamic range of the output neuron, and will tend to make the response of a neuron to a given input independent of the initial firing rate. The theoretical analysis thus indicated that a high proportional activation of inhibitory input confers appropriate characteristics upon the responses of magnocellular neurons to osmotic inputs. Thus this model produced the highly counter-intuitive prediction that when magnocellular neurons are excited in response to an increase in osmotic pressure, that increase reflects not only an increase in excitatory input but also an increase, of equal or greater magnitude, in inhibitory input. This counter-intuitive prediction was then tested, and supported first by direct measurement of a large increase in the release of the inhibitory neurotransmitter GABA in the supraoptic nucleus during osmotic stimulation, and second by studies of the effects of blocking GABA neurotransmission on the responses of oxytocin cells to osmotic stimulation.

7.3.6 Firing rate analysis

As can be seen above, we might conclude from the shape of the interspike interval distribution that spontaneous firing of an oxytocin cell is a renewal process effectively Poissonian except for the effect of the HAP. However, weak, slow activity-dependent influences might not be apparent from the interspike interval histogram, as to exert a significant influence, they would require summation of the effects of several spikes occurring in a short period. Since we are interested in periods of activity as opposed to individual spikes, we can take one of two approaches: we can anal-

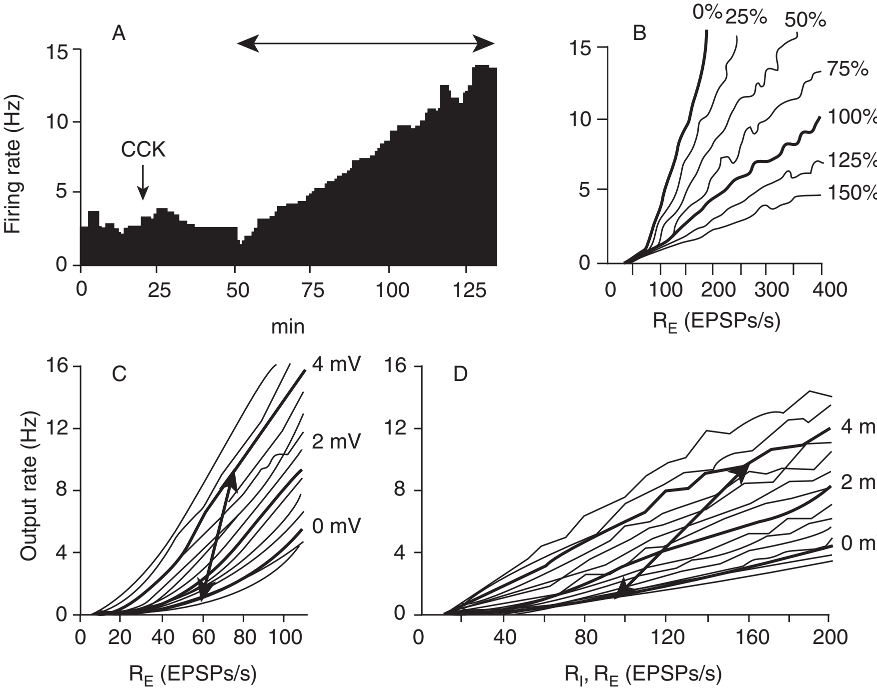

Figure 7.5

(A) Response of a typical oxytocin cell to a linear increase in plasma osmotic pressure, induced by an slow intravenous infusion of hypertonic saline administered in the period marked by the double headed arrow. This recording also shows the transient excitation induced by intravenous injection of CCK. Note the striking linearity of the increase in firing rate. (B) Relationship between the output rate of a model oxytocin cell and EPSP rate R_E. The different lines show the effect of adding IPSPs in proportion to the EPSPs. For a balanced input, comprising equal average numbers of EPSPs and IPSPs, the input-output relationship is shallower and more linear than for EPSPs alone. (C,D) Relationship between output rate and R_E, for values of resting potential varying in 0.4mV steps from the standard value (62mV), The double-headed arrows connect points corresponding to an output rate of 1Hz at the initial resting potential to points corresponding to 10Hz at a membrane potential depolarised by 4mV. This line thus indicates the apparent dynamic range of oxytocin cells in response to osmotic stimulation *in vivo*. (C) shows simulations for a cell stimulated by EPSPs alone, (D) shows simulations for a cell stimulated by an equal number of EPSPs and IPSPs. Adapted from [16].

yse serial dependence between average firing rate measured in successive short time intervals; or we can look at serial dependence of instantaneous interspike interval length on the past discharge activity.

What would we expect if spikes were generated wholly independently of the previous incidence of spikes? For a renewal process, the probability of an event occurring at any given time is independent of the timing of preceding events. In particular:

1. The tail of the inter-spike interval distribution should be well described by a single negative exponential (i.e. except where the refractory period of the cell prevents firing).

2. Data should show invariant statistical characteristics when shuffled randomly.

3. As the variance of the event frequency (σ^2) equals the mean of the event frequency (μ) for a Poisson process, the index of dispersion (σ^2/μ) should be close to 1 (if the relative refractory period is relatively small), and should be independent of bin-width.

We analysed recordings from oxytocin cells to investigate how they deviate from randomness by each of these criteria. For each cell, the sampled firing rates (sampled in successive short time intervals, e.g., 1s bins) were expressed as a distribution, showing the relative frequency of the occurrence of particular firing rates. These distributions are typically bell-shaped, with the peak around the mean firing rate, and a rather symmetrical spread. In the next step the interspike intervals were shuffled randomly creating a new *recording*. We recalculated the firing rates - again for 1s bins - and compared the new distribution of firing rates to the original distribution. If the original distribution is based on randomly occurring intervals, than further randomisation should have no effect. However, in the overwhelming majority of cases, the distribution of randomised firing rates was wider than the distribution of observed firing rates (Figure 7.6). In other words, the observed firing rate distribution was more uniform than would be expected if there were no serial dependence between interspike intervals.

If randomisation did not consistently affect the shape of the distribution we might have reasonably concluded that there were no activity-dependent influences that had a significant influence on spike patterning beyond those that influence the interspike interval distribution. Since it does, we can conclude that activity-dependent mechanisms underlie the spontaneous activity of oxytocin cells. The mechanisms are weak in the sense that their effects are not readily apparent in the interspike interval distribution, but are apparent on a time scale of 1s. In other words, variations in average activity of the order observed from second to second are enough to produce discernible feedback effects on spike activity.

7.3.7 Index of dispersion

We stated above what we expect from the index of dispersion if spike arrival would be generated independently of previous spikes. The index of dispersion is the ratio of the variance of firing rate to the mean firing rate. For a Poisson distribution, the variance of the event frequency equals the mean, and the index of dispersion = 1, and is independent of bin width. The closer to 0 the index of dispersion, the more ordered

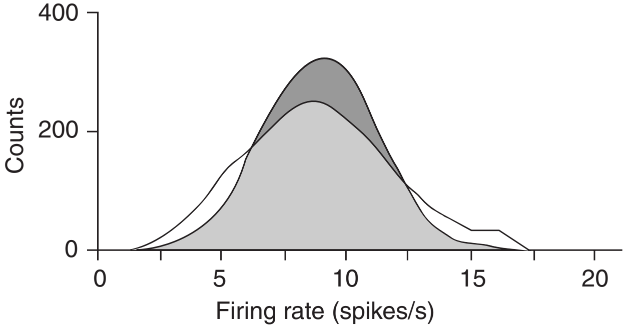

Figure 7.6

Observed firing rate distribution (dark area) and randomised firing rate distribution (white area). The figure shows the analysis of a stationary stretch of a recording of the activity of an oxytocin neuron. Firing rates were calculated in 1 s bins and the shaded curve shows the distribution of firing rates in these 1s bins. The original interspike intervals were randomised and from the randomised data new second by second firing rates were calculated. The white curve shows the distribution of firing rates from the randomised data. The firing rate distribution has a higher peak, but is narrower than the randomised firing rate distribution, indicating that the sample is more uniform than would be expected from a completely random sample. The discrepancy between the two distributions is a first indication for structure in the firing pattern.

the underlying series. An index of dispersion above 1 suggests that the series is more irregular than a Poisson process, and could be an indication of heavy clustering. We calculated the index of dispersion for oxytocin cells, using different bin sizes; the index of dispersion differed with varied bin widths in a characteristic way: for very small bins the index of dispersion was high, (average value 0.7 ± 0.05), but decreases when the bin width increases. Thus, when looking at the recording at a very short time scale such as 0.06s, firing appears to be near random, but, when looking at increasingly longer periods (up to 2s), the firing appears to be more and more ordered. The firing appears to be most ordered when looking at a time scale of 4 to 8s, where the index of dispersion is the smallest (0.27 ± 0.4). Since the index of dispersion depends on bin size, the generation of spikes is not independent of previous activity.

7.3.8 Autocorrelation analysis

Autocorrelation refers to the serial dependence of observations in a stationary time series; autocorrelation coefficients measure the correlation between observations at different lag times. In our case, the observations are firing rates for a specified period of time, or window, and to calculate the autocorrelation coefficient of lag n, the observation at time t is compared to the observation at time $t+n$. For small windows

(0.06 to 1s), a negative autocorrelation was found for most oxytocin cells, but for larger window sizes, this negative autocorrelation disappears. Thus, at small time scales (0.06 s to 1s), periods of relatively high activity are likely to be followed by periods with relatively low average activity, and vice versa. Thus the average firing rate over long periods of time is much more regular than would be expected from the local variability in firing rate.

This type of analysis can be logically extended to serial dependence of interval length. The simplest approach is to consider how the length of a given interval δt depends upon the history of previous spike activity the preceding intervals $t_2, t_3, t_4, t_5, \cdots$ When oxytocin cell recordings are analysed in this way, there is a negative relationship between t_1 and $t_2 + \cdots + t_n$, the slope of which is typically maximal for n between 5 and 10, and which approaches 0 as n exceeds about 20. This indicates that for oxytocin cells firing at typical background rates, there is a negative effect of discharge rate upon spike activity through mechanisms slow enough to summate over 5 to 10 spikes in 1 to 2s (Figure 7.7).

It should be remembered that the biological message encoded by a single oxytocin cell is only the contribution that that cell makes to the secretion from the oxytocin population as a whole. The overall secretion rate has considerable biological significance, and the rate must be maintained at an appropriate and steady average level for prolonged periods when necessary for regulating sodium excretion (natriuresis), but local second-by-second variability in the secretory rate of individual neurons is of no biological significance unless such changes occur synchronously throughout the population (as during reflex milk ejection). For oxytocin cells therefore, what is important for natriuresis is only that they accurately maintain an average steady state activity when measured over long periods. Oxytocin cells clearly have activity-dependent negative-feedback mechanisms that ensure long term stability of average firing rate.

7.4 Summary and conclusions

Although the interspike interval distributions of oxytocin cells seem to suggest that spike arrival times (during spontaneous activity) are largely independent of previous firing activity (except for the refractory period), closer analysis shows otherwise. Firing rate analysis demonstrated that the index of dispersion did not equal 1 and was not independent of bin width. Further, while the activity appears to be nearly random at a small time scale, over a scale of several seconds it appears much more ordered. The analysis of serial dependence showed that on a small time scale the activity is clustered, but on a larger time scale the activity is more homogenous. Thus on a short and medium time scale the cell possesses a *memory* and balances the activity, whereby periods of short intervals tend to be followed by periods with longer ones, and vice versa. However, on a long time scale the activity is rather homogenous.

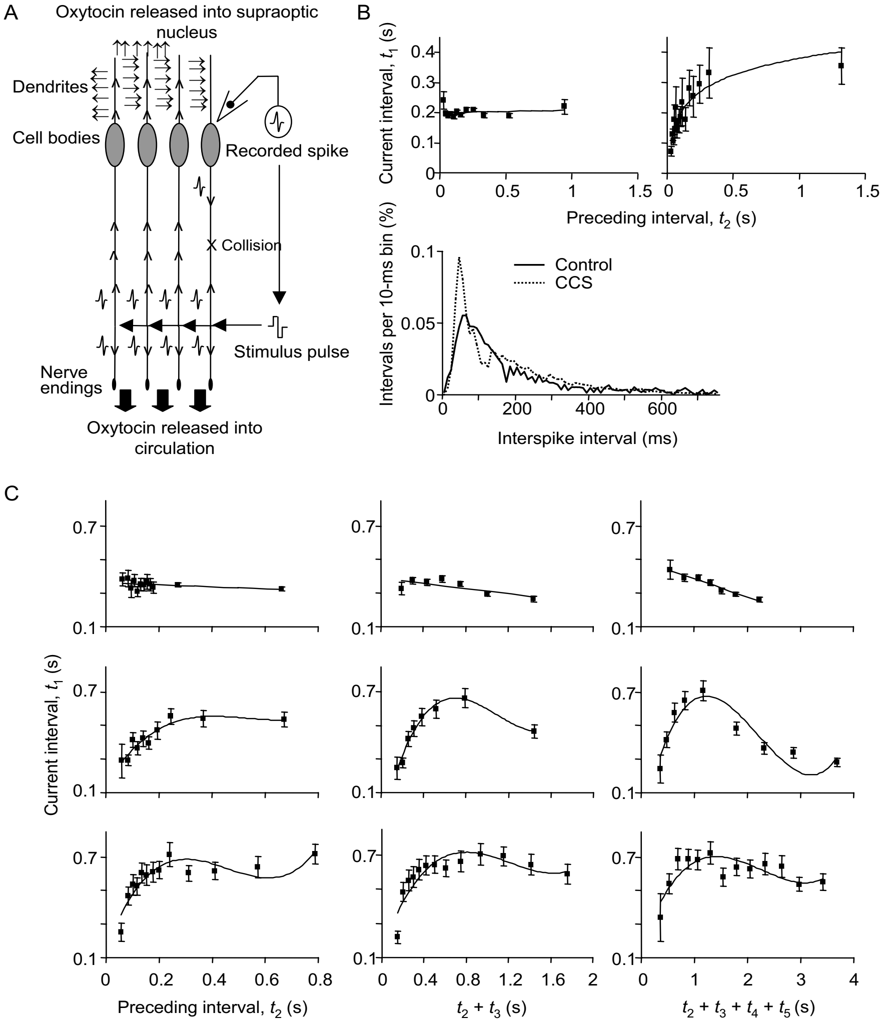

Figure 7.7

(A) Schematic diagram illustrating the technique of constant-collision stimulation (CCS). Spontaneous extracellular spikes are recorded from a supraoptic neuron. (continued).

Figure 7.7

Each spontaneous spike triggers the application of an electrical stimulus pulse to the neural stalk, which initiates an antidromic spike in the axon of every supraoptic neuron, since all supraoptic neurons project to the pituitary via the neural stalk. The antidromic spike evoked in the axon of the recorded neuron is extinguished by collision but other antidromic spikes persist to invade the cell bodies of most neighbouring neurons, and thence activates intranuclear connections and dendritic release of oxytocin. [19]. (B) Every interspike interval (t_1) in a selected recording period was paired with its predecessor (t_2) and preceding intervals t_3, $t_4, \cdots$. to study the dependence of current activity upon preceding activity-B shows the analysis of a representative oxytocin neuron, the mean t_1 (± standard error) is plotted against t_2 before (left upper panel) and after (right upper panel) CCS, and the corresponding interspike interval distributions before and after CCS (bottom panel). CCS stimulation induces an increase in the proportion of short intervals, as seen in the interspike interval histograms, and an increase in clustered firing, as shown by the positive relationship between t_1 and t_2. (C) Example of the mean t_1 (± standard error) against t_2, or $t_2 + t_3$, or $t_2 + t_3 + t_4 + t_5$ (with linear or polynomial trend lines) for a representative neuron in control conditions (upper panels), during CCS (middle panels), and during CCS + thapsigargin (bottom panels). From the top left panel in C it seems there is little influence of t_2 upon t_1, there is a negative regression, but with a very shallow slope. However, this weak influence is long-lasting and so summates with successive spikes, because in looking at t_1 vs. $t_2 + \cdots + t_5$, there is now a strong inverse correlation. During CCS this negative relationship is still present but is preceded by a positive relationship, indicating a short-lasting positive feedback action superimposed upon the normal slow negative feedback.

These results demonstrate structure in sequences of interspike intervals, and from its characteristics we may conclude it to be the effect of the AHP.

Thus sufficient information seems to be available from the characteristics of spontaneous discharge activity to produce concise computational models that can mimic this behavior closely when they incorporate features that appropriately describe the impact of intrinsic, activity-dependent mechanisms. Such models are unique descriptors of a particular neuronal phenotype, and are by design well-matched to experimental data, but are also capable of generating fresh insight into cell properties, testing the coherence and feasibility of biological hypotheses, and capable of generating novel and counter-intuitive predictions.

We have concentrated on demonstrating this approach for an example neuron with limited network connectivity. The oxytocin cell is an output neuron, with few axonal collaterals to make any recurrent connection with other neurons in the CNS, hence activity-dependent influences in activity primarily reflect intrinsic cell properties in normal circumstances. However, this approach is potentially particularly appropriate for analysing the behavior of neurons where activity-dependent influences are mediated by interactions with other neurons. As far as the analytical approach is concerned, this is indifferent to whether activity-dependent influences reflect intrinsic properties or external feedbacks, and concise models can be equally indifferent to this where it helps to collapse mini-neuronal networks into single elements.

As considered above, oxytocin neurons normally function autonomously, but during suckling in lactating rats, oxytocin cells show dramatically different behavior that reflects weak pulse coupling through dendro-dendritic interactions [19, 20]. These weak, mutually excitatory interactions have been extensively studied experimentally. The underlying mechanisms are complex: oxytocin is released from the dendrites in response to a rise in intracellular Ca^{2+}, but little oxytocin release normally results from spike activity. However, agents that liberate Ca^{2+} from intracellular stores also cause a large mobilisation of dendritic stores of oxytocin into readily-releasable stores that are subsequently available for activity-dependent release. One such agent is oxytocin itself, which triggers Ca^{2+}mobilisation from thapsigargin-sensitive stores in the endoplasmic reticulum after binding to specific oxytocin receptors on oxytocin cells. Thus dendritic oxytocin can enable subsequent activity-dependent oxytocin release. Oxytocin released around the dendrites has multiple other effects, in particular inhibiting glutamate release from afferent excitatory synapses and attenuating the effect of synaptically-released GABA by post-synaptic actions. Oxytocin is released in very high concentrations around the dendrites and in CSF has a long half-life, making it a neurohormonal-like messenger within the brain that can potentially act at distant sites and over a prolonged time scale, though local expression of peptidases may protect some sites of potential action.

Dendritic release of signalling molecules is far from unique to the oxytocin cells; dendritic release has been demonstrated in a number of systems and for a variety of molecules including vasopressin and dopamine, and may be a wholly general phenomenon, at least for peptidergic transmitters. The capacity of peptides to act at a distance through their long half-life in the brain, their ability to act at low concentrations through G-protein coupled receptors linked to a diversity of neuromodulatory actions, and the remarkable ability of peptides to selectively induce the expression of coherent behaviors, makes it important to integrate their effects into models of brain function. To model dendritic influences within biophysical frameworks is of course possible, but it may be as revealing and helpful to analyse and model their impact, rather than the underlying mechanisms.

The impact of activity-dependent positive feedback on the normal spontaneous activity of oxytocin cells should be apparent from the above-described statistical analyses; but there is no visible impact, or any impact overlaps with and is fully occluded by activity-dependent negative feedback (mediated by the AHP). To look at the potential for activity-dependent positive feedback in the network we can, however, look at how the structure of discharge activity is altered in defined experimental conditions. First, we can look at the effect on the activity of an oxytocin cell of synchronised activation of its neighbours through the technique of constant-collision stimulation. This reveals the existence of a rapid and transient mutual excitation that is normally masked by the HAP except when synchronous activation of neurons enhances this effect. Second we can look at the consequences of priming the releasable pool of oxytocin in the dendrites by treatment with the intracellular Ca^{2+} mobilising agent thapsigargin. Thapsigargin, like constant-collision stimulation, reveals weak mutual excitation, though whereas constant-collision stimulation amplifies this effect by the artificial synchronisation of electrical discharge, thapsigargin does so by amplifying the releasable pool of oxytocin in the dendrites. As these actions are independent, they can be combined with additive or synergistic effect, giving rise to a clear and strong appearance of positive-feedback excitation, and with that, clustered firing, including occasional intense bursts of activity.

In oxytocin cells, priming of dendritic release switches the behavior of oxytocin cells from being a population of autonomous neurons whose individual activity is governed by their synaptic inputs independently of their neighbours, to being a loosely-coupled population in which synchronous bursts of activity erupt through mutual excitatory excitation, while at the same time, external influences on activity are suppressed. This capacity for functional re-wiring of neuronal networks provides a possible explanation for how peptides can initiate long-lasting, coherent behavioral responses.

To summarise: neurons exhibit a wide diversity of electrophysiological phenotypes that have important consequences for how they process information. Neurons have several modes of communication with other neurons, as well as fast transmitter-mediated interactions via conventional synapses, neurons can release peptides and other substances from dendrites that have a neurohormonal-like action over a wide target area, the specific targets being defined by their expression of specific receptors. Neurohormonal actions have limited target specificity compared to neurotransmitters, where the target is restricted to the particular synapses. Neurohormonal actions also have limited temporal specificity, requiring integrated release over extended periods of time. The actions of neurohormones, however, can include organisational influences on networks, changing the strength of interactions by priming releasable reserves. While we have shown this for the releasable pools of oxytocin in dendrites, similar mechanisms may apply generally, perhaps including priming of release from synapses.

To model such a functional architecture we need concise models of individual neurons that accurately encapsulate their electrophysiological phenotype. For a model, the phenotype of a neuron may express not merely its intrinsic properties but also those properties conferred upon its behavior that are intrinsic to the network; models must be concerned with activity-dependent influences on a cell's behavior, but need not be concerned whether those influences result from a cell's intrinsic membrane properties or from recurrent network connections. However, the electrophysiological phenotype may be functionally plastic under the influence of certain types of signal, including in particular neurohormonal signals.

This chapter has discussed in particular examples of hypothalamic neurons. We have argued that the diversity of neuronal phenotypes is great, and we need models to understand the functional implications of the differences. We have argued that for this we do not necessarily need complex biophysical models, but we do need quantitatively accurate, computationally concise representations of the electrophysiological phenotypes. It may be questioned whether there is not some difference between the hypothalamus and other regions of the brain, such as the neocortex. Of course there are differences. Most importantly, we know some of the things that the hypothalamus does; sometimes in great detail, including why, and what we don't know is generally amenable to hypothesis and testing. We also know that what the hypothalamus does is important. But we need strong conceptual frameworks, with predictive power, to build our understanding.

References

[1] McKinley M.J., Bicknell R.J., Hards D., McAllen R.M., and Vivas L. (1992) Efferent neural pathways of the lamina terminalis subserving osmoregulation. *Prog. Brain Res.* **91**: 395-402.

[2] Wray S. (2001) development of luteinising hormone releasing hormone neurones. *J. Neuroendocrinol.* **13**: 3-12.

[3] Douglas A.J., Leng G., Ludwig M., and Russell J.A. (2000) (Editors) *Oxytocin and Vasopressin from molecules to function.* Special Edition of *Exp. Physiol.* (Volume 85S).

[4] Leng G. (1988) *Pulsatile Release of Hormones and Bursting Activity in Neuroendocrine Cells* CRC Press; Boca Raton, Florida. 261 pp.

[5] Leng G., and Brown D. (1997) The origins and significance of pulsatility in hormone secretion from the pituitary. *J. Neuroendocrinol.* **9**: 493-513.

[6] Leng G., Brown C.H., and Russell J.A. (1999) Physiological pathways regulating the activity of magnocellular neurosecretory cells. *Prog. Neurobiol.* **57**: 625-655.

[7] Bourque C.W., and Renaud L.P. (1990) Electrophysiology of mammalian magnocellular vasopressin and oxytocin neurosecretory neurons. *Front. Neuroendocrinol.* **11**: 183-212.

[8] Armstrong W.E. (1995) Morphological and electrophysiological classification of hypothalamic supraoptic neurons. *Prog. Neurobiol.* **47**: 291-339.

[9] Bourque C.W., Oliet S.H., and Richard D. (1994) Osmoreceptors, osmoreception, and osmoregulation. *Front. Neuroendocrinol.* **15**: 231-274.

[10] Lambert R., Dayanithi G., Moos F., and Richard P. (1994) A rise in intracellular Ca^{2+} concentration of isolated rat supraoptic cells in response to oxytocin. *J. Physiol.* **478**: 275-288.

[11] Dayanithi G., Widmer H., and Richard P. (1996) Vasopressin-induced intracellular Ca^{2+} increase in isolated rat supraoptic cells. *J. Physiol.* **490**: 713-727.

[12] Sabatier N., Richard P., and Dayanithi G. (1998) Activation of multiple intracellular transduction signals by vasopressin in vasopressin-sensitive neurones of the rat supraoptic nucleus. *J. Physiol.* **513**: 699-710.

[13] Li Z., and Hatton G. (1997) Ca^{2+} release from internal stores: role in generating depolarising after-potentials in rat supraoptic neurones. *J. Physiol.* **498**: 339-350.

[14] Li Z., Decavel C., and Hatton G. (1995) Calbindin-D28k: role in determining intrinsically generated firing patterns in rat supraoptic neurones. *J. Physiol.* **488**: 601-608.

[15] Leng G., Brown C.H., Bull P.M., Brown D., Scullion S., Currie J., Blackburn-Munro R.E., Feng J.F., Onaka T., Verbalis J.G., Russell J.A., and Ludwig M. (2001) Responses of magnocellular neurons to osmotic stimulation involves coactivation of excitatory and inhibitory input: an experimental and theoretical analysis. *J. Neurosci.* **21**: 6967-6977.

[16] Tuckwell H.C. (1988) *Introduction to Theoretical Neurobiology*, Vol. 2, Cambridge UK, Cambridge University Press.

[17] Oliet S.H., and Bourque C.W. (1993) Mechanosensitive channels transduce osmosensitivity in supraoptic neurons. *Nature* **364**: 341-343.

[18] Feng J.F., and Brown D. (1999) Coefficient of variation of interspike intervals greater than 0.5. How and when? *Biol. Cybern.* **80**:291-297.

[19] Ludwig M., Sabatier N., Bull P.M., Landgraf R., Dayanithi G., and Leng G. (2002) Intracellular calcium stores regulate activity-dependent neuropeptide release from dendrites. *Nature* **418**: 85-89.

[20] Ludwig M. (1998) Dendritic release of vasopressin and oxytocin *J. Neuroendocrinol.* **10**: 881-895.

Chapter 8

Bursting Activity in Weakly Electric Fish

Rüdiger Krahe[1] **and Fabrizio Gabbiani**[2]

[1]*Beckman Institute for Advanced Science and Technology, Department of Molecular and Integrative Physiology, University of Illinois at Urbana/Champaign, 405 N. Mathews Ave. Urbana, IL 61801, U.S.* [2]*Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.*

CONTENTS

8.1 Introduction

Neurons in many sensory systems tend to fire action potentials intermittently with spikes grouped into bursts of high-frequency discharge. Functionally, bursts have been implicated in many different phenomena, such as efficient transmission of sensory information [76], regulation of information flow during slow-wave sleep [116], selective communication between neurons [58], epileptic seizures [84], and synaptic plasticity [94]. In recent years, evidence has accumulated that bursts indeed encode sensory information and that they may even be more reliable indicators of important sensory events than spikes fired in tonic mode [47, 76, 82, 86, 99, 107, 128]. To understand the biological relevance of bursts and the cellular mechanisms underlying their generation, a wide variety of approaches are needed. *In vivo* recordings from neurons in awake/behaving animals allow investigating how different firing modes affect behavioral performance. *In vitro* experiments, on the other hand, offer a greater control over the preparation and are best suited to study cellular mechanisms of bursting. Finally, various levels of modeling can summarize experimental findings, test our understanding of mechanisms, and inspire new experiments. In this chapter, we will follow this line of investigation and review a number of recent studies of burst firing in weakly electric fish.

The electrosensory system of South American weakly electric fish has proven to be extremely well suited for combined neuroethological and computational studies of information processing from systems neuroscience to the characteristics of ion channels. In this review, we will give a brief introduction to the electrosensory system, describe in more detail the *in vivo* firing properties of electrosensory pyramidal cells in the hindbrain of these fish, and report on the potential behavioral role of bursts. Next, we present results of *in vitro* studies that have elucidated some of the cellular mechanisms underlying burst generation in pyramidal cells. This is followed by a discussion of detailed compartmental models that successfully reproduce *in vitro* bursting and reduced models offering a dynamical systems perspective on burst mechanisms. We conclude by comparing burst firing in weakly electric fish to other systems.

8.1.1 What is a burst?

Spike bursts have been described in a large number of systems. Voltage traces from a selection of bursting neurons are displayed in Figures 8.1 and 8.5. As is evident from these examples, bursts can occur on a wide range of time scales and vary in their fine temporal structure. Because the biophysical mechanisms underlying bursts can be so diverse, it comes as no surprise that no unique definition of bursts exists. We will use the term here for the basic event that is part of every burst definition: a burst is a series of action potentials fired in rapid succession, set off in frequency against the rest of a spike train. In an interspike interval (ISI) histogram, burst spikes will typically fall into one peak at short intervals with the rest of the intervals forming either a shoulder to this peak, a low plateau or a second, smaller peak at larger values (Figure 8.2a) [40, 86]. This very general definition has been used in many systems to classify spike sequences as belonging to bursts or not. However, other criteria can be applied as well, as illustrated in Figure 8.2b-d (see, e.g., [13, 28, 37, 49, 51, 60, 78, 95, 127, 131]). The specific choice of criterion will largely depend on the properties of the system under study.

8.1.2 Why bursts?

We can now ask in more detail why some nerve cells generate bursts. The answer may not be the same for every cell type, and there may even be different uses for burst firing within the same neuron under different behavioral conditions. At a mechanistic level, evidence has been accumulating that the reliability of synaptic transmission can be significantly enhanced for spikes arriving in rapid succession at the presynaptic terminal [76, 113, 120, 128, 129, 132]. The physiological consequence of increased transmission probability for burst spikes is noise filtering, where isolated presynaptic spikes can be conceived as noise and bursts as signal [41, 76]. Under this scheme, burst firing can improve the reliability of information transmission across synapses. A recent alternative and complementary proposal states that bursts may be a means of selective communication between neurons [58]. If postsynaptic neurons display membrane oscillations with cell-specific frequencies, the interspike intervals within a given presynaptic burst may determine which of the postsynaptic cells will be induced to spike.

But what is it that is signaled by bursts? In the case of relay cells of the lateral geniculate nucleus, it has been shown that bursts as well as spikes generated in tonic mode encode visual information [99]. A current hypothesis states that bursts may signal the detection of objects to the cortex while tonic firing may serve in the encoding of object details [46, 99, 107]. Another possibility is heightened selectivity of burst spikes compared to isolated spikes as observed in cells in primary auditory cortex that show sharpened frequency tuning for bursts [35]. In Section 8.3 of this chapter we will review recent work on weakly electric fish showing that spike bursts of pyramidal cells at an early stage of electrosensory processing extract behaviorally relevant stimulus features more reliably than isolated spikes.

8.2 Overview of the electrosensory system

Electrosensation may seem exotic to us, but it forms an essential part of the sensory world for a number of animal taxa. It allows them to navigate, detect approaching predators and prey, and to communicate (for recent reviews see [10, 124]). Furthermore, some of its properties make for interesting comparisons with other, less *exotic*, sensory systems: Similar to the auditory system, the electrosensory system

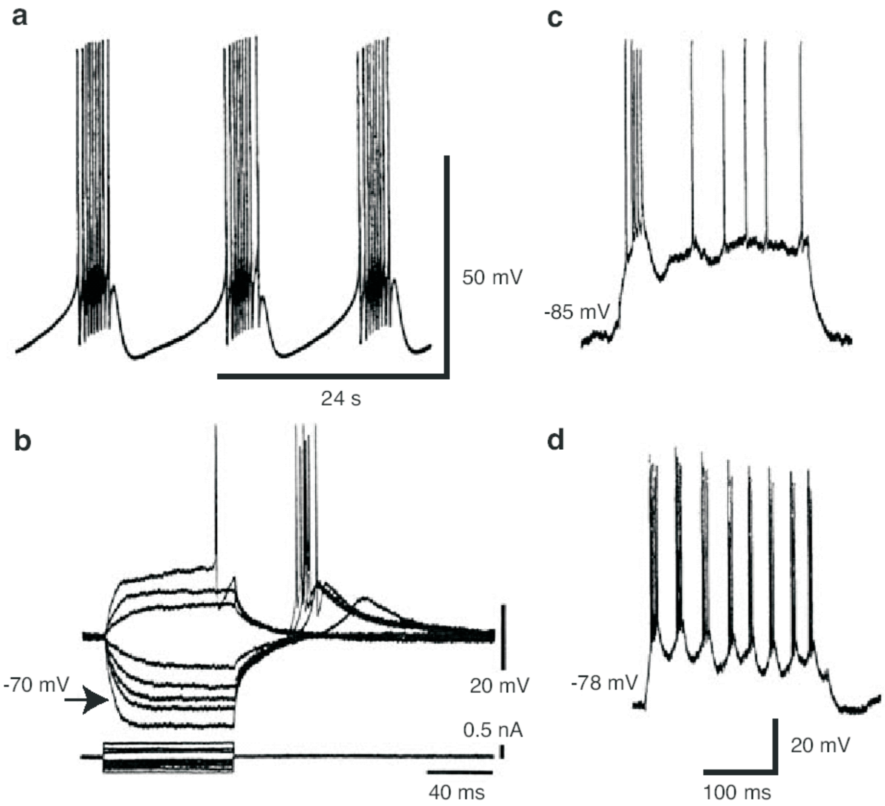

Figure 8.1

Examples of spike bursts observed in various preparations. a) The R15 neuron of *Aplysia* generates slow membrane potential oscillations on which bursts of action potentials ride (adapted from [6]). b) Rebound bursts in response to hyperpolarizing current pulses from a depolarizing holding potential in thalamic relay neurons (arrow indicates resting potential; adapted from [59]). c), d) Two types of bursting behavior in cortical neurons of the cat (called intrinsically bursting and chattering cells, respectively; adapted from [43]).

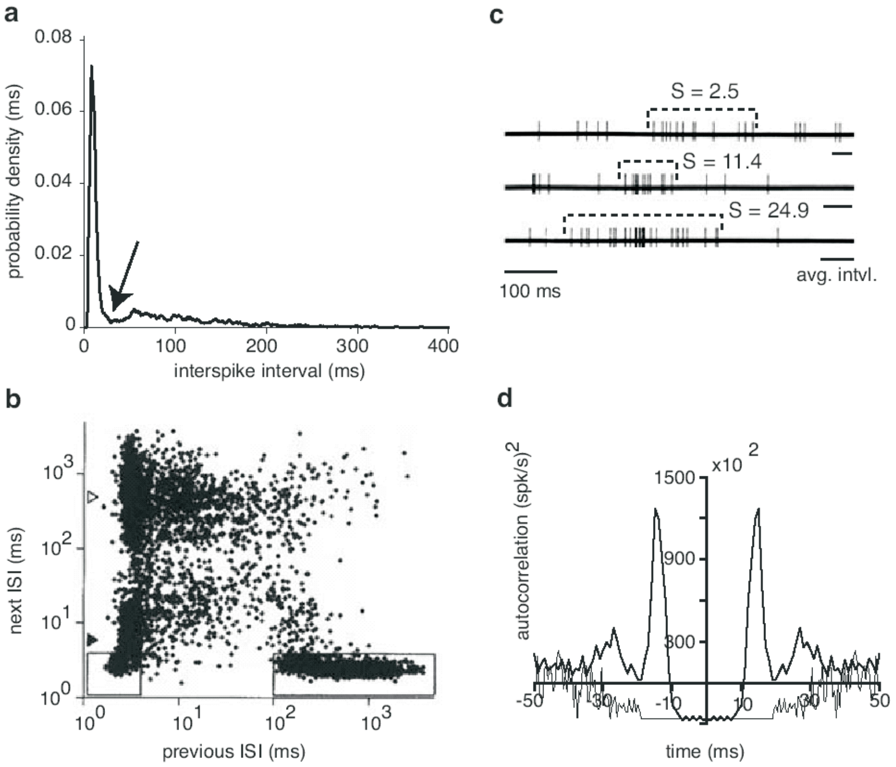

Figure 8.2

Examples of criteria used to assign spikes to bursts. a) A dip in the ISI histogram separates burst interspike intervals from longer interburst intervals (arrow; same cell as in Figure 8.5). b) Joint ISI plots clearly identify initial spikes of a burst (right rectangle) from intraburst spikes (left square; adapted from [99]). c) Bursts may be defined by computing a surprise factor that measures their deviations from the expected patterns of spontaneous independent spikes (adapted from [73]). d) Spike train autocorrelation functions of bursting neurons sometimes show clear peaks that are eliminated by treating bursts as single events (gray line; adapted from [86]). A similar definition has used the power spectrum of spike trains (Fourier transform of the autocorrelation function; see [4]).

is specialized in processing fast variations in stimulus amplitude and phase. It is quite fascinating that electrosensory processing in fish and auditory processing in barn owls and bats have evolved similar computational algorithms for time coding (e.g., [21, 66, 67]). The multiple two-dimensional topographical representations of the sensory surface (electroreceptors in the skin of the fish) within the brain are found similarly in the visual system where there are multiple topographical representations of the retina [62]. Additionally, the principal electrosensory neurons in the hindbrain come as ON- and OFF-types, have center-surround receptive fields, and as in the case of mammalian thalamic neurons (e.g., [29, 107, 111]), their responses are shaped by descending feedback.

8.2.1 Behavioral significance of electrosensation

Electroreception comes in two types, passive and active. The passive sense takes advantage of the electric fields generated by living organisms or, as has been shown in sharks, the electromagnetic field of the earth (e.g., [63]). Unlike passive electrosensation and most other sensory modalities, active electrosensation relies on signals originating from the animal itself. The fish generates an electric field through discharge of an electric organ extending along most of the caudal part of its body (Figure 8.3). The Gymnotiformes are one of two groups of teleosts that independently evolved active electrosensing [88]. Fish of the two Gymnotiform genera treated here, *Eigenmannia* and *Apteronotus*, produce a quasi-sinusoidal electric organ discharge (EOD) waveform with frequencies between 200 and 1200 Hz, the exact range being species-specific.

Objects or animals with impedance different from that of water perturb the electric field surrounding a fish. Electroreceptors in the skin monitor these distortions and thus provide information about obstacles, approaching predators, or prey (Figure 8.3; [2, 89, 90]). Nearby conspecifics also engage in electric communication, for example in the context of courtship [48, 55, 87]. Thus, the active electrosense allows weakly electric fish to forage and to communicate under conditions when other senses are more or less useless as is the case in their natural habitat: They are nocturnal animals and live in turbid tropical freshwaters, which strongly limits the usefulness of vision. Similar to echolocation in bats, active electrosensation opens an ecological niche that is safe from most diurnal predators. Additionally, it opens a new channel for intraspecific communication.

8.2.2 Neuroanatomy of the electrosensory system

Two sets of primary afferents transmit information on electric field perturbations from electroreceptors in the skin to the first central processing stage in the hindbrain, the electrosensory lateral line lobe (ELL). So-called T-receptor afferents fire strictly phase-locked to each cycle of the EOD, thus carrying information about phase distortions [103]. We will, however, focus on the amplitude-coding pathway that involves a different set of afferents, P-receptor afferents. These nerve fibers fire action potentials in a probabilistic fashion (thus the ”P”) depending on EOD amplitude (see

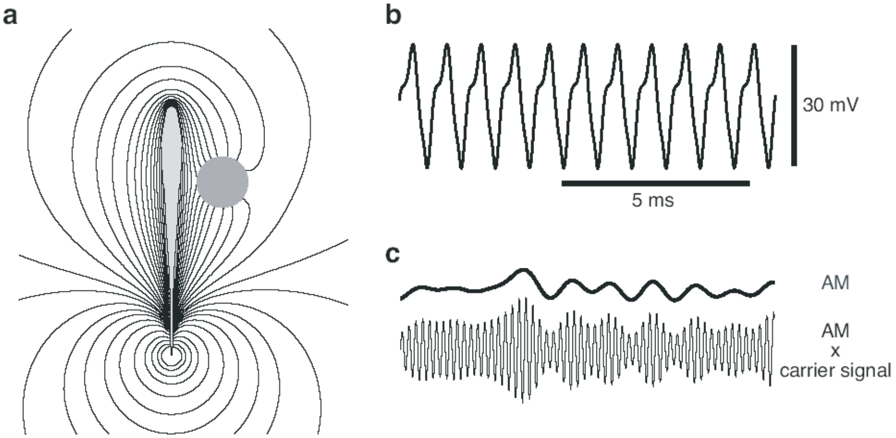

Figure 8.3

Objects in the vicinity of a weakly electric fish distort the self-generated electric field. The ensuing change in current flow across the skin - the electrosensory image of the object - is monitored by electroreceptors. a) The sketch is a snapshot of the isopotential lines of the electric field at the peak of an EOD cycle with an object of low conductivity distorting the field. b) Short section of the quasi-sinusoidal EOD waveform of *Apteronotus albifrons* recorded as the potential difference between an electrode next to the head and one close to the tail. c) Sketch illustrating the relationship between amplitude modulation waveform (AM) and the underlying EOD carrier signal.

Figures 8.3c and 8.4b). Several thousand P-receptor afferents carry information from all parts of the body to the ELL [22]. There, each individual fiber trifurcates and terminates in three adjoining somatotopic representations of the fish's skin, the centromedial (CMS), centrolateral (CLS), and lateral (LS) segments of the ELL [53] (Figure 8.4a).

P-receptor afferents directly synapse onto one set of principal output cells of the ELL, the basilar pyramidal cells or E-units (*E-xcited*; Figure 8.4). The other set of output neurons, the non-basilar pyramidal cells, or I-units ('I-inhibited'), receives indirect feedforward input from the afferents via inhibitory interneurons [81]. Consequently, E-units fire action potentials in response to increases in electric field amplitude, whereas I-units fire in response to decreases [102] (Figure 8.4b). The spatial receptive fields of pyramidal cells are more complex than their direct connections with primary afferents would suggest: E-units have an excitatory center and an inhibitory surround and vice versa for I-units. This is analogous to ON- and OFF-cells in the visual system [7, 11, 102, 109]. A prominent feature of both types of pyramidal cells is their extensive apical dendrites that extend far into the molecular layer of the ELL (Figure 8.4a). Here, pyramidal cells receive proprioceptive input and massive feedback from higher centers of electrosensory processing. Descending control via the apical dendrites has been shown to play a role in oscillatory responses of pyramidal cells, in gain control, in shaping receptive field size, in adaptive filtering of predictable sensory patterns, and may also be involved in a sensory searchlight mechanism [9, 14, 16, 31].

8.2.3 Electrophysiology and encoding of amplitude modulations

Behaviorally relevant amplitude modulations of the electric field induced by objects, prey, or conspecifics cover a frequency range of up to 80 Hz [8]. With their tonic response properties and firing rates in the range from 50 to 600 spikes per second, P-receptor afferents appear well suited to encode these amplitude modulations by changes in instantaneous firing rate [8, 91, 98, 133] (see Figure 8.4b). This was confirmed in studies applying linear stimulus-estimation algorithms to the responses of P-receptor afferents to stochastic modulations of electric field amplitude [23, 40, 70, 86, 130]. Up to 80% of the stimulus time course can be recovered from single primary afferent spike trains. Therefore, it seems that, prior to entering the hindbrain, electrosensory information is faithfully encoded and undergoes very little processing.

What kind of processing takes place in the ELL? One hypothesis could be that single pyramidal cells perform even better at transmitting detailed information on the stimulus time course than P-receptor afferents by averaging out noise over 5 to 20 primary afferents converging onto them [7, 110]. This does not seem to be the case when amplitude modulations are presented over large areas of the body surface, mimicking communication signals. Linear stimulus estimation from pyramidal cell spike trains in *Eigenmannia* yielded poor results compared to primary afferents [40, 68, 86]. Since neighboring pyramidal cells receive input from overlapping areas of the fish's skin, it is conceivable that the information is conveyed in a distributed manner. However, even when stimulus estimation was based on pairs of spike trains from simultaneously recorded pyramidal cells with overlapping receptive fields, the fraction of the stimulus recovered was still well below the fraction encoded by single primary afferents [68]. Recent studies in the CLS and LS of the related weakly electric fish *Apteronotus leptorhynchus*, however, indicate that pyramidal cells may not act as a homogeneous population in this respect. Bastian and coworkers found that the efficiency for encoding global amplitude modulations scales with the spontaneous firing rate of pyramidal cells (3-50 Hz) [11]. Furthermore, the spatial extent of the stimulus seems to affect how much information a cell can transmit about the amplitude modulations [11]. Thus, it seems possible that a subset of pyramidal cells is able to transmit information on the electric stimulus time course, and that the spatial extent of stimuli affects the response properties, probably via feedback input to the apical dendrites [31]. However, even the best performing cells observed so far do not improve on the performance of P-receptor afferents [11, 39, 40, 86, 130].

In summary, compared to the primary afferents, pyramidal cells of the ELL are poor encoders of the stimulus time-course. Hence, the question remains, what kind of information do most ELL pyramidal cells transmit to the next stage of electrosensory processing?

8.3 Feature extraction by spike bursts

8.3.1 Bursts reliably indicate relevant stimulus features

Despite their generally poor performance at encoding the time course of amplitude modulations, inspection of pyramidal cell spike trains shows that their responses are selective (Figure 8.5). E-units typically fire isolated spikes or short spike bursts in response to upstrokes in stimulus amplitude whereas I-units fire in response to downstrokes. Bursts consist of 2 to 10 spikes with a mean of about 3 spikes per burst. On average, roughly 60% of the spikes fired by a given cell occur in bursts [40]. Spatially extended upstrokes and downstrokes in amplitude are known to be integral parts of the electrosensory input eliciting the so-called *Jamming Avoidance Response* (JAR [52]). In case of the JAR, the signals of two conspecifics interfere, creating a beat pattern extending over a large part of the body. To avoid low-frequency beats, which affect the fish's ability to electrolocate, nearby animals can actively increase the difference between their EOD frequencies. Localized upward and downward deflections in EOD amplitude moving across the sensory surface, on the other hand, may signal the presence of prey [89]. Thus, global as well as local up- and downstrokes in amplitude are presumably important electrosensory events. It therefore seems plausible that pyramidal cells could signal the occurrence of these temporal stimulus features without transmitting detailed information on the stimulus time course. Various methods are available to quantify neuronal classification performance, for example neural network models that learn the optimal stimulus pattern eliciting spikes (e.g., [74]). A more direct approach derived from signal detection theory uses a linear op-

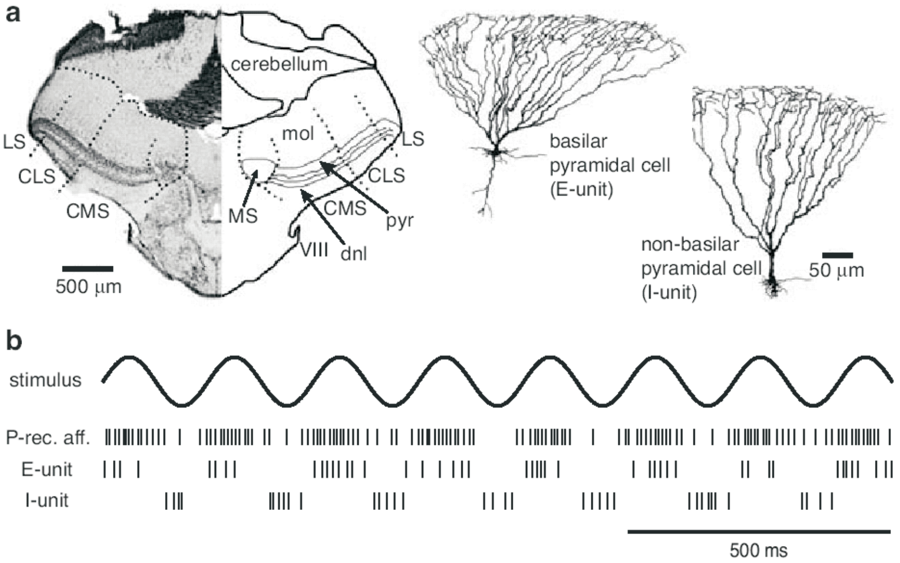

Figure 8.4

Processing of amplitude modulations of the electric field by P-receptor afferents and pyramidal cells in the ELL. a) P-receptor afferents enter the hindbrain via the octavolateral nerve (VIII) and trifurcate to form three somatotopic maps of the body surface (LS: lateral segment; CLS: centrolateral segment; CMS: centromedial segment). A fourth map (medial segment, MS) is formed by passive electrosensory input, which is not treated here. The cross-section through the hindbrain of Eigenmannia shows the layered organization of the ELL maps with the deep neuropil layer (dnl) containing the primary afferent fibers, and the somata of the pyramidal cells forming a distinct dark layer (pyr). Basilar pyramidal cells receive direct input from P-receptor afferents onto their basilar dendrite, while non-basilar pyramidal cells receive indirect inhibitory input via interneurons. In the molecular layer (mol) descending inputs connect onto the apical dendrites of pyramidal cells (adapted from [86]). b) Raster plots of spike trains of P-receptor afferents, E- and I-cells in response to sinusoidal amplitude modulations (top trace).

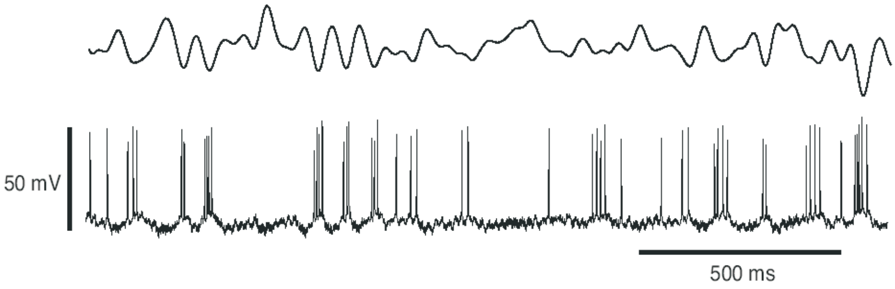

Figure 8.5

Pyramidal cells tend to fire spikes in short bursts. Intracellular recording of an I-type pyramidal cell in the CMS stimulated with random amplitude modulations (top trace). Note the coupling of spike bursts and isolated spikes to downstrokes in amplitude (adapted from [86]).

eration on the input signal followed by a threshold computation. Thus, the specific issue of interest is whether burst spikes perform better at extracting stimulus features than spikes occurring in isolation.

8.3.2 Feature extraction analysis

To quantify how well a spike train discriminates stimulus patterns, one first needs to estimate the optimal stimulus feature for eliciting spikes. Here, we describe the application of a Euclidian pattern classifier to this problem (see [40, 86] for a slightly more general method). First, the spike train, $x(t)$, and the stimulus, $s(t)$, are binned so as to allow a maximum of one spike per bin. A variable r_t is defined to take the value 1 if the time bin ending at t contains a spike and the value 0 if it does not contain a spike. Stimulus segments, $\mathbf{s}_t$, ending at time t and comprising $\sim$ 100 bins prior to time t are assigned to one of two ensembles, $P(\mathbf{s}|r=1)$ and $P(\mathbf{s}|r=0)$, depending on whether $\mathbf{s}_t$ preceded a time bin containing a spike or not (i.e., $r_t = 0$ or 1; Figure 8.6). The feature, $\mathbf{f}$, is computed from the means, $\mathbf{m}_1$ and $\mathbf{m}_0$, of these conditional distributions $P(\mathbf{s}|r=1)$ and $P(\mathbf{s}|r=0)$: $\mathbf{f} = \mathbf{m}_1 - \mathbf{m}_0$. For E-units, the typical feature is a strong upstroke in stimulus amplitude preceded by a small downstroke (Figure 8.6 bottom), for I-units it is a strong downstroke preceded by a small upstroke (Figure 8.7a) [40, 86]. It is important to note, however, that the exact shape of the classifier depends not only on the individual cell studied but also on the bandwidth of the stimulus [86]. Typically, only the time bins between 0 ms (spike occurrence) and -300 ms show significant deviations from an amplitude of 0 mV suggesting that pyramidal cells do not integrate over longer time spans of sensory input.

To assess the separation between the two ensembles of stimulus segments, each segment is projected onto the feature vector, $\mathbf{f}$, and compared to a threshold value, θ: $h_{\mathbf{f},\theta}(\mathbf{s}) = \langle \mathbf{f}; \mathbf{s} \rangle - \theta$, where $\langle ; \rangle$ denotes the scalar product. The projection, $h_{\mathbf{f},\theta}(\mathbf{s})$,

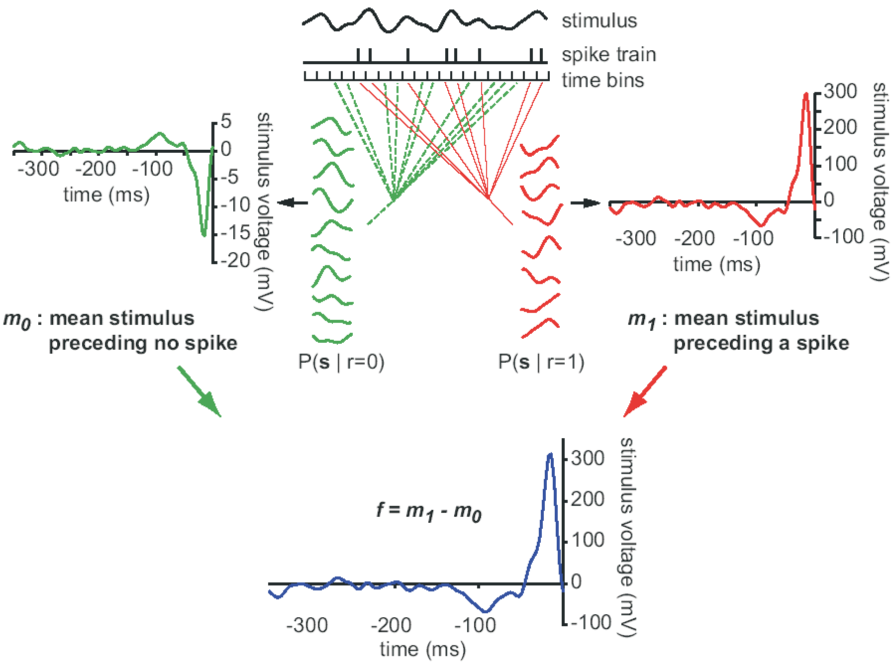

Figure 8.6

Computation of the Euclidian pattern classifier. For each time bin of a given spike train the stimulus vector preceding this bin is assigned to one of two ensembles ($P(\mathbf{s}|r=0)$ and $P(\mathbf{s}|r=1)$) depending on whether the time bin contains a spike or not. The Euclidian classifier is defined as the mean stimulus preceding spikes ($\mathbf{m}_1$, right) minus the mean stimulus preceding time bins without a spike ($\mathbf{m}_0$, left): $\mathbf{f} = \mathbf{m}_1 - \mathbf{m}_0$. For this E-unit, the feature is a strong upstroke in amplitude, peaks at around -25 ms, and then returns to 0 mV. Bandwidth of the stimulus: 0 to 44 Hz. Adapted from [86].

can be conceived of as a measure of similarity between a stimulus segment and the feature vector.

The performance of this Euclidian classifier in predicting the occurrence of spikes is quantified using a Receiver Operating Characteristic (ROC) analysis [38, 40, 44, 86]. First, the conditional probability distributions of the projections,

$$P(h_{\mathbf{f},\theta}(\mathbf{s})|r=1)$$

and

$$P(h_{\mathbf{f},\theta}(\mathbf{s})|r=0),$$

are plotted and compared to threshold, θ (Figure 8.7b). A spike is detected if $h_{\mathbf{f},\theta}(\mathbf{s}) > 0$, that is if $\langle \mathbf{f};\mathbf{s} \rangle$ is larger than the threshold (to the right of the dashed vertical line in Figure 8.7b). Integrating the tail of the distribution $P(h_{\mathbf{f},\theta}(\mathbf{s})|r=1)$

to the right of the threshold, θ, yields the probability of correct detection, P_D. The right tail of the distribution $P(h_{\mathbf{f},\theta}(\mathbf{s})|r=0)$ corresponds to the probability of false alarms, P_{FA}. By varying the threshold value, θ, P_D can be determined as a function of P_{FA}. The resulting curves are called ROC curves (Figure 8.7c). The larger the area under a given curve the better is the detection performance. However, false alarms are not the only kind of error that can occur. The second type of error happens when a spike is missed because the corresponding projection value is below threshold ($P(h_{\mathbf{f},\theta}(\mathbf{s})|r=1)$ to the left of θ). Therefore, a measure of the misclassification error has to incorporate both, the probability of false alarms and the probability of missed events: $P_E = 0.5[P_{FA} + (1 - P_D)]$. The best classification performance of an ideal observer corresponds to the minimum of the plot of P_E versus P_{FA} (Figure 8.7d).

As is evident from Figure 8.7b, the probability distribution of stimulus projections for burst spikes is more clearly separated from the distribution of stimuli preceding a spikeless bin than is the one for isolated or all spikes. Consequently, the ROC curve for burst spikes rises more steeply than the one for isolated spikes and all spikes (Figure 8.7c) yielding the lowest misclassification errors (Figure 8.7d). The superior feature extraction performance of burst spikes was typical for all cells studied so far in the CMS and LS of the weakly electric fish, *Eigenmannia* (overall 133 pyramidal cells [40, 68, 69, 86]).

When the same analysis was applied to spike trains of primary afferents, they consistently performed worse than pyramidal cells (Figure 8.7e) [86] suggesting that information is filtered in different ways at the first two stages of electrosensory processing. Feature extraction analysis also revealed differences in performance between cells recorded in different maps of the ELL. Cells from the CMS displayed lower misclassification errors than cells from the LS (Figure 8.7f) [86]. This finding correlates well with the different behavioral significance attributed to the two maps. The CMS has been shown by lesion experiments [85] to be necessary and sufficient for JAR behavior, which is known to involve the correlation of up- and downstrokes in stimulus amplitude with advances or delays in EOD phase [52]. The LS, on the other hand, was shown to be necessary and sufficient for the processing of electrocommunicatory signals [85], which may involve a more complex analysis of the electrosensory input.

Recently, the analysis of electrosensory information transmission was extended to simultaneously recorded spike trains of pairs of pyramidal cells with overlapping receptive fields [68]. Cross-correlation analysis showed that correlations in spike timing between cells of the same type (two E-units or two I-units) were broad (tens of milliseconds) and were not caused by shared synaptic input, but were induced by the independent coupling of both cells to the stimulus. Feature extraction analysis demonstrated that spikes of two nearby cells occurring within a coincidence time window of 5 to 10 ms significantly improved the reliability of feature extraction compared to burst spikes of the individual neurons (Figure 8.8b,c). Interestingly, a large fraction of the coincident spikes occurred in bursts (for a coincidence time window of 5 ms, 63±15%, mean ± standard deviation; see Figure 8.8a). This finding supports the thesis that coincident bursts of spikes may constitute the most reliable neural code [76]. The similar time scales of the typical intraburst interspike interval

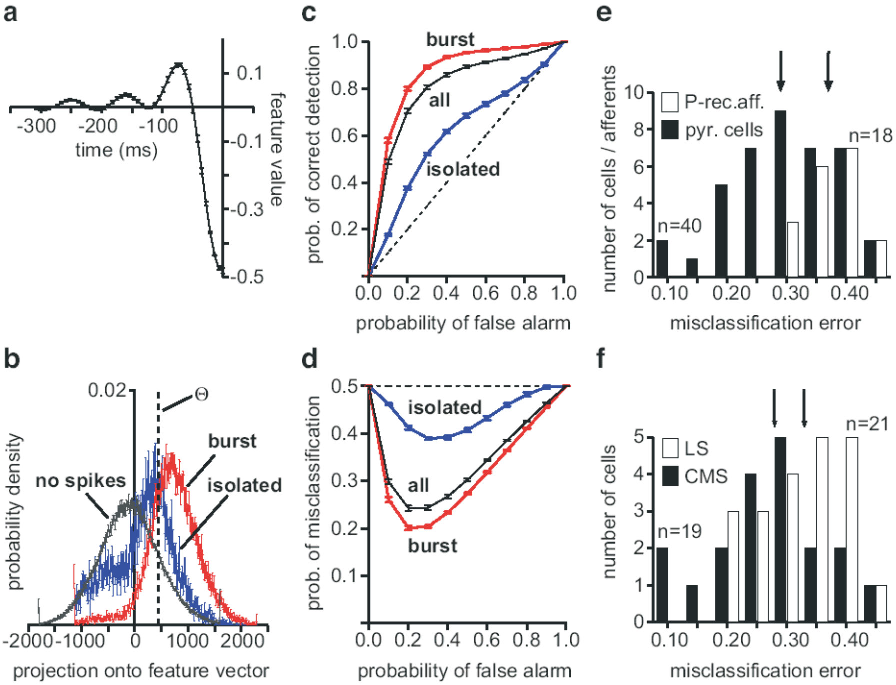

Figure 8.7

ROC analysis of feature extraction performance. a) A representative optimal stimulus feature of an I-unit. Bandwidth of the stimulus: 0-12 Hz. b) Probability density distributions of the projections of stimulus segments preceding time bins containing a spike and of stimulus segments preceding time bins without a spike (black curve). Spikes were assigned to two classes, isolated spikes (blue) and burst spikes (red), based on the ISI histogram (Figure 8.2). The probabilities of correct detection and of false alarms are computed by integrating the tails of the probability distributions to the right of threshold, θ (dashed vertical line): $P_D = P(\langle \mathbf{f}; \mathbf{s}_t \rangle > \theta | r_t = 1)$, $P_{FA} = P(\langle \mathbf{f}; \mathbf{s}_t \rangle > \theta | r_t = 0)$. c) ROC curves obtained by varying the threshold θ along the abscissa in b. The dashed line indicates chance performance. d) Probability of misclassification, P_E, versus probability of false alarm. The best performance of the Euclidian classifier can be read from the minimum of this plot. e) Comparison of feature extraction performance by P-receptor afferents (white bars) and pyramidal cells (black bars). The arrows indicate the respective median values of the two distributions. f) Distributions of the misclassification errors for pyramidal cells from the CMS (black bars) and LS (white bars). (a)-(d) adapted from [39]. (e) and (f) adapted from [86].

(10-15 ms) and of the best coincidence time window (5-10 ms) suggest that, from the viewpoint of the postsynaptic target, coincident spikes may be considered as *distributed bursts* (see also [128]).

ROC analysis has been applied before to compare signal detection performance by burst and tonic response modes of relay cells in the lateral geniculate nucleus of anesthetized cats [46]. Cells were found to indicate visual stimuli more reliably when firing in burst mode than when in tonic mode. While the role of burst firing in the thalamus remains debated [108, 115], evidence is mounting that bursts in thalamic relay cells do occur in the awake animal and that they convey stimulus-related information (reviewed in [107], see also [120, 121, 131]). It seems, however, that bursts are much less prevalent in thalamic relay cells of awake mammals than they are in pyramidal cells of awake weakly electric fish. Thalamic bursts often appear to be transient responses to the beginning of sensory events, which are then followed by tonic encoding of stimulus details [47, 106, 107]. In contrast, bursts in electric fish pyramidal cell do not abate over the course of a long stimulus but seem to be the major signaling mode employed by those cells. The feature extraction analysis developed by Gabbiani et al. [40] moves beyond the method employed by Guido et al. [46] by yielding information on the optimal feature driving a given cell and on how reliably the occurrence of this feature is indicated by different subsets of spikes in a spike train.

In conclusion, it appears that, at least for global modulations of stimulus amplitude as used in the studies of weakly electric fish described above, electrosensory information transmission undergoes a dramatic transformation at the earliest stages of processing. The primary afferents reliably encode the stimulus time course by their instantaneous firing rate. At the first central nervous stage of electrosensory processing pyramidal cells extract behaviorally relevant features from the persistent stream of afferent input and indicate their times of occurrence to higher-order nuclei by firing short bursts of spikes and by stimulus-induced coincident activity of groups of cells.

8.4 Factors shaping burst firing *in vivo*

As described for other systems [24, 26, 80, 83], the propensity of ELL pyramidal cells to burst is related to their morphology and seems to be under descending control from higher centers of sensory processing. Bastian and coworkers studied spontaneous burst firing by pyramidal cells of the CLS and LS in *Apteronotus leptorhynchus* [12, 13]. Spontaneous firing rate of these neurons is negatively correlated with the size of their apical dendrite, whereas the probability to generate spontaneous spike bursts increases with the length of the dendritic arbor. The largest apical dendrites reach high up into the molecular layer of the ELL (Figure 8.2a) [12, 13]. There, the apical dendrites are contacted by parallel fibers originating from the pos-

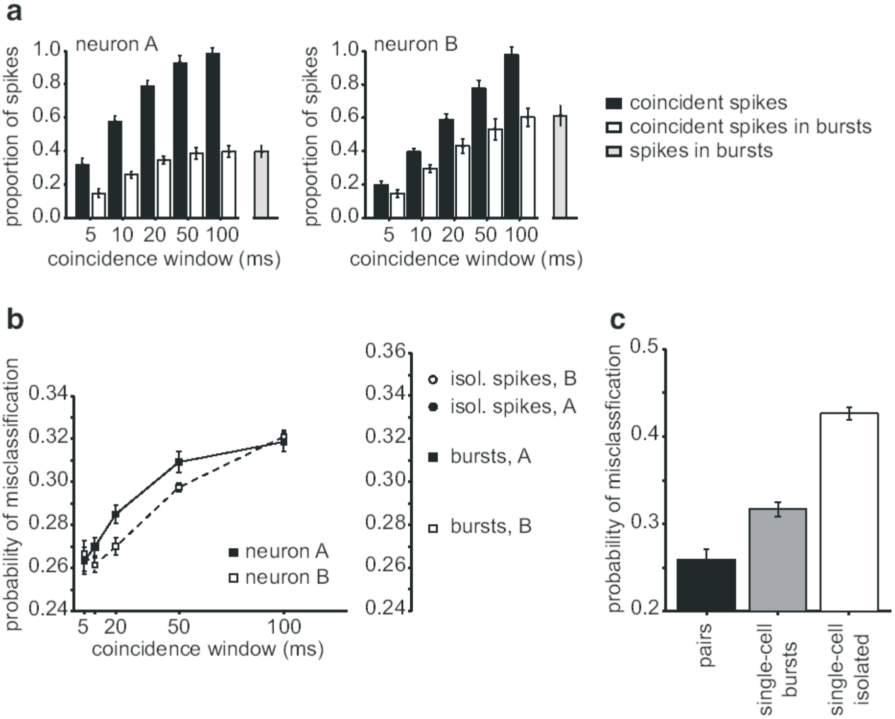

Figure 8.8

Feature extraction by distributed bursts. a) Fraction of coincident spikes of two simultaneously recorded I-units from CMS with overlapping receptive fields. Black bars: proportion of spikes of neuron A (left) and B (right) coinciding with spikes on the respective other neuron within the time window displayed on the abscissa. White bars: proportion of coincident spikes that occurred in bursts. Grey bars: overall percentage of spikes that occurred in bursts. b) Left: Minimum probability of misclassification by coincident spikes of neurons A and B as a function of the size of the coincidence time window. Spikes coinciding within a time window of 5-10 ms performed significantly better at feature extraction than did isolated or even burst spikes of the individual neurons (right). c) Summary diagram of feature extraction performance by coincident spikes of pairs of pyramidal cells of the same type (E-E pairs and I-I pairs pooled; n=16), by burst spikes of individual cells, and by isolated spikes of single cells (n=58). Adapted from [68].

terior eminentia granularis of the cerebellum [9, 16]. These parallel fibers control the spontaneous firing rate of pyramidal neurons as well as their probability to produce spontaneous bursts [13]. They are part of an indirect electrosensory feedback pathway, which is thought to be involved in gain control [16]. Therefore, it is conceivable that this indirect feedback could switch pyramidal cell responses between a bursting and a tonic mode. Firing in burst mode would improve feature extraction performance, whereas in tonic mode pyramidal cells might function as encoders of stimulus time course [13]. Switching between tonic and burst mode, however, has so far not been demonstrated for stimulus-driven pyramidal cell responses. Recent evidence suggests that not only indirect feedback to the dorsal molecular layer but also direct inhibitory feedback to the proximal apical dendrites of pyramidal cells affects their firing patterns [31]. This inhibitory direct feedback pathway supports an oscillatory component of burst responses. It is spatially diffuse and is strongest when amplitude modulations occur over large areas of the body surface as they do when fish engage in electrocommunication. For localized, prey-like stimuli, however, the inhibition is only weak and does not support oscillatory burst responses.

8.5 Conditional action potential backpropagation controls burst firing *in vitro*

Slice preparations of the ELL of *Apteronotus leptorhynchus* have proven enormously fruitful in elucidating cellular and network mechanisms of electrosensory processing (reviewed in [16, 123]). The laminar organization of the ELL allows for accurate placement of recording and stimulating electrodes in various layers along the pyramidal cell axis (see Figure 8.4a). Deprived of the natural barrage of primary sensory and feedback inputs, and only stimulated by intracellular constant current injection, pyramidal cells *in vitro* display rhythmic oscillations of the membrane potential, which periodically trigger series of high-frequency spike bursts (30 to over 300 Hz) [123]. The frequency characteristics of this oscillatory burst discharge (burst frequency and intra-burst spike frequency) vary across the three topographic maps of the ELL roughly correlating with pyramidal cell tuning properties observed *in vivo* [109, 124, 126].

8.5.1 Experimental evidence for conditional backpropagation

It was shown early on [125] that active backpropagation of Na^+ spikes into the apical dendrite is an integral part of high-frequency burst generation by pyramidal cells, similar to what has been described for several other systems [50]. Spikes are initiated at the soma or axon hillock and travel back into the apical dendrite up to the first major branch points ($\sim$200 μm). Membrane depolarization and repolarization in the dendrite are slower than in the soma and therefore dendritic spikes are longer in duration than somatic ones. A fast afterhyperpolarization (AHP) of the somatic membrane increases the potential difference between the soma and the still depolarized dendrite and leads to a sizable amount of current being sourced back into the soma where it supports a depolarizing afterpotential (DAP; Figure 8.9). In the course of a burst, somatic DAP amplitude is potentiated because of frequency-dependent broadening of dendritic spikes. Consecutive DAPs sum up and cause the frequency of somatic spike generation to increase. Eventually, the DAP itself will reach threshold for spike initiation and a high-frequency somatic spike doublet will be generated (ISI typically < 6 ms). Since the refractory period of the apical dendrite is longer ($\sim$4.5 ms) than that of the soma ($\sim$ 2 ms), the dendrite does not support active backpropagation of the second spike of the doublet, and the corresponding DAP at the soma fails allowing the AHP to terminate the burst (Figure 8.9b). This mechanism of burst generation and termination has been termed *conditional backpropagation* [75], because backpropagation is essential for burst production, and it is conditional on sufficiently low spike frequencies. When spike frequency exceeds the dendritic refractory period, backpropagation fails and the burst is terminated.

A number of cellular components of the burst mechanism have been identified. Na^+ channels are distributed in a punctate manner along the proximal 200 μm of the apical dendrite consistent with the finding that active backpropagation of TTX-sensitive spikes terminates at about this distance from the soma [125]. A candidate mechanism for the broadening of dendritic spikes is cumulative inactivation of a dendritic K^+-conductance [75]. The inactivation would slow the repolarization of the dendritic membrane potential in a spike-frequency-dependent manner, thus increasing the amplitude of the somatic DAP. A likely candidate for this current is the Apteronotid homologue of the mammalian Kv3.3 K^+-channel (*Apt*Kv3.3), which is extensively distributed along the entire axis of pyramidal cells [96, 97]. Local blockade of dendritic *Apt*Kv3.3 led to slowing of spike repolarization and increase in somatic DAP with a time-course similar to that of a regular burst. This manipulation also lowered the threshold for burst discharge evoked by current injection into the soma [97]. Therefore, it seems likely that this high-voltage-activated K^+ channel is either directly involved in the mechanism of burst discharge or at the very least can modulate the threshold for burst generation [93]. Another contribution to the potentiation of the somatic DAP in the course of a burst comes from a persistent Na^+ current which is activated by the increasing dendritic spike duration [34]. In contrast to other systems (for review see [56]), Ca^{2+} currents or Ca^{2+}-dependent K^+ currents do not appear to be necessary for burst generation [34, 75, 93].

The detailed knowledge of pyramidal cell morphology, the organization of primary sensory and feedback input, and of the conductances shaping burst firing *in vitro*, makes pyramidal cells ideally suited for detailed modeling of the mechanism underlying burst firing. This mechanism differs in interesting ways from burst generation as described in several other systems. One obvious peculiarity of ELL pyramidal cell bursts is that ISI duration decreases in the course of a burst (Figure 8.9b), a phenomenon that has not been described in any other system so far. *In vivo*, however, this ISI pattern can be observed only rarely (Krahe, unpublished observations). With natural synaptic input, other factors like inhibition and the interplay between afferent and feedback input may also shape the bursts and contribute to their termination. Furthermore, the basilar dendrites of E-units warrant closer investigation since they have been shown to be equipped with Na^+ channels as well as *Apt*Kv3.3 K^+ channels, and might thus also support backpropagation and bursting in a way similar to the apical dendrite [96, 97, 125] (see also [100] for similar conclusions in neocortical pyramidal neurons).

8.5.2 Multicompartmental model of pyramidal cell bursts

Based on the detailed spatial reconstruction of a dye-filled E-type pyramidal cell [17], Doiron et al. [33, 34] developed a multicompartmental model that successfully reproduces burst firing as it is observed *in vitro* (Figure 8.9). The main goal of these studies was to identify the components of the burst mechanism that underlie dendritic spike broadening and somatic DAP potentiation since those are responsible for the progressive decrease in ISI duration and eventual burst termination. A key feature of the model was the presence of fast Na^+ and K^+ currents in both somatic and dendritic compartments, to account for Na^+ action potential generation and backpropagation (Figure 8.10a). In order to achieve the narrow somatic and broader dendritic spike shapes (see 5.5.1), the time constants of the active conductances in the dendrite had to be increased relative to the soma. This also yielded a relatively longer refractory period for the dendritic spike compared to the somatic one.

While the core model outlined above reproduced key features of the somatic and dendritic response, it failed to generate spike bursts. Doiron et al. [33] were able to exclude a number of potential burst mechanisms described for other systems: Ca^{2+}- or voltage-dependent slowly activating K^+ channels, slow inactivation of the dendritic Na^+ channel, and slow activation of the persistent Na^+ current. Finally, modification of the dendritic delayed rectifier channel yielded burst properties corresponding to the *in vitro* findings: A low-threshold slow inactivation of the K^+ conductance led to dendritic spike broadening in the course of a burst and to a corresponding increase in the DAP amplitude, which eventually triggered a doublet, leading to dendritic spike failure and burst termination due to the AHP. Whereas slow activation of the persistent Na^+ current proved insufficient to elicit proper bursting, it was recently shown to be an important component of the DAP potentiation [34]. It is activated by the broadening of dendritic spikes and boosts the sub-threshold depolarization of the somatic membrane. Thereby it largely determines the time it takes to reach threshold for doublet firing. Since the doublet terminates the burst, the persistent Na^+ current thus controls burst duration. With the interburst period being largely fixed by the duration of the AHP, the persistent Na^+ current also determines burst oscillation period [34]. Since it can be activated by descending feedback to the apical dendrites [15, 17], this provides a potential mechanism for controlling burst firing depending on behavioral context.

To summarize, the key features of the pyramidal cell burst mechanism are i) a dendritic Na^+ conductance that supports active backpropagation of spikes into the dendrite and that feeds the somatic DAP, ii) a slow cumulative inactivation of a delayed rectifier current which leads to dendritic spike broadening in the course of a

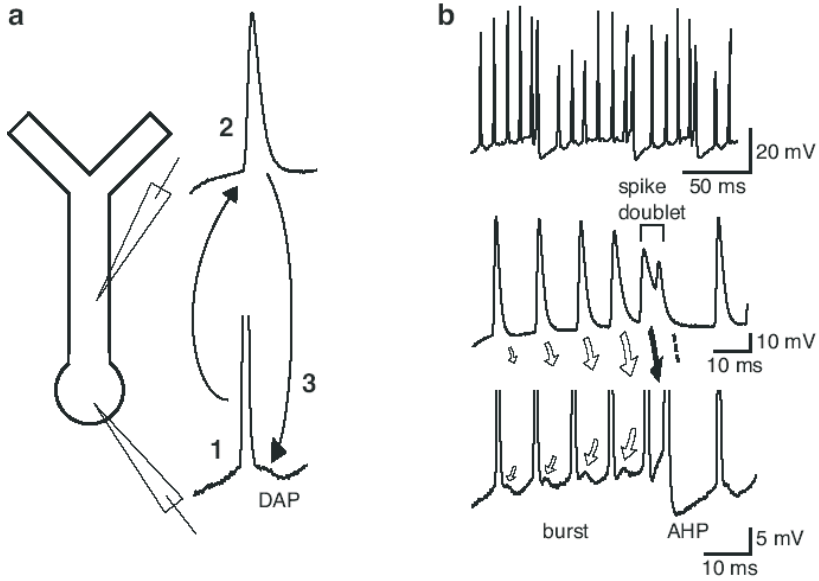

Figure 8.9

Summary of the mechanism underlying high-frequency burst generation in pyramidal cells *in vitro*. a) Schematic diagram of a pyramidal cell with a narrow spike recorded in the soma (1). The somatic spike is actively propagated back into the apical dendrite where a much broader version of the same spike can be recorded (2). Current sourcing from the dendrite back into the soma causes a DAP (3). b) Top: Oscillatory burst discharge recorded in the soma of a pyramidal cell with 0.74 nA depolarizing current injection. Middle and bottom: Somatic and dendritic spike burst recorded separately in two different cells. The time scales are adjusted to allow alignment of spikes. Somatic spikes are truncated. As evident from the dendritic recording, spike repolarization slows down in the course of a burst allowing the DAP at the soma to potentiate. Eventually, the DAP reaches threshold and causes a high-frequency spike doublet. Since the dendritic refractory period is longer than the somatic one, the dendrite cannot support active propagation of the second spike of the doublet. The DAP fails and allows the afterhyperpolarization (AHP) to terminate the spike burst. (a) adapted from [75], (b) adapted from [34].

burst, thus potentiating the somatic DAP, iii) a shorter refractory period for somatic spikes compared to dendritic ones renders backpropagation conditional on the instantaneous firing rate, iv) the rate of the DAP potentiation, which is part of a positive feedback loop in which dendritic spike broadening activates a persistent Na^+ current, which further boosts depolarization. The slow dynamics of the persistent Na^+ current largely control burst duration and burst frequency.

8.5.3 Reduced models of burst firing

Detailed biophysical models are powerful tools for probing the understanding of cellular mechanisms at a microscopic scale. However, they are computationally too complex for modeling of large networks or for analyzing the behavior of single cells from a dynamical systems perspective. Having understood the key mechanisms, it is often possible to reduce a detailed biophysical model to its essential components and then apply dynamical systems analysis to the lower-dimensional model [101]. The multi-compartmental model described above has undergone two such reductions, first to a two-compartment model, termed a *ghostburster* for reasons explained in more detail below [32], and then to an even simpler two-variable delay-differential-equation model [71].

To model the generation of the somatic DAP, only a somatic and one dendritic compartment representing the entire apical dendritic tree are needed (Figure 8.11a) [32]. Soma and dendrite were equipped with fast Na^+ channels, delayed rectifier K^+ currents, and passive leak current. Current flow between the compartments followed simple electrotonic gradients determined by the coupling coefficient between the two compartments, scaled by the ratio of somatic to total model surface (see also [64, 80, 129]). Thus, the entire system was described by only six nonlinear differential equations using modified Hodgkin/Huxley kinetics [54]. To achieve the relatively longer refractory period of the dendrite [75], the time constant of dendritic Na^+ inactivation was chosen to be longer than somatic Na^+ inactivation and somatic K^+ activation. The key element for the burst mechanism was the introduction of a slow inactivation variable for the dendritic delayed rectifier current, whose time constant was set to about 5 times slower than the mechanisms of spike generation. In this configuration, the two-compartment model reliably reproduced the potentiation of the somatic DAP, which eventually triggers the firing of a spike doublet, the burst termination due to failure of backpropagation, and the rapid onset of the AHP [32] (see Figure 8.9b).

To study the burst dynamics, the ghostburster model was treated as a fast-slow burster [57, 101], separating it into a fast subsystem representing all variables related to spike generation, and a slow subsystem representing the dendritic K^+ inactivation variable, p_d. The fast subsystem could then be investigated using the slow variable as a bifurcation parameter. The dashed lines in Figure 8.11b show the quasi-static bifurcation diagram with maximum dendritic membrane voltage as a representative state variable of the fast subsystem, and p_d as the slow subsystem. For constant values of $p_d > p_{d1}$, there exists a stable period-one solution. At $p_d = p_{d1}$ the fast subsystem transitions to a period-two limit cycle. This corresponds to intermittent doublet firing

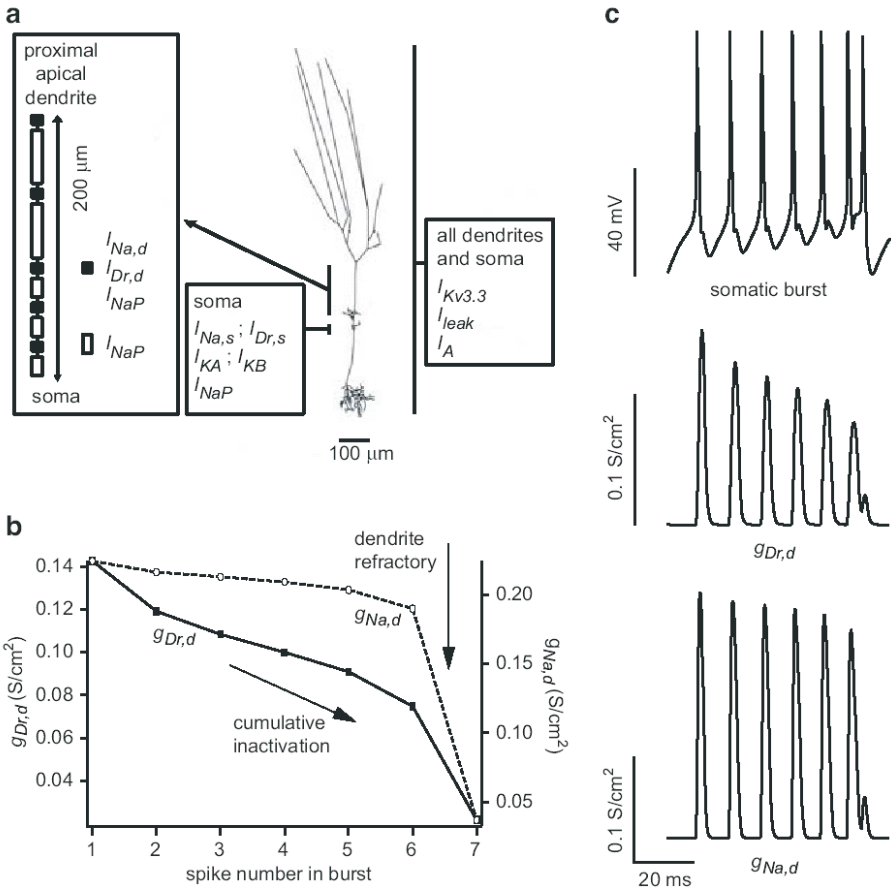

Figure 8.10

Multi-compartmental model of burst generation. a) The model was based on the detailed reconstruction of a dye-filled E-type pyramidal cell [17]. The distribution of ionic channels along the neuron's axis is indicated in the insets. The detailed placement of Na$^+$ and K$^+$ channels in separate compartments of the proximal dendrite is shown on the left. b) The model reproduces the increasing firing frequency in the course of a burst with a doublet at the end and a burst AHP (top). The dendritic delayed-rectifier conductance, $g_{Dr,d}$, shows cumulative inactivation as the burst evolves (middle). The dendritic voltage-gated Na$^+$ conductance, $g_{Na,d}$, fails when the somatic ISI is within its refractory period (bottom). c) Summary graph showing the decrease in peak conductance of $g_{Dr,d}$ and $g_{Na,d}$ as a function of spike number for the burst shown in b. Whereas $g_{Dr,d}$ inactivates in a cumulative way, $g_{Na,d}$ decays much more gradually but is completely shut off by the high-frequency doublet. Adapted from [33].

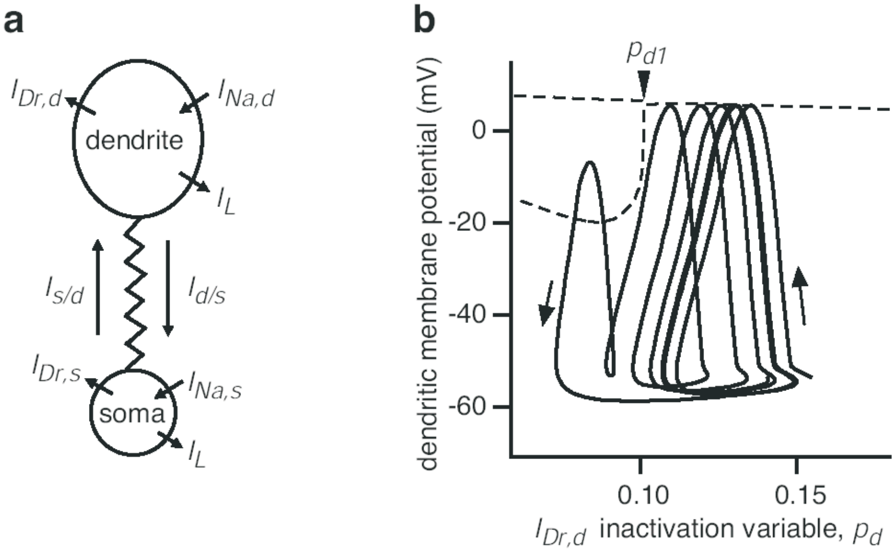

Figure 8.11

Two-compartment model of burst generation. a) Sketch of the somatic and dendritic compartments linked by an axial resistance. b) The dashed lines show the quasi-static bifurcation diagram with a representative of the fast subsystem, the maximum dendritic membrane voltage, as a function of the slow subsystem, the dendritic K^+ inactivation variable, p_d. Overlaid is a single burst trajectory (solid line; burst begins with the upwards pointing arrow on the right). Adapted from [32].

with dendritic spike failure, since for $p_d < p_{d1}$ dendritic repolarization is sufficiently slow to cause very strong somatic DAPs capable of eliciting a second somatic spike after a small time interval (∼3 ms). The overlaid burst trajectory (solid line) shows the beginning of the burst on the right side (upwards arrow). The maximum of the dendritic membrane voltage decreases for the second spike of the doublet (compare Figure 8.9b), which occurs at $p_d < p_{d1}$. The short doublet ISI is followed by the long interburst ISI, the slow variable recovers until the next burst begins. Because p_d is reinjected near an infinite-period bifurcation (saddle-node bifurcation of fixed points responsible for spike excitability), Doiron et al. [32] termed this burst mechanism *ghostbursting* (*sensing* the ghost of an infinite-period bifurcation [118]). Thus, the two-compartment model nicely explains the dynamics of pyramidal cell bursting observed *in vitro* by the interplay between fast spike-generating mechanisms and slow dendritic K^+-channel inactivation.

In a further reduction of the model, Laing and Longtin [71] replaced the six ordinary differential equation model by an integrate-and-fire model consisting of a set of two discontinuous delay-differential equations. An interesting aspect of this model is that it uses a discrete delay to mimic the ping-pong effect between soma and dendrite. When a spike occurs, the somatic membrane potential is boosted by a variable amount but only if the preceding ISI was longer than the dendritic refractory period and only after a certain delay. The amount of somatic boosting depends on the firing history of the neuron. For long ISIs, it decays towards zero, for short ISIs it builds up.

Bifurcation analysis of both the ghostburster and the delay model revealed properties that contrast with other models of burst generation. When increasing amounts of current are injected into the soma, both reduced models move from quiescence for subthreshold current through a range of tonic periodic firing into irregular bursting (Figure 8.12) [32, 71]. The transition from quiescence to tonic firing is through a saddle-node bifurcation of fixed points after which the systems follow a stable limit cycle. The periodic attractor increases monotonically in frequency as current is increased. The fact that the models pass from quiescence to repetitive firing through a saddle-node bifurcation is characteristic of class I excitability [57, 101]. Accordingly, the neurons are able to fire at arbitrarily low rates close to the bifurcation, which is also observed when injecting small amounts of current into pyramidal cells in the slice preparation [75]. At higher current the models move through a saddle-node bifurcation of limit cycles after which they follow a chaotic attractor corresponding to burst firing. For very large input currents, the cells periodically discharge spike doublets (right of the dotted line in Figure 8.12 a, b). This progression from quiescence through periodic firing and bursting to periodic doublet discharge closely reproduces the behavior of pyramidal cells in the slice preparation [75]. Similar to the ghostburster model, the delay integrate-and-fire model also allows the generation of a wide *gallery* of bursts of different shapes indicating that pyramidal cells may be able to adjust burst duration and frequency depending on context.

The simplicity of the delay model also allowed examination of the effects of periodic forcing corresponding to injection of sinusoidal current at the soma. Depending on the frequency of sinusoidal forcing, the threshold for burst firing could be increased or decreased relative to the threshold in the unforced system. This finding suggests that depending on the frequency of amplitude modulations of the electric field, the threshold for burst firing of pyramidal cells might shift.

The most appealing aspect of the delay model is its simplicity and computational efficiency. Since the model captures the basic properties of burst firing described by the more elaborate ionic models [32, 33], it may be suitable for use in larger-scale models of electrosensory processing.

8.6 Comparison with other bursting neurons

Bursting neurons have been described in a variety of systems including the crustacean stomatogastric ganglion [5], the lamprey spinal cord [45], dorsal root ganglion cells [1], thalamic reticular and relay cells [107, 116], and pyramidal neurons in sev-

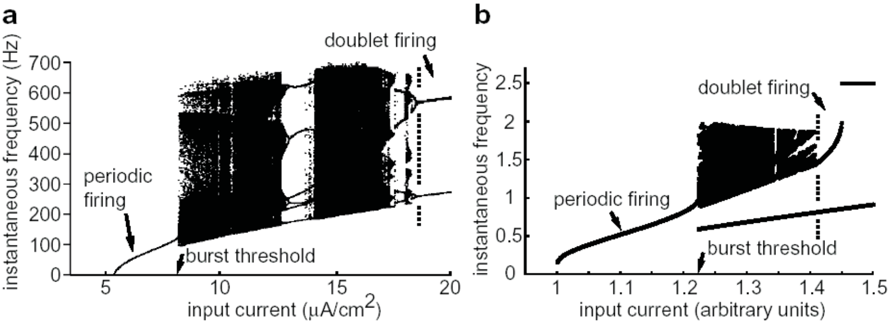

Figure 8.12

Instantaneous firing frequency versus amount of injected current for a) ghostburster model, and b) two-variable delay-differential-equation model. Both models show an absolute threshold for firing, above which they discharge periodically. At some intermediate current ($I \sim 8.5$ for the ghostburster model and $I \sim 1.22$ for the delay model), the models transition through a saddle-node bifurcation of limit cycles into irregular bursting. At very high input currents they begin to fire doublets (right of the dotted line in a and b). Doublet firing involves two distinct ISI values, the long inter-doublet ISI (upper line) and the short doublet ISI (lower line). (a) adapted from [32], (b) adapted from [71].

eral cortical areas and layers [26, 43, 83]. Naturally, the depth of understanding of the underlying ionic mechanisms is not the same for every system. However, modeling approaches based on experimental findings have been helpful in elucidating cellular and dynamical aspects of burst firing in a number of different preparations. In the following, we discuss three aspects of burst firing to which the electric fish preparation has brought new perspectives: 1) burst firing can be caused by a *ping-pong* interplay between soma and dendrite; 2) ghostbursting offers novel dynamics for oscillatory bursting; 3) the underlying ionic mechanisms shape the ISI sequence within the burst.

8.6.1 Ping-pong between soma and dendrite

The term *ping-pong* [129] refers to the interplay between soma and dendrite that has been shown to be an essential part of the burst mechanism in a number of cell types. The idea that soma-dendritic interactions shape neuronal response properties became prominent when intracellular labeling and electrophysiology were combined (e.g., [72, 83]). Based on reconstructions of various neocortical cell types, Mainen and Sejnowski [80] showed that neurons sharing the same ionic channel distributions but differing in dendritic morphology displayed a wide range of response properties from regular firing to rhythmically bursting. Dendritic Na$^+$ channels proved to be necessary for bursting since they support backpropagation of spikes into the dendrite and the subsequent current flow back into the soma. The somatic DAP can then feed further spikes, similar to the mechanism described above for ELL pyramidal cells [33, 75]. Two basic mechanisms for boosting the DAP seem to be realized in bursting neurons. First, voltage-activated dendritic Ca^{2+} channels have been found to increase the somatic DAP in pyramidal cells in layer 5 of neocortex [104, 132], in the subiculum [61] and at least in a fraction of CA1 pyramidal cells of the hippocampus [42, 79]. Second, the somatic DAP can be enhanced by persistent Na^{+} currents as observed in cortical chattering cells [19], in layer 3 sensorimotor cortical neurons [92], some hippocampal CA1 neurons [3, 119], and in ELL pyramidal cells [34]. In these latter cases, Ca^{2+} has been shown not to be a necessary component for bursting. Wang [129] suggested that spike-triggered Ca^{2+} influx might be too slow to support bursting at high $\hat{\gamma}$ -frequencies (20-70 Hz) observed in chattering cells [19, 43, 117]. The same reasoning could apply to bursting of ELL pyramidal cells *in vitro*, which shows oscillations in the $\hat{\gamma}$-range [125, 126], and which is Ca^{2+}-independent [75].

These systems all share a somatic DAP induced by current flow from the dendrite. They differ, however, in several other aspects such as, for example, mechanisms of burst termination. In layer 3 cells of sensorimotor cortex, Ca^{2+}-activated K^{+} channels repolarize the dendrite and stop the current flow towards the soma [92]. This mechanism had been predicted by a multi-compartmental modeling study of layer 5 intrinsically bursting pyramidal cells [100]. Based on a two-compartment model of neocortical chattering cells, Wang [129] suggested that bursts are terminated when a dendritic voltage-dependent K^{+} channel is sufficiently activated to repolarize the dendritic membrane. Hence, the above described burst termination by failure of backpropagation due to the relatively long dendritic refractory period constitutes a hitherto unknown mechanism [75].

For thalamic relay cells it was long believed that dendrites did not play a major role in burst generation since bursting persists in acutely isolated cells devoid of dendrites [56]. In a recent combined *in vitro* and modeling study, however, Destexhe and coworkers showed that the low-threshold Ca^{2+} channels underlying burst generation had to have a roughly 5 times higher density in the dendrite than in the soma to yield Ca^{2+} spikes comparable to those seen in intact relay cells [30]. The actual burst consists of fast Na^{+} and K^{+} activity riding the crest of the Ca^{2+} spike. At depolarized membrane potentials, the underlying I_T channel is inactivated and the cells respond in tonic mode [107]. Deinactivation requires hyperpolarization for at least 50-100 ms. Therefore, thalamic bursting is characterized by very long ISIs preceding the actual burst.

It should be mentioned that soma-dendritic interactions are not the only route to bursting. Some cell types, such as cerebellar granule cells, seem to be electrotonically too compact to support a ping-pong mechanism [27, 80]. Instead, a persistent Na^{+} current in conjunction with a slow Ca^{2+}-independent K^{+} current can cause oscillations, with fast Na^{+} spikes riding on top of the oscillations [27].

8.6.2 Dynamical properties of burst oscillations

On a more macroscopic scale, bursting in ELL pyramidal cells seems unique in two respects. First, ISI duration decreases within bursts, which is atypical. Second, the changes in firing properties with increasing input current exhibit an unusual bifurcation structure. As shown in the slice preparation [75] and in both the reduced models [32, 71], the firing properties pass from quiescence for subthreshold input current through tonic firing for intermediate current levels to bursting. Other systems, in contrast, have been shown to pass from quiescence through bursting to tonic firing (e.g. [27, 36, 43, 100, 104, 112, 116, 129]). Accordingly, for a given input current, bursting systems are usually described as switching between quiescence (fixed point) and spiking (limit cycle) [57]. As shown by the reduced ELL pyramidal cell models, however, the fast subsystem can always follow a limit cycle [32, 71]. Since the slow subsystem is itself oscillating, it modulates the period of the fast subsystem and forces it to pass near the ghost of an infinite-period bifurcation, which yields the long interburst intervals, as opposed to bifurcating to a fixed point solution.

8.6.3 Intra-burst ISI sequences

Within a burst fired by an ELL pyramidal cell, instantaneous firing rate increases until a spike doublet eventually terminates the burst. Due to its long refractory period, the dendrite fails to actively backpropagate the action potential allowing the AHP to set in and repolarize the soma (Figure 8.9b). From a dynamical systems point of view, the burst termination can be understood as a bifurcation from a period-one to a period-two limit cycle of the fast, spike-generating, system (Figure 8.11b). In all other models of bursting neurons, bursts end with a transition form a period-one limit cycle to a fixed point (quiescence; [57]). This corresponds to the observation that, in most systems, bursts begin with a very high instantaneous firing rate and then slow down. One reason for the slow-down can be the gradual activation of a dendritic K^+ channel which reduces current flow to the soma and increases the time to reach threshold for action potential firing [92, 100, 129]. Alternatively, spike backpropagation, and with it the somatic DAP, can fail when dendritic Na^+ channels cumulatively inactivate [25, 61, 114] or when synaptic inhibition sufficiently hyperpolarizes the dendritic membrane [20, 77, 122].

8.7 Conclusions

Two main lines of evidence indicate that bursts can play an important role in neuronal information transmission. First, bursts have been shown to surpass single spikes in their information carrying performance [35, 40, 46, 86, 99]. Besides acting as unitary events, burst duration, that is the number of spikes, may also be a mode of information transmission [28, 65]. Second, high-frequency burst firing increases the reliability of synaptic transmission at unreliable synapses [76, 113, 120, 128, 129, 132]. The development of the technique of feature extraction analysis has given us a powerful tool to quantify how reliably neurons indicate the occurrence of certain stimulus features without prior knowledge of what these features look like [39, 40, 86]. Its application to neuronal responses in various behavioral contexts may teach us how the possible contribution of burst firing (or other firing patterns) to information transmission changes with changing behavioral context.

Compartmental modelling based on detailed reconstructions of neuronal morphology has demonstrated that dendritic structure is a major determinant of a neuron's firing properties [26, 50, 80, 83, 105]. From a mechanistic point of view, reduced models, such as two-compartment models and point neurons, have been effective at revealing the underlying dynamics of burst generation. As illustrated here, detailed multi-compartmental modeling can aid in understanding the ionic and structural mechanisms underlying particular neuronal firing patterns [33, 34] and, when that is achieved, simplified models can help in elucidating the dynamic properties of these mechanisms [32, 71]. The ghostburster model and the delay model reproduce burst discharge as it is observed *in vitro* in spite of their simplicity, suggesting that the essential components of the intrinsic burst mechanism are understood.

For pyramidal cells in the ELL of weakly electric fish, there are first indications that the probability of burst generation is under descending control and depends on the spatial geometry of the stimulus [13, 31, 75, 97]. Similar observations have been made for thalamic neurons (e.g., [29, 37, 107]) and nerve cells in the subthalamic nucleus [127]. Modeling studies will be key in the exploration of how descending control shapes burst firing. Interestingly, ELL pyramidal cells possess a number of spatially distinct input areas that could be controlled separately depending on behavioral context [9, 16].

One of the most urgent questions to be addressed is whether or not the mechanisms that shape bursting under *in vitro* conditions are also the key determinants of burst firing in the intact animal. Of course, the ionic channels responsible for conditional backpropagation will be at work *in vivo*, too. Nevertheless, most pyramidal cells when recorded *in vivo* do not show shortening of ISIs in the course of a burst, at least under the stimulus conditions studied so far [13, 40, 86]. The models described above therefore need to be refined. They need to address the effects of naturalistic synaptic input from the sensory afferents, including the effects of indirect inhibitory input via local interneurons, and the possible contributions of descending control. Descending control could act directly via synaptic excitation and inhibition, but also indirectly by modulations of synaptic transmission [9] or by inducing phosphorylation of the AptKv3.3 K^+ channel [97]. The development of the reduced models may also make it possible to construct larger network models that could still incorporate naturalistic spike train statistics. Two construction blocks for such a network model could be the delay-differential equation model of an ELL pyramidal cell [71] and a recently developed simple model of P-receptor afferents that captures much of the experimentally observed firing dynamics [18].

Acknowledgments: We gratefully acknowledge the very generous provision of figure material by Ray Turner and Brent Doiron. We also would like to thank Ling Chen for help with Figures 8.3A and 8.12b, and Mark Nelson and Brent Doiron for critically reading the manuscript. The work on stimulus encoding and feature extraction in the electrosensory system was a collaboration with Walter Metzner, Gabriel

Kreiman, and Christof Koch, and was supported by NSF.

References

[1] Amir R., Michaelis M., and Devor M. (2002) Burst discharge in primary sensory neurons: triggered by subthreshold oscillations, maintained by depolarizing afterpotentials. *J. Neurosci.* **22**: 1187-1198.

[2] Assad C., Rasnow B., and Stoddard P.K. (1999) The electric organ discharges and electric images during electrolocation. *J. Exp. Biol.* **202**: 1185-1193.

[3] Azouz R., Jensen M.S., and Yaari Y. (1996) Ionic basis for spike after-depolarization and burst generation in adult rat hippocampal CA1 pyramidal cells. *J. Physiol.* (London) **492**: 211-223.

[4] Bair W., Koch C., Newsome W., and Britten K. (1994) Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J. Neurosci.* **14**: 2870-2892.

[5] Bal T., Nagy F., and Moulins M. (1988) The pyloric central pattern generator in crustacea: a set of conditional neuronal oscillators. *J. Comp. Physiol. A* **163**: 715-727.

[6] Barker J.L., and Gainer H. (1975) Studies on bursting pacemaker potential activity in molluscan neurons. I. Membrane properties and ionic contributions. *Brain Res.* **84**: 461-477.

[7] Bastian J. (1981a) Electrolocation. II. The effects of moving objects and other electrical stimuli on the activities of two categories of posterior lateral line lobe cells in *Apteronotus albifrons. J. Comp. Physiol. A* **144**: 481-494.

[8] Bastian J. (1981b) Electrolocation. I. How the electroreceptors of Apteronotus albifrons code for moving objects and other electrical stimuli. *J. Comp. Physiol. A* **144**: 465-479.

[9] Bastian J. (1999) Plasticity of feedback inputs in the apteronotid electrosensory system. *J. Exp. Biol.* **202**: 1327-1337.

[10] Bastian J. (2003) Electrolocation. In: *The Handbook of Brain Theory and Neural Networks.* 2nd ed. (Arbib MA, ed.), pp 391-394. Cambridge, MA: MIT Press.

[11] Bastian J., Chacron M.J., and Maler L. (2002) Receptive field organization determines pyramidal cell stimulus-encoding capability and spatial stimulus selectivity. *J. Neurosci.* **22**: 4577-4590.

[12] Bastian J., and Courtright J. (1991) Morphological correlates of pyramidal cell adaptation rate in the electrosensory lateral line lobe of weakly electric fish. *J.*

Comp. Physiol. A **168**: 393-407.

[13] Bastian J., and Nguyenkim J. (2001) Dendritic modulation of burst-like firing in sensory neurons. *J. Neurophysiol.* **85**: 10-22.

[14] Bell C.C. (2001) Memory-based expectations in electrosensory systems. *Curr. Opin. Neurobiol.* **11**: 481-487.

[15] Berman N.J., Dunn R.J., and Maler L. (2001) Function of NMDA receptors and persistent sodium channels in a feedback pathway of the electrosensory system. *J. Neurophysiol.* **86**:

[16] Berman N.J., and Maler L. (1999) Neural architecture of the electrosensory lateral line lobe: Adaptations for coincidence detection, a sensory searchlight and frequency-dependent adaptive filtering. *J. Exp. Biol.* **202**: 1243-1253. 1612-1621.

[17] Berman N.J., Plant J., Turner R.W., and Maler L. (1997) Excitatory amino acid receptors at a feedback pathway in the electrosensory system: implications for the searchlight hypothesis. *J. Neurophysiol.* **78**: 1869-1881.

[18] Brandman R., and Nelson M.E. (2002) A simple model of long-term spike train regularization. *Neural Comput.* **14**: 1575-1597.

[19] Brumberg J.C., Nowak L.G., and McCormick D.A. (2000) Ionic mechanisms underlying repetitive high-frequency burst firing in supragranular cortical neurons. *J. Neurosci.* **20**: 4829-4843.

[20] Buzsáki G., Penttonen M., Nadasdy Z., and Bragin A. (1996) Pattern- and inhibition-dependent invasion of pyramidal cell dendrites by fast spikes in the hippocampus *in vivo*. *Proc. Natl. Acad. Sci. USA* **93**: 9921-9925.

[21] Carr C.E., and Friedman M.A. (1999) Evolution of time coding systems. *Neural Comput.* **11**: 1-20.

[22] Carr C.E., Maler L., and Sas E. (1982) Peripheral organization and central projections of the electrosensory nerves in gymnotiform fish. *J. Comp. Neurol.* **211**: 139-153.

[23] Chacron M.J., Longtin A., and Maler L. (2001) Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli. *J. Neurosci.* **21**: 5328-5343.

[24] Chagnac-Amitai Y., Luhmann H.J., and Prince D.A. (1990) Burst generating and regular spiking layer 5 pyramidal neurons of rat neocortex have different morphological features. *J. Comp. Neurol.* **296**: 598-613.

[25] Colbert C.M., Magee J.C., Hoffman D.A., and Johnston D. (1997) Slow recovery from inactivation of Na$^+$ channels underlies the activity-dependent attenuation of dendritic action potentials in hippocampal CA1 pyramidal neurons. *J. Neurosci.* **17**: 6512-6521.

[26] Connors B.W., and Gutnick M.J. (1990) Intrinsic firing patterns of diverse neo-

cortical neurons. *Trends. Neurosci.* **13**: 99-104.

[27] D'Angelo E., Nieus T., Maffei A., Armano S., Rossi P., Taglietti V., Fontana A., and Naldi G. (2001) Theta-frequency bursting and resonance in cerebellar granule cells: experimental evidence and modeling of a slow- K^{+}-dependent mechanism. *J. Neurosci.* **21**: 759-770.

[28] DeBusk B.C., DeBruyn E.J., Snider R.K., Kabara J.F., and Bonds A.B. (1997) Stimulus-dependent modulation of spike burst length in cat striate cortical cells. *J. Neurophysiol.* **78**: 199-213.

[29] Destexhe A. (2000) Modelling corticothalamic feedback and the gating of the thalamus by the cerebral cortex. *J. Physiol.* (Paris) **94**: 391-410.

[30] Destexhe A., Neubig M., Ulrich D., and Huguenard J. (1998) Dendritic low-threshold calcium currents in thalamic relay cells. *J. Neurosci.* **18**: 3574-3588.

[31] Doiron B., Chacron M.J., Maler L., Longtin A., and Bastian J. (2003) Inhibitory feedback required for network oscillatory responses to communication but not prey stimuli. *Nature* **421**: 539-543.

[32] Doiron B., Laing C., Longtin A., and Maler L. (2002) Ghost bursting: a novel neuronal burst mechanism. *J. Comput. Neurosci.* **12**: 5-25.

[33] Doiron B., Longtin A., Turner R.W., and Maler L. (2001) Model of gamma frequency burst discharge generated by conditional backpropagation. *J. Neurophysiol.* **86**: 1523-1545.

[34] Doiron B., Noonan L., Lemon N., and Turner R.W. (2003) Persistent Na^{+} current modifies burst discharge by regulating conditional backpropagation of dendritic spikes. *J. Neurophysiol.* **89**: 324-337.

[35] Eggermont J.J., and Smith G.M. (1996) Burst-firing sharpens frequency-tuning in primary auditory cortex. *Neuroreport* **7**: 753-757.

[36] Falcke M., Huerta R., Rabinovich M.I., Abarbanel H.D.I., Elson R.C., and Selverston A.I. (2000) Modeling observed chaotic oscillations in bursting neurons: the role of calcium dynamics and IP3. *Biol. Cybern.* **82**: 517-527.

[37] Fanselow E.E., Sameshima K., Baccala L.A., and Nicolelis M.A. (2001) Thalamic bursting in rats during different awake behavioral states. *Proc. Natl. Acad. Sci. USA* **98**: 15330-15335.

[38] Gabbiani F., Koch C. (1998) Principles of spike train analysis. In: *Methods in neuronal modeling* (Koch C, Segev I, eds.), pp. 313-360. Cambridge, MA: MIT Press.

[39] Gabbiani F., and Metzner W. (1999) Encoding and processing of sensory information in neural spike trains. *J. Exp. Biol.* **202**: 1267-1279.

[40] Gabbiani F., Metzner W., Wessel R., and Koch C. (1996) From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**:564-567.

[41] Goense J.B.M., Ratnam R., and Nelson M.E. (in press) Burst firing improves the detection of weak signals in spike trains. *Neurocomp.*

[42] Golding N.L., Jung H.-Y., Mickus T., and Spruston N. (1999) Dendritic calcium spike initiation and repolarization are controlled by distinct potassium channel subtypes in CA1 pyramidal neurons. *J. Neurosci.* **19**: 8789-8798.

[43] Gray C.M., and McCormick D.A. (1996) Chattering cells: Superficial pyramidal neurons contributing to the generation of synchronous firing in the visual cortex. *Science* **274**: 109-113.

[44] Green D.M., and Swets J.A. (1966) *Signal Detection Theory and Psychophysics.* New York, NY: Wiley.

[45] Grillner S., Ekeberg O., El Manira A., Lansner A., Parker D., Tegner J., and Wallen P. (1998) Intrinsic function of a neuronal network - a vertebrate central pattern generator. *Brain Res. Rev.* **26**: 184-197.

[46] Guido W., Lu S.M., Vaughan J.W., Godwin D.W., and Sherman S.M. (1995) Receiver operating characteristic (ROC) analysis of neurons in the cat's lateral geniculate nucleus during tonic and burst response mode. *Vis. Neurosci.* **12**: 723-741.

[47] Guido W., and Weyand T. (1995) Burst responses in thalamic relay cells of the awake behaving cat. *J. Neurophysiol.* **74**: 1782-1786.

[48] Hagedorn M., and Heiligenberg W. (1985) Court and spark: electric signals in the courtship and mating of gymnotoid electric fish. *Anim. Behav.* **33**: 254-265.

[49] Harris K.D., Hirase H., Leinekugel X., Henze D.A., and Buzsáki G. (2001) Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**: 141-149.

[50] Häusser M., Spruston N., and Stuart G.J. (2000) Diversity and dynamics of dendritic signaling. *Science* **290**: 739-744.

[51] He J., and Hu B. (2002) Differential distribution of burst and single-spike responses in auditory thalamus. *J. Neurophysiol.* **88**: 2152-2156.

[52] Heiligenberg W. (1991) *Neural Nets in Electric Fish.* Cambridge, MA: MIT Press.

[53] Heiligenberg W., and Dye J. (1982) Labeling of electroreceptive afferents in a gymnotoid fish by intracellular injection of HRP: the mystery of multiple maps. *J. Comp. Physiol. A* **148**: 287-296.

[54] Hodgkin A.L., and Huxley A.F. (1952) A quantitative description of membrane currents and its application to conduction and excitation in nerve. *J. Physiol.* (London) **117**: 500-544.

[55] Hopkins C.D. (1988) Neuroethology of electric communication. *Annu. Rev. Neurosci.* **11**: 497-535.

[56] Huguenard J.R. (1996) Low-threshold calcium currents in central nervous system neurons. *Annu. Rev. Physiol.* **58**: 329-348.

[57] Izhikevich E.M. (2000) Neural excitabilty, spiking and bursting. *Int. J. Bif. Chaos.* **10**: 1171-1266.

[58] Izhikevich E.M. (2002) Resonance and selective communication via bursts in neurons having subthreshold oscillations. *BioSyst.* **67**:95-102.

[59] Jahnsen H., and Llinas R. (1984) Ionic basis for the electroresponsiveness and oscillatory properties of guinea-pig thalamic neurones *in vitro*. *J. Physiol.* (Lond.) **349**: 227-247.

[60] Jung H-Y, Mickus T., and Spruston N. (1997) Prolonged sodium channel inactivation contributes to dendritic action potential attenuation in hippocampal pyramidal neurons. *J. Neurosci.* **17**: 6639-6646.

[61] Jung H-Y., Staff N.P., and Spruston N. (2001) Action potential bursting in subicular pyramidal neurons is driven by a calcium tail current. *J. Neurosci.* **21**: 3312-3321.

[62] Kaas J.H. (1997) Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* **44**: 107-12.

[63] Kalmijn J. (1982) Electric and magnetic field detection in elasmobranch fishes. *Science* **218**: 916-918.

[64] Kepecs A., and Wang X-J (2000) Analysis of complex bursting in cortical pyramidal neuron models. *Neurocomp.* **32-33**: 181-187.

[65] Kepecs A., Wang X-J., and Lisman J. (2002) Bursting neurons signal input slope. *J. Neurosci.* **22**: 9053-9062.

[66] Konishi M. (1990) Similar algorithms in different sensory systems and animals. *Cold Spring Harb. Symp. Quant. Biol.* **55**: 575-584.

[67] Konishi M (1991) Deciphering the brain's codes. *Neural Comput* **3**: 1-18.

[68] Krahe R., Kreiman G., Gabbiani F., Koch C., and Metzner W. (2002) Stimulus encoding and feature extraction by multiple sensory neurons. *J. Neurosci.* **22**: 2374-2382.

[69] Krahe R., Kreiman G., Gabbiani F., Koch C., and Metzner W. (in prep.) Feature extraction from global and local stimuli by electrosensory neurons.

[70] Kreiman G., Krahe R., Metzner W., Koch C., and Gabbiani F. (2000) Robustness and variability of neuronal coding by amplitude-sensitive afferents in the weakly electric fish *Eigenmannia*. *J. Neurophysiol.* **84**: 189-204.

[71] Laing C.R., and Longtin A. (2002) A two-variable model of somatic-dendritic interactions in a bursting neuron. *Bull. Math. Biol.* **64**: 829-860.

[72] Larkman A., and Mason A. (1990) Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. I. Establish-

ment of cell classes. *J. Neurosci.* **10**: 1407-1414.

[73] Legéndy C.R., and Salcman M. (1985) Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.* **53**: 926-939.

[74] Lehky S.R., Sejnowski T.J., and Desimone R. (1992) Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *J. Neurosci.* **12**: 3568-3581.

[75] Lemon N., and Turner R.W. (2000) Conditional spike backpropagation generates burst discharge in a sensory neuron. *J. Neurophysiol.* **84**: 1519-1530.

[76] Lisman J.E. (1997) Bursts as a unit of neural information: making unreliable synapses reliable. *Trends. Neurosci.* **20**:38-43.

[77] Lowe G. (2002) Inhibition of backpropagating action potentials in mitral cell secondary dendrites. *J. Neurophysiol.* **88**: 64-85.

[78] Lu S.M., Guido W., and Sherman S.M. (1992) Effects of membrane voltage on receptive field properties of lateral geniculate neurons in the cat: contributions of the low-threshold Ca^{2+} conductance. *J. Neurophysiol.* **68**: 2185-2198.

[79] Magee J.C., and Carruth M. (1999) Dendritic voltage-gated ion channels regulate the action potential firing mode of hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **82**: 1895-1901.

[80] Mainen Z.F., and Sejnowski T.J. (1996) Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* **382**: 363-366.

[81] Maler L., Sas E.K., and Rogers J. (1981) The cytology of the posterior lateral line lobe of high-frequency weakly electric fish (Gymnotidae): dendritic differentiation and synaptic specificity in a simple cortex. *J. Comp. Neurol.* **195**: 87-139.

[82] Martinez-Conde S., Macknik S.L., and Hubel D.H. (2002) The function of bursts of spikes during visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. *Proc. Natl. Acad. Sci. USA* **99**:13920-13925.

[83] Mason A., and Larkman A. (1990) Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. II. Electrophysiology. *J. Neurosci.* **10**: 1415-1428.

[84] McCormick D.A., and Contreras D (2001) On the cellular and network bases of epileptic seizures. *Annu. Rev. Physiol.* **63**:815-846.

[85] Metzner W., and Juranek J. (1997) A sensory brain map for each behavior. *Proc. Natl. Acad. Sci. USA* **94**: 14798-14803.

[86] Metzner W., Koch C., Wessel R., and Gabbiani F. (1998) Feature extraction by burst-like spike patterns in multiple sensory maps. *J. Neurosci.* **18**: 2283-2300.

[87] Metzner W., and Viete S. (1996) The neuronal basis of communication and orientation in the weakly electric fish, *Eigenmannia*. I. Communication behavior or: Seeking a conspecific's response. *Naturwissenschaften* **83**: 6-14.

[88] Moller P. (1995) *Electric Fishes. History and Behavior.* Fish and Fisheries Series, vol. 17 (Pitcher T.J., ed.) London: Chapman and Hall.

[89] Nelson M.E., and MacIver M.A. (1999) Prey capture in the weakly electric fish *Apteronotus albifrons*: Sensory acquisition strategies and electrosensory consequences. *J. Exp. Biol.* **202**: 1195-1203.

[90] Nelson M.E., MacIver M.A., and Coombs S. (2002) Modeling electrosensory and mechanosensory images during the predatory behavior of weakly electric fish. *Brain Behav. Evol.* **59**: 199-210.

[91] Nelson M.E., Xu Z., and Payne J.R. (1997) Characterization and modeling of P-type electrosensory afferent responses to amplitude modulations in a wave-type electric fish. *J. Comp. Physiol. A* **181**: 532-544.

[92] Nishimura Y., Asahli M., Saitoh K., Kitagawa H., Kumazawa Y., Itoh K., Lin M., Akamine T., Shibuya H., Asahara T., and Yamamoto T. (2001) Ionic mechanisms underlying burst firing of layer III sensorimotor cortical neurons of the cat: an *in vitro* slice study. *J. Neurophysiol.* **86**: 771-781.

[93] Noonan L., Doiron B., Laing C., Longtin A., and Turner R.W. (2003) A dynamic dendritic refractory period regulates burst discharge in the electrosensory lobe of weakly electric fish. *J. Neurosci.* **23**: 1524-1534.

[94] Paulsen O., and Sejnowski T.J. (2000) Natural patterns of activity and long-term synaptic plasticity. *Curr. Opin. Neurobiol.* **10**:172-179.

[95] Ramcharan E.J., Gnadt J.W., and Sherman S.M. (2000) Burst and tonic firing in thalamic cells of unanesthetized, behaving monkeys. *Vis. Neurosci.* **17**: 55-62.

[96] Rashid A.J., Dunn R.J., and Turner R.W. (2001) A prominent soma-dendritic distribution of Kv3.3 K^+ channels in electrosensory and cerebellar neurons. *J. Comp. Neurol.* **441**: 234-247.

[97] Rashid A.J., Morales E., Turner R.W., and Dunn R.J. (2001) The contribution of dendritic Kv3 K^+ channels to burst threshold in a sensory neuron. *J. Neurosci.* **21**: 125-135.

[98] Ratnam R., and Nelson M.E. (2000) Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J. Neurosci.* **20**: 6672-6683.

[99] Reinagel P., Godwin D., Sherman S.M., and Koch C. (1999) Encoding of visual information by LGN bursts. *J. Neurophysiol.* **81**:2558-2569.

[100] Rhodes P.A., and Gray C.M. (1994) Simulations of intrinsically bursting pyramidal neurons. *Neural Comput.* **6**: 1086-1110.

[101] Rinzel J., and Ermentrout B. (1998) Analysis of neural excitability and oscillations. In: *Methods in Neuronal Modeling* (Koch C, Segev I, eds.), pp. 251-291. Cambridge, MA: MIT Press.

[102] Saunders J., and Bastian J. (1984) The physiology and morphology of two types of electrosensory neurons in the weakly electric fish, *Apteronotus leptorhynchus. J. Comp. Physiol. A* **154**: 199-209.

[103] Scheich H., Bullock T.H., and Hamstra R.H.J. (1973) Coding properties of two classes of afferent nerve fibers: high frequency electroreceptors in the electric fish, *Eigenmannia. J. Neurophysiol.* **36**: 39-60.

[104] Schwindt P., and Crill W. (1999) Mechanisms underlying burst and regular spiking evoked by dendritic depolarization in layer 5 cortical pyramidal neurons. *J. Neurophysiol.* **81**: 1341-1354.

[105] Segev I., and Rall W. (1998) Excitable dendrites and spines: earlier theoretical insights elucidate recent direct observations. *Trends. Neurosci.* **21**: 453-460.

[106] Sherman S.M. (1996) Dual response modes in lateral geniculate neurons: mechanisms and functions. *Vis. Neurosci.* **13**: 205-13.

[107] Sherman S.M. (2001a) Tonic and burst firing: dual modes of thalamocortical relay. *Trends. Neurosci.* **24**:122-126.

[108] Sherman S.M. (2001b) A wake-up call from the thalamus. *Nature Neurosci.* **4**: 344-346.

[109] Shumway C. (1989a) Multiple electrosensory maps in the medulla of weakly electric gymnotiform fish. I. Physiological differences. *J. Neurosci.* **9**: 4388-4399.

[110] Shumway C. (1989b) Multiple electrosensory maps in the medulla of weakly electric gymnotiform fish. II. Anatomical differences. *J. Neurosci.* **9**: 4400-4415.

[111] Sillito A.M., Jones H.E., Gerstein G.L., and West D.C. (1994) Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* **369**: 479-482.

[112] Silva L.R., Amitai Y., and Connors B.W. (1991) Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. *Science* **251**: 432-435.

[113] Snider R.K., Kabara J.F., Roig B.R., and Bonds A.B. (1998) Burst firing and modulation of functional connectivity in cat striate cortex. *J. Neurophysiol.* **80**: 730-744.

[114] Spruston N., Schiller Y., Stuart G., and Sakmann B. (1995) Activity-dependent action potential invasion and calcium influx into hippocampal CA1 pyramidal cells. *Science* **268**: 297-300.

[115] Steriade M. (2001) To burst, or rather, not to burst. *Nature Neurosci.* **4**: 671.

[116] Steriade M., McCormick D.A., and Sejnowski T.J. (1993) Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**:679-685.

[117] Steriade M., Timofeev I., Dürmüller N., and Grenier F. (1998) Dynamic properties of corticothalamic neurons and local cortical interneurons generating fast rhythmic (30-40 Hz) spike bursts. *J. Neurophysiol.* **79**: 483-490.

[118] Strogatz S.H. (1994) *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering*. Reading, MA: Addison-Wesley.

[119] Su H., Alroy G., Kirson E.D., and Yaari Y. (2001) Extracellular calcium modulates persistent sodium current-dependent burst-firing in hippocampal pyramidal neurons. *J. Neurosci.* **21**: 4173-4182.

[120] Swadlow H.A., and Gusev A.G. (2001) The impact of 'bursting' thalamic impulses at a neocortical synapse. *Nature Neurosci.* **4**:402-408.

[121] Swadlow H.A., Gusev A.G., and Bezdudnaya T. (2002) Activation of a cortical column by a thalamocortical impulse. *J. Neurosci.* **22**: 7766-7773.

[122] Tsubokawa H., and Ross W.N. (1996) IPSPs modulate spike backpropagation and associated [Ca^{2+}] changes in the dendrites of hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **76**: 2896-2906.

[123] Turner R.W., and Maler L. (1999) Oscillatory and burst discharge in the *Apteronotid electrosensory* lateral line lobe. *J. Exp. Biol.* **202**: 1255-1265.

[124] Turner R.W., Maler L., and Burrows M. (1999) Electrolocation and electrocommunication. *J. Exp. Biol.* **202**.

[125] Turner R.W., Maler L., Deerinck T., Levinson S.R., and Ellisman M.H. (1994) TTX-sensitive dendritic sodium channels underlie oscillatory discharge in a vertebrate sensory neuron. *J. Neurosci.* **14**: 6453-6471.

[126] Turner R.W., Plant J.R., and Maler L. (1996) Oscillatory and burst discharges across electrosensory topographic maps. *J. Neurophysiol.* **76**: 2364-2382.

[127] Urbain N., Rentero N., Gervasoni D., Renaud B., and Chouvet G. (2002) The switch of subthalamic neurons from an irregular to a bursting pattern does not solely depend on their GABAergic inputs in the anesthetic-free rat. *J. Neurosci.* **22**: 8665-8675.

[128] Usrey W.M., Alonso J-M, Reid R.C. (2000) Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. *J. Neurosci.* **20**:5461-5467.

[129] Wang X-J. (1999) Fast burst firing and short-term synaptic plasticity: a model of neocortical chattering neurons. *Neuroscience* **89**:347-362.

[130] Wessel R., Koch C., and Gabbiani F. (1996) Coding of time-varying electric field amplitude modulations in a wave-type electric fish. *J. Neurophysiol.* **75**: 2280-2293.

[131] Weyand T.G., Boudreaux M., and Guido W. (2001) Burst and tonic response

modes in thalamic neurons during sleep and wakefulness. *J. Neurophysiol.* **85**:1107-1118.

[132] Williams S.R., and Stuart G.J. (1999) Mechanisms and consequences of action potential burst firing in rat neocortical pyramidal neurons. *J. Physiol.* (London) **521**: 467-482.

[133] Xu Z., Payne J.R., and Nelson M.E. (1996) Logarithmic time course of sensory adaptation in electrosensory afferent nerve fibers in a weakly electric fish. *J. Neurophysiol.* **76**: 2020-2032.

Chapter 9

Likelihood Methods for Neural Spike Train Data Analysis

Emery N. Brown, Riccardo Barbieri, Uri T. Eden, and Loren M. Frank

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, U.S., Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology, U.S.

CONTENTS

9.1 Introduction

Computational neuroscience uses mathematical models to study how neural systems represent and transmit information. Although modeling in computational neuroscience spans a range of mathematical approaches, the discipline may be divided approximately into two schools. The first school uses detailed biophysical (Hodgkin and Huxley and their variants) models of individual neurons, networks of neurons or artificial neural network models to study emergent behaviors of neural systems.

The second school, and the one we discuss here, develops signal processing algorithms and statistical methods to analyze the ever-growing volumes of data collected in neuroscience experiments. The growing complexity of neuroscience experiments makes use of appropriate data analysis methods crucial for establishing how reliably specific system properties can be identified from experimental measurements. In particular, careful data analysis is an essential complement to neural network modeling; it allows validation of neural network model predictions in addition to feeding back biologically relevant constraints and parameter values for further analytic and simulation studies. Neuroscience experiments and neural spike train data have special features that present new, exciting challenges for statistical research.

Neuroscience data analyses as well as research on new data analysis methods should exploit established statistical paradigms wherever possible. Several standard statistical procedures, widely used in other fields of science have been slow to find their way into mainstream application in neuroscience data analysis. One such set of procedures are those based on the likelihood principle [9, 31]. The likelihood function is a central tool in statistical theory and modeling, typically based on a parametric model of an experimental data set. The likelihood is formulated by deriving the joint distribution of the data, and then viewing this joint distribution as a function of the model parameters with the data fixed. This function serves as a criterion function for estimating the model parameters, assessing goodness-of-fit, constructing confidence statements, and eventually, for making inferences about the particular problem under study. The several optimality properties of the likelihood approach is one of the main reasons this paradigm is central to statistical theory and data analysis. Neural spike trains are point process measurements. Therefore, to help better acquaint neuroscientists with likelihood-based methods, we review the likelihood paradigm for point process observations.

The remainder of this chapter is organized as follows. In Section 9.2, we show how any point process model may be characterized in terms of its conditional intensity function. The conditional intensity function is a history-dependent generalization of the rate function for a Poisson process. It provides a canonical representation of the stochastic properties of a neural spike train. We use the conditional intensity function to derive the joint probability density of the neural spike train and hence, its likelihood function. We next review briefly the optimality properties of the likelihood approach and we show how the conditional intensity function may be used to derive goodness-of-fit tests based on the time-rescaling theorem. In Section 9.3 we apply our likelihood methods in three actual data analyses. In the first example we compare the fits of exponential, gamma and inverse Gaussian interspike interval distribution models to a spike train time-series from a retinal ganglion neuron. In the second example, we use likelihood and non-likelihood based methods to analyze the spatial receptive fields of a hippocampal neuron recorded while a rat executes a behavioral task on a linear track. In the third example, we show how the likelihood function may be used to construct a criterion function for adaptive estimation that makes it possible to track plasticity in a neural receptive field on a millisecond time-scale. We illustrate the method by performing a dynamic analysis of the spatial receptive field of a hippocampal neuron from the same linear track experiment studied in the second example. Section 9.4 presents a set of conclusions.

9.2 Theory

9.2.1 The conditional intensity function and interspike interval probability density

As mentioned in Section 9.1, the key to deriving the likelihood function for a parametric model of a neural spike train is defining the joint probability density. The joint probability density of a neural spike train can be characterized in terms of the conditional intensity function. Therefore, we first derive the conditional intensity function for a point process and review some of its properties.

Let $(0,T]$ denote the observation interval and let $0 < u_1 < u_2 < \cdots < u_{J-1} < u_J \leq T$ be a set of J spike time measurements. For $t \in (0,T]$ let $N_{0:t}$ be the sample path of the point process over $(0,t]$. It is defined as the event $N_{0:t} = \{0 < u_1 < u_2 \cdots < u_j \leq t \cap N(t) = j\}$, where $N(t)$ is the number of spikes in $(0,t]$ and $j \leq J$. The sample path is a right continuous function that jumps 1 at the spike times and is constant otherwise [11, 33]. The function $N_{0:t}$ tracks the location and number of spikes in $(0,t]$ and hence, contains all the information in the sequence of spike times (Figure 9.1A).

We define the conditional intensity function for $t \in (0,T]$ as

$$\lambda(t|H_t) = \lim_{\triangle \to 0} \frac{Pr(N(t+\triangle) - N(t) = 1|H_t)}{\triangle} \tag{9.1}$$

where H_t is the history of the sample path and of any covariates up to time t. In general $\lambda(t|H_t)$ depends on the history of the spike train and therefore, it is also termed the stochastic intensity [11]. In survival analysis, the conditional intensity function is called the hazard function [20]. This is because the hazard function can be used to define the probability of an event in the interval $[t, t+\triangle)$ given that there has not been an event up to t. It follows that $\lambda(t|H_t)$ can be defined in terms of the inter-event or spike time probability density at time t, $p(t|H_t)$, as

$$\lambda(t|H_t) = \frac{p(t|H_t)}{1 - \int_0^t p(u|H_u)du} \tag{9.2}$$

We gain insight into the definition of the conditional intensity function in Equstion (9.1) by considering the following heuristic derivation of Equation (9.2) based on the definition of the hazard function. We compute explicitly the probability of the event,

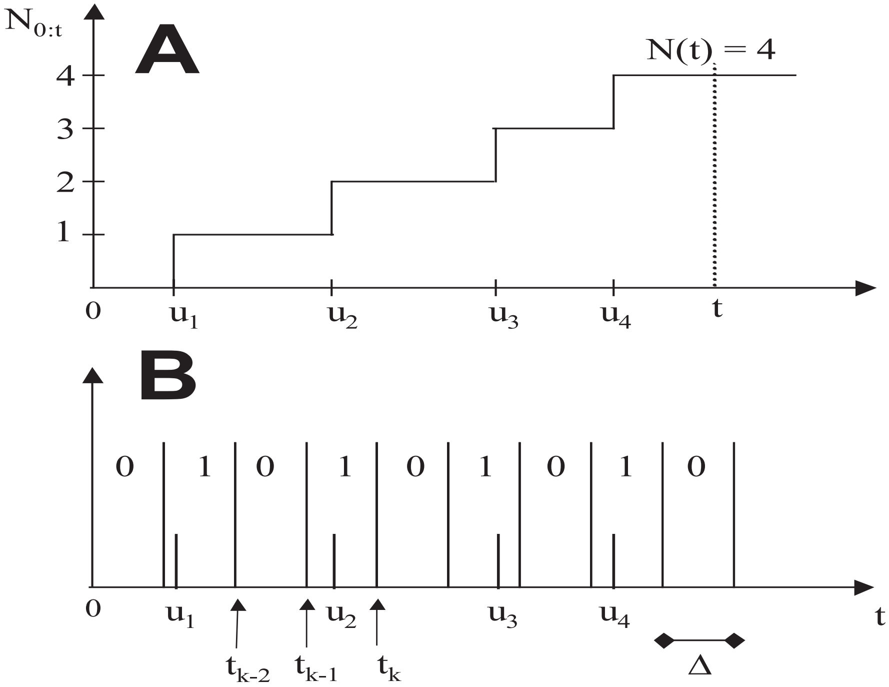

Figure 9.1

A. The construction of the sample path $N_{0:t}$ from the spike times $u_1, \cdots, u_4$. At time t, $N_{0:t} = \{u_1, u_2, u_3, u_4 \cap N(t) = 4\}$ B. The discretization of the time axis to evaluate to evaluate the probability of each spike occurrence or non-occurrence as a local Bernoulli process. By Equation (9.10) the probability of the event u_2, i.e., a 1 between t_{k-1} and t_k, is $\lambda(t_k|H_k)\triangle$ whereas the probability of the event immediately prior to u_2, i.e., a 0 between t_{k-2} and t_{k-1}, is $1 - \lambda(t_{k-1}|H_{k-1})\triangle$. In this plot we have taken $\triangle_k = \triangle$ for all $k = 1, \cdots, K$.

a spike in $[t,t+\triangle)$ given H_t and that there has been no spike in $(0,t)$. That is,

$$\begin{aligned} Pr(u\in[t,t+\triangle)|u>t,H_t) &= \frac{Pr(u\in[t,t+\triangle)\cap u>t|H_t)}{Pr(u>t|H_t)} \\ &= \frac{Pr(u\in[t,t+\triangle)|H_t)}{Pr(u>t|H_t)} \\ &= \frac{\displaystyle\int_t^{t+\triangle} p(u|H_u)du}{1-\displaystyle\int_0^t p(u|H_u)du} \\ &= \frac{p(t|H_t)\triangle}{1-\displaystyle\int_0^t p(u|H_u)du} + o(\triangle) \\ &= \lambda(t|H_t)\triangle + o(\triangle) \end{aligned} \tag{9.3}$$

where $o(\triangle)$ refers to all events of order smaller than $\triangle$, such as two or more events (spikes) occurring in an arbitrarily small interval. This establishes Equation (9.2). The power of the conditional intensity function is that if it can be defined as Equation (9.3) suggests then, it completely characterizes the stochastic structure of the spike train. In any time interval $[t,t+\triangle), \lambda(t|H_t)\triangle$ defines the probability of a spike given the history up to time t. If the spike train is an inhomogeneous Poisson process then, $\lambda(t|H_t) = \lambda(t)$ becomes the Poisson rate function. Thus, the conditional intensity function (Equation (9.1)) is a history-dependent rate function that generalizes the definition of the Poisson rate. Similarly, Equation (9.1) is also a generalization of the hazard function for renewal processes [15, 20].

We can write

$$\lambda(t|H_t) = -\frac{d\left[\log[1-\displaystyle\int_0^t p(u|H_u)du]\right]}{dt} \tag{9.4}$$

or on integrating we have

$$-\int_0^t \lambda(u|H_u)du = \log\left[1-\int_0^t p(u|H_u)du\right]. \tag{9.5}$$

Finally, exponentiating yields

$$\exp\left\{-\int_0^t \lambda(u|H_u)du\right\} = 1-\int_0^t p(u|H_u)du. \tag{9.6}$$

Therefore, by Equations. (9.2) and (9.6) we have

$$p(t|H_t) = \lambda(t|H_t)\exp\left\{-\int_0^t \lambda(u|H_u)du\right\}. \tag{9.7}$$

Together Equations (9.2) and (9.7) show that given the conditional intensity function the interspike interval probability density is specified and vice versa. Hence, defining one completely defines the other. This relation between the conditional intensity or hazard function and the inter-event time probability density is well known in survival analysis and renewal theory [15, 20]. Equations (9.2) and (9.7) show that it holds for a general point process model. This relation is exploited in the data analysis examples we discuss.

9.2.2 The likelihood function of a point process model

The likelihood of a neural spike train, like that of any statistical model, is defined by finding the joint probability density of the data. We show in the next proposition that the joint probability of any point process is easy to derive from the conditional intensity function.

Proposition 1. Given $0 < u_1 < u_2 < \cdots < u_J < T$, a set of neural spike train measurements, the sample path probability density of this neural spike train, i.e. the probability density of exactly these J events in $(0,T]$, is

$$\begin{aligned} p(N_{0:T}) &= \prod_{j=1}^{J} \lambda(u_j|H_{u_j}) \exp\left\{-\int_0^T \lambda(u|H_u)du\right\} \\ &= \exp\left\{\int_0^T \log\lambda(u|H_u)dN(u) - \int_0^T \lambda(u|H_u)du\right\}. \end{aligned} \tag{9.8}$$

Proof. Let $\{t_k\}_{k=1}^{K}$ be a partition of the observation interval $(0,T]$. Take $\triangle_k = t_k - t_{k-1}$, where $t_0 = 0$. Assume that the partition is sufficiently fine so that there is at most one spike in any $(t_{k-1}, t_k]$. For a neural spike train choosing $\triangle_k \leq 1$ msec would suffice. We define $dN(k) = 1$ if there is a spike in $(t_{k-1}, t_k]$ and 0 otherwise, and the events

$$\begin{aligned} A_k &= \{\text{spike in } (t_{k-1}, t_k] | H_k\} \\ E_k &= \{A_k\}^{dN(k)} \{A_k^c\}^{1-dN(k)} \\ H_k &= \left\{\cap_{j=1}^{k-1} E_j\right\} \end{aligned} \tag{9.9}$$

for $k = 1, \cdots, K$. In any interval $(t_{k-1}, t_k]$ we have (Figure 9.1B)

$$\begin{aligned} Pr(E_k) &= \lambda(t_k|H_k)\triangle_k + o(\triangle_k) \\ Pr(E_k^c) &= 1 - \lambda(t_k|H_k)\triangle_k + o(\triangle_k). \end{aligned} \tag{9.10}$$

By construction of the partition we must have $u_j \in (t_{k_j-1}, t_{k_j}], j = 1, \cdots, J$ for a subset of the intervals satisfying $k_1 < k_2 \cdots < k_J$. The remaining $K - J$ intervals have no spikes. The spike events form a sequence of correlated Bernoulli trials. It follows from Equation (9.10) and the Lemma in the Appendix, that given the partition, the probability of exactly J events in $(0,T]$ may be computed as

$$\begin{aligned}
p(N_{0:T})\prod_{j=1}^{J}\triangle_{k_j} &= p(u_j \in (t_{k_j-1},t_{k_j}], j=1,\cdots,J\cap N(T)=J)\prod_{j=1}^{J}\triangle_{k_j} \\
&= Pr(\cap_{k=1}^{K}E_k) \\
&= \prod_{k=2}^{K} Pr(E_k|\cap_{j=1}^{k-1}E_j)Pr(E_1) \\
&= \prod_{k=1}^{K}[\lambda(t_k|H_k)\triangle_k]^{dN(t_k)}[1-\lambda(t_k|H_k)\triangle_k]^{1-dN(t_k)}+o(\triangle_*) \\
&= \prod_{j=1}^{J}[\lambda(t_{k_j}|H_{k_j})\triangle_{k_j}]^{dN(t_{k_j})}\prod_{l\neq k_j}[1-\lambda(t_l|H_l)\triangle_l]^{1-dN(t_l)}+o(\triangle_*) \\
&= \prod_{j=1}^{J}[\lambda(t_{k_j}|H_{k_j})\triangle_{k_j}]^{dN(t_{k_j})}\prod_{l\neq k_j}\exp\{-\lambda(t_l|H_l)\triangle_l\}+o(\triangle_*) \\
&= \exp\left\{\sum_{j=1}^{J}\log\lambda(t_{k_j}|H_{k_j})dN(t_{k_j})-\sum_{l\neq k_j}\lambda(t_l|H_l)\triangle_l\right\} \\
&\quad \cdot\exp\left\{\sum_{j=1}^{J}\log\triangle_{k_j}\right\}+o(\triangle_*)
\end{aligned} \tag{9.11}$$

where, because the $\triangle_k$ are small, we have used the approximation $[1-\lambda(k)\triangle_k]\approx\exp\{-\lambda(k)\triangle_k\}$ and $\triangle_*=\max_k\triangle_k$. It follows that the probability density of exactly these J spikes in $(0,T]$ is

$$\begin{aligned}
p(N_{0:T}) &= \lim_{\triangle_*\to 0}\left[\frac{\exp\left\{\sum_{j=1}^{J}\log\lambda(t_{k_j}|H_{k_j})dN(t_{k_j})-\sum_{l\neq k_j}\lambda(t_l|H_l)\triangle_l\right\}}{\prod_{j=1}^{J}\triangle_j}\right. \\
&\qquad \left.\cdot\exp\left\{\sum_{j=1}^{J}\log\triangle_{k_j}\right\}+\frac{o(\triangle_*)}{\prod_{j=1}^{J}\triangle_j}\right] \\
&= \exp\left\{\int_0^T\log\lambda(u|H_u)dN(u)-\int_0^T\lambda(u|H_u)du\right\}. \qquad \textbf{Q.E.D.}
\end{aligned} \tag{9.12}$$

Proposition 1 shows that the joint probability density of a spike train process can be written in a canonical form in terms of the conditional intensity function [3, 8, 11]. That is, when formulated in terms of the conditional intensity function, all point process likelihoods have the form given in Equation (9.8). The approximate probability density expressed in terms of the conditional intensity function (Equation (9.11d)) was given in [5]. The proof of Proposition 1 follows the derivation in [1]. The insight provided by this proof is that correct discretization for computing the local probability of a spike event is given by the conditional intensity function. An alternative derivation of Equation (9.8) can be obtained directly using Equation (9.7) [3].

If the probability density in Equation (9.8) depends on an unknown q-dimensional parameter θ to be estimated then, Equation (9.8) viewed as a function of θ given $N_{0:T}$ is the likelihood function defined as

$$\begin{aligned} L(\theta|N_{0:T}) &= p(N_{0:T}|\theta) \\ &= \exp\left\{\int_0^T \log\lambda(u)|H_u,\theta)dN(u) - \int_0^T \lambda(u|H_u)du\right\}. \end{aligned} \tag{9.13}$$

The logarithm of Equation 9.13 is the log likelihood function defined as

$$\log L(\theta|N_{0:T}) = \int_0^T l_u(\theta)du \tag{9.14}$$

where $l_t(\theta)$ is the integrand in Equation (9.14) or the *instantaneous* log likelihood defined as

$$l_t(\theta) = \log[\lambda(t|H_t,\theta)]\frac{dN(t)}{dt} - \lambda(t|H_t,\theta). \tag{9.15}$$

Given a model for the spike train, defined either in terms of the conditional intensity function or the interspike interval probability density, the likelihood is an objective quantity that offers a measure of *rational belief* [9, 31]. Specifically, the likelihood function measures the relative preference for the values of the parameter given the observed data $N_{0:T}$. Similarly, the instantaneous log likelihood in Equation (9.15) may be viewed as measuring the instantaneous accrual of *information* from the spike train about the parameter θ. We will illustrate in the applications in Section 9.3 how methods to analyze neural spike train data may be developed using the likelihood function. In particular, we will use the instantaneous log likelihood as the criterion function in the point process adaptive filter algorithm presented in Section 9.3.3.

9.2.3 Summarizing the likelihood function: maximum likelihood estimation and Fisher information

If the likelihood is a one or two-dimensional function it can be plotted and completely analyzed for its information content about the model parameter θ. When the dimension of θ is greater than 2, a complete analysis of the likelihood function by graphical methods is not possible. Therefore, it is necessary to summarize this function. The most common way to summarize the likelihood is to compute the maximum likelihood estimate of the parameter θ. That is, we find the value of this parameter that is most likely given the data. This corresponds to the value of θ that makes Equation (9.13) or equivalently, Equation (9.14) as large as possible. We define the maximum likelihood estimate $\hat{\theta}$ as

$$\hat{\theta} = \arg\max_\theta L(\theta|N_{0:T}) = \arg\max_\theta \log L(\theta|N_{0:T}). \tag{9.16}$$

With the exception of certain elementary models the value of θ that maximizes Equation (9.16) has to be computed numerically. In most multidimensional problems it is difficult to insure that the numerical analysis will yield a global maximum. Most often a local rather than a global estimate is obtained. Several different starting values of parameters should be used in the numerical optimization procedure to increase the probability of obtaining a global maximum of the likelihood.

A second standard statistic computed to summarize the likelihood is the Fisher Information. The Fisher Information is defined as

$$I(\theta) = -E[\nabla^2 \log L(\theta|N_{0:T})], \tag{9.17}$$

where ∇^2 is the Hessian of the log likelihood with respect to θ and E denotes the expectation taken with respect to $p(N_{0:T}|\theta)$. The Fisher Information matrix can be used to measure the uncertainty in the maximum likelihood estimate. This is because under not too stringent regularity conditions, the asymptotic (large sample) distribution of the maximum likelihood estimate $\hat{\theta}$ is the Gaussian distribution whose mean is the true value of the parameter θ, and whose covariance matrix is $I(\theta)^{-1}$ [9, 31]. Because θ is unknown we evaluate $I(\theta)$ as $I(\hat{\theta})$ or $I(\hat{\theta})_{N_{0:T}} = -\nabla^2 \log L(\hat{\theta}|N_{0:T})$ where the latter is termed the observed Fisher information. Under the Gaussian approximation to the asymptotic distribution of the maximum likelihood estimate, the Fisher information may be used to compute confidence intervals for the components of the true parameter vector given by

$$\hat{\theta}_i \pm z_{1-\alpha/2}[I(\hat{\theta}_i)]^{-1/2}, \tag{9.18}$$

where $\hat{\theta}_i$ is the i^{th} component of $\hat{\theta}$ for $i = 1, \cdots, q$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard Gaussian distribution.

Another way of viewing the maximum likelihood estimate along with the Fisher information is as a means of constructing a Gaussian approximation to the likelihood function. By expanding $\log L(\theta|N_{0:T})$ in a Taylor series about $\hat{\theta}$ we obtain the following Gaussian approximation to $L(\theta|N_{0:T})$ namely,

$$L(\theta|N_{0:T}) \sim [(2\pi)^q |I(\hat{\theta})|]^{-1/2} \exp\left\{ -\frac{(\theta-\hat{\theta})^T I(\hat{\theta})^{-1}(\theta-\hat{\theta})}{2} \right\}. \tag{9.19}$$

While Equation (9.19) is functionally equivalent to the statement that the maximum likelihood estimate has an approximate Gaussian probability density, this equation has a Bayesian interpretation. This is because in the classical or frequentist statement that the asymptotic distribution of the maximum likelihood estimate is Gaussian, $\hat{\theta}$ is a random variable and θ, the true parameter value is a fixed quantity. In Equation (9.19) $\hat{\theta}$ and $I(\hat{\theta})$ are fixed quantities and θ is a random variable [31, 35].

9.2.4 Properties of maximum likelihood estimates

One of the most compelling reasons to use maximum likelihood estimation in neural spike train data analyses is that for a broad range of models, these estimates have other important optimality properties in addition to being asymptotically Gaussian.

First, there is consistency which states that the sequence of maximum likelihood estimates converges in probability (or more strongly almost surely) to the true value as the sample size increases. Second, the convergence in probability of the estimates means that they are asymptotically unbiased. That is, the expected value of the estimate $\hat{\theta}$ is θ as the sample size increases. For some models and some parameters, unbiasedness is a finite sample property. The third property is invariance. That is, if $\hat{\theta}$ is the maximum likelihood estimate of θ, then $\tau(\hat{\theta})$ is the maximum likelihood estimate of $\tau(\theta)$. Finally, the maximum likelihood estimates are asymptotically efficient in that as the sample size increases, the variance of the maximum likelihood estimate achieves the Cramer-Rao lower bound. This lower bound defines the smallest variance that an unbiased or asymptotically unbiased estimate can achieve. Like unbiasedness, efficiency for some models and some parameters is achieved in a finite sample. Detailed discussions of these properties are given in [9, 31].

9.2.5 Model selection and model goodness-of-fit

In many data analyses it is necessary to compare a set of models for a given neural spike train. For models fit by maximum likelihood, a well-known approach to model selection is the Akaike's Information Criterion (AIC) [31]. The criterion is defined as

$$AIC = -2\log L(\hat{\theta}|N_{0:T}) + 2q, \tag{9.20}$$

where q is the dimension of the parameter vector θ. The AIC measures the trade-off between how well a given model fits the data and the number of model parameters needed to achieve this fit. The fit of the model is measured by the value of $-2x$ the maximized likelihood and the cost of the number of fitted parameters is measured by $2q$. Under this formulation, i.e. considering the negative of the maximized likelihood, the model that describes the data best in terms of this trade-off will have the smallest AIC.

Evaluating model goodness-of-fit, i.e., measuring quantitatively the agreement between a proposed model and a spike train data series, is a more challenging problem than for models of continuous-valued processes. Standard distance discrepancy measures applied in continuous data analyses, such as the average sum of squared deviations between recorded data values and estimated values from the model, cannot be directly computed for point process data. Berman [4] and Ogata [28] developed transformations that, under a given model, convert point processes like spike trains into continuous measures in order to assess model goodness-of-fit. One of the transformations is based on the time-rescaling theorem.

A form of the time-rescaling theorem is well known in elementary probability theory. It states that any inhomogeneous Poisson process may be rescaled or transformed into a homogeneous Poisson process with a unit rate [36]. The inverse transformation is a standard method for simulating an inhomogeneous Poisson process from a constant rate (homogeneous) Poisson process. Meyer [26] and Papangelou [30] established the general time-rescaling theorem, which states that any point process with an integrable rate function may be rescaled into a Poisson process with a unit rate. Berman and Ogata derived their transformations by applying the general form of the theorem. An elementary proof of the time-rescaling theorem is given in [8].

We use the time-rescaling theorem to construct goodness-of-fit tests for a neural spike data model. Once a model has been fit to a spike train data series we can compute from its estimated conditional intensity the rescaled times

$$\tau_j = \int_{u_{j-1}}^{u_j} \lambda(u|H_u)du, \tag{9.21}$$

for $j = 1, \cdots, J$. If the model is correct then, according to the theorem, the τ_js are independent exponential random variables with mean 1. If we make the further transformation

$$z_j = 1 - \exp(-\tau_j), \tag{9.22}$$

then z_js are independent uniform random variables on the interval [0,1). Because the transformations in Eqs. (9.21) and (9.22) are both one-to-one, any statistical assessment that measures agreement between the z_js and a uniform distribution directly evaluates how well the original model agrees with the spike train data. We use Kolmogorov-Smirnov tests to make this evaluation [8].

To construct the Kolmogorov-Smirnov test we order the z_js from smallest to largest, denoting the ordered values as $z_j s$ and then plot the values of the cumulative distribution function of the uniform density defined as $b_j = (j - 1/2)/J$ for $j = 1, \cdots, J$ against the z_js. If the model is correct, then the points should lie on a 45^o line. Confidence bounds for the degree of agreement between the models and the data may be constructed using the distribution of the Kolmogorov-Smirnov statistic [19]. For moderate to large sample sizes the 95% confidence bounds are well approximated as $b_j \pm 1.36/J^{1/2}$ [19]. We term these plots Kolmogorov-Smirnov (KS) plots.

9.3 Applications

9.3.1 An analysis of the spiking activity of a retinal neuron

In this first example we study a spike train data series from a goldfish retinal ganglion cell neuron recorded in vitro (Figure 9.2). The data are 975 spikes recorded over 30 seconds from neuron 78 in [18]. They were provided by Dr. Satish Iyengar from experiments originally conducted by Dr. Michael Levine at the University of Illinois [22, 23]. The retinae were removed from the goldfish and maintained in a flow of moist oxygen. Recordings of retina ganglion cells were made with an extracellular microelectrode under constant illumination.

The plot of the spikes from this neuron (Figure 9.2) reveals a collection of short and long interspike intervals (ISI). To analyze these data we consider three ISI probability models: the gamma, exponential and inverse Gaussian probability densities.

The gamma probability density is a probability model frequently used to describe renewal processes. It is the ISI probability model obtained from a simple stochastic integrate-and-fire model in which the inputs to the neuron are Poisson with a constant rate [37]. It has the exponential probability density, the interspike interval model associated with a simple Poisson process, as a special case. The inverse Gaussian probability density is another renewal process model that can be derived from a stochastic integrate-and-fire model in which the membrane voltage is represented as a random walk with drift [37]. This model was first suggested by Schroedinger [32] and was first applied in spike train data analyses by Gerstein and Mandelbrot [14]. Because the gamma and inverse Gaussian ISI probability densities can be derived from elementary stochastic integrate-and-fire models, these probability densities suggest more plausible points of departure for constructing statistical models of neural spike trains than the Poisson process.

The gamma and inverse Gaussian probability densities are, respectively,

$$p_1(w_j|\theta) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} w_j^{\alpha-1} \exp\{-\lambda w_j\}, \tag{9.23}$$

where $\theta = (\alpha, \lambda), \alpha > 0, \lambda > 0,$

$$p_2(w_j|\theta) = \left(\frac{\lambda}{2\pi w_j^3}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{\lambda(w_j-\mu)^2}{\mu^2 w_j}\right\}, \tag{9.24}$$

where $\theta = (\mu, \lambda), \mu > 0, \lambda > 0$ and $w_j = u_j - u_{j-1}$ for $j = 1, \cdots, J$. For the gamma (inverse Gaussian) model $\alpha(\mu)$ is the location parameter and $\lambda(\lambda)$ is the scale parameter. If $\alpha = 1$ then the gamma model is the exponential probability density. The mean and variance of the gamma model are respectively $\alpha\lambda^{-1}$ and $\alpha\lambda^{-2}$ whereas the mean and variance of the inverse Gaussian model are respectively μ and $\mu^3\lambda^{-1}$. Fitting these models to the spike train data requires construction of the likelihoods and estimation of θ for the three models. By our results in Section 9.2, the log likelihood can be represented either in terms of the conditional intensity or the ISI probability model. Here we use the latter. Given the set of ISIs, $w = (w_1, \cdots, w_J)$, then, under the assumption that the ISIs are independent, the likelihood functions for the two models are respectively

$$L_1(\theta|w) = \prod_{j=1}^{J} p_1(w|\theta) = \left[\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right]^J \prod_{j=1}^{J} w_j^{\alpha-1} \exp\{-\lambda w_j\} \tag{9.25}$$

$$L_2(\theta|w) = \prod_{j=1}^{J} p_2(w|\theta) = \prod_{j=1}^{J} \left(\frac{\lambda}{2\pi w_j^3}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{\lambda(w_j-\mu)^2}{\mu^2 w_j}\right\}. \tag{9.26}$$

The maximum likelihood estimate of θ for the gamma model cannot be computed in

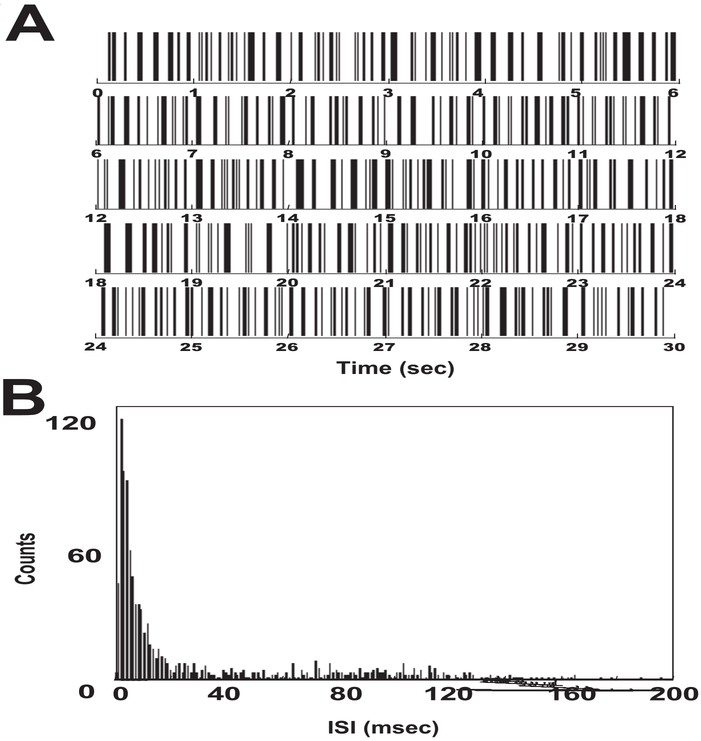

Figure 9.2

A. Thirty seconds of spike times from a retinal ganglion neuron recorded *in vitro* under constant illumination. There is an obvious mixture of short and long interspike intervals. B. Interspike interval histogram for the neural spike train in A. While most of the spikes occur between 3 to 40 msec, there are many intervals longer than 70 msec.

closed form, but rather numerically as the solution to the equations

$$\hat{\lambda} = \frac{\hat{\alpha}}{\bar{w}} \tag{9.27}$$

$$J \log \Gamma(\hat{\alpha}) = J\alpha \log\left(\frac{\hat{\alpha}}{\bar{w}}\right) + (\hat{\alpha} - 1)\sum_{j=1}^{J} \log w_j \tag{9.28}$$

where $\bar{w} = J^{-1}\sum_{j=1}^{J} w_j$ and the $\hat{}$ denotes the estimate. Equations (9.27) and (9.28) are obtained by differentiating Equation (9.25) with respect to α and λ and setting the derivatives equal to zero. It follows from Equation (9.17) that the Fisher Information matrix is

$$I(\theta) = J\begin{bmatrix} \dfrac{\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)\Gamma'(\alpha)}{\Gamma^2(\alpha)} & -\dfrac{1}{\lambda} \\ -\dfrac{1}{\lambda} & \dfrac{\alpha}{\lambda^2} \end{bmatrix}. \tag{9.29}$$

Similarly, differentiating Equation (9.26) with respect to μ and λ and setting the derivatives equal to zero yields as the maximum likelihood estimate of the inverse Gaussian model parameters

$$\hat{\mu} = J^{-1}\sum_{j=1}^{J} w_j \tag{9.30}$$

$$\hat{\lambda}^{-1} = J^{-1}\sum_{j=1}^{J}(w_j^{-1} - \hat{\mu}^{-1}). \tag{9.31}$$

On evaluating Equation (9.17) for this model we find that the Fisher Information matrix is

$$I(\theta) = J\begin{bmatrix} \lambda\mu^{-3} & 0 \\ 0 & (2\lambda^2)^{-1} \end{bmatrix}. \tag{9.32}$$

We compare the fits to the spike train of the exponential, gamma and inverse Gaussian models by comparing the model probability density estimate for each to the normalized histogram of the ISIs (Figure 9.3). The exponential model underpredicts the number of short ISIs ($<$ 10 msec), overpredicts the number of intermediate ISIs (10 to 50 msec) (Figure 9.3B), and underpredicts the number of long ISIs ($>$ 120 msec), (Figure 9.3C). While the gamma model underpredicts the number of short ISIs, ($<$ 10 msec) more than the exponential model, it predicts well the number of intermediate ISIs (10 to 50 msec) (Figure 9.3B), and also underpredicts the number of long ISIs ($>$ 120 msec), (Figure 9.3C). Because the gamma model estimate of α is $\hat{\alpha} = 0.81$ (Table 1), the mode of this probability density lies at zero where the probability density is infinite. Zero lies outside the domain of the probability density as it is defined only for ISIs that are strictly positive. This explains the monotonically decreasing shape of this probability density seen in the plot. Although not completely accurate, the inverse Gaussian model gives the best fit to the short ISIs (Figure 9.3B). The inverse Gaussian model also describes well the intermediate ISIs

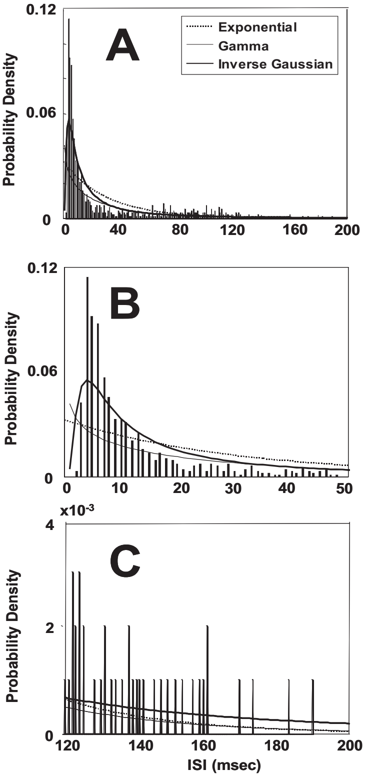

Figure 9.3

A. Maximum likelihood fits of the exponential (dotted line), gamma (solid line), inverse Gaussian (solid bold line) models to the retinal neuron spike trains in Figure 9.2A displayed superimposed on a normalized version of the interspike interval histogram in Figure 9.2B. B. Enlargement from (A) of the interspike interval histogram from 0 to 50 msec to display better the data, and the three model fits over this range. C. Enlargement from (A) of the interspike interval histogram from 120 to 200 msec and the three model fits.

from 25 to 50 msec (Figure 9.3B) and of the three models, underpredicts the long ISIs the least (Figure 9.3C).

Because of Equation (9.2), specifying the spike time probability density is equivalent to specifying the conditional intensity function. From Equation (9.2) and the invariance of the maximum likelihood estimate discussed in Section 9.2.5, it follows that if $\hat{\theta}$ denotes the maximum likelihood estimate of θ then the maximum likelihood estimate of the conditional intensity function for each model can be computed from Equation (9.2) as

$$\lambda(t|H_t,\hat{\theta}) = \frac{p(t|u_{N(t)},\hat{\theta})}{1 - \int_{u_{N(t)}}^{t} p(u|u_{N(t)},\hat{\theta})du}, \tag{9.33}$$

for $t > u_{N(t)}$ where $u_{N(t)}$ is the time of the last spike prior to t. The estimated conditional intensity from each model may be used in the time-rescaling theorem to assess model goodness-of-fit as described in Section 9.2.5.

An important advantage of the KS plot is that it allows us to visualize the goodness-of-fit of the three models without having to discretize the data into a histogram (Figure 9.4). While the KS plot for neither of the three models lies entirely within the 95% confidence bounds, the inverse Gaussian model is closer to the confidence bounds over the entire range of the data. These plots also show that the gamma model gives a better fit to the data than the exponential model.

The AIC and KS distance are consistent with the KS plots (Table 1). The inverse Gaussian model has the smallest AIC and KS distance, followed by the gamma and exponential models in that order for both. The approximate 95% confidence interval for each model parameter was computed from maximum likelihood estimates of the parameters (Equations (9.27), (9.28), (9.30), (9.31)) and the estimated Fisher information matrices (Equations (9.29), (9.32)) using Equation (9.18). Because none of the 95% confidence intervals includes zero, all parameter estimates for all three models are significantly different from zero. While all three models estimate the mean ISI as 30.73 msec, the standard deviation estimate from the inverse Gaussian model of 49.0 msec is more realistic given the large number of long ISIs (Figure 9.2B).

In summary, the inverse Gaussian model gives the best overall fit to the retinal ganglion spike train series. This finding is consistent with the results of [18] who showed that the generalized inverse Gaussian model describes the data better than a lognormal model. Our inverse Gaussian model is the special case of their generalized inverse Gaussian model in which the index parameter of the Bessel function in their normalizing constant equals −0.5. In their analyses Iyengar and Liao estimated the index parameter for this neuron to be −0.76. The KS plots suggests that the model fits can be further improved. The plot of the spike train data series (Figure 9.2) suggests that the fit may be improved by considering an ISI model that would specifically represent the obvious propensity of this neuron to burst as well as produce long ISIs in a history-dependent manner. Such a model could be derived as the mixture model that Iyengar and Liao [18] suggest as a way of improve their generalized inverse Gaussian fits.

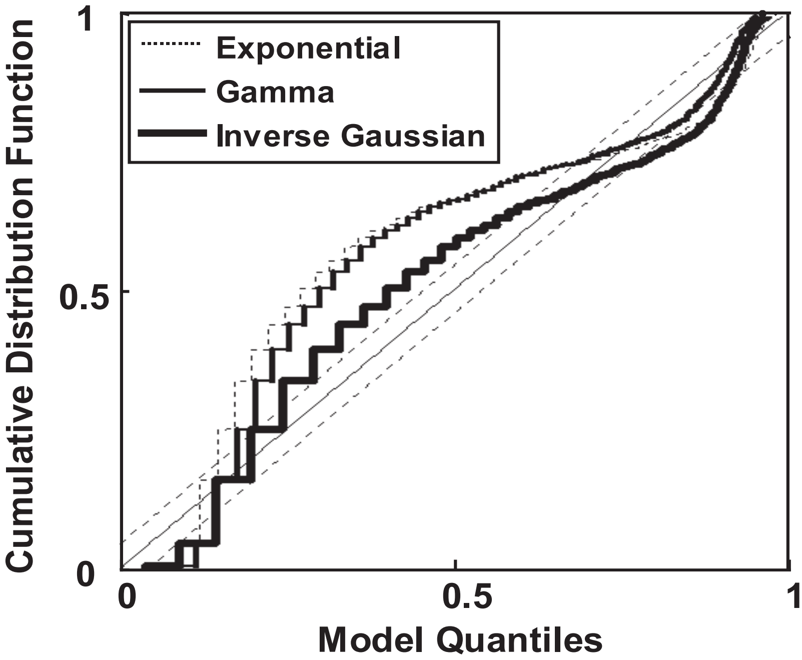

Figure 9.4

Kolmogorov-Smirnov plots for the fits of the exponential (dotted line), gamma (solid line), and inverse Gaussian (solid bold line) models to the neural spike train in Figure 9.2. The parallel diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data. By this criterion, statistically acceptable agreement between a model and the data would be seen if the KS plot for that model fell entirely within the confidence bounds.

	Exponential	Gamma		Inverse Gaussian	
	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\lambda}$
$\hat{\theta}$	0.0325	0.805	0.262	30.76	12.1
CI	[0.0283 0.0367]	[0.678 0.931]	[0.208 0.316]	[24.46 37.06]	[9.9 14.3]
ISI	30.77±30.77	30.73± 34.74		30.76± 49.0	
AIC	8598	8567		8174	
KS	0.233	0.2171		0.1063	

Table 1: The row is the maximum likelihood estimate $\hat{\theta}$. CI (95% confidence interval for the parameter); ISI (interspike interval mean and SD); AIC (Akaike's Information Criterion); KS (Kolmogorov-Smirnov statistic).

9.3.2 An analysis of hippocampal place-specific firing activity

As a second example of applying likelihood methods, we analyze the spiking activity of a pyramidal cell in the CA1 region of the rat hippocampus recorded from an animal running back and forth on a linear track. Hippocampal pyramidal neurons have place-specific firing [29]. That is, a given neuron fires only when the animal is in a certain sub-region of the environment termed the neuron's place field. Because of this property these neurons are often called place cells. The neuron's spiking activity correlates most closely with the animal's position on the track ([38]). On a linear track these fields approximately resemble one-dimensional Gaussian functions. The data series we analyze consists of 4,265 spikes from a place cell in the CA1 region of the hippocampus recorded from a rat running back and forth for 1200 seconds on a 300-cm U-shaped track. In Figure 9.5, we show the linearized track plotted in time so that the spiking activity during the first 400 seconds can be visualized on a pass-by-pass basis [12]. We compare two approaches to estimating the place-specific firing maps of a hippocampal neuron. In the first, we use maximum likelihood to fit a specific parametric model of the spike times to the place cell data as in [3, 6, 8]. In the second approach we compute a histogram-based estimate of the conditional intensity function by using spatial smoothing of the spike train [12, 27]. The analysis presented here parallels the analysis performed in [8].

We let $x(t)$ denote the animal's position at time t, we define the spatial function for the one-dimensional place field model as the Gaussian function

$$s(t) = \exp\left\{\alpha - \frac{(x(t)-\mu)^2}{2\sigma^2}\right\}, \tag{9.34}$$

where μ is the center of the place field, σ^2 is a scale factor, and $\exp(\alpha)$ is the maximum height of the place field at its center. We define the spike time probability density of the neuron as either the inhomogeneous gamma (IG) model

$$p(u_j|u_{j-1},\theta) = \frac{\psi s(u_j)}{\Gamma(\psi)}\left[\int_{u_{j-1}}^{u_j} \psi s(u)du\right]^{\psi-1} \exp\left\{-\int_{u_{j-1}}^{u_j} \psi s(u)du\right\}, \tag{9.35}$$

or as the inhomogeneous inverse Gaussian (IIG) model

$$p(u_j|u_{j-1},\theta) = \frac{s(u_j)}{\left[2\pi\left[\int_{u_{j-1}}^{u_j} s(u)du\right]^3\right]^{1/2}} \exp\left\{-\frac{1}{2}\frac{\left(\int_{u_{j-1}}^{u_j} s(u)du-\psi\right)^2}{\psi^2\int_{u_{j-1}}^{u_j} s(u)du}\right\} \tag{9.36}$$

for $j = 1,\cdots,J$, where $\psi > 0$ is a location parameter for both models and $\theta = (\mu,\alpha,\sigma^2,\psi)$ is the set of model parameters to be estimated from the spike train. If we set $\psi = 1$ in Equation (9.35) we obtain the inhomogeneous Poisson (IP) model as a special case of the IG model. The models in Equations (9.35) and (9.36) are Markov because the current value of either the spike time probability density or the conditional intensity (rate) function (see Equation (9.3)) depends only on the time

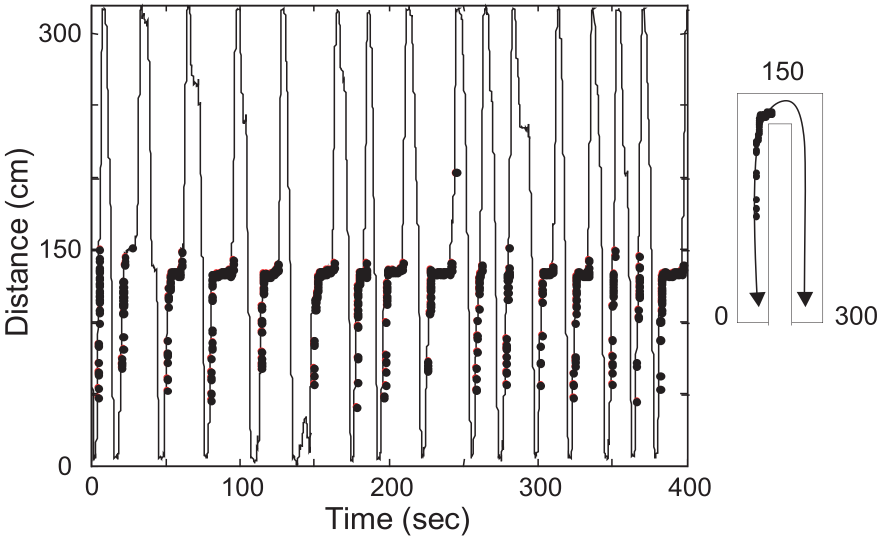

Figure 9.5

Place-specific spiking activity of a hippocampal pyramidal neuron recorded from a rat running back and forth for 400 sec of a 1200 sec experiment on a 300 cm U-shaped track (outset on the right). The track has been linearized and plotted in time so that the spiking activity on each pass can be visualized. The black dots show the spatial locations at which the neuron discharged a spike. The place field of this neuron extends from approximately 50 to 150 cms. In addition to having place-specific firing, this neuron is directional in that it spikes only as the animal moves from bottom to top (from left to right in the outset) between 50 to 150 cms.

of the previous spike. The IP, IG and IIG models are inhomogeneous analogs of the simple Poisson (exponential), gamma and inverse Gaussian models discussed in Section 9.3.1. These inhomogeneous models allow the spiking activity to depend on a temporal covariate, which, in this case, is the position of the animal as a function of time.

The parameters for all three models, the IP, IG and the IIG can be estimated from the spike train data by maximum likelihood [3, 8]. The log likelihoods for these two models have the form

$$\log L(\theta|N_{0:T}) = \sum_{j=1}^{J} \log p(u_j|u_{j-1},\theta) \tag{9.37}$$

To compute the spatial smoothing estimate of the conditional intensity function, we proceed as in [8]. We divide the 300 cm track into 4.2 cm bins, count the number of spikes per bin, and divide the count by the amount of time the animal spends in the bin. We smooth the binned firing rate with a six-point Gaussian window with a standard deviation of one bin to reduce the effect of running velocity [12]. The smoothed spatial rate function is the spatial conditional intensity estimate. The spatial smoothing procedure yields a histogram-based estimate of $\lambda(t)$ for a Poisson process because the estimated spatial function makes no history dependence assumption about the spike train. The IP, IG and IIG models were fit to the spike train data by maximum likelihood whereas the spatial rate model was computed as just described. As in Section 9.3.1 we use the KS plots to compare directly goodness-of-fit of the four models of this hippocampal place cell spike train.

The smoothed estimate of the spatial rate function and the spatial components of the rate functions for the IP, IG and IIG models are shown in Figure 9.6. The smoothed spatial rate function most closely resembles the spatial pattern of spiking seen in the raw data (Figure 9.5). While the spiking activity of the neuron is confined between approximately 50 and 150 cms, there is, on nearly each upward pass along the track, spiking activity between approximately 50 to 100 cms, a window of no spiking between 100 to 110 or 125 cms and then, a second set of more intense spiking activity between 125 to 150 cms. These data features are manifested as a bimodal structure in the smoothed estimate of the spatial rate function. The first mode is 10 spikes/sec and occurs at 70 cms, whereas the second mode is approximately 27 spikes/sec and occurs at approximately 120 cms. The spatial components of the IP and IG models were identical. Because of Equation (9.34), this estimate is unimodal and suggests a range of non-zero spiking activity which is slightly to the right of that estimated by the smoothed spatial rate function. The mode of the IP/IG model fits is 20.5 spikes/sec and occurs at approximately 110 cms. The IIG spatial component is also unimodal by virtue of Equation (9.34). It has its mode of 20.5 at 114 cms. This model significantly overestimates the width of the place field as it extends from 0 to 200 cm. The scale parameter, σ, is 23 cm for the IG model and 43 cm for the IIG model.

For only the IP and the smoothed spatial rate models do the curves in Figure 9.6 represent a spatial rate function. For the IG and IIG models the rate function, or

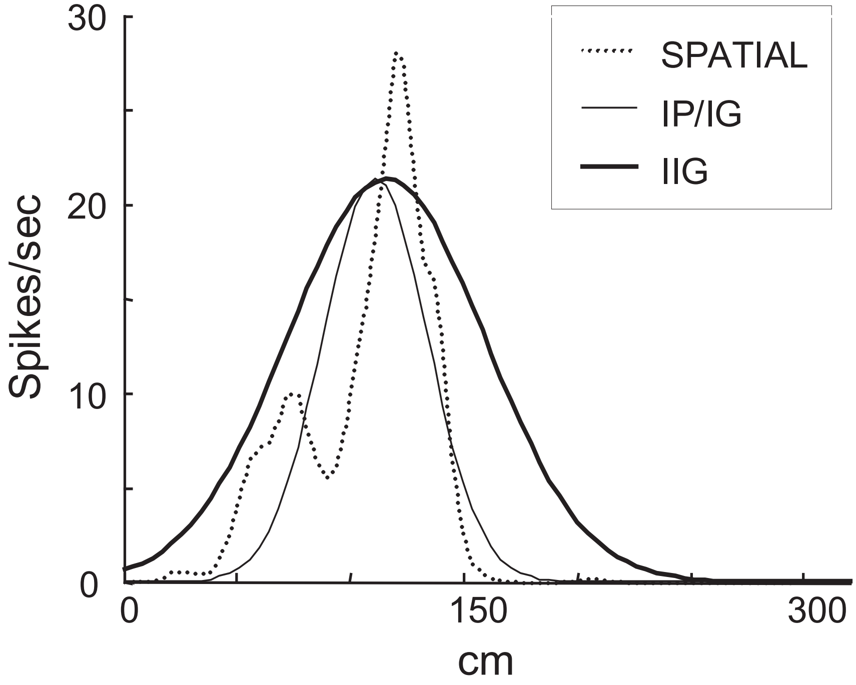

Figure 9.6

The place field estimates derived from the spatial smoothing model (dotted line), and the maximum likelihood fits of the inhomogeneous Poisson (IP) (thin solid line), inhomogeneous gamma (IG) (thin solid line), and inhomogeneous inverse Gaussian (IIG) models (thick solid line). The units of spikes/sec only apply to the spatial and IP model fits. For the IG and the IIG models the graphs show the spatial components of the rate models. The conditional intensity (rate) function for these two models is obtained from Equation (9.2) in the way Equation (9.33) was used to compute this quantity in Section 9.3.1.

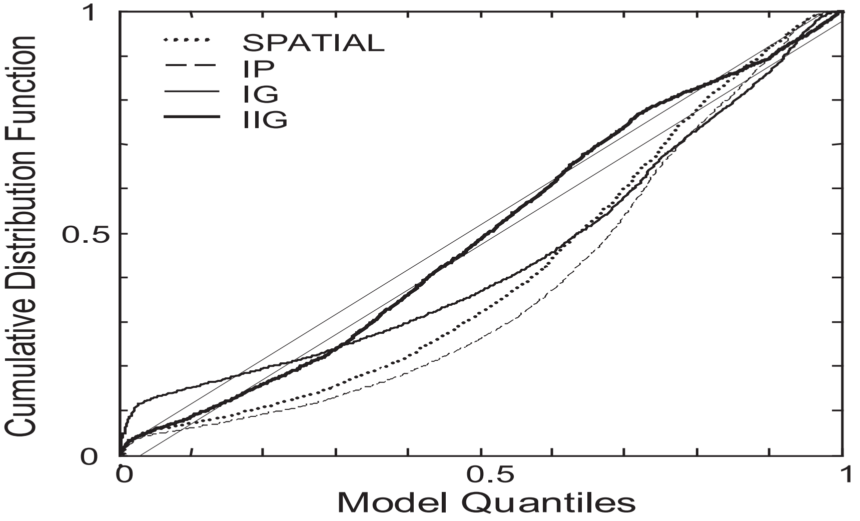

Figure 9.7

Kolmogorov-Smirnov plots of the spatial smoothing model (dotted line), and the maximum likelihood fits of the inhomogeneous Poisson (IP) (dashed line), inhomogeneous gamma (IG) (thin solid line), and inhomogeneous inverse Gaussian (IIG) (thick solid) models. As in Figure 9.4, the parallel diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data.

conditional intensity function, is computed from Equation (9.2) using the maximum likelihood estimate of θ. For both of these models this equation defines a spatio-temporal rate function whose spatial component is defined by Equation (9.34). This is why the spatial components of these models are not the spatial rate function. For the IP model Equation (9.2) simplifies to Equation (9.34). The smoothed spatial rate model makes no assumption about temporal dependence, and therefore, it implicitly states that its estimated rate function is the rate function of a Poisson process.

The KS plot goodness-of-fit comparisons are shown in Figure 9.7. The IG model overestimates at lower quantiles, underestimates at intermediate quantiles, and lies within the 95% confidence bounds at the upper quantiles (Figure 9.7). The IP model underestimates the lower and intermediate quantiles, and like the IG model, lies within the 95% confidence bounds in the upper quantiles. The KS plot of the spatial rate model is similar to that of the IP model, yet closer to the confidence bounds. This analysis suggests that the IG, IP and spatial rate models are most likely over-smoothing this spike train because underestimating the lower quantiles corresponds to underestimating the occurrence of short ISIs [3]. The fact that the IP and IG models estimate a different temporal structure for this spike train data series is evidenced by the fact that while they both have the same spatial model components (Figure 9.6), Their KS plots differ significantly (Figure 9.7). This difference is due entirely to the fact that $\hat{\psi} = 0.61$ for the IG model whereas for the IP model $\psi = 1$ by assumption. The IP, the IG, and the smoothed spatial rate function have the greatest lack of fit in that order. Of the 4 models, the IIG is the one that is closest to the confidence bounds. Except for an interval around the 0.30 quantile where this model underestimates these quantiles, and a second interval around the 0.80 quantile where it overestimates these quantiles, the KS plot of the IIG model lies within the 95% confidence bounds.

The findings from the analysis of the spatial model fits (Figure 9.6) appear to contradict the findings of the overall goodness-of-fit analysis (Figure 9.7). The IIG gives the poorest description of the spatial structure in the data yet, the best overall description in terms of the KS plot. The smoothed spatial rate function model seems to give the best description of the data's spatial structure however, its overall fit is one of the poorest. To reconcile the findings, we first note that the better overall fit of the smoothed spatial rate function relative to the IP (Figure 9.7) is to be expected because the IP estimates the spatial component of a Poisson model with a three parameter model that must have a Gaussian shape. The smoothed spatial rate function model, on the other hand, uses a smoothed histogram that has many more parameters to estimate the spatial component of the same Poisson model. The greater flexibility of the smoothed spatial rate function allows it to estimate a bimodal structure in the rate function. Both the IP and smoothed spatial rate models use an elementary temporal model in that both assume that the temporal structure in the data is Poisson. For an inhomogeneous Poisson model the counts in non-overlapping intervals are independent whereas the interspike interval probability density is Markov (Equation (9.35)). The importance of correctly estimating the temporal structure is also seen in comparing the IP and IG model fits. These two models have identical spatial components yet, different KS plots because $\hat{\psi} = 0.61$ for the IG and $\psi = 1$ for the IP model by assumption. The KS plots suggest that while the IIG model does not describe the spatial component of the data well, its better overall fit comes because it does a better job at describing the temporal structure in the spike train. In contrast, the smoothed spatial rate function fits exclusively the spatial structure in the data to the exclusion of the temporal structure.

In summary, developing an accurate model of the place-specific firing activity of this hippocampal neuron requires specifying correctly both its spatial and temporal components. Our results suggest that combining a flexible spatial model, such as in the smoothed spatial rate function model with non-Poisson temporal structure as in the IG and IIG models, should be a way of developing a more accurate description. Another important consideration for hippocampal pyramidal neurons is that place-specific firing does not remain static. The current models would not capture this dynamic behavior in the data. In the next example we analyze a place cell from this same experiment using a point process adaptive filter algorithm to estimate the dynamics of the place cell spatial receptive fields using a model with flexible spatial and temporal structures.

9.3.3 An analysis of the spatial receptive field dynamics of a hippocampal neuron

The receptive fields of neurons are dynamic in that their responses to relevant stimuli change with experience. This plasticity, or experience-dependent change, has been established in a number of brain regions. In the rat hippocampus the spatial receptive fields of the CA1 pyramidal neurons evolve through time in a reliable manner as an animal executes a behavioral task. When, as in the previous example, the experimental environment is a linear track, these spatial receptive fields tend to migrate and skew in the direction opposite the cell's preferred direction of firing relative to the animal's movement, and increase in both maximum firing rate and scale [24, 25]. This evolution occurs even when the animal is familiar with the environment. As we suggested in Section 9.3.2, this dynamic behavior may contribute to the failure of the models considered there to describe the spike train data completely.

We have shown how the plasticity in neural receptive fields can be tracked on a millisecond time-scale using point process adaptive filter algorithms [7, 13]. Central to the derivation of those algorithms were the conditional intensity function (Equation (9.1)) and hte instantaneous log likelihood function (Equation (9.15)). We review briefly the derivation of the point process adaptive filter and illustrate its application by analyzing the spatial receptive field dynamics of a second pyramidal neuron from the linear track experiment discussed in Section 9.3.2.

To derive our adaptive point process filter algorithm we assume that the q-dimensional parameter θ in the instantaneous log likelihood (Equation (9.15)) is time varying. We choose K large, and divide $(0,T]$ into K intervals of equal width $\triangle = T/K$, so that there is at most one spike per interval. The adaptive parameter estimates will be updated at $k\triangle$. A standard prescription for constructing an adaptive filter algorithm to estimate a time-varying parameter is instantaneous steepest descent [17, 34] . Such an algorithm has the form

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial J_k(\theta)}{\partial \theta}|_{\theta=\hat{\theta}_{k-1}} \tag{9.38}$$

where $\hat{\theta}_k$ is the estimate at time $k\triangle$, $J_k(\theta)$ is the criterion function at $k\triangle$, and ε is a positive learning rate parameter. If for continuous-valued observations $J_k(\theta)$ is chosen to be a quadratic function of θ then, it may be viewed as the instantaneous log likelihood of a Gaussian process. By analogy, the instantaneous steepest descent algorithm for adaptively estimating a time-varying parameter from point process observations can be constructed by substituting the instantaneous log likelihood from Equation (9.15) for $J_k(\theta)$ in Equation (9.38). This gives

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial l_k(\theta)}{\partial \theta}|_{\theta=\hat{\theta}_{k-1}} \tag{9.39}$$

which, on rearranging terms, gives the instantaneous steepest descent adaptive filter algorithm for point process measurements

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial \log \lambda(k\triangle|H_k, \hat{\theta}_{k-1})}{\partial \theta} \left[dN(k\triangle) - \lambda(k\triangle|H_k, \hat{\theta}_{k-1})\triangle\right]. \tag{9.40}$$

Equation (9.40) shows that the conditional intensity function completely defines the instantaneous log likelihood and therefore, a point process adaptive filtering algorithm using instantaneous steepest descent. The parameter update $\hat{\theta}_k$ at $k\triangle$ is the previous parameter estimate $\hat{\theta}_{k-1}$ plus a dynamic gain coefficient,

$$-\frac{\varepsilon\partial \log \lambda(k\triangle|H_k,\hat{\theta}_{k-1})}{\partial\theta},$$

multiplied by an innovation or error signal $[dN(k\triangle) - \lambda(k\triangle|H_k,\hat{\theta}_{k-1})\triangle]$. The error signal provides the new information coming from the spike train and it is defined by comparing the predicted probability of a spike, $\lambda(k\triangle|\hat{\theta}_{k-1})\triangle$, at $k\triangle$ with $dN(k\triangle)$, which is 1 if a spike is observed in $((k-1)\triangle, k\triangle]$ and 0 otherwise. How much the new information is weighted depends on the magnitude of the dynamic gain coefficient. The parallel between the error signal in Equation (9.40) and that in standard recursive estimation algorithms suggests that the instantaneous log likelihood is a reasonable criterion function for adaptive estimation with point process observations. Other properties of this algorithm are discussed in [7].

Our objective is to identify plasticity related to both the spatial and temporal properties of the place receptive fields. Therefore, because given the learning rate, defining the conditional intensity function is sufficient to define the adaptive algorithm, we set

$$\lambda(k\triangle|H_k,\theta_k) = \lambda^S(k\triangle|x(k\triangle),\theta_k)\lambda^T(k\triangle-\zeta_k|\theta_k), \tag{9.41}$$

where $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ is a function of the rat's position $x(k\triangle)$ at time $k\triangle$, $\lambda^T(k\triangle - \zeta_k|\theta_k)$ is a function of the time since the last spike, ζ_k is the time of the last spike prior to $k\triangle$ and θ_k is a set of time-dependent parameters. These two functions $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ and $\lambda^T(k\triangle-\zeta_k|\theta_k)$ are respectively the spatial and temporal components of the conditional intensity function. To allow us to capture accurately the complex shape of place fields and the ISI structure of CA1 spike trains, we define $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ and $\lambda^T(k\triangle-\zeta_k|\theta_k)$ as separate cardinal spline functions. A spline is a function constructed of piecewise continuously differentiable polynomials that interpolate between a small number of given coordinates, known as control point. The parameter vector θ_k contains the heights of the spline control points at time $k\triangle$. These heights are updated as the spikes are observed. As in [13] the spatial control points were spaced 1 every 10 cm plus one at each end for 32 total. The temporal control points were spaced 1 every 4 msec from 0 to 25 msec, and then every 25 msec from 25 to 1000 msec for 50 in total. Hence, the dimension of θ is $q = 82$ in this analysis.

In Figure 9.8, we show the first 400 sec of spike train data from that experiment displayed with the track linearized and plotted in time as in Figure 9.5. The spiking activity of the neuron during the full 1200 sec of this experiment is used in the analysis. There were 1,573 spikes in all. The place-specific firing of the neuron is readily visible as the spiking activity occurs almost exclusively between 10 and 50 cms. As in the previous example, the spiking activity of the neuron is entirely unidirectional; the cell discharges only as the animal runs up and not down the track. The intensity of spiking activity (number of spikes per pass) increases from the start of the experiment to the end.

We used the spline model of the conditional intensity function (Equation (9.41)) in the adaptive filter algorithm (Equation (9.40)) to estimate the dynamics of the receptive field of the neuron whose spiking activity is shown in Figure 9.8. The parameter updates were computed every 2 msec and the learning rate parameters were chosen based on the sensitivity analysis described in [13]. Examples of the spatial and temporal components of the conditional intensity function are shown in Figure 9.9. The migration of the spatial component during the course of the experiment is evidenced by the difference between these functions on the first pass compared with the last pass (Figure 9.9A). On the first pass the spatial function has a height of 12, is centered at approximately 40 cm and extends from 15 to 55 cms. By the last pass, the center of the spatial function has migrated to 52 cm, its height has increased to almost 20 and the range of the field extends from 15 to 70 cms. The migration of this spatial function is an exception. Typically, the direction of field migration is in the direction opposite the one in which the cell fires relative to the animal's motion [24, 25]. This place field migrates in the direction that the neuron fires relative to the animal's motion.

The temporal component of the intensity function characterizes history dependence as a function of the amount of time that has elapsed since the last spike. The temporal function shows increased values between 2 to 10 msec and around 100 msec. The former corresponds to the bursting activity of the neuron whereas the latter is the modulation of the place specific firing of the neuron by the theta rhythm [13]. For this neuron the modulation of the spiking activity due to the bursting activity is stronger than the modulation due to the approximately 6 to 14 Hz theta rhythm. Between the first and last pass the temporal component of the conditional intensity function increases slightly in the burst range and decreases slightly in the theta rhythm range. By definition, the rate function, i.e., the conditional intensity function based on the model in Equation (9.41) is the product of the spatial and temporal components at a given time. This is the reason why the units on the spatial and temporal components (Figure 9.9) are not spikes/sec. However, the product of the spatial and temporal components at a given time gives the rate function with units of spikes/sec. A similar issue arose in the interpretation of the spatial components of the conditional intensity functions for the IG and IIG models in Section 9.3.2 (Figure 9.6).

As in the previous two examples we used the KS plots based on the time-rescaling theorem to assess the goodness-of-fit of the adaptive point process filter estimate of the conditional intensity function (Figure 9.10). We compared the estimate of the conditional intensity function with and without the temporal component. The model without the temporal component is an implicit inhomogeneous Poisson process. The impact of including the temporal component is clear from the KS plots. For the model without the temporal component the KS plot does not lie within the 95% confidence bounds, whereas with the temporal component the plot is completely within the bounds. The improvement of the model fit with the temporal component is not surprising given that this component is capturing the effect of theta rhythm and

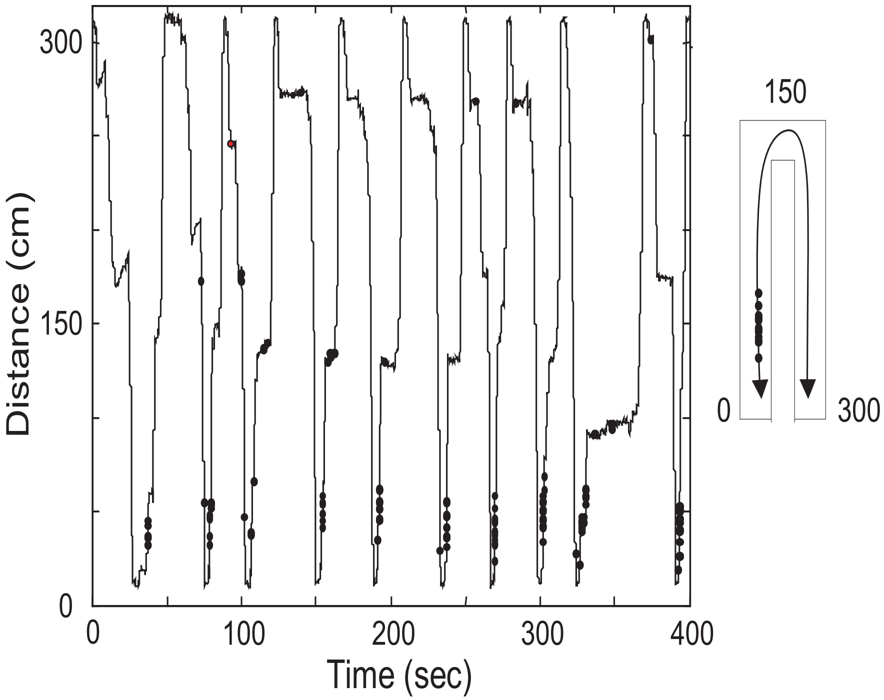

Figure 9.8

As in Figure 9.5, place-specific spiking activity of a second hippocampal pyramidal neuron recorded from a rat running back and forth for 400 sec of a 1200 sec experiment on a 300 cm U-shaped track (outset on the right). As in Figure 9.5, the track has been linearized and plotted in time so that the spiking activity on each pass can be visualized. The black dots show the spatial locations at which the neuron discharged a spike. The place field of this neuron extends from approximately 10 to 50 cms. Along with place-specific firing, this neuron is also directional in that it spikes only as the animal moves from bottom to top (from left to right in the outset) between 10 to 50 cms. The intensity of spiking increases from the start to the end of the experiment. In the data analyses we use the spiking activity during the entire 1200 sec.

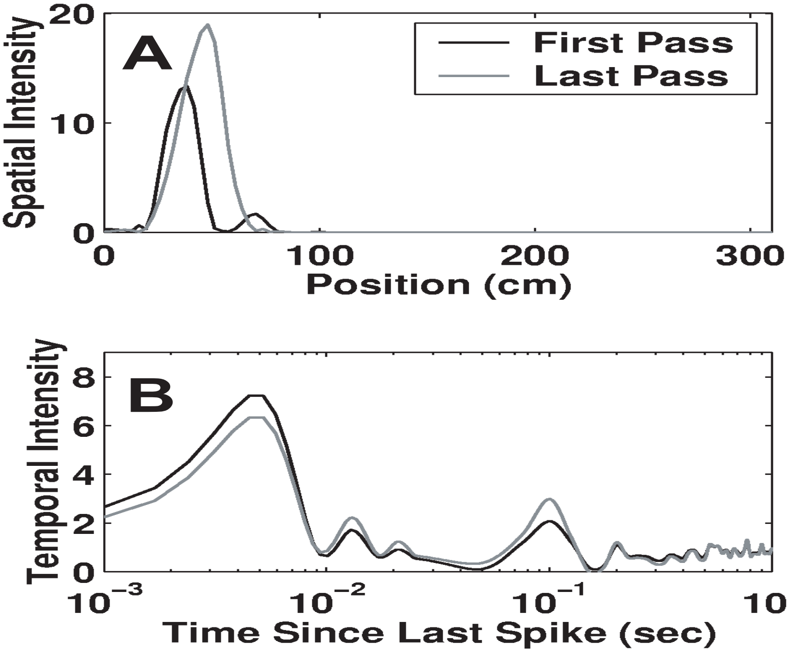

Figure 9.9

A. Point process adaptive filter estimates of the spatial component of the conditional intensity (rate) function on the first (solid black line) and last pass (solid gray line). B. Point process adaptive filter of the temporal component of the conditional intensity (rate) function on the first (solid black line) and last pass (solid gray line).

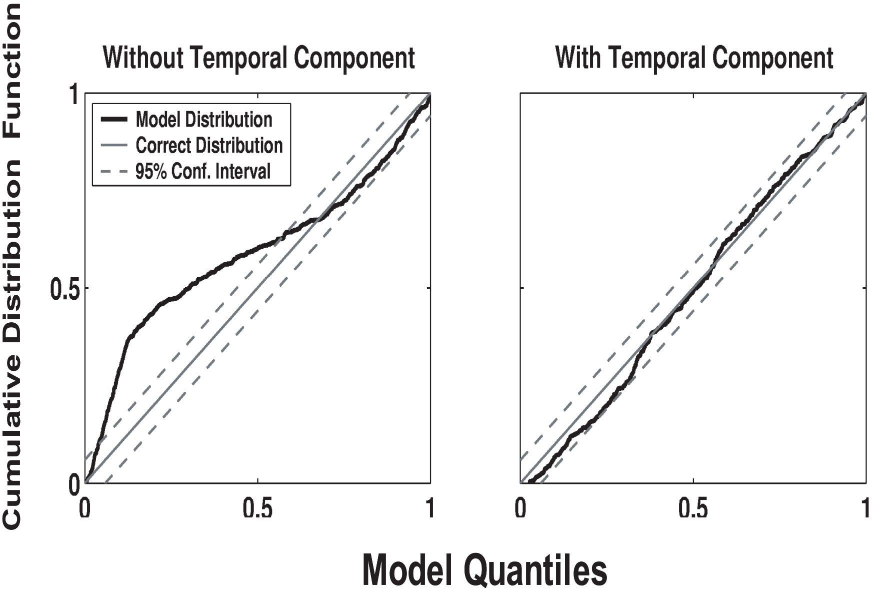

Figure 9.10

Kolmogorov-Smirnov (KS)plots for point process adaptive filter estimates of the conditional intensity (rate) function A. Without the temporal component of the model and B. With the temporal component of the model. The parallel dashed diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data. The solid 45° line represents exact agreement between the model and the data. The adaptive filter estimate of the conditional intensity function with the temporal component gives a complete statistical description of this neural spike train based on the KS plot.

the bursting activity of this neuron (Figure 9.9B).

In summary, using the adaptive filter algorithm with the spline model to characterize the conditional intensity function we computed a dynamic estimate of the place receptive field. The updating was carried out on a 2 msec time-scale. We have found that the dynamics of these fields are best analyzed using videos. Videos of these types of analyses can be found on the websites cited in [7, 13]. These analyses show that use of a flexible model can lead to an accurate characterization of the spatial and temporal features of the hippocampal neuron's place receptive field. The results in this example illustrate an important improvement over the model fits in Section 9.3.2. These improvements can be measured through our KS plot goodness-of-fit tests. We believe these dynamic estimation algorithms may be used to characterize receptive field plasticity in other neural systems as well.

9.4 Conclusion

Neural spike trains are point processes and the conditional intensity function provides a canonical characterization of a point process. Therefore, we used the conditional intensity function to review several concepts and methods from likelihood theory for point process models that are useful for the analysis neural spike trains. By using the conditional intensity it was easy to show that the likelihood function of any point process model of a neural spike train has a canonical form given by Equation (9.8). The link between the conditional intensity function and the spike time probability model (Equation (9.2)) shows that defining one explicitly defines the other. This relation provided important modeling flexibility that we exploited in the analyses of three actual neural spike train data examples. In the first example, we used simple (renewal process) ISI models. In the second example, we applied spike time probability models that were modulated by a time-dependent covariate whereas in the third example, we formulated the spike train model directly in terms of the conditional intensity function. This allowed us to model explicitly history dependence and analyze the dynamic properties of the neurons receptive field. The conditional intensity function was also fundamental for constructing our goodness-of-fit tests based on the time-rescaling theorem.

The likelihood framework is an efficient way to extract information from a neural spike train typically by using parametric statistical models. We showed in the third example that it may also be used to develop dynamic estimation algorithms using a semiparametric model. Likelihood methods are some of the most widely used paradigms in statistical modeling due the extensive theoretical framework and the extensive applied experience that now lies behind these techniques.

Our analyses showed a range of ways of constructing and fitting non-Poisson models of neural spike train activity using the likelihood framework. While the different examples illustrated different features of the likelihood principles, we included in each a goodness-of-fit analysis. We believe the goodness-of-fit assessment is a crucial, yet often overlooked, step in neuroscience data analyses. This assessment is essential for establishing what data features a model does and does not describe. Perhaps, most importantly, the goodness-of-fit analysis helps us understand at what point we may use the model to make an inference about the neural system being studied and how reliable that inference may be. We believe that greater use of the likelihood based approaches and goodness-of-fit measures can help improve the quality of neuroscience data analysis. Although we have focused here on analyses of single neural spike train time-series, the methods can be extended to analyses of multiple simultaneously recorded neural spike trains. These latter methods are immediately relevant as simultaneously recording multiple neural spike trains is now a common practice in many neurophysiological experiments.

9.5 Appendix

Lemma 1. Given n events $E_1, E_2, \cdots, E_n$ in a probability space, then

$$Pr\left(\cap_{i=1}^{n} E_i\right) = \prod_{i=2}^{n} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \tag{9.42}$$

Proof: By the definition of conditional probability for $n = 2$, $Pr(E_1 \cap E_2) = Pr(E_2|E_1)Pr(E_1)$. By induction

$$Pr\left(\cap_{i=1}^{n-1} E_i\right) = \prod_{i=2}^{n-1} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \tag{9.43}$$

Then

$$\begin{aligned} Pr\left(\cap_{i=1}^{n} E_i\right) &= Pr\left(E_n | \cap_{i=1}^{n-1} E_i\right) Pr\left(\cap_{i=1}^{n-1} E_i\right) \\ &= Pr\left(E_n | \cap_{i=1}^{n-1} E_i\right) \prod_{i=1}^{n-1} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1) \\ &= \prod_{i=1}^{n} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \qquad \textbf{Q.E.D.} \end{aligned} \tag{9.44}$$

Acknowledgments. Support for this work was provided in part by NIMH grants MH59733, MH61637, NIDA grant DA015664 and NSF grant IBN-0081458. We are grateful to Satish Iyengar for supplying the retinal spike train data analyzed in Section 9.3.1.

References

[1] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models based on Counting Processes*. New York: Springer-Verlag.

[2] Barbieri, R., Frank, L. M. Quick, M. C., Wilson, M. A., and Brown, E. N. (2001) Diagnostic methods for statistical models of place cell spiking activity. *Neurocomputing*, **38-40**, 1087-1093.

[3] Barbieri, R., Quirk, M.C., Frank, L. M., Wilson, M. A., and Brown, E. N. (2001). Construction and analysis on non-Poisson stimulus-response models of neural spike train activity. *J. Neurosci. Meth.*, **105**, 25-37.

[4] Berman, M. (1983). Comment on "Likelihood analysis of point processes and its applications to seismological data" by Ogata. *Bulletin Internatl. Stat. Instit.,* **50**, 412-418.

[5] Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cyber.*, **59**, 189-200.

[6] Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, **18**, 8411-7425.

[7] Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., and Solo V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *PNAS*, **98**, 12261-12266.

[8] Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., and Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.*, **14**, 325-346.

[9] Casella, G., and Berger, R. L. (1990). *Statistical Inference*. Belmont, CA: Duxbury.

[10] Chhikara, R. S., and Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* New York: Marcel Dekker.

[11] Daley, D., and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* New York: Springer-Verlag.

[12] Frank, L. M., Brown, E. N., and Wilson, M. A., (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron,* **27**, 169-178.

[13] Frank, L. M., Eden U. T., Solo, V., Wilson, M. A., and Brown, E. N., (2002). Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *Journal of Neuroscience*, **22**, 3817-3830.

[14] Gerstein, G. L. and Mandelbrot, B. (1964) Random walk models for the spike activity of a single neuron. *J. Biophys.,* **4**, 41-68.

[15] Gerstner, W., and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: University Press.

[16] Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. London: Chapman and Hall.

[17] Haykin, S. (1996). *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall.

[18] Iyengar, S., and Liao, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cyber.,* **77**, 289-295.

[19] Johnson, A., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions-2*. New York: Wiley.

[20] Kalbfleisch, J., and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

[21] Kass, R. E., and Ventura, V. (2001). A spike train probability model. *Neural Comput.*, **13**, 1713-1720.

[22] Levine, M. W., Saleh, E. J., and Yamold, P. (1988). Statistical properties of the maintained discharge of chemically isolated ganglion cells in goldfish retina. *Vis. Neurosci.*, **1**, 31-46.

[23] Levine, M. W. (1991). The distribution of intervals between neural impulses in the maintained discharges of retinal ganglion cells. *Biol. Cybern.,* **65**, 459-467.

[24] Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997). *Proc. Natl. Acad. Sci. USA,* **94**, 8918-8921.

[25] Mehta, M. R., Quirk, M. C., and Wilson, M. A. (2000). *Neuron,* **25**, 707- 715.

[26] Meyer, P. (1969). Démonstration simplifiée d'un théoréme de Knight. In *Séminaire Probabilité V* (pp. 191-195). New York: Springer-Verlag.

[27] Muller, R. U., and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience,* **7**, 1951-1968.

[28] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of American Statistical Association,* **83**, 9-27.

[29] O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.,* **34**, 171-175.

[30] Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.*, **165**, 483-506.

[31] Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* London: Oxford.

[32] Schroedinger, E. (1915). Zur Theorie der fall- und steigversuche an teilchen mit Brownscher bewegung. *Phys. Ze.*, **16**, 289-295.

[33] Snyder, D., and Miller, M. (1991). *Random Point Processes in Time and Space* (2nd ed.). New York: Springer-Verlag.

[34] Solo, V. and Kong, X. (1995). *Adaptive Signal Processing Algorithms: Stability and Performance.* Upper Saddle River, NJ: Prentice-Hall.

[35] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* In Springer Series in Statistics, New York: Springer-Verlag.

[36] Taylor, H. M., and Karlin, S. (1994). *An Introduction to Stochastic Modeling* (rev. ed.) San Diego, CA: Academic Press.

[37] Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology: Nonlinear and Stochastic Theories.* New York; Cambridge.

[38] Wilson, M. A., and McNaughton, B. L., (1993). Dynamics of the hippocampal ensemble code for space. *Science*, **261**, 1055-1058.

[39] Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, **397**, 613- 616.

Chapter 10

Biologically-Detailed Network Modelling

Andrew Davison
Yale University School of Medicine, Section of Neurobiology, P.O. Box 208001, New Haven, CT 06520-8001, U.S.

CONTENTS

10.1 Introduction

The appropriate level of detail for a computational neuroscience model is determined by the nature of the system or phenomenon under investigation, by the amount of experimental data available and by the aims of the investigator. For example, models of certain network phenomena, e.g., synchronization, do not require details of cell morphology, and perhaps not even ionic currents – integrate and fire neurons will suffice. On the other hand, studying dendritic processing in active dendrites requires details of dendritic morphology and of the active channels in the dendrites. However, although a detailed model may be desirable, such a model must be well constrained by experimental data or else be over-parameterised and suffer from lack of predictive or explanatory power. The final factor affecting the level of detail is the purpose of the model: to provide support for a specific hypothesis about the function of a neural system – such models tend to be more abstract with less detail – or to be guided by experimental data to discover unknown properties of a system – such models tend to be more detailed, aiming at as realistic a model as possible.

The advantages of simplified, "abstract" models (the simplifications might include all-to-all connectivity, integrate-and-fire neurons or non-spiking, rate models) are that they (i) may be amenable to mathematical analysis, allowing conclusions to be generalized; (ii) are simpler to understand and analyze; (iii) have fewer unknown parameters, potentially allowing the model to be better constrained by the available data, and so making the hypothesis driving the model more easily falsifiable; and (iv) require little computing power, so that the model behaviour can be more thoroughly investigated. Disadvantages of such models are that (i) simplifications may conflict with experimental data; (ii) the model may be over-simplified such that it does not represent the real system; (iii) they are usually difficult to relate to experimental data, so may be less constrained by data, and less able to generate experimentally-testable predictions.

The advantages of detailed, "realistic" models are that they (i) can be related directly to experimental data; (ii) constitute "a compact and self-correcting database of neurobiological facts and functional relationships." [29]; (iii) do not pre-judge the properties and functioning of the system (are guided by the data), and so may enable discovery of unknown properties of the system; and (iv) are more able to make experimentally testable predictions. The disadvantages are that they (i) may not be hypothesis-driven; (ii) are always incomplete – an omitted element may be crucial in the functioning of the system; (iii) may be almost as complex as the system being modelled, and so just as hard to analyze; (iv) usually require many unknown or estimated parameters – several different parameter sets may produce the same behaviour; (v) require considerable computing power; (vi) are not amenable to mathematical analysis.

Given the notable advantages and disadvantages of both classes of model, it would appear desirable, in the model-based investigation of a neural system, to develop a hierarchy of models with different levels of complexity. Thus, although a simplified, abstract model which can be easily simulated and analyzed may be difficult to relate to/constrain by experimental data, it may be related directly to a more detailed, realistic model with which it is consistent, and the detailed model may be related directly to/constrained directly by the data.

"Realistic" models, with detailed ion channel kinetics and cell morphologies, based on experimental data, have mainly been used in studying single cells, because it is for single cells that most of the experimental data is available and because of limited computer power. Increasingly, however, experimentalists have begun to obtain data suitable for realistic modelling of networks, and the relentless increase in the ratio of computer power to cost has made modelling of medium-sized networks (of a few hundred to tens of thousands of neurons) with detailed, realistic neurons and synapses feasible.

In this chapter I will discuss the requirements for biologically-detailed, realistic network modelling. The requirements are divided into those for the neuron models, the synaptic models, the pattern of connections between cells, the network inputs. I will then discuss implementation of the model – choice of simulation environment and numerical issues. Finally I will discuss putting-it-all-together – validation of the network model as a whole. To illustrate these requirements I will use a model of the granule cell layer in the cerebellum by Maex and De Schutter [23] and a model of olfactory bulb developed by the author and collaborators [7, 8].

10.2 Cells

Most if not all biologically-realistic single neuron models are based on Hodgkin-Huxley-like ion channel models [18] and the compartmental/cable modelling of dendrites introduced by Rall [30]. Examples include models of cerebellar Purkinje cells by De Schutter and Bower [11, 12], of olfactory bulb mitral cells by Bhalla and Bower [3] and CA3 hippocampal pyramidal cells by Traub et al. [39]. For further examples see the Senselab ModelDB website (http://senselab.med.yale.edu/SenseLab /ModelDB).

These models are in general very complex and their simulation requires solution of thousands of differential equations. The Bhalla and Bower mitral cell model, for example, has almost 300 compartments with up to six ion channel types in each. The De Schutter and Bower Purkinje cell model has 1600 compartments and nine ion channels. Simulation of such models requires large amounts of computer power and they are therefore in practice unsuitable for use in network models where hundreds or thousands of cells must be simulated.

For network modelling we therefore require neuron models with a lower level of complexity that nevertheless retain as much fidelity to the biology as possible. A number of strategies have been used to construct such intermediate-complexity single cell models. All take as their starting point a detailed compartmental neuron model and attempt to simplify it while retaining the electrotonic properties and/or input-output behaviour of the detailed model. One strategy is to concentrate on the electrotonic properties and reduce the number of compartments in the cell while conserving the membrane time constants and the cell input resistance [6, 37]. A more drastic strategy is to attempt to abstract the key features of the cell into as few compartments and channel types as possible, and constrain the simplified model to have the same input-output properties as the detailed model, in terms of firing rate response to synaptic or electrical stimulation [27]. Both strategies give shorter simulation times than the fully-detailed models.

10.2.1 Modelling olfactory bulb neurons

In developing our model of the mammalian olfactory bulb we used the second strategy to construct simplified models of mitral and granule cells [9] based on the detailed models of Bhalla and Bower [3]. The method relies on there being regions of the dendritic tree within which ion channel densities are uniform. For example, in the Bhalla-Bower mitral cell model the primary dendrite shaft has 6 compartments, but the peak conductances for the different ion channels are the same in all those compartments. The primary dendrite apical tuft has 94 compartments, and the peak conductances of the different ion channels are the same in all those compartments, although different to the values in the dendrite shaft. Therefore, we reduced the primary dendrite shaft to a single, iso-potential compartment with the same peak conductance densities as in the original model, and carried out a similar reduction for the primary dendrite tuft, for the secondary dendrites and for the soma/axon region. These four iso-potential compartments were linked by purely resistive elements. The relative surface areas of the four compartments, and the values of the three linking resistances, were varied in order to obtain the best fit between the reduced model and the original model.

For the purposes of our network model, the details of signal transmission within the dendritic tree of an individual neuron are not of interest. What is important is the input-output relationship, i.e., we treat the neuron as a black box. It is important for the reduced model to mimic the original model over as wide a range of input conditions as possible. Therefore, we compared the models with inputs at two different sites, the soma and the primary dendrite apical tuft, and with different input amplitudes, such that the weakest input gave an output firing rate of about 10 Hz and the strongest input gave an output firing rate of almost 100 Hz. We added an extra free parameter, scaling the amplitude of the input to the reduced model relative to that to the original model, to take account of differences in input resistance of the models.

With an error function based on the relative timing of the first four action potentials in response to current injection, we used the Simplex optimization method to find the values of the free parameters that gave the best agreement between the reduced and original models. Part of the results are shown in Figure 10.1. The four-compartment mitral cell model gave a good qualitative and quantitative fit to the fully-detailed model, and ran 75 times faster than the full model, making its use in a large network model practical.

10.2.2 Modelling cerebellum neurons

For the model granule cell, Maex and De Schutter reduced a 13 compartment model of an *in vitro* turtle granule cell [16] to a single, isopotential, spherical compartment in order to reduce the computational requirements for the network model. This reduction was justified by the electric compactness of rat granule cells. The rate constants describing the voltage-gated channels were increased to give *in vivo*, rather than room temperature, kinetics. To compensate for the resultant reduced charge transfer the channel peak conductances were also increased. A number of changes

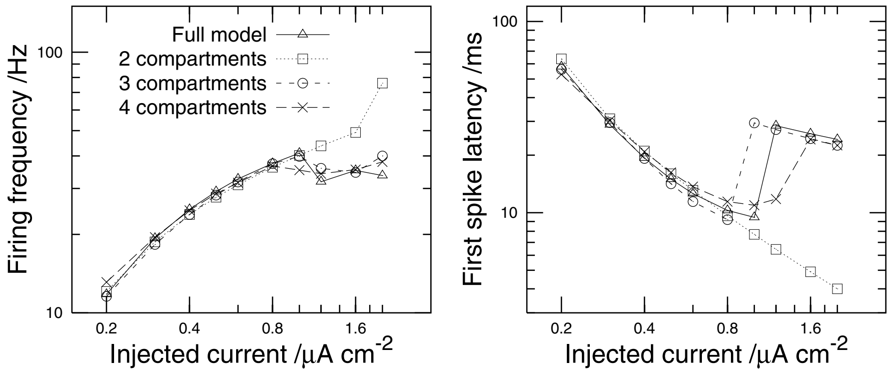

Figure 10.1

Comparison of the firing rate and first-spike latency of the reduced mitral cell models with the original Bhalla-Bower model. (A) Firing frequency. (B) Time from current onset to first action potential. From [10] with permission.

were made to convert the model from turtle to rat (see [23]). The model granule cell responses agreed qualitatively with experimental recordings. The model Golgi cell was developed just for the network model. In this case, a single-compartment model was dictated by the lack of morphological reconstructions. Because of insufficient voltage clamp data the Golgi cell model used the same ion channels as the granule cell. Ion channel peak conductances were then tuned to give qualitative agreement with current clamp recordings.

10.3 Synapses

A complete model of synaptic transmission would incorporate calcium release following action potential invasion of the pre-synaptic terminal, vesicle transport docking, release and recycling, neurotransmitter diffusion and removal from the synaptic cleft, binding of neurotransmitter, conformational changes of ion channel proteins, and entry of ions through open channels. A minimal model of synaptic transmission would consist of a pre-synaptic impulse triggering a step change in membrane potential in the post-synaptic compartment. An appropriate level of description for network models is likely to lie between these two extremes. Two common simplifications of the pre-synaptic mechanism are (i) an AP triggers, after a certain delay, a square pulse of neurotransmitter at the post-synaptic mechanism; the post-synaptic current is then affected by the pulse amplitude and duration; (ii) more drastically, a presynaptic AP triggers a stereotyped post-synaptic response and details of neurotransmitter release/diffusion are ignored completely. An excellent discussion of the simplification of synaptic models is given by Destexhe et al. (1998) [10]. They also present optimized algorithms for calculating synaptic currents in a network model with high computational efficiency.

Values for peak synaptic conductances and synaptic time constants can be obtained from voltage-clamp recordings.

10.4 Connections

The pattern of connections within a neural system can strongly influence its function. This presents a major challenge to modelling, since many connectivity parameters are not known and may change with the size of the network that is modelled.

Important factors include the size (both number of cells and spatial extent) and shape of the network, the number of synapses per cell, and the frequency of synapses as a function of distance and direction from the cell soma. These determine the degree of interaction between any two cells, i.e., how many synapses there are in the shortest path between two cells, or, more comprehensively, the distribution of path lengths between any two cells.

10.4.1 Network topology

For most brain areas the topology is well characterised. Cortical regions are characterised by a laminar structure, often with a columnar organization in the direction perpendicular to the laminae. The olfactory bulb has a well defined laminar structure with each cell type restricted to a single layer. It is therefore natural to describe it by a two-dimensional network. To represent the actual shape of the olfactory bulb the network topology should be defined by the surface of some ellipsoid. A planar network, however, makes calculation of location and distance much easier. For the granular layer of the cerebellum, Maex and De Schutter found that a one-dimensional network displayed the same dynamics as a two-dimensional one. Reducing the dimensionality of the network has the advantage that the spatial extent of the remaining dimension(s) can be made larger.

Edge effects are a particular problem. Where two regions of cortex with different functions abut one another there are presumably connections across the ‘boundary’. How should these be incorporated in the model? In modelling it is common practice to wrap-round the array of cells (whether one- or two-dimensional) so that cells at one edge of the array are adjacent to cells at the opposite edge. This avoids the system having different properties in the centre from those at the edge, but it should be remembered that real systems do not wrap around in this way. We used this technique in our olfactory bulb model, giving a toroidal topology, which is not too dissimilar to the incomplete-ellipsoidal topology of the real bulb. Maex and De Schutter did not use wrap-around, instead increasing the size of the network sufficiently to give a central region with no edge effects.

10.4.2 Number of connections

There are two principal problems regarding the number of connections in a detailed network model. The first is finding experimental data on the statistics of connections between the different cell types in a model. There is a paucity of such data for many brain regions. The second is that almost all models must have fewer cells and connections than the biological system because of insufficient computer resources.

10.4.2.1 Estimation of number of connections

The number of synapses per olfactory bulb mitral cell, n_{syn}, has not been experimentally determined. However, it can be estimated from other measurements. The number of synapses in the EPL of adult mice has been estimated by electron microscopy as $(1.1 \pm 0.3) \times 10^9$ [28]. It is unclear whether this estimate is of reciprocal synapses or of individual synapses (a reciprocal synapse consists of an excitatory-inhibitory pair). In the latter case, the number of reciprocal synapses will be half the above estimate. An indirect estimate gives a very similar result: the number of spines on the peripheral dendrites of a single granule cell has been measured as 144–297 in mice [42] and 158–420 in rabbits [25]. Assuming 200 spines per cell and one reciprocal synapse per spine, taking the number of mitral cells per mouse OB as 38400 [32] and the ratio of granule:mitral cells as 150 [36], then the number of synapses in the mouse EPL $= 200 \times 38400 \times 150 = 1.15 \times 10^9$. The number will be slightly higher for rabbit, as there are about 60000 mitral cells [32].

Assuming a constant density of synapses on mitral/tufted cell dendrites, it would be expected that mitral cells, which are larger, would have more synapses than tufted cells. The total secondary dendrite lengths for rabbit mitral and middle tufted cells have been measured as 15016 μm and 4050 μm respectively [25]. Therefore, taking the ratio of tufted:mitral cells as 2.5 [36],

$$\text{Synaptic density} = \frac{1.1 \times 10^9}{38400\,(15016 + 2.5 \times 4050)} \simeq 1.14\ \mu\mathrm{m}^{-1} \qquad (10.1)$$

so the number of synapses per mitral cell is approximately $1.14 \times 15016 = 17000$ and the number per tufted cell is 4600.

These calculations assume that all synapses involve a mitral/tufted cell, and so they ignore centrifugal inputs onto interneurons in the EPL. Therefore these are probably slight overestimates.

In practice, it was not possible to simulate a network with several thousand synapses per mitral cell, and a value of 500 synapses per cell was used. The reduced number of synapses was compensated for by a granule cell model that was more excitable than found experimentally.

In the cerebellar granule cell layer model of Maex and De Schutter there were an average number of 602 synapses from granule cells onto each Golgi cell, corresponding to a connection probability of 0.2. It is not stated, however, how this latter figure was determined.

10.4.2.2 Reducing the network size

Simulating large, highly connected networks in which the individual elements are themselves complex is extremely computationally intensive. Therefore it is desirable to simulate smaller networks and to infer the behaviour of the full-scale network from the behaviour of the smaller simulations.

In principle there are two ways to shrink a network: it can be made smaller in extent or be made more sparse. A smaller network, which may represent a sub-region of the neural structure, has the same connectivity within the sub-region, but ignores any connections from outside. Such connections may be very significant, but experimental conditions can be simulated that minimize the degree of activation of external connections, for example focal stimulation.

A sparser network does not suffer from such effects of missing external connections, but will have different connectivity to the full-scale network: either the number of connections per cell will be reduced or the probability of connecting to a neighbouring cell will be increased.

In our model of the olfactory bulb we used both methods of reducing the network size. In one version of the model, a single glomerulus was stimulated, thus reducing the number of cells that needed to be simulated by about 1000-2000 (the number of glomeruli in a single bulb) so that a realistic number of mitral and granule cells could be used. In another version, with 25-100 glomeruli, the network was shrunk less drastically in extent, but was made sparser, with only a single mitral cell per glomerulus. The number of synaptic connections per cell was maintained the same; therefore the connectivity between cells (e.g., the probability of two given mitral cells contacting the same granule cell) was increased.

As mentioned previously, the Maex and De Schutter cerebellum model is of a one-dimensional array of cells, a ‘beam’ of the granule cell layer. This greatly reduces the number of cells without, for the phenomena considered in their model, affecting the network dynamics. The ratio of granule:Golgi cells is estimated as 1000:1 [23], but a reduced ratio of about 150:1 was used in the model. The number of connections was reduced accordingly, but the peak synaptic conductance was normalized according to the number of connections in order to maintain a fixed level of synaptic input. This method preserves the connectivity but increases the influence that a single connection has on the target cell response.

10.4.3 Distribution of connections

Without knowing the exact biochemical signals that regulate dendritic growth and synapse formation, the most realistic method of specifying the connections between cells in a model would be to generate a realistic dendritic morphology for all neurons, form a synapse wherever two dendrites come sufficiently close together, and use activity-dependent pruning to eliminate some synapses and strengthen others.

Ascoli (2000) [1] reviews two approaches to generating realistic dendritic morphologies. The L-Neuron program [2] is one example of a system for generating virtual neurons whose morphological parameters (e.g. length of branches) have the same statistical distribution as real neurons, based on a few so-called 'fundamental parameters' measured from an experimental data set. An alternative is to generate an entire population of neurons at once, using a growth algorithm that distributes dendrites among neurons in a manner that reflects competition for metabolic resources [35].

A simpler alternative is to use a probabilistic method, with the probability of two neurons being connected based on the distance and possibly on the direction between the neurons. To specify the reciprocal, dendrodendritic synapses between mitral and granule cells in the olfactory bulb we supposed that each mitral cell has a probability density field $p(r,\phi)$ (polar coordinates in the plane of the secondary dendrites), such that the probability of forming a synaptic connection within a region of size $r\delta r\delta\phi$ at point (r,ϕ) is

$$P(r,\phi) = p(r,\phi)r\delta r\delta\phi \tag{10.2}$$

The identity of the granule cell to which the connection is made could also be determined by such a probabilistic method. However, since the radius of the granule cell dendritic field is much smaller than that of the mitral cell, it is simplest to make the connection to the granule cell whose soma is nearest to the point (r,ϕ). We made the simplifying assumption that the probability of a mitral cell forming a synapse at a point depends only on the radial distance of the point from the soma, and not on the direction, i.e.,

$$p(r,\phi) = p(r) \tag{10.3}$$

This is an approximation, since of course synapses must occur on dendrites, and these project in definite directions from the cell. However, the dendrites branch copiously, so the approximation appears reasonable.

What is $p(r)$? First, we assumed that synapses are approximately evenly spaced along the dendrites. If the dendrites have no branches, then $p(r) \propto 1/r$ (i.e., the average *number* of synapses at any given radial distance is constant, so the *density* of synapses declines with distance). If the dendrites branch copiously such that dendritic density is constant within the arbour, then $p(r) =$ constant. In practice, $p(r)$ is likely to lie somewhere within these limits. For any single cell, $p(r)$ will have a discontinuity at each branch point, but these can be smoothed out by taking an ensemble average from many cells. For simplicity we used $p(r) \propto 1/r$ but it would be of interest in future to examine the effect of the $p(r)$ distribution on the network behaviour.

In summary, the connections are specified as follows: The number of synapses per mitral cell is fixed. For each synapse, a direction ϕ and radius r are chosen at random from uniform distributions, and a connection made to the granule cell whose soma is located closest to the point (r,ϕ). Therefore the number of synapses per granule cell is not constant, but follows some distribution. An example of the pattern of connections between a mitral cell and the granule cell population is shown in Figure 10.2.

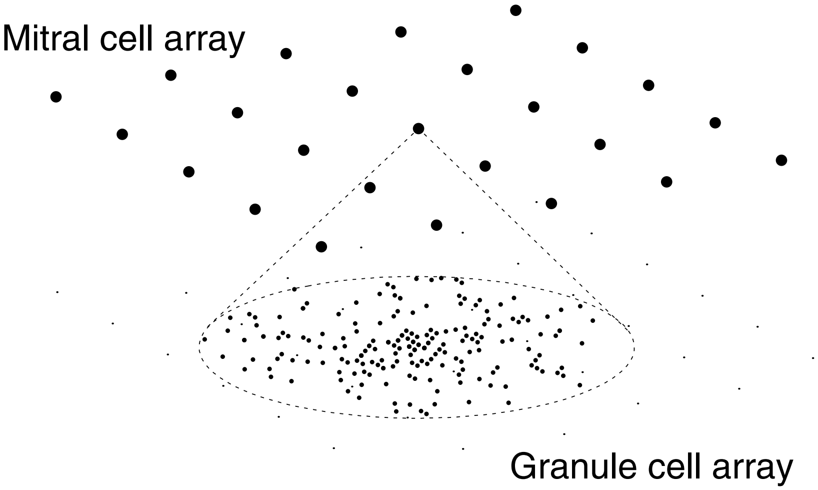

Figure 10.2

The mitral–granule cell network. Mitral cells are shown as large dots. Each mitral cell makes n_{syn} connections to granule cells selected randomly from all the cells within its ‘arbor’ (shown as a circle). $n_{\text{syn}} = 200$ in this figure. The granule cells connected to the central mitral cell are shown as intermediate-size dots. There are fewer connected cells than connections, since some granule cells have more than one connection to the mitral cell. The remaining granule cells are shown as the smallest dots. From [8] with permission.

A probabilistic method was also used in Maex and De Schutter's model of the cerebellar granule cell layer to determine granule–Golgi cell connections. Since the network in this case was one-dimensional, only r had to be considered and not ϕ. The probability of a connection was constant within a fixed range, so $p(r) = \text{constant}$.

10.5 Inputs

It is pointless to go to great lengths to accurately model a neural system if similar care is not taken in modelling its inputs ('rubbish in, rubbish out'). However, it is often the inputs to a system about which least is known. In modelling the inputs one must consider two main issues, first, the spatial and temporal pattern of inputs across the network as a whole, and second, the nature of the inputs to individual cells.

10.5.1 Spatiotemporal pattern of inputs

Are the network inputs organised spatially, or are they uniform across the network? The spatial organisation may depend on the regime being simulated, e.g., natural vs artificial stimulation. Are the inputs stationary in time, or do we wish to simulate a transient response? Again, depending on the regime we wish to investigate, the same network may receive either type of temporal input. From the point-of-view of inputs, it is easier to model sensory systems, since the inputs are in general better understood, and are easier to control experimentally.

10.5.1.1 Spatial pattern of OB inputs

The olfactory bulb receives inputs from the olfactory nerve (sensory inputs from olfactory receptor neurons (ORNs) or electrical stimulation of the nerve) and centrifugal inputs from various brain regions. Little is known about the centrifugal inputs although they are thought to be mainly modulatory in nature. When considering experiments performed in bulb slices centrifugal inputs can be ignored entirely since the connections are not preserved in this preparation. For these reasons we did not include centrifugal inputs in our model.

In one sense, modelling the spatial distribution of inputs is easy: each ORN-type projects to one or a pair of glomeruli and each glomerulus receives input from only one ORN-type [5, 24, 40]. However, any given receptor responds to a broad range of odorants, and any given odorant activates receptors of more than one type. There is as yet no satisfactory model for this, although features of the odour molecule such as shape and chemical identity appear to be important [22, 26]. There is some evidence that receptors which respond to chemically-similar odorants project to nearby glomeruli [15, 19] but it is not established that this arrangement has a functional purpose [21].

We chose to model 10 odour 'features' and 36 receptor types/glomeruli (in a 6×6

array). Let a_{ij} be the degree of activation of each receptor type j by each odour feature (OF) i ($0 \leq a_{ij} \leq 1$). All the a_{ij}'s form a 10×36 matrix, $\mathbf{A}$. An odour is then represented by a 10 element vector $\vec{x}$. Applying odour $\vec{x}$ to the bulb model produces a receptor-activation vector $\vec{y}$:

$$\vec{y} = \mathbf{A}\vec{x} \tag{10.4}$$

This assumes no interaction between odour features at the receptor level and so is a simplification of the real situation [20]. The current applied to the glomerular compartment of mitral cell j is then proportional to the element y_j of the receptor activation vector. The proportionality includes a scaling for odour intensity. There are no good statistical data available which would allow determination of the matrix $\mathbf{A}$. In developing a procedure to specify $\mathbf{A}$, we adopted the criterion that a large minority of the elements should be zero (no response), and that a small minority should have strong responses. This criterion is based on imaging studies of glomerular activation (e.g., [14, 15, 33]). Full details of how $\mathbf{A}$ was generated are given in Davison (2001) [7].

10.5.1.2 Spatial pattern of inputs to cerebellar granule cell layer

The input to the granular layer of the cerebellum comes from the mossy fibres, forming synapses onto Golgi cells and granule cells. In the model the input distribution was spatially homogeneous, although with random variability in the set of innervating mossy fibres for each granule cell, and random variability in synaptic weights.

10.5.1.3 Temporal pattern of OB inputs

The temporal variation of olfactory stimuli is slow compared to auditory or visual stimuli. A typical experimental stimulus is exposure to a single odour for an extended period. We replicated this by stimulating the network with a low intensity 'background' odour for 1000 ms then applied a fixed odour at a fixed intensity for the remainder of the simulation (steady-state conditions). We also used a sinusoidally varying stimulus since the experimental odour intensity may vary periodically due to sniffing or breathing.

10.5.1.4 Temporal pattern of inputs to cerebellar granule cell layer

Steady-state conditions were also used by Maex and De Schutter for the mossy fibre inputs. The mossy fibre inputs were modelled as pulses representing the arrival of a spike. Due to the lack of experimental information about the fine structure of the spike trains in mossy fibres the simplest model was used, that the probability of firing of a mossy fibre was independent of its firing history (this corresponds to a Poisson distribution of the instantaneous firing rate).

10.5.2 Inputs to individual cells

The possibilities for inputs to individual cells are (i) continuous current injection, (ii) synaptic inputs and (iii) current pulses.

10.5.2.1 Continuous current injection

Each cell has one or more current sources at approximately the same location as the synapses in the real cell. The amplitude of the current injection may vary from cell to cell. The spatial pattern of the inputs consists of specifying the current amplitude at each input node in the network. The temporal pattern is generally simple – constant or periodically varying – possibly overlaid with a random element. It has been shown for the Hodgkin-Huxley model that under a wide range of conditions a random continuous current, based on the diffusion process, is equivalent to discrete pulses from a Poisson process [13]. This method has the advantage of simplicity. It is preferable to use this method if little is known about the details of the input synapses or of the temporal structure of the inputs, since few assumptions must be made. In general this method is the least computationally expensive, which may be an important factor. The exception to this is if there is a random element, in which case fixed time-step integration methods must be used, which may be slower than variable time-step methods. With pulsed inputs a random element may be introduced in the pattern of input triggers, and variable time-step methods may still be used. This method was used in our olfactory bulb model.

10.5.2.2 Synaptic inputs

If there is sufficient experimental data about the input synapses and sufficient computer power then this method is preferable. It has the obvious advantage that it matches the biological system most closely and therefore one can have greater confidence in the results of the model. The input is a sequence of input spikes. The spikes may have a constant interval, the interval may follow some functional form, or may be drawn from a specified random distribution. A commonly used input distribution is the Poisson. The global pattern is then the spatial distribution of the mean interval and of the synaptic amplitude. This method is in general the most computationally expensive as calculations are required for each synapse, particularly so if there is more than one input synapse per cell. It does, however, allow the combination of variable time step methods and noisy inputs. This method was used in the cerebellum model of Maex and De Schutter (1998).

10.5.2.3 Current pulses

Modelling the inputs as current pulses combines some of the advantages of the previous two methods. A current pulse requires less computation than a synaptic conductance, and less knowledge about the details of the synapse. In contrast to continuous current injection, more of the fine temporal structure of the inputs is retained, for networks in which spike timing is important, and variable time step methods may be used with random input.

10.6 Implementation

While it is possible to implement complex, realistic neuronal models in a general purpose programming language such as C++ (a class library, “CONICAL” for compartmental neuron models is available from http://www.strout.net/conical/ [38]), there are several software packages specifically for simulating such models. Neuron (www.neuron.yale.edu) [17], GENESIS (www.genesis-sim.org) [4, 41] and Surf-Hippo (www.cnrs-gif.fr/iaf/iaf9/surf-hippo.html). These programs are suitable for modelling at every level from sub-cellular to systems, although they are most commonly used for single neuron and small-network models. There are several advantages to using these programs:

- Avoids reinventing the wheel. The built-in algorithms for solving the systems of differential equations have been optimized and thoroughly tested.

- Better conceptual control and faster development. Models can be written in high level languages in which neuronal elements such as synapses or sections of cell membrane are fundamental objects. This makes the structure of the model easier to understand from reading the source code. The programs also have graphical user interfaces which facilitate model development, simulation and display of results.

- Code portability. GENESIS and Surf-Hippo run on several variants of Unix. Neuron runs on Unix, Microsoft Windows and Apple Macintosh operating systems. In each case, model code written for one platform will run unmodified on any another platform for which the simulator is available, e.g., a GENESIS model developed under Solaris can be easily transferred to Linux, or a Neuron model developed under MacOS transferred to Windows, with no need for compiling.

10.7 Validation

Almost by definition a model implies simplification, approximation, reduction in scale. In developing a detailed model of a neuronal circuit approximations and informed guesses are required in every part of the model, due either to gaps in experimental knowledge (this is an important outcome of detailed quantitative modelling – gaps in our knowledge are starkly exposed) or to limitations of computer power.

It is therefore especially important to demonstrate that these lacks do not invalidate the model, i.e., that the behaviour of the model is consistent with the majority of experimental knowledge about the system being modelled.

The first necessary step is to test that the model can reproduce the experimental results that were used in developing the model, the ‘built-in’ properties. This test is important first to check for mistakes in the implementation of the model (such as ‘bugs’ in the code) but can also expose inconsistencies between different sources of experimental data. This alone is not sufficient, however. It is essential to attempt to reproduce experimental results that were not used in developing the model. If the ‘emergent’ properties of the model match experimental findings, this validates the model. (For more discussion on built-in vs emergent properties of computational models, see Protopapas et al. (1998) [29]).

As an example, one way in which we tested our olfactory bulb model was by attempting to reproduce published data on dendrodendritic synaptic currents [34]. Many of the model parameters were derived from this same publication, but the amplitude and time constant of the mitral cell IPSC were not incorporated directly in the model: they are emergent properties. The simulated IPSC matched the experimental one closely, although with a number of discrepancies. Because of these discrepancies we would regard the model as only partially validated by these results. The resolution of these discrepancies suggests further lines of enquiry.

10.8 Conclusions

What is the future of biologically-detailed network modelling? As computers become more powerful, more detail can be incorporated into models, but with this comes the need for more detailed, carefully-designed biological experiments to constrain and test the models. As the complexity of models approaches that of real systems, more sophisticated analysis tools will be required. As discussed in the Introduction, it will in most cases be desirable to develop a hierarchy of models to link abstract, conceptual models, via models of intermediate complexity, to detailed models and thence to biological data.

References

[1] G. A. Ascoli (1999), Progress and perspectives in computational neuroanatomy, *Anatomical Record,* **257**: 195–207.

[2] G. A. Ascoli, and J. L. Krichmar (2000), L-Neuron: a modeling tool for the efficient generation and parsimonious description of dendritic morphology, *Neurocomputing*, **32–33**: 1003–1011.

[3] U. S. Bhalla, and J. M. Bower (1993), Exploring parameter space in detailed single cell models: simulations of the mitral and granule cells of the olfactory bulb, *J. Neurophysiol.*, **69**: 1948–1965.

[4] J. M. Bower, and D. Beeman (1997), *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System*, TELOS: New York.

[5] L. B. Buck (1996), Information coding in the vertebrate olfactory system, *Ann. Rev. Neurosci.*, **19:** 517–544.

[6] P. C. Bush, and T. J. Sejnowski (1993), Reduced compartmental models of neocortical pyramidal cells, *J. Neurosci. Meth.*, **46**: 159–166.

[7] A. P. Davison (2001), *Mathematical Modelling of Information Processing in the Olfactory Bulb*, University of Cambridge.

[8] A. P. Davison, J. Feng, and D. Brown (2003), Dendrodendritic inhibition and simulated odor responses in a detailed olfactory bulb network model. *J. Neurophysiol.* In Press.

[9] A. P. Davison, J. Feng, and D. Brown (2000), A reduced compartmental model of the mitral cell for use in network models of the olfactory bulb, *Brain Research Bulletin*, **51**: 393–399.

[10] A. Destexhe, Z. F. Mainen, and T. J. Sejnowski (1998), Kinetic models of synaptic transmission, in *Methods in Neuronal Modeling: From Ions to Networks*, C. Koch and I. Segev (eds.), MIT Press: Cambridge, Massachusetts, 1-25.

[11] E. De Schutter, and J. M. Bower (1994), An active membrane model of the cerebellar Purkinje cell. I. Simulation of current clamps in slice, *J. Neurophysiol.*, **71**: 375–400.

[12] E. De Schutter, and J. M. Bower (1994), An active membrane model of the cerebellar Purkinje cell. II. Simulation of synaptic responses, *J. Neurophysiol.*, **71**: 401–419.

[13] S. Feerick, J. Feng, and D. Brown (2000), Random pulse input versus continuous current plus white noise: Are they equivalent?, *Neurocomputing* **32-33**: 127–132.

[14] R. W. Friedrich, and S. I. Korsching (1987), Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging, *Neuron*, **18**: 737–752.

[15] R. W. Friedrich, and S. I. Korsching (1998), Chemotopic, combinatorial and noncombinatorial odorant representations in the olfactory bulb revealed using a voltage-sensitive axon tracer, *J. Neurosci.*, **23**: 9977–9988.

[16] F. Gabbiani, J. Midtgaard, and T. Knopfel (1994), Synaptic integration in a model of cerebellar granule cells, *J. Neurophysiol.*, **72**: 999–1009.

[17] M. L. Hines, and N. T. Carnevale (1997), The NEURON simulation environment, *Neural Computation*, **9**: 1179–1209.

[18] A. L. Hodgkin, and A. F. Huxley (1952), A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol.*, **117**: 500–544.

[19] K. Imamura, N. Mataga, and K. Mori (1992), Coding of odor molecules by mitral/tufted cells in rabbit olfactory bulb. I. Aliphatic compounds, *J. Neurophysiol.*, **68**: 1986–2002.

[20] T. Kurahashi, G. Lowe, and G. H. Gold(1994), Suppression of odorant responses by odorants in olfactory receptor cells, *Science*, **265:** 118–120.

[21] G. Laurent(1999), A systems perspective on early olfactory coding, *Science*, **286**: 723–728.

[22] B. A. Johnson, and M. Leon(2000), Odorant molecular length: one aspect of the olfactory code, *J. Computat. Neurosci*, **426**: 330–338.

[23] R. Maex, and E. De Schutter(1998), Synchronization of Golgi and granule cell firing in a detailed network model of the cerebellar granule cell layer, *J. Neurophysiol.*, **80**: 2521–2537.

[24] P. Mombaerts, F. Wang, C. Dulac, S. K. Chao, A. Nemes, M. Mendelsohn, J. Edmondson, and R. Axel (1996), Visualizing an olfactory sensory map, *Cell*, **87**: 675–686.

[25] K. Mori, K. Kishi, and H. Ojima(1983), Distribution of dendrites of mitral, displaced mitral, tufted, and granule cells in the rabbit olfactory-bulb, *J. Computat. Neurosci.*, **219**: 339-355.

[26] K. Mori, H. Nagao, and Y. Yoshihara(1999), The olfactory bulb: coding and processing of odor molecule information, *Science*, **286**: 711–715.

[27] P. F. Pinsky, and J. Rinzel(1994), Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons, *Journal of Computational Neuroscience*, **1**: 39–60.

[28] S. L. Pomeroy, A. -S. LaMantia, and D. Purves(1990), Postnatal construction of neural circuitry in the mouse olfactory bulb, *J. Neurosci.*, **10**: 1952–1966.

[29] A. D. Protopapas, M. Vanier, and J. M. Bower (1998), Simulating large networks of neurons, in *Methods in Neuronal Modeling: From Ions to Networks*, C. Koch and I. Segev (eds.), MIT Press: Cambridge, Massachusetts, 461–498.

[30] W. Rall (1959), Branching dendritic trees and motoneuron membrane resistivity, *Exp. Neurol.*, **1**: 491–527.

[31] R. Ritz, and T. J. Sejnowski (1997), Synchronous oscillatory activity in sensory systems: new vistas on mechanisms, *Current Opinion in Neurobiology*, **7**: 536–546.

[32] J.-P. Royet, H. Distel, R. Hudson, and R. Gervais(1998), A re-estimation of the number of glomeruli and mitral cells in the olfactory bulb of rabbit, *Brain Research*, **788**: 35–42.

[33] B. D. Rubin, and L. C. Katz (1999), Optical imaging of odorant representations in the mammalian olfactory bulb, *Neuron*, **23**: 499–511.

[34] N. E. Schoppa, J. M. Kinzie, Y. Sahara, T. P. Segerson, and G. L. Westbrook (1998), Dendrodendritic inhibition in the olfactory bulb is driven by NMDA receptors, *J. Neurosci.*, **18**: 6790–6802.

[35] S. L. Senft (1997), A statistical framework for presenting developmental neuroanatomy., in *Neural Network Models of Cognition: Biobehavioral Foundations*, J. Donahoe and V. P. Dorsel (eds.) Elsevier Press: New York.

[36] G. M. Shepherd (1972), Synaptic organisation of the mammalian olfactory bulb, *Physio. Rev.*, **52**: 864–917.

[37] K. Stratford, A. Mason, A. Larkman, G. Major, and J. Jack (1989), The modelling of pyramidal neurones in the visual cortex, in *The Computing Neuron*, R. Durbin and C. Miall and G. Mitchison (eds.), Addison-Wesley: Wokingham, 296–321.

[38] J. Strout (1977), A library for the compartmental simulation of neurons, *Neuroscience-Net*, Article #10013.

[39] R. D. Traub, R. K. S. Wong, R. Miles, and H. Michelson (1991), A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances, *J. Neurophysiol.*, **66**: 635–650.

[40] F. Wang, A. Nemes, M. Mendelsohn, and R. Axel (1998), Odorant receptors govern the formation of a precise topographic map, *Cell*, **93**: 47–60.

[41] M. A. Wilson, U. S. Bhalla, J. D. Uhley, and J. M. Bower (1989), GENESIS: A system for simulating neural networks., in *Advances in Neural Information Processing Systems*, D. Touretzky (ed.), Kaufmann: San Mateo, CA, 485–492.

[42] T. B. Woolf, G. M. Shepherd, and C. A. Greer (1991), Local information processing in dendritic trees: subsets of spines in granule cells of the mammalian olfactory bulb, *J. Neurosci.*, **11**: 1837–1854.

Chapter 11

Hebbian Learning and Spike-Timing-Dependent Plasticity

Sen Song

Freeman Building, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY 22734, U.S.

CONTENTS

11.1 Hebbian models of plasticity

> When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place on one or both cells so that A's efficiecncy as one of

the cells firing B is increased. — Donald Hebb, 1949

In his famous book published a little more than half a century ago, *The Organization of Behavior,* Hebb proposed a hypothetical mechanism by which 'cell assemblies', which are a group of cells that could act like a form of 'short term memory' and support self-sustaining reveberatory activity outlasting the input, could be constructed [37]. These suggestions were later extended into other areas and now serve as the basis for a large body of thinking concerning activity-dependent processes in development, learning, and memory [7, 13, 28, 35, 52, 74, 100]. What Hebb proposed was an elegant way for correlated, i.e., interesting, features of an input stimulus to become permanently imprinted in the architecture of neural circuits to alter subsequent behavior, which is the hallmark of learning. It is similar in form to classical conditioning in the psychology literature. Many models have subsequently been constructed based on extensions of this simple rule, now commonly called the Hebbian rule. These models have given reasonable accounts of many aspects of development and learning [44, 48, 58, 59, 62, 70, 73, 75, 77, 90, 95, 97]. In this chapter, we will not attempt to review the literature on Hebbian learning exhaustively. Instead, we will try to review some relevant facts from the Hebbian learning literature and discuss their connections to spike-timing-dependent plasticity (STDP), which are based on recent experimental data. To discuss Hebbian learning and STDP in a coherent mathematical framework, we need to introduce some formalism. Let us consider one neuron receiving many inputs labelled 1 to N and denote the instantaneous rate for the *ith* input as r_i^{in} and the output as (r^{out}). The integration performed by the neuron could be written as

$$\tau_m \frac{dr^{out}(t)}{dt} = G \sum_i w_i r_i^{in}(t), \tag{11.1}$$

where $r^{out}(t)$ is the instantaneous firing rate of the output neuron at time t, G is a constant gain factor for the neuron, w_i is the synaptic strength of the ith input, and $r_i^{in}(t)$ is the instantaneous firing rate of the ith input at time t. Solving the differential equation, we have

$$r^{out}(t) = G \int_0^\infty dt' K(t') \sum_i w_i r_i^{in}(t-t') - \theta, \tag{11.2}$$

with

$$K(t) = \frac{1}{\tau_m} e^{-t/\tau_m}. \tag{11.3}$$

$K(t)$ is a kernel function used to simulate the membrane integration performed by the neuron and θ is the threshold. Therefore, the rate of a given neuron is linearly dependent on the total amount of input into the neuron over the recent past, with exponentially more emphasis on the most recent inputs. For the sake of simplicity, we do not include the rectifying nonlinearity introduced by the threshold and only consider the regime above threshold. If we assume that plasticity is on a slower time scale than changes in firing rates and only depend on the average firing rates, we can further simplify Equation (11.2) to

$$r^{out}(t) = G\sum_i w_i r_i^{in}(t) - \theta. \tag{11.4}$$

This simplification is however not appropriate when we consider plasticity rules that depend on spike time later in this chapter. Computationally, the simplest rule that follows from Hebb's idea is probably

$$\tau_w \frac{dw_i(t)}{dt} = r^{out}(t) r_i^{in}(t). \tag{11.5}$$

We will drop the '(t)' term in subsequent equations and the reader should note that all these entities represent functions in time. After plugging in Equation (11.4), we have

$$\tau_w \frac{dw_i}{dt} = G\sum_j w_j r_j^{in} r_i^{in}, \tag{11.6}$$

If we average over a long time, we can write the rule as

$$\tau_w \frac{dw_i}{dt} = G\sum_j w_j Q_{ij}, \tag{11.7}$$

where $Q_{ij}(t) =< r_j^{in}(t) r_i^{in}(t) >$ represents the average correlation between the inputs over the training sets. If this rule is applied for a long enough time, the weight vectors would pick out the principal eigenvector of the correlation matrix of the inputs the neuron experienced [62].

11.2 Spike-timing dependent plasticity

More recently, many studies have focused on spike-timing dependent plasticity rules (STDP). These rules are inspired by recent studies that have tested the role of timing in synaptic plasticity by directly controling the timing of pre- and postsynaptic spikes. Markram et al. [51], using dual intracellular recordings, found that if neuron A repeatedly fired 10ms before neuron B, the connection from neuron A to neuron B was strengthened. However, if neuron A consistently fired 10 ms after neuron B, the connection was weakened. Time separations of 100ms between pre- and postsynaptic spikes were ineffective in inducing synaptic plasticity. Similar results have been found in hippocampal slice [21, 49], hippocampal culture [21], somatosensory cortical slice [26], and visual cortical slice [78]. Zhang et. al. [101], working with an *in vivo* preparation of the optic tectum of frogs, have documented the relationship between changes in synaptic strength and relative timing in great detail. Earlier studies done with field recordings suggest essentially the same rule [36, 46]. An inverse form of STDP was found for inhibitory synapses in the cerebellum of the electrical fish [5]. Egger *et al.* [23] found a symmetric form of STDP for spiny stellate neurons in visual cortical slices. The mainly temporally asymmetric form of the spike-timing dependent plasticity (STDP) rule found in these experimental studies has great attractions. The inclusion of both potentiation and depression in one rule addresses the problem of coordinating LTP and LTD at one synapse by stressing causality as a condition for strengthening. Chance coincidences occur with roughly equal positive and negative time delays and only truly causal inputs have a consistent timing relationship to the postsynaptic spike. Only the causal inputs will be strengthened under STDP. Furthermore, the dependence on time also makes this rule useful for learning tasks involving temporal patterns. Recently, Yao and Dan [98] used pairs of gratings at slightly different orientations presented in rapid succession to induce bi-directional changes in orientation selectivity that depended on the timing of the two gratings. The time scale of stimulus separation for effective orientation plasticity is similar to the time scale of STDP. They went on to characterize this effect physiologically [30]. Schuett *et al.* [71] paired a brief visual stimulus with electrical stimulation in the visual cortex and found alteration in the cortical orientation maps using optical imaging. These studies furnish an important link between STDP on the cellular level and plasticity on the physiological and perceptual levels, and directly relate to theoretical work reviewed in this chapter. In the framework of rates outlined in the previous section, we can represent STDP as:

$$\frac{dw_i}{dt} = \int_{-\infty}^{\infty} dt' P(t') r^{out}(t) r_i^{in}(t+t'). \tag{11.8}$$

When compared with Equation (11.5), the explicit dependency on the the temporal relationship of presynaptic and postsynaptic rates is apparent. The τ_w term is absent because a factor of similar nature is present in $P(t')$ for the amount of change per pair of spikes. In this review, we consider STDP rules of the temporally asymmetric kind and write

$$P(\Delta t) = \begin{cases} \frac{A_+}{\tau_+} e^{(\Delta t/\tau_+)} & \text{if } \Delta t < 0 \\ \frac{-A_-}{\tau_-} e^{(-\Delta t/\tau_-)} & \text{if } \Delta t \geq 0\,, \end{cases} \tag{11.9}$$

where $P(\Delta t)$ indicates the amount of of change in synaptic strength for a pair of pre- and postsynaptic spikes seperated by time $\Delta t (t_{pre} - t_{post} = \Delta t)$. we impose a lower bound of zero and a upper bound of g_{max} on the synapses, and $P(\Delta t)$ is expressed as a fraction of the maximal synaptic strength g_{max}. We decided to use an exponential function based on curve fits to experimental data, especially that contained in Zhang *et al.* [101]. The constants τ_+ and τ_- determine the time scale of spike pair separation over which STDP is operating. Bi and Poo [7] gives an excellent review of this issue. They have noted that the time scale over which the synaptic strengthening portion of STDP operates (τ_+) is mostly around 20-30 ms, while the synaptic weakening portion (τ_-) has a more variable time course. Most of the studies cited in this chapter are preformed with a τ_- of around 20 ms, while we have also studied effects of STDP with $\tau_- = 100$ms. A_+ and A_- determine the maximum amounts of synaptic modification for each spike pair. The studies cited in this chapter have assumed that the modification of each spike pair sum linearly. Some recent experiments have demonstrated nonlinear summations [29, 78], which needs to be addressed in future investigations (see [43]) for some attempts). The STDP functions with both τ_- values are plotted in Figure 11.1.

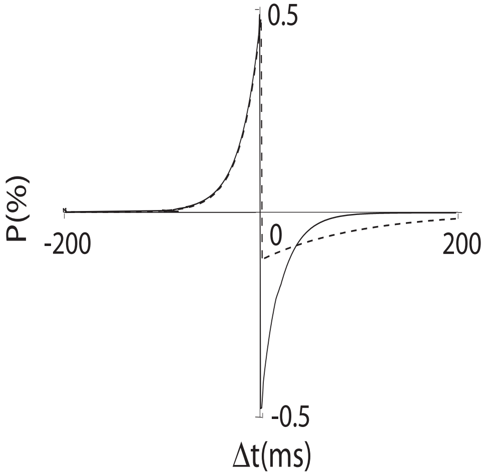

Figure 11.1

The STDP modification function. The change of the peak conductance at a synapse due to a single pre- and postsynaptic action potential pair is $F(\Delta t)$ times the maximum value g_{max}. Δt is the time of the presynaptic spike minus the time of the postsynaptic spike. In this figure, P is expressed as a percentage. Solid line represents the form of STDP with $\tau_- = 20ms$ and dotted line has $\tau_- = 100ms$. Adapted with permission from [79].

11.3 Role of constraints in Hebbian learning

The simple Hebbian rule outlined in Section 11.1 is in general unstable. The total synaptic strengths will continue to grow over time. The connection strength from correlated inputs would continually increase and eventually drive the postsynaptic cell to fire at catastrophic rates. An upper bound imposed on synaptic strengths alone cannot ameliorate this situation as all connection strengths would eventually reach the upper bound, thus rendering the postsynaptic neuron equally responsive to all inputs. Furthermore, the simple Hebbian rule lacks competition among inputs, strengthening of one set of inputs does not automatically lead to decreased synaptic strengths for other inputs. Several modifications to the basic Hebbian rule have been suggested to address these issues.

11.3.1 Covariance rule

Because neuronal firing rates are always positive, there is no synaptic weakening under the simple Hebbian rule. As a first step, let us consider rules where the synapse is strengthened when the presynaptic rate or the postsynaptic rate is above average. Otherwise, the synapse is weakened. Mathematically,

$$\tau_w \frac{dw_i}{dt} = (r_i^{in} - \overline{r_i^{in}})(r^{out} - \overline{r^{out}}), \tag{11.10}$$

where the overlined quantities represent the average rates [75, 20]. Plugging in Equation (11.4), this leads to

$$\tau_w \frac{dw_i}{dt} = G \sum_j w_j C_{ij}, \tag{11.11}$$

where $C_{ij} =< (r_j^{in} - \overline{r_j^{in}})(r_i^{in} - \overline{r_i^{in}}) >$ represents the average covariance between the inputs over the training sets and the rule is called the covariance rule. Notice that $Q_{ij} = C_{ij} + \overline{r_i^{in} r_j^{in}}$, so the original simple Hebbian rule can be written as

$$\tau_w \frac{dw_i}{dt} = G \sum_j w_j (C_{ij} + \overline{r_i^{in} r_j^{in}}). \tag{11.12}$$

The mean rate is subtracted out under the covariance rule compared to the simple Hebbian rule. In all subsequent sections, we will express the rule in terms of the covariance C rather than the correlation Q. The covariance rule, unfortunately, is still not stable by itself. Although the mean activity alone does not make the synapses grow, correlated activity among inputs will continue to drive all the synapses towards the upper bound.

11.3.2 Constraints based on postsynaptic rate

In order to have a stable learning rule and introduce competition among inputs, most of the studies applying the Hebbian rule rely on some additional mechanism to ensure competition among different inputs. The are two main ways to achieve it, either by constraining the postsynaptic activity or by constraining the total synaptic weights. In this section, we will introduce rules which constrain the synapses based on postsynaptic rates. Later on, we will also see that STDP can also constrain the synapses to keep the postsynatic rate constant. In the next section, we will introduce rules that constrain the total synaptic weights. It should be noted that if the presynaptic rate is fixed, constraining the total synaptic strengths will also constrain the postsynaptic rate and vice versa. One example of constraints based on the postsynaptic rate is the BCM model. In the BCM model, correlated pre- and postsynaptic activity induces LTP when the postsynaptic firing rate is above a threshold and induces LTD when it is below the threshold. Mathematically,

$$\tau_w \frac{d}{dt} w_i = \eta \Phi(r^{out}, r_\theta) r_i^{in}, \tag{11.13}$$

where η is a constant, Φ is a nonlinear function and r_θ is the threshold. Φ is zero when r^{out} is smaller than a certain value r_0. It is negative when r^{out} is between r_0 and r_θ. It is positive when r^{out} is larger than r^θ. An often used choice is $\Phi = r^{out}(r^{out} - r_\theta)$. Sliding the threshold based on the average postsynaptic firing rate confers stability [8]. This introduces competition as inputs strive to raise the postsynaptic firing rate above the threshold which is determined by the average rate given by the mean activity of all inputs. Another example is synaptic scaling based on postsynaptic rates [22, 47, 86, 87, 88, 89, 92]. Mathematically,

$$\begin{aligned} &\tau_w \tfrac{dw_i}{dt} = r_i^{in} r^{out} - \beta w_i (a^{out} - a^{out}_{goal}) \\ &= \textstyle\sum_j Q_{ij} w_j - \beta w_i (a^{out} - a^{out}_{goal}) \\ &= \textstyle\sum_j C_{ij} w_j + \sum_j \overline{r_i^{in} r_j^{in}}) w_j - \beta w_i (a^{out} - a^{out}_{goal}) \\ &= \textstyle\sum_j C_{ij} w_j + \overline{r_i^{in}}\,\overline{r^{out}} - \beta w_i (a^{out} - a^{out}_{goal}), \end{aligned} \tag{11.14}$$

where a^{out} is an averaged version of r^{out} and a^{out}_{goal} is a constant goal value for a^{out}. The synaptic weights stop changing when the right side is equal to zero. The final r^{out} will depend on the correlation and value of a^{out}_{goal} , but it is constrained and would not grow without bound.

11.3.3 Constraints on total synaptic weights

Another way to achieve stability is to constrain the total synaptic weights to be constant. Miller classified these types of constraints into two classes: multiplicative and subtractive constraints [61]. These rules also enforce competition because increased strength in one set of synapses will lead to decrease in synaptic strengths in other inputs as the total synaptic weight is kept constant. In multiplicative constraints, the underlying mechanism effectively multiplies the strength of each synapse by a factor after each application of the Hebbian rule, so the total synaptic strength is kept constant. Mathematically,

$$\begin{aligned} &\tau_w \tfrac{dw_i}{dt} = \textstyle\sum_j Q_{ij} w_j - \gamma(\overrightarrow{w}) w_i \\ &= \textstyle\sum_j C_{ij} w_j + \sum_j \overline{r_i^{in} r_j^{in}} w_j - \gamma(\overrightarrow{w}) w_i \\ &= \textstyle\sum_j C_{ij} w_j + \overline{r_i^{in}}\,\overline{r^{out}} - \gamma(\overrightarrow{w}) w_i, \end{aligned} \tag{11.15}$$

where $\gamma(\overrightarrow{w})$ is a function of all the synaptic weights. For an explanation of how the subtractive factor in the equation above is the same as multiplying w_i by a factor after each time step, please see [61]. Under multiplicative constraints, the principal eigenvalue is amplified and the final weight eventually approaches the principal eigenvalue. A famous example of multiplicative constraint is the Oja rule, where

$$\gamma(\overrightarrow{w}) = \alpha^2 (r^{out})^2, \tag{11.16}$$

and α is a constant that determines the total synaptic weight [62]. In subtractive constraints, a factor independent of current strength is subtracted from the strength of each synapse after each application of the Hebbian rule. Mathematically,

$$\begin{aligned} \tau_w \frac{dw_i}{dt} &= \sum_j Q_{ij} w_j - e(\overrightarrow{w}) \\ &= \sum_j C_{ij} w_j + \sum_j \overline{r_i^{in} r_j^{in}} w_j - e(\overrightarrow{w}) \\ &= \sum_j C_{ij} w_j + \overline{r_i^{in}}\,\overline{r^{out}} - e(\overrightarrow{w}), \end{aligned} \tag{11.17}$$

where $e(\overrightarrow{w})$ is a function of all the synaptic weights. Under subtractive constraints, if the principal eigenvalue is zero-summed, then the principal eigenvalue is amplified and leads to saturation. Otherwise, a zero sum vector that grows to complete saturation imposed on the constraint background is the final result [61]. As an example, let us consider a neuron receiving 1000 inputs. The correlation between inputs i and j is given by

$$< r_i(t) r_j(t') >= \bar{r}^2 + \bar{r}_{in}^2 (\sigma^2 \delta_{ij} + (1 - \delta_{ij}) c_i c_j) e^{-|t-t'|/\tau_{corr}} \tag{11.18}$$

with $\bar{r}_{in} = 10Hz$, $\sigma = 0.5$, $\tau_{corr} = 20ms$, and c_i, which we call the correlation parameter, varied from zero to 0.2 uniformly across the 1000 excitatory synapses ($c_i = 0.2(i-1)/(N-1)$). This graded variation in c_i values produces a gradation of correlation across the population of synapses. Pairs of synapses with large c values are more highly correlated than synapses with smaller c values. This is achieved in the simulations by changing the presynaptic rates after each interval of time chosen from an exponential distribution with mean interval τ_{corr}. For every interval, we generate $N+1$ random numbers, y and x_i for $i = 1, 2, \ldots, N$ from Gaussian distributions with zero mean and standard deviation one and σ_i respectively, where $\sigma_i = \sqrt{\sigma^2 - c_i^2}$. At the start of each interval, the firing rate for synapse i is set to $r_i = \bar{r}_{in}(1 + x_i + c_i y)$, and it is held at this value until the start of the next interval (see [79] for details). The final synaptic strengths under multiplicative constraints is plotted in Figure 11.2A and reproduces a scaled version of the principal eigenvector of the correlation matrix quite faithfully. Synapses show a graded transition from weak to strong as the correlation strength grows. The final synaptic strengths under subtractive constraints is plotted in Figure 11.2B and the situation is quite different. Synapses with correlation strength above a threshold all had maximal synaptic strengths while synapses with correlation strength below the threshold had synaptic strengths of zero. we have also plotted the final synaptic strengths under the BCM rule introduced in the previous section in Figure 11.2C. STDP gives a gradient of synapse strength, and will be discussed in detail in Section 11.4.5.3 (Figure 11.2D).

Another example is the formation of ocular dominant cells. In this scenario, the postsynaptic cell receives inputs from two eyes. Inputs from the same eye are correlated while inputs from different eyes are either anti-correlated or uncorrelated. Under subtractive constraints, inputs from different eyes need only to be uncorrelated for the postsynaptic cell to finally receive inputs exclusively from one eye while anti-correlation is needed in the case of multiplicative constraints [61].

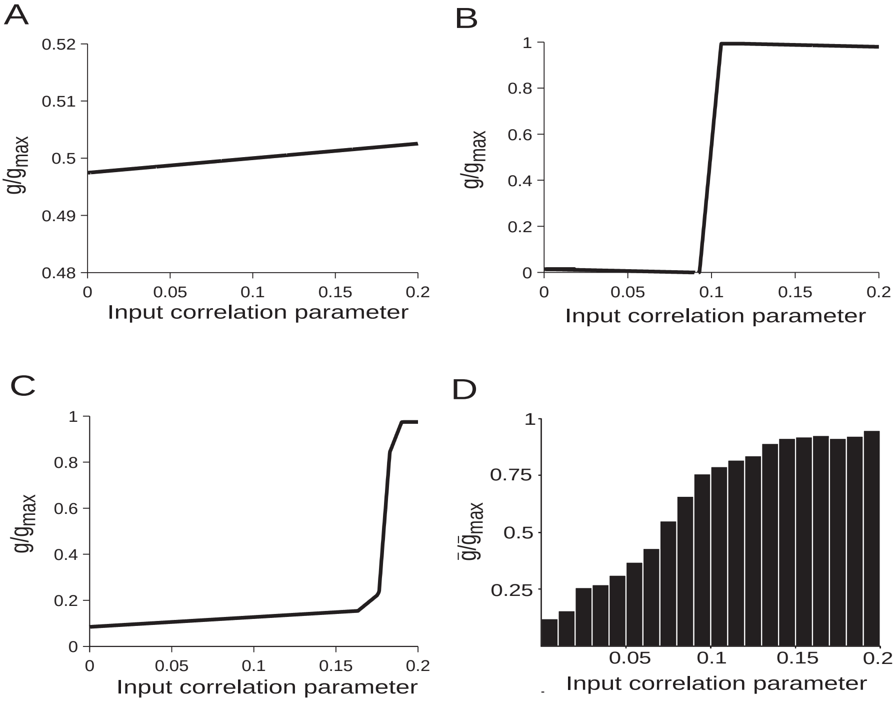

Figure 11.2

Final synaptic strengths for different learning rules for inputs with gradient of correlations. (A) Rate-based Hebbian rule and multiplicative constraint; (B) rate-based Hebbian rule and subtractive constraint; (C) BCM rule; (D) STDP. (D) Adapted with permission from [79].

11.4 Competitive Hebbian learning through STDP

11.4.1 STDP is stable and competitive by itself

A perhaps surprising result is that STDP by itself is stable and competitive [79]. How are synapses constrained and how is competition achieved in this case? Intuitively, changes in the strength of one synapse would shift postsynaptic spike timing and affect the synaptic strengths of other synapses. All the inputs therefore compete for control of the postsynaptic spike timing. In response to Poisson inputs, individual synapses tend to the bounds imposed on the synapse. However, the overall distribution of synaptic strengths is stable and bimodal (Figure 11.3A) [79] (However, see [43, 68, 89] for conditions where this might not hold). Furthermore, the postsynaptic rate is kept constant if we vary the average input rates (Figure 11.3C). The coefficient of variation (CV) is also kept at a constant high value near 1.0 indicating an irregu-

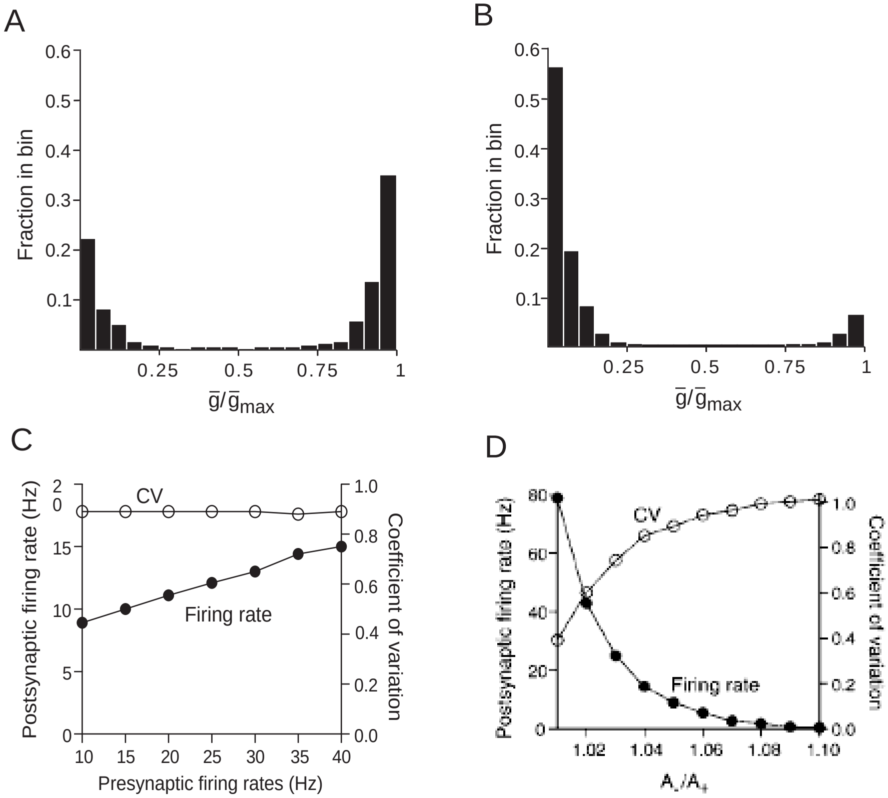

Figure 11.3

STDP results in stable distribution of synaptic strength and constrains postsynaptic firing rate. (A) Histogram of synaptic strength for input rate of 10Hz. (B) Histogram of synaptic strength for input rate of 20Hz. (C) Postsynaptic firing rate and coefficient of variation (CV) stay constant for different input firing rates. (D) Postsynaptic firing rate is affected by A_{-}/A_{+}. Adapted from [79].

lar spike train. This is achieved by shifting the proportion of synapses near the two bounds (compare Figure 11.3B and A). The postsynaptic rate is however affected by the ratio A_{-}/A_{+} in the STDP function P [79] (Figure 11.3D).

11.4.2 Temporal correlation between inputs and output neuron

To understand how this is achieved in detail, let us see what insights the rate framework can provide us. In order to calculate the distribution of synaptic strengths under STDP, we first need to know the correlation between the presynaptic and postsynaptic spike trains. We can write it as

$$\begin{aligned}
&< r^{out}(t')r_i^{in}(t'+t) >_{t'} \\
&= G < \int_0^\infty dt'' K(t'') \textstyle\sum_j w_j r_j^{in}(t'-t'') r_i^{in}(t'+t) >_{t'} - \theta < r_i^{in} > \\
&= G \int_0^\infty dt'' K(t'') \textstyle\sum_j w_j < r_i^{in}(t'+t) r_j^{in}(t'-t'') >'_{t'} - \theta < r_i^{in} > .
\end{aligned} \tag{11.19}$$

If we define $< r_i^{in}(t'-t) r_j^{in}(t'-t'') >_{t'}$ to be $Q_{ij}(t-t'')$, then

$$< r^{out}(t')r_i^{in}(t'+t) >_{t'} = G \int_0^\infty dt'' K(t'') \sum_j w_j Q_{ij}(-t-t'') - \theta < r_i^{in} > . \tag{11.20}$$

Let us assume that all the inputs have the same mean firing rate $\overline{r_{in}}$ and the correlation between any two inputs i and j is

$$Q_{ij}(t) = \overline{r_{in}}^2 (C_{ij} e^{-|t|/\tau_c} + 1) + \delta(t) \overline{r_{in}} \delta_{ij}. \tag{11.21}$$

Therefore the correlation between the presynaptic and postsynaptic spike trains is

$$\begin{aligned}
&< r^{out}(t')r_i^{in}(t'+t) >_{t'} \\
&= G \int_0^\infty dt'' \tfrac{1}{\tau_m} e^{-t''/\tau_m} \textstyle\sum_j w_j \overline{r_{in}}^2 (C_{ij} e^{-|t+t''|/\tau_c} + 1) + \delta(t) \overline{r_{in}} \delta_{ij} - \theta \overline{rth_{in}} \\
&= -\theta \overline{r_{in}} \\
&+G \begin{cases} \int_0^{-t} dt'' \tfrac{1}{\tau_m} e^{-t''/\tau_m}\, e^{t-t''))/\tau_c} \textstyle\sum_j w_j \overline{r_{in}}^2 C_{ij} + \tfrac{w_i \overline{r_{in}}}{\tau_m} e^{t/\tau_m} & \\ \quad + \int_{-t}^\infty dt'' \tfrac{1}{\tau_m} e^{-t''/\tau_m} e^{t''-t/\tau_c} \textstyle\sum_j w_j \overline{r_{in}}^2 C_{ij} + \textstyle\sum_j w_j \overline{r_{in}}^2 & \text{if} \quad t < 0 \\ \int_0^\infty dt'' \tfrac{1}{\tau_m} e^{-t''/\tau_m} e^{t-t''/\tau_c} \textstyle\sum_j w_j C_{ij} \overline{r_{in}}^2 + \textstyle\sum_j w_j \overline{r_{in}}^2 & \text{if} \quad t \geq 0 \end{cases} \\
&= -\theta \overline{r_{in}} \\
&+G \begin{cases} \textstyle\sum_j w_j C_{ij} \overline{r_{in}}^2 [\tfrac{\tau_c}{\tau_c - \tau_m} (e^{t/\tau_c} - e^{t/\tau_m}) & \\ + \tfrac{\tau_c}{\tau_c + \tau_m} e^{t/\tau_m}] + \textstyle\sum_j w_j \overline{r_{in}}^2 + \tfrac{w_i \overline{r_{in}}}{\tau_m} e^{t/\tau_m} & \text{if} \quad t < 0 \\ \textstyle\sum_j w_j C_{ij} \overline{r_{in}}^2 \tfrac{\tau_c}{\tau_c + \tau_m} e^{t/\tau_m} + \textstyle\sum_j w_j \overline{r_{in}}^2 & \text{if} \quad t \geq 0 \end{cases}
\end{aligned} \tag{11.22}$$

11.4.3 Mean rate of change in synaptic strength

We can now calculated the mean rate of change in synaptic strength induced by STDP.

$$\frac{dw_i(t)}{dt} = \int_{-\infty}^\infty dt' P(t') r^{out}(t) r_i^{in}(t+t'), \tag{11.23}$$

where $P(t)$ is the STDP plasticity function and $r_i^{in}(t)$ and $r^{out}(t)$ are the firing rates for the ith input and the output respectively. From Section 11.2,

$$P(t) = \begin{cases} \frac{A_+}{\tau_+} e^{t/\tau_+} & \text{if} \quad t < 0 \\ -\frac{A_-}{\tau_-} e^{-t/\tau_-} & \text{if} \quad t \geq 0 \end{cases} \tag{11.24}$$

From Section 11.1,

$$r^{out}(t) = G \int_0^\infty dt' K(t') \sum_i w_i r_i^{in}(t-t') - \theta. \tag{11.25}$$

Therefore,

$$\begin{aligned}&\tfrac{dw_i(t)}{dt}\\ &= G\textstyle\int_0^\infty dt' \int_{-\infty}^\infty dt'' K(t')P(t'')\sum_j w_j r_j^{in}(t-t')r_i^{in}(t+t'') - \theta\overline{r_{in}}(A_+ - A_-).\end{aligned} \tag{11.26}$$

For simplicity, let us assume that the scale of STDP is slow compared to the spiking rate of the neurons, we can therefore replace $r_j^{in}(t-t')r_i^{in}(t+t'')$ with its average $< r_j^{in}(t-t')r_i^{in}(t+t'') >_t$, or $Q_{ij}(-t'-t'')$. We assume that $Q_{ij}(-t'-t'')$ is stationary for all subsequent calculations. We can now calculate the rate of change in synaptic strengths for correlated inputs. From Section 11.1,

$$Q_{ij}(t) = C_{ij}\overline{r_{in}}^2 e^{-|t|/\tau_c} + \overline{r_{in}}^2 + \delta(t)\overline{r_{in}}\delta_{ij}, \tag{11.27}$$

and

$$K(t) = \frac{1}{\tau_m}e^{-t/\tau_m}. \tag{11.28}$$

Upon collecting terms,

$$\begin{aligned}&\tfrac{dw_i}{dt} = G(H\overline{r_{in}}^2\textstyle\sum_j C_{ij}w_j - (A_- - A_+)\overline{r_{in}}^2\sum_j w_j + \overline{r_{in}}w_iA_+\frac{\tau_+ + \tau_m}{\tau_+\tau_m})'')\\ &-\theta\overline{r_{in}}(A_+ - A_-),\end{aligned} \tag{11.29}$$

where

$$\begin{aligned}H &= \int_0^\infty dt' \frac{1}{\tau_m}e^{-t'/\tau_m}\{\frac{A_+}{\tau_+}[\int_{-t'}^0 dt'' e^{t''/\tau_+}e^{-(t'+t'')/\tau_c}\\ &\quad + \int_{-\infty}^{-t'} dt'' e^{t''/\tau_+}e^{t''+t')/\tau_c}] - \frac{A_-}{\tau_-}\int_0^\infty dt'' e^{-t''/\tau_-}e^{-(t'+t'')/\tau_c}\}\\ &= \frac{A_+\tau_c^2}{(\tau_c - \tau_+)(\tau_c + \tau_m)} - \frac{A_+\tau_c\tau_+}{(\tau_c - \tau_+)(\tau_+ + \tau_m)} + \frac{A_+\tau_c\tau_+}{(\tau_c + \tau_+)(\tau_+ + \tau_m)}\,.\\ &\quad -\tfrac{A_-\tau_c^2}{(\tau_c+\tau_-)(\tau_c+\tau_m)}\\ &= \frac{\tau_c^3}{(\tau_c + \tau_m)(\tau_c + \tau_+)(\tau_c + \tau_-)}[A_+ - A_- + A_+(\frac{\tau_-}{\tau_c}) - A_-(\frac{\tau_+}{\tau_c})\\ &\quad + \frac{2A_+\tau_m\tau_+(\tau_c + \tau_-)}{\tau_c^2(\tau_+ + \tau_m)}].\end{aligned} \tag{11.30}$$

The first term in the brackets in Equation (11.29) corresponds to the effect of correlations in the inputs, the second term corresponds to the effect of the average input, and the last term takes into account the correlation introduced by spiking. It is interesting that STDP can switch from a Hebbian to an anti-Hebbian rule depending on the time scale of the input correlation. In order for the rule to be Hebbian, H has to be greater than zero. Otherwise, it is anti-Hebbian. Solving $H > 0$ for τ_c yields

$$\tau_c < \frac{x - \sqrt{x^2 - 4(A_- - A_+)y}}{2(A_- - A_+)}, \tag{11.31}$$

where

$$x = A_+\tau_- - A_-\tau_+ + 2\frac{A_+\tau_m\tau_+}{\tau_+ + \tau_m}, \tag{11.32}$$

and

$$y = \frac{2A_+ \tau_m \tau_+ \tau_-}{\tau_+ + \tau_m}. \tag{11.33}$$

For $\tau_m = 6$ms, $\tau_+ = \tau_- = 20$ms, and $A_-/A_+ = 1.05$, τ_c needs to be less than approximately 250 ms for the rule to be Hebbian. As discussed before, Miller and MacKay [61] noted that normalization of the postsynaptic weights is essential for Hebbian learning to function properly and suggested two general ways that normalization could be achieved, multiplicative and subtractive normalizations. They defined subtractive normalization to be

$$\frac{dw_i}{dt} = \sum_j C_{ij} w_j - e(\overrightarrow{w}). \tag{11.34}$$

Therefore, STDP could be viewed as a form of subtractive constraint if we define $e(\overrightarrow{w}) = \frac{(A_- - A_+)}{GH} \sum_j w_j$ and drop the last two terms, which does not depend on input correlations. If all the inputs do not have the same average rate, the normalization is not exact and STDP seems to act to keep the postsynaptic neuron at a constant firing rate (See also [42]). However, there is an important difference between STDP and rate based Hebbian rule with subtractive constraint. As shown in the previous section, under subtractive constraint, if a group of synapses have correlations above a threshold, they tend to go to maximal synaptic strengths and synapses with correlations below threshold tend to have zero strength. Individual synapses under STDP still tend to adopt maximal or zero strength. However, as a small group, the average synaptic strengths can show gradation as the correlation is varied as demonstrated in Figure 11.2D.

11.4.4 Equilibrium synaptic strengths

To analytically calculate the resulting synaptic weight distribution from the STDP rule and the neuron model, several authors have taken a Fokker-Planck approach [16, 68, 89]. Here we adapted their results to fit the present formalism. Any single synapse continuously undergoes weight changes according to the plasticity rule and the timing relationships between the presynaptic and postsynaptic neurons. It is therefore most appropriate to write down a probabilistic equation for its strength. We denote its distribution with $P(w,t)$, where t denotes the time and w denotes its weight. If we assume the changes in synaptic weights is small during each time step, the weight of the synapse can be described as a biased random walk. We can write the following partial differential equation in the synaptic weight distribution $P(w,t)$,which basically counts the influx and outflux of weights for a given bin in the histogram of synaptic strengths:

$$\frac{\partial P(w,t)}{\partial t} = -\frac{\partial}{\partial w}[A(w)P(w,t)] + \frac{\partial^2}{\partial w^2}[D(w)P(w,t)]. \tag{11.35}$$

The drift term A(w) indicates the net weight ‘force field’ experienced by an individual synapse, which is calculated in the previous section as dw/dt (Equation 11.29).

Whether a synapse increases or decreases in strength depends on the sign of $A(w)$ for that particular weight value. The diffusion term of the random walk can be calculated as

$$D(w) = \int_{-\infty}^{\infty} dt' P(t')^2 r^{out}(t) r_i^{in}(t+t'). \tag{11.36}$$

A full derivation of A(w) and D(w) for the spiking model can be found in [16]. Because we have imposed both a lower bound of zero and upper bound of g_{max} on the synaptic strengths, this equation has the following boundary conditions.

$$J(0,t) = J(g_{max},t) = 0, \tag{11.37}$$

where $J(w,t)$ is the probability current defined by

$$J(w,t) = A(w)P(w,t) - \frac{1}{2}\frac{\partial}{\partial w}(D(w)P(w,t)). \tag{11.38}$$

According to Fokker-Planck theory, the equilibrium density $P(w)$ can be described as a Gibbs distribution with a plasticity potential $U(w)$, where in the limit of small step size

$$U(w) \approx -2\int_0^w dw' A(w')/D(w'). \tag{11.39}$$

So, $P(w)$ will be concentrated near the global minima of $U(w)$ and have a spread that is proportional the drift term and thus the step size. The minimum can be located at an interior point where the drift term vanishes or at the boundaries. In the situations described in this chapter, the equilibrium is typically located at the boundaries([68]). If A_-/A_+ is set to be slightly larger than 1, the weak negative bias in the plasticity function can balance the positive correlations. Under these conditions, the diffusion constant D is approximately constant and given by

$$D \approx (G\overline{r_{in}}^2 \sum_{j\neq i} w_j - \theta\overline{r_{in}})((A_-^2/\tau_- + A_+^2/\tau_-)/2. \tag{11.40}$$

The potential is given by

$$\begin{aligned} &U(w_i) \approx 2(-G(H\overline{r_{in}}^2 C_{ii} - (A_- - A_+)\overline{r_{in}}^2 + \overline{r_{in}}A_+ \tfrac{\tau_+ + \tau_m}{\tau_+\tau_m})w_i^2 \\ &-2(GH\overline{r_{in}}^2 \textstyle\sum_{j\neq i} C_{ij}w_j - G(A_- - A_+)\overline{r_{in}}^2 \sum_{j\neq i} w_j + \theta\overline{r_{in}}(A_- - A_+)w_i) \\ &/(G\overline{r_{in}}^2 \textstyle\sum_{j\neq i} w_j - \theta\overline{r_{in}})((A_-^2/\tau_- + A_+^2/\tau_-)). \end{aligned} \tag{11.41}$$

$P(w_i)$ is given by $P(w_i) = Ne^{-U(w_i)}$, where N is a normalization factor. The expected value of w_i is then

$$E_i = \int_0^{g_{max}} dw' w' P(w'), \tag{11.42}$$

Since the distribution of synaptic strength for each synapse depends on the strength of other synapses(interpreted as the expected value), these equations need to be solved for self-consistency.

Before we dive into detailed solutions of those quite complex equations, let us try to gain some intuitive understanding of these equations. Let us write $A(w)$ as $aw+b$, where $a=G(H\overline{r_{in}}^2C_{ii}-(A_{-}-A_{+})\overline{r_{in}}^2+\overline{r_{in}}A_{+}\frac{\tau_{+}+\tau_m}{\tau_{+}\tau_m})$,and $b=(GH\overline{r_{in}}^2\sum_{j\neq i}C_{ij}w_j-G(A_{-}-A_{+})\overline{r_{in}}^2\sum_{j\neq i}w_j+\theta\overline{r_{in}}(A_{-}-A_{+})$. In Figure 11.4A, we have plotted $A(w)$ for different values of a and b. In this figure, w is expressed as a fraction of g_{max}. Under the conditions considered in this chapter, a is always positive, the fix point where the line crosses the x axis is thus unstable and the only stable fix points are located at the boundaries. For $D=0.05$, the corresponding $U(w)$ and $P(w)$ are given in Figure 11.4B and C. $U(w)$ has local minima at both boundaries and $P(w)$ shows corresponding concentration of synaptic weights in the histogram. Under the conditions considered in this chapter, solutions of $P(w)$ are unsaturated and $P(w)$ has significant weights at both boundaries. Therefore $U(0)\approx U(g_{max})$, which means $g_{max}(2b+ag_{max})/D\approx 0$. The total weight $\sum_j w_j$ will be adjusted until this condition is approximately satisfied. A slight deviation of this quantity from 0 will shift the relative proportion of weights near lower and upper bounds. a determines the width of two lobes of the distribution near lower and upper bound. We plot the expected synaptic strength versus this quantity in Figure 11.4D to demonstrate how changing this quantity will shift the relative proportion of weights in the two lobes and determine the mean synaptic strength. The curve is fairly steep, therefore a slight change in the term involving C_{ij} will have big effects on the mean synaptic strength. Notice that the central portion of the curve is almost linear. Therefore we could write, for $2b/D+ag_{max}/D\approx 0$,

$$\begin{aligned}
&E(P(w_i))=Sg_{max}(2b+ag_{max})/D+0.5g_{max}\\
&=2Sg_{max}(-G(H\overline{r_{in}}^2C_{ii}-(A_{-}-A_{+})\overline{r_{in}}^2+\overline{r_{in}}A_{+}\tfrac{\tau_{+}+\tau_m}{\tau_{+}\tau_m})w_i^2\\
&-2(GH\overline{r_{in}}^2\textstyle\sum_{j\neq i}C_{ij}w_j-G(A_{-}-A_{+})\overline{r_{in}}^2\sum_{j\neq i}w_j+\theta\overline{r_{in}}(A_{-}-A_{+})w_i)\\
&/(G\overline{r_{in}}^2\textstyle\sum_{j\neq i}w_j-\theta\overline{r_{in}})(A_{-}^2/\tau_{-}+A_{+}^2/\tau_{-})+0.5g_{max},
\end{aligned}\tag{11.43}$$

where S is a scaling factor around 0.25.

11.4.5 Three common scenarios and comparison to simulations

To make the mathematical expressions derived in the previous sections more intuitively apparent, we will consider three common scenarios and compare the results derived from the analytical calculations to those from simulations performed with an integrate and fire neuron receiving 1000 Poisson inputs [79].

11.4.5.1 Constant Poisson inputs

The first situation we will consider is that of Poisson inputs with the same mean rates and no correlations between them. From Section 11.1,

$$Q_{ij}(t)=\overline{r_{in}}^2+\delta(t)\overline{r_{in}}\delta_{ij},\tag{11.44}$$

and

$$K(t)=\frac{1}{\tau_m}e^{-t/\tau_m}.\tag{11.45}$$

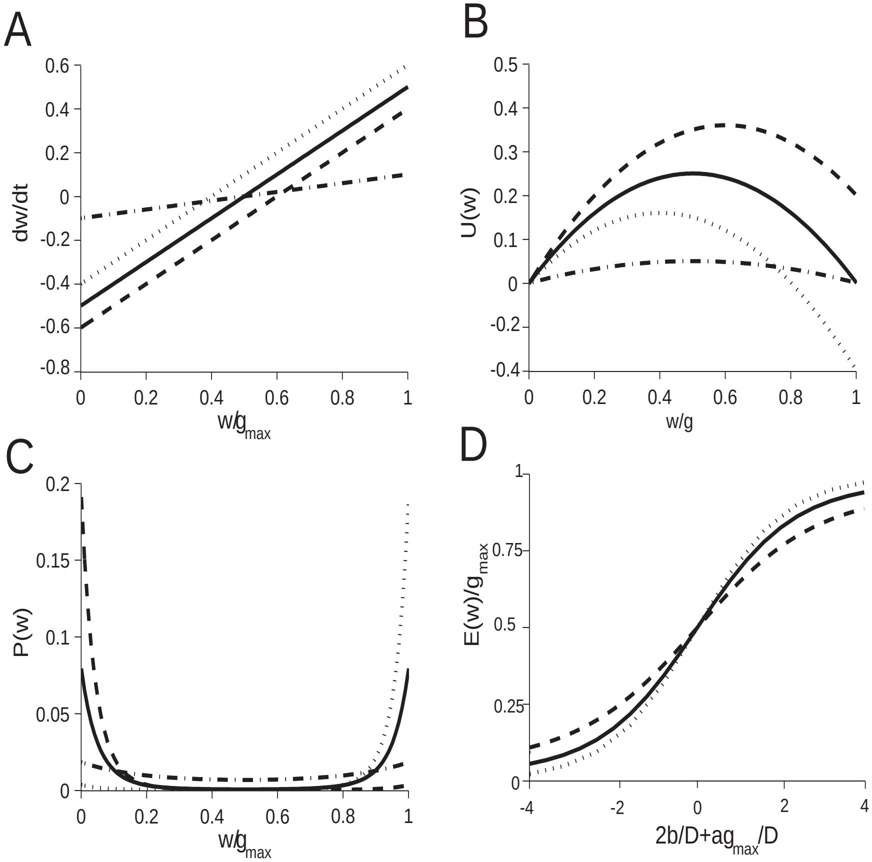

Figure 11.4

Calculation of synaptic weight distribution. (A) Mean rate of synaptic weight change for different values of a and b. Solid line a=1, b=-0.5. Dotted line a=1, b=-0.4. Dashed line a=1, b=-0.6. Dash dot line a=0.2, b=-0.5. (B) Same as A except for the potential U(w). (C) Same as A except for the normalization distribution P(w). (D) Expectation value of P(w) versus $2b/D + ag_{max}/D$. Solid line, a=1. Dashed line, a=2. Dashed line, a=0.5.

Upon collecting terms,

$$\frac{dw_i}{dt} = G\overline{r_{in}}w_iA_+\frac{\tau_+ + \tau_m}{\tau_+\tau_m} - G(A_- - A_+)\overline{r_{in}}^2\sum_j w_j + \theta\overline{r_{in}}(A_- - A_+), \tag{11.46}$$

or,

$$\frac{dw_i}{dt} = \overline{r_{in}}(Gw_iA_+\frac{\tau_+ + \tau_m}{\tau_+\tau_m} - (A_- - A_+)\overline{r_{out}}). \tag{11.47}$$

and

$$\begin{aligned} U(w) \approx -G(-(A_{-} - A_{+})\overline{r_{in}}^{2} + \overline{r_{in}}A_{+}\tfrac{\tau_{+}+\tau_{m}}{\tau_{+}\tau_{m}})w_i^2/D \\ -2(-G(A_{-} - A_{+})\overline{r_{in}}^{2}\textstyle\sum_{j\neq i} w_j + \theta\overline{r_{in}}(A_{-} - A_{+}))w_i/D. \end{aligned} \tag{11.48}$$

As we noted in the previous section, $U(0) \approx U(g_{max})$, so

$$g_{max}\overline{r_{in}}A_{+}\frac{\tau_{+}+\tau_{m}}{\tau_{+}\tau_{m}} \approx 2(A_{-} - A_{+})(\overline{r_{in}}^{2}\sum_{j} w_j - \theta\overline{r_{in}}), \tag{11.49}$$

or

$$g_{max}GA_{+}\frac{\tau_{+}+\tau_{m}}{\tau_{+}\tau_{m}} \approx 2(A_{-} - A_{+})\overline{r^{out}}, \tag{11.50}$$

In the steady state situation, the left-hand term is not affected by the postsynaptic rate, so the output rate is effectively kept constant. This effect was noted in Song and Abbott [79] and demonstrated in Figure 11.3. An intuitive explanation was given in Song et al. based on balances of excitation and inhibition. Neuron can operate in two different modes with distinct spike-train statistics and input-output correlations [1, 14, 85]. If the total amount of excitation overwhelms the amount of inhibition, the mean input to the neuron would bring it well above threshold if action potentials were blocked (Figure 11.5A). In this situation, the neuron operates in an input-averaging or regular-firing mode. The postsynaptic spike sequences produced in this mode are quite regular (Figure 11.5C) because the timing of the postsynaptic spikes is not sensitive to presynaptic spike times. There are roughly equal numbers of presynaptic action potentials before and after each postsynaptic spike [1, 14] (Figure 11.6A). Because the area under the STDP curve is slightly negative ($A_{-} - A_{+} > 0$), for a flat correlation curve,

$$\frac{dw_i(t)}{dt} = \int_{-\infty}^{\infty} dt' P(t') r^{out}(t) r_i^{in}(t+t') = K\int_{-\infty}^{\infty} dt' P(t'), \tag{11.51}$$

where K is a constant, is still negative. Thus the synapses are weakened.

As the excitatory synapses are weakened by STDP, the postsynaptic neuron enters a balanced mode of operation in which it generates a more irregular sequence of action potentials, and the timing of the postsynaptic spikes becomes more tightly correlated with the timing of the presynaptic spikes. The total synaptic input in the balanced mode is, on average, slightly below threshold (Figure 11.5B), so the postsynaptic neuron fires irregularly, primarily in response to statistical fluctuations in the total input (Figure 11.5D). Because action potentials occur preferentially after a random positive fluctuation, there tend to be more excitatory presynaptic spikes before than after a postsynaptic response [1, 14, 85] (Figure 11.6B). The small excess of presynaptic spikes just before a postsynaptic spike is described in the rate model as $\frac{\frac{w_i\overline{r_{in}}}{\tau_m}e^{t/\tau_m}}{\sum_j w_j\overline{r_{in}}^2}$ if $\quad t < 0$ It can be gathered in Figure 11.6 that the excess calculated from simulations is indeed well approximated by an exponential function. The STDP rule achieves a steady-state distribution of peak synaptic conductances when the excess of

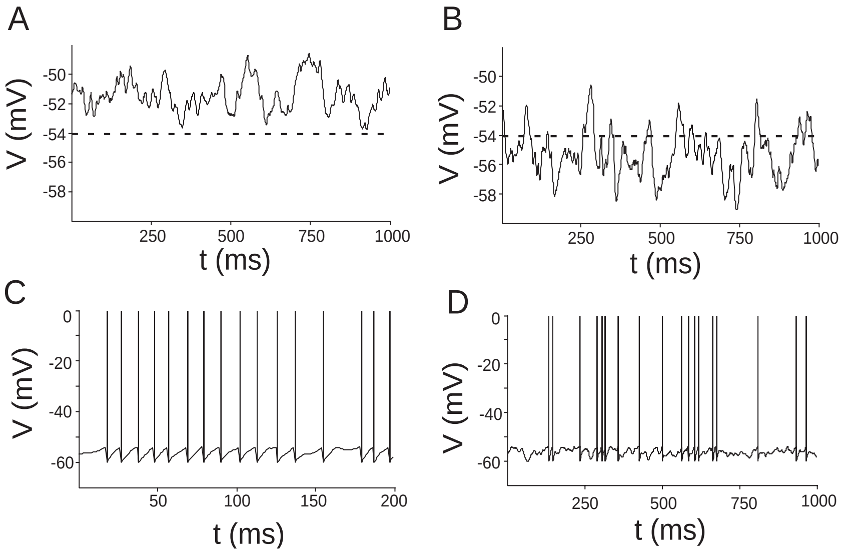

Figure 11.5

Regular and irregular firing modes of a model integrate-and-fire neuron. Upper panels show the model with action potentials deactivated, and the dashed lines show the action potential threshold. The lower figures show the model with action potentials activated. (A) In the regular firing mode, the average membrane potential without spikes is above threshold. (B) In the irregular firing mode, the average membrane potential without spikes is below threshold. (C) In the regular firing mode, the firing pattern is fast and regular (note the different time scale in the lower panel). (D) In the irregular firing mode, the firing pattern is slower and irregular. Adapted with permission from [1].

presynaptic action potentials prior to postsynaptic firing compensates for the asymmetry in the areas under the positive and negative portions of the STDP modification curve [1, 79] (Figure 11.6B). This condition is mathematically described in Equation 11.50.

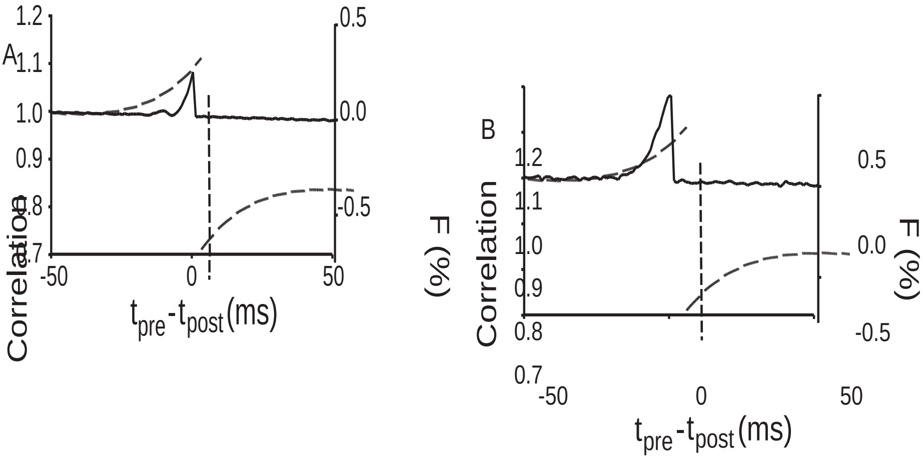

Figure 11.6

Correlation between pre- and postsynaptic action potentials before and after STDP. The solid curves indicate the relative probability of a presynaptic spike occurring at time t_{pre} when a postsynaptic spike occurs at time t_{post}. A correlation of one is the value due solely to chance occurrences of such pairs. The dashed curves show the STDP modification function from Figure 11.1. (A) Before STDP, the neuron is in the unbalanced mode with large excess excitatory drive. There is only a small excess of presynaptic spikes prior to a postsynaptic action potential. (B) After STDP, the neuron is in the balanced mode. There is a larger excess of presynaptic spikes prior to a postsynaptic action potential. Adapted with permission from [79].

11.4.5.2 Correlations with different time constants

We performed a simulation where 500 of the 1000 inputs were Poisson trains with a fixed rate and the other 500 inputs were correlated with each other with the correlation function

$$Q_{ij}(t) = C_{ij}\overline{r_{in}}^2 e^{-|t|/\tau_c} + \overline{r_{in}}^2 + \delta(t)\overline{r_{in}}\delta_{ij}, \tag{11.52}$$

and

$$C_{ij} = \begin{cases} C^2 \text{ if } \quad 0 < i < 500, 0 < j < 500 \\ 0 \quad \text{otherwise.} \end{cases} \tag{11.53}$$

Therefore,

$$\begin{aligned} &< r^{out}(t')r_i^{in}(t'+t) >_{t'} \\ &= -\theta\overline{r_{in}} + G \begin{cases} \sum_j w_j C_{ij}\overline{r_{in}}^2[\frac{\tau_c}{\tau_c-\tau_m}(e^{t/\tau_c} - e^{t/\tau_m}) + \frac{\tau_c}{\tau_c+\tau_m}e^{t/\tau_m}] & \\ +\sum_j w_j\overline{r_{in}}^2 + \frac{w_i\overline{r_{in}}}{\tau_m}e^{t/\tau_m} & \text{if} \quad t < 0 \\ \sum_j w_j C_{ij}\overline{r_{in}}^2 \frac{\tau_c}{\tau_c+\tau_m}e^{t/\tau_m} + \sum_j w_j\overline{r_{in}}^2 & \text{if} \quad t \geq 0 \end{cases} \end{aligned} \tag{11.54}$$

we plotted the average correlogram between the spike trains of the inputs and the postsynaptic neuron from the simulations for $\tau_c = 20ms$ in Figure 11.7A. We have also plotted in thin dotted line the mirror image of the portion of the correlation curve for $t_{pre} - t_{post} > 0$ for visual guidance. The correlograms display a 'spike' of excess presynaptic action potentials just before a postsynaptic spike, which reflects the excess of spikes needed to push the postsynaptic neuron above threshold and is represented by the $\frac{w_i}{\sum_j w_j \overline{r_{in}} \tau_m} e^{t/\tau_m}$(for t¡0) term in the rate model and is discussed in the previous section. We have also plotted in Figure 11.7B the correlogram after removing the $\frac{w_i}{\sum_j w_j \overline{r_{in}} \tau_m} e^{t/\tau_m}$(for t¡0) term in solid lines along with the predicted values from the rate model after appropriate normalization in dotted lines. The remaining excess represents excess of presynaptic spikes before a postsynaptic spike contributed by correlations in the inputs. If $\tau_+ = \tau_-$, the symmetric portion of the correlation leads to weakening if the total area under the STDP curve is negative as assumed in this chapter. The asymmetric portion shown as the difference between the solid line and the thin dotted line contributes to strengthening of synapses and has to cancel out the weakening resulting from the symmetric portion. If $\tau_- > \tau_+$, the symmetric portion can also result in synaptic strengthening. To calculate the final synaptic strengths, we could use approximation in Equation (11.43), and write

$$\begin{aligned} & w_{corr} = 2Sg_{max}(-G(H\overline{r_{in}}^2 C^2 - (A_- - A_+)\overline{r_{in}}^2 + \overline{r_{in}} A_+ \tfrac{\tau_+ + \tau_m}{\tau_+ \tau_m})w_i^2) \\ & -2(G499H\overline{r_{in}}^2 C^2 w_{corr} - G(A_- - A_+)\overline{r_{in}}^2 (499 w_{corr} + 500 w_{uncorr}) \\ & -\theta(A_- - A_+)\overline{r_{in}})) \\ & /((G\overline{r_{in}}^2 (499 w_{corr} + 500 w_{uncorr}) - \theta \overline{r_{in}})(A_-^2/\tau_- + A_+^2/\tau_-)) + 0.5 g_{max}, \end{aligned} \tag{11.55}$$

and

$$\begin{aligned} & w_{uncorr} = 2Sg_{max}(-G(-(A_- - A_+)\overline{r_{in}}^2 + \overline{r_{in}} A_+ \tfrac{\tau_+ + \tau_m}{\tau_+ \tau_m})w_i^2) \\ & -2(-G(A_- - A_+)\overline{r_{in}}^2 (499 w_{uncorr} + 500 w_{corr}) - \theta(A_- - A_+)\overline{r_{in}})) \\ & /((G\overline{r_{in}}^2 (499 w_{uncorr} + 500 w_{corr}) - \theta \overline{r_{in}})(A_-^2/\tau_- + A_+^2/\tau_-)) + 0.5 g_{max}, \end{aligned} \tag{11.56}$$

where w_{corr} is the average synaptic strength from the correlated cluster, w_{uncorr} is the synaptic strength from the uncorrelated cluster, and w is the synaptic weight the synapses are held at. Therefore, we can define R, where

$$R = \frac{(w_{corr} - w_{uncorr})\overline{r_{out}}}{-w_{corr}} = 499HSGC^2 \overline{r_{in}}/(A_-^2/\tau_- + A_+^2/\tau_-). \tag{11.57}$$

The quantity R is proportional to H, we therefore computed the ratio R for data obtained from the simulations for a range of τ_c and compared it with the prediction of H from the rate model. We use $C = 0.3$ in the simulations (see [79]), and $A_+ = 0.005$, $A_- = 1.05 * 0.005$, $\tau_+ = \tau_- = 20ms$. In Figure 11.7 C, we plot R calculated from data collected from the simulations in solid line, and H in dotted line after scaling according to the value at $\tau_c = 10$ms.

11.4.5.3 Gradient of correlations

Finally, let us reconsider the case briefly discussed in Section 11.3.3 where a neuron received inputs with a gradient correlations. The correlation between inputs i and j is given by $< r_i(t) r_j(t') >= \overline{r}_{in}^2 + \overline{r}_{in}^2 (\sigma^2 \delta_{ij} + (1 - \delta_{ij}) c_i c_j) e^{-|t-t'|/\tau_{corr}}$, with $\overline{r}_{in} = 10Hz$, $\sigma = 0.5$, $\tau_{corr} = 20ms$, and $c_i = a(i - 499.5)/999 + b$. We can solve the

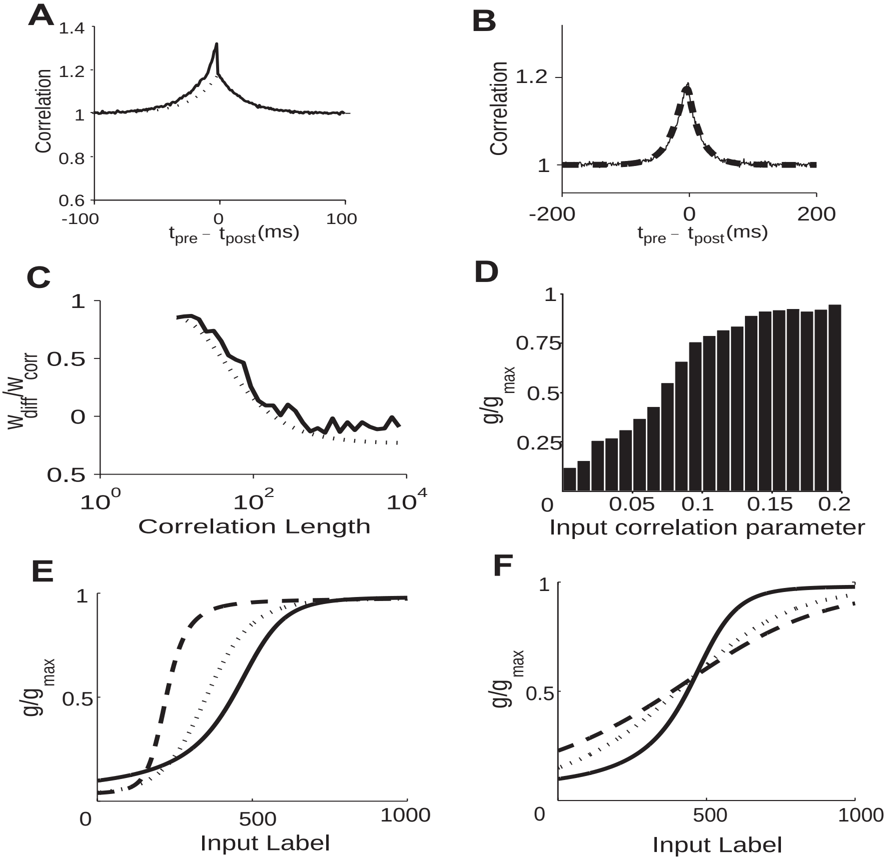

Figure 11.7

Correlations between input and output and final synaptic strengths are well predicted by the rate model. (A) Correlation between input and output for $\tau_c = 20ms$. Thin dotted line is the mirror image of the portion of the correlation curve for $t_{pre} - t_{post} > 0$. If $\tau_+ = \tau_-$, the symmetric portion of the correlation leads to weakening while the asymmetric portion contributes to strengthening of synapses. (B) Solid line same as A with correlation caused by spiking removed in solid line alone with prediction from rate model in dotted line. (C) Normalized difference in average equilibrium synaptic strength between correlated and uncorrelated groups (R) for different input correlation length. (D) Equilibrium synaptic strength for different input correlation parameters from simulation (E) Equilibrium synaptic strength for different input correlation gradients from rate model. a is 0.1 for all curves. Solid line b=0.1. Dotted line b=0.1. Dashed line b=0.2. (F) Same as E except for different a. b=0.1 for all curves. Solid line a=0.1. Dotted line a=0.75. Dashed line a=0.5.

set of equations for the expected values of *w* from the Fokker-Planck formulation numerically for self-consistency and the results are plotted in Figure 11.7E for different values of b and Figure 11.7F for different values of a. Notice that increasing a which is the slope of the gradient of correlations result in increased slope of the final synapse values with some saturation. While increasing the mean correlation of the all the inputs shift the curves to the left resulting in higher mean synaptic strengths. This shift might not be a desirable feature and can be removed by introducing a variable A_- to make STDP insensitive to mean correlation but rather the difference in correlations, as discussed in [43]. In Figure 11.7D, we have binned the resulting synaptic weights from simulations for inputs with graded correlations into 20 bins and plotted the mean synaptic weight for each bin. These values thus reflect the mean expected value of synaptic strength for synapses in that bin. It is very similar to the solid curve in Figure 11.7E and F which correspond to the same parameters. Although STDP is similar to subtractive normalization with a rate-based Hebbian rule, it produces a graded mean expected value of synaptic strengths rather than a bimodal one. This is because although each synapse still tends to the boundaries, the stochasticity allows the expected value for each synapse to be graded. So for synapses with similar driving force, some will be at the lower boundary and some will be at the upper boundary and the average of them will be close to the expected value of the probability distribution $P(w)$. Therefore STDP combines the desired features of competitiveness of subtractive normalization with sensitivity to correlation magnitude of multiplicative normalization in one rule.

11.5 Temporal aspects of STDP

Since STDP incorporated timing into the plasticity rule itself, it is natural to investigate its utility in learning temporal patterns. It has been suggested as a form of temporal difference learning to learn predictive coding by Rao and colleagues [67]. They have also used it to explain directional sensitivity in cortical cells. Similar ideas were used earlier by Abbott and colleagues as a basis for a model of place cells and spatial navigation in rats [11]. STDP reduces latency in the inputs. If a cell receiving inputs with different latencies, the inputs with shorter latency will tend to precede postsynaptic firing while the inputs with longer latency will tend to lag behind. STDP will lead to strengthening of the synapses of the inputs with shorter latency and weakening of the synapses of the inputs with longer latency. The final effect of this is a reduction in the response latency of the postsynaptic cell [79]. This was used as an explanation for the asymmetry expansion in place fields during training [54, 55]. Since STDP is very sensitive to synchrony in the inputs, when coupled with delay lines, it can be used to learn arbitrary temporal patterns by strengthening the appropriate delay lines so all the inputs arrive at the postsynaptic cell at the same time. This forms the basis of a model of tuning of delay lines in the barn owl auditory system [32]. It has also been used for sequence learning by other authors. [69]

11.6 STDP in a network

In this section, we will consider several network models using STDP. However, before we dive into the models, let us consider some general characteristics of STDP in the network setting. We have already noted in the previous section that STDP prefers inputs with shorter latencies, so if there are multiple pathways conveying similar information feeding into oa cell in the network, the pathway with the shortest latency will eventually dominate. Secondly, temporally asymmetric STDP discourages the formation of mutually excitatory loops. If neuron A is predictive of the firing of neuron B, neuron B cannot be predictive of the firing of neuron A and must lag behind. Therefore, if the synapses from neuron A to neuron B are strengthened, the synapses from neuron B and neuron A will be weakened, making a mutually excitatory configuration unstable. However, under some circumstances, a temporally asymmetric STDP could switch to a temporally symmetric STDP if the firing rates are sufficiently high, which could explain why many mutually excitatory loops nonetheless exist where temporally asymmetric STDP has been observed [78]. Experimental studies of STDP in a network have been carried out by Bi and Poo [7], highlighting the sensitive of patterns of firing in networks to the specific timing of the inputs. However, also apparent from these studies, studies of STDP in a network will especially difficult because of the intricate interplay of dynamics of network activity and the sensitivity to timing of STDP itself. The computational meaning of the network activities also needs to be investigated. It seems that a coherent computational framework is especially lacking here and will be a fruitful area of investigation.

11.6.1 Hebbian models of map development and plasticity

Since Hubel and Wiesel's pioneering studies of monocular deprivation [39], formation and alteration of columns and maps have been one of the favorite models for the study of activity-dependent changes in cortical circuits. It is widely recognized that activity is critical for refining synaptic connections to give adult patterns of connectivity and function (for reviews, see [17, 41, 100]). In development, spontaneous correlated activity is present from an early stage. Examples include waves of activity that propagate in the retina and LGN [27, 31, 56, 63, 65, 96, 99]. Later on, sensory inputs further refine the synaptic connections for cortical circuits [41, 81, 100]. Alterations of spontaneous and sensory induced activity can change both the degree of formation and the specific shape of cortical maps [39, 76, 94]. Mechanisms of synaptic modification include both changes in synaptic strengths and sprouting of new synapses. Changes in synaptic strengths have been linked to activity through mechanisms like LTP and LTD [2, 3, 9, 10, 50]. More recently, studies of spike-timing dependent plasticity have provided a more direct link between activity and modification of synaptic strength [5, 21, 23, 26, 36, 46, 49, 51, 78, 101]. On the other hand, the local release of neurotrophins and other molecules has been proposed to translate patterns of activity into patterns of synaptogenesis and neuronal survival and growth (reviewed in [12, 53, 72, 66, 83]). It is not clear at this stage whether activity is required for the initial development of cortical maps. Some studies suggest that the maps might be genetic and do not require either visual experience or even the retina [18, 19, 38]. However, endogenous correlated firing from other sources has not been systematically ruled out [99].

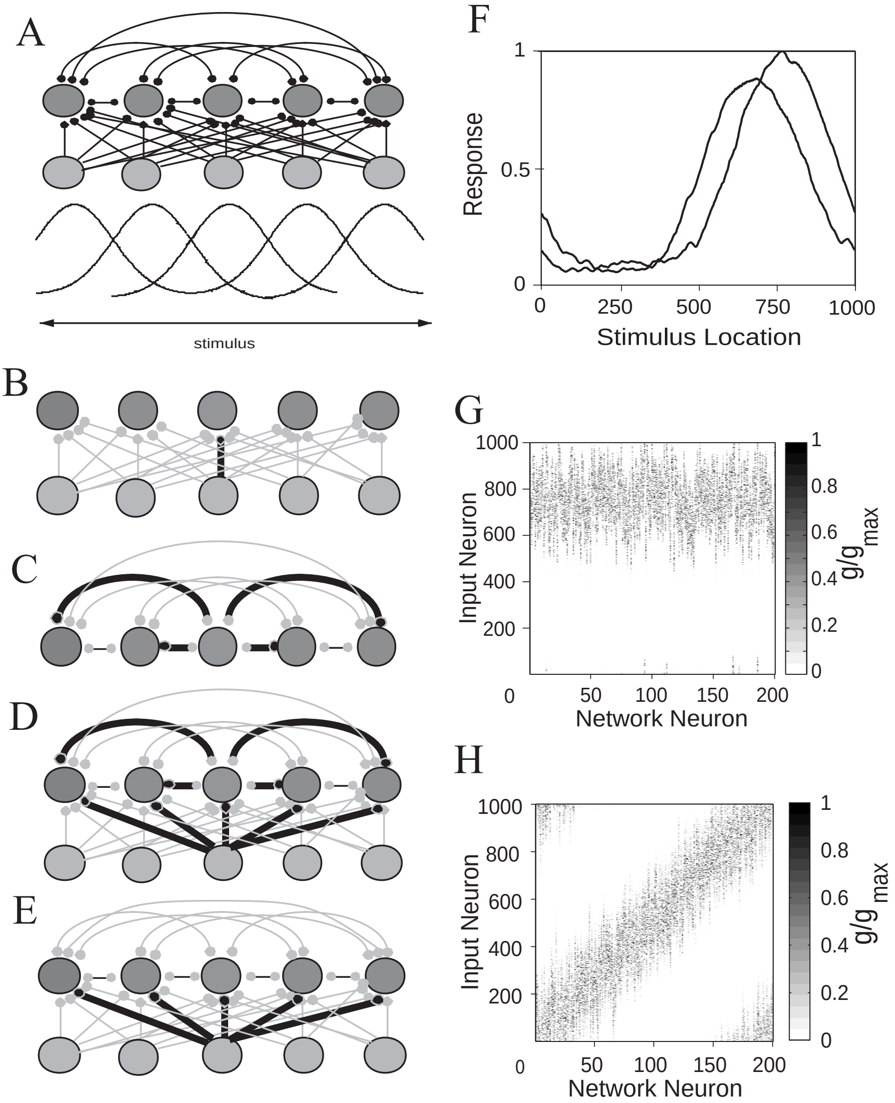

Figure 11.8
Continued.

Figure 11.8

Formation of columns and maps under STDP. A) Network Diagram. B) First stage of column formation. Seed placed in the feedforward connections creates a correlated group of network neurons. C) Second stage of column formation. Correlated group of network neurons send out connections to other neurons in the network. D) Third stage of column formation. Transfer of information from recurrent layer to feedforward layer. E) Last stage of column formation. Recurrent connections go away as feedforward connections become well formed. F) Receptive fields of two network neurons. G) Feedforward synaptic strengths define a column. The shade of each dot represents the strength of one synapse. The horizontal stripe indicates group of network neurons having similar input connections. H) Feedforward synaptic strengths define a map. This is formed when short range excitatory and global inhibitory recurrent connections are introduced in the network layer. The diagonal bar reflects the progression of receptive field centers as we move across the the sheet of network neurons. Adapted from [43, 80].

In the remainder of the section, we will describe a recent model of column and map formation and plasticity using STDP and compare it to earlier models of column and map formation using Hebbian rule based on rates [80]. In this model, an input and a network layer of neurons are simulated (Figure 11.8A). The inputs neurons generate Poisson spike trains in response to inputs presented. The network layer contained integrate-and-fire neurons. Every neuron in the input layer is randomly connected to one fifth of the neurons in the network layer and all the neurons in the network layer are recurrently connected. All of the synaptic connections were governed by STDP. Neurons in the input layer have a Gaussian receptive field so it will respond to inputs presented at a particular location best and respond less vigorously if the input is presented slightly away from the center of the receptive field and not respond at all far away from the receptive field center. Let us also label the inputs according to the receptive field center to simplify the discussion. We present stimulus at random locations for a period of time drawn from an exponential distribution with a mean on the order of the STDP time window and switch to another random location at the end of the period. This results in a moving Gaussian hill of activity on the input layer and introduces temporal correlations among the inputs that effectively drives STDP.

As mentioned previously, ocularly dominant cells can develop with a rate-based Hebbian rule if a global constraint is used [60, 61]. Under multiplicative constraints, there needs to be anti-correlation between inputs from different eyes, while lack of correlation will be sufficient under subtractive constraints [61]. Under STDP, the situation is similar to a rate-based Hebbian rule under subtractive constraints. If there are multiply sources of similar correlation among inputs from the same source but no correlation among inputs from different sources, under STDP, eventually only one source will drive the cell as STDP is a very competitive Hebbian rule. This competition will however take a much longer time than competition among sources with different strengths of correlations. In the context of this model, it allows neurons in the network layer to acquire receptive fields centered around only one location.

After STDP, those neurons are connected to only inputs from a small neighborhood in the input space. An example of such a receptive field is plotted in Figure 11.8F.

Without the recurrent connections, we would expect each neuron in the network layer to develop a receptive field but at a random location. When the recurrent connections are present, a single column of neurons all tuned to the same location is formed (Figure 11.8G). The sequence of events that leads to the formation of the column depends on the timing sensitive property of STDP. Let us examine it more closely. We have plotted the sequence of events leading to the formation of such a column in a diagram form in Figure 11.8B-E. Each circle in the graph represents a group of neurons and will be labelled from left to right 1 to 5 respectively.

For the ease of visualization, a seed has been added to the original network consisting of increased strengths for neurons from input group 3 to network group 3 (Figure 11.8B). This makes group 3 of network neurons become correlated with each other. This seed is not needed for the development of a single column. Without the seed, a group of correlated neurons in the network layer will automatically form and drive the remainder of the process. However, the neurons need not to be neighboring ones, which complicates the visualization. The process of symmetry breaking can also be quite time consuming. Once a correlated group is established, such a group of correlated network neurons will send out connections to other neurons in the network. This is perhaps easier to understand if we look at it from the perspective of a network neuron that is not part of the correlated group (Figure 11.8C). From that perspective, it is just the strengthening of the synapses corresponding to the most correlated inputs. This process differs from models based on rate-based Hebbian rules. In such models, connections between neurons that have high firing rates are strengthened resulting in reciprocal connectivity. But under asymmetric STDP, as discussed before, such reciprocal loops are discouraged and a layered structure is favored. At this stage of the development of a column, activity originates in the input layer, passes unto the correlated group of neurons, and then unto other neurons in the network layer. However, the firing of the input layer precedes that of the other neurons in the network layer. Therefore, the connections from the input layer unto the other neurons in the network layer would be strengthened (Figure 11.8D). This starts a process of transfer of information contained in the connectivity of the recurrent layer to the feedforward connections. Once the feedforward pathway reaches sufficient strength, the recurrent connections will weaken because the feedforward pathway has a shorter latency and therefore is the preferred pathway. Eventually, a column defined wholly by the feedforward connections will result (Figure 11.8E), completing the process of transfer. This can be viewed as a form of self-supervised learning enabled by the temporal asymmetry and predictive coding properties of STDP.

Next we would like to consider the formation of maps. Many models based on a rate-based Hebbian rule have been proposed to account for various aspects of map formation [44, 48, 77, 73, 75, 90, 95, 97]. In particular, we would like to consider a series of models by Miller and colleagues which are the most biologically sophisticated [57, 58, 59]. In these models, they were able to get formation of ocular dominance stripes by using an arbor function [60, 61]. Orientation selective columns can also develop and can be matched to the ocular dominance column map

[24, 25]. What seems important is excitatory connections to neighboring neurons and inhibitory connections to neurons that are further away. However, they always had fixed recurrent connections. Instead, all connections will be allowed to be plasticity at all time in the model with STDP. A map consisting of a collection of columns in orderly progression of receptive field center can be formed if the excitatory recurrent connections are restricted to a neighborhood and a global all to all inhibition is imposed on the network layer (Figure 11.8H). The columns have a tendency to spread to nearby neurons but the inhibitory connections force different neurons to fire at different times and adopt different receptive fields. These two tendencies balance each other, leading to an orderly map with nearby neurons having similar receptive fields while ones further away have different receptive fields.

Similar mechanisms can also explain adult plasticity in maps following lesions. Cortical circuits remain malleable in adulthood and can be altered by either deprivation of inputs or repeated presentation of specific stimulus schedules [15, 33, 40, 64, 91, 93]. For example, cutting a nerve innervating one finger will result in loss of activity in the part of cortex representing this finger initially. However, neurons having receptive fields in lesioned areas will over time adopt receptive field structures of neighboring neurons [15]. This transfer first happens on the recurrent neurons and the information eventually gets encoded in the feedforward connections.

Some of the predictions made by this model are in accord with recent experimental data. This model predicts that the development and plasticity of column and map structure on the network layer would precede the development and plasticity of such structures in the feedforward inputs. Some recent experimental evidence showed that receptive field structures appear in layer 2/3 first before they appear in layer 5 [84]. In adult plasticity experiments, cells first acquire new receptive fields with longer latency presumably going through the recurrent connections. The latency later drops, presumably reflecting the transfer to the feedforward connections [6, 34, 82]. More recently, experiments have been performed that indicate possible involvement of STDP in adult plasticity [71, 98]. Taken together, these make the general ideas outlined in this model a viable theory of activity dependent column and map formation. Other processes like axon and dendrite growth are undoubtedly also involved and the specific involvement of different processes will need to be investigated in future studies.

11.6.2 Distributed synchrony in a recurrent network

It has been noted before in the chapter that symmetric connections are not favored under STDP. Therefore, network models that assume symmetric synaptic connections like the Hopfield network are not favored under STDP. Instead, synfire chains seem to develop in recurrently connected networks with STDP.

Horn and colleagues provide an analysis of such a network [45]. They have noted that under some parameter regimes, the network operates in a distributed synchrony mode. The assembly of cells spontaneously breaks up into groups that fire synchronously. Then those groups take turns in firing in a cyclic manner. They speculated that such mechanisms might allow the learning of synfire chains.

11.7 Conclusion

Spike-timing-dependent plasticity (STDP) is a plasticity rule based on the timing of pre- and postsynaptic spikes. Recent experiments provide ample biological support for this plasticity rule. STDP gives detailed predictions for naturalistic inputs and makes it feasible for the first time to directly compare plasticity rules for naturalistic inputs with experimental data [29]. Therefore it is important to develop a theory to establish the fundamental properties of this plasticity rule, and the favorable interplay between theory and experiments will likely make STDP an important area of study. This chapter seeks to summarize recent results in these directions and place them in a coherent framework in comparison to Hebbian rules based on rates. STDP is a stable and competitive Hebbian rule by itself, with competition for the control of postsynaptic spike timing providing competition among inputs. STDP favors inputs with correlations among them as under the Hebbian rule. STDP can be viewed as striving to keep the postsynaptic firing rate roughly constant and achieves it by subtracting a factor from all synaptic weights. A Fokker-Planck formulation is introduced to predict the distribution of synaptic weights under STDP. STDP also has properties that make it attractive in learning temporal patterns. It favors synchronous inputs and this has been used in a model of the barn owl auditory system to tune delay lines [32]. In a network setting, STDP favors pathways with the shortest latency. The temporally asymmetric STDP disfavors reciprocal loops. Instead, it seems to be able to support the formation of synfire chains [45]. It has also been used in a recent model of column and map formation which makes use of the timing sensitivity of STDP to enable transfer of connectivity information from the recurrent connections to the feedforward connections. It is also unique in that very few additional conditions need to be imposed besides the plasticity rule itself. It is my suspicion that investigation of STDP in a network setting with naturalistic input spike trains will be a fertile ground of modelling for years to come.

However, there are still several issues that need to be addressed in future studies. First of all, synapses under STDP is bimodal in that synaptic strengths of individual synapses are saturated at the bounds. Some attempts to add a homeostatic rule to the synaptic strength apparently destroys the properties of competition and output rate normalization of STDP [43, 68, 89]. Experiments are needed to establish the distribution of synaptic strengths *in vivo* and potential mechanisms that keep synapses from saturating need to be discovered. Secondly, STDP in this chapter is modelled as if all spike pairs are independent and their effects on synaptic strengths are additive. Recent experiments have shown that this assumption to be incorrect [29, 78]. In particular, at low rates, spike pairs seem to sum sublinearly and at high rates potentiation is favored over depression. Consequences of those nonlinear summation effects need to be studied in future theoretical studies.

References

[1] Abbott, L. F. and Song, S. (1999), Temporally asymmetric Hebbian learning, spike timing and neuronal response variability, in *Advances in Neural Information Processing Systems 11*, Kearns, M. S., Solla, S. A., and Cohn, D. A. (eds.), 69–75, MIT Press.

[2] Bear, M. F. and Abraham, W. C. (1996), Long-term depression in hippocampus, *Annual Reviews in Neuroscience*, **19**: 437–462.

[3] Bear, M. F. and Malenka, R. C. (1994), Synaptic plasticity: LTP and LTD, *Current Opinion in Neurobiology*, **4**: 389–99.

[4] Bekkers, J. M. and Stevens, C. F. (1996), Cable properties of cultured hippocampal neurons determined from sucrose-evoked miniature EPSCs, *Journal of Neurophysiology*, **75**:1250–5.

[5] Bell, C. C., Han, V. Z., Sugawara, Y. and Grant, K. (1997), Synaptic plasticity in a cerebellum-like structure depends on temporal order, *Nature*, **387**: 278–81.

[6] Benuskova, L., Diamond, M. E., and Ebner, F. F. (1994), Dynamic synaptic modification threshold - computational model of experience-dependent plasticity in adult-rat barrel cortex, *Proceedings of the National Academy of Sciences, United States of America*, **91**: 4791–4795.

[7] Bi, G. Q., and Poo, M. M. (2001), Synaptic modification by correlated activity: Hebb's postulate revisited, *Annual Reviews in Neuroscience*, **24**: 139–166.

[8] Bienenstock, E. L., Cooper, L. N. and Munro, P. W. (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *Journal of Neuroscience*, **2**: 32–48.

[9] Bliss, T. V. P. and Lomo, T. (1973), Long-lasting potentiation of synaptic transmission in dentate area of anesthetized rabbit following stimulation of perforant path, *Journal of Physiology*, **232**: 331–356.

[10] Bliss, T. V. P., and Collingridge, G. L. (1993), A synaptic model of memory - long-term potentiation in the hippocampus, *Nature*, **361**: 31–39.

[11] Blum, K.I., and Abbott, L.F. (1996), A model of spatial map formation in the hippocampus of the rat., *Neural Computation*, **8**: 85-93.

[12] Bonhoeffer, T. (1996), Neurotrophins and activity-dependent development of the neocortex, *Current Opinion in Neurobiology*, **6**: 119–126.

[13] Brown, T. H., Kairiss, E. W. and Keenan, C. L. (1990), Hebbian synapses – biophysical mechanisms and algorithms, *Annual Reviews in Neuroscience*, **13**, 475–511.

[14] Bugmann, G., Christodoulou, C., and Taylor, J. G. (1997), Role of temporal

integration and fluctuation detection in the highly irregular firing of a leaky integrator neuron model with partial reset, *Neural Computation*, **9**: 985-1000.

[15] Buonomano, D. V., and Merzenich, M. M. (1998), Cortical plasticity: From synapses to maps, *Annual Reviews in Neuroscience*, **21**: 149–186.

[16] Cateau, H., and Fukai, T. (2003), A stochastic method to predict the consequence of arbitrary forms of spike-timing-dependent plasticity., *Neural Computation*, **15**: 597-620.

[17] Constantine-Paton, M., Cline, H. T., and Debski, E. (1990), Patterned activity, synaptic convergence, and the nmda receptor in developing visual pathways, *Annual Reviews in Neuroscience*, **13**: 129–154.

[18] Crair, M. C., Gillespie, D. C. and Stryker, M. P. (1998), The role of visual experience in the development of columns in cat visual cortex, *Science*, **279**: 566–570.

[19] Crowley, J. C., and Katz, L. C. (1999), Development of ocular dominance columns in the absence of retinal input, *Nature Neuroscience*, **2**: 1125–1130.

[20] Dayan, P., and Abbott, L. F. (2001), *Theoretical Neuroscience*, MIT Press: Cambridge.

[21] Debanne, D., Gahwiler, B. H., and Thompson, S. M. (1998), Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures, *Journal of Physiology*, **507 (Pt 1)**: 237–47.

[22] Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999), Plasticity in the intrinsic excitability of cortical pyramidal neurons, *Nature Neuroscience*, **2**: 515–520.

[23] Egger, V., Feldmeyer, D., and Sakmann, B. (1999), Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in vat barrel cortex, *Nature Neuroscience*, **2**: 1098–1105.

[24] Erwin, E., and Miller, K. D. (1996), The role of neural activity in rearranging connections in the central visual system., in *Computational Neuroscience: Trends in Research 1995*, Bower, J. M. (ed.), 179–184, Academic Press.

[25] Erwin, E., and Miller, K. D. (1998), Correlation-based development of ocularly matched orientation and ocular dominance maps: Determination of required input activities, *Journal of Neuroscience*, **18**: 9870–9895.

[26] Feldman, D. E. (2000), Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex, *Neuron*, **27**: 45–56.

[27] Feller, M. B. (1999), Spontaneous correlated activity in developing neural circuits, *Neuron*, **22**: 653–656.

[28] Fregnac, Y., and Shulz, D. E. (1999), Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning, *Journal of Neurobiology*, **41**: 69–82.

[29] Froemke, R. C., and Dan, Y. (2002), Spike-timing-dependent synaptic modification induced by natural spike trains, *Nature*, **416**: 433-8.

[30] Fu, Y., Djupsund, K., Gao, H., Hayden, B., Shen, K., and Dan, Y. (2002), Temporal specificity in the cortical plasticity of visual space representation, *Science*, **296**: 1999–2003.

[31] Galli, L., and Maffei, L. (1998), Spontaneous impulse activity of rat retinal ganglion-cells in prenatal life, *Science*, **242**: 90–91.

[32] Gerstner, W., Kempter, R., van Hemmen, J. L. and Wagner, H. (1996), A neuronal learning rule for sub-millisecond temporal coding, *Nature*, **383**: 76–81.

[33] Gilbert, C. D. (1996), Plasticity in visual perception and physiology, *Current Opinion in Neurobiology*, **6**: 269–274.

[34] Grajski, K., and Merzenich, M. (1990), Hebb-type dynamics is sufficient to account for the inverse magnification rule in cortical somatotopy, *Neural Computationation*, **2**: 71–84.

[35] Guillery, R. W. (1972), Binocular competition in the control of geniculate cell growth, *Journal of Comparative Neurology*, **144**, 117–29.

[36] Gustafsson, B., Wigstrom, H., Abraham, W. C., and Huang, Y. Y. (1987), Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials, *Journal of Neuroscience*, **7**: 774–80.

[37] Hebb, D. O. (1949), *The Organization of Behavior: A Neuropsychological Theory*, Wiley: New York.

[38] Horton, J. C., and Hocking, D. R. (1996), An adult-like pattern of ocular dominance columns in striate cortex of newborn monkeys prior to visual experience, *Journal of Neuroscience*, **16**: 1791–1807.

[39] Hubel, D. H., and Wiesel, T. N. (1965), Binocular interaction in striate cortex of kittens reared with artificial squint., *Journal of Neurophysiology*, **28**: 1041–1965.

[40] Kaas, J. H. (1991), Plasticity of sensory and motor maps in adult mammals, *Annual Reviews in Neuroscience*, **14**: 137–167.

[41] Katz, L. C., and Shatz, C. J. (1996), Synaptic activity and the construction of cortical circuits, *Science*, **274**: 1133–1138.

[42] Kempter, R., Gerstner, W., and van Hemmen, J. L. (2001), Intrinsic stabilization of output rates by spike-based Hebbian learning, *Neural Computation*, **13**: 2709–2741.

[43] Kepecs, A., van Rossum, M., Song, S., and Tegner, J. (2002), Spike-timing dependent Plasticity: Common themes and divergent vistas, *Biological Cybernetics*, **87**: 446-58.

[44] Kohonen, T. (1984), *Self-Organization and Associative Memory,* Springer-Verlag: Berlin.

[45] Levy, N., Horn, D., Meilijson, I., and Ruppin, E. (2001), Distributed synchrony in a cell assembly of spiking neurons, *Neural Networks*, 14: 815–824.

[46] Levy, W. B., and Steward, O. (1983), Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus, *Neuroscience,* **8**: 791–7.

[47] Leslie, K. R., Nelson, S. B., and Turrigiano, G. G. (2001), Postsynaptic depolarization scales quantal amplitude in cortical pyramidal neurons, *Journal of Neuroscience*, **21**: RC170.

[48] Linsker, R. (1986), From basic network principles to neural architecture: Emergence of spatial-opponent cells, *Proceedings of the National Academy of Sciences, United States of America,* **83**: 7508-7512.

[49] Magee, J. C., and Johnston, D. (1997), A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons, *Science*, **275**: 209–13.

[50] Malenka, R. C., and Nicoll, R. A. (1999), Neuroscience – Long-term potentiation – A decade of progress?, *Science*, **285**: 1870–1874.

[51] Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997), Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, *Science*, **275**:213–5.

[52] Marr, D. (1969), A theory of cerebellar cortex, *Journal of Physiology*, **202**: 437–470.

[53] McAllister, A. K., Katz, L. C., and Lo, D. C. (1999), Neurotrophins and synaptic plasticity, *Annual Reviews in Neuroscience*, **22**: 295–318.

[54] Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997), Experience-dependent, asymmetric expansion of hippocampal place fields, *Proceedings of the National Academy of Sciences, United States of America,* **94**, 8918–21.

[55] Mehta, M. R., Quirk, M. C., and Wilson, M. A. (2000), Experience-dependent asymmetric shape of hippocampal receptive fields, *Neuron*, **25**: 707–15.

[56] Meister, M., Wong, R. O. L., Baylor, D. A., and Shatz, C. J. (1991), Synchronous bursts of action-potentials in ganglion-cells of the developing mammalian retina, *Science*, **252**: 939-943.

[57] Miller, K. D. (1990), Correlation-based mechanisms of neural development, in *Neuroscience and Connectionist Theory,* Gluck, M. A. and Rumelhart, D. E. (eds.), 267–353, Lawrence Erlbaum Associates: Hillsdale NJ.

[58] Miller, K. (1996a), Receptive fields and maps in the visual cortex: models of ocular dominance and orientation columns, in *Models of Neural Networks III*, Domany, E., van Hemmen, J., and Schulten, K. (eds.) 55–78, Springer-Verlag: New York.

[59] Miller, K. D., Erwin, E., and Kayser, A. (1999), Is the development of orientation selectivity instructed by activity?, *Journal of Neurobiology*, **41**: 44–57.

[60] Miller, K. D., Keller, J. B., and Stryker, M. P. (1989), Ocular dominance column development – analysis and simulation, *Science,* **245**: 605-615.

[61] Miller, K. D., and MacKay, D. J. C. (1994), The role of constraints in Hebbian learning, *Neural Computation*, **6**: 100–126.

[62] Oja, E.(1992), Principal components, minor components, and linear neural networks, *Neural Networks,* **5**: 927–935.

[63] O'Donovan, M., Chub, N., and Wenner, P. (1998), Mechanisms of spontaneous activity in developing spinal networks, *Journal of Neurobiology,* **37**: 131–145.

[64] O'Leary, D., Ruff, N., and Dyck, R. (1994), Developmental, critical period plasticity, and adult reorganization of mammalian somatosensory systems, *Current Opinions in Neurobiology,* **4**: 535–44.

[65] Penn, A. A., Riquelme, P. A., Feller, M. B., and Shatz, C. J. (1998), Competition in retinogeniculate patterning driven by spontaneous activity, *Science,* **279**: 2108–2112.

[66] Poo, M. M. (2001), Neurotrophins as synaptic modulators, *Nature Rev. Neuroscience,* **2**: 24–32.

[67] Rao, R. P. N., and Sejnowski, T. J. (2001), Spike-timing-dependent Hebbian plasticity as temporal difference learning, *Neural Computation*, **13**: 2221–2237.

[68] Rubin, J., Lee, D. D., and Sompolinsky, H. (2001), Equilibrium properties of temporally asymmetric Hebbian plasticity, *Physical Review Letters,* **86**: 364–367.

[69] Ruf, B., and Schmitt, M. (1997), Unsupervised learning in networks of spiking neurons using temporal coding, in *Proc. 7th Int. Conf. Artificial Neural Networks (ICANN'97),* Gerstner W., Germond A., Hasler, M., and Nicoud, J. D. (eds.), 361–366, Springer-Verlag: Heidelberg.

[70] Sanger, T. D. (1989), Optimal unsupervised learning in a single-layer linear feedforward neural network, *Neural Networks,* **2**: 459–473.

[71] Schuett, S., Bonhoeffer, T., and Hubener, M. (2001), Pairing-induced changes of orientation maps in cat visual cortex, *Neuron,* **32**: 325–337.

[72] Schuman, E. M.(1999), Neurotrophin regulation of synaptic transmission, *Current Opinion in Neurobiology,* **9**: 105–109

[73] Sejnowski, T. J. (1977), Storing covariance with nonlinearly interacting neurons, *Journal of Mathematical Biology,* **4**: 303–321.

[74] Sejnowski, T. J. (1999), The book of Hebb, *Neuron,* **24**: 773–776.

[75] Sejnowski, T. J., and Tesauro, G. (1989), The Hebb rule for synaptic plasticity:

Algorithms and implementations, in *Neural models of plasticity: Experimental and theoretical approaches* Byrne, J. H. and Berry, W. O. (eds.), 94–103, Academic Press: San Diego.

[76] Sharma, J., Angelucci, A., and Sur, M. (2000), Induction of visual orientation modules in auditory cortex, *Nature*, **404**: 841–847.

[77] Shouval, H. Z., and Perrone, M. P. (1995), Post-Hebbian learning rules, in *The Handbook of Brain Theory and Neural Networks*, Arbib, M. A. (ed.), 745–748, MIT Press: Cambridge, MA.

[78] Sjostrom, P. J., Turrigiano, G. G., and Nelson, S. B. (2001), Rate, timing, and cooperativity jointly determine cortical synaptic plasticity, *Neuron*, **32**: 1149–1164.

[79] Song, S., Miller, K. D., and Abbott, L. F. (2000), Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, *Nature Neuroscience*, **3**: 919–926.

[80] Song, S., and Abbott, L. F. (2001), Cortical development and remapping through spike timing-dependent plasticity, *Neuron*, **32**: 339-50.

[81] Sur, M., and Leamey, C. A.(2001), Development and plasticity of cortical areas and networks, *Nature Review Neuroscience*, **2**: 251–262.

[82] Sutton, G. G., Reggia, J. A., and Armentrout, S. L. and Dautrechy, C. L. (1994), Cortical map reorganization as a competitive process, *Neural Computation*, **6**: 1–13.

[83] Thoenen, H. (1995), Neurotrophins and neuronal plasticity, *Science*, **270**: 593–598.

[84] Trachtenberg, J. T., Trepel, C., and Stryker, M. P. (2000), Rapid extragranular plasticity in the absence of thalamocortical plasticity in the developing primary visual cortex, *Science*, **287**: 2029–2032.

[85] Troyer, T. W., and Miller, K. D. (1997), Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell, *Neural Computation*, **9**: 971–983.

[86] Turrigiano, G. G. (1999), Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same, *Trends in Neuroscience*, **22**: 221–227.

[87] Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998), Activity-dependent scaling of quantal amplitude in neocortical neurons, *Nature*, **391**: 892–896.

[88] Turrigiano, G. G., and Nelson, S. B.(2000), Hebb and homeostasis in neuronal plasticity, *Current Opinion in Neurobiology*, **10**: 358–364.

[89] van Rossum, M. C. W., Bi, G. Q., and Turrigiano, G. G. (2000), Stable Hebbian learning from spike timing-dependent plasticity, *Journal of Neuroscience*, **20**:

8812–8821.

[90] von der Malsburg, C. (1973), Self-organization of orientation sensitive cells in the striate cortex, *Kybernetik*, **14**: 85–100.

[91] Wall, J. T. (1988), Variable organization in cortical maps of the skin as an indication of the lifelong adaptive capacities of circuits in the mammalian brain, *Trends in Neuroscience*, **11**: 549–557.

[92] Watt, A. J., van Rossum, M. C. W., MacLeod, K. M., Nelson, S. B., and Turrigiano, G. G.(2000), Activity coregulates quantal AMPA and NMDA currents at neocortical synapses, *Neuron*, **26**: 659–670.

[93] Weinberger, N. (1995), Dynamic regulation of receptive fields and maps in the adult sensory cortex, *Annual Reviews of Neuroscience*, **19**: 129–58.

[94] Weliky, M., and Katz, L. C. (1997), Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity, *Nature*, **386**: 680–685.

[95] Wimbauer, S., Gerstner, W., and Van Hemmen, J. L. (1994), Emergence of spatiotemporal receptive-fields and its application to motion detection, *Biological Cybernetics*, **72**: 81–92.

[96] Wong, R. O. L., Chernjavsky, A., Smith, S. J., and Shatz, C. J.(1995), Early functional neural networks in the developing retina, *Nature*, **374**: 716–718.

[97] Wiskott, L., and Sejnowski, T. (1998), Constrained optimization for neural map formation: A unifying framework for weight growth and normalization, *Neural Computation*, **10**: 671–716.

[98] Yao, H. S., and Dan, Y. (2001), Stimulus timing-dependent plasticity in cortical processing of orientation, *Neuron*, **32**: 315–323.

[99] Yuste, R., Peinado, A., and Katz, L. C. (1992), Neuronal domains in developing neocortex, *Science*, **257**: 666–669.

[100] Zhang, L. I., and Poo, M. M.(2001), Electrical activity and development of neural circuits, *Nature Neuroscience*, **4**: 1207–1214.

[101] Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M. (1998) A critical window for cooperation and competition among developing retinotectal synapses, *Nature*, **395**:37–44.

Chapter 12

Correlated Neuronal Activity: High- and Low-Level Views

Emilio Salinas[1], and Terrence J. Sejnowski[2,3]

[1]*Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1010, U.S.,* [2]*Computational Neurobiology Laboratory, Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, U.S.,* [3]*Department of Biology, University of California at San Diego, La Jolla, CA 92093, U.S.*

CONTENTS

12.1 Introduction: the timing game

Correlated firing is a common expression used in Neuroscience. It refers to two or more neurons that tend to be activated at the same time. It is used so frequently in part because there are so many timescales at which one may analyze neural activity. In a sense, correlation might appear as a trivial phenomenon. For instance, if one looks at day-long activity, practically the whole cerebral cortex fires in a correlated manner, because of the sleep-wake cycle. Similarly, whenever an object appears within the visual field, many neurons in visual cortex are expected to respond throughout the same time interval. Clearly, such correlations are to be expected. However, as the observation time window becomes smaller, explaining the presence of correlations becomes more difficult and, at the same time, potentially much more useful. Suppose the activity of two visual neurons is monitored during presentation of a visual stimulus, after its onset. Suppose also that within a short time window of, say, a few hundred milliseconds, spikes from the two neurons tend to appear at the same time. Why is this? Neither the sensory information nor the state of the subject are changing in an appreciable way, so the correlation must reflect something about the internal dynamics of the local circuitry or its connectivity. This is where correlations become interesting.

Thus, correlations at relatively short timescales become useful probes for understanding what neural circuits do, and how they do it. This is what this chapter is about. This analysis goes down to the one millisecond limit (or even further), where correlation changes name and becomes synchrony. Even at this point, the significance of correlated activity cannot be taken for granted. Some amount of synchrony is practically always to be expected simply because cortical neurons are highly interconnected [14, 91]. The question is not just whether there is any correlated activity at all, but whether timing is an issue and correlations make any difference. In other words, given the function of a particular microcircuit or cortical area, if the system were able to control the level and timescale of correlated activity, what would the optimal values be? For example, in a primary sensory area, stimulus representation is of paramount importance, so maybe measuring an excess of coincident spikes in this case is not an accident, but a consequence of the *algorithm* that local circuits use to encode stimulus features. This is just an example; the broader question is whether neurons exploit the precise coincidence of spikes for specific functions. There are several theoretical proposals that revolve around this concept; we discuss some of them below.

Asking about the functional implications of correlated activity is one way to attack the problem; this is a top-down approach. Another alternative is to take a bottom-up view and investigate the biophysical processes related to correlated firing. These come in two flavors, mechanisms by which correlations are generated, and mechanisms by which a postsynaptic neuron is sensitive to correlated input. In this case valuable information can be obtained about possible correlation patterns and timescales, and in general about the dynamics of correlated activity.

This approach is also important because it sheds some light on a fundamental question: how does a single cortical neuron respond under realistic stimulation conditions? The reason this is a problem is the interaction between spike-generating mechanisms, which are inherently nonlinear, and the input that drives the neuron, which typically has a complicated temporal structure. The major obstacle is not the accuracy of the single-neuron description; in fact, classic conductance-based models [24] like the Hodgkin-Huxley model [48] are, if anything, too detailed. The larger problem is the complexity of the total driving input, which is mediated by thousands of synaptic contacts [14, 91]. What attributes of this input will the postsynaptic response be most sensitive to? Correlations between synaptic inputs are crucial here because they shape the total input and hence the postsynaptic response. Determining what exactly is their role is a key requisite for understanding how neurons interact dynamically and how the timing of their responses could be used for computational purposes.

This chapter reviews work related to both perspectives on correlated activity: the high-level approach at which function serves to guide the analysis, and the low-level approach that is bound to the biophysics of single neurons. Eventually (and ideally), the two should merge, but currently the gap between them is large. Nevertheless, comparing results side by side provides an interesting panorama that may suggest further clues as to how neurons and neural circuits perform their functions.

12.2 Functional roles for spike timing

There is little doubt that the correct timing of action potentials is critical for many functions in the central nervous system. The detection of inter-aural time differences in owls and the electrosensory capabilities of electric fish are two well-known examples [21]. In these cases it is not surprising that timing is important; it is ingrained in the nature of the sensory signals being detected. The issue of timing also arises naturally in the rodent somatosensory system [7]. To explore their surroundings, rats move their whiskers periodically. To locate an object, whisker deflections need to be interpreted relative to whisker position, which can be determined from the phase of the motor signal. Thus, the latencies of stimulus-evoked responses relative to such internal signal can be used to encode spatial information. This mechanism by which sensory-triggered activity is interpreted relative to an internal, reference signal may be applicable to other circuits and in a more general way [6, 57].

12.2.1 Stimulus representation

Spike timing, however, has been discussed in an even wider sense than implied by the above examples. No doubt, this is partly because oscillations at various frequencies and synchronous activity are so widespread [8–86]. One proposal that has received considerable attention is that the coordinated timing of action potentials may be exploited for stimulus representation [71–44]. Specifically, neurons that have different selectivities but fire synchronously may refer to the same object or concept, binding its features. The following experiment [51] illustrates this point. The receptive fields of two visual neurons were stimulated in two ways, by presenting a single object, and by presenting two objects. Care was taken so that in the two conditions practically the same firing rates were evoked. The synchrony between pairs of neurons varied across conditions, even when the firing rates did not. Thus, correlations seemed to code whether one or two stimuli were shown [51].

In general, changes in firing rate pose a problem when interpreting variations in synchrony or correlations, first, because the latter can be caused by the former, and second, because the impact of a change in correlation upon downstream neurons becomes uncertain given a simultaneous change in firing rate. When neural activity is compared in two conditions involving different stimuli, it is likely that the evoked firing rates from the recorded neurons will change; even the populations that respond within a given area may be different. This is one of the main factors that muddles the interpretation of experiments in which correlations have been measured [72]. The most solid paradigms for investigating correlated activity are those in which variations in correlation are observed without variations in stimulation and without parallel changes in firing rate, but fulfilling all of these conditions requires clever experimental design and analysis.

There are many other studies in which correlations have been interpreted as additional coding dimensions for building internal representations. The following are cases in which the confounding factors just mentioned were minimized. Consider two neurons with overlapping receptive fields, and hence a considerable degree of synchrony. Analysis of the activity of such visual neurons in the lateral geniculate nucleus has shown [23] that significantly more information about the stimulus (around 20% more) can be extracted from their spike trains if the synchronous spikes are analyzed separately from the nonsynchronous ones. In a similar vein, recordings from primary auditory cortex indicate that, when a stimulus is turned on, neurons respond by changing their firing rates and their correlations [26]. In many cases the firing rate modulations are transient, so they may disappear if the sound is sustained. However, the evoked changes in correlation may persist [26]. Thus, the correlation structure can signal the presence of a stimulus in the absence of changes in firing rate.

Finally, the antennal lobe of insects is an interesting preparation in which this problem can be investigated. Spikes in this structure are typically synchronized by 20 Hz oscillations [90]. When these neurons are artificially desynchronized [55], the specificity of downstream responses is strongly degraded, selectivity for different odors decreases, and responses to new odors arise, even though this loss of information does not occur upstream. Apparently, what happens is that the downstream cells — Kenyon cells in the mushroom bodies — act as coincidence detectors that detect synchronized spikes from projection neurons in the antennal lobe. Kenyon cells have very low firing rates and are highly selective for odors, so in effect they sparsify the output of the antennal lobe [62]. In addition, disrupting synchrony in this system has a real impact on behavior: it impairs odor discrimination [79]. This preparation is also convenient for studying the biophysical mechanisms underlying such oscillatory processes [9, 10].

These examples show that the neural codes used to represent the physical world can be made more efficient by taking into account the pairwise interactions between neural responses. The degree to which this is actually a general strategy used by neurons is uncertain; the key observation is that, under this point of view, correlations are stimulus-dependent, just like sensory-evoked firing rates. The studies discussed below suggest a different alternative in which correlations change rapidly as functions of internal events and may regulate the flow of neural information, rather than its meaning [68].

12.2.2 Information flow

The regulation of information flow is illustrated by the following result [70]. When intracortical microstimulation is applied during performance of a visiual-motion discrimination task, the subject's response is artificially biased, but the bias depends strongly on the time at which the microinjected current is delivered relative to stimulus onset. Microstimulation has a robust effect if applied during presentation of the visual stimulus, but it has no effect if applied slightly earlier or slightly later than the natural stimulus [70]. This suggests that even a simple task is executed according to an internal schedule, such that the information provided by sensory neurons is effectively transmitted only during a certain time window. How does this internal schedule work? One possibility is that changes in correlations are involved [68]. This is suggested by a number of recent experiments in which correlations were seen to vary independently of stimulation conditions. To work around the usual problems with stimulus-linked correlations, investigators have studied correlated activity in paradigms where, across trials, stimulation conditions remain essentially constant and the most significant changes occur in the internal state of a subject.

Riehle and colleagues trained monkeys to perform a simple delayed-response task where two cues were presented sequentially [66]. The first cue indicated a target position and instructed the animal to get ready, while the second cue gave the go signal for the requested hand movement. Crucially, the go signal could appear 600, 900, 1200 or 1500 ms after the first cue, and this varied randomly from trial to trial. Neurons recorded in primary motor cortex increased their synchrony around the time of the actual sensory stimulus or around the time when the animal expected the go signal but it did not appear [66]. The latter case is the most striking, because there the firing rates did not change and neither did the stimulus; the synchronization depended exclusively on the internal state of the monkey.

Fries and colleagues [35] used attention rather than expectation to investigate the synchrony of visual neurons in area V4. They used conditions under which firing rates varied minimally, taking advantage of the finding that, although attention may have a strong effect on the firing rates evoked by visual stimuli, this modulation is minimized at high contrast [65]. Monkeys were trained to fixate on a central spot and to attend to either of two stimuli presented simultaneously and at the same eccentricity. One of the stimuli fell inside the receptive field of a neuron whose activity was recorded. Thus the responses to the same stimulus could be compared in two conditions, with attention inside or outside the neuron's receptive field. At the same time, the local field potential (LFP) was recorded from a nearby electrode. The LFP measures the electric field caused by transmembrane currents flowing near the electrode, so it gives an indication of local average activity [38]. The correlation that was studied in these experiments [35] was that between the LFP and the recorded neuron's spikes. The key quantity here is the spike-triggered average of the LFP, or STA. The STA is obtained by adding, for each spike recorded, a segment of the LFP centered on the time of the spike; the final sum is then divided by the total number of spikes. The result is the average LFP waveform that is observed around the time of a spike. STAs were computed for attention outside and inside the receptive field. They were similar, but not identical: rapid fluctuations were more pronounced when attention was directed inside the receptive field; in the Fourier decomposition, power in the low frequency band (0–17 Hz) decreased while power in the high frequency band (30–70 Hz) increased. Because the STA reflects the correlation between one neuron and the neighboring population, the interpretation is that, as attention shifts to the receptive fields of a cluster of neurons, these become more synchronized at high frequencies and less so at low frequencies. Although the changes in synchrony were modest — on average, low-frequency synchronization decreased by 23% and high-frequency synchronization increased by 19% — changes in firing rate were also small; these were enhanced by a median of 16% with attention inside the receptive field. Under these conditions the changes in synchrony could be significant in terms of their impact on the responses of downstream neurons.

The study just discussed [35] suggests that synchrony specifically in the gamma band (roughly 30–80 Hz) may enhance the processing of information in some way. But what exactly is the impact of such synchronization? Another recent study [34] suggests at least one measurable consequence: the latencies of synchronized neurons responding to a stimulus may shift in unison. In this case the paradigm was very simple: oriented bars of light were flashed and the responses of two or more neurons in primary visual cortex (V1) were recorded, along with LFPs. Neurons were activated by the stimuli, and the key quantity examined was the time that it took the neurons to respond — the latency — which was calculated on each trial. Latencies covaried fairly strongly from trial to trial (mean correlation coefficient of 0.34, with a range from 0.18 to 0.55), so pairs of neurons tended to fire early or late together. This tendency depended on the amount of gamma power in the LFPs right before the stimulus. When the LFPs from two electrodes both had a strong gamma component, the latency covariation between the two recorded neurons from the same pair of electrodes was high. Note that the spectral composition of the LFPs was only weakly related to changes in firing rate, so short latencies were probably not due to changes in excitability. This means that, if neurons get synchronized around 40 Hz right before a stimulus is presented, they will respond at about the same time [34]. In other words, while the mean firing rates are mostly insensitive to shifts in oscillation frequencies, the time spread in the evoked spikes from multiple neurons is much smaller when the gamma oscillations are enhanced. This could certainly have an impact on a downstream population driven by these neurons [18–67]. Thus, the modulation of latency covariations [34] is a concrete example of how the synchrony of a local circuit may be used to control the strength of a neural signal.

Finally, we want to mention two other studies [36, 37] that also investigated the synchronization of V1 neurons, this time using an interocular rivalry paradigm. In rivalry experiments, different images are shown to the two eyes but only one image is perceived at any given moment [52]. The perception flips from one image to the other randomly, with a characteristic timescale that depends on the experimental setup. The studies in question [36, 37] were done in awake strabismic cats, a preparation with two advantages: V1 neurons are dominated by a single eye, so their firing rates essentially depend on what their dominant eye sees regardless of the other one, and it is relatively easy to know which of the two images is perceived (at equal contrasts for the two images, one eye always suppresses the other, and this can be measured by tracking the cat's eye movements in response to conflicting moving stimuli). The two conditions compared were: a single image presented to the eye driving the recorded neurons, or the same stimulus shown to the driving eye plus a conflicting image presented to the other eye. The firing rates in these two conditions should be the same, because strabismus makes most neurons monocular; indeed, the rates did not change very much across conditions and did not depend on which image was perceived. However, synchrony within the 40 Hz band did change across conditions [36, 37]. When neurons were driven by the eye providing the percept, synchrony was much stronger in the rivalrous condition than in the monocular one. In contrast, when neurons were driven by the eye whose image became suppressed, synchrony was much lower in the rivalrous condition than in the monocular one. In other words, when conflicting images were presented, neurons responding to the image being perceived were always more synchronized. In this case, stronger synchronization in the high frequency band (30–70 Hz) is suggested to be a neural correlate of stimulus selection [36, 37].

In summary, it is possible that correlations between neurons can be controlled independently of firing rate. Two ideas that have been put forth are: that this may serve to generate more efficient neural codes [71, 44], which follows from theoretical arguments and experiments in which correlations vary in a stimulus-dependent way, or to regulate the flow of information [68], which follows from experiments in which correlations have been linked to expectation, attention, sensory latencies and rivalry — all processes that regulate the strength but not the content of sensory-derived neural signals. Other alternatives may become apparent in the future.

Next we discuss some common types of correlated activity patterns. In part, the goal is to describe them mathematically, at least to a first-order approximation.

12.3 Correlations arising from common input

As mentioned above, oscillations and synchronous responses are commonly observed throughout the nervous system [8–86]. This is not particularly surprising; in fact, correlations are to be expected simply because neurons in the brain are extensively interconnected [14, 91]. Now we will discuss two major mechanisms that give rise to correlated activity, common input and recurent connectivity. The distinction between them is somewhat artificial, but it is useful in portraying the range of correlation patterns that may arise. Although they will not be included, it should be kept in mind that intrinsic oscillatory properties of neurons are also important in determining global rhythmic activity [53–54].

An important analytical tool used to study the joint activity of neurons is the cross-correlation histogram or cross-correlogram [63–15], which is constructed from pairs of spike trains. This function shows the probability (or some quantity proportional to it) that neuron B fires a spike t milliseconds before or after a spike from neuron A, where t is called the time shift or time lag. When the two spike trains are independent, the cross-correlogram is flat; when they covary in some way, one or more peaks appear [15]. A peak at zero time shift means that the two neurons tend to fire at the same time more often than expected just by chance. Interpreting a cross-correlogram constructed from experimental data can be quite difficult because any covariation during data collection will show up as a peak [15]. Two neurons, for example, may respond at the same time to changes in stimulation conditions even if they are independent; this will produce a peak that has nothing to do with the functional connectivity of the circuit, which is what one is usually interested in. Another problem with this technique is that it requires large amounts of data. These disadvantages, however, have much lesser importance with simulated spike trains because they can be very long and their statistics can be constant.

Figure 12.1 shows synthetic, computer-generated spike trains from neurons that share some of their driving inputs but are otherwise disconnected. Responses from 20 neurons are displayed in each panel. Continuous traces superimposed on the spike rasters show the mean spike density or instantaneous firing rate, averaged over all neurons; this quantity is proportional to the probability of observing a spike from any of the neurons at any given time. Cross-correlograms are shown below. As mentioned above, the y-axis indicates the probability of observing a pair of spikes separated in time by the amount on the x-axis. The normalization is such that the probability expected by chance is equal to 1. The spikes shown were produced by integrate-and-fire model neurons [24, 67, 82], each driven by two time-varying signals, $g_E(t)$ and $g_I(t)$, representing the total excitatory and inhibitory conductances generated by large numbers of synaptic inputs. Details of the model are given in the Appendix. To generate synchronous activity between postsynaptic responses, the conductances $g_E(t)$ and $g_I(t)$ were correlated across neurons. This is exactly what would happen if pairs of postsynaptic neurons shared some fraction of all presynaptic spike trains driving them. In Figure 12.1a the mean correlation between conductances was 0.2. This means that, for any pair of neurons i and k, the correlation coefficient between excitatory conductances,

$$\rho_E^{ik} = \frac{\left\langle \left(g_E^i - \left\langle g_E^i \right\rangle\right)\left(g_E^k - \left\langle g_E^k \right\rangle\right)\right\rangle}{\sqrt{\left\langle \left(g_E^i - \left\langle g_E^i \right\rangle\right)^2 \right\rangle}\sqrt{\left\langle \left(g_E^k - \left\langle g_E^k \right\rangle\right)^2 \right\rangle}}, \tag{12.1}$$

was approximately 0.2. In this expression the angle brackets <> indicate an average over time, and neurons are indexed by a superscript. Inhibitory conductances also had a correlation of 0.2 across neurons, but all excitatory and inhibitory conductances were independent of each other. The nonzero correlation between conduc-

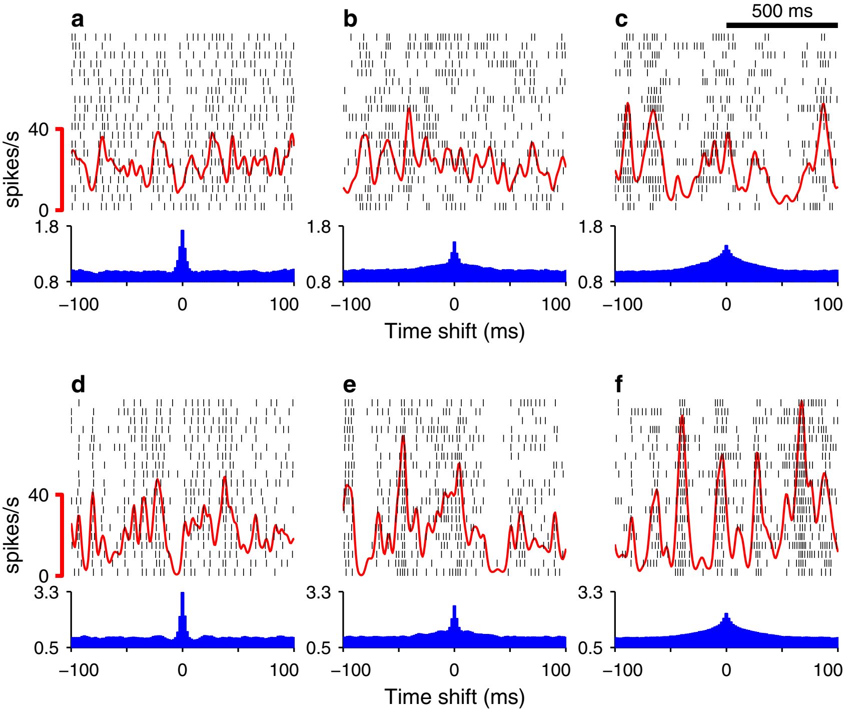

Figure 12.1

Spike trains correlated by common input. Each panel includes 20 computer-generated spike trains. Each row represents one neuron and each small, vertical line one spike. Neurons were modeled as leaky integrate-and-fire units disconnected from each other but driven by synaptic conductances that co-fluctuated across neurons. Continuous traces superimposed on the rasters are firing rates, averaged over all neurons, obtained by smoothing the spike trains with a Gaussian function with σ=10 ms. Plots below the rasters are cross-correlation histograms averaged over multiple distinct pairs of units. These were based on longer spike trains that included the segments shown.

tances gives rise to the sharp peak in the histogram of Figure 12.1a.

Figure 12.1b was generated using the same correlation values, but the excitatory signals $g_E(t)$ varied more slowly (in addition, their magnitude was adjusted so that similar output rates were produced). The characteristic time at which $g_E(t)$ varies is its correlation time. Below we describe this quantity more accurately; for the moment the crucial point is that in Figure 12.1 the correlation time of $g_E(t)$ corresponds to the time constant of excitatory synapses, τ_E. This is essentially the duration of a unitary synaptic event. In Figure 12.1a the synaptic time constants for excitation and inhibition were both equal to 2 ms. In Figure 12.1b τ_I stayed the same but τ_E was increased to 20 ms. As can be seen in the raster, this changed the postsynaptic responses: the spike trains are more irregular; the spikes of a single neuron tend to appear in clusters. The two timescales show up in the cross-correlogram as a sharp peak superimposed on a wider one. Figure 12.1c shows what happens when both synaptic time constants are set to 20 ms. Now the clustering of spikes in individual spike trains is even more apparent and the cross-correlogram shows a single, wide peak.

The correlations between conductances parameterize the degree of synchrony among output responses. When the correlations are 0 the responses are independent and the cross-correlogram is flat; when the correlations are equal to 1 all neurons are driven by the exact same signals and thus produce the same spike train — this is perfect synchrony. Figures 12.1a–12.1c were generated with correlations of 0.2, whereas Figures 12.1d–12.1f were generated with correlations of 0.5. Notice that the shapes of the histograms in the top and bottom rows are the same, but the y-axis scales in the latter are much larger. Larger correlations always produce more synchrony and larger fluctuations in instantaneous firing rates (continuous traces). In addition, they may also alter the postsynaptic firing rates, but this effect was intentionally eliminated in Figure 12.1 so that different synchrony patterns could be compared at approximately equal firing rates.

These examples show that there are at least two important factors determining the synchronous responses caused by common input: the amount of common input, which corresponds to the magnitudes of the correlations between conductances, and the timescales of the input signals, which in this case are determined by synaptic parameters. Analogously, there are two aspects of the cross-correlation function that are important, the height of the peak and its width.

12.4 Correlations arising from local network interactions

Networks of recurrently interconnected neurons may naturally give rise to oscillatory and synchronous activity at various frequencies; this is a well documented finding [94–17]. The type of activity generated depends on the network's architecture, on its inputs, and on single-cell parameters. Here we illustrate this phenomenon with a highly simplified network with the following properties. (1) Model neurons, excitatory and inhibitory, are of the integrate-and-fire type, without any intrinsic oscillatory mechanisms. (2) Synaptic connections between them are all-to-all and random, with strengths drawn from a uniform distribution between 0 and a maximum value g_{max}; this is both for excitatory and inhibitory contacts. (3) All neurons receive an external input drive implemented through fluctuating conductances $g_E(t)$ and $g_I(t)$, which are uncorrelated across neurons.

Figure 12.2 illustrates some of the firing patterns produced by such a network. For Figure 12.2a the recurrent connections were weak, i.e., g_{max} was small. The peak in the cross-correlogram is also small, indicating that the neurons fired nearly independently. The peak is narrow because all synaptic time constants were set to

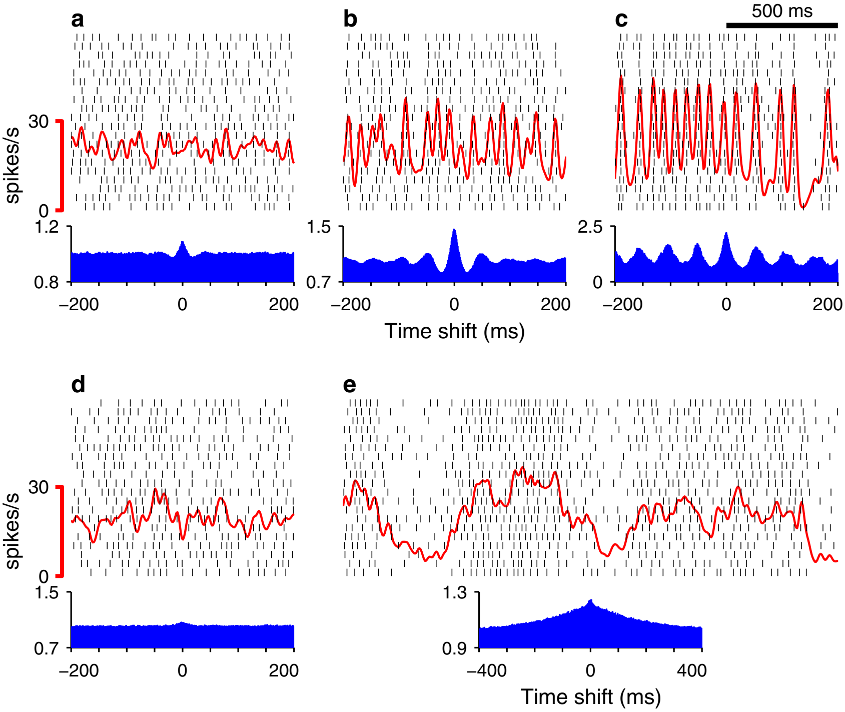

Figure 12.2

Spike trains correlated by local network interactions. Same format as in Figure 12.1. Neurons were modeled as integrate-and-fire units receiving two types of inputs: a background synaptic drive that was independent across neurons, and recurrent synaptic input from other units in the network. The full network consisted of 100 excitatory and 25 inhibitory model neurons. Synaptic connections were all-to-all, with conductances chosen randomly (uniformly) between 0 and a maximum value g_{max}. Note the different y-axes for cross-correlation histograms.

3 ms. Figures 12.2b and 12.2c show what happens when the connections are made progressivly stronger. The central peak becomes much taller (notice the different y-axes), and secondary peaks, indicating oscillatory activity, become apparent. In contrast to the rest of the cross-correlograms, the one in Figure 12.2c was generated from a short segment of data. This enhanced the secondary peaks, which practically disappeared when longer stretches of data were used (not shown). This is because the frequency of the oscillatory activity is not constant so, over a long time, many

phases are averaged out, making the correlation flat everywhere except in the central region. As in Figure 12.1, compensatory adjustments were made so that average firing rates remained approximately the same; in this case the external excitatory drive was slightly decreased as the connection strengths increased.

Figures 12.2d and 12.2e show that even such a simplified network may have quite complex dynamics. Parameters in Figure 12.2d were identical to those of Figure 12.2b, except for two manipulations. First, for recurrent excitatory synapses only, the synaptic time constant was increased from 3 to 10 ms; and second, to compensate for this, the synaptic conductances were multiplied by 3/10. This generated approximately the same firing rates and also preserved the average recurrent conductance level. However, as a consequence of these changes the correlations between postsynaptic spikes almost disappeared. Thus, the tendency to fire in phase is much larger when the characteristic timescales for excitatory and inhibitory synaptic events are the same. This is reminiscent of resonance.

Figure 12.2e illustrates another interesting phenomenon. In this case the timescales of both excitatory and inhibitory recurrent synapses were set to 10 ms, while the characteristic time of all external input signals stayed at 3 ms. The mean conductance levels, averaged over time, were the same as in Figure 12.2c, so the connections were relatively strong. Now the peak in the cross-correlation histogram (Figure 12.2e) is much wider than expected, with a timescale on the order of hundreds of milliseconds. Such long-term variations are also apparent in the spike raster and in the firing rate trace. This is quite surprising: firing fluctuations in this network occur with a characteristic time that is at least an order of magnitude longer than any intrinsic cellular or synaptic timescale. Discussion of the underlying mechanism is beyond the scope of this chapter, but in essence it appears that the network makes transitions between two pseudo steady-state firing levels, and that the times between transitions depend not only on cellular parameters but also on how separated the two firing levels are.

In any case, a key point to highlight is that, in all examples we have presented, the cross-correlation histograms show a common feature: a central peak with a shape that resembles a double-exponential. That is, the correlation function can be described as

$$C(t) = C_{max} \exp\left(-\frac{|t|}{\tau_{corr}}\right), \tag{12.2}$$

where t is the time lag and τ_{corr}, which determines its width, is the correlation time. This function is related to many types of random processes [87, 42]. Indeed, below we will use it to characterize the total input that drives a typical cortical neuron.

So far, we have looked at a variety of correlation patterns that a network may display. Next, we take the point of view of a single downstream neuron that is driven by this network. From this perspective we return to an important question posed in the introduction: how does the response of a postsynaptic neuron depend on the full set of correlated spike trains that typically impinge on it? Equation 12.2 will be used as a rough characterization of those correlations. The answer will be presented in two parts. First we will discuss some of the main factors determining whether input correlations have an impact on the postsynaptic response, and roughly to what degree. Then we will present a mathematical model that is somewhat abstract but that can be solved analytically and can provide some quantitative insight into the problem.

12.5 When are neurons sensitive to correlated input?

The goal of this section is to identify some of the main factors that determine the sensitivity of a postsynaptic cortical neuron to the presence of correlations in its inputs.

Synapses generate discrete events that are localized in time. Hence the basic intuition suggesting that timing is important: if action potentials from two excitatory neurons arrive simultaneously or within a short time window of each other to the same postsynaptic neuron, the two synaptic events may add up, producing a larger conductance change. Roughly, depending on the time interval between their arrivals, two presynaptic action potentials may act as two separate events of unit amplitude and duration, as one event of unit amplitude but lasting twice as long, or as one event of double amplitude and unit duration. If excitatory spikes have a tendency to arrive simultaneously more often than expected by chance, the target neuron might respond more vigorously.

This idea has been confirmed through simulation studies [13, 60]. Compared to independent spike trains, synchronous spikes may evoke stronger responses, but only up to a point, after which further synchronization actually decreases the response [13, 60]. This decrease occurs for two reasons. First, only a certain number of simultaneous excitatory synaptic events are required to trigger an action potential, so, once this number is reached, other simultaneous spikes cannot enhance the response. Second, excitatory spikes that arrive while the postsynaptic cell is in its refractory period are wasted. Thus, there is a tradeoff between two effects: on one hand, grouping excitatory spikes in time so that synaptic events summate; on the other, spreading them so that refractory effects are avoided.

This line of argument, however, has serious limitations. Refractory effects become important only when the output neuron is firing near its maximum rate, which is rarely the case. And more importantly, inhibition is not considered. Inhbition alters the scenario in three ways. (1) It may affect the sensitivity of the postsynaptic neuron to synchronous excitatory spikes. (2) Synchrony affects not only the average response of the cell but also the variability of the output spike train, and this too may depend on the level of inhibition. (3) Additional questions arise about the effects of synchrony between pairs of inhibitory input spikes and between excitatory-inhibitory pairs as well. In short, the situation gets considerably more complicated. The spectrum of possible firing modes of a neuron is often split between two extreme cases, integration and coincidence detection.

12.5.1 Coincidence detection

The classic mechanism underlying a neuron's sensitivity to temporal patterns is coincidence detection [2–50]. Neurons can certainly be sensitive to the arrival of spikes from two or more inputs within a short time window; the most notable examples are from the auditory system [5, 21]. The question is, however, whether this mechanism is commonly used throughout the cortex.

In the traditional view, coincidence detection is based on a very short membrane time constant [2–50]. However, it may be greatly enhanced by the spatial arrangement of synapses and by nonlinear processes. For instance, nearby synapses may interact strongly, forming clusters in which synaptic responses to simultaneous activation are much stronger than the sum of individual, asynchronous responses [58]. A neuron could operate with many such clusters which, if located on electrotonically distant parts of the dendritic tree, could act independently of each other. Voltage-dependent channels in the dendrites may mediate or boost such nonlinear interactions between synapses [58–64]. These nonlinearities could in principle increase the capacity for coincidence detection to the point of making the neuron selective for specific temporal sequences of input spikes, and the very idea of characterizing those inputs statistically would be questionable. However, the degree to which the cortex exploits such nonlinearities is uncertain.

The coincidence detection problem can also be posed in terms of the capacity of a network to preserve the identity of a volley of spikes fired by multiple neurons within a short time window [18, 31]. Suppose a neuron receives a volley of input spikes; what is the likelihood of evoking a response (reliability), and what will its timing be relative to the center of mass of the input volley (precision)? Theoretical studies suggest that the temporal precision of the response spikes is not limited by the membrane time constant, but rather by the up-slope of excitatory synaptic events. Thus, under the right conditions a volley of synchronized action potentials may propagate in a stable way through many layers [31]. Whether areas of the cortex actually exchange information in this way is still unclear, and other modes of information transmission are possible [88].

12.5.2 Fluctuations and integrator models

The flip side of coincidence detection is integration. Neurons may also sum or average many inputs to generate an action potential [2, 50, 73]. Earlier theoretical arguments suggested that neurons acting as integrators would not be sensitive to temporal correlations [74], or that these would only matter at high firing rates, where refractory effects become important [13, 60]. However, later results [67, 69] show that neurons may still be highly sensitive to weak correlations in their inputs even if there is no spatial segregation along the dendritic tree and no synaptic interactions beyond the expected temporal summation of postsynaptic currents.

A key quantity in this case is the balance of the neuron, which refers to the relative strength between inhibitory and excitatory inputs [67, 82, 73]. When the neuron is not balanced, excitation is on average stronger than inhibition, such that the net synaptic current is depolarizing and the mean steady-state voltage is near or above threshold. In this case the main driving force is the drift toward steady state, and input fluctuations have a small effect on the rate of output spikes [67, 33]. On the other hand, when the neuron is balanced, both excitation and inhibition are strong, the mean input current is zero or very small, and the mean steady-state voltage remains below threshold. However, the neuron may still fire because there are large voltage fluctuations that lead to random threshold crossings. In this mode, any factor that enhances the fluctuations will produce more intense firing [67, 32].

There is a subtle but important distinction between mechanisms that may alter input fluctuations. Higher rates should be seen in a balanced neuron if fluctuations increase without affecting the mean synaptic conductances, as when only the correlations change [67]. But if stronger fluctuations are accompanied by increases in total conductance, as when both excitatory and inhibitory inputs fire more intensely, the firing rate may actually decrease [32–22]. In a complex network these effects may be hard to disentangle.

Figure 12.3 compares the responses of balanced (upper traces) and unbalanced (lower traces) model neurons [67]. These were driven by excitatory and inhibitory input spike trains similar to those illustrated in Figure 12.1. For the balanced neuron both excitatory and inhibitory synaptic conductances were strong, and the combined current they generated near threshold was zero. In contrast, for the unbalanced unit both conductances were weak, but their combined current near threshold was excitatory. The four panels correspond to different correlation patterns in the inputs. In Figure 12.3a all inputs are independent, so all cross-correlograms are flat. The voltage traces reveal a typical difference between balanced and unbalanced modes: although the output rate is approximately the same, the subthreshold voltage of the balanced neuron is noisier and its interspike intervals are more variable [67, 82]. Figure 12.3b shows what happens when the excitatory inputs fire somewhat synchronously due to common input. The firing rate of the balanced neuron always increases relative to the response to independent inputs, whereas the rate of the unbalanced neuron may show either a smaller (although still substantial) increase or a decrease [13, 60]. Another effect of synchrony is to increase the variability of the output spike trains, both for balanced and unbalanced configurations [67, 69, 78, 80]; this can be seen by comparing Figures 12.3b and 3d with Figure 12.3a. Correlations between inhibitory inputs can also produce stronger responses. When the inhibitory drive oscillates sinusoidally, as in Figure 12.3c, the balanced neuron practically doubles its firing rate compared to no oscillations; in contrast, the unbalanced does not change.

The balance of a neuron is important in determining its sensitivity to correlations, but there is another key factor [67]. There are three correlation terms: correlations between pairs of excitatory neurons, between pairs of inhibitory neurons, and between excitatory-inhibitory pairs. The first two terms increase the voltage fluctuations but the last one acts in the opposite direction, decreasing them. The total effect on the postsynaptic neuron is a function of the three terms. In Figure 12.3 d, all inputs to the model neurons are equally correlated, but the balanced model shows no change in firing rate. Thus, it is possible to have strong correlations between all

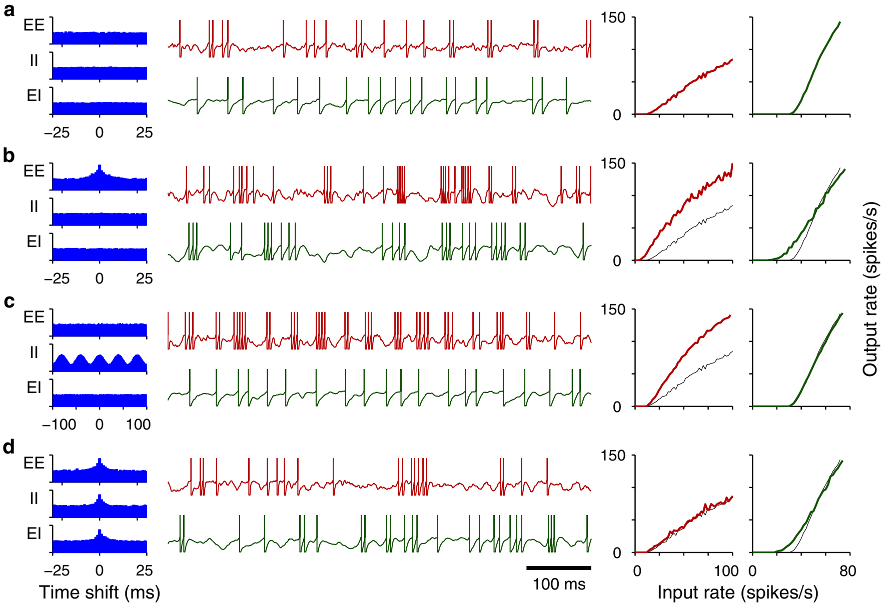

Figure 12.3

Responses of two model neurons to four input correlation patterns. Histograms on the left show average cross-correlations between pairs of excitatory input spike trains (EE), between inhibitory pairs (II), and between excitatory-inhibitory pairs (EI). Y-axes in the correlograms go from 0.7 to 1.4. Upper and lower traces in each panel show the responses of balanced and unbalanced neurons, respectively. The rate of inhibitory inputs was always equal to 1.7 times the excitatory rate. For all middle traces the excitatory input rate was 42 spikes/s. The plots on the right show the firing rates of the two (postsynaptic) model neurons versus the mean firing rate of the (presynaptic) excitatory inputs. Thin black lines are the curves obtained with independent inputs (top panel). The two output neurons were leaky integrate-and-fire units with identical parameters; they differed in the relative strength of their excitatory and inhibitory inputs. (Adapted from [67] and [68].)

inputs but still not see a change in the firing rate of the postsynaptic neuron relative to the case of independent inputs.

In summary, a balanced neuron is much more sensitive to input correlations than an unbalanced one because correlations affect the fluctuations in synaptic drive, which cause the balanced neuron to fire. However, the postsynaptic response depends on the relative values of the three correlation terms, which may cancel out. The key point here is that even when neurons act as integrators they can, in a statistical sense, be highly sensitive to the temporal patterns of their input spikes.

Interestingly, at least in some pyramidal neurons, distal dendrites seem to act much more like coincidence detectors than proximal dendrites [93], so real neurons may,

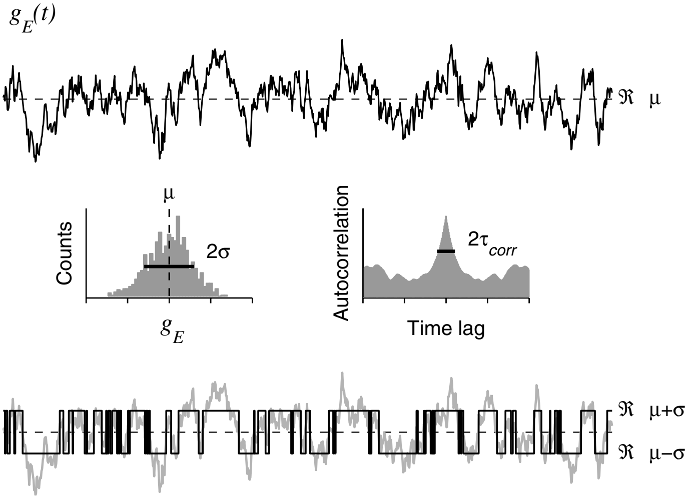

Figure 12.4

Parameterization of a continuous conductance trace. The top graph represents the total synaptic conductance generated by excitatory spikes driving a postsynaptic neuron. This conductance can be characterized statistically by its mean μ, standard deviation σ, and correlation time τ_{corr}. The left histogram is the distribution of conductance values of the top trace. The histogram on the right is its autocorrelation function. The width of the peak is parameterized by τ_{corr}. The bottom graph shows a binary variable that approximates the continuous trace. The binary function has the same mean, standard deviation and correlation time as the original function [69].

to some extent, combine both types of firing modes.

12.6 A simple, quantitative model

Now we discuss a simple model for which the responses to correlated input can be calculated analytically [69]. The first step is to describe its input.

12.6.1 Parameterizing the input

The input to a neuron consists of two sets of spike trains, ones that are excitatory and others that are inhibitory. What are the total synaptic conductances generated by these spikes? How can they be characterized? An example generated through a computer simulation is shown in Figure 12.4. The top trace represents the total excitatory conductance $g_E(t)$ produced by the constant bombardment of excitatory synapses onto a model neuron. When plotted versus time, the time course looks noisy, random. Because $g_E(t)$ is the result of thousands of individual synaptic events, the distribution of conductance values should be approximately Gaussian, with mean and standard deviation

$$\begin{aligned} \mu &= \langle g_E \rangle \\ \sigma &= \sqrt{\left\langle (g_E - \mu)^2 \right\rangle}, \end{aligned} \tag{12.3}$$

where the angle brackets <> indicate an average over time. The histogram on the left in Figure 12.4 shows the distribution of g_E values for the top trace. Indeed, it is close to a Gaussian, even though the trace is relatively short (1 s long sampled at 1 ms intervals). The mean and standard deviation, however, are not enough to characterize the conductance trace because it fluctuates with a typical timescale that has to be determined independently. The histogram on the right shows the autocorrelation function of the trace. This is akin to the cross-correlation functions discussed earlier, except that the correlation is between a continuous function and itself. Now a peak centered at zero time lag indicates that $g_E(t)$ and $g_E(t + \Delta t)$ tend to be similar to each other, and the width of the peak tells how fast this tendency decreases. A flat autocorrelation means that all values of g_E were drawn randomly and independently of each other. Thus, an autocorrelation function that is everywhere flat except for a peak centered at zero is the signature of a stochastic function that varies relatively smoothly over short timescales but whose values appear entirely independent when sampled using longer intervals. The autocorrelation function can be computed analytically for a variety of noise models, and it is typically a double exponential, as in Equation 12.2, with $C_{max} = \sigma^2$. Identical considerations apply to the conductance generated by inhibitory synapses.

From Figures 12.1 and 12.2 and from these observations, it appears that a reasonable framework to describe the total excitatory and inhibitory conductances that drive a cortical neuron is to model them using two random signals with given means, standard deviations and correlation times. Indeed, this approach has been tested experimentally, with highly positive results [29, 30]. This is also what was done to generate the spikes in Figure 12.1 (see Appendix). As explained below, this approximation is very good; for the leaky integrate-and-fire model the responses obtained using this method versus actual spike trains are virtually identical within a large parameter range (not shown).

In general, calculating the three parameters for $g_E(t)$ or $g_I(t)$ from the quantities that parameterize the corresponding input spike trains is difficult. However, this can be done under the following simplifying assumptions. Suppose there are N_E excitatory spike trains that are independent, each with Poisson statistics and a mean rate r_E. Also suppose that the synapses operate like this: whenever an input spike arrives, $g_E(t)$ increases instantaneously by an amount G_E; otherwise, $g_E(t)$ decreases exponentially toward zero with a time constant τ_E (see ref. [24]). For this simple scheme it can be shown [81] that

$$\begin{aligned}\mu &= G_E N_E r_E \tau_E \\ \sigma^2 &= \frac{G_E^2 N_E r_E \tau_E}{2} \\ \tau_{corr} &= \tau_E \,, \end{aligned} \tag{12.4}$$

with the correlation function having a double-exponential shape. Thus, for this situation, a model neuron in a simulation can be driven by two methods. First, by generating independent Poisson spikes and increasing the conductance every time one such spike arrives, exactly as described above. In this case parameters G_E, r_E, N_E, τ_E, and the corresponding quantities for inhibitory inputs need to be specified. The second method is to generate the fluctuating signals $g_E(t)$ and $g_I(t)$ directly by combining random numbers, in which case only the respective μ, σ and τ_{corr} are strictly required. Neuronal responses evoked using this type of model can match experimental data quite well [29, 30].

When the assumptions of the case just discussed are violated, for instance, when the spikes driving a neuron are not independent, determining μ, σ and τ_{corr} analytically becomes much more difficult. However, in general one should expect correlations to increase σ, and the correlation time should be equal to the synaptic time constant, although it may increase further if the input spikes are correlated over longer timescales.

Next, we ask how each of the three key parameters, μ, σ and τ_{corr}, affects the response of a postsynaptic neuron.

12.6.2 A random walk in voltage

The model neuron we consider is the non-leaky, integrate-and-fire neuron [67, 69, 40], whose dynamics resemble those of random walk models used to study diffusion in physical systems [87, 42, 41, 12]. The voltage V of this unit changes according to the input $I(t)$ that impinges on it, such that

$$\tau \frac{dV}{dt} = I(t) \,, \tag{12.5}$$

where τ is its integration time constant. In this model an action potential is produced when V exceeds a threshold V_θ. After this, V is reset to an initial value V_{reset} and the integration process continues evolving according to the equation above. This model is related to the leaky integrate-and-fire model [24, 67, 82] but it lacks the term proportional to $-V$ in the right-hand side of the differential equation. An additional and crucial constraint is that V cannot fall below a preset value, which acts as a barrier. For convenience the barrier is set at V=0, so only positive values of V are allowed. This choice, however, makes no difference in the model's dynamics. Except for the barrier and the spike-generating mechanism, this model neuron acts as an ideal integrator, with an integration time constant τ.

Here the input $I(t)$ is the total current, including excitatory and inhibitory components. To simplify things even further, we will consider $I(t)$ to be a noisy function with mean μ, standard deviation σ, and correlation time τ_{corr}. Note, however, that these quantities now refer to the total current, not to the conductances, as before; this is just to simplify the notation. In this scheme it is not clear how exactly $I(t)$ is related to $g_E(t)$ and $g_I(t)$, which in principle are the measurable parameters of real neurons. However, evidently μ should depend on the means of the conductances, σ should depend on their standard deviations, and τ_{corr} should depend on their correlation times. This qualitative relationship is good enough to proceed because the model is somewhat abstract anyway.

The quantity that we are interested in is T, the time that it takes for V to go from reset to threshold. T is known as the first passage time or the interspike interval; what we want to know are its statistics. The key for this [69] is to rewrite Equation 12.5 as follows

$$\tau \frac{dV}{dt} = \mu + \sigma Z(t)\,, \tag{12.6}$$

where $Z(t)$ is a binary variable that can only be either $+1$ or -1 and whose correlation function is a double exponential with correlation time τ_{corr}. Thus $I(t)$ has been replaced by a stochastic binary function that indicates whether $I(t)$ is above or below its average. This approximation is illustrated in Figure 12.4 (bottom trace). The binary function has the same mean, standard deviation and correlation time as $I(t)$. This substitution allows us to solve Equation 12.6 analytically [69]. Notice also that the neuron's time constant τ simply acts as a scale factor on the input. Hereafter it will be considered equal to 1.

Figure 12.5 shows examples of spike trains produced by the model when driven by the binary, temporally correlated input. In this figure μ was negative, so on average the voltage tended to drift away from threshold, toward the barrier. In this case the spikes are triggered exclusively by the random fluctuations, as measured by σ; without them the neuron would never reach threshold. In Figures 12.5a–12.5c the correlation time is τ_{corr}=1 ms. For a binary variable like Z, which switches between +1 and -1, the correlation time corresponds to the average time one needs to wait to observe a change in sign. In other words, the correlation time is equal to half the average time between sign changes. Thus, the input in Figure 12.5a (lower trace) flips state approximately every 2 ms. Figure 12.5c shows that, under these conditions, the neuron fires at a relatively low rate and irregularly; the times between spikes or interspike intervals are quite variable, which can also be seen from the interspike-interval distribution in Figure 12.5b.

When τ_{corr} is increased to 5 ms, as in Figures 12.5d–12.5f, the changes in input state occur approximately every 10 ms (Figure 12.5d, lower trace). This produces a large increase in mean firing rate and, to a lesser extent, an increase in variability. This can be seen by comparing the spike trains from Figures12.5c and 12.5f. The respective mean rates are 10 and 37 spikes/s. Notice that there is a short time interval that appears very frequently. The short interval results when the input stays positive for a relatively long time, as is the case with the pair of spikes in Figure 12.5d. This interval is equal to $(V_\theta - V_{reset})/(\mu + \sigma)$, which is the minimum separation between

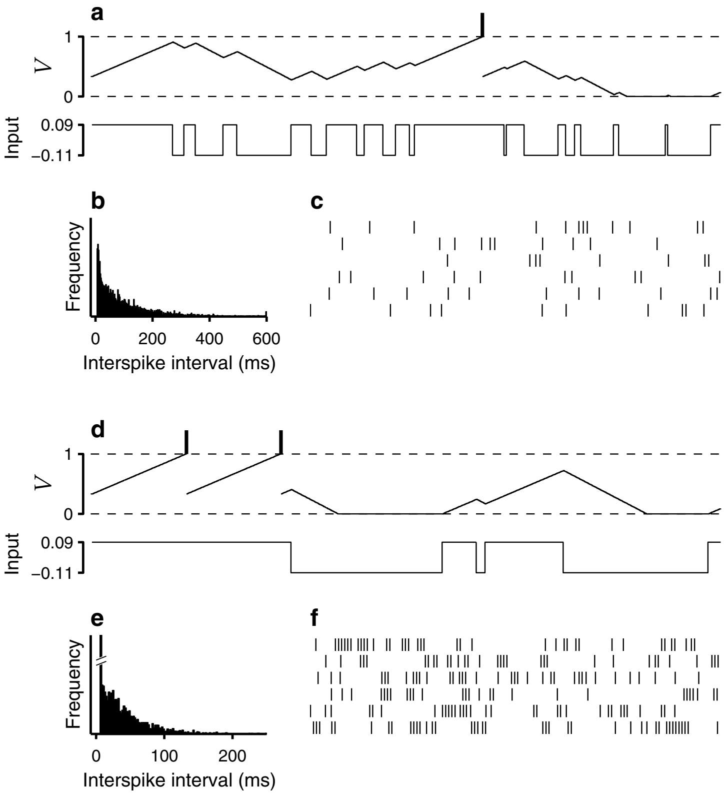

Figure 12.5

Responses of the nonleaky integrate-and-fire model driven by correlated, binary noise. The input switches states randomly, but on average the same state is maintained for $2\tau_{corr}$ ms. Sample voltage and input time courses are 50 ms long. Raster plots show 6 seconds of continuous simulation time. For the three top panels the correlation time τ_{corr} was 1 ms; for the lower panels it was 5 ms. (Adapted from [69].)

spikes in the model given μ and σ. The number of spikes separated by this interval grows as the correlation time increases. At the same time, however, longer correlation times also give rise to long interspike intervals, which occur because the input can stay in the low state for longer stretches of time. This is why correlation time increases variability: it produces both short and long interspike intervals. The quantity that is most often used to measure the regularity of a spike train is the coefficient of variation, or CV_{ISI}, which is equal to the standard deviation of the interspike intervals divided by their mean. The CV_{ISI} in Figures 12.5c is equal to 1, as for a Poisson process; in Figure 12.5f it is equal to 1.18, which reflects the higher variability. Note that μ and σ are the same for all panels. This demonstrates that the input correlation time may have a very strong impact on the response of a postsynaptic neuron [69]. This is an interesting observation because little is known about the dynamic role of this parameter.

12.6.3 Quantitative relationships between input and output

The solution to the non-leaky model of Equation 12.6 consists in the moments of T, $\langle T\rangle$, $\langle T^2\rangle$ and so forth. For each of these moments there are three sets of analytic expressions, because details of the solutions depend on the relative values of μ and σ. Here we only discuss the expressions for the average interspike interval $\langle T\rangle$, which is the inverse of the mean firing rate, but $\langle T^2\rangle$ and therefore the CV_{ISI} can also be obtained in closed form [69].

When $\mu > \sigma$,

$$\langle T\rangle = \frac{V_\theta - V_{reset}}{\mu}\,. \tag{12.7}$$

In this case there is a strong positive drift toward threshold. Even when Z is equal to -1 the total input is positive; in other words, the voltage gets closer to threshold in every time step, whether the fluctuating component is positive or negative. The mean firing rate behaves as if the input were constant and there were no fluctuations. This can be seen in Figure 12.6, which plots the mean firing rate and the CV_{ISI} of the model neuron as a function of σ for various combinations of the other two input parameters. The values of μ are indicated in each column, and the three curves in each plot correspond to τ_{corr} equal to 1, 3 and 10 ms, with higher correlation values always producing stronger responses and higher variability. Continuous lines and dots correspond to analytic solutions and simulation results, respectively. Notice how, when μ=0.02, the firing rate stays constant for σ below 0.02, although the variability increases most sharply precisely within this range.

When μ=0,

$$\langle T\rangle = \frac{2\left(V_\theta - V_{reset}\right)}{\sigma} + \frac{V_\theta^2 - V_{reset}^2}{2\tau_{corr}\sigma^2}\,. \tag{12.8}$$

Clearly, the average interspike interval decreases with both τ_{corr} and σ. In this case there is no drift, no net displacement; the voltage advances toward threshold when Z=+1 and retreats toward the barrier when $Z = -1$. Under these conditions the neuron is driven exclusively by fluctuations. The middle column of Figure 12.6corresponds to this regime. As can be seen, the variability of the neuron also increases monotonically with σ and τ_{corr}.

Finally, when $\mu \leq \sigma$,

$$\langle T\rangle = \frac{V_\theta - V_{reset}}{\mu} + \tau_{corr}(c-1)^2\left(\exp\left(-\alpha V_\theta\right) - \exp\left(-\alpha V_{reset}\right)\right)\,, \tag{12.9}$$

where we have defined

$$c \equiv \frac{\sigma}{\mu}$$

$$\alpha \equiv \frac{1}{\mu\tau_{corr}(c^2-1)}\,. \tag{12.10}$$

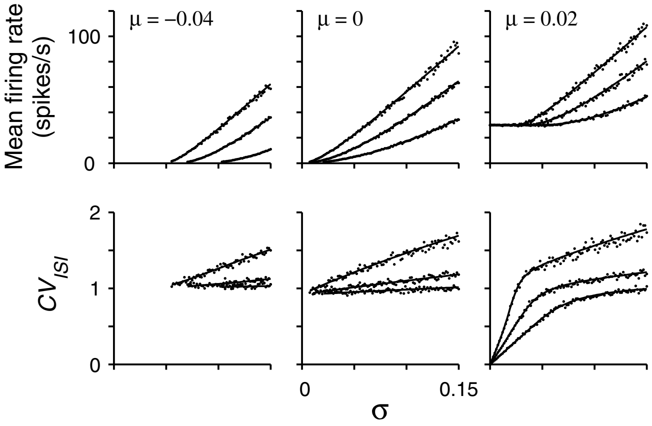

Figure 12.6

Mean firing rate and coefficient of variation for the nonleaky integrate-and-fire neuron driven by correlated binary noise. Continuous lines are analytic expressions and dots are results from computer simulations. Each simulation data point was based on spike trains containing 2000 spikes. The three curves in each graph are for different values of τ_{corr}: 1 ms (lower curves), 3 ms (middle curves), and 10 ms (upper curves). (Adapted from [69].)

As with the above equations, Figure 12.6 reveals the excellent agreement between this expression and computer simulations. An interesting special case is obtained when σ=μ, or c=1. Then the total input is zero every time that Z equals -1, so half the time V does not change and half the time V increases by 2μ in each time step. Therefore, the average time to threshold should be equal to $(V_\theta - V_{reset})/\mu$, which is precisely the result from Equation 12.9. This quantity does not depend on the correlation time, but the CV_{ISI} does. The analytic expression for the CV_{ISI} is particularly simple in this case:

$$\sqrt{\frac{2\,\mu\,\tau_{corr}}{V_\theta - V_{reset}}}. \tag{12.11}$$

Thus, the variability of the output spike train diverges as τ_{corr} increases, but the mean rate does not.

This last observation is valid in a more general sense, and is an important result regarding the effects of correlations. In most cases, the limit behaviors of the firing rate and the CV_{ISI} as the correlation time increases are quite different: the rate tends to saturate, whereas the variability typically diverges. This is illustrated in Figure 12.7. The one condition in which the variability saturates as the correlation time tends to infinity is when μ is larger than σ (thickest line on right column). The asymptotic value of the CV_{ISI} in this case is $c/\sqrt{1-c^2}$. In this parameter regime the drift is strong, so it usually produces high firing rates as well.

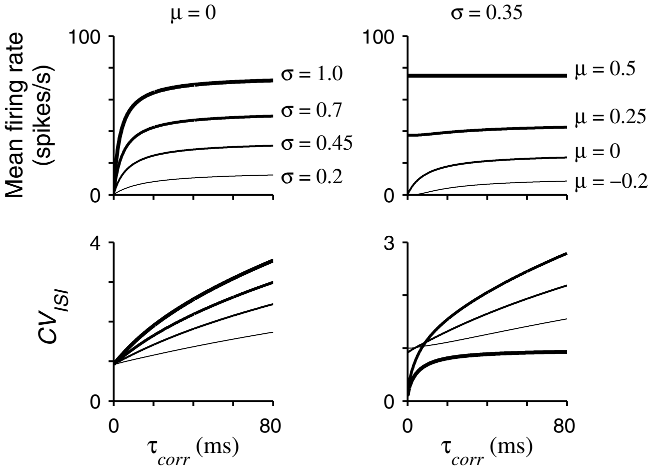

Figure 12.7

Responses of the nonleaky integrate-and-fire neuron as functions of input correlation time τ_{corr}. Only analytic results are shown. As the correlation time increases, the firing rate always tends to an asymptotic value. In contrast, the CV_{ISI} diverges always, except when $\mu > \sigma$; this case corresponds to the thickest line in the plots on the right. (Adapted from [69].)

The key to obtain all the analytic expressions was the use of a binary input. One may wonder, however, whether the results are valid with a more realistic input signal. It turns out that, for this model, the mean firing rate and the CV_{ISI} obtained using correlated Gaussian noise are very similar to those obtained with binary noise. This is not entirely unexpected, first, because the neuron essentially adds its inputs, and second, because Gaussian noise can be properly approximated as the sum of multiple binary random samples, as a consequence of the central limit theorem. This is strictly true when all binary and Gaussian samples are independent, that is, when the autocorrelation functions are everywhere flat, but the approximation works quite well even when there is a correlation time. For example, the rate and the CV_{ISI} still increase as functions of correlation time, and the same asymptotic behaviors are seen [69].

12.7 Correlations and neuronal variability

The spike trains of neurons recorded in awake animals are highly variable [25, 75–78]. However, spike generation mechanisms themselves seem to be highly reliable [20, 49, 56]. The contrast between these two observations stirred a fair amount of discussion, especially after the work of Softky and Koch [76], who pointed out that although the CV_{ISI} of typical cortical neurons is close to 1, this number should be much lower for an integrator that adds up many small contributions in order to fire, especially at high output rates. However, their arguments applied in the absence of inhibition, and later work [82, 73] showed that including incoming inhibitory spikes produces higher CV_{ISI} values even in integrator models without any built-in coincidence detection mechanisms [2–50] or similar nonlinearities [58–64], a result that is consistent with early stochastic models [41, 84]. So-called 'balanced' models, in which inhibition is relatively strong, typically bring the CV_{ISI} to the range between 0.5 and 1 [73, 82], which is still lower than reported from recorded data [25, 75–78]. Other intrinsic factors have also been identified as important in determining spike train variability; for instance, combining the proper types of conductances [11], tuning the cellular parameters determining membrane excitability [47, 82], and bistability [92].

However, several lines of evidence point to correlations in the conductances (or currents) that drive a neuron as a primary source of variability. First, correlated firing is ubiquitous. This has been verified through a variety of techniques, including *in vivo* experiments in which pairs of neurons are recorded simultaneously. The widths of the corresponding cross-correlograms may go from a few to several hundred milliseconds [43–77], so they may be much longer than the timescales of common AMPA and GABA-A synapses [27]. Second, *in vitro* experiments in which neurons are driven by injected electrical current suggest that input correlations are necessary to reproduce the firing statistics observed *in vivo* [28, 29, 30, 78]. This is in line with the suggestion that fluctuations in eye position are responsible for a large fraction of the variability observed in primary visual neurons, because they provide a common, correlating signal [46]. Third, this also agrees with theoretical studies [33, 67]; in particular with results for the non-leaky integrate-and fire model showing that the CV_{ISI} depends strongly on the correlation time of the input [69]. In addition, similar analyses applied to the traditional leaky integrate-and-fire model reveal the same qualitative dependencies [69]. This, in fact, can be seen in Figure 12.1, where the model with leak was used: increases in the synaptic time constants give rise to longer correlation times and to higher CV_{ISI} values (compare Figures 12.1a and 12.1c), an effect that has nothing to do with the synchronization between output spike trains. Finally, high variability is also observed in simulation studies in which network interactions produce synchronized recurrent input [85–89], as in Figure 12.2.

12.8 Conclusion

The activity of a local cortical microcircuit can be analyzed in terms of at least two dimensions, its intensity, which is typically measured by the mean firing rates of the neurons, and its coherence across neurons, which is often described in terms of synchrony or cross-correlations between pairs of units. These correlations serve as probes for the organization and dynamics of neural networks. There is strong evidence, both theoretical and experimental, indicating that correlations may be important dynamic components of cortical microcircuits. Here we have discussed two general hypotheses, the encoding of stimulus features and the gating of information from one structure to another. Although quite different, both are based on the premise that correlations have a specific functional role. Interestingly, there is an interpretation that is entirely opposite to the sensory-coding hypothesis, which suggests that correlations between cortical neurons limit the accuracy with which neural populations may encode stimulus features [95] (see also [1]). These top-down ideas have generated considerable debate, but the crucial question remains unresolved as to whether correlations have a specific, separate functional role, or whether they simply participate in all functions, just as firing rates do. It is also conceivable that there is no generic strategy, and that the meaning and impact of correlations vary from one local microcircuit to another.

A second, critical question is whether correlations can be controlled independently of firing rates. That is, a group of neurons M may affect another group A in two, not necessarily exclusive, ways: by changing the firing rates of A or the correlations between local neurons in A. There are two knobs that can be turned, and the question is whether these can be turned independently of each other. Here we reviewed some studies that begin to address this issue by taking the point of view of a single neuron: what intrinsic properties make it sensitive to correlations? How do correlations affect its response? Can changes in input correlations and input firing rates be distinguished? The hope is that this bottom-up perspective will eventually help clarify the top-down ideas by identifying and constraining the role of correlations in local circuit dynamics. A good example of this is the above section on neuronal variability. The highly variable discharge of cortical neurons is observed and characterized in recordings from awake, behaving preparations; and experiments *in vitro*, as well as computational and theoretical studies, identify a variety of biophysical mechanisms responsible for the observation. In this particular case, input correlations seem to play a major role because they can generate highly variable output spike trains in the absence of any additional intrinsic mechanisms [28, 46, 69, 78].

In conclusion, the two questions just pondered may represent high- and low-level interpretations of the same phenomenon, but a conceptual framework providing a unified view of this problem is still lacking. Establishing such framework, however, may serve as a guidelight for future investigations.

12.9 Appendix

Here we describe the leaky integrate-and-fire model [24, 67, 82, 84] driven by conductance changes that was used to generate Figure 12.1. In this model, the membrane potential V evolves according to

$$\tau_m \frac{dV}{dt} = -V - g_E(t)(V - V_E) - g_I(t)(V - V_I), \tag{12.12}$$

where the resting potential has been set to 0 mV. The spike-generating currents are substituted by a simple rule: whenever V exceeds a threshold (20 mV), a spike is emitted and V is clamped to a reset value (10 mV) for a refractory period (1.8 ms). After that, V continues evolving according to the above equation. The excitatory and inhibitory conductances, $g_E(t)$ and $g_I(t)$, were generated by combining Gaussian random numbers [69], so that the resulting traces would have the desired mean, standard deviation and correlation time. These parameters were related to input rates and model synaptic conductances through Equations 12.4. For Figures 12.1a and 1d, $N_E r_E$=27.5 spikes/ms, G_E=0.02, τ_E=2 ms, $N_I r_I$=12.15 spikes/ms, G_I=0.06, and τ_I=2 ms, with G_E and G_I in units of the leak conductance (i.e., where the leak conductance equals 1). For Figures 12.1b and 12.1e, τ_E=20 ms. For Figures 12.1c and 12.1f, τ_I=τ_E=20 ms. Correlations between the conductances of different neurons were generated by drawing correlated Gaussian samples during generation of the $g_E(t)$ and $g_I(t)$ traces for different neurons. The correlation coefficient for a pair of conductances, Equation 12.1, is equal to the correlation coefficient between the corresponding Gaussian samples. Other parameters were: τ_m=20 ms, V_E=74 mV, V_I=-10 mV, Δt=0.1 ms.

Acknowledgments Research was supported by the Howard Hughes Medical Institute and by startup funds from Wake Forest University School of Medicine. Parts of this chapter were adapted from [68].

References

[1] Abbott L.F., and Dayan P. (1999) The effect of correlated activity on the accuracy of a population code. *Neural Comput. 11:* 91–101.

[2] Abeles M. (1982) Role of the cortical neuron: Integrator or coincidence detector? *Israel J Med Sci* **18:** 83–92.

[3] Aertsen A., and Arndt M. (1993) Response synchronization in the visual cortex. *Curr. Opin. Neurobiol.* **3:** 586–594.

[4] Aertsen A.M.H.J., Gerstein G.L., Habib M.K., and Palm G. (1989) Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J Neurophysiol* **61:** 900–917.

[5] Agmon-Snir H., Carr C.E., and Rinzel J. (1998) The role of dendrites in auditory coincidence detection *Science* **393:** 268–272.

[6] Ahissar E., and Arieli A. (2001) Figuring space by time. *Neuron* **32**: 185–201.

[7] Ahissar E., Sosnik R., and Haidarliu S. (2000) Transformation from temporal to rate coding in a somatosensory thalamocortical pathway. *Nature* **406**: 302–306.

[8] Barlow J.S. (1993) *The Electroencephalogram: Its Patterns and Origins.* Cambridge, MA: MIT Press.

[9] Bazhenov M., Stopfer M., Rabinovich M., Abarbanel H.D.I., Sejnowski T.J., and Laurent G. (2001) Model of cellular and network mechanisms for odor-evoked temporal patterning in the locust antennal lobe. *Neuron* **30:** 569–581.

[10] Bazhenov M., Stopfer M., Rabinovich M., Huerta R., Abarbanel H.D.I., Sejnowski T.J., and Laurent G. (2001) Model of transient oscillatory synchronization in the locust antennal lobe. *Neuron* **30:** 553–567.

[11] Bell A.J., Mainen Z.F., Tsodyks M., and Sejnowski T.J. (1995) 'Balancing' of conductances may explain irregular cortical spiking. Technical Report INC-9502, Institute for Neural Computation, UCSD, San Diego, CA, 92093–0523.

[12] Berg H.C. (1993) *Random Walks in Biology.* Princeton, NJ: Princeton University Press.

[13] Bernander Ö., Koch C., and Usher M. (1994) The effects of synchronized inputs at the single neuron level. *Neural Comput.* **6:** 622–641.

[14] Braitenberg V., and Schüz A. (1997) *Cortex: Statistics and Geometry of Neuronal Connectivity.* Berlin: Springer-Verlag.

[15] Brody C.D. (1999) Correlations without synchrony. *Neural Comput.* **11:** 1537–1551.

[16] Brosch M., and Schreiner C.E. (1999) Correlations between neural discharges are related to receptive field properties in cat primary auditory cortex. *Eur. J. Neurosci.* **11:** 3517–3530.

[17] Brunel N., and Hakim V. (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* **11:** 1621-1671.

[18] Burkitt A.N., and Clark G.M. (1999) Analysis of integrate-and-fire neurons: synchronization of synaptic input and spike output. *Neural Comput.* **11:** 871–901.

[19] Bush P., and Sejnowski T.J. (1996) Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *J.*

Comput. Neurosci. **3:** 91–110.

[20] Calvin W.H., and Stevens C.F. (1968) Synaptic noise and other sources of randomness in motoneuron interspike intervals. *J. Neurophysiol.* **31:** 574–587.

[21] Carr C.E., and Friedman M.A. (1999) Evolution of time coding systems. *Neural Comput.* **11**: 1–20.

[22] Chance F.S., Abbott L.F., and Reyes A.D. (2002) Gain modulation from background synaptic input. *Neuron* **35:** 773–782.

[23] Dan Y., Alonso J.M., Usrey W.M., and Reid R.C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neurosci.* **1:** 501–507.

[24] Dayan P., and Abbott L.F. (2001) *Theoretical Neuroscience*. Cambridge, MA: MIT Press.

[25] Dean A. (1981) The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* **44:** 437–440.

[26] DeCharms R.C., and Merzenich M.M. (1995) Primary cortical representation of sounds by the coordination of action potential timing. *Nature* **381:** 610–613.

[27] Destexhe A., Mainen Z.F., and Sejnowski T.J. (1998) Kinetic models of synaptic transmission. In: *Methods in Neuronal Modeling* (second edition), C. Koch, I. Segev, eds., pp. 1–25, Cambridge, MA: MIT Press.

[28] Destexhe A., and Paré D. (1999) Impact of network activity on the integrative properties of neocortical pyramidal neurons *in vivo*. *J. Neurophysiol.* **81:** 1531–1547.

[29] Destexhe A., and Paré D. (2000) A combined computational and intracellular study of correlated synaptic bombardment in neocortical pyramidal neurons *in vivo*. *Neurocomputing* **32-33:** 113–119.

[30] Destexhe A., Rudolph M., Fellous J.M., and Sejnowski T.J. (2001) Fluctuating synaptic conductances recreate *in vivo*-like activity in neocortical neurons. *Neuroscience* **107:** 13–24.

[31] Diesmann M., Gewaltig M.-O., and Aertsen A. (1999) Stable propagation of synchronous spiking in cortical neural networks. *Nature* **402:** 529–533.

[32] Doiron B., Longtin A., Berman N., and Maler L. (2001) Subtractive and divisive inhibition: effect of voltage-dependent inhibitory conductances and noise. *Neural Comput.* **13:** 227–248.

[33] Feng J., and Brown D. (2000) Impact of correlated inputs on the output of the integrate-and-fire model. *Neural Comput.* **12:** 671-692.

[34] Fries P., Neuenschwander S., Engel A.K., Goebel R., and Singer W. (2001) Rapid feature selective neuronal synchronization through correlated latency

shifting. *Nature Neurosci.* **4:** 194–200.

[35] Fries P., Reynolds J.H., Rorie A.E., and Desimone R. (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291:** 1560–1563.

[36] Fries P., Roelfsema P.R., Engel A.K., König P., and Singer W. (1997) Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 12699–12704.

[37] Fries P., Schröder J.-H., Roelfsema P.R., and Singer W., Engel AK (2002) Oscillatory neural synchronization in primary visual cortex as a correlate of stimulus selection. *J. Neurosci.* **22:** 3739–3754.

[38] Frost Jr. J.D. (1967) An averaging technique for detection of EEG-intracellular potential relationships. *Electroenceph. Clin. Neurophysiol.* **23:** 179–181.

[39] Fuentes U., Ritz R., Gerstner W., and Van Hemmen J.L. (1996) Vertical signal flow and oscillations in a three-layer model of the cortex. *J. Comput. Neurosci.* **3:** 125–136.

[40] Fusi S., and Mattia M. (1999) Collective behavior of networks with linear (VLSI) Integrate and Fire Neurons *Neural Comput.* **11:** 633–652.

[41] Gerstein G.L., and Mandelbrot B. (1964) Random walk models for the spike activity of a single neuron. *Biophys. J.* **4:** 41–68.

[42] Gardiner C.W. (1985) *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Berlin: Springer-Verlag.

[43] Gochin P.M., Miller E.K., Gross C.G., and Gerstein G.L. (1991) Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp. Brain Res.* **84:** 505–516.

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

[45] Gray C.M., and McCormick D.A. (1996) Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* **274:** 109–113.

[46] Gur M., Beylin A., and Snodderly D.M. (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. *J. Neurosci.* **17:** 2914–2920.

[47] Gutkin B.S., and Ermentrout G.B. (1998) dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms and cortical spike train statistics. *Neural Comput.* **10:** 1047–1065.

[48] Hodgkin A.L., and Huxley A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**: 500–544.

[49] Holt G.R., Softky W.R., Koch C., and Douglas R.J. (1996) Comparison of

discharge variability *in vitro* and *in vivo* in cat visual cortex neurons. *J. Neurophysiol.* **75:** 1806–1814.

[50] König P., Engel A.K., and Singer W. (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* **19:** 130–137.

[51] Kreiter A.K., and Singer W. (1996) Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J. Neurosci.* **16:** 2381–2396.

[52] Leopold D.A., and Logothetis N.K. (1999) Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* **3:** 254–264

[53] Llinás R.R. (1988) The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. *Science* **242:** 1654–1664.

[54] Lüti A., and McCormick D.A. (1998) H-current: properties of a neuronal and network pacemaker. *Neuron* **21:** 9–12.

[55] MacLeod K., Bäcker A., and Laurent G. (1998) Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* **395:** 693–698.

[56] Mainen Z.F., and Sejnowski T.J. (1995) Reliability of spike timing in neocortical neurons. *Science* **268:** 1503–1506.

[57] Mehta M.R., Lee A.K., and Wilson M.A. (2002) Role of experience and oscillations in transforming a rate code into a temporal code *Nature* **417:** 741–746.

[58] Mel B.W. (1993) Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.* **70:** 1086–1101.

[59] Mel B.W. (1999) Why have dendrites? A computational perspective. In: *Dendrites,* Stuart G, Spruston N, Hausser M, eds., pp. 271–289. Oxford: Oxford University Press.

[60] Murthy V.N., and Fetz E.E. (1994) Effects of input synchrony on the firing rate of a three-conductance cortical neuron model. *Neural Comput.* **6:** 1111–1126.

[61] Nelson J.I., Salin P.A., Munk M.H.-J., Arzi M., and Bullier J. (1992) Spatial and temporal coherence in cortico-cortical connections: a cross-correlation study in areas 17 and 18 in the cat. *Vis. Neurosci.* **9:** 21–37.

[62] Perez-Orive J., Mazor O., Turner G.C., Cassenaer S., Wilson R.I., and Laurent G. (2002) Oscillations and sparsening of odor representations in the mushroom body. *Science* **297:** 359–65.

[63] Perkel D.H., Gerstein G.L., and Moore G.P. (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* **7:** 419–440.

[64] Poirazi P., and Mel B.W. (2001) Impact of active dendrites and structural plas-

ticity on the memory capacity of neural tissue. *Neuron* **29:** 779–796.

[65] Reynolds J.H., Pasternak T., and Desimone R. (2000) Attention increases sensitivity of V4 neurons. *Neuron* **26:** 703–714.

[66] Riehle A., Grün S., Diesmann M., and Aertsen A. (1997) Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278:** 1950–1953.

[67] Salinas E., and Sejnowski T.J. (2000) Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* **20:** 6193–6209.

[68] Salinas E., and Sejnowski T.J. (2001) Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* **2:** 539–550.

[69] Salinas E., and Sejnowski T.J. (2002) Integrate-and-fire neurons driven by correlated stochastic input. *Neural Comput.* **14:** 2111–2155.

[70] Seidemann E., Zohary U., and Newsome W.T. (1998) Temporal gating of neural signals during performance of a visual discrimination task. *Nature* **394:** 72–75.

[71] Singer W., and Gray C.M. (1995) Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18:** 555–586.

[72] Shadlen M.N., and Movshon J.A. (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* **24:** 67–77.

[73] Shadlen M.N., and Newsome W.T. (1994) Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4:** 569–579.

[74] Shadlen M.N., and Newsome W.T. (1998) The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *J. Neurosci.* **18:** 3870–3896.

[75] Shinomoto S., Sakai Y., and Funahashi S. (1999) The Ornstein-Uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. *Neural Comput.* **11:** 935–951.

[76] Softky W.R., and Koch C. (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13:** 334–350.

[77] Steinmetz P.N., Roy A., Fitzgerald P.J., Hsiao S.S., Johnson K.O., and Niebur E. (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404:** 187–190.

[78] Stevens C.F., and Zador A.M. (1998) Input synchrony and the irregular firing of cortical neurons. *Nature Neurosci.* **1:** 210–217.

[79] Stopfer M., Bhagavan S., Smith B.H., and Laurent G. (1997) Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Na-*

ture **390:** 70–74.

[80] Svirskis G., and Rinzel J. (2000) Influence of temporal correlation of synaptic input on the rate and variability of firing in neurons. *Biophys. J.* **79:** 629–637.

[81] Tiesinga P.H.E., José J.V., and Sejnowski T.J. (2000) Comparison of current-driven and conductance-driven neocortical model neuron with Hodgkin-Huxley voltage-gated channels. *Phys. Rev. E* **62:** 8413–8419.

[82] Troyer T.W., and Miller K.D. (1997) Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Comput.* **9:** 971–983.

[83] Tsodyks M.V., and Sejnowski T.J. (1995) Rapid state switching in balanced cortical network models. *Network* **6:** 111–124.

[84] Tuckwell H.C. (1988) *Introduction to Theoretical Neurobiology,* Volumes 1 and 2. New York: Cambridge University Press.

[85] Usher M., Stemmler M., Koch C., and Olami Z. (1994) Network amplification of local fluctuations causes high spike rate variability, fractal firing patterns and oscillatory local field potentials. *Neural Comput.* **6:** 795–836.

[86] Usrey W.M., and Reid R.C. (1999) Synchronous activity in the nervous system. *Annu. Rev. Neurosci.* **61:** 435–456.

[87] Van Kampen N.G. (1981) *Stochastic Processes in Physics and Chemistry.* Amsterdam: North Holland.

[88] van Rossum M.C.W., Turrigiano G.G, and Nelson S.B. (2002) Fast propagation of firing rates through layered networks of noisy neurons. *J. Neurosci.* **22:** 1956–1966.

[89] Van Vreeswijk C., and Sompolinsky H. (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274:** 1724–1726.

[90] Wehr M., and Laurent G. (1999) Relationship between afferent and central temporal patterns in the locust olfactory system. *J. Neurosci.* **19:** 381–390.

[91] White E.L. (1989) *Cortical Circuits.* Boston: Birkhäuser.

[92] Wilbur W.J., and Rinzel J. (1983) A theoretical basis for large coefficient of variation and bimodality in neuronal interspike interval distribution. *J. Theo. Biol.* **105:** 345–368.

[93] Williams S.R., and Stuart G.J. (2002) Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons. *Science* **295:** 1907–1910.

[94] Wilson M., and Bower J.M. (1992) Cortical oscillations and network interactions in a computer simulation of piriform cortex. *J. Neurophysiol.* **67:** 981–995.

[95] Zohary E., Shadlen M.N., and Newsome W.T. (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370:** 140–143.

Chapter 13

A Case Study of Population Coding: Stimulus Localisation in the Barrel Cortex

Rasmus S. Petersen,[1] **and Stefano Panzeri**[2]

[1]*Cognitive Neuroscience Sector, International School for Advanced Studies, Via Beirut 2/4, 34014 Trieste, Italy,* [2]*UMIST, Department of Optometry and Neuroscience, Manchester M6O 1QD, U.K.*

CONTENTS

13.1 Introduction

Animals are under strong selective pressure to react as quickly as possible to relevant stimuli. For example, flies can change their flight trajectory with a reaction time of about 30 ms. Similarly, primate pattern recognition can be so rapid (∼200ms) that the processing at each of the many stages that intervene between photoreceptors and motor neurons can take probably only ∼ 10 ms, before the results must be passed on to the next stage [37]. Since neurons rarely fire more than a single spike in such short periods of time, it seems that neuronal computation is capable of operating on the basis of very small numbers of action potentials.

In these situations, each stimulus must be encoded by few spikes per neuron. This observation is important because it strongly limits the complexity of the neuronal response and encourages us that it might be possible to get a thorough understanding of the neural code.

Cracking the neural code for a given stimulus set means understanding the manner in which the space of spike trains is divided into clusters representing the different stimuli. We would like to answer two basic questions: (1) Does this clustering depend on the precise timing of spikes, or only on the number of spikes that each neuron emits? (2) Does the clustering depend on correlations across multiple spikes, or only on individual spikes? To address these questions, it is necessary to quantify and compare the stimulus discriminability that the different candidate codes afford – the natural framework for which is Shannon's information theory.

Recently, the *series expansion* method of estimating mutual information has been developed, which is specifically tailored to the case of sparsely responding neurons [21, 22]. We have applied it to a popular model system – the somatosensory cortex of the rat – where the number of evoked spikes per stimulus is also small, and have thereby been able to study issues of spike timing and population coding in a rigorous manner [18, 25, 26]. These developments are reviewed in the following sections.

13.2 Series expansion method

13.2.1 Quantifying neuronal responses and stimuli

To measure the stimulus discriminability of a given neural code, the first step is to categorise the neuronal response in a manner that reflects the code. We consider spike times in the interval $[0,T]$ ms relative to the onset of the stimulus. For a *spike count* code, the response of a neuron on a given trial is simply the number of spikes emitted in the interval. To evaluate spike timing, the interval is subdivided into a sequence of L bins of width dt ms ($dt = T/L$). If each bin contains at most 1 spike, then the response on a given trial is the one of the 2^L possible sequences that occurs. It is not, however, necessary for the analyses described below that the sequence to be binary [28]. If a population of C neurons is being considered, the response on a given trial is the set of C simultaneously recorded spike counts or spike sequences. Having quantified the response, we can ask how well it discriminates between the different stimuli.

13.2.2 Mutual information and sampling bias

The mutual information quantifies how well an ideal observer of neuronal responses can, on average, discriminate which stimulus occurred based on a response observed on a single trial (Shannon, 1948):

$$I(S,R) = \langle \sum_n P(n|s) \log_2 \frac{P(n|s)}{P(n)} \rangle_s \tag{13.1}$$

Here $S = \{s\}$ is the set of stimuli, $R = \{n\}$ is the set of responses; n can be a single cell response, or a population response and either a spike count or a spike sequence. $P(n|s)$ is the posterior probability of a response n given stimulus s; $P(n)$ is the stimulus-average response probability and $P(s)$ is the prior stimulus probability.

Mutual information quantifies *diversity* in the set of probabilities $P(n|s)$. If these probabilities are all equal, for a given response n, and hence equal to $P(n)$, the argument of the logarithm is one and the response n contributes nothing to $I(S,R)$. Conversely, the more that the $P(n|s)$ differ, the greater the contribution of response n to the information.

In principle, to obtain an estimate of mutual information, we simply measure these probabilities from recorded data and substitute them into Equation (13.1). The problem is that it is difficult to estimate the conditional probabilities accurately, given the number of stimulus repetitions presented in a typical physiological experiment. The consequent fluctuations in the estimated conditional probabilities lead to spurious diversity that mimics the effect of genuine stimulus-coding responses. Hence, the effect of limited sampling is an *upward* bias in the estimate of the mutual information [38]. Since this sampling problem worsens as the number of response categories increases, the bias is intrinsically greater for timing codes compared to count codes. Hence it is important to exercise considerable care when assessing the information content of complex codes (many response categories) compared to simple ones (few response categories).

One way to approach this issue is to use bias correction procedures [14, 38]. The basis for these methods is the fact that, provided the number of trials is not too small, the information bias depends on the data in a surprisingly simple way. Essentially, the bias is proportional to the number of response categories divided by the number of trials. More precisely [23],

$$Bias = -\frac{1}{2N \ln 2}(R - 1 - \sum_s (R_s - 1)) \tag{13.2}$$

This expression depends only on the total number of trials N and on the number of *relevant* response categories (R, R_s). A given stimulus will evoke different responses with different probabilities: some will have high probability, some low probability, and others will have zero probability. A response is *relevant* if its probability, conditional to the stimulus, is non-zero. R_s is the number of responses that are relevant to stimulus s; R is the number of responses that are relevant considering all the stimuli together. In the simplest case, R is simply the total number of response categories (e.g., 2^L for a single neuron with L binary time bins). For a review of bias correction, including more sophisticated methods, see [29].

In principle, the expected bias of Equation (13.2) can be subtracted from the naive information estimate of Equation (13.1) to yield a bias-free estimate. In practice, of course, we need to know how many trials is *not too small*. Extensive experiments with simulated data indicate that the number of trials should be at least a factor of two times greater than the number of response categories [23, 30]. With 50 trials, for example, codes with up to about 25 response categories can typically be accurately evaluated. For a single neuron, this implies at most 4 time bins (each containing 0 or 1 spikes); for a pair of neurons, at most 2 time bins each.

Hence, even with bias correction, it is difficult to evaluate spike timing codes, and extremely difficult to evaluate spike timing population codes, using so-called *brute force* application of Equation (13.1). As a consequence, it is usually difficult to study neural coding in a truly systematic way. However, progress has recently been made in the important special case of neurons that fire small numbers of spikes; the theoretical foundation for which is described next.

13.2.3 Series expansion approach to information estimation

The variety of possible spike sequences, and hence the potential complexity of the neural code, increases rapidly with the number of spikes emitted per trial; conversely, low firing rates limit the complexity. Since typical firing rates in the barrel cortex are just 0–3 spikes per whisker deflection, the mutual information can be well-approximated by a second order power series expansion in the time window T, that depends only on PSTHs and pair-wise correlations between spikes at different times [18, 21, 33]. These quantities are far easier to estimate from limited experimental data than are the full conditional probabilities required by the direct method. Provided that (i) the number of spikes in the response window is < 1 (averaged over stimuli) and (ii) spikes are not locked to one another with infinite time precision, the mutual information can be approximated by a Taylor series expansion in the duration of the response window T [21, 22]. To second order:

$$I(S,R) = I_t + I_{tta} + I_{ttb} + I_{ttc} + O(T^3) \tag{13.3}$$

Here I_t is first order in T and I_{tta}, I_{ttb} and I_{ttc} are second order. The first order term depends only on the PSTH of each neuron; the second order term depend also on pairwise correlations. Provided that the approximation is accurate, mutual information can thus be estimated from knowledge of only first and second order statistics - it is not necessary to measure the full conditional probabilities demanded by Equation (13.1). This property makes the series expansion approach much less prone to sampling bias than the brute force approach. In practice, this means that coding can be studied to significantly better temporal precision than would otherwise be possible.

An important feature of the method is that the contribution of individual spikes (I_t and I_{tta}) is evaluated separately from that of correlated spike patterns (I_{ttb} and I_{ttc}). The amount of information that a neuronal population conveys by the timing of individual spikes is the sum of an *independent spike timing* term (I_t) and a PSTH *similarity* term (I_{tta}). The former term expresses the information that would be conveyed were the spikes to carry independent information; the latter term corrects this for any redundancy arising from similarity of PSTHs across stimuli. The remaining two terms (I_{ttb} and I_{ttc}) express any further effect that correlated spike patterns might have beyond that of individual spikes. By evaluating these terms separately, the series expansion yields direct insight into a question of great current interest in neural coding – the role of individual spikes compared to correlated spike patterns.

13.2.3.1 Independent spike timing

If within-trial spike patterns do not convey information, then all information must be in the timing of individual spikes. Under these circumstances, the time-varying firing rate (PSTH) is a complete description of the neuronal response, and is the only statistic required in order to estimate the information. The greater the diversity in PSTH structure across stimuli, the greater is the information. If each spike provides independent information about the stimulus set, Equation (13.4) (the independent spike timing term) gives the total information available in the response [4, 7, 21].

$$I_t = \sum_{a,i} \left\langle n_{\bar{ais}} \log_2 \frac{\bar{n}_{ais}}{\langle \bar{n}_{ais'} \rangle_{s'}} \right\rangle_s \tag{13.4}$$

n_{ais} is the number of spikes in time bin i of cell a elicited by stimulus s on a particular trial. The bar $\bar{\cdot}$ means an average over trials, thus $n_{\bar{ais}}$ is simply the corresponding PSTH. The angle brackets $\langle \cdots \rangle_s$ denote an average over stimuli, weighted by the stimulus probabilities $P(s)$.

13.2.3.2 PSTH Similarity

If there is any redundancy between spikes, Equation (13.4) can overestimate the information. Redundancy is present if the PSTH value at a given time bin correlates across the stimulus set with the PSTH value at a different time bin for the same cell, or correlates with the PSTH value at any time bin for a different cell. This type of correlation has been termed *signal correlation* and Equation (13.5) quantifies the amount of redundancy that it introduces. The PSTH similarity term is:

$$I_{tta} = \frac{1}{2\log_e 2} \sum_{a,b,i,j} \left[CS_{aibj} \left(1 - \log_e \frac{CS_{aibj}}{MS_{ai} MS_{bj}} \right) - ECS_{aibj} \right] \tag{13.5}$$

Here $MS_{ai} = \langle \bar{n}_{ais} \rangle_s$ is the average of the PSTH over stimuli for time bin i of cell a; $CS_{aibj} = \langle \bar{n}_{ais} \bar{n}_{bjs} \rangle_s$ is the signal correlation between time bin i of cell a and bin j of cell b; $ECS_{aibj} = MS_{ai} MS_{bj}$ is the expected value of CS_{aibj} for PSTHs that are uncorrelated across stimuli. I_{tta} is always negative or zero. I_{tta} and I_t together express any information that the population conveys purely by the timing of individual spikes (time-varying firing rate).

13.2.3.3 Stimulus-dependent spike patterns

A neuronal population can carry information by emitting patterns of spikes that *tag* each stimulus, without the differences in the patterns being expressed in the PSTHs [6, 11, 32, 39, 41]. When the assumptions of the series expansion are satisfied, the only types of spike pattern it is necessary to consider are spike pairs – higher order interactions can be neglected. Patterns involving pairs of spikes are quantified, for each stimulus, as the correlation between the number of spikes occurring in each of two time bins. For within-cell patterns, the bins come from the same cell; for cross-cell patterns, they come from different cells. This quantity is sometimes known as the noise correlation. In the case of cross-cell synchrony, for example, the noise correlation will be greater than expected from the PSTHs. In the case of within-cell refractoriness, where the presence of a spike in one bin predicts the absence of a spike in the next bin, the noise correlation will be less than that expected from the PSTH.

The amount of information conveyed by stimulus-dependent spike patterns depends, analogously to the PSTH information, on how much the noise correlations (normalised by firing rate) vary across the stimulus set: the greater the diversity, the greater the information available. This effect is quantified by Equation (13.6):

$$I_{ttc} = \frac{1}{2}\sum_{a,b,i,j}\langle CN_{aibjs}\log_2\left[\frac{CN_{aibjs}}{ECN_{aibjs}} \div \frac{\langle CN_{aibjs'}\rangle_{s'}}{\langle ECN_{aibjs'}\rangle_{s'}}\right]\rangle_s \tag{13.6}$$

CN_{aibjs} (noise correlation) is the *joint* PSTH of bin i of cell a and bin j of cell b given stimulus s. It is equal to $\overline{n_{ais}n_{bjs}}$, unless $a = b$ and $i = j$. In the latter case, CN_{aibjs} is equal to zero if each bin contains at most one spike, or equal to $\overline{n_{ais}^2} - \overline{n_{ais}}$ otherwise [28]. $ECN_{aibjs} = \bar{n}_{ais}\bar{n}_{bjs}$ is the expected value of CN_{aibjs} for statistically independent spikes. I_{ttc} is positive or zero.

13.2.3.4 Stimulus-independent spike patterns

Even if not stimulus-dependent, spike patterns can exert an effect on the neuronal code through a subtle interaction between signal correlation and noise correlation. In contrast to stimulus-dependent patterns, this less intuitive coding mechanism has received little attention in experimental work – it has been noted in theoretical papers by [1, 17, 22, 36]. As shown schematically in Figure 13.1, correlated noise can serve to sharpen the distinction in responses to different stimuli under certain circumstances, and to blur those distinctions under other circumstances. In general, this term – Equation (13.7) – is positive if signal correlations and noise correlations have different signs, negative if they have the same signs. If signals are uncorrelated, the term is zero.

$$I_{ttb} = -\frac{1}{2\ln 2}\sum_{a,b,i,j}\langle CN_{aibjs} - ECN_{aibjs}\rangle_s \ln\frac{CS_{aibj}}{ECS_{aibj}} \tag{13.7}$$

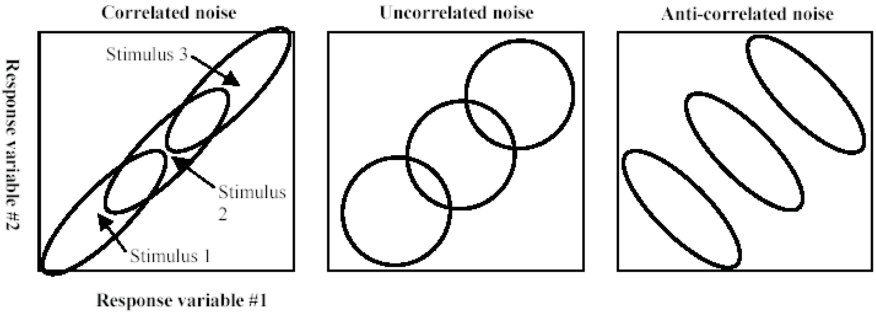

Figure 13.1

Effect of stimulus-independent patterns on neural coding. Each panel sketches hypothetical distributions of responses to three different stimuli. The response variables can be considered either to be different bins within the same cell or bins across different cells. Each ellipse indicates the set of responses elicited by a given stimulus. In each of these examples, signal correlations are positive whereas the sign of noise correlation differs. In the middle panel, noise correlation is zero, and stimulus-independent patterns exert no effect on the total information. When noise correlation is positive (left panel), responses to the stimuli are less discriminable and stimulus-independent spike patterns cause a redundant effect. When noise correlation is negative (right panel), responses are more discriminable and the contribution of stimulus-independent spike patterns is thus synergistic. In general, if signal and noise correlations have the same sign, the effect of stimulus-independent patterns is redundant, if they have opposite signs, it is synergistic. Reproduced with permission from [25].

13.2.4 Generalised series expansion

Is the breakdown of mutual information into individual spike terms and correlation-dependent terms a general property of neural encoders or a peculiarity of systems firing few spikes? Pola et al. [30] have recently investigated this issue, and have proved that the decomposition is completely general. The break-down of mutual information into I_t, I_{tta}, I_{ttb} and I_{ttc} terms generalises in a natural way to the exact case of Equation (13.1) Moreover, each of the terms in the exact breakdown has a very similar mathematical expression to their analogues in the second order series expansion. The main difference is that the exact components are expressed in terms of an interaction coefficient that takes all moments of the spiking response into account, not just pairwise correlations. Of course, the exact decomposition has the same sampling characteristics as Equation (13.1), so the second order series expansion is often more convenient to use in practice.

13.2.5 Sampling bias of the series expansion

In general, if C neurons are considered, each with L time bins, the total number of response categories is 2^{CL}. For example, with two cells and four (binary) time bins, there are 256 response categories, and, using bias correction, one would typically require at least 500 trials per stimulus with the brute force method in order to obtain meaningful information estimates. The reason for the improved sampling properties of the series expansion is that the number of free parameters is much less – it is the sum of the number of PSTH bins and the number of correlation bins. Instead of being exponential in C and L like the brute force method, it is only quadratic.

The bias of the series expansion has been investigated separately for the different terms I_t, I_{tta}, I_{ttb} and I_{ttc} [21]. Mathematical analysis revealed that the bias is of order T^2 for I_t and I_{ttc} but of order T^3 for I_{tta} and I_{ttb}. Therefore, essentially *all* the bias comes from the independent spike term and from the stimulus-dependent spike pattern term, and it is only these terms that need bias correction. I_t depends only on the PSTHs, which have CL parameters. I_{ttc} depends also on auto-correlations and cross-correlations. Cross-correlations have $L^2C(C-1)/2$ parameters. For binary bins, auto-correlations have $CL(L-1)/2$ parameters. For the multiple-spikes-per-bin case, auto-correlations have $CL(L+1)/2$ parameters. Hence it is I_{ttc} that contributes most to the bias – particularly its cross-correlational part. For a cell pair I_{ttc} has a total of $2L^2 - L$ parameters in the binary bin case, so that $L = 4$ can safely be analysed with 50 trials – a factor of 10 less than the brute force method.

We have applied the series expansion method extensively to the study of cortical coding in the whisker modality of the rat. The following section briefly introduces some essential background concerning this widely studied sensory system.

13.3 The whisker system

13.3.1 Whisking behaviour

The natural environment of the rat consists of dark, confined spaces where vision is of little use. Apart from smell and taste, the major sensory system of the rat is touch – particularly the whiskers. Each of these specialised hairs operates somewhat like the pick-up of a record player, converting object surface characteristics into mechanical vibrations that are then transduced by mechanoreceptors located in the whisker follicle. It is a sensitive system, endowing the rat with tactile acuity comparable to that of human fingertips [5]. The rat has about 30 large whiskers (macrovibrissae) on each side of the snout in addition to hundreds of smaller ones arranged arround and inside the mouth (microvibrissae). The macrovibrissae, which are about 10-40 mm long, are under active, muscular control. Rats typically explore objects of interest by sweeping their whiskers backwards and forwards across the object at 4-10Hz, a movement known as *whisking*.

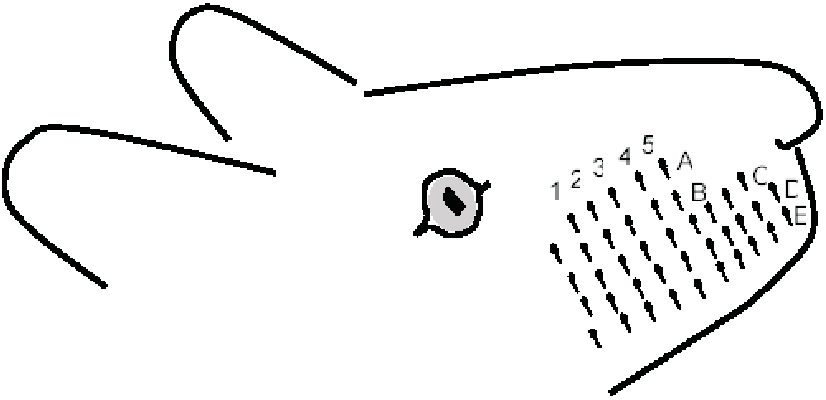

Figure 13.2
Arrangement of whiskers on the rat's snout. Rows, labelled by letters A-E and arcs, labelled by numbers 1-5 are indicated. The four most posterior whiskers that lie between the rows are labelled with the Greek letters: from top to bottom, α, β, γ, and δ.

The macrovibrissae are arranged in a highly consistent, grid-like pattern. Their locations are conventionally specified by *row* and *arc* coordinates. Rows are conventionally labelled by the letters A-E and arcs by numbers 1-5 (Figure 13.2).

13.3.2 Anatomy of the whisker system

Each macrovibrissa is innervated by about 250 mechanoreceptors, located on the base of the whisker follicle, inside the skin [42]. These respond to the mechanical deformation of a whisker, caused by contact with an external object. The cell bodies are located in the trigeminal ganglion, which lies at the base of the skull. The major projection of these cells is to the principal trigeminal nucleus, located in the brainstem, which projects in turn to the ventral posterior nucleus (VPM) of the thalamus. Both these structures are believed to contain on the order of a few hundred neurons per whisker. The main output of the VPM is to layer 4 of the primary somatosensory cortex (S1). The part of S1 that specifically represents the whiskers is known as the posterior medial barrel subfield (PMBSF). Like the primate visual system, there is a massive expansion in cell numbers at the level of the barrel cortex. Each whisker has a corresponding *barrel-column* of 300-400 micron diameter, containing about 10000 neurons. The number of cells per whisker is thus about 100 times greater than at subcortical levels, and may reflect a radical change in coding strategy.

Since barrels can be visualised histologically [44], it is possible to relate circuitry to sensory structure in a precise, quantitative manner; reviewed by Miller et al. [15].

This makes the whisker system a popular preparation for studying the operation of cortical columns.

13.3.3 Whisker physiology

Most neurons in the whisker system can be easily activated by delivering a rapid *tap* to the whiskers – but this response is largely invariant to the steady-state position of the whiskers. The standard experimental protocol for studying the system is to deliver low frequency step-like deflections to each individual whisker, often using a piezoelectric wafer. Such studies revealed that the anatomical mapping between whiskers and cortical barrel-columns has a simple physiological counterpart [35, 43]. Deflection of the whisker anatomically associated with a given barrel-column – the *principal whisker* (PW) – typically elicits a robust response. Deflection of the surrounding whiskers (SW) can often also evoke a response. Although individual neurons in a given barrel-column might respond well to particular SWs, different neurons prefer different SWs so that the average SW response is considerably weaker than the PW response. With standard, extracellular recording, the average response elicited by deflecting a neuron's PW is about 1 spike per trial – that for immediately neighbouring SWs about 0.5 spikes per trial, that for distant SWs even less [3]. Responses in barrel cortex are therefore low enough for the series expansion to be applicable.

In fact, these data probably reflect a subpopulation of particularly active cortical neurons. Recent studies using the whole cell patch method (where activity is not used as a criterion for selecting cells) have reported substantial numbers of neurons firing at rates 0.1 spikes per deflection or less under standard *in vivo* conditions (M. Brecht, personal communication).

13.4 Coding in the whisker system

13.4.1 Introduction

Classical whisker physiological has shown that the primary somatosensory cortex contains a topographic map for the location of a whisker on the rat's snout. Since deflection of a whisker evokes responses for neurons both in the topographically matching barrel-column as well as in surrounding barrel-columns [3, 24], whisker location seems to be a *population code*. Due to the relatively sparse nature of the neuronal response and to the modular organisation of the whisker system, this seemed to us a good opportunity for getting thorough insight into the nature of a cortical population code.

We asked two basic questions. (1) The role of spike timing: is whisker location coded simply by the number of spikes that occur over 100s of milliseconds (spike count coding), or, is the millisecond-precision spike timing crucial (spike time coding)? (2) The role of spike patterns: is the basic unit of whisker location coding the individual spike, or (synergistic) spike patterns?

13.4.2 Role of spike timing

We measured the neuronal response and its information content in terms of both spike count and spike timing, as described above. Vibrissae C1, C2, C3, D1, D2, D3, E1, E2 and E3 were stimulated one at a time in order to study how populations of cortical neurons encode stimulus location. The stimulus was an up-down step function of 80 cm amplitude and 100 ms duration, delivered once per second, 50 times for each vibrissa. Stimulus onset was defined as time = 0 ms. For single neurons, we discretised the response in the post-stimulus interval 0-40 ms into 5 ms bins; for neuron pairs, 10 ms bins.

Having defined the spike count response and the spike timing response, we compared the amount of mutual information that each of them conveys about whisker location, using the series expansion method described above.

Results averaged over 106 single units located in barrel-column D2 are shown in Figure 13.3 [18]. Early in the response (0-10 ms), the spike count provided almost as much information (90% on average) as spike timing. Later, however, there was a significant advantage for the timing code. Whereas spike timing information continued to increase gradually, spike count information saturated. Indeed, at longer time windows (data not shown), spike count information actually decreased. This occurred for two reasons. First, for some cells, the essential temporal structure of the PSTH could not be reduced simply to counting spikes. Second, since the evoked response was transient (< 50 ms), in long time windows, the spike count signal was degraded by *spontaneous* firing. At 40 ms, spike timing provided 44% more information than did total spike count.

What degree of precision underlies the spike timing code? To answer this question, we varied the resolution at which spike times were binned and computed the average information across stimuli as a function of bin size. If information increases as bin size is decreased, timing must be precise on the scale of the smaller bin size. Using a single, 20 ms bin, the average information across all 106 sampled neurons was 0.10 ± 0.006 bits (mean± SEM). Reducing bin size to 10 ms, the information present in the 0-20 ms interval increased to 0.14 ± 0.008 bits. Further reductions of bin size to 5 ms and finally to 2.5 ms, yielded additional increases in information to 0.146 ± 0.008 and 0.154 ± 0.008 bits respectively. Hence, the precision of the code was at least 5 ms.

Subsequently [25], we investigated whether these results generalise to the case of pairs of neurons, recorded either within the same barrel-column (D2-D2, 52 pairs) or from different barrel-columns (D1-D2, 80 pairs; D1-D3, 39 pairs; D2-D3, 93 pairs). For cell pairs in barrel-column D2, the average information in spike timing and spike count at 20 ms post-stimulus was similar, 0.27 ± 0.09 bits and 0.25 ± 0.08 bits, respectively. However, by 40 ms post-stimulus, D2 cell pairs conveyed 0.31 ± 0.10 bits by spike timing – 25% more than by spike count. The advantage of spike timing compared to spike count for cell pairs located in different barrel-columns tended to

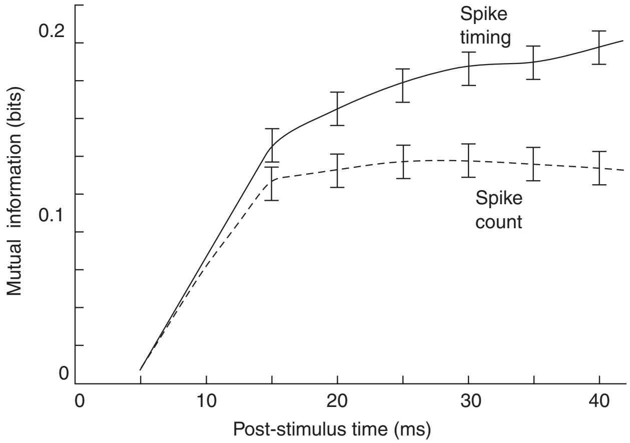

Figure 13.3

Coding by spike count vs. coding by spike timing. Mutual information plotted as a cumulative function of time, averaged over single neurons located in barrel-column D2. The bars denote SEM. Reproduced with permission from [18].

be greater: for D1-D2 pairs, the advantage was 29%; for D2-D3 pairs, 33%. For D1-D3 pairs, the advantage was 52%.

These results show that spike timing is important for the population coding of stimulus location. Spike timing is particularly informative for populations that encompass separate barrel-columns.

13.4.2.1 Coding of specific stimuli

The preceding section refers to information averaged over all nine stimulus sites. Another issue of interest concerns how well each particular whisker is represented in the spikes of a given barrel-column. To answer this question, we estimated the information that neurons convey about whether the stimulus site was, for example, *D2* or *not-D2* [18, 19]. For this analysis, all 8 non-D2 whiskers were considered to be in the same category. By computing this quantity selecting one whisker at a time, we obtained a *whisker-specific* information function. For neuron pairs located in barrel-column D2, the most reliably encoded whisker was the principal one, D2. At 40 ms, D2-specific information accounted for, on average, 65% of the total information about all 9 whiskers. Information specific to any given surround whisker was at least 6 times smaller than that to the principal one, on average.

13.4.2.2 Role of spike patterns

The information in spike timing could be generated in two ways. The simplest is if all the information were coded by stimulus-dependent differences in the timing of individual spikes; within-trial correlations between spike times not being informa-

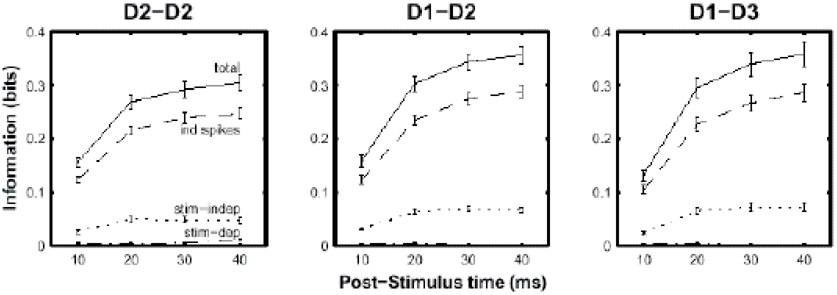

Figure 13.4

Role of spike patterns in population coding. Labels above the graphs refer to neuronal locations. Total information in spike timing (solid line) is compared to the contribution of each component in the series expansion, averaged over cell pairs. Bars denote SEM. Reproduced with permission from [25].

tive. In this case, information can only be coded by variations in the PSTH structure across stimuli. The second way is if particular spike patterns were to occur within the same trial, which could code information even in the absence of stimulus-dependent PSTH structure. As discussed above, these contributions are quantified separately by the series expansion, permitting us to compare their importance for cortical stimulus location coding.

Figure 13.4 shows how these different, timing-dependent components contributed to the coding of stimulus location. The left panel shows results averaged over all pairs of neurons located in barrel-column D2. At 40 ms post-stimulus, the timing of individual spikes (dashed line) accounted for $83 \pm 14\%$ of the total information in spike timing (solid line). Stimulus-dependent spike patterns (dash-dotted line) accounted for $5\pm7\%$, stimulus-independent patterns (dotted line) for $12\pm14\%$. Similar results were obtained for pairs of neurons located in different barrel-columns: D1-D2 pairs (middle panel) conveyed $17\pm6\%$ by spike patterns (stimulus-dependent and stimulus-independent patterns considered together), D2-D3 pairs (not shown) conveyed $15\pm7\%$, D1-D3 pairs (right panel) conveyed $18\pm6\%$. Thus spike patterns conveyed about 15-18% of the total information in the population spike train.

To probe the nature of the spike pattern contribution further, we split the information components into separate *within-cell* and *cross-cell* parts. The major finding was that *within-cell* spike patterns gave a significant positive contribution to the information in the population code (0.07 ± 0.04 bits at 40 ms post-stimulus), and this contribution was a stimulus-*independent* one (I_{ttb}). In addition, there was a very small, positive contribution from stimulus-dependent patterns across cells (0.007 ± 0.02 bits) and a small negative effect of stimulus-independent patterns across cells (-0.02 ±0.02 bits). These results were robust to changes in both the time window and the bin size [27]. Similar results were also obtained for neurons located in different barrel-columns.

Overall, neither within nor across barrel-columns, did cross-cell spike patterns seem to code information about stimulus location: the net contribution of spike patterns to the population code was almost entirely attributable to within-cell patterns.

What is the nature of these informative, within-cell patterns? We found that the PSTHs of neurons in the same barrel-column tend to be similar across different stimuli (positive signal correlation) and that the trial-by-trial variability is negative (negative noise correlation). Hence, as discussed in more detail in [25], there was a positive contribution to the information from the stimulus-independent term I_{ttb}.

13.4.2.3 Role of the first spike

The previous sections showed that the coding of stimulus location is achieved mainly by the timing of individual spikes. We asked whether it is possible to further specify the nature of the code: is a similar quantity of information transmitted by any single spike or, alternatively, is a particular subset of individual spikes crucial? We repeated the above analyses considering only the first, second, or third spikes per cell recorded on each stimulus trial. The information conveyed by the individual spike terms of the series expansion – Equations (13.4) and (13.5) – was compared to the corresponding data for the whole spike train [18, 25]. For both single neurons and neuron pairs in barrel-column D2, the first spikes conveyed almost as much information as the entire spike train. For cell pairs, the mean first spike information was 91±7% of that in the entire 40 ms spike trains. For neurons in different barrel-columns, the corresponding values were 87±7% (D1-D2 pairs), 91±9% (D2-D3 pairs) and 89±9% (D1-D3 pairs). The mean information conveyed by D2-D2 pairs in the second and third spikes was 43±18% and 18±14%, respectively, of that present in the individual spikes of the whole spike train. Similar results for second and third spikes were obtained for cell pairs distributed across different barrel-columns. Since nearly all the information in the entire spike train was already present in the first post-stimulus spike, the later spikes were almost completely redundant, both for neuron pairs within and across barrel-columns.

All the observations reviewed above point toward the conclusion that, to a large extent, the barrel cortex population code for stimulus location consists of the time of individual cells' first spike after whisker deflection. We characterise this as a simple, spike-time population code. Under our experimental conditions, the basic functional unit of barrel cortex for stimulus localisation seems to be the single neuron rather than the neuronal ensemble.

13.4.2.4 Pooling

Given that barrel cortex employs population coding of stimulus location, any brain area that wants to make use of its output is faced with a significant *decoding* problem. Consider a downstream target neuron receiving afferent inputs from a set of neurons with widely differing tuning properties. In order to make sense of this stream of signals, the target neuron faces the daunting task of keeping track of which neuron caused each distinct post-synaptic potential. We term this most general situation *labelled line* decoding.

A simple, alternative decoding strategy has been proposed. The idea of *pooling* is that the target neuron simply sums up all afferent inputs, regardless of their origin. The pooling hypothesis can be tested by quantifying the amount of information that is lost in a pooled representation, compared to that available in the full (labelled line) population response.

We used the generalised, multiple-spikes series expansion [28] to quantify the mutual information between the responses of cell pairs and whisker stimulus location [19]. For this analysis, we considered the time window 0-40 ms post-stimulus, divided into 10 ms bins. Neurons were located either within the same barrel-column (e.g., D2-D2) or across two different ones (e.g., D1-D2). For D2-D2 pairs, pooling caused an average information loss of 5%, depending on the precise time window. In contrast, for cross-columnar pairs, the information loss was 25–39%. Thus pooling causes markedly less information within a column compared to across columns.

Is the pooling information loss attributable to particular elements of the stimulus set, or due to an overall degradation? We addressed this by estimating the information specific to each of the 9 whiskers. The result was clear. For cell pairs located in D2 barrel-column, pooling did not cause any loss of information about whether or not the principal whisker was stimulated (1%, not significant). The loss was much higher for surround whiskers (up to 32%). In contrast, for D1-D2 pairs, pooling caused 42% loss specific to whisker D1 and 31% loss specific to whisker D2. For D1-D3 pairs, there was 54% loss specific to D3. Thus, information about the principal whisker was fully preserved only by within–column pooling, not by cross–columnar pooling.

These results suggest that, despite the enormous potential complexity of the cortical code, the information carried about the location of a stimulated whisker may be read–off in a highly efficient manner by a simple mechanism – pooling the afferent activity of neurons with similar sensory tuning. We speculate that the anatomical substrate supporting this decoding is the cortical column.

13.5 Discussion

We briefly discuss two issues arising from these results: (1) whether such a *first spike time code* can be utilised by the rest of the brain and (2) the relation to other work on the role of correlations in population coding.

13.5.1 Decoding first spike times

First spike time also has an important role in other sensory systems. The first spike time of cortical responses encodes visual contrast [10, 31] and sound source location [9]. Although information is thus *available* in first spike times, it is not always safe to assume that such information can be *used* by the rest of the animal's brain. Unlike the experimenter, an animal likely does not have independent knowledge of stimulus time. There are two potential solutions to this problem. (1) Since the collection of vibrissal sensory data under natural conditions is an active process initiated by a motor command, the sensory system could use the output from the motor system as an estimate of stimulus time. However, this motor efference signal acting alone would probably not possess sufficient temporal precision to permit the representation of information by first spike times. (2) The sensory system could use the *relative* timing between spikes in the neuronal population [12, 40]. When whisker D2 is deflected, many neurons within its barrel-column fire spikes within a few milliseconds of each other. Thus a simple way that the rat might decode the deflection of this whisker would be to detect the occurrence of large-scale simultaneous firing within the barrel-column. We showed that this *barrel activation* algorithm can lead to very good discrimination of whether or not a given whisker was deflected, based purely on the relative spike times of neurons within the column (Panzeri, Petroni, Diamond & Petersen, in preparation). The critical feature of the decoding algorithm is that the simultaneous activity characteristic of whisker deflection can only be detected if spike times are registered to high temporal precision. Thus, in order to decode the stimulus in the absence of information about when the stimulus occurred, precise spike timing is a crucial aspect of the neural code. Spike count codes do not permit simultaneous firing events to be accurately identified and hence are extremely uninformative. The advantage for spike timing codes over spike count codes is even more marked than in the previous analyses that assumed knowledge of stimulus onset.

13.5.2 Role of cross-correlations in population codes

There have been a number of previous reports that correlated spike patterns across different neurons play an important part in neural population coding [2, 6, 11, 32, 39, 41]. The strategy in all these studies was to demonstrate the existence of some stimulus-linked cross-correlation structure that could not be accounted for by the null hypothesis of independent firing. In the simplest case, the cross-correlogram was shown to differ significantly from its shift predictor [11]. In other cases, more sophisticated statistics were deemed necessary (e.g., [13]), but the logic was similar.

The importance of these studies is that they showed that cross-correlated spike patterns might play a role in neural coding. However, they did not *quantify* the information in spike patterns and compare it to that available in individual spikes. This is important, since comparison of cross-correlogram to its shift predictor shows neither how much the cross-correlations contribute to coding, nor whether any contribution is additive or redundant with the individual spike contribution. The simplest way that one might seek to quantify the role of spike patterns is to estimate the information conveyed by a given neuronal population and to repeat the calculation with trial-shuffled responses. If shuffling significantly reduces the information, cross-correlated spike patterns play an important part in the population code. The problem with this method is that the converse result, where shuffling has no effect, is ambiguous.

The series expansion framework is helpful for clarifying the ambiguity. The effect of shuffling is to set all correlations equal to the values expected from statistical independence (apart from fluctuations due to finite sampling). Since the two spike-pattern terms I_{ttb} and I_{ttc} essentially quantify how far the correlations are from being independent, shuffling will make them close to zero (after bias correction). However, since I_{ttb} can be positive (synergistic) or negative (redundant), the lack of any effect of shuffling could be due to a negative I_{ttb} contribution cancelling out a positive I_{ttc} contribution.

The series expansion addresses the problem of shuffling ambiguity by quantifying stimulus-dependent and stimulus-independent correlation contributions separately. Recently another approach has been proposed by Nirenberg, Latham and their colleagues [16] which casts spike train analysis in a decoding framework. Comparison of this approach with the series expansion substantially clarifies the physical significance of I_{ttb} and I_{ttc}. The task of decoding is, given the neural activity n evoked on a particular trial, to guess which stimulus caused it. In the *maximum likelihood* framework, you choose the most likely stimulus – that is, the s that maximises $P(s|n)$. By Bayes rule,

$$P(s|n) = \frac{P(n|s)P(s)}{P(n)} \tag{13.8}$$

The heart of the decoder is thus an internal model of what responses are caused by each of the possible stimuli $P(n|s)$. In general, such a model should take into account all possible statistical dependencies between the spike trains of different neurons, and will be extremely complex. Indeed, since the brain is faced with a similar sampling problem during learning to the one the experimenter faces when analysing physiological data, it may be that the ability of neural systems to use general $P(n|s)$ models is rather limited.

Nirenberg et al. [16] asked how accurately stimuli can be decoded by a model that ignores correlations; that is:

$$P(n|s) = \prod_c P(n_c|s) \tag{13.9}$$

where n_c is the response of cell c. They showed that the information loss, compared to the most general, correlation-dependent decoder, can be expressed by a term that they call ΔI. This information theoretic term measures the average increase in the length of the binary digit code needed to characterize the stimulus given the response when Equation (13.9) is used to approximate the true $P(n|s)$. It quantifies the information cost of neglecting correlated activity in decoding. The authors applied this method to the responses of pairs of retinal ganglion cells recorded from the isolated mouse retina during presentation of movies of natural scenes. The result was that ΔI is typically only 2–3% of the total information in the neuronal responses.

The result shows that the population activity can be decoded very accurately even by a relatively simple model that has no knowledge of the cross-correlated activity that different stimuli might evoke. Note, however, that although the model ignores correlations, the spike trains that the neuronal population is producing may still exhibit correlations. Although the decoder can do nothing to change these correlations (on-line trial-shuffling would be necessary), they might still have a significant influence on the neural code. This is the essential difference between the series expansion method and the ΔI method. In fact, we can characterise correlations as being *decoder-sensitive* and *decoder-insensitive*. ΔI focuses on the decoder-sensitive type, the series expansion considers both.

How do decoder-sensitive and -insensitive correlations relate to the spike pattern components I_{ttb} and I_{ttc} of the series expansion? The answer is remarkably simple. It turns out that ΔI, the decoder-sensitive contribution of cross-correlations to the population code, is precisely equal to the (cross-correlational) stimulus-dependent term of the series expansion I_{ttc} [20, 30]. This is true also of the general, exact decomposition. In other words, one can think of the stimulus-dependent spike pattern term as expressing a decodable spike pattern effect. In contrast, the stimulus-independent term of the series expansion I_{ttb} quantifies the effect of the intrinsic correlations in the spike train, which cannot affect the performance of the decoder.

Finally, we note a second difference between ΔI and the series expansion; namely that, as defined in [16], ΔI considers only correlations between different neurons, whereas the series expansion assesses also the role of spike correlations within neurons. However, the difference is minor, since the ΔI formalism could easily be extended to the case of either within-cell correlations alone, or within-cell and cross-cell correlations together.

13.6 Conclusions

Although single trial discriminability (mutual information) is widely agreed to be the right framework for addressing neural coding, information theory has been applied mainly to *single* neuron coding, and little to the more general case of population coding. This is due to the problem of limited sampling. In this chapter, we have argued that, for the class of sparsely responding neuronal ensembles, the series expansion approach to information estimation allows population coding to be studied in a rigorous, comprehensive manner. For rat barrel cortex, this method has revealed that there is a temporal code of a simple kind: about 85% of the total information available in the spike trains of neuron pairs concerning whisker location can be attributed to the timing of individual spikes. Moreover, about 90% of this information is captured by the first post-stimulus spike fired by each neuron.

Acknowledgments: Supported by European Community IST-2000-28127, Telethon Foundation GGP02459, J.S. McDonnell Foundation 20002035, Wellcome Trust grant 066372 /Z /01 /Z, Italian MURST, Consiglio Nazionale delle Ricerche 02. 00536. ST97 and Regione Friuli Venezia Giulia. S.P. is supported by an MRC Research Fellowship.

References

[1] Abbott, L.F. and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11**, 91-101.

[2] Abeles, M., Bergman, H., Margalit, E., and Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* **70**, 1629-1638.

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

[4] Brenner, N., Strong, S.P., Koberle, R., Bialek, W., and de Ruyter van Steveninck R.R. (2000). Synergy in a neural code. *Neural Comput.* **12**, 1531-1552.

[5] Carvell, G.E., and Simons, D.J. (1990). Biometric analyses of vibrissal tactile discrimination in the rat. *J. Neurosci.* **10**, 2638-2648.

[6] deCharms, R.C. (1998). Information coding in the cortex by independent or coordinated populations. *Proc. Natl. Acad. Sci. USA* **95**, 15166-15168.

[7] DeWeese, M.R. (1996). Optimization principles for the neural code. *Network* **7**, 325-331.

[8] Furukawa, S., Xu, L., and Middlebrooks, J.C. (2000). Coding of sound-source location by ensembles of cortical neurons. *J. Neurosci.* **20**, 1216-1228.

[9] Furakawa S., Xu L., and Middlebrooks J.C. (2002) Coding of sound-source location by ensembles of cortical neurons. *J. Neurosci.* **20**, 1216-1228.

[10] Gawne, T.J., Kjaer, T.W., Hertz, J.A., and Richmond, B.J. (1996). Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb. Cortex* **6**, 482-489.

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

[12] Jenison, R.L. (2001). Decoding first spike latency: A likelihood approach. *Neurocomputing* **38-40**, 239-248.

[13] Konig, P. (1994). A method for the quantification of synchrony and oscillatory properties of neuronal activity. *J. Neurosci. Methods* **54**, 31-37.

[14] Miller, G.A. (1955). Note on the bias on information estimates. *Information Theory in Psychology: Problems and Methods II-B*, 99-100.

[15] Miller, K.D., Pinto, D.J., and Simons, D.J. (2001). Processing in layer 4 of the neocortical circuit: new insights from visual and somatosensory cortex. *Curr. Opin. Neurobiol.* **11**, 488-497.

[16] Nirenberg, S., Carcieri, S.M., Jacobs, A.L., and Latham, P.E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature* **411**, 698-701.

[17] Oram, M.W., Foldiak, P., Perrett, D.I., and Sengpiel, F. (1998). The 'Ideal Homunculus': decoding neural population signals. *Trends Neurosci.* **21**, 259-265.

[18] Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., and Diamond, M.E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* **29**, 769-777.

[19] Panzeri, S., Petroni, F., Petersen, R.S., and Diamond, M.E. (2003). Decoding neuronal population activity in rat somatosensory cortex: role of columnar organization. *Cereb. Cortex.* **13**, 45-52.

[20] Panzeri, S., Pola, G., Petroni, F., Young, M.P., and Petersen, R.S. (2002). A critical assessment of the information carried by correlated neuronal firing. *Biosystems*, **67**, 187-193.

[21] Panzeri, S. and Schultz, S. (2001). A unified approach to the study of temporal, correlational and rate coding. *Neural Comput.* **13**, 1311-1349.

[22] Panzeri, S., Schultz, S.R., Treves, A., and Rolls, E.T. (1999). Correlations and the encoding of information in the nervous system. *Proc. R. Soc. Lond. B Biol. Sci.* **266**, 1001-1012.

[23] Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network* **7**, 87-107.

[24] Petersen, R.S. and Diamond, M.E. (2000). Spatio-temporal distribution of whisker-evoked activity in rat somatosensory cortex and the coding of stimulus location. *J. Neurosci.* **20**, 6135-6143.

[25] Petersen, R.S., Panzeri, S., and Diamond, M.E. (2001). Population coding of stimulus location in rat somatosensory cortex. *Neuron* **32**, 503-514.

[26] Petersen, R.S., Panzeri, S., and Diamond, M.E. (2002a). Population coding in somatosensory cortex. *Curr. Opin. Neurobiol.* In press.

[27] Petersen, R.S., Panzeri, S., and Diamond, M.E. (2002b). The role of individual spikes and spike patterns in population coding of stimulus location in rat somatosensory cortex. *Biosystems*. In press.

[28] Petroni, F. (2002). A computational analysis of the functional role of timing of action potentials in the cerebral cortex. Ph.D. Thesis. University of Newcastle upon Tyne. In preparation.

[29] Pola, G., Schultz, S., Petersen, R.S., and Panzeri, S. (2002). A practical guide to information analysis of spike trains. *In Tools for Neurscience Databases*, R. Kotter, ed. Kluwer Academic Press.

[30] Pola, G., Thiele, A., Hoffman, K.-P., and Panzeri, S (2003). An exact method to quantify the information transmitted by different mechanisms of correla-

tional coding. *Network,* **14**, 35-60.

[31] Reich, D.S., Mechler, F., and Victor, J.D. (2001). Temporal coding of contrast in primary visual cortex: when, what, and why. *J. Neurophysiol.* **85**, 1039-1050.

[32] Riehle, A., Grun, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950-1953.

[33] Schultz, S.R. and Panzeri, S. (2001). Temporal correlations and neural spike train entropy. *Phys. Rev. Lett.*

[34] Shannon, C. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.* **27**, 379-423.

[35] Simons, D.J. (1978). Response properties of vibrissa units in rat SI somatosensory neocortex. *J. Neurophysiol.* **41**, 798-820.

[36] Snippe, H.P. (1996). Parameter extraction from population codes: a critical assessment. *Neural. Comput.* **8**, 511-529.

[37] Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* **381**, 520-522.

[38] Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Comput.* **7**, 399-407.

[39] Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515-518.

[40] Van Rullen, R. and Thorpe, S.J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* **13**, 1255-1283.

[41] Villa, A.E., Tetko, I.V., Hyland, B., and Najem, A. (1999). Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. *Proc. Natl. Acad. Sci. USA* **96**, 1106-1111.

[42] Welker E. and Van der Loos H. (1986). Quantitative correlation between barrel-field size and the sensory innervation of the whiskerpad: a comparative study in six strains of mice bred for different patterns of mystacial vibrissae. *J. Neurosci.* **6**, 3355-3373.

[43] Welker, C. (1971). Microelectrode delineation of fine grain somatotopic organization of (SmI) cerebral neocortex in albino rat. *Brain Res.* **26**, 259-275.

[44] Woolsey, T.A. and Van der Loos, H. (1970). The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. *Brain Res.* **17**, 205-242.

Chapter 14

Modelling Fly Motion Vision

Alexander Borst
Max-Planck-Institute of Neurobiology, Department of Systems and Computational Neurobiology, Am Klopferspitz 18a D-82152 Martinsried, Germany

CONTENTS

14.1 The fly motion vision system: an overview

Whenever an animal is moving in its environment, moves its eyes or an object moves in front of its eyes, the visual system is confronted with motion. However, this motion information is not explicitly represented in the two-dimensional brightness pattern of the retinal image. Instead, motion has to be computed from the temporal brightness changes in the retinal image. This is one of the first and most basic processing steps performed by the visual system. This primary process of motion detection has become a key issue in computational neuroscience, because it represents a neural computation well described at the algorithmic level that has not been understood at the cellular level in any species so far, yet simple enough to be optimistic in this respect for the future. The development of models of motion detection has been experimentally driven in particular by investigations on two systems, the rabbit retina [1] and the insect visual system [49]. Vice versa, there is probably no other field in system neuroscience where experiments were more influenced by theory than in the study of motion vision. Motion vision has thus become a classical problem in computational neuroscience, which many laboratories around the world have embarked on. In the following I will give an overview of what is known about the computations underlying motion vision in the fly, where a lot of experimental results are available (for review see: [13, 36]) and where modelling efforts have reached a rather detailed biophysical level at many processing steps.

This chapter summarizes our current understanding of fly motion vision with an emphasis on modelling rather than on the large set of available experimental data. After giving an overview of the fly motion vision system, the next part of the chapter introduces the correlation-type of motion detector, a model for local motion detection that has been successfully applied to explain many features of motion vision, not only in flies but also in higher vertebrates including man. This is followed by an outline of how local motion signals become spatially processed by large-field neurons of the lobula plate in order to extract meaningful signals for visual course control. In a final section, the article will discuss in what directions current research efforts are pointing to fill in the missing pieces.

The processing of visual motion starts in the eye. In flies, like in most invertebrates, this structure is built from many single elements called facets or ommatidia. Each ommatidium possesses its own little lens and its own set of photoreceptors. The latter send their axons into a part of the brain exclusively devoted to image processing called the *visual ganglia*. Within these ganglia, images become processed by an array of local motion detectors (Figure 14.1, top). Such motion detectors are thought to exist for horizontal as well as for vertical image motion and to cover the whole visual field of the animal.

In the next processing step the output of such local motion detectors become spatially integrated by various large field elements. Anatomically, this happens on the dendrites of tangential cells located in the posterior part of the third visual ganglion, called the *lobula plate*. There exists a limited set of such tangential cells that can be grouped according to their preferred direction of image motion (Figure 14.1, bottom): some of the cells respond preferentially to horizontal image motion from front to back (e.g., the three HS-cells, i.e., HSN, HSE and HSS, both CH-cells, i.e., dCH and vCH), others to horizontal image motion in the opposite direction (H1 and H2), others respond selectively to vertical image motion from top to bottom (the VS-cells VS1, VS2, VS3) etc. All in all, there are only about 60 such neurons per hemisphere in the blowfly *Calliphora* that collectively cover the whole visual field of the animal.

However, these neurons do not integrate the output signals of local motion detectors independently but interact with each other. Specific connections have been determined between tangential neurons of the left and the right lobula plate as well as between neurons within one lobula plate (Figure 14.1, colored lines). These connections tune many tangential cells responsive to specific motion signals in front of both eyes, and others that are selectively responsive to motion of small moving objects or relative motion. Tangential cells have been shown to synapse onto descending neurons (e.g., [86]) which connect either to the flight motor in the thoracic ganglion of the animals controlling the various flight maneuvers, or to specific neck muscles controlling head movements (not shown).

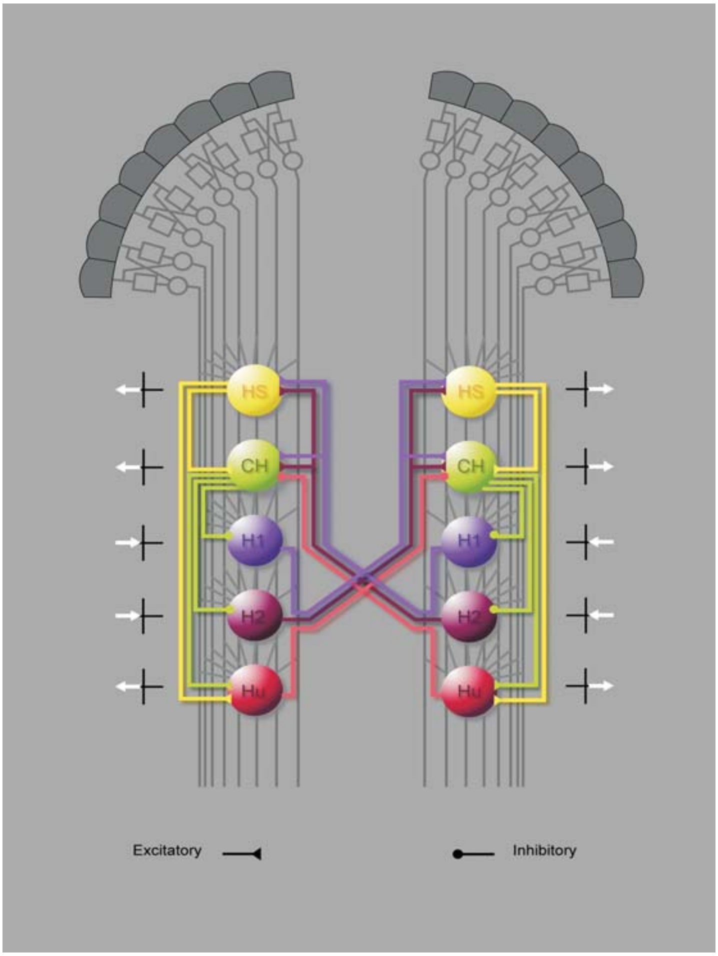

Figure 14.1

Processing of visual motion information in the fly visual system. As a first step, visual motion information is processed by a retinotopic array of local motion detectors. Their output signals are spatially integrated in parallel by tangential cells of the lobula plate. The central circuit diagram shows the connectivity between several tangential cells sensitive to horizontal image motion of the lobula plates on both sides of the brain. Excitatory and inhibitory connections are displayed as triangles and circles, respectively. White arrows indicate preferred directions of each cell group for visual motion on its ipsilateral side. Graded potential neurons making ipsilateral connections only (HS, CH) are shown in bright colors. Spiking neurons connecting to the contralateral lobula plate (H1, H2, Hu) are shown in dark colors. (See color insert.)

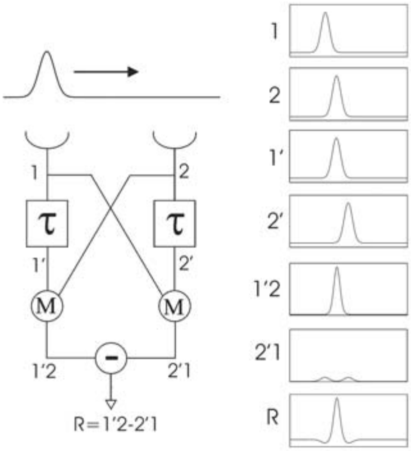

Figure 14.2

Minimal circuit diagram of a correlation detector. It consists of two subunits. In each subunit, the retinal signals from two neighboring locations are multiplied with each other (M), after one or both of them have been fed through a temporal filter with a time constant τ. This operation is done twice in a mirror-symmetrical way in both subunits. The output signals of both subunits are finally subtracted.

14.2 Mechanisms of local motion detection: the correlation detector

The process of local motion detection has been successfully described by the so-called correlation-type of motion detector or, in brief, correlation detector. This model has been first proposed on the basis of experimental studies on optomotor behavior of insects [49, 72, 73, 74, 75]. In subsequent studies, the correlation detector has also been applied to explain motion detection in different vertebrate species including man (e.g., [34, 35, 87, 88, 89, 90, 96], for review, see [6, 8]).

In its most parsimonious form, such a correlation detector consists of two mirror-symmetrical subunits (Figure 14.2, left). In each subunit, the signals derived from two neighboring inputs are multiplied with each other after one of them has been shifted in time with respect to the other by a delay line or a temporal low-pass filter. The final detector response is given by the difference of the output signals of both subunits. The combination of a temporal delay and a multiplication is the reason why this type of detector measures the degree of coincidence of the signals in its input channels or, in other words, performs on average a spatio-temporal cross-correlation. The basic operations of the correlation detector are summarized on the right side of Figure 14.2. Here it is assumed, for simplicity, that the brightness distribution of the retinal image is not filtered spatially or temporally but directly feeds the movement detector. If an object passes the detector from left to right, the left input channel (1) is activated first, and then some time later the right input channel is activated (2). This time interval depends on the velocity of the object and the spatial distance between both input channels, i.e., the sampling base of the detector. By delaying the left input signal (1′), the time-shift between the two input signals is reduced, and ideally accurately compensated for. Then, both signals arrive simultaneously at the multiplication stage of the left subunit resulting in a large response (1′2). For the right subunit, the temporal filter increases the time-shift between the input signals. This leads to a comparatively small detector response (2′1). After subtracting the output signals of both subunits, the final output response (R) is obtained. In the example shown here, the object motion from left to right is called the *preferred direction* of the motion detector. Motion in the opposite direction will result in a sign-inverted, i.e., negative response. This is called the *anti-preferred* or *null direction* of the detector.

14.2.1 Steady-state response properties

In the study of biological motion vision, periodic stimuli have played a dominant role. Therefore, the response of a correlation detector shall now be calculated for periodic sine gratings moving at a constant velocity. We consider a sine-wave of wavelength λ and contrast ΔI travelling at a velocity $v[^o/s]$. The inputs are spaced by $\Delta\phi$, the temporal filter has an amplitude and phase response denoted A and Φ, respectively. Thus, the signals entering the detector at the left and right input are both sinusoids with a DC value corresponding to the mean luminance, an amplitude ΔI and a temporal frequency $\omega = 2\pi v/\lambda$ that are phase shifted by $\Delta\phi/\lambda$ with respect to each other. At the output of the low-pass filters, these sinusoids, once in steady state, simply have an additional amplitude factor $A(\omega)$ and an additional phase shift $\Phi(\omega)$. Multiplication of the respective signals and subtracting the result of the left and right multiplier leads to the following expression of the time-averaged detector response [4, 19]:

$$R = \Delta I^2 A(2\pi v/\lambda)\sin(-\Phi(2\pi v/\lambda))\sin(2\pi\Delta\phi/\lambda) \tag{14.1}$$

If we assume the temporal filter to be a low-pass of 1st order, we obtain:

$$R = \Delta I^2 \frac{\tau 2\pi v/\lambda}{1+\tau^2(2\pi v/\lambda)^2} \cdot \sin(2\pi\Delta\phi/\lambda) \tag{14.2}$$

The steady-state response of a motion detector can thus be seen to depend on many internal and stimulus parameters in a non-trivial way. These dependencies will now be discussed in detail.

14.2.1.1 Velocity tuning

The dependence of the detector response on the velocity of the moving grating is shown in Figure 14.3 for three different time-constants. First of all, we note that the detector response, unlike a speedometer, does not increase linearly with increasing pattern velocity. Rather, it exhibits an optimum at a certain velocity. For velocities

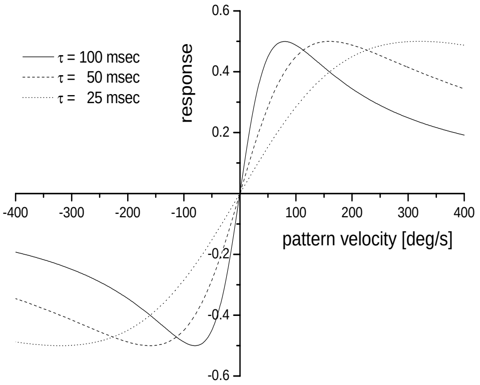

Figure 14.3

Steady-state responses of the minimal correlation model shown in Figure 14.2 as a function of pattern velocity for three different low-pass time-constants.

beyond that optimal velocity the response falls off gradually towards zero. Furthermore, the shape of the response curve depends on the time-constant of the detector low-pass filter: the larger the time-constant, the smaller the velocity at which the response is maximum.

The time-constant of the detector filter thus becomes a decisive model parameter that sets the operating range of the motion detection system in the velocity domain. In particular, it determines the slope of the detector response around velocity zero and, thus, the gain of the sensor.

14.2.1.2 Pattern dependence

Another peculiar response characteristic of the correlation detector is its dependence on pattern properties such as its contrast and spatial wavelength. As can be seen in Equations (14.1) and (14.2), the response depends on the square of the pattern contrast. This has indeed been experimentally confirmed using low stimulus contrasts [57]. For higher contrasts, the experimental data fail to follow a quadratic contrast dependence. Rather, the response of fly motion sensitive neurons as well as the strength of the optomotor following behavior saturates for contrasts higher than about 50% [27]. This might be explained by adaptive changes of internal gain factors in the detection system [14, 15, 31, 51, 67].

Another feature of correlation detectors, which is obvious in the above formulas, is that the steady-state response is proportional to the sine of the ratio of the sampling base and the pattern wavelength. The sampling base of the motion detector, i.e., the spatial separation of its input lines, limits the spatial resolution of the system. As is the case for any discrete sampling system, wavelengths can only be resolved up to a certain limit. The smallest wavelength that can be resolved is given by twice the

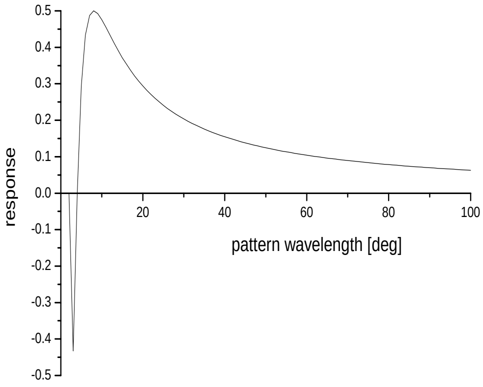

Figure 14.4

Steady-state responses of the minimal correlation model shown in Figure 14.2 as a function of pattern wavelength assuming a sampling base of 2 deg.

sampling interval (the *nymquist limit*). For a sampling base of 2 deg. of visual angle, this dependence on the pattern wavelength is illustrated in Figure 14.4.

As can be derived from Equations (14.1) and (14.2), the response is maximum if the ratio of sampling base and pattern wavelength equals $1/4$, in our example at a wavelength of $\lambda = 8$ deg. For larger wavelengths, the response gradually declines towards zero. For smaller wavelengths, the response starts oscillating between positive and negative values: it is zero for $\lambda = 4$ deg and negative between $\lambda = 4$ and $\lambda = 2$ deg. Here, the correlation detector signals negative values for pattern motion along its preferred direction. This behavior is due to the sampling theorem (see above) and called *spatial aliasing*. It can indeed be experimentally observed in the laboratory, but usually does not affect the system performance since higher spatial frequencies are much reduced by the optics at the front end of the detector [39, 40, 41, 57].

The spatial pattern wavelength not only influences the overall amplitude of the response, it also affects its velocity tuning in an intricate way. This is shown in Figure 14.5 for 3 different pattern wavelengths, again assuming a sampling base of 2 deg. The larger the wavelength, the higher the pattern velocity at which the response becomes optimum. More precisely, the response optimum is directly proportional to the pattern wavelength; doubling the wavelength leads to a doubling of the optimum velocity. If we introduce the temporal frequency of the stimulus as being the ratio of the velocity and the pattern wavelength, the response curves for all different pattern velocities can be seen to coincide at one frequency (Figure 14.6).

Formally, this can be deduced from the above formulas by replacing v/λ with f_t, the temporal frequency of the motion stimulus. This frequency can be understood as the number of pattern cycles passing by one point in space per second. Introduction of the circular frequency $\omega = 2\pi f_t$ and rewriting Equation (14.2) accordingly results

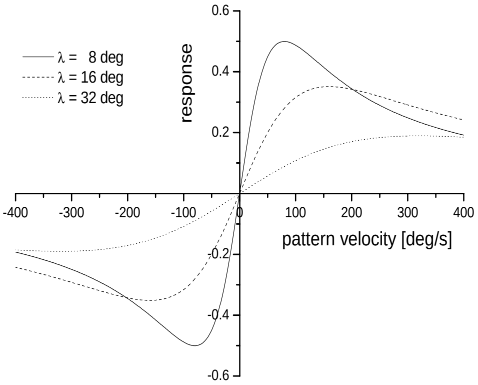

Figure 14.5

Steady-state responses of the minimal correlation model shown in Figure 14.2 as a function of pattern velocity for three different pattern wavelengths.

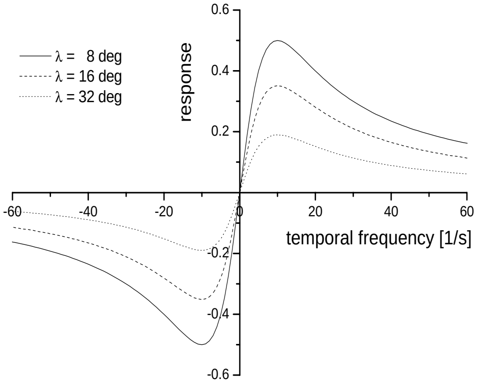

Figure 14.6

Steady-state responses of the minimal correlation model shown in Figure 14.2 as a function of temporal frequency for three different pattern wavelengths.

in:

$$R = \Delta I^2 \frac{\tau\omega}{1+(\tau\omega)^2} \cdot \sin(2\pi\Delta\phi/\lambda) \tag{14.3}$$

Setting $\delta R/\delta\omega = 0$, the response can be calculated to be optimum at the following frequency ω:

$$\omega = \frac{1}{\tau} \tag{14.4}$$

All the various properties of the correlation detector summarized above have been investigated in fly motion vision using either behavioral or electrical responses of motion-sensitive neurons [19, 22, 23, 27, 41, 53, 56], for review, see [30]. As predicted by the model, responses have been found to be maximum at a given stimulus velocity which depended on the spatial wavelength. The temporal frequency optimum was determined to be between 2-5 Hz, indicating a time-constant of about 20-100 msec, depending on the specific kind of low-pass filter assumed in the model calculations. Also, spatial aliasing did occur exactly between once and twice the interommatidial angle (the angle between the optical axes of two neighboring facets called *ommatidia*) indicating a nearest neighborhood interaction as the predominant input to the system. However, under low light levels, larger sampling bases have been also seen to contribute [19, 80].

Besides moving sine gratings, the correlation detector response to patterns with an arbitrary spatial luminance distribution $F(x)$ translating at a time varying velocity function $v(t)$ can be approximated. In this so-called *continuous approach*, the sampling base is assumed to be sufficiently small so that the signal in the right input line $(F(x_1))$ of the detector (see Figure 14.2) relates to the one in the left input $(F(x_2))$ by adding the first term of a Taylor series around x_1: $F(x_2) \sim F(x_1) + (dF/dx)_{x=x_1}dx$. If the temporal low-pass filter is replaced by a sufficiently small time delay ε, the output of the filter $F(t-\varepsilon)$ can be approximated again by a Taylor expansion. This leads to the following expression of the detector response at place x and time t [74]:

$$R(x,t) = v(t)\varepsilon(F \cdot F_{xx} - (F_x)^2) \tag{14.5}$$

Here, the dependence of the detector response on the local intensity $F(x)$ and its first (F_x) and second (F_{xx}) spatial derivative can be directly seen and calculated for any kind of spatial intensity profile. When applied to the examples considered above, i.e., moving sine gratings, the approximation reflects the central part of the velocity response curve around zero only (Figure 14.3) where the response is a quasi linear function of pattern velocity, but not the decline for higher velocities.

14.2.1.3 Orientation tuning

Another interesting feature of the correlation detector pertains to its orientation tuning. Intuitively, one might expect a regular cosine dependence of the detector to variation of the orientation of the moving pattern. This indeed is a good approximation for large pattern wavelengths (large with respect to the sampling base). However, in detail, the response as a function of the orientation is given by the following equation [97]:

$$R = \Delta I^2 \frac{\tau\omega}{1+(\tau\omega)^2} \cdot \sin(\cos\Theta \cdot 2\pi\Delta\phi/\lambda) \tag{14.6}$$

Equation (14.6) can be intuitively understood as the response being proportional to the sine of the phase difference between the two detector input lines. This phase difference is given as the ratio of the sampling base and the pattern wavelength as long as the pattern moves orthogonal to the grating $(\Theta = 0)$. Rotation of the pattern

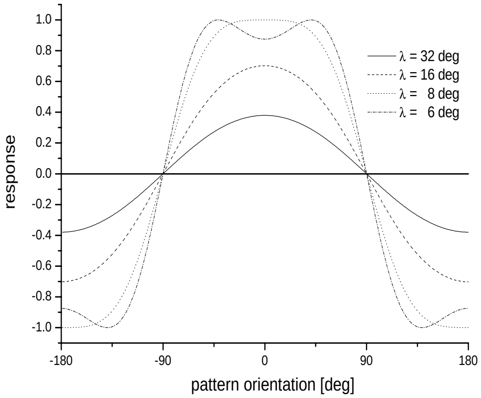

Figure 14.7

Steady-state responses of the minimal correlation model shown in Figure 14.2 as a function of pattern orientation for four different pattern wavelengths assuming a detector sampling base of 2 deg.

away from this orientation leads to a reduced phase difference but leaves the temporal modulation frequency of each input line unaffected. For pattern motion parallel to the grating, i.e., rotated by 90 deg, both input lines become activated in synchrony. The differences between a regular cosine tuning curve and the actual behavior of a correlation detector become obvious when the pattern wavelength is getting close to twice the sampling base (near the resolution limit), i.e., at wavelengths of 8 and 6 degrees assuming a sampling base of 2 deg (Figure 14.7). At $\lambda = 8$ deg, the tuning curve seems much more saturated than a simple cosine, and at $\lambda = 6$ deg, the tuning curves reveal dips where otherwise their maxima are located.

14.2.2 Dynamic response properties

In all the characteristics of the correlation detector discussed so far the response has always been considered as an average over time, i.e., during steady-state. However, in any living organism the velocity stimulus is changing as a function of time as it is moving, and responses occur as a function of time again. Often, responses have to occur in as short a time as possible to ensure the survival of the observer. Therefore, beside the steady-state properties, the dynamics of the local motion detector response are of great interest, too.

Figure 14.8 shows the time course of the detector response to a moving sine grating the velocity of which was stepped from zero to a constant value (corresponding to a temporal frequency of 2 Hz) at t = 1 sec and back to zero at t = 4 sec. Shown is the spatially summated output of all 8 motion detectors used in this simulation together with the individual output signals of the first 3 motion detectors. There are several

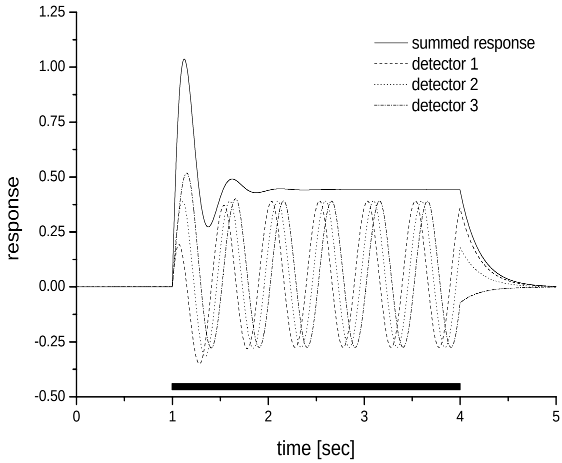

Figure 14.8

Dynamic response profile of the minimal correlation model shown in Figure 14.2 in response to the on- (at time = 1 sec) and off-set (at time = 4sec) of constant pattern motion. Shown are the summated output signals (in black) of an array of 8 such motion detectors (sampling base = 2 deg) together covering one spatial wavelength ($\lambda = 16$ deg) of the pattern, together with the output signals of three neighboring individual detector signals (in red, green and blue, respectively). The pattern was moving at 32 deg/sec resulting in a temporal frequency of 2 Hz.

notable things here.

First of all, the spatially summated response changes over time in the initial period of the stimulation. Before assuming its steady-state response given by Equation (14.2), the response is transiently ringing at the temporal frequency of the stimulus. The ringing is damped and settles over time to the steady-state value. Indeed, after being postulated based on modelling [4], such ringing behavior has been observed experimentally in fly motion sensitive neurons [27, 66]. Beside the velocity step, sinusoidal velocity modulation represents another example case where the dynamic responses of correlation detectors have been investigated [26]. Again, the spatially integrated response was found to deviate from being proportional to the velocity signal when either the amplitude or the frequency of the velocity modulation exceeded certain values. In summary, thus, the spatially integrated motion detector output follows the dynamics of the velocity input within certain limits; beyond that, significant distortions may occur. Those limits are in general determined by the time-course of the intrinsic filters of the motion detection system.

Another feature that can be seen in Figure 14.8 pertains to the individual detector signals. In contrast to the overall response that exhibits a ringing only initially, the individual detector output signals continue to oscillate at the temporal frequency of the pattern as long as the pattern is moving. The modulations of these signals are phase shifted due to the fact that they receive input from different spatial parts of the pattern. Only after spatially integrating the signals over different spatial phases, these oscillations become cancelled resulting in a smooth, constant steady-state response. Again this prediction has been experimentally verified using different techniques. In one set of experiments, presenting the pattern to the fly through a narrow slit restricted spatial integration. Recording from a lobula plate tangential cell revealed ongoing temporal oscillations at the temporal frequency of the stimulus [28]. Using a different approach to record local signals in an integrating neuron, Single and Borst filled one of the tangential cells with a calcium-sensitive dye and optically recorded the neural calcium responses in a spatially resolved way. Presenting a full-field spatial grating resulted in local calcium oscillations in the dendrite of the integrating neuron which had all the postulated properties [84]. For the further processing of local motion detector signals, it is important to note that meaningful information about the velocity cannot be deduced from the local detector output signals. Their signals as a function of time are only partially determined by the stimulus velocity, but additionally superimposed by signals reflecting local pattern properties. Therefore, some sort of postprocessing like spatial integration has to take place before such signals can be used for visual course control.

14.2.3 Additional filters and adaptive properties

As we have seen above, many properties of fly motion vision can be successfully captured by the correlation detector in its simplest form (Figure 14.2). Using Occam's razor, this model consists of only one temporal filter and one multiplier per subunit. This, however, is not the original form the model took when first proposed [73]. As we will see in the following, besides front-end spatial filters accounting for the optics, at least one more temporal filter with an adaptive time-constant needs to be postulated in order to account for a series of observations using the impulse response of the system as an indicator.

Fly neurons respond to an instantaneous displacement of the visual pattern with a sudden rise in activity followed by an exponential decay. In the language of system theory, such a stimulus represents a velocity pulse, and the response to that has been named the impulse response, accordingly. Such a behavior is correctly predicted by the minimal model [79]:

$$\langle R \rangle_\phi = \Delta I^2 \cdot \exp(-t/\tau_l) \cdot \sin(\Delta\phi)\sin(\psi) \tag{14.7}$$

Here, $\langle R \rangle_\phi$ denotes the spatial average over the responses of an array of motion detectors to a pattern that has been shifted by a visual angle ψ. Interestingly, the time constant of this decay has been shown to shorten when tested after presentation of an adapting motion stimulus [5, 79]. In terms of the minimal detector model this inevitably implies that the time constant of the low-pass filter has changed. Given that, one would expect a concomitant shift of the steady-state response towards higher velocities (see Equation (14.4)). This, however, could not be verified experimentally

[50]. Hence, the minimal model has to be extended by at least one more temporal filter in order to account for these experimental observations. Inserting an additional first-order high-pass filter with time-constant τ_h in the cross-arms of the detector (see Figure 14.9a) results in the following steady-state response [27]:

$$R = \Delta I^2 \sin(2\pi\Delta\phi/\lambda) \cdot \frac{\tau_h\omega \cdot (1+\tau_l\tau_h\omega^2)}{(1+\tau_l^2\omega^2)\cdot(1+\tau_h^2\omega^2)} \tag{14.8}$$

If the time-constants of both the high- and the low-pass filter are equal, the steady-state response becomes identical to the one of the minimal model (compare with Equation (14.3)). Interestingly, in such a detector model, only the high-pass filter determines the time course of the impulse response [12]:

$$\langle R \rangle_\phi = \Delta I^2 \exp(-t/\tau_h)\sin(\Delta\phi)\sin(\psi) \tag{14.9}$$

Consequently assuming that the time constant of the high-pass filter is the locus of adaptation led to the formulation of a detector model with an adaptive high-pass time-constant. To describe the time-constant of the high-pass filter as adaptive within a range of max τ_h and min τ_h, the following differential equation was used:

$$d\tau_h/dt = -(\tau_h - \min\tau_h)S + (\max\tau_h - \tau_h)K \tag{14.10}$$

Here, the time-constant decreases faster the further away it is from the minimum value it can assume, and this decrease is proportional to a signal S that we will define later. The time-constant increases faster the further it is away from its maximum value and this relaxation is also proportional to a constant factor K. Both S and K incorporate the time-constants for the adaptation and relaxation, respectively. This results in the following steady state value of $\tau_h(d\tau_h/dt = 0)$:

$$\tau_h = (\max\tau_h \cdot K + \min\tau_h \cdot S)/(K+S) \tag{14.11}$$

From Equation (14.11), one can see that for $S = 0$, $\tau_h = \max\tau_h$, while for $S >> K$, $\tau_h = \min\tau_h$. The final step in the formulation of an adaptive detector model is to define the signal driving adaptation of the time-constant. The shortening of the time-constant has been found to exhibit a peaked velocity dependence and to be the more pronounced the higher the stimulus contrast [78]. One way to obtain such a signal is from the rate of change of the low-pass output. This signal is larger the higher the contrast, and, up to a given temporal frequency, the higher the velocity of the moving grating. The high-frequency cut-off will be set by the low-pass time-constant. In the simulation shown in Figure 14.9, the output of the low-pass was smoothed by a 1st order filter with a 30 msec time-constant prior to differentiation. The resulting signal was then smoothed again by a 1st order filter with a 300 msec time-constant and finally squared.

The minimal model elaborated in this way (Figure 14.9a) can resolve the conflicts mentioned above. The detector displays an adaptive impulse response (Figure 14.9c) without altering the detector's temporal frequency tuning at different stimulus contrasts (Figure 14.9b). Besides revealing an adaptive impulse response, the elaborated

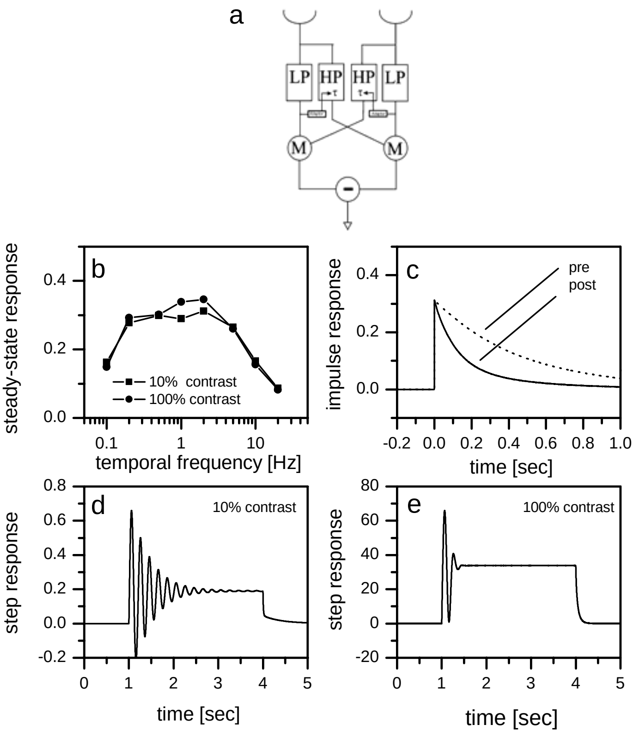

Figure 14.9

Basic properties of an adaptive high-pass detector. (a) Circuit diagram of the detector model. The structure is identical to the detector shown in Figure 14.2 except for the additional high-pass filter and the adaptor which shortens the high-pass filter time-constant. For the simulations, the high-pass filter had an adaptable time-constant between 50 (τ_{min}, fully adapted) and 500 (τ_{max}, undadapted) msec, the low-pass filter time-constant was fixed at 50 msec. (b) Steady-state responses of the adaptive detector to constantly drifting gratings of low (i.e., 10%) and high (i.e., 100%) contrasts. (c) Impulse responses of the adaptive detector before (red) and after (black) exposure to an adaptive stimulus. (d,e) Step response of the adaptive detector at low (d) and high (e) pattern contrast.

model with an adaptive high-pass time-constant also faithfully mimics a particular contrast-dependency of transient response oscillations observed in fly motion sensitive neurons. The experimental observations demonstrate that the transient ringing is most pronounced at the onset of stimulus motion at low pattern low-contrasts, whereas it significantly shortens with higher pattern contrasts [78]. In terms of the model presented above, this implies a shortening of the high-pass time-constant during the first stimulus period since the high-pass time-constant predominates the length of the ringing period. Indeed, the elaborated model as shown in Figure 14.9a displays this particular response feature of fly visual interneurons as well (Figures 14.9d,e). As an alternative, the high-pass filter was also proposed at the front-end of the detector instead of the cross arms [52]. This elaboration has the appeal to mimic the adaptive response features of fly motion-sensitive neurons without any further additions to the detector, but fails to show an impulse response [12]. In summary, as is shown above, only few elaborations of the minimal correlation detector are needed to account for many response peculiarities of fly motion vision in astonishing detail [12].

14.3 Spatial processing of local motion signals by Lobula plate tangential cells

Following the structure of the fly visual ganglia, the columnar organization reflecting the pixelation of the image by the facet eye is given up at the level of the third visual ganglion, the so-called lobula complex. In its posterior part, called the *lobula plate*, the tangential cells extend their large dendrites, integrating across many hundreds to thousands of columnar elements. Due to the large diameter of their processes, these cells have been studied intracellularly by electrophysiological recordings in great detail [22, 53, 54, 55, 56, 59, 60]. These integrating neurons can today be described and modeled at a quite detailed biophysical level based on direct recordings of their membrane properties and synaptic connectivity [10, 43, 45]. Modeling these lobula plate tangential cells not only helped to understand the details underlying their electrical signals but also turned out to be indispensable for the analysis of the local motion detection mechanisms reviewed above, because it allowed, through combined experimental/modelling studies, to dissect out the input signals in an indirect way and to assess the potential contribution of network interactions to various phenomena such as gain control (see below).

14.3.1 Compartmental models of tangential cells

A compartmental model describes the spatio-temporal potential distribution within single neurons based on the exact anatomy of the cell and all ionic currents flowing across its membrane. Once these data are known, the neuron is turned into an electrical equivalent circuit which can be represented as a set of coupled differential equations which in turn can be solved numerically by standard software packages [81]. As a first step a digital base of various tangential cells was built by 3D-reconstructing individual neurons from cobalt-stained material [10]. This database includes both CH-cells (vCH- and dCH-cell), all three HS-cells (HSN-, HSE- and HSS-cell) and

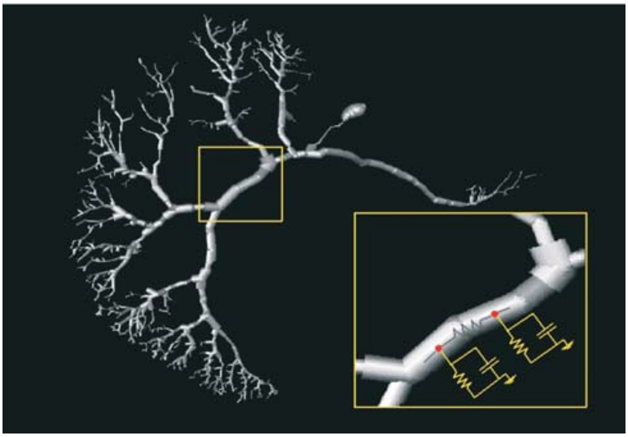

Figure 14.10

Compartmental model of a VS1-cell. The inset shows a simplified version of an electrical equivalent circuit representing each compartment in the simulation. The cell is not shown plane-parallel, but rotated by 20-30° around the dorso-ventral axis with the dendrites towards the viewer (Modified from [10]).

most members of the VS-cell family from the blowfly lobula plate. In contrast to the spiking neurons H1, H2 and Hu, the above neurons do not produce full-blown action potentials but respond to visual motion mainly by a graded shift of their membrane potential. In HS- and VS-cells, but not CH-cells, small amplitude action potentials can be superimposed on the graded response [58, 44]. Once the anatomical data were collected, the specific membrane properties had to be determined from current- and voltage-clamp experiments. In a first series of experiments, hyperpolarizing and depolarizing currents were injected to determine steady-state I-V curves. It appeared that at potentials more negative than resting, a linear relationship holds, whereas at potentials more positive than resting, an outward rectification was observed. Therefore, in all subsequent experiments, when a sinusoidal current of variable frequency was injected, a negative DC current was superimposed to keep the neurons in a hyperpolarized state. The resulting amplitude and phase spectra revealed an average steady-state input resistance of 4-5 $M\Omega$ and a cut-off frequency between 40 and 80 Hz. To determine the passive membrane parameters R_m (specific membrane resistance), R_i (specific internal resistivity) and C_m (specific membrane capacitance), the experiments were repeated in computer simulations on compartmental models of the cells. Assuming a spatially homogeneous distribution of these parameters, the 3-dimensional parameter space was screened through. In comparing the model response with the experimental dataset a single optima for each neuron was found (see table 14.3.1). No characteristic differences between different members of the same cell class were detectable. We also applied an error analysis of the fitting procedure to see how much the different membrane parameters could be varied away from the point of best fit and still lead to acceptable behavior of the model as compared to the

	CH-cells	HS-cells	VS-cells
R_m	$2.5k\Omega\text{cm}^2$	$2.0k\Omega\text{cm}^2$	$2.0k\Omega\text{cm}^2$
R_i	60 Ωcm	40 Ωcm	40 Ωcm ‘
C_m	$1.5\mu F/cm^2$	$0.9\mu F/cm^2$	$0.8\mu F/cm^2$
Delayed rectifying K-current	+	+	+
Fast Na-dependent K-current	-	+	+
Fast Na-current	-	+	+
Non-inactivating LVA Ca-current	Fast	Slow	Slow

Table 1: Summary of passive membrane parameters and voltage-activated ionic currents in lobula plate tangential cells. Data from Borst and Haag, 1996, and Haag et al., 1997.

experimental data set given the statistical fluctuations inherent in the experiments.

In the next step, voltage-activated membrane currents were studied using the switched electrode voltage clamp technique [43]. In CH-cells, two currents were identified: a slow calcium inward current, and a delayed rectifying, non-inactivating potassium outward current. HS- and VS-cells appeared to possess similar currents as did CH-cells but, in addition, exhibited a fast activating sodium inward current and a sodium-activated potassium outward current. While the delayed rectifying potassium current in all three cell classes was responsible for the observed outward rectification, the sodium inward current produced the fast and irregular spike-like depolarizations found in HS- and VS-cells but not in CH-cells [58, 44]. When blocking the sodium current by either TTX or intracellular QX314, action potentials could no longer be elicited in HS-cells under current-clamp conditions. Again, as in the analysis of passive membrane properties, voltage-activated currents were incorporated with the appropriate characteristics into the compartmental models of the cells.

Calcium currents were further analyzed combining the switched-electrode voltage clamp technique with optical recording using calcium sensitive dyes [46]. This allowed the study of Calcium currents, which were too small to be detectable in voltage-clamp experiments and, in addition revealed, their spatial distribution. For all three cell types considered, CH-, HS- and VS-cells, the activation curve turned out to be rather flat covering a voltage range from -60 to -20 mV in dendritic as well as presynaptic areas of the cells. The calcium increase was fastest for CH-cells with a time constant of about 70 ms. In HS- and VS-cells the time constant amounted to 400 – 700 ms. The calcium dynamics as determined in different regions of the cells were similar, except for a small segment between the axon and the dendrite in HS- and VS-cells, where the calcium increase was significantly faster. In summary, these results show the existence of a low-voltage-activated (LVA) calcium current in dendritic as well as presynaptic regions of fly LPTCs with little or no inactivation.

Beside ionic currents another crucial property of neurons is the repertoire of transmitter receptors on their dendrites. Transmitter-gated currents were studied in HS- and VS-cells using an *in vitro* preparation of the fly brain. Pharmacological and im-

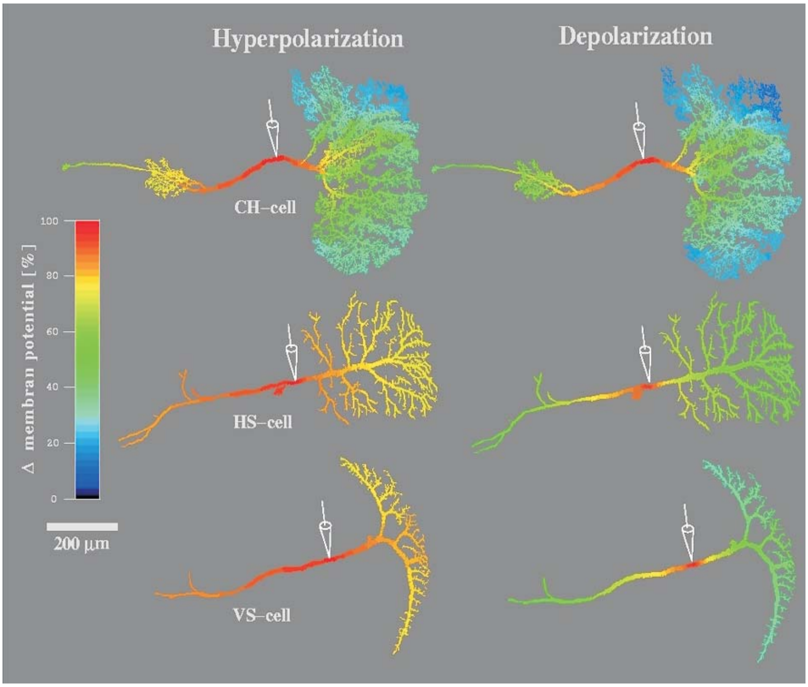

Figure 14.11

Steady-state voltage distribution in compartmental models of fly LPTCs after simulated current injection into the axon. False-color code represents the local voltage in percent of voltage at injection site (From [43]). (See color insert.)

munohistochemical revealed that HS- and VS-cells receive their excitatory input via nicotinic acetylcholine receptors and become inhibited via GABA receptors [16, 17].

Using all the information summarized above and incorporating them into compartmental models allowed, amongst other things, to visualize the steady-state potential distribution following injection of de- or hyperpolarizing currents into the axon of the neuron (Figure 14.11). In addition, knowing the precise kinetics of voltage-activated currents led the models reproduce many of the experimentally observed dynamic response properties like e.g., frequency-dependent amplification of synaptic input in HS-cells [44, 45].

14.3.2 Dendritic integration and gain control

One response feature of LPTCs studied intensively in the past concerns their spatial integration characteristics: when enlarging the area in which the motion stimulus is displayed the response saturates significantly [9, 42, 55, 83]. The interesting fact is the observation that such a saturation occurs not only for motion along the preferred, but also along the null direction of the cell. Furthermore, for patterns moving at different velocities, different saturation plateaus are found. This latter phenomenon has also been called *gain control* and was observed at the behavioral level in flies, too [76]. In general, modelling studies along with measurements of motion induced changes of input conductances revealed that the circuit model of a single tangential cell and its presynaptic array of motion detectors is fully sufficient to produce the observed saturation of membrane response.

For understanding the phenomenon of gain control, one needs to realize that a) the tangential cells receive excitatory and inhibitory synaptic input from elements equivalent to the subunits of the correlation detector, and b) these presynaptic elements have only a weak direction selectivity: they not only respond to motion along their preferred direction but also, to a lesser extent, to motion along the opposite direction (see panel 2′1 in Figure 14.6). These conclusions are based on various electrophysiological, calcium-imaging and pharmacological experiments [7, 9, 11, 70, 71, 83, 85]. Given that, motion along one direction leads to a joint, though differently weighed activation of excitatory and inhibitory input, resulting in a mixed reversal potential at which the postsynaptic response settles for large field stimuli. This can be seen by the following calculation where the membrane potential (V) of a tangential cell is approximated as an isopotential compartment (E_e, g_e denoting excitatory reversal potential and conductance, respectively, subscript i for inhibitory, $E_{leak} = 0$):

$$V = (E_e g_e + E_i g_i)/(g_e + g_i + g_{leak}) \qquad (14.12)$$

Assuming $E_e = -E_i$ and introducing $c = g_i/g_e$ to denote the ratio of inhibitory and excitatory conductances being co-activated during preferred direction motion, one obtains:

$$V = E_e g_e (1-c)/(g_e + c g_e + g_{leak}) \qquad (14.13)$$

With increasing pattern size, g_e become large compared to g_{leak}: The membrane potential tends towards a saturation level. This level can be expressed as:

$$\lim_{g_e \to \infty}(V) = E_e(1-c)/(1+c) \qquad (14.14)$$

As can be calculated from such correlation-type input elements, the activation ratio of these opponent inputs is a function of pattern velocity:

$$c \cong \cos(R - \Phi(\omega))/\cos(R + \Phi(\omega)) \qquad (14.15)$$

with R denoting 2π times the ratio of the EMD's sampling base and the spatial pattern wavelength λ, Φ denoting the phase response of the EMD's temporal filter and $\omega = 2\pi v/\lambda$. Consequently, motion in one direction jointly activates excitatory and inhibitory inputs with a ratio that depends on velocity. This explains how the postsynaptic membrane potential saturates with increasing pattern size at different levels for different pattern velocities.

Taking into account the membrane properties of real tangential cells (Figure 14.12a), simulations of detailed compartmental models indeed revealed the phenomenon of gain control. Stimulating the neuron by pattern motion of increasing size lead to a

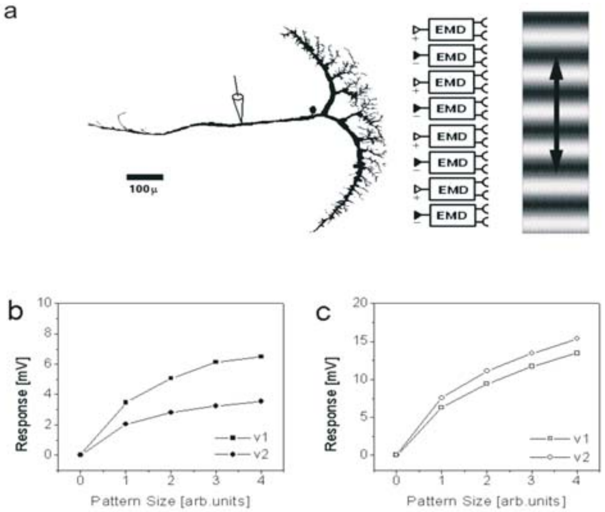

Figure 14.12

Spatial integration properties of compartmental models of LPTCs. (a) Simulation of a LPTC receiving input from two arrays of elementary motion detectors (EMD) tuned to opposite directions of motion and forming excitatory (+) and inhibitory (-) inputs onto the dendrite of the cell, respectively. A VS-cell was 3D-reconstructed from cobalt-stained material and was simulated as having only passive membrane properties. (b,c) Gain control in the model cell before and after blockade of inhibition. When stimulated by patterns of increasing size at two different velocities, the axonal membrane potential saturates at different levels (gain control). After blocking the inhibitory inputs, both velocities yield similar responses (Modified from [83]).

spatial saturation of the resulting membrane potential in the axon. If the velocity of pattern motion is changed, a different saturation level is assumed (Figure 14.12b). Blocking the inhibitory input still resulted in a spatial saturation but abolished the phenomenon of gain control (Figure 14.12c): now, the same level was approximated for increasing pattern size independent of the pattern velocity [83].

In summary, thus, gain control is produced without the need of any further network interactions. These assumptions were experimentally verified by blocking the inhibitory input with picrotoxinin [83] resulting in three observations: a) the preferred direction response grew larger and the null direction response changed its sign from a hyper- to a depolarization, b) the change of input resistance induced by preferred direction motion decreased showing that, before, inhibitory currents were activated as well and c) as a final proof of the above explanation, gain control was abolished.

Models of HS- and VS-cells produced a spatial saturation curve indistinguishable from their natural counter parts. Real CH-cells, however, saturated much more strongly leading to a discrepancy between experimental and model data. Originally, this finding led to the assumption that CH-cells might have spatially inhomogeneous membrane parameters with very high values of membrane resistance in their dendrite [45]. A detailed modelling study following up on this revealed, however, that even when allowing for such complex model no satisfying fit between all available data sets and the respective model behavior can be achieved [20]. A possible explanation might come from the observation that in contrast to the model outline, real CH-cells do not seem to receive direct input from retinotopically arranged arrays of motion-sensitive elements but rather are connected indirectly to the visual surround via HS-cells [48]. Whether this finding can resolve all existing discrepancies is currently being investigated.

14.3.3 Binocular interactions

Many of the LPTCs have been found to be sensitive to image motion in front of the contralateral eye, in addition to motion in front of their ipsilateral eye. In the case of LPTCs of the graded or mixed response type, i.e., HS and CH-cells, this was found to be true for HSN, HSE, dCH and vCH, but not for HSS-cells. All these cells are excited on their ipsilateral side by motion from the front to the back and receive additional excitatory input by contralateral back-to-front motion [24, 45, 47, 53, 56, 61]. Hence, they are tuned to rotational flow-fields.

Using dual recording techniques, one extracellular recording from the spiking neuron and one intracellular recording from the CH-cell, two heterolateral neurons were identified as providing the excitatory input to HS- as well as CH-cells: the H1- and the H2-cell. As is indicated in Figure 14.1, these cells have the appropriate preferred direction to tune the CH-cells to rotatory motion. In addition, an hitherto unidentified neuron (Hu) with an opposite preferred direction is inhibiting the contralateral CH-cells. While these findings provide a sufficient explanation for the selectivity of CH-cells for rotational cues, additional connections between LPTCs within one lobula plate were recently discovered [47]. Again using dual recording techniques, CH- and HS-cells were found to excite those cells that have identical preferred directions like Hu, and CH-cell were found to inhibit those neurons with opposite preferred direction, like H1 and H2 (see Figure 14.1). Through this kind of ipsilateral connections onto heterolateral neurons, HS- and CH-cells could be demonstrated to inhibit their contralateral counterparts; excitation e.g., in the left dCH-cell should inhibit the right dCH-cell, and vice versa, while inhibition in the left CH-cell should facilitate excitation in the right one.

Moreover, close inspection of the circuit diagram reveals the existence of feedback loops bringing back the signals onto the cell where they started from. Several of these predictions could indeed be experimentally verified [47]. The conclusion from these experiments is that the intrinsic connectivity between the different tangential cells within one lobula plate and between the lobula plates in both hemispheres favors an asymmetrical distribution of excitation. Such an asymmetry will be imposed on the network from the sensory input when rotational flow-fields stimulate the eyes, but not when translational stimuli occur. The flow-field selectivity of LPTCs therefore seems not only determined by their feed-forward connectivity, but also by the intrinsic wiring within the network formed by LPTCs in both hemispheres. Of course, the circuit diagram presented in Figure 14.1 only contains a small number of the existing lobula plate neurons (5 out 60 on each side!) where such connections have been studied in detail to date. In particular, no neurons sensitive to vertical image motion are included there. More investigations concerning the connectivity amongst such VS-cells and between the horizontal and vertical cells are presently under way (Haag and Borst, in preparation) possibly providing an explanation for the complex flow-fields measured in many neurons of the lobula plate [63, 64].

To gain a further understanding in the functional consequences of such network interactions, modelling work was started. As a first step, a network model of the lobula plate was built using single-compartment models of each cell in the circuit. After adjusting the connectivities to the experimental data, the individual circuit elements revealed similar responses to binocular motion stimuli as their natural counterparts (Figures 14.13 and 14.14). First of all, in response to rotational motion stimuli, all neurons modulate their response strongly when the stimulus is switched from clockwise to counter clockwise rotation (Figure 14.13). In this case, the internal connections of the circuit amplify the excitation levels imposed onto the neurons by the local motion detectors, i.e., when a neuron on one side is excited, its counterpart on the other side becomes inhibited. The situation changes when instead of rotational motion translational stimuli are presented (Figure 14.14): now, the modulation is weaker in HS-cells and CH-cells when the stimulus switches from contraction to expansion. In particular in CH-cells, the membrane potential departs only little from resting during stimulation either way. In this case, the internal connections work against the feed forward signals coming from the local motion detectors and, thus, reduce the responses substantially. While these simulations represent only a starting point, future studies will investigate how these internal connectivities affect the response behavior of the circuit elements once more critical stimuli are presented e.g., low light levels or low contrast patterns where noise becomes an issue.

14.3.4 Dendro-dendritic interactions

In contrast to the lobula plate neurons covered so far which all respond strongest to large-field motion, another group of tangential cells has been described which respond best to small moving objects or relative motion (called FD- or CI-neurons; [25, 38]. These cells receive inhibitory input when contralateral back-to-front motion is additionally displayed to an excitatory ipsilateral front-to-back stimulus. There is evidence that the vCH-cell is responsible for conveying this type of inhibitory input since the FD1-cell was shown to loose its inhibitory input after photo-inactivation of the vCH-cell [95]. To what extend the CI-cells are also inhibited by the vCH-cell and whether the inhibition of the FD1- and/or CI-neurons by the vCH-cell is also responsible for their ipsilateral small-field tuning, i.e., their preference for small objects moving, and thus represent the ‘pool cell’ postulated in the original models

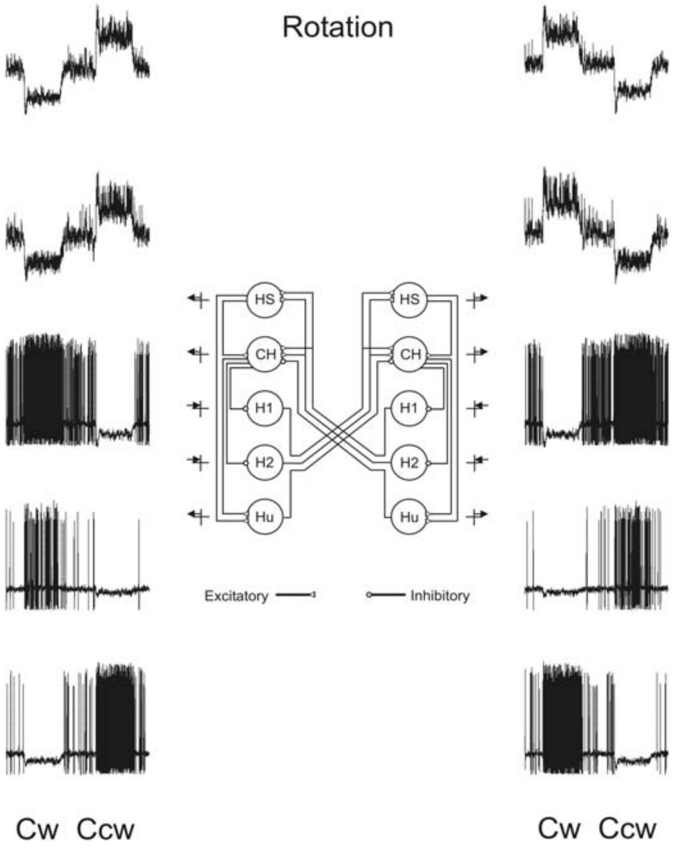

Figure 14.13

Simulation of the network of tangential cells with each neuron modeled as a single compartment. The circuit, without the input from local motion detectors, is shown in the center. To the left and to the right, the signals of 5 different tangential cells sensitive to horizontal image motion are displayed as a function of time. The stimulus consists of a binocular image rotation, first clockwise (CW) for 1 sec, then counterclockwise (CCW) for 1 sec.

of figure-ground discrimination [76, 77], is not clear at the moment.

However, the following facts lead to some interesting speculations of how the CH-cell might be responsible for the small-field tuning in FD- and CI-neurons. First of all, the CH-cell is known to be GABAergic with chemical output synapses within its dendrite in the lobula plate [68, 37]. Since there are no chemical output synapses found within the protocerebral ramifications of CH-cells [37], it seems reasonably justified to assume that the CH-cell inhibits other lobula plate tangential cells within

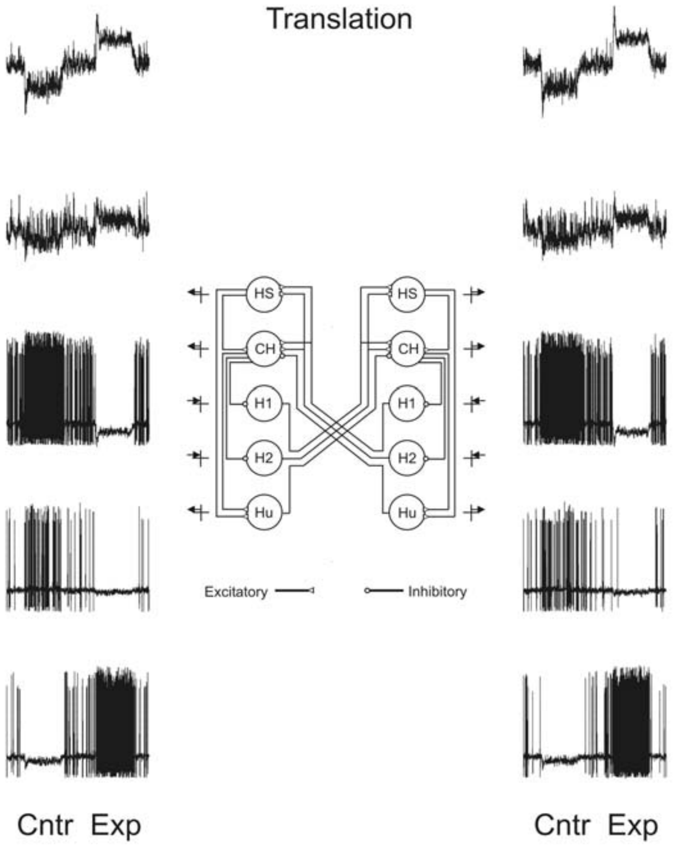

Figure 14.14

Same as Figure 14.13 except that here the stimulus consists of a binocular image translation, first contraction (Cntr) for 1 sec, then expansion (Exp) for 1 sec.

their dendrites, where they also receive the retinotopic input from local motion detectors. If CH-cells received the same type of retinotopic input like all other tangential cells, this might lead to a complete cancellation of any type of excitation within the dendrite of the postsynaptic partner cells. This, however, turned out not to be true. Recent findings suggest that the CH-cells receive their motion input only indirectly through dendro-dendritic electrical synapses from HS-cells [48] leading to a spatially blurred motion image on their dendrite [21]. As is shown in Figure 14.15 and Figure 14.16, inhibition of FD-cells by such a blurred activity pattern from CH-cell dendrites indeed enhances motion edges as exist during movement of small objects in front of a stable background (Figure 14.15), but cancel the signals almost completely

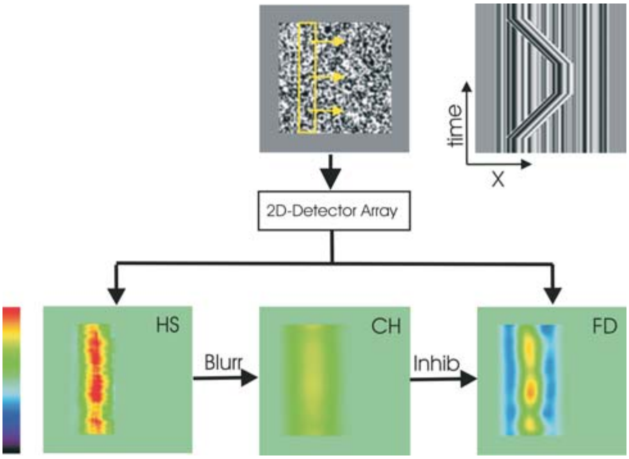

Figure 14.15

Possible mechanism underlying the selectivity of FD-cells for relative motion in the fly visual system. A The stimulus consists of a 2D random dot pattern covering both a bar and a background (top center). The bar is moving in front of the background to the right, stops, and to the left again. Motion is shown as an x-t plot for a center cross section through the stimulus pattern (top right). This stimulus is processed by a 2D array of motion detectors which then feed onto the dendrites of HS-cells (bottom left) and FD-cells (bottom right). HS-cell dendrites contact the dendrites of CH-cells (bottom center) which in turn inhibit FD-cells in a dendro-dendritic manner. The excitation level of the dendrite of all three tangential cells is shown as a false-color image (scale bar bottom left) (See color insert.).

when large-field background motion is presented (Figure 14.16). However, before any selective responses of FD-cells to small-field motion can be obtained from such dendritic activity patterns, additional nonlinear operations have to be postulated like e.g., rectification of dendritic signals and/or local spike generation [76, 77]. Nevertheless, the present experimental findings support such a biophysical mechanism to underlie the particular response properties of FD- and CI-cells, representing a rather unique example of how image processing is done within the dendrites of individual neurons (Cuntz et al., in preparation).

14.4 Conclusions

Fly motion vision represents one of the best studied biological systems where the modelling efforts summarized above went hand in hand with the experimental investigations. However, as the reader might have noticed, the models of local motion

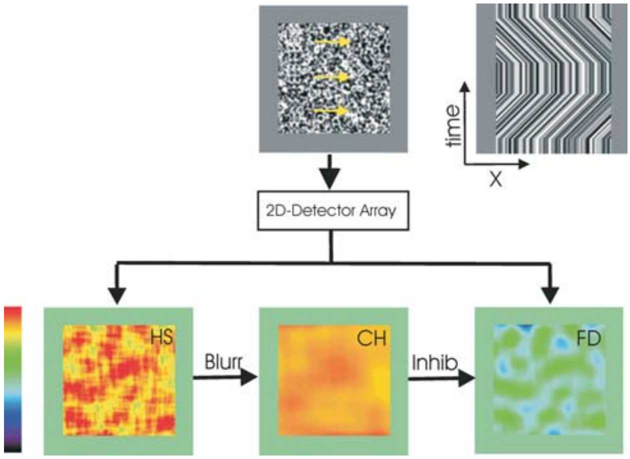

Figure 14.16

The same as Figure 14.15, except for the fact that this time there is only large-field motion of the background (see xt-plot in top row). Under these circumstances, inhibition of the motion signals by the CH-cell dendrite leads to an annulment of excitation on the FD-cell dendrite (See color insert.).

detection are settled at a different level than those of the spatially integrating stages, i.e., the tangential cells of the lobula plate. While the properties of local motion detectors have been inferred from a black box analyses using visual motion stimuli as input and, in most cases, their spatially and/or temporally integrated signals as output, as manifested either in the tangential cells or in optomotor response behavior, spatial processing of local motion information was studied directly by electrophysiological recordings of the neural elements. This has to do with the trivial fact that the columnar elements in the medulla where local motion detection is likely to take place [2, 3, 18, 33] are much smaller than the large tangential cells of the lobula plate and, thus, are much harder to record from intracellularly. As a consequence, the motion detection system as simulated today represents a hybrid model with purely algorithmic processing stages at the level of local motion detection (low-pass filtering, multiplication) and biophysically realistic compartmental models of the neurons thereafter.

Thus, a major thrust of present research efforts is to uncover the cellular identities of the neurons constituting the local motion detector and, hopefully, the biophysical processes underlying their elementary operations using e.g., genetically encoded indicators of neural activity [69]. However, independent of that, our understanding of the functional properties of fly motion vision will be tremendously improved by considering the fact that the circuits which usually are tested in the laboratory under open-loop conditions, operate naturally under closed-loop conditions [94]. Furthermore, the characteristics of the fly's flight maneuvers [91, 92, 93] together with the spatial frequency content in natural scenes [32] is likely to shape the input to the fly's motion detection system in a specific way which in turn will have led to specific adaptations of the system analyzing it. Since present technology does not yet allow to record from single neurons while the fly is free to fly around, present efforts aim at recording the visual flow the fly is experiencing during free flight [82] which can be later used as a stimulus presented to the tethered fly for electrophysiological recording [62]. Along these lines, future modelling efforts that make use of neuromorphic chips implementing the fly motion detection circuits offer the advantage to be testable in closed-loop situations in real time (e.g., [65]).

Acknowledgments: I would like to thank Hermann Cuntz, Karl Farrow and Juergen Haag for critically reading the manuscript, Robert Schorner for the artwork in Figure 14.1 and Agnes Haid for secretarial help. The work was supported by the Max-Planck-Society.

References

[1] Barlow H.B., and Levick W.R., (1965). The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* **178**, 477-504.

[2] Bausenwein B., Dittrich A.P.M., and Fischbach K.F. (1992) The optic lobe of Drosophila melanogaster. II. Sorting of retinotopic pathways in the medulla. *Cell Tissue Res.,* **267:** 17-28.

[3] Bausenwein B., and Fischbach K.F. (1992) Activity labeling patterns in the medulla of Drosophila melanogaster caused by motion stimuli. *Cell Tissue Res.,* **270:** 25-35.

[4] Borst A., and Bahde S. (1986) What kind of movement detector is triggering the landing response of the housefly? *Biol. Cybern.,* **55:** 59-69.

[5] Borst A., and Egelhaaf M. (1987) Temporal modulation of luminance adapts time constant of fly movement detectors. *Biol. Cybern.,* **56:** 209-215.

[6] Borst A., and Egelhaaf M. (1989) Principles of visual motion detection. *Trends Neurosci.,* **12:** 297-306.

[7] Borst A., and Egelhaaf M. (1990) Direction selectivity of fly motion-sensitive neurons is computed in a two-stage process. *Proc. Natl. Acad. Sci. Am.,* **87:** 9363-9367.

[8] Borst A., and Egelhaaf M. (1993) Detecting visual motion: Theory and models. In: Miles F.A., Wallman J. (eds.) *Visual Motion and its Role in the Stabilization of Gaze*. Elsevier, pp 3-27.

[9] Borst A., Egelhaaf M., and Haag J. (1995) Mechanisms of dendritic integration underlying gain control in fly motion-sensitive interneurons. *J. Comput. Neurosci.,* **2:** 5-18.

[10] Borst A., and Haag J. (1996) The intrinsic electrophysiological characteristics of fly lobula plate tangential cells. I. Passive membrane properties. *J. Computat. Neurosci.,* **3:** 313-336.

[11] Borst A., and Single S. (2000) Local current spread in electrically compact neurons of the fly. *Neuroscience Letters,* **285:** 123 - 126.

[12] Borst A., Reisenman C., and Haag J. (2003) Adaptation of response transients in fly motion vision. II. model studies *Vision Res.*, **43**: 1303-1322.

[13] Borst A., and Haag J. (2002) Neural networks in the cockpit of the fly. *J. Comp. Physiol.,* **188:** 419-437.

[14] Borst A. (2002) Noise, not stimulus entropy, determines neural information rate. *J. Computat. Neurosci.,* **14**: 23-31.

[15] Brenner N., Bialek W., and de Ruyter van Steveninck R.R. (2000) Adaptive rescaling maximizes information transmission. *Neuron,* **26:** 695-702.

[16] Brotz T., and Borst A. (1996) Cholinergic and GABAergic receptors on fly tangential cells and their role in visual motion detection. *J. Neurophysiol.,* **76:** 1786-1799.

[17] Brotz T., Gundelfinger E., and Borst A. (2001) Cholinergic and GABAergic pathways in fly motion vision. *BioMedCentral Neuroscience,* **2:** 1.

[18] Buchner E., Buchner S., and Buelthoff I. (1984) Deoxyglucose mapping of nervous activity induced in Drosophila brain by visual movement. *J. Comp. Physiol. A,* **155:** 471-483.

[19] Buchner E. (1984) Behavioural analysis of spatial vision in insects. In *Photoreception and Vision in Invertebrates*, M.A. Ali, (Ed.) Plenum Press: New York, London, pp. 561-621.

[20] Cuntz H. (2000) Raeumliche Verteilung von Membranparametern in *Fliegenneuronen: eine Simulationsstudie*. Diploma Thesis, Tuebingen.

[21] Duerr V., and Egelhaaf M. (1999) *In vivo* calcium accumulation in presynaptic and postsynaptic dendrites of visual interneurons. *J. of Neurophysiol.,* **82:** 3327-3338.

[22] Eckert H. (1973) Optomotorische Untersuchungen am visuellen System der Stubenfliege *Musca domestica L. Kybernetik,* **14:** 1-23.

[23] Eckert H. (1980) Functional properties of the H1-neurone in the third optic ganglion of the blowfly, *Phaenicia. J. Comp. Physiol.,* **135:** 29-39.

[24] Eckert H., and Dvorak D.R. (1983) The centrifugal horizontal cells in the lobula plate of the blowfly *Phaenicia sericata. J. Insect. Physiol.,* **29:** 547-560.

[25] Egelhaaf M. (1985) On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. II. Figure-Detection Cells, a new class of visual interneurons. *Biol. Cybern.,* **52:** 195-209.

[26] Egelhaaf M, and Reichardt W (1987) Dynamic response properties of movement detectors: Theoretical analysis and electrophysiological investigation in the visual system of the fly. *Biol. Cybern.,* **56:** 69-87.

[27] Egelhaaf M., and Borst A. (1989) Transient and steady-state response properties of movement detectors. *J. Opt. Soc. Am. A,* **6:** 116-127.

[28] Egelhaaf M., Borst A., and Reichardt W. (1989) Computational structure of a biological motion detection system as revealed by local detector analysis in the fly's nervous system. *J. Opt. Soc. Am. A,* **6:** 1070-1087.

[29] Egelhaaf M., and Borst A. (1992) Are there separate ON and OFF channels in fly motion vision? *Vis. Neurosci.,* **8:** 151-164.

[30] Egelhaaf M., and Borst A. (1993) Movement detection in arthropods. In: Miles F.A., Wallman J. (eds.) *Visual Motion and its Role in the Stabilization of Gaze.* Elsevier, pp 53-77.

[31] Fairhall A.L., Lewen G.D., Bialek W., and de Ruyter van Steveninck R.R. (2001) Efficiency and ambiguity in an adaptive neural code. *Nature,* **412:** 787-792.

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

[33] Fischbach K.F., and Dittrich A.P.M. (1989) The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.,* **258:** 441-475.

[34] Foster D.H. (1969) The response of the human visual system to moving spatially-periodic patterns. *Vision Res.,* **9:** 577-590.

[35] Foster D.H. (1971) The response of the human visual system to moving spatially-periodic patterns: Further analysis. *Vision Res.,* **11:** 57-81.

[36] Frye M.A., and Dickinson M.H. (2001) Fly flight: a model for the neural control of complex behavior. *Neuron,* **32:** 385-388.

[37] Gauck V., Egelhaaf M., and Borst A. (1997) Synapse distribution on VCH, an inhibitory, motion-sensitive interneuron in the fly visual system. *J. Comp. Neurol.,* **381:** 489-499.

[38] Gauck V., and Borst A. (1999) Spatial response properties of contralateral inhibited lobula plate tangential cells in the fly visual system. *J. Comp. Neurol.,* **406:** 51-71.

[39] Goetz K.G. (1964) Optomotorische Untersuchungen des visuellen Systems einiger Augenmutanten der Fruchtfliege Drosophila. *Kybernetik,* **2:** 77-92.

[40] Goetz K.G. (1965) Die optischen bertragungseigenschaften der Komplexaugen von Drosophila. *Kybernetik,* **2:** 215-221.

[41] Goetz K.G. (1972) Principles of optomotor reactions in insects. *Bibl. Ophthal.*

82: 251-259.

[42] Haag J., Egelhaaf M., and Borst A. (1992) Dendritic integration of motion information in visual interneurons of the blowfly. *Neuroscience Letters,* **140:** 173-176.

[43] Haag J., Theunissen F., and Borst A. (1997) The intrinsic electrophysiological characteristics of fly lobula plate tangential cells. II. Active membrane properties. *J. Computat. Neurosci.,* **4:** 349-369.

[44] Haag J., and Borst A. (1996) Amplification of high-frequency synaptic inputs by active dendritic membrane processes. *Nature,* **379:** 639-641.

[45] Haag J., Vermeulen A., and Borst A. (1999) The intrinsic electrophysiological characteristics of fly lobula plate tangential cells. III. Visual response properties. *J. Computat. Neurosci.,* **7:** 213-234.

[46] Haag J., and Borst A. (2000) Spatial distribution and characteristics of voltage-gated calcium currents within visual interneurons. *J. Neurophysiol.,* **83:** 1039-1051.

[47] Haag J., and Borst A. (2001) Recurrent network interactions underlying flow-field selectivity of visual interneurons. *J. Neurosci.,* **21:** 5685-5692.

[48] Haag J., and Borst A. (2002) Dendro-dendritic interactions between motion-sensitive large-field neurons in the fly. *J. Neurosci.,* **22:** 3227-3233.

[49] Hassenstein B., and Reichardt W. (1956) Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rsselkfers Chlorophanus. *Z. Naturforsch.,* **11b:** 513-524.

[50] Harris R.A., O'Carroll D.C., and Laughlin S.B. (1999) Adaptation and the temporal delay filter of fly motion detectors. *Vision Res.,* **39:** 2603-2613.

[51] Harris R.A., O'Carroll D.C., and Laughlin S.B. (2000). Contrast gain reduction in fly motion adaptation. *Neuron,* **28:** 595-606.

[52] Harris R.A., and O'Carroll D.C. (2002). Afterimages in fly motion vision. *Vision Research,* **42:** 1701-1714.

[53] Hausen K. (1981) Monocular and binocular computation of motion in the lobula plate of the fly. *Verh. Dtsch. Zool. Ges.,* **74:** 49-70.

[54] Hausen K. (1982a) Motion sensitive interneurons in the optomotor system of the fly. I. The horizontal cells: structure and signals. *Biol. Cybern.,* **45:** 143-156.

[55] Hausen K. (1982b) Motion sensitive interneurons in the optomotor system of the fly. II. The horizontal cells: receptive field organization and response characteristics. *Biol. Cybern.,* **46:** 67-79.

[56] Hausen K. (1984) The lobula-complex of the fly: Structure, function and significance in visual behaviour. In M.A. Ali (ed): *Photoreception and Vision in*

Invertebrates. New York, London: Plenum Press, pp. 523-559.

[57] Hengstenberg R., and Gtz K.G. (1967) Der Einflub des Schirmpigmentgehalts auf die Helligkeits- und Kontrastwahrnehmung bei Drosophila-Augenmutanten. *Kybernetik,* **3:** 276-285.

[58] Hengstenberg R. (1977) Spike response of non-spiking visual interneurone. *Nature,* **270:** 338-340.

[59] Hengstenberg R. (1982) Common visual response properties of giant vertical cells in the lobula plate of the blowfly *Calliphora. J. Comp. Physiol. A,* **149:** 179-193.

[60] Hengstenberg R., Hausen K., and Hengstenberg B. (1982) The number and structure of giant vertical cells (VS) in the lobula plate of the blowfly *Calliphora erytrocephala. J. Comp. Physiol. A,* **149:** 163-177.

[61] Horstmann W., Egelhaaf M., and Warzecha A.K. (2000) Synaptic interactions increase optic flow specificity. *Eur. J. Neurosci.,* **12:** 2157-2165.

[62] Kern R., Petereit C., and Egelhaaf M. (2001) Neural processing of naturalistic optic flow. *J. Neurosci.,* **21:** RC139.

[63] Krapp H.G., and Hengstenberg R. (1996). Estimation of self-motion by optic flow processing in single visual interneurons. *Nature,* **384:** 463-466.

[64] Krapp H.G., Hengstenberg R., and Egelhaaf M. (2001) Binocular contributions to optic flow processing in the fly visual system. *J. Neurophysiol.,* **85:** 724-734.

[65] Liu S.C. (2000) A neuromorphic VLSI model of global motion processing in the fly. *IEEE Trans. Pattern Anal. Machine Intell.,* **47:** 1458-1467.

[66] Maddess T. (1986) Afterimage-like effects in the motion-sensitive neuron H1. *Proc. R. Soc. Lond. B,* **228:** 433-459.

[67] Maddess T., and Laughlin S.B. (1985) Adaptation of the motion-sensitive neuron H1 is generated locally and governed by contrast frequency. *Proc. R. Soc. Lond. B,* **225:** 251-275.

[68] Meyer E.P., Matute C., Streit P., and Nssel D.R. (1986) Insect optic lobe neurons identifiable with monoclonal antibodies to GABA. *Histochemistry,* **84:** 207-216.

[69] Miyawaki A., Llopis J., Heim R., McCaffery J.M., Adams J.A., Ikura M., and Tsien R.Y. (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature,* **388:** 882-887.

[70] Oertner T.G., Single S., and Borst A. (1999) Separation of voltage- and ligand-gated calcium influx in locust neurons by optical imaging. *Neuroscience Letters,* **274:** 95-98.

[71] Oertner T.G., Brotz T., and Borst A. (2001) Mechanisms of dendritic calcium signaling in fly neurons. *J. Neurophysiol.,* **85:** 439-447.

[72] Poggio T., and Reichardt W. (1973) Considerations on models of movement detection. *Kybernetik,* **13**: 223-227.

[73] Reichardt W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication*, W.A. Rosenblith, (Ed.) The M.I.T. Press and John Wiley & Sons: New York, London, pp. 303-317.

[74] Reichardt W. (1987) Evaluation of optical motion information by movement detectors. *J. Comp. Physiol. A,* **161:** 533-547.

[75] Reichardt W., and Varju D. (1959) Uebertragungseigenschaften im Auswertesystem fr das Bewegungssehen. *Z. Naturforsch.,* **14b:** 674-689.

[76] Reichardt W., Poggio T., and Hausen K. (1983) Figure-ground discrimination by relative movement in the visual system of the fly. II. Towards the neural circuitry. *Bio. Cybern.,* **46:** 1-30.

[77] Reichardt W., Egelhaaf M., and Guo A. (1989). Processing of figure and background motion in the visual system of the fly. *Biol. Cybern.,* **61:** 327-345.

[78] Reisenman C., Haag J., and Borst A. (2003) Adaptation of response transients in fly motion vision. I. Experiments. *Vision Res.,* **14**: 1293-1309.

[79] de Ruyter van Steveninck R.R., Zaagman W.H., and Mastebroek H.A.K. (1986) Adaptation of transient responses of a movement-sensitive neuron in the visual system of the blowfly *Calliphora erytrocephala. Biol. Cybern.,* **53:** 451-463.

[80] Schuling F.H., Masterbroek H.A.K., Bult R., and Lenting B.P.M. (1989) Properties of elementary movement detectors in the fly *Calliphora erythrocephala. J. Comp. Physiol. A,* **165:** 179-192

[81] Segev I., Fleshman J.W., and Burke R.E. (1989) Compartmental models of complex neurons. In C. Koch and I. Segev (eds): *Methods in Neuronal Modeling: From Synapses to Networks*. Cambridge, London: MIT Press, pp. 63-96.

[82] Schilstra C., and van Hateren J.H. (1998) Stabilizing gaze in flying blowflies. *Nature,* **395:** 654-654.

[83] Single S., Haag J., and Borst A. (1997) Dendritic computation of direction selectivity and gain control in visual interneurons. *J. Neurosci.,* **17:** 6023-6030.

[84] Single S., and Borst A. (1998) Dendritic integration and its role in computing image velocity. *Science,* **281:** 1848-1850.

[85] Single S., and Borst A. (2002) Different mechanisms of calcium entry within different dendritic compartments. *J. Neurophys.,* **87:** 1616-1624.

[86] Strausfeld N.J., and Bassemir U.K. (1985) Lobula plate and ocellar interneurons converge onto a cluster of descending neurons leading to neck and leg motor neuropil in *Calliphora erythrocephala. Cell Tissue Res.,* **240:** 617-640.

[87] van Doorn A.J., and Koenderink J.J. (1982a) Temporal properties of the visual

detectability of moving spatial white noise. *Brain Res.,* **45:** 179-188.

[88] van Doorn A.J., and Koenderink J.J. (1982b) Spatial properties of the visual detectability of moving spatial white noise. *Brain Res.,* **45:** 189-195.

[89] van Santen J.P.H., and Sperling G. (1984) Temporal covariance model of human motion perception. *J. Opt. Soc. Am. A,* **1:** 451-473.

[90] van Santen J.P.H., and Sperling G. (1985) Elaborated Reichardt detectors. *J. Opt. Soc. Am. A,* **2:** 300-320.

[91] Wagner H. (1986a) Flight performance and visual control of flight of the free-flying housefly (*Musca domestica*). I: Organization of the flight motor. *Phil. Trans. R. Soc. Lond. B,* **312:** 527-551.

[92] Wagner H. (1986b) Flight performance and visual control of flight of the free-flying housefly *(Musca domestica).* II: Pursuit of targets. *Phil. Trans. R. Soc. Lond. B.,* **312:** 553-579.

[93] Wagner H. (1986c) Flight performance and visual control of flight of the free-flying housefly *(Musca domestica).* III: Interactions between angular movement induced by wide-and smallfield stimuli. *Phil. Trans. R. Soc. Lond. B,* **312:** 581-595.

[94] Warzecha K., and Egelhaaf M. (1996) Intrinsic properties of biological motion detectors prevent the optomotor control system from getting unstable. *Phil. Trans. R. Soc. Lond. B,* **351:** 1579-1591.

[95] Warzecha A.K., Egelhaaf M., and Borst A. (1993) Neural circuit tuning fly visual interneurons to motion of small objects. I. Dissection of the circuit by pharmacological and photo-inactivation techniques. *J. Neurophysiol.,* **69:** 329-339.

[96] Wilson H.R. (1985) A model for direction selectivity in threshold motion perception. *Biol. Cybern.,* **51:** 213-222.

[97] Zanker J.M. (1990) On the directional sensitivity of motion detectors. *Biol. Cybern.,* **62:** 177-183.

Chapter 15

Mean-Field Theory of Irregularly Spiking Neuronal Populations and Working Memory in Recurrent Cortical Networks

Alfonso Renart,[1] **Nicolas Brunel,**[2] **and Xiao-Jing Wang**[1]

[1] *Volen Center for Complex Systems, Brandeis University, Waltham, MA 02254, U.S.,* [2] *CNRS, Neurophysique et Physiologie du Système Moteur, Université Paris René Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France*

CONTENTS

15.1 Introduction

In cortical neural circuits, the biophysics of neurons and synapses and the collective network dynamics produce spatiotemporal spike patterns that presumably are optimized for the functional specialization of the system, be it sensory, motor or memory. Therefore, different systems might use different codes. For example, the 'spike timing code' or 'correlation code' that relies on precise spike timing is critical for the computation of coincidence detection in the brainstem auditory pathways, and may also contribute to information processing in other neural systems. A 'burst code' is prevalent in central pattern generators of the motor systems, where rhythmicity is produced by oscillatory repetition of brief clusters of spikes (bursts). Neurons can also signal information using a 'rate code', by virtue of the frequency at which the spikes are discharged. The idea of rate coding originated from the work of [3], who discovered that a stimulus feature (such as intensity) could be accurately read out from the firing rate of a sensory neuron. Since then, many studies have shown that firing rates convey a large amount of stimulus-related information in neurons.

In a small neural network, such as the visual system of flies or the electrosensory system of electric fish, there are a few synaptic connections per cell and each spike has a large impact on the post-synaptic cell. Hence spike timing is expected to be important. Moreover, a reliable estimate of the firing rate of one or a few pre-synaptic inputs requires a long-time average of spike counts and is, hence, not adequate to subserve fast perceptual or motor behaviors in these systems at fast time scales (∼ 100 milliseconds). The situation, however, is drastically different in a cortical circuit, where a huge number of neurons are available and organized into columns of functionally similar neurons [84]. A typical cortical neuron receives thousands of synapses, most of them from neighboring neurons [4, 76]; the impact of a single pre-synaptic spike onto a post-synaptic cell is relatively small. Moreover, spike trains of cortical neurons are highly stochastic and irregular (see e.g., [30, 108, 110], but see [59]), hence there is a lot of noise in spike timing. This fact raised the question of whether the observed spike train irregularity conveyed information or was rather a reflection of the various sources of noise present at the cellular and network levels [105]. Even if the spike times from single cells are noisy, information can still be conveyed in the average activity of pools of weakly correlated neurons. Suppose that a neuron receives connections from C_{cell} other neurons in a column. Being from the same column, the average activity of these inputs is similar, but since their spike trains are irregular the number $N_i(\Delta t)$ of spikes emitted by each cell i in the time interval $[t, t+\Delta t]$ is random. The total input to the post-synaptic neuron

$$f(t) \sim \sum_{i}^{C_{cell}} N_i(\Delta t)$$

provides an estimate of the average activity across the population. Since C_{cell} is large (100-1000) [17], and neurons are only weakly correlated [14, 31, 74, 130], noise can be largely (though not completely) averaged out [105, 106], and the estimate of the average activity of the pre-synaptic pool can be quite accurate even with a small Δt. In other words, the population firing rate can be defined (almost) instantaneously in real time. Moreover, such a rate code can be readily decoded by a post-synaptic cell: the summation of thousands of synaptic inputs provides a means to readout the population firing rate at any time.

Population firing rate models were introduced in the early 1970s and have since become widely popular in theoretical neuroscience. These models are described as non-linear differential equations, to which tools of mathematical analysis are applicable. Thus, concepts like attractor dynamics, pattern formation, synchronous network oscillations, etc, have been introduced in the field of neurobiology (See [37] for a review and references). Early models, such as associative memory models, were formulated in terms of firing-rates [27, 64]. Broadly speaking, two different approaches can be used to construct a firing-rate model. A rate model can be built *heuristically*: for example, a unit is assumed to have a threshold-linear or sigmoid input-output relation [5, 125, 126]. This class of rate models is valuable for its simplicity; important insights can be gained by detailed analysis of such models. The drawback is that these models tend to be not detailed enough to be directly related to electrophysiology. For example, the baseline and range of firing rates are arbitrarily defined so they cannot be compared with those of real neurons. It is therefore difficult to use the available data to constrain the form of such models. On the other hand, a firing-rate model can also be *derived*, either rigorously or approximately, from a spiking neuron model. To do that, the dynamics of spiking neurons must be well understood. The analytical study of the dynamics of spiking neuron models was pioneered by [68], and has witnessed an exponential growth in recent years. Up to date, most of the work was done with the leaky-integrate-and-fire (LIF) neuron model [1, 8, 11, 18, 19, 21, 24, 51, 68, 77, 88, 114, 120]. The LIF model is a simple spiking model that incorporates basic electrophysiological properties of a neuron: a stable resting potential, sub-threshold integration, and spikes. A network model can be constructed with LIF neurons coupled by realistic synaptic interactions. Such models have been developed and studied for many problems, such as synchronization dynamics, sensory information processing, or working memory. In some instances, firing-rate dynamics can be derived from the underlying spiking neuron models [26, 36, 40, 109]. These firing rate models provide a more compact description that can be studied in a systematical way.

Analytical studies of networks of neurons are usually performed in the context of 'mean-field' theories. In such theories, the synaptic input of a neuron in the network is traditionally only described by its average: the 'mean-field'. This first class of models is applicable to networks in which neurons are weakly coupled and fire in a regular fashion. More recently, mean-field theories have been introduced in which the synaptic inputs are described not only by their mean, but also by the fluctuations of their synaptic inputs, which come potentially both from outside the network, and from the recurrent inputs. This second class of models is applicable to strongly coupled networks in which neurons fire irregularly [11, 118, 119].

The objective of this chapter is to provide a pedagogical summary of this latter type of mean-field theory in its current state. We will introduce the theory in several steps, from single neurons, to self-consistent theory of a recurrent network with simple synapses, to realistic synaptic models. The chapter is organized into two parts, which address the two general ingredients of a mean-field theory for network models based on biophysics. First, a method is needed for the analytical description of a neuron's output in response to a large number of highly noisy pre-synaptic inputs, including realistic synaptic interactions (time course, voltage dependence) which are critical in determining the network behavior. This will be described in Section 15.2. Second, in a recurrent circuit, any neuron both receives inputs from, and sends output to, other neurons in the same network. Therefore, the pre-synaptic and post-synaptic firing rates are related to each other. The mean-field theory provides a procedure to calculate the neural firing rate in a self-consistent manner, in the steady-state. It also can also be extended to a description of the temporal dynamics of the neural firing rate. This will be discussed in Section 15.3. The self-consistent theory is then applied to a strongly recurrent network model of working memory which displays multi-stability between a resting state and memory-related persistent activity states.

15.2 Firing-rate and variability of a spiking neuron with noisy input

The first part of the present paper is devoted to the firing properties of a leaky integrate-and-fire (LIF) neuron in response to stochastic synaptic inputs. After the introduction of the LIF neuron, we proceed as follows: First, the statistical properties of the input current will be described, given certain assumptions about the stochastic activity of the pre-synaptic inputs to the neuron. Second, we will discuss the conditions under which the dynamics of the depolarization can be approximated by a diffusion equation. Third, we will show how to calculate the output mean firing rate and coefficient of variation (CV) of the cell given our assumptions. Next the effect of finite synaptic time constants will be explained. Finally, we provide a discussion on how realistic synaptic transmission, including voltage-dependent conductances and non-linear summation of inputs, can be incorporated into this framework.

15.2.1 The leaky integrate-and-fire neuron

In the LIF model, the voltage difference $V(t)$ across the membrane changes in response to an injected current $I(t)$ according to

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) + I(t), \tag{15.1}$$

where $C_m = 0.2$ nF is the total membrane capacitance, $g_L = 20$ nS is the leak conductance and $V_L = -70$ mV is the leak, or resting potential of the cell in the absence of input (see e.g., [69]). According to this equation, the membrane is seen as a simple RC circuit, with a time constant τ_m given by

$$\tau_m = \frac{C_m}{g_L} = 10 \text{ ms}. \tag{15.2}$$

Spiking is implemented in the model by defining a threshold voltage V_{th} such that the neuron is said to emit a spike at time t_{spk} whenever $V(t = t_{spk}) = V_{th} = -50$ mV. Refractoriness is taken into account by clamping the voltage to a reset value $V_r = -60$ mV for a time $\tau_{ref} = 2$ ms after each spike, i.e., if $V(t = t_{spk}) = V_{th}$, then $V(t') = V_r$ for $t' \in (t^+_{spk}, t_{spk} + \tau_{ref})$. When the neuron is inserted in a network, $I(t)$ represents the total synaptic current, which is assumed to be a linear sum of the contributions from each individual pre-synaptic cell.

15.2.2 Temporal structure of the afferent synaptic current

We will start with the simplest description of the interaction between the pre- and post-synaptic neurons. It amounts to assuming that each pre-synaptic spike causes an instantaneous change in post-synaptic voltage which is independent of the current value of this voltage, and depends only on a parameter J measuring the strength of the synapse (more precisely, J is the amount of positive charge entering the membrane due to the spike). If C neurons synapse onto this cell, each with an efficacy $J_i \quad (i = 1, \ldots, C)$, then the current into the cell can be represented as

$$I(t) = \sum_{i=1}^{C} J_i \sum_{j} \delta(t - t^i_j), \tag{15.3}$$

where t^i_j is the time of the j^{th} spike from the i^{th} pre-synaptic neuron. If the neuron is initially at rest, and a pre-synaptic cell fires a single spike at time $t = 0$, then by integrating Equation (15.1) one obtains

$$V(t) = V_L + \frac{J_i}{C_m} \exp\left(-\frac{t}{\tau_m}\right) \Theta(t), \tag{15.4}$$

where $\Theta(t)$ is the Heaviside function, $\Theta(t) = 0$ if $t < 0$ and 1 if $t > 0$. Thus, the post-synaptic potential (PSP) produced by each pre-synaptic spike consists of an instantaneous “kick” of size $\bar{J}_i = J_i/C_m$ followed by an exponential decay with time constant τ_m. For example, if the unitary charge is $J_i = 0.04$ pC, and $C_m = 0.2$ nF, then the kick size is $\bar{J}_i = 0.04/0.2 = 0.2$ mV.

We consider a neuron receiving synaptic input from a large pool of C_E excitatory and C_I inhibitory cells. We make two important assumptions regarding the activity of these inputs: first, that each of them fires spikes according to a stationary Poisson process, i.e., with a constant probability of emitting a spike per unit time. Second, that these Poisson processes are independent from cell to cell, i.e., the occurrence of a spike from any given cell does not give any information about the firing probability of any other neuron. These assumptions will need to be verified at the network level for the theory to be self-consistent (see Section 15.3).

We denote the average firing rate of each excitatory (inhibitory) input $j = 1, \ldots, C_{E,I}$, as ν_{E_j} (ν_{I_j}), and the efficacy of the corresponding excitatory (inhibitory) synapse as J_{E_j} (J_{I_j}). For simplicity, we first assume that all the rates and synapse from each pre-synaptic population are identical, i.e., $\nu_{E_j} = \nu_E$ and $J_{E_j} = J_E$ for all j, and similarly for the inhibitory population.

In this simple situation, the temporal average of the total current is constant in time and given by

$$< I(t) > \equiv \mu_C = \sum_{j=1}^{C_E} J_{E_j} \nu_{E_j} - \sum_{i=1}^{C_I} J_{I_i} \nu_{I_i} = C_E J_E \nu_E - C_I J_I \nu_I. \tag{15.5}$$

For a Poisson process $s(t)$ of rate ν, $< (s(t)-\nu)(s(t')-\nu) >= \nu\delta(t-t')$. Thus, using the fact that the inputs are Poisson and independent, the connected two point correlation function of the total current is given by

$$\begin{aligned} < (I(t)- < I >)(I(t')- < I >) > &= \left[\sum_j^{C_E} J_{E_j}^2 \nu_{E_j} + \sum_i^{C_I} J_{I_i}^2 \nu_{I_i} \right] \delta(t-t') \\ &= (C_E J_E^2 \nu_E + C_I J_I^2 \nu_I)\delta(t-t') \\ &\equiv \sigma_C^2 \delta(t-t'). \end{aligned} \tag{15.6}$$

15.2.3 The diffusion approximation

In principle, the next step would be to solve the dynamics of the depolarization as described in Equation (15.1) in the presence of the stochastic current $I(t)$. As it is, this task is still too difficult, so we will make one further approximation, namely to replace the *point* process $I(t)$, by a process $\bar{I}(t)$ with the same mean and two-point correlation function as $I(t)$, such that the voltage response $V(t)$ to $\bar{I}(t)$ becomes continuous (instead of discrete as a result of the synaptic kicks) in time. The idea is to make the size of the voltage kicks $\bar{J}_{E,I} \equiv J_{E,I}/C_m$ small, while at the same time increasing their overall frequency by increasing $C_{E,I}$ (notice that since the sum of two Poisson processes is another Poisson process, $I(t)$ can be considered the difference of two Poisson processes of rates $C_E \nu_E$ and $C_I \nu_I$ respectively). For a cortical neuron, since it receives a large number of pre-synaptic contacts, each of which contributes only to a small fraction of the voltage distance between rest and threshold, one expects this approximation to be plausible and give accurate results.

Since the inputs to our cell are assumed to be stochastic, the temporal evolution of $V(t)$ is probabilistic. The fundamental object for the description of the dynamics of the membrane potential is the probability density $\rho(V,t|V_0,t_0)$ for $V(t) \in [V, V+dV]$ given that $V(t_0) = V_0$. If we consider our averages to be carried out over an *ensemble* of identical neurons, each with a different realization of the stochasticity, $\rho(V,t|V_0,t_0)$ can be considered a "population" density, so that $\rho(V,t|V_0,t_0)dV$ is the fraction of neurons among the ensemble with membrane potentials in $[V, V+dV]$ given that all neurons were at V_0 at $t = t_0$. In the Appendix, we present an intuitive derivation of a differential equation which governs the temporal evolution of

$\rho(V,t|V_0,t_0)$ in the presence of the stochastic input $I(t)$ (see e.g., [49, 54, 97, 99] for more details). Such equation reads

$$\frac{\partial}{\partial t}\rho(V,t|V_0,t_0) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial V^n}[A_n\, \rho(V,t|V_0,t_0)], \tag{15.7}$$

where

$$\begin{aligned} A_1(V) &= -\frac{(V-V_L)}{\tau_m} + (\bar{J}_E C_E \nu_E - \bar{J}_I C_I \nu_I) = \\ &= -\frac{(V-V_L)}{\tau_m} + \frac{\mu_C}{C_m} \equiv -\frac{(V-V_{ss})}{\tau_m} \\ A_2 &= (\bar{J}_E{}^2 C_E \nu_E + \bar{J}_I{}^2 C_I \nu_I) = (\frac{\sigma_C}{C_m})^2 \equiv \frac{\sigma_V^2}{\tau_m} \\ A_n &= (\bar{J}_E{}^n C_E \nu_E + (-1)^n \bar{J}_I{}^n C_I \nu_I) \qquad n = 3,4,\ldots \end{aligned} \tag{15.8}$$

where A_n are the infinitesimal moments of the stochastic process. The infinitesimal moments completely specify the dynamics of $\rho(V,t|V_0,t_0)$. The drift coefficient A_1 captures the deterministic component of the temporal evolution of $V(t)$; $V_{ss} = V_L + \mu_C/g_L$ is the steady-state voltage in the absence of stochasticity. The diffusion coefficient A_2 measures the fluctuations of $V(t)$. In the absence of threshold, the variance of the depolarization is $\sigma_V^2/2 = \sigma_C^2\tau_m/(2C_m^2)$.

In what is referred to as the diffusion approximation, A_n for $n > 2$ are assumed to be negligible and set to zero [97, 117]. Looking at Equations (15.8), one can see under which conditions this will be a valid approximation. Since the infinitesimal moments depend on powers of the kick size times their overall rate, one expects the approximation to be appropriate if the kick size is very small but the overall rate is very large, in such a way that the size of all moments of order higher than two become negligible in comparison with the drift and diffusion coefficients. In particular, in the next sections we show how, in the limit of infinitely large networks, if the synaptic efficacies are scaled appropriately with the network size, the approximation can become exact.

We will for now take it for granted, and focus on the properties of Equation (15.7) when only the first two infinitesimal moments are non-zero. The resulting equation is called the Fokker-Planck equation for $\rho(V,t|V_0,t_0)$, and reads

$$\frac{\partial}{\partial t}\rho(V,t|V_0,t_0) = \frac{\partial}{\partial V}[\frac{(V-V_{ss})}{\tau_m}\rho(V,t|V_0,t_0)] + \frac{\sigma_V^2}{2\tau_m}\frac{\partial^2}{\partial V^2}[\rho(V,t|V_0,t_0)]. \tag{15.9}$$

The process described by this equation, characterized by a constant diffusion coefficient $D = \sigma_V^2/(2\tau_m)$ and a linear drift, is called the Ornstein-Uhlenbeck (O-U) process (see e.g., [123]). It describes the temporal evolution of $V(t)$ when the input to the neuron is no longer $I(t)$, but

$$\bar{I}(t) \equiv \mu_C + \sigma_C \eta(t), \tag{15.10}$$

where $\eta(t)$ is called a *white noise* process. It can be defined heuristically as a random variable taking values

$$\eta(t) = \lim_{dt \to 0} \mathbf{N}(0, \frac{1}{\sqrt{dt}}) \tag{15.11}$$

for all t independently, where we have defined $\mathbf{N}(\alpha, \beta)$ is a Gaussian random variable of mean α and variance β^2. The mean and two-point correlation function of the white noise process are therefore, $< \eta(t) >= 0$ and $< \eta(t)\eta(t') >= \delta(t-t')$ respectively. In effect, we are now replacing Equation (15.1) by

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) + \mu_C + \sigma_C \eta(t), \tag{15.12}$$

or

$$\tau_m \frac{dV(t)}{dt} = -(V(t) - V_{ss}) + \sigma_V \sqrt{\tau_m} \eta(t). \tag{15.13}$$

This is called the white-noise form of the Langevin equation of the process $V(t)$. It has the appeal that it is written as a conventional differential equation so that the dynamics of $V(t)$ is described in terms of its sample paths, rather than in terms of the temporal evolution of its probability distribution, as in the Fokker-Planck Equation (15.9). In general, the practical use of the Langevin equation is that it provides a recipe for the numerical simulation of the sample paths of the associated process. Developing Equation (15.13) to first order one obtains

$$V(t+dt) = (1 - \frac{dt}{\tau_m})V(t) + V_{ss}\frac{dt}{\tau_m} + \sigma_V \sqrt{\frac{dt}{\tau_m}} \mathbf{N}(0,1). \tag{15.14}$$

Assuming that dt/τ_m is small but finite, Equation (15.14) provides an iterative procedure which gives an *approximate* description of the temporal evolution of $V(t)$. This scheme is general and can be used for any diffusion process. For the O-U process in particular, in the absence of threshold Equation (15.9) can be solved exactly. The population density of this process is a Gaussian random variable with a time-dependent mean and variance [97, 123], so that

$$\begin{aligned}\rho(V,t|V_0,t_0) = \mathbf{N}\Big(&V_{ss} + (V_0 - V_{ss})\exp(-\tfrac{t-t_0}{\tau_m}), \tfrac{\sigma_V}{\sqrt{2}} \\ &\left[1 - \exp(-\tfrac{2(t-t_0)}{\tau_m})\right]^{1/2}\Big).\end{aligned} \tag{15.15}$$

Using this result one can find an exact iterative procedure for the numerical simulation of the process. Assuming V_0 is the value of the depolarization in the sample path at time t, e.g., $V_0 = V(t)$, the depolarization at a latter time $t + \Delta t$ will be

$$\begin{aligned}V(t+\Delta t) = {}& V_{ss} + (V(t) - V_{ss})\exp(-\tfrac{\Delta t}{\tau_m}) \\ &+ \tfrac{\sigma_V}{\sqrt{2}} \left[1 - \exp(-\tfrac{2\Delta t}{\tau_m})\right]^{1/2} \mathbf{N}(0,1).\end{aligned} \tag{15.16}$$

This update rule is exact for all Δt [54].

In Figure 15.1, sample paths of $V(t)$ in the presence of the original input current $I(t)$ obtained by numerical integration of Equation (15.1) are compared with sample paths in the presence of the effective input $\bar{I}(t)$, obtained using Equation (15.16). As illustrated in Figure 15.1, $\tau_{ref} = 2$ ms after emitting a spike, $V(t)$ begins to integrate its inputs again starting from V_r until it reaches V_{th}. The first time $V(t)$ reaches V_{th} is called the 'first-passage time' (denoted by T_{fp}). Taking the refractory period into account, the whole interval between consecutive spikes is called the inter-spike interval (ISI). Therefore, the statistics of ISIs can be analyzed using the theory of first-passage times of the Ornstein-Uhlenbeck process [97, 117].

15.2.4 Computation of the mean firing rate and CV

The Fokker-Planck Equation (15.9) can be rewritten as a continuity equation by defining

$$S(V,t|V_0,t_0) \equiv -\frac{(V - V_{ss})}{\tau_m}\rho(V,t|V_0,t_0) - \frac{\sigma_V^2}{2\tau_m}\frac{\partial}{\partial V}\rho(V,t|V_0,t_0)], \qquad (15.17)$$

so that Equation (15.9) becomes

$$\frac{\partial}{\partial t}\rho(V,t|V_0,t_0) = -\frac{\partial}{\partial V}S(V,t|V_0,t_0). \qquad (15.18)$$

Thus, $S(V,t|V_0,t_0)$ is the flux of probability (or probability current) crossing V at time t. To proceed, a set of boundary conditions on t and V has to be specified for $\rho(V,t|V_0,t_0)$. First one notices that, if a threshold exists, then the voltage can only be below threshold and can only cross it from below (the threshold is said to be an absorbing barrier). The probability current at threshold gives, by definition, the average firing rate of the cell. Since $\rho(V > V_{th},t|V_0,t_0) = 0$, the probability density must be zero at $V = V_{th}$, otherwise the derivative would be infinite at $V = V_{th}$ and so would be the firing rate according to Equation (15.17). Therefore, we have the following boundary conditions

$$\rho(V_{th},t|V_0,t_0) = 0 \qquad \text{and} \qquad \frac{\partial}{\partial V}\rho(V_{th},t|V_0,t_0) = -\frac{2\nu(t)\tau_m}{\sigma_V^2}, \qquad (15.19)$$

for all t. The conditions at $V = -\infty$ ensure that the probability density vanishes fast enough to be integrable, i.e.,

$$\lim_{V\to-\infty}\rho(V,t|V_0,t_0) = 0 \qquad \text{and} \qquad \lim_{V\to-\infty}V\rho(V,t|V_0,t_0) = 0. \qquad (15.20)$$

Since the threshold is an absorbing boundary, a finite probability mass is constantly leaving the interval $(-\infty, V_{th})$. Under this condition, there is no stationary distribution for the voltage, i.e., $\rho(V,t|V_0,t_0) \to 0$ as $t \to \infty$. In order to study the steady-state of the process, one can keep track of the probability mass leaving the integration interval at t, and re-inject it at the reset potential at $t + \tau_{ref}$. This injection

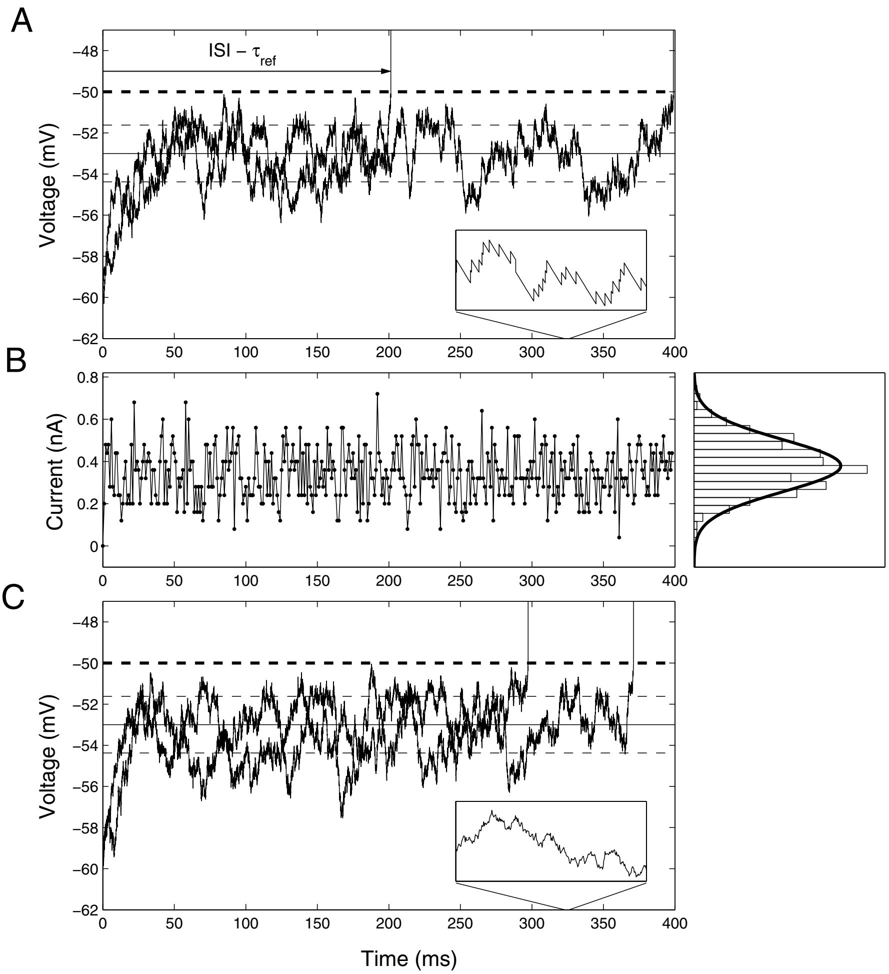

Figure 15.1

Leaky integrate-and-fire neuron model in response to stochastic inputs. **(A)**. Sample paths of the membrane potential $V(t)$ in response to a stochastic current $I(t)$ obeying Equation (15.3) with Poisson spike trains, with $\bar{J}_E = \bar{J}_I = 0.2$ mV, $C_E = C_I = 1000$, $\nu_E = 9$ Hz, $\nu_I = 0.5$ Hz. The resulting steady-state voltage is $V_{ss} = -53$ mV (thin solid line) with a standard deviation of $\sigma_V^2/2 = 1.38$ mV (thin dotted line). Three sample paths (different colors) are shown from the moment when the neuron starts to integrate its inputs ($\tau_{ref} = 2$ ms after the previous spike) until $V(t)$ reaches threshold for the first time. The time it takes for this to happen is called the first passage time and the total time in between two consecutive spikes is the inter-spike interval (ISI). Threshold ($V_{th} = -50$ mV) is shown as a thick dashed line (Continued).

Figure 15.1

Inset: Snapshot of the blue sample path from 323 to 326 ms shows the 0.2 mV discontinuities in $V(t)$ due to the synaptic inputs. **(B)** Current in top panel averaged over 1 ms time bins. Each point represents the average current into the neuron in the previous millisecond. For clarity of presentation, consecutive points are joined by lines. **Right**. Histogram of currents from left panel. The smooth blue line represents the distribution of $(1/\Delta t)\int_t^{t+\Delta t}\bar{I}(t')dt'$, where $\bar{I}(t)$ is the current into the cell in the diffusion approximation, Equation (15.10), and $\Delta t = 1$ ms. **(C)** Same as **A**, but with inputs now described by the diffusion approximation. The macroscopic structure of the sample paths is very similar. The differences between the Poisson input and the diffusion approximation can only be appreciated by looking at the inset.

of probability represents an extra probability current $S^{reset}(V,t)$, that adds to the current $S(V,t|V_0,t_0)$ associated to the sub-threshold dynamics of $V(t)$. Taking this into account, one can rewrite the Fokker-Planck equation like

$$\frac{\partial}{\partial t}\rho(V,t|V_0,t_0) = -\frac{\partial}{\partial V}[S(V,t|V_0,t_0) + S^{reset}(V,t)]. \tag{15.21}$$

Since this injection only results in a change of probability mass in $V = V_{reset}$, the new current is given by

$$S^{reset}(V,t) = \nu(t-\tau_{ref})\Theta(V - V_{reset}). \tag{15.22}$$

To find the solution for the steady-state distribution $\rho_{ss}(V)$, we insert expression (15.22) into the Fokker-Planck equation (15.21), and look for time independent solutions by setting the left hand side of this equation to zero,

$$\frac{\partial}{\partial V}[\frac{(V-V_{ss})}{\tau_m}\rho_{ss}(V)] + \frac{\sigma_V^2}{2\tau_m}\frac{\partial^2}{\partial V^2}\rho_{ss}(V) = -\nu\,\delta(V - V_{reset}). \tag{15.23}$$

Solving this equation with the boundary conditions (15.19-15.20), one obtains the following expression for the steady-state distribution [24]

$$\rho_{ss}(V) = \frac{2\nu\tau_m}{\sigma_V}\exp\left(-\frac{(V-V_{ss})^2}{\sigma_V^2}\right)\int_{\frac{V-V_{ss}}{\sigma_V}}^{\frac{V_{th}-V_{ss}}{\sigma_V}}\Theta(x - \frac{V_r - V_{ss}}{\sigma_V})e^{x^2}dx. \tag{15.24}$$

The function $\rho_{ss}(V)$ gives the fraction of cells in a non-refractory state with depolarizations in $(V, V+dV)$ in the steady state. Taking into account also the fraction $\nu\tau_{ref}$ of neurons in a refractory state, the steady state firing rate ν can be found by the normalization condition

$$\int_{-\infty}^{V_{th}}\rho_{ss}(V)dV \; + \; \nu\tau_{ref} = 1. \tag{15.25}$$

Plugging expression (15.24) into this equation and solving for ν one gets

$$\frac{1}{\nu} = \tau_{ref} + \tau_m\sqrt{\pi}\int_{\frac{V_r-V_{ss}}{\sigma_V}}^{\frac{V_{th}-V_{ss}}{\sigma_V}} e^{x^2}(1+\text{erf}(x))dx, \tag{15.26}$$

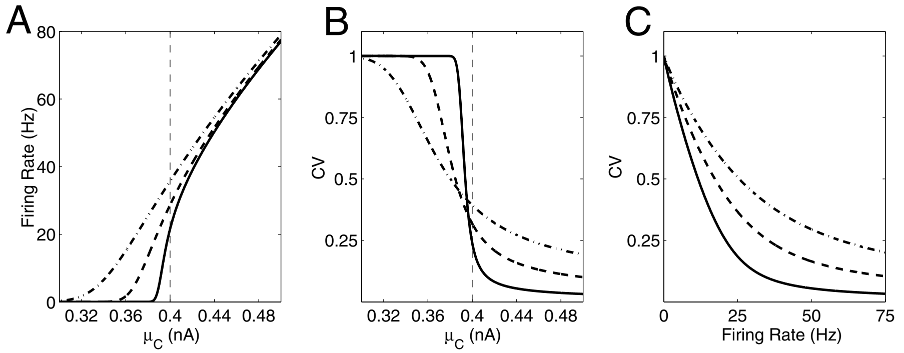

Figure 15.2

Firing rate **(A)** and CV **(B)** of the LIF neuron as a function of the mean current μ_C for three values of the effective standard deviation in the voltage $\sigma_V = 0.1$ mV (solid), 1 mV (dashed) and 4 mV (dot-dashed). **(C)** CV as a function of the mean firing rate when μ_C is varied as in the three curves in **A-B**. The parameters of the cell are $C_m = 0.2$ nF, $g_L = 20$ nS ($\tau_m = 10$ ms), $\tau_{ref} = 2$ ms, $V_L = -70$ mV, $V_{th} = -50$ mV and $V_r = -60$ mV.

where $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{u^2} du$.

Once the firing rate is known, the following recursive relationship between the moments of the first-passage time distribution of the Ornstein-Uhlenbeck process (see e.g., [117]) can be used to find $< T_{fp}^2 >$

$$\frac{\sigma_V^2}{2}\frac{d^2 < T_{pf}^k >}{dx^2} + (V_{ss} - x)\frac{d < T_{fp}^k >}{dx} = -k < T_{fp}^{k-1} >, \qquad (15.27)$$

where $x = V_r$. Given that $< T_{fp} >= 1/\nu$ in Equation (15.26), the CV of the ISI is given by [19]

$$CV^2 \equiv \frac{< T_{fp}^2 > - < T_{fp} >^2}{< T_{fp} >^2} = 2\pi\nu^2 \int_{\frac{V_r - V_{ss}}{\sigma_V}}^{\frac{V_{th} - V_{ss}}{\sigma_V}} dx e^{x^2} \int_{-\infty}^{x} dy e^{y^2}(1 + \mathrm{erf}(y)). \quad (15.28)$$

In Figures 15.2A-B we plot the mean firing rate and CV of the LIF neuron as given by Equations (15.26,15.28), as a function of the mean current μ_C for various values of the effective standard deviation in the voltage σ_V. The firing rate (Figure 15.2A) is a monotonic increasing function of the average current. Qualitatively, it starts to rise when the average current comes within a standard deviation of the current threshold, defined as $I_{th} \equiv g_L(V_{th} - V_L)$, and shown as a vertical dashed line in Figure 15.2A. It increases supra-linearly with μ_C for sub-threshold mean currents, and sub-linearly when the mean current is above threshold, eventually saturating at $1/\tau_{ref}$. Therefore, for a wide range of values of σ_V, $\mu_C \sim I_{th}$ is close to the point where the curvature of $\nu(\mu_C)$ changes sign. The deterministic current threshold I_{th} also marks the transition between two different behaviors of the CV (Figure 15.2B). When the mean input is sub-threshold ($\mu_C < I_{th}$), spike discharge is triggered by fluctuations in the current, and spike trains are irregular. Therefore, in this regime, the CV is high, close to one. For supra-threshold mean current ($\mu_C > I_{th}$), the CV decays to zero and spiking becomes regular. The sharpness of this transition depends on σ_V. When the fluctuations are small, the transition is very sharp, so as soon as the neuron starts firing, it does so in a regular fashion. For large values of σ_V the transition is smooth (and if σ_V is large enough the CV can first increase noticeably from one for $\mu_C < I_{th}$, not shown), so the neuron fires initially irregularly and becomes progressively more regular as μ_C becomes much larger than I_{th}. In Figure 15.2C, the CV is plotted as a function of the firing rate for three values of σ_V as μ_C is increased gradually. In general, when the firing rate increases as a result of an increase in the mean current, the CV decreases. This decrease is faster with smaller fluctuations.

15.2.5 Effect of synaptic time constants

So far, we have assumed that the post-synaptic currents (PSPCs) are delta-functions, without any duration or temporal kinetics. In reality, synaptic currents rise and decay with time constants that range from 1 ms to several hundred ms. To incorporate a finite time constant of post-synaptic currents into the theoretical framework described in the previous section, we consider a variable $I(t)$ which, upon arrival of a spike at t_{spk}, evolves according to

$$\tau_{syn}\frac{dI(t)}{dt} = -I(t) + J\delta(t - t_{spk}). \qquad (15.29)$$

For non-zero τ_{syn}, it is now $I(t)$ that has a discontinuous jump of size J/τ_{syn} when a pre-synaptic spike arrives. For $t > t_{spk}$, $I(t)$ decays exponentially back to zero with a time constant τ_{syn}. Importantly, the total area under this PSC is J independently of the value of τ_{syn}. The previous scheme can therefore be recovered continuously by letting $\tau_{syn} \to 0$. Now, instead of injecting an instantaneous charge J for every spike, we spread this same amount of charge over a time τ_{syn}. The effect of this on the voltage is to smoothen the rise of the PSP. Now the PSP is continuous and given by

$$V(t) = V_L + \left(\frac{J}{C_m}\right)\frac{\tau_m}{\tau_m - \tau_{syn}}\left[\exp(-\frac{t}{\tau_m}) - \exp(-\frac{t}{\tau_{syn}})\right]\Theta(t - t_{spk}), \qquad (15.30)$$

with a rise time τ_{syn} and a decay time τ_m. If the membrane and the synaptic time constants are equal, the PSP has the shape of an α-function

$$V(t) = V_L + \left(\frac{J}{C_m}\right)\frac{t}{\tau_m}\exp(-\frac{t}{\tau_m})\Theta(t - t_{spk}). \qquad (15.31)$$

The most important effect of a non-zero synaptic constant on the Fokker-Planck scheme presented above is the appearance of temporal correlations in the afferent current. The total synaptic input becomes a process with mean $< I >= \mu_C$, and an exponential two-point correlation function $C_C(t,t') \equiv < (I(t) - < I >)(I(t') - < I >) >$ given by

$$C_C(t,t') = (\sigma_C^2 / 2\tau_{syn}) \exp(-|t-t'|/\tau_{syn}). \tag{15.32}$$

Using once again the diffusion approximation to replace the input to $I(t)$ by a Gaussian process with the same mean and correlation function, and defining $\delta I(t) = I(t) - \mu_C$, the Langevin equations of the process now read

$$\tau_m \frac{dV(t)}{dt} = -(V(t) - V_{ss}) + \frac{\delta I(t)}{g_L} \tag{15.33}$$

$$\tau_{syn} \frac{d}{dt} \delta I(t) = -\delta I(t) + \sigma_C \eta(t). \tag{15.34}$$

Although $V(t)$ is not Markovian anymore (knowledge of $I(t)$, in addition to $V(t)$, is needed to determine $V(t+dt)$ probabilistically), $V(t)$ and $I(t)$ *together* constitute a bi-variate Markov process [54, 99]. From Equations (15.33, 15.34) one can therefore derive a Fokker-Planck equation characterizing the evolution in time of the joint probability of V and I. However, the presence of temporal correlations in $I(t)$ makes the calculation of the firing rate much more involved than for the simple Ornstein-Uhlenbeck case and, indeed, the mean first-passage time can only be obtained in the case where $\tau_{syn} \ll \tau_m$, using perturbation theory on the parameter $k \equiv \sqrt{\tau_{syn}/\tau_m} \ll 1$ [23, 34, 40, 61, 67]. We present here only the final result: the firing rate is given by

$$\nu_{syn}(k) = \nu + k\alpha\sigma_V \left(\frac{\partial \nu}{\partial V_{th}} + \frac{\partial \nu}{\partial V_r} \right) + O(k^2), \tag{15.35}$$

where ν is the firing rate of the white noise case, Equation (15.26),

$$\alpha = -\zeta(1/2)/\sqrt{2} \sim 1.03$$

and ζ is the Riemann zeta function [2]. Note that the firing rate calculated in [23] does not include the term proportional to $\partial \nu / \partial V_r$, because of the approximation made in that paper, namely the neuron was assumed to be in the sub-threshold regime, in which the dependency of the mean firing rate on the reset potential is very weak.

Another way to write Equation (15.35) is to replace the threshold V_{th} and V_r in the expression for the mean first-passage time obtained for a white noise current (15.26), by the following effective k-dependent expressions

$$V_{th}^{eff} = V_{th} + \sigma_V \alpha k \tag{15.36}$$

$$V_r^{eff} = V_r + \sigma_V \alpha k. \tag{15.37}$$

This first order correction is in good agreement with the results from numerical simulations for $\tau_{syn} < 0.1\ \tau_m$. To extend the validity of the result to larger values of τ_{syn}, a second order correction can be added to the effective threshold, with coefficients determined by a fit to numerical simulations with values of τ_{syn} up to τ_m

[23]. A finite synaptic time constant leads to synaptic filtering of the pre-synaptic inputs which, in general, leads to a reduction in post-synaptic firing rates [23]. This effect is more pronounced for sub-threshold mean currents, since in this regime the neuronal firing results from the fluctuations in the current which can be filtered out by the synapse. Note that the effect of increasing τ_{syn} is to spread out in time the same amount of charge influx into the cell. Since charge is constantly leaking out of the cell membrane, the longer τ_{syn}, the lower the overall magnitude of the voltage fluctuations.

Finally, one can also compute perturbatively the firing rate in the large synaptic time constant limit [83]. An interpolation between the two limits gives rather accurate results in the whole range of all synaptic time constants. A similar approach has been used to compute the firing rate of another simple spiking neuron, the quadratic neuron [22].

15.2.6 Approximate treatment of realistic synaptic dynamics

Real synaptic currents can depart in at least three ways from the currents considered until now: (i) individual post-synaptic currents can in some circumstances sum non-linearly, due to receptor saturation; (ii) post-synaptic currents are voltage dependent, because synaptic activation occurs as conductance change rather than current increase, and because the maximal conductance can itself be voltage-dependent; (iii) multiple synaptic time scales are present, due to the different kinetics of the AMPA, GABA$_A$, and NMDA receptors. We first describe the standard biophysical model for describing post-synaptic currents (see also e.g., [33, 120]), and then discuss separately how the three issues can be dealt with using approximate treatments.

15.2.6.1 Biophysical models of post-synaptic currents

Synaptic activation opens ion channels in the post-synaptic cell. The amount of current flowing through these channels is proportional to the product of the number of open channels times the driving force of the synaptic current:

$$I_{syn}(t) = g_{syn}(V)\, s(t)(V(t) - V_{syn}), \tag{15.38}$$

where $g_{syn}(V)$ is the (possibly voltage-dependent) maximal conductance, $s(t)$ is a gating variable measuring the fraction of open channels at the synapse and V_{syn} is the synaptic reversal potential. The term $V - V_{syn}$ is the driving force of the synapse, and it determines its polarity, i.e., whether a synaptic current is depolarizing ($V - V_{syn} < 0$) or hyper-polarizing ($V - V_{syn} > 0$). In the presence of a driving force term, all synaptic inputs are voltage-dependent.

We consider two types of kinetic schemes for the gating variable $s(t)$. If the underlying dynamics of the synaptic channels is fast compared with the typical firing rates of the spike trains at the synapse, the synapse is usually far from saturation and a linear kinetic scheme is appropriate. Additionally, in this situation the rise time of the post-synaptic currents (PSCs) is so fast that it can be considered instantaneous, so that the kinetics can also be approximated by a first order system, i.e.,

$$\frac{ds(t)}{dt} = -\frac{s(t)}{\tau_s} + \sum_k \delta(t - t_k), \tag{15.39}$$

where $\sum_k \delta(t - t_k)$ represents the pre-synaptic spike train arriving at the synapse. The average fraction of open channels is linear in the firing rate ν across the synapse $\bar{s} = \tau_s \nu$. Since $s(t)$ is a fraction, i.e., necessarily less than one, this description is appropriate as long as $\nu \ll 1/\tau_s$. We will use it for the description of GABA$_A$R- and AMPAR-mediated transmission, which have synaptic time constants of $\tau_{\text{GABA}_\text{A}} = 10$ ms, and $\tau_{\text{AMPA}} = 2$ ms [13, 15, 58, 72, 127, 129]. This approximation is reasonable if $\nu < 1/\tau_{\text{GABA}_\text{A}} = 100$ Hz.

If the underlying channel dynamics is of the order of, or slower, than the typical inter-spike intervals of the spike trains crossing the synapse (which is the case for the NMDAR-mediated PSPcs, with a time constant of $50 - 100$ ms), the channels decay slowly in between spikes, and a few spikes in a train at hight frequencies can recruit a fraction of open channels close to unity. In this case, the effect of subsequent spikes is bounded by the saturation of all post-synaptic receptors, hence spikes sum non-linearly. Also, for slow channel dynamics, the PSC rise times are on the order of the fastest time-scales of the system (a few milliseconds), and can no longer be neglected. A non-linear, second order scheme, provides an accurate description of the kinetics of the gating variable $s(t)$ in these conditions:

$$\frac{ds(t)}{dt} = -\frac{s(t)}{\tau^{decay}} + \alpha x(t)(1 - s(t)) \tag{15.40}$$

$$\frac{dx(t)}{dt} = -\frac{x(t)}{\tau^{rise}} + \sum_k \delta(t - t_k). \tag{15.41}$$

For slow synaptic dynamics, the average gating variable is no longer a linear function of the pre-synaptic rate unless the firing rate is only a few Hz.

15.2.6.2 Average gating variable vs. rate in the non-linear model

An immediate consequence of this non-linear summation is that the average value of the gating variable becomes a non-linear function of the average firing rate of the spike train through the synapse. This function depends on the statistics of the spike train. If the spike train is regular, an approximation can be obtained by replacing $\sum_k \delta(t - t_k)$ by the mean firing rate ν of the spike train in Equation (15.41). In this case the average of the gating variable becomes

$$\bar{s} = \frac{\tau\nu}{1 + \tau\nu}, \tag{15.42}$$

where the effective time constant is equal to $\tau = \tau^{rise}\tau^{decay}\alpha$.

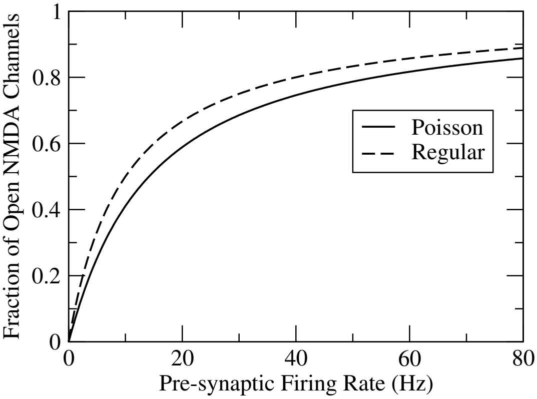

Figure 15.3

Average fraction of open NMDA channels as a function of the pre-synaptic firing rate. The solid (dashed) line, calculated with Equation (15.43) (Equation (15.42)), corresponds to the case when the spike train at the synapse is Poisson (periodic).

If the spike train is Poisson, the expression for $\bar{s}$ is [25]

$$\bar{s} = \frac{\nu\tau}{1+\nu\tau}\left(1 + \frac{1}{1+\nu}\sum_{n=1}^{\infty}\frac{(-\alpha\tau^{rise})^n T_n}{(n+1)!}\right) \equiv \psi(\nu)$$

$$T_n = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{\tau^{rise}(1+\nu\tau)}{\tau^{rise}(1+\nu\tau)+k\tau^{decay}}. \qquad (15.43)$$

We will use this description for NMDAR-mediated transmission, with parameters $\tau^{decay}_{\text{NMDA}} = 100$ ms, $\tau^{rise}_{\text{NMDA}} = 2$ ms, and $\alpha = 0.5$ KHz. The effective NMDA time constant is thus $\tau_{\text{NMDA}} = 100$ ms. In Figure 15.3, the average gating variable of an NMDA synapse is plotted as a function of the average pre-synaptic firing rate ν, for the case of a regular and a Poisson input spike train. Note that, due to the saturation term on the right of Equation (15.40), the gating variable starts to saturate when the pre-synaptic rate becomes larger than $\sim 1/\tau_{\text{NMDA}} \sim 10$ Hz.

15.2.6.3 Voltage-dependence of the post-synaptic currents

A post-synaptic current is voltage-dependent because of its driving force; in addition the maximal conductance can also be voltage-dependent, as in the case of the NMDA channels [87].

In general, even if the maximal conductance does not depend on the voltage, the voltage dependence induced by the driving force term in the unitary synaptic current (15.38) modifies the previous framework for calculating the output firing rate of the cell in several ways.

Let us separate the time course of the gating variable into a deterministic component, associated to its temporal average (we assume stationary inputs, in which case the average of the gating variable is constant), and a fluctuating component due to the stochastic nature of the spike trains in our model, i.e., $s_{syn}(t) = \bar{s}_{syn} + \delta s_{syn}(t)$. The unitary synaptic current, Equation (15.38), now becomes

$$I_{syn}(t) = g_{syn}\bar{s}_{syn}(V(t) - V_{syn}) + g_{syn}\delta s_{syn}(t)(V(t) - V_{syn}). \tag{15.44}$$

The main complication due to the driving force is that now the fluctuating component of the synaptic current (second term in the right-hand side of Equation (15.44)) becomes voltage dependent. This multiplicative dependence of the fluctuations on the membrane potential renders a rigorous treatment of the fluctuations in the current difficult. To avoid this complication, we replace the voltage by its average $\bar{V}$ in the driving force for the fluctuating component of the synaptic current, so that

$$I_{syn}(t) \sim g_{syn}\bar{s}_{syn}(V(t) - V_{syn}) + g_{syn}\delta s_{syn}(t)(\bar{V} - V_{syn}). \tag{15.45}$$

The deterministic part of the current $g_{syn}\bar{s}_{syn}(V(t) - V_{syn})$ can be dealt with easily by noting that $g_{syn}\bar{s}_{syn}$ can be absorbed in the leak conductance, and $g_{syn}\bar{s}_{syn}V_{syn}$ can be absorbed in the resting membrane potential V_L.

The resulting effect on neuronal properties is an increase in the total effective leak conductance of the cell

$$g_L \longrightarrow g_L + g_{syn}\bar{s}_{syn}, \tag{15.46}$$

which is equivalent to a decrease of the membrane time constant from $\tau_m = C_m/g_L$, to $\tau_m^{eff} = C_m/(g_L + g_{syn}\bar{s}_{syn}) = \tau_m/\alpha_{\tau_m}$. Thus, the synaptic input makes the neuron leakier by a factor equal to the relative increase in conductance due to synaptic input ($\alpha_{\tau_m} = 1 + \bar{s}_{syn}g_{syn}/g_L$). The resting (or steady-state) membrane potential is also re-normalized

$$V_L \longrightarrow \frac{g_L V_L + g_{syn}\bar{s}_{syn}V_{syn}}{g_L + g_{syn}\bar{s}_{syn}}, \tag{15.47}$$

and becomes a weighted average of the different reversal potentials of the various synaptic currents, where each current contributes proportionally to the relative amount of conductance it carries.

Voltage-dependence of NMDA channels. For NMDA channels to open, binding of neurotransmitter released by the pre-synaptic spike is not enough. The post-synaptic cell must also be sufficiently depolarized to remove their blockade by magnesium. It is conventional to model this using a voltage-dependent maximal conductance [66]:

$$g_{\text{NMDA}}(V) = \frac{g_{\text{NMDA}}}{(1 + ([\text{Mg}^{2+}]/\gamma)\exp(-\beta V(t))} \equiv g_{\text{NMDA}}\frac{1}{J(V(t))}, \tag{15.48}$$

with $[\text{Mg}^{2+}] = 1$ mM, $\gamma = 3.57$ and $\beta = 0.062$. To be able to incorporate this effect into the framework described in the previous sections, we linearize the voltage dependence of the NMDA current around the average voltage $\bar{V}$, obtaining

$$\frac{V(t) - V_E}{J(V(t))} \sim \frac{V(t) - V_E}{J(\bar{V})} + (V(t) - \bar{V})\frac{J(\bar{V}) - \beta(\bar{V} - V_E)(1 - J(\bar{V}))}{J^2(\bar{V})} + O((V(t) - \bar{V})^2) + \ldots \tag{15.49}$$

This linear approximation is very accurate for the range of values of $V(t)$ between reset and threshold [25]. Using it, the non-linear voltage-dependent NMDA current can be emulated by a linear current with a renormalized maximal conductance and reversal potential. Defining

$$I_{\text{NMDA}}(t) \equiv g^{eff}_{\text{NMDA}} s_{\text{NMDA}}(V(t) - V^{eff}_E), \tag{15.50}$$

the renormalized parameters read

$$\begin{aligned} g^{eff}_{\text{NMDA}} &= g_{\text{NMDA}} \frac{J(\bar{V}) - \beta(\bar{V} - V_E)(1 - J(\bar{V}))}{J^2(\bar{V})} \\ V^{eff}_E &= \bar{V} - \frac{g_{\text{NMDA}}}{g^{eff}_{\text{NMDA}}} \left(\frac{\bar{V} - V_E}{J(\bar{V})} \right). \end{aligned} \tag{15.51}$$

To give a qualitative idea of the properties of the linearized NMDA current, using $V_E = 0$ mV and $\bar{V} = -55$ mV, one obtains $g^{eff}_{\text{NMDA}} \sim -0.22 g_{\text{NMDA}}$ and $V^{eff}_E \sim -81.8$ mV. Since the slope of the $I - V$ plot for the original current is negative at voltages near the average depolarization of the neuron, the effective NMDA conductance is negative. However, since the effective reversal potential is lower than the cell's typical depolarization, the total effect of the effective NMDA current is depolarizing, as it should.

Calculation of the average voltage $\bar{V}$. To complete our discussion of the voltage-dependence, we need to compute the average voltage $\bar{V}$, that enters in Equation (15.45) and Equations (15.51). This can easily be done using Equation (15.24). The result is

$$\begin{aligned} \bar{V} &= \int_{-\infty}^{V_{th}} V[\rho_{ss}(V) + \nu\tau_{ref}\delta(V - V_r)]dV = \\ &= V_{ss} - (V_{th} - V_r)\nu\tau^{eff}_m - (V_{ss} - V_r)\nu\tau_{ref}. \end{aligned} \tag{15.52}$$

15.2.6.4 Fluctuations in the synaptic current in the case of multiple synaptic time scales

The results of Section 15.2.5 can be applied when a single time scale is present in synaptic currents. This is obviously not the case when fluctuations are due to AMPA, GABA$_\text{A}$ and NMDA currents. In the absence of rigorous results for fluctuations with multiple time scales, one has to resort to an approximation. The approximation is based on the fact that the longer the synaptic time constant, the more the fluctuations of the gating variable will be filtered out (see Section 15.2.5). Therefore, we expect the fluctuations in the GABA$_\text{A}$ and NMDA currents to be smaller in magnitude than those associated to the AMPA currents. We thus neglect their contribution and assume that $\delta s_{\text{NMDA}}(t) = \delta s_{\text{GABA}}(t) \sim 0$.

15.2.6.5 Summary: firing statistics of a neuron with realistic AMPA, GABA$_\text{A}$ and NMDA synaptic inputs

Here, we summarize the description of a LIF neuron that receives C_E excitatory synaptic inputs and C_I inhibitory synaptic inputs (Figure 15.1), with synapses described by individual conductances $g_{j_{\text{AMPA}}}$ and $g_{j_{\text{NMDA}}}$, $j = 1,2,...,C_E$; $g_{j_{\text{GABA}}}$, $j = 1,2,...C_I$). In the presence of these inputs, Equation (15.1) now reads

$$\begin{aligned}C_m \frac{dV(t)}{dt} = &- g_L(V(t) - V_L) - \\ &- \left[\sum_{j=1}^{C_E} g_{j_{\text{AMPA}}} s_{j_{\text{AMPA}}}(t) + \frac{g_{j_{\text{NMDA}}} s_{j_{\text{NMDA}}}(t)}{J(V(t))}\right](V(t) - V_E) - \\ &- \left[\sum_{j=1}^{C_I} g_{j_{\text{GABA}}} s_{j_{\text{GABA}}}(t)\right](V(t) - V_I). \qquad (15.53)\end{aligned}$$

For simplicity, we again assume that the synaptic conductances and the firing rates of all pre-synaptic inputs from the same sub-population are identical. Using the approximations described in the previous sections, this equation becomes

$$\begin{aligned}C_m \frac{dV(t)}{dt} = &- g_L(V(t) - V_L) - \\ &- C_E\left[g_{\text{AMPA}}\bar{s}_{\text{AMPA}}\right](V(t) - V_E) - \\ &- C_E\left[g^{eff}_{\text{NMDA}}\bar{s}_{\text{NMDA}}\right](V(t) - V_E^{eff}) - \\ &- C_I\left[g_{\text{GABA}}\bar{s}_{\text{GABA}}\right](V(t) - V_I) + \delta I(t), \qquad (15.54)\end{aligned}$$

where $\bar{s}_{\text{AMPA}} = \nu_E \tau_{\text{AMPA}}$, $\bar{s}\text{GABA}_{\text{A}} = \nu_I \tau_{\text{GABA}}$ and $\bar{s}_{\text{NMDA}} = \psi(\nu_E)$ where the function ψ is defined in Equation (15.43), and the fluctuations are described by

$$\tau_{\text{AMPA}} \frac{d}{dt}\delta I(t) = -\delta I(t) + \sigma_{eff}\,\eta(t) \qquad (15.55)$$

$$\sigma^2_{eff} = g^2_{\text{AMPA}}(\bar{V} - V_E)^2 C_E \bar{s}_{\text{AMPA}} \tau_{\text{AMPA}}. \qquad (15.56)$$

Since all the deterministic components of the current are now linear in the voltage, the equations describing the membrane potential dynamics can be expressed as

$$\tau_m^{eff} \frac{dV(t)}{dt} = -(V(t) - V_{ss}) + \frac{\delta I(t)}{g_L^{eff}} \qquad (15.57)$$

$$\tau_{\text{AMPA}} \frac{d}{dt}\delta I(t) = -\delta I(t) + \sigma_{eff}\,\eta(t). \qquad (15.58)$$

The effective membrane time constant is

$$\tau_m^{eff} = \frac{C_m}{g_L^{eff}} = \tau_m \frac{g_L}{g_L^{eff}}, \qquad (15.59)$$

and the effective leak conductance of the cell is the sum of the passive leak conductance plus the increase in the conductances associated to all the synaptic inputs to the cell

$$g_L^{eff} = g_L + g_{\text{AMPA}} C_E \bar{s}_{\text{AMPA}} + g^{eff}_{\text{NMDA}} C_E \bar{s}_{\text{NMDA}} + g_{\text{GABA}} C_I \bar{s}_{\text{GABA}}. \qquad (15.60)$$

In *in vivo* experiments, it was estimated that, even when neurons fire at low rates (a few hertz), g_L^{eff} is at least 3-5 times larger than g_L [32], therefore τ_m^{eff} is $3-5$ shorter than τ_m. For example, if $\tau_m = 10$ ms, then $\tau_m^{eff} \simeq 2-3$ ms. When neurons fire at higher rates (leading to larger synaptic conductances), the value of g_L^{eff} would be significantly larger and τ_m^{eff} would be even smaller.

The steady-state voltage V_{ss} now becomes

$$\begin{aligned} V_{ss} = [\, g_L V_L \,+\, (C_E g_{\text{AMPA}} \bar{s}_{\text{AMPA}}) V_E + (C_E g_{\text{NMDA}}^{eff} \bar{s}_{\text{NMDA}}) V_E^{eff} \\ +\, (C_I g_{\text{GABA}} \bar{s}_{\text{GABA}}) V_I] \,/ g_L^{eff}. \end{aligned} \tag{15.61}$$

Note that the steady state potential V_{ss} is bounded between the highest and the lowest reversal potentials of the four currents to the neuron. In particular, it can never become lower than V_I. Thus, no matter how strong inhibition is, in this model the average membrane potential will fluctuate around a value not lower than the reversal potential of the inhibitory synaptic current, e.g., at approximately -70 mV.

Since Equations (15.57) and (15.58) can be mapped identically to Equations (15.33) and (15.34), one can now use equation (15.26) to compute the firing rate of a neuron,

$$\nu_{post} = \left[\tau_{ref} + \tau_m^{eff} \sqrt{\pi} \int_{\frac{V_r^{eff} - V_{ss}}{\sigma_{eff}}}^{\frac{V_{th}^{eff} - V_{ss}}{\sigma_{eff}}} e^{x^2} (1 + \text{erf}(x)) \right]^{-1} . \tag{15.62}$$

where τ_m^{eff} and V_{ss} are given by Equations (15.59-15.61); V_{th}^{eff} and V_r^{eff} are given by Equations (15.36-15.37). Note that now, the average voltage in the steady state $\bar{V}$ plays a role in determining the firing rate, through both V_{ss} and σ_{eff}. Since $\bar{V}$ is related linearly to the firing rate (Equation (15.52)), the firing rate is not an explicit function of the synaptic input. Even if the inputs are entirely external (feedforward), and all the synaptic conductances are fixed, $\bar{V}$ still depends on the post-synaptic firing rate ν itself. Therefore, ν must be determined self-consistently.

Equation (15.62) constitutes a non-linear input-output relationship between the firing rate of our post-synaptic neuron and the average firing rates ν_E and ν_I of the pre-synaptic excitatory and inhibitory neural populations. This input-output function is conceptually equivalent to the simple threshold-linear or sigmoid input-output functions routinely used in firing-rate models. What we have gained from all these efforts is a firing-rate model that captures many of the underlying biophysics of the real spiking neurons. This makes it possible to quantitatively compare the derived firing-rate model with detailed numerical simulations of the irregularly firing spiking neurons, an important step to relate the theory with neurophysiological data.

15.3 Self-consistent theory of recurrent cortical circuits

A cortical microcircuit receives afferent inputs and sends efferent outputs downstream, thereby information processing is carried out in a 'feedforward' fashion. At the same time, interesting computations may be accomplished by horizontal or recurrent synaptic connections within the local network. The relative importance of feedforward versus recurrent processing is likely to be different for each specific task, and vary from one cortical area to another. In the primary visual cortex (V1), recurrent synaptic connections are quite abundant [76]; their functional importance (such as to the generation of orientation selectivity) has been the subject of intense debate [39, 111]. Recently, there is growing interest in the recurrent networks of association cortical areas, such as the parietal cortex or prefrontal cortex. This interest was primarily motivated by the observation of 'working memory neurons' in these cortices. In experiments when an animal is required to remember a transient stimulus cue across a delay period of a few seconds, between the cue presentation and behavioral response, neurons in association areas display stimulus-selective, elevated persistent activity across the delay period [44, 55]. Since the elevated neural activity can be triggered by a brief input but outlast it for many seconds, persistent activity cannot be explained by a feedforward mechanism. It has been hypothesized that persistent activity can be self-sustained by synaptic 'reverberations' within a strongly recurrent local network (see for a review [6, 121]).

We will now discuss how a recurrent network of neurons can be described by mean-field theory. We will first consider how stationary states of such networks can be obtained in a self-consistent way. Next we discuss dynamical approaches which allow an assessment of the stability of the stationary states. Then, an example from an one-population network of excitatory cells is analyzed in detail, introducing the concept of bistability by means of a graphical analysis, and relating it to the phenomenon of persistent neural activity in working memory. A more detailed model of a network for object working memory is then described. Finally, we discuss the possibility of multi-stability in cortical networks in which both excitation and inhibition are strong, but roughly cancel each other out.

15.3.1 Self-consistent steady-state solutions in large unstructured networks

In a recurrent network, the post-synaptic neuron and its pre-synaptic inputs are part of the same network, and hence, if the activity in the network is not changing, their firing activity must, in a statistical sense, be the same. If, as we discussed in Section 15.2, the output firing rate only depends on the average rate of the inputs, then equalizing pre- and post-synaptic activity will yield an equation that determines the firing rates in the possible stationary states of the network. This is a very general necessary condition that has to be met in any steady-state solution of the network dynamics.

However, in order for us to be able use the input-output relationship found in Section 15.2, the dynamics of the network should be such that the network properties in this stationary states are consistent with the assumptions we made in Section 15.2. Thus, these assumptions impose several additional conditions that the steady-states should obey to be truly self-consistent:

- The fluctuations in the inputs must be approximately independent from neuron to neuron. This condition will be trivially satisfied when the major part of the noise comes from external independent sources. It will also be satisfied when the network is sparsely connected, i.e., when the connection probability between any pair of neurons is weak. In this case, the 'noise' term coming from the recurrent network itself becomes uncorrelated from neuron to neuron [19, 24, 118, 119].

- The probability of a spike being emitted in the network at any moment must be constant in time. Thus, the steady state must be stable with respect to any instability that leads to non-stationary global network activity, such as synchronized oscillations.

- The neurons must emit approximately as Poisson processes for the input-output relationship to be valid. This is in general expected to be true when the average total input to the neurons is sub-threshold, which will be the situation of interest in our discussion.

Several types of local network connectivity and synaptic structure are conceivable. They differ mainly in the source of the fluctuations in the synaptic current to the neurons. One approach is to investigate the behavior of the network as a function of its size N and of the number of connections per cell C. The strategy is to scale the PSP size $\bar{J} \equiv J/C_m$ (a measure of the synaptic strength) with C, and study the behavior of the network as $C \to \infty$. An advantage of this procedure is that **a)** the behavior of the network is much simpler and easier to analyze in the $C = \infty$ limit, and **b)** network behaviors which are only quantitatively different for finite C, become qualitatively different as C becomes infinite. Additionally several of the technical assumptions we had to make in Section 15.2 become exact in this limit.

Alternatively, one can assume that N and C are large but finite. In this case one does not assume any specific scaling of the PSP size with C, but rather uses the formulas for arbitrary values of these parameters (as in the previous sections) and studies the behavior of the resulting equations when they take realistic values informed by the available experimental data. Some of the hypothesis made in the calculations will only be verified approximately, but the theory will be more directly comparable to experiments, where C and J/C_m are finite.

In the following subsections we describe the self-consistency equations obtained in each of these scenarios. For ease of exposition, our discussion will be carried out in the simplest case of neurons connected through instantaneous synapses, unless specified otherwise.

15.3.1.1 Fully connected networks; External noise

In a fully connected network $C = N$. In such a network, all neurons see essentially the same recurrent input. In order to obtain a finite mean synaptic input (the 'mean-field') the PSPs are usually assumed to scale as $1/N$. In this case, the total synaptic input has a mean of order 1, and noise of order $1/\sqrt{N}$. This is apparent in the equations for the moments of the diffusion process (15.8), where it is clear that in this case only the first moment remains non-zero as $N \to \infty$. Thus, the recurrent component of the synaptic current becomes deterministic. In this framework, noise is assumed to come from unspecified external sources, and is assumed to be independent for each neuron.

Let us consider the simple case of a single neural population. We express the input-output relationship of the cell, Equation (15.62), as $\nu_{\text{post}} = \phi(\mu(\nu_{\text{pre}}), \sigma)$, making explicit the dependency of the rate of the post-synaptic cell on the rate of its pre-synaptic inputs through the mean μ and standard deviation σ of the fluctuations in the total afferent current. In the steady-state $\nu_{\text{post}} = \nu_{\text{pre}} \equiv \nu$, so the self-consistent relationship can be written as

$$\nu = \phi(\mu(\nu), \sigma), \tag{15.63}$$

where the mean input current $\mu(\nu) = \mu_{ext} + \mu_{rec}$ is the sum of an external tonic current μ_{ext} and of a mean recurrent synaptic current $\mu_{rec}(\nu)$. The noise component of the current comes exclusively from outside the network, i.e., $\sigma = \sigma_{ext}$. The solution of Equation (15.63) can be obtained graphically by plotting $\phi(\mu(\nu), \sigma)$ vs. ν and by looking at its intersections with the diagonal line [8, 11, 120]. Alternatively, one can plot the f-I curve $\nu = \phi(\mu, \sigma)$ vs. μ and look at its intersections with $\nu = \mu^{-1}(\mu)$ vs. μ [20]. An example of this kind of analysis is given below. When several populations are present, the framework is extended by adding one self-consistency equation per population (again see specific example below).

A general feature of an all-to-all network is that the level of noise is independent of the activity in the recurrent network. Thus the activity of the neurons is modulated by changes in the mean current they receive. As we shall see below, this has consequences on the statistics of their spike trains, a consequence that can be tested experimentally.

15.3.1.2 The balanced state

The all-to-all network architecture with $1/N$ couplings, though simple, is not very realistic. In the cortex, synaptic couplings are much stronger than $1/N$ and neurons are not fully connected. This motivates the study of networks which are sparsely connected, and with stronger coupling. Let us consider a network in which each neuron receives C random connections from a total of N neurons. If the network is very sparse, i.e., if $N \gg C$, the probability that two neurons receive a substantial fraction of common inputs becomes very small, so the recurrent inputs in this network will be effectively uncorrelated between any pair of post-synaptic cells. Since the second infinitesimal moment of the diffusion process, which measures the fluctuations in the synaptic current, scales as $C\bar{J}^2$ (see Equation (15.8)), to keep the fluctuations finite as $C \to \infty$, one should scale the synaptic couplings as $\bar{J} \sim 1/\sqrt{C}$. On the other hand, the mean synaptic input scales as $\bar{J}C \sim \sqrt{C}$ and diverges to plus or minus infinity (if the coupling is excitatory or inhibitory, respectively). Thus, one immediately sees that if the network is composed of a single excitatory population, the neuron will either be at saturation or totally silent for large C. To obtain plausible levels of activity in this framework, one needs, therefore, to introduce an inhibitory population.

Let us write $C_E = c_E C$, $C_I = c_I C$, where c_E and c_I are finite. As already anticipated, in order to keep the diffusion coefficient A_2 finite, the scaling $\bar{J}_{E,I} = j_{E,I}/\sqrt{C}$ must be used. The infinitesimal moments of the stochastic process become

$$\begin{aligned} A_1(V) &= -\frac{(V - V_{ss})}{\tau_m} + \sqrt{C}\,[j_E c_E \nu_E - j_I c_I \nu_I] \\ A_2 &= j_E^2 c_E \nu_E + j_I^2 c_I \nu_I \\ A_n &= C^{1-\frac{n}{2}}[j_E^n c_E \nu_E + (-1)^n j_I^n c_I \nu_I] \qquad n = 3, \ldots \end{aligned} \tag{15.64}$$

In this case, as $C \to \infty$, all terms of order $n > 2$ vanish, and Equation (15.7) becomes identical to the Fokker-Planck Equation (15.9). Therefore, in this case the diffusion approximation becomes exact. The second term in the drift coefficient A_1, which gives the mean current into the cell, diverges as $\sqrt{C}$. Thus, unless the excitatory and the inhibitory drives into the cell *balance* each other to within $1/\sqrt{C}$, the resulting massive excitatory or inhibitory drive will drive the neuron towards saturation, or total silence. [118] showed, in a recurrent network of binary neurons, that this balanced state can arise as a dynamically stable state in a very robust way. Using that neuronal model, a complete description of the temporal fluctuations of activity, beyond the Poisson assumption that we have been using, can be performed. As we shall now see, the equations which determine the average firing rate of the excitatory and inhibitory populations in the balanced state are very general, and applicable to any single neuron model that assumes that the different synaptic inputs to the cell are summed linearly.

Let us consider the two population network in which each population receives $C_{E,I}^{ext} = c_{E,I}^{ext} C$ excitatory Poisson inputs of rate ν^{ext} through synapses of strength $\bar{J}_{E,I}^{ext} = j_{E,I}^{ext}/\sqrt{C}$ from outside the network. We know that the fluctuating component of the current will be of order one and the mean inputs will be

$$\begin{aligned} \frac{\mu_E}{C_m} &= \sqrt{C}\,[j_{EE} c_{EE} \nu_E - j_{EI} c_{EI} \nu_I + j_E^{ext} c_E^{ext} \nu^{ext}] \\ \frac{\mu_I}{C_m} &= \sqrt{C}\,[j_{IE} c_{IE} \nu_E - j_{II} c_{II} \nu_I + j_I^{ext} c_I^{ext} \nu^{ext}]. \end{aligned} \tag{15.65}$$

Following the arguments presented in the previous sections, one can now *impose* that $\mu_{E,I}$ be order one, and see whether there is a stable self-consistent solution arising from this constraint. The simplest case is when $C = \infty$, which one expects to be also qualitatively correct for large but finite networks. In this case, a finite mean current can only be obtained if the balance is perfect, i.e., if the total excitation and inhibition cancel each other precisely. This means that the terms in square brackets in Equation (15.65) have to vanish identically, leading to a set of two coupled linear equations [118]

$$\begin{aligned} j_{EE}c_{EE}\nu_E - j_{EI}c_{EI}\nu_I + j_E^{ext}c_E^{ext}\nu^{ext} &= 0 \\ j_{IE}c_{IE}\nu_E - j_{II}c_{II}\nu_I + j_I^{ext}c_I^{ext}\nu^{ext} &= 0, \end{aligned} \tag{15.66}$$

which implies that the self-consistent rates of the two populations become a linear function of the external input

$$\nu_E = k_E\ \nu^{ext} + O\left(\frac{1}{\sqrt{C}}\right); \qquad\qquad \nu_I = k_I\ \nu^{ext} + O\left(\frac{1}{\sqrt{C}}\right). \tag{15.67}$$

In contrast to a fully connected network, a balanced network dynamics is an intrinsic source of noise. In fact, even if the network is purely deterministic (with constant inputs instead of stochastic Poisson trains), and the external afferents are assumed to be regular, the balanced network can give rise to chaotic network dynamics and highly irregular neural activities [118, 119]. The firing rates in this network can, therefore, be determined self-consistently without making any assumptions about the specific single neuron model as long as the synaptic currents from different inputs are summed linearly. Although this is a quite remarkable result, the linearity of the self-consistent rates on the external input raises a fundamental problem from a computational perspective: does this mean that a balanced network cannot subserve non-linear behaviors such as bistability? Put it differently, can a network be both bistable and generate its own noise? We will come back to this issue below.

The arguments presented above are valid for synapses modeled as voltage-independent synaptic currents. With conductance-based synaptic currents, the situation is quite different, since in the large C limit, the total synaptic conductance diverges to infinity, hence effective neuronal time constant tends to zero. In this limit, the membrane potential is slaved to an effective 'steady-state' potential that stays finite in that limit [107]. Thus, there is no more 'balance' condition to be fulfilled in this situation, unless additional hypothesis are used. In any case, the simple balanced network picture is useful as a metaphor for networks with strong coupling and highly irregular firing of its constituent neurons. We will use the term 'balanced network' in this loose sense in the following.

15.3.1.3 Large but finite sparse networks

We again assume a large sparse network ($N \gg C$) so that the recurrent inputs to different neurons can still be assumed independent, but we now assume that the number of connections per neuron C is large ($C \gg 1$) but finite, and the coupling strength to be small (the unitary PSP size $J/C_m \ll (V_{th} - V_L)$) but finite [11]. For example, $N \sim 10{,}000$ and $C \sim 1{,}000$; $J/C_m \sim 0.1 - 0.3$ mV whereas $(V_{th} - V_L) \sim 10 - 15$ mV. Thus, in this case, both the mean synaptic input and the fluctuations around it depend on the firing rate of the pre-synaptic neurons. For the case of a single population the self-consistent equation becomes

$$\nu = \phi(\mu(\nu), \sigma(\nu)) \tag{15.68}$$

where the mean and variance of the current are given respectively by

$$\mu(\nu) = \mu_{ext} + CJ\nu \tag{15.69}$$
$$\sigma^2(\nu) = \sigma^2_{ext} + CJ^2\nu. \tag{15.70}$$

In this 'extended' mean-field theory, not only the mean inputs are included in the description, but also the fluctuations around the 'mean-field' are relevant. As emphasized above, this approach is only applicable to network states in which neurons fire in an approximately Poissonian way, and when the low connection probability makes the emission processes of neurons essentially uncorrelated. Moreover, since J and C are finite, this approach is only approximate. However, simulations show it gives very accurate results when $C\nu\tau_m$ is large (several hundreds) and $J/(C_m V_{th})$ is small (less than several percent), as seem to be the case in cortex [10, 24, 19]. Equation (15.68) can again be solved graphically to obtain the self-consistent, steady-state firing rates in the network (see below).

It is straightforward to extend this description to a two population network of excitatory and inhibitory neurons. The equations are, for finite C_E, C_I, (E-to-E) J_{EE}, (I-to-E) J_{EI}, (E-to-I) J_{IE}, and (I-to-I) J_{II}:

$$\nu_E = \phi(\mu_E, \sigma_E) \tag{15.71}$$
$$\nu_I = \phi(\mu_I, \sigma_I) \tag{15.72}$$
$$\begin{aligned}
\mu_E &= \mu_{extE} + [C_E J_{EE}\nu_E - C_I J_{EI}\nu_I] \\
\mu_I &= \mu_{extI} + [C_E J_{IE}\nu_E - C_I J_{II}\nu_I] \\
\sigma^2_E &= \sigma^2_{extE} + \left[C_E J^2_{EE}\nu_E + C_I J^2_{EI}\nu_I\right] \\
\sigma^2_I &= \sigma^2_{extI} + \left[C_E J^2_{IE}\nu_E + C_I J^2_{II}\nu_I\right].
\end{aligned} \tag{15.73}$$

The stationary states of these two population networks and their stability properties have been studied extensively [11, 19]. Since the number of connections per neuron in these networks is large, they behave qualitatively like the balanced networks discussed in the previous section.

15.3.1.4 Spatial distribution of activity in finite heterogeneous networks

Mean-field equations have been derived for heterogeneous networks of binary neurons [119] and for heterogeneous networks of noisy LIF neurons [10]. Consider a network of N neurons in which the probability that two neurons are connected is $C/N \ll 1$. Each neuron will receive C connections on average, and the cell-to-cell fluctuations in the number of afferents will be order $\sqrt{C}$. In principle, when C is large, the fluctuations in the number of connections are small compared to the mean. However, since networks of excitatory and inhibitory cells settle down in a balanced state in which excitation and inhibition cancel each other out to within $1/\sqrt{C}$ (see above), the effective average input to the neurons becomes of the same order as its fluctuations, and this is reflected in wide distributions of firing rates in the steady states. To calculate this distributions self-consistently one proceeds as follows: The *temporal* average currents can be written as

$$\begin{aligned}\mu_E &= J_{EE}\sum_{j=1}^{N_E} c_j \nu_j^E - J_{EI}\sum_{j=1}^{N_I} c_j \nu_j^I \\ \mu_I &= J_{IE}\sum_{j=1}^{N_E} c_j \nu_j^E - J_{II}\sum_{j=1}^{N_I} c_j \nu_j^I,\end{aligned} \tag{15.74}$$

where c_j is a binary random variable such that $\mathrm{Prob}(c_j = 1) = C_{E,I}/N_{E,I} \equiv \varepsilon$, and where we have assumed for simplicity that the excitatory (inhibitory) synaptic efficacies are uniform for each type of connections (J_{EE}, J_{EI}, J_{IE}, J_{II}). The temporal averages of the current $\mu_{E,I}$ are now random variables due to the randomness in the connectivity. Their *spatial* averages are equal to

$$\begin{aligned}\bar{\mu_E} &= \varepsilon[J_{EE}N_E\bar{\nu}_E - J_{EI}N_I\bar{\nu}_I] \\ \bar{\mu_I} &= \varepsilon[J_{IE}N_E\bar{\nu}_E - J_{II}N_I\bar{\nu}_I],\end{aligned} \tag{15.75}$$

where $\bar{\nu}_{E,I}$ are the average excitatory and inhibitory rates across the population. The variance of $\mu_{E,I}$ across the population is

$$\begin{aligned}\sigma_{\mu_E}^2 &= \varepsilon\left[(1-\varepsilon)\left(J_{EE}^2 N_E\bar{\nu}_E^2 + J_{EI}^2 N_I\bar{\nu}_I^2\right) + J_{EE}^2 N_E\sigma_{\nu_E}^2 + J_{EI}^2 N_I\sigma_{\nu_I}^2\right] \\ \sigma_{\mu_I}^2 &= \varepsilon\left[(1-\varepsilon)\left(J_{IE}^2 N_E\bar{\nu}_E^2 + J_{II}^2 N_I\bar{\nu}_I^2\right) + J_{IE}^2 N_E\sigma_{\nu_E}^2 + J_{II}^2 N_I\sigma_{\nu_I}^2\right],\end{aligned} \tag{15.76}$$

with $\sigma_{\nu_{E,I}}^2$ equal to the variance of the spatial distribution of rates across the network. Since $\mu_{E,I}$ are the sum of many independent contributions, their distribution will be approximately Gaussian, so we can write

$$\mu_{E,I}(z) = \bar{\mu}_{E,I} + \sigma_{\mu_{E,I}}\, z, \tag{15.77}$$

where $z = \mathbf{N}(0,1)$. Rigorously speaking, the randomness in the connectivity will also induce cell-to-cell variability in the temporal fluctuations in the synaptic current. However, this effect is weak compared to the effect on the mean, so one can neglect it and still get very accurate results [10]. We will therefore assume that they are constant across the population, and close the self-consistency loop by writing the rates as a function of the mean and variance of the synaptic current

$$\nu_{E,I}(z) = \phi(\mu_{E,I}(z), \sigma_{E,I}). \tag{15.78}$$

To estimate the spatial distribution of rates across the network $\rho_{E,I}(\nu)$, we thus write

$$\rho_{E,I}(\nu) = \int \rho(z)\rho_{E,I}(\nu|z)dz = \int \rho(z)\delta(\nu - \phi(\mu_{E,I}(z), \sigma_{E,I}))dz. \tag{15.79}$$

The firing rate distributions obtained in this way agree very well with the results from numerical simulations [10] and are also qualitatively similar to the wide distributions of firing rates seen in cortex [70].

15.3.2 Stability and dynamics

Are the states in which rates are solutions of equations (15.63), (15.68) or (15.71) and (15.72) stable? In order to answer rigorously this question, one must come back to the Fokker-Planck approach of Section 15.2.3, and write down the corresponding equation for the distribution of membrane potentials of neurons in the network, $\rho(V,t|V_0,t_0)$, coupled to the average firing rate $\nu(t)$ through the boundary conditions [1, 24]. A brief sketch of this approach is provided in the Appendix. Although this is the rigorous way to assess the stability of the steady-state solutions within the strong noise framework, analytical investigations of the Fokker-Planck equations are rather involved, and their numerical resolutions are also complicated [88]. Furthermore, their generalization to noisy situations with realistic synaptic dynamics is even more involved [21, 40]. Thus, it is of interest to investigate the possibility of approximating the dynamics by simpler dynamical equations, such as the Wilson-Cowan-type equations [126]. It is important to emphasize, however, that each dynamical description is suitable only for certain types of instabilities. For instance, an approximate dynamics in terms of firing rates cannot predict the instabilities of the network to a state where neurons are synchronized 'spike-to-spike'. Once a particular dynamical description is selected, the stability of the steady states against perturbations which comply with the assumptions of the chosen dynamical picture can be assessed.

Approximate firing rate dynamics have been found in some situations. For example, in weakly coupled networks with long synaptic time constants, one can derive an equation for the synaptic gating variable $s(t)$ [36, 37]. Let us write Equation (15.62) as $\nu = \phi(\mu_V, \sigma_V)$, where $\mu_V = V_{ss} - V_L = \mu_V^{ext} + C\bar{J}s(t)$. We consider a single population, fully connected network, with $\sigma_V = \sigma_C^{ext}\sqrt{\tau_m}/C_m = 5$ mV. The dynamics for $s(t)$ reads

$$\begin{aligned} \frac{ds(t)}{dt} &= -\frac{s(t)}{\tau_{syn}} + \nu(t) \\ \nu(t) &= \phi(\mu_V^{ext} + C\bar{J}s(t), \sigma_V), \end{aligned} \tag{15.80}$$

where $\mu_V^{ext} = \bar{J}_{ext}C_{ext}s_{ext}$ is the contribution of the external input to the steady state voltage. According to this scheme, the firing rate is always at its steady state value given the input, whereas the synaptic gating variable only follows the rate with its characteristic time constant τ_{syn}. This approximation is justified for the LIF model, since it has been shown that the population firing rate follows the synaptic input instantaneously [21, 40], provided that there is sufficient input noise, and that the synaptic time constant is comparable to, or longer than, the effective membrane time constant τ_m^{eff}. To what extent this approximation holds true for the Hodgkin-Huxley neuron model, or for real neurons, remains to be established. Qualitatively, the reason is that when there is enough input noise, there is always a significant fraction of the neurons across the population close enough to threshold and 'ready' to respond immediately to a change in current. Thus, it is appropriate to use the steady-state relationship $\nu = \phi(\mu_V, \sigma_V)$ even if the inputs are not stationary, e.g., $\nu(t) = \phi(\mu_V(t), \sigma_V) = \phi(\mu_V^{ext} + C\bar{J}s(t), \sigma_V)$, as in Equation (15.80).

The fixed point of this system is

$$s_{ss} = \tau_{syn}\phi(\mu_V^{ext} + C\bar{J}s_{ss}, \sigma_V). \tag{15.81}$$

To check the stability of the fixed points of this network, the standard procedure is to consider a small perturbation of a steady state

$$s = s_{ss} + \delta s \exp(\lambda t), \tag{15.82}$$

where δs is a small perturbation that grows at a rate λ. Stability of the steady state implies $Re(\lambda) < 0$ for all possible perturbations. Inserting Equation (15.82) in Equation (15.80), we get

$$\lambda = -\frac{1}{\tau_{syn}} + \left.\frac{d\phi(\mu_V(s), \sigma_V)}{ds}\right|_{s=s_{ss}}. \tag{15.83}$$

Thus, the stability condition $\lambda < 0$ is

$$\left.\frac{d\phi(\mu_V(s), \sigma_V)}{ds}\right|_{s=s_{ss}} < \frac{1}{\tau_{syn}}. \tag{15.84}$$

Equation (15.84) is a condition on the slope of the input-output function ϕ at the value of the input current given by s_{ss}. Since it is more intuitive to work with firing rates, we can express it as a condition on the slope of ϕ as a funcion of ν if we note that we only need the value of this slope at the steady state. In general

$$\frac{d\phi}{d\nu} = \left(\frac{d\phi}{ds}\right)\left(\frac{ds}{d\nu}\right). \tag{15.85}$$

Since the output rate is an instantaneous function of s, we can calculate the first term on the right hand side for all $s(t)$. The second term we do not know in principle, but on the steady state $s_{ss} = \tau_{syn}\nu_{ss}$. Thus

$$\left.\frac{d\phi}{ds}\right|_{s=s_{ss}} = \frac{1}{\tau_{syn}} \left.\frac{d\phi}{d\nu}\right|_{\nu=\nu_{ss}}, \tag{15.86}$$

so that the stability condition becomes

$$\left.\frac{d\phi(\mu_V(\nu), \sigma_V)}{d\nu}\right|_{\nu=\nu_{ss}} < 1. \tag{15.87}$$

Thus, for a fixed point to be stable, the slope of the output rate as a function of the input rate, should be less than one at the fixed point. This is shown graphically in Figure 15.4, where $\phi(\mu_V(\nu), \sigma_V)$ is plotted as a function of ν, both for an excitatory network ($\bar{J} = 0.5$ mV) and an inhibitory network ($\bar{J} = -0.5$ mV). The external inputs are adjusted so that there is an intersection with the diagonal at 1 Hz in both cases. When the network is excitatory (Figure 15.4A; in this case we use $\mu_V^{ext} = 0$ mV), the function $\phi(\mu_V(\nu), \sigma_V)$ raises very fast from zero rates to saturation. Thus, the slope

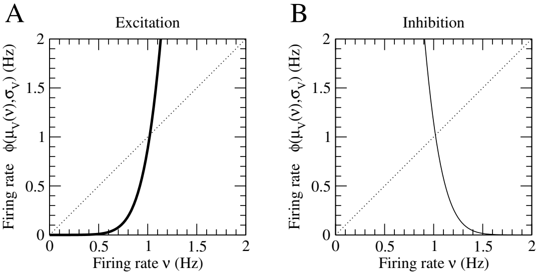

Figure 15.4

Self-consistent solution of Equation (15.80) and its stability properties. The output firing rate $\phi(\mu_V(\nu), \sigma_V)$ is plotted versus ν in two situations. **(A)** Excitatory network with $\mu_V^{ext} = 0$ mV, $\bar{J} = 0.5$ mV. **(B)** Inhibitory network with $\mu_V^{ext} = 20.4$ mV, $\bar{J} = -0.5$ mV. Other parameters are: $C = 1000$, $\tau_m = \tau_{syn} = 20$ ms ($C_m = 0.5$ nF, $g_L = 25$ nS), $\tau_{ref} = 2$ ms, $V_{th} = -50$ mV, $V_r = -60$ mV, $\sigma_V = 5$ mV. For both types of networks, there is a self-consistent solution around 1Hz. In the excitatory network, this self-consistent solution is highly unstable, because the slope of $\phi(\mu_V(\nu), \sigma_V)$ vs. ν is much larger than one; in the inhibitory network, the self-consistent solution is highly stable because of the large negative slope. In a balanced network where inhibition strongly dominates the recurrent circuit, the slope becomes infinite negative. Note that in the excitatory network, there are two other solutions: one at zero rate, and one close to saturation rates (about 500 Hz).

of $\phi(\mu_V(\nu), \sigma_V)$ is much larger than one at the self-consistent rate ν_{ss} and this steady state is, thus, unstable. The conclusion here is that low firing rates are expected to be hard to achieve in purely excitatory networks unless they are weakly coupled.

On the other hand, when the network is inhibitory (Figure 15.4B; $\mu_V^{ext} = 20.4$ mV), a supra-threshold external input is required to obtain an active network. The function ϕ now decreases as a function of ν (due to the fact that the mean decreases with ν). Equation (15.87) is now trivially satisfied when the coupling is predominantly inhibitory, $\bar{J} < 0$. Hence, a network state at this rate is stable. In the balanced network of Section 15.3.1.2, inhibition strongly dominates recurrence because it has to compensate for the external inputs. In this limit, the slope becomes infinitely negative. Note, however, that this strong stability of the purely inhibitory network is peculiar to synaptic couplings without latency. In presence of a latency, oscillatory instabilities appear even in strongly noisy networks [19, 24, 26].

Another simplified rate dynamics which has been frequently used is

$$\begin{aligned} \tau \frac{d\nu(t)}{dt} &= -\nu(t) + \phi(\mu_V^{ext} + C\bar{J}s(t), \sigma_V) \\ s(t) &= \tau_{syn}\nu(t), \end{aligned} \tag{15.88}$$

where τ remains unspecified. Although the fixed points of the systems described by Equations (15.80) and (15.88) are the same, this latter scheme neglects the dynamics for the synaptic variable, and instead uses an arbitrary time constant for the process by which the firing rate attains its steady state value. In conditions of high noise, Equations (15.80) seem, therefore, better suited to describe the time course of network activity than Equations (15.88).

We can extend Equations (15.80) to allow the description of a network with two (excitatory and inhibitory) neural populations, with firing rates ν_E and ν_I, and with synaptic latencies. If synaptic activation has a latency of τ_{lE} for excitation and τ_{lI} for inhibition, then we have

$$\begin{aligned} \frac{ds_E(t)}{dt} &= -s_E(t)/\tau_{syn,E} + \nu_E(t - \tau_{lE}) \\ \nu_E(t) &= \phi_E(\mu_V^{extE} + C_E\bar{J}_{EE}s_E(t) - \bar{J}_{EI}C_I s_I(t), \sigma_V^E) \\ \frac{ds_I(t)}{dt} &= -s_I(t)/\tau_{syn,I} + \nu_I(t - \tau_{lI}) \\ \nu_I(t) &= \phi_I(\mu_V^{extI} + C_E\bar{J}_{IE}s_E(t) - \bar{J}_{II}C_I s_I(t), \sigma_V^I). \end{aligned} \tag{15.89}$$

Equations (15.89) are Wilson-Cowan type dynamical mean-field equations which are ‘derived’ from an underlying biophysical description of neurons and synapses. In principle, the behavior of this model can be compared quantitatively (albeit approximately) with that of the original large-scale network of irregularly spiking LIF neurons. One should bear in mind, however, that when time delays are included, the dynamics become significantly richer, and the analysis more complicated. In fact, rigurously speaking, in the presence of temporal delays the system becomes infinite-dimensional, even if one deals with a single population (a function evaluated at $t+\tau$, i.e., displaced in time, can be expressed as an infinite series of the time-derivatives of the function evaluated at t). Still, simplified descriptions of this complicated dynamical system can be used which produce results in good agreement with those from simulations [21, 26, 40]. In general, the Wilson-Cowan type equations do a good job of predicting the mean rate instabilities. They predict oscillatory instabilities only in the above mentioned conditions (large amplitude noise filtered by synapses with time constants comparable to membrane time constants, see [21, 26, 40]). For small noise, instabilities not predicted by the rate equations occur, in which neurons are synchronized 'spike-to-spike' (see e.g., [1, 62, 122] and refs therein). For other discussions about reductions to firing rate equations, see [7, 52, 109]. In a network in which a significant part of the fluctuations is generated by the network itself, Equations (15.80-15.88) can also be generalized to include the variance as a dynamical variable [11]. This approach gives rather accurately the mean rate instabilities in such networks, but not the oscillatory instabilities induced by the interplay between the dynamics of the variance and the mean, where the full Fokker-Planck approach must be used [19].

15.3.3 Bistability in a single population network

We now come back to the fully connected network of N_E excitatory cells. As we have seen in Section 15.3.2 (Figure 15.4), such networks can have more than one steady-state for the firing rates, provided the excitatory coupling is strong enough. Thus, an excitatory network can be *bistable*. In absence of external inputs, and with linear synapses, bistability typically occurs between one steady state at zero rate and another one at rates close to saturation. How is this picture affected by realistic synaptic dynamics? A situation of interest is when synaptic currents are mediated by saturating NMDA receptors [120]. By the arguments presented above, the fluctuations in these currents are negligible, because of the long decay time constant. In addition to N_E recurrent contacts, each neuron also receives C_{ext} AMPAR-mediated excitatory synaptic inputs from outside the network. Each of these external inputs provides spikes according to an independent Poisson process, and the firing rate of each input is random from a distribution with mean ν_{ext}. Thus, in this particular example, all the noise is generated outside the network, and is independent from cell to cell by construction.

We consider a slightly different dynamical picture from the one introduced in the previous section in Equation (15.80). The main difference is that, since we want to include the realistic synaptic dynamics, Equation (15.40), appropriate for the slow NMDA channel dynamics, the synaptic currents depend now non-linearly on the firing rates. Following the arguments presented in Section 15.3.2, we use the mean level of synaptic activity at the recurrent synapses $\bar{s}_{\text{NMDA}}$ as the dynamical variable, since the time constant of the NMDA-mediated currents is the slower time scale of the system. However, the synaptic activity depends now in a non-linear way on the input rate. Equations (15.40) and (15.80) suggest seeking a dynamical equation for the variable s of the form

$$\frac{d\bar{s}_{\text{NMDA}}}{dt} = -\frac{\bar{s}_{\text{NMDA}}}{\tau_{\text{NMDA}}} + (1-\bar{s}_{\text{NMDA}})F(\nu), \tag{15.90}$$

where the function $F(\nu)$ is determined in a self-consistent way by the steady state dependency of $\bar{s}_{\text{NMDA}}$ on ν (note that we are going directly from the firing rate ν to the $\bar{s}_{\text{NMDA}}$ variable. Thus we are neglecting the dynamics associated to the variable $x(t)$ in Equation (15.41)).

Imposing $d\bar{s}_{\text{NMDA}}/dt = 0$ we obtain

$$\bar{s}^{ss}_{\text{NMDA}} = \frac{F(\nu)}{\frac{1}{\tau_{\text{NMDA}}} + F(\nu)} \equiv \psi(\nu), \text{ so that } F(\nu) = \frac{\psi(\nu)}{\tau_{\text{NMDA}}(1-\psi(\nu))}. \tag{15.91}$$

Inserting this expression back into Equation (15.90), we obtain

$$\frac{d\bar{s}_{\text{NMDA}}}{dt} = -\frac{1}{\tau^{eff}_{\text{NMDA}}}\left[\bar{s}_{\text{NMDA}} - \psi(\nu)\right], \tag{15.92}$$

where $\tau^{eff}_{\text{NMDA}} = \tau_{\text{NMDA}}(1-\psi(\nu))$, and where ν is given by

$$\nu = \left[\tau_{ref} + \tau^{eff}_{m}\sqrt{\pi}\int_{\frac{V_r - V_{ss}}{\sigma_V}}^{\frac{V_{th}-V_{ss}}{\sigma_V}} e^{x^2}(1+\text{erf}(x))dx\right]^{-1}, \tag{15.93}$$

which depends on $\bar{s}_{\text{NMDA}}$ through τ^{eff}_m and V_{ss} (see Equations (15.61) and (15.59)). Thus, due to the saturation implicit in Equation (15.90), the effective time constant of this dynamics depends on the firing rate, and becomes faster at higher pre-synaptic activity. When the firing rates change slowly enough, this dynamics produces quantitative agreement with the results from simulations of the full spiking network [95].

Looking at these expressions, one notices that the dependence of the firing rate on the mean recurrent synaptic activity is always through the product $N_E\bar{g}_{\text{NMDA}}\bar{s}_{\text{NMDA}} \equiv g_{\text{tot}}\bar{s}_{\text{NMDA}} \equiv \tilde{s}$. The steady states of our dynamics are thus given by the solutions of the following equation

$$\frac{\tilde{s}}{g_{\text{tot}}} = \psi(\nu(\tilde{s})), \tag{15.94}$$

which correspond to the intersections of the curves given by each side of this equation plotted as a function of $\tilde{s}$. This equation generalizes Equation (15.63) to the situation of non-linear synapses. Qualitatively, these intersections correspond to the points in which the activity at the synapse (the right-hand side of Equation (15.94)) is equal to the feedback provided by the network (the left-hand side of the same equation), which is a necessary condition for the network to be at a steady state. The advantage of using $\tilde{s}$ as our variable, is that now the right-hand side of the self-consistency equation is no longer dependent on the total synaptic conductance g_{tot}, which measures the gain, or amplification, of the mean activity at a single synapse by the network. Thus, as g_{tot} is varied, the self-consistent solutions of the dynamics move like the intersections of the two curves in Equation (15.94) as the slope of the straight line measuring the network feedback is changed. In Figure 15.5A, the function $\psi(\nu(\tilde{s}))$ and the line $\tilde{s}/g_{\text{tot}}$ have been plotted for three values of g_{tot}. This figure shows that, depending on the value of the gain, the two curves can intersect either once or three times, allowing for the possibility of several coexisting steady state solutions.

The next step is to look at the stability of these solutions. It can be done along the lines of Section 15.3.2. Let us, for brevity use $\tilde{s} \equiv \tilde{s}_{\text{NMDA}}$. We rewrite the dynamical equation as

$$\frac{d\tilde{s}}{dt} = \frac{1}{\tau_{\text{NMDA}}^{eff}}(g_{\text{tot}}\psi(\nu(\tilde{s})) - \tilde{s}) \equiv G(\tilde{s}), \tag{15.95}$$

The stability of a steady-state $\tilde{s}_{ss}$ is given by the slope of $G(\tilde{s})$ evaluated at $\tilde{s}_{ss}$. If

$$\left.\frac{dG(\tilde{s})}{d\tilde{s}}\right|_{\tilde{s}=\tilde{s}_{ss}} < 0 \qquad \text{or} \qquad \left.\frac{d\psi(\nu(\tilde{s}))}{d\tilde{s}}\right|_{\tilde{s}=\tilde{s}_{ss}} < \frac{1}{g_{\text{tot}}} \tag{15.96}$$

then $\tilde{s}_{ss}$ is stable. Recall that $G(\tilde{s_{ss}}) = 0$. If $G(\tilde{s})$ has a negative slope at $\tilde{s}_{ss}$, then it is positive for $\tilde{s}$ slightly less than $\tilde{s}_{ss}$, therefore $\tilde{s}$ will increase in time according to Equation (15.95), converging towards $\tilde{s}_{ss}$. Similarly, $G(\tilde{s})$ is negative for $\tilde{s}$ slightly larger than $\tilde{s}_{ss}$, again $\tilde{s}$ will converge back to $\tilde{s}_{ss}$. Therefore, $\tilde{s}_{ss}$ is stable. Conversely, if the derivative of $G(\tilde{s})$ is positive at a steady state, the latter is unstable.

Equation (15.96) implies that to assess the stability of a steady state solution graphically by looking at the intersections of the two functions in Equation (15.94), the stable fixed points will be those in which the sigmoid function has a lower slope than the straight line at the intersection.

Figure 15.5A shows that if the recurrent gain g_{tot} is lower than the dashed line marked by g_{tot}^{Low}, the slope $1/g_{\text{tot}}$ is too high and there is only one fixed point with low, but non-zero activity, which is always stable. On the other hand, when the recurrent gain is higher than the dashed line marked by g_{tot}^{High}, the only fixed point, which is also always stable, corresponds to a state of high activity. In between, there is a range of values of g_{tot} in which three fixed points coexist. The ones with higher and lower activity are stable (marked with a filled circle), and the intermediate one (open circle) is unstable. When g_{tot} lies within this range, the network is said to be bistable. The intermediate unstable point corresponds to the steady-state which we showed in Figure 15.4 with excitatory connections.

As shown in Figure 15.5C, when the network is bistable, transient inputs can switch the state of the network between its two stable states. The network can, in this sense, be used as a working memory device (see next section), as the presence or absence of an input to the network can be read out from its activity *after* the stimulus is no longer physically present. One atractive feature of this encoding scheme is its robustness. Indeed, the activity state of the network does not reflect the occurrence of a single stimulus, but rather of a *class* of stimuli. In our example, quite a different range of amplitudes of the applied current will lead to the same steady-state. The network is said to use ‘attractor dynamics’, since each fixed point attracts the state of the network from a wide range of initial conditions. This concept can be understood by imagining that the state of the network, in our example the gating variable $\bar{s}$, slides

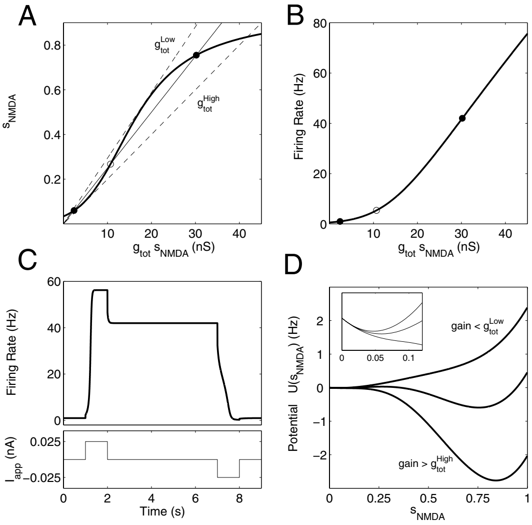

Figure 15.5

Bistability in a simple, one-population recurrent network. **(A)**. Mean synaptic activity as a function of the total recurrent synaptic input $\tilde{s} = g_{\text{tot}}\bar{s}_{\text{NMDA}}$ (sigmoid; thick solid line) and same quantity when $\tilde{s}$ is interpreted as the total recurrent network feedback (thin dashed lines). $g_{\text{tot}}^{High} = 50$ nS and $g_{\text{tot}}^{Low} = 34$ nS correspond to the highest and lowest values of the network gain in which the network is bistable for this network parameters. The crossings of a straight line corresponding to an intermediate value of $g_{\text{tot}} = 40$ nS with the sigmoid, correspond to the 3 steady state solutions in the bistable regime. Stable (unstable) solutions are marked with a filled (open) circle. **(B)**. Mean firing rate of the neurons as a function of the total recurrent synaptic input. The fixed point solutions in **A** are shown. Note the low rate ($\sim$ 40 Hz) of the stable high activity state, and the non-zero rate ($\sim$ 1 Hz) of the stable low activity state. Network parameters are: $C_{ext}\nu_{ext} = 0.7$ KHz, $g_{\text{AMPA}}^{ext} = 3.1$ nS. Parameters for the single cells are like in Figure 15.1. Synaptic parameters are given in Section 15.2.6.2. **(C)**. Time course of activity (top) in the network when $g_{\text{tot}} = 40$ nS. Brief current pulses (bottom) switch the network between its two stable states. **(D)**. Potential function associated to the dynamics of Equation (15.92) for three different values of g_{tot}, corresponding to the situations where there is only one low activity steady-state (upper curve), one high activity steady-state (lower curve) and for a bistable network (middle curve). The values of s_{NMDA} at the steady states are given by the minima of the potential function. **Inset**: blown-up version of the low activity region of U shows the disappearance of the low activity minimum of the potential as g_{tot} increases.

down a hilly landscape. The valleys (or minima) correspond to the steady-states, and the class of stimuli which are attracted to each steady-state (called its basin of attraction) are the set of all locations in the landscape which roll down to the same minima. Indeed, the dynamics (15.92) can be re-written as

$$\frac{d\bar{s}}{dt} = -\frac{dU(\bar{s})}{d\bar{s}}. \tag{15.97}$$

This dynamics describes the movement of a point particle at location $\bar{s}$ sliding down the landscape defined by the function $U(\bar{s})$ in the presence of high friction. $U(\bar{s})$ is also called the potential of the dynamics, and it is such that the speed of the particle at location $\bar{s}$ is equal to minus the slope of U at that location. In Figure 15.5D we show three examples of the $U(\bar{s})$ for values of the gain at the recurrent connections g_{tot} such that the network has either a single high ($g_{\text{tot}} > g_{\text{tot}}^{High}$) or low ($g_{\text{tot}} < g_{\text{tot}}^{Low}$) activity steady-state, and for an intermediate value of g_{tot}, where the network is bistable. For low enough gain, $U(\bar{s})$ has a single minimum, and as the gain increases, a second minimum at a higher value of $\bar{s}$ appears. These two minima coexist for a range of values of the gain, but if the gain is high enough, the low activity minimum disappears (see inset in Figure 15.5D).

Several features deserve comments: First, in contrast to the network of linear synapses of Section 15.3.2, the firing rate in the high activity fixed point is about 40 Hz (Figure 15.5B), much less than saturation rates, even in the absence of inhibition. This rate is in the upper range of the available physiological data for persistent activity(20-50 Hz). This relatively low rate is due to the saturation properties of the NMDA receptor. Second, the low activity state has a low firing rate of about 1Hz. This is due to the presence of noise in the system. In the absence of noise, i.e., when the synaptic current is constant in time, the input-output function of the neuron becomes a sigmoid with a 'hard' threshold. For currents below this threshold the output rate is identically zero (see trend for decreasing noise levels in Figure 15.2) and in the supra-threshold regime it increases as a sub-linear function of the input current until saturation at $1/\tau_{ref}$ is reached. In these conditions, when the network is bistable, the low activity state is necessarily zero. When noise is included in the description, the firing rate can be non-zero even in the sub-threshold regime: the membrane potential, which hovers around its steady state below threshold, crosses this threshold once in a while as a result of the random fluctuations in the input current [8, 118]. Such a state of low, fluctuation-driven activity has been suggested to correspond to the background or spontaneous activity state found in the cortex [11]. The fact that the low activity state is stable in Figure 15.5 is due to the fact that the excitatory feedback is unrealistically weak (see caption of Figure 15.5). When excitatory feedback is stronger, the non-zero low rate state becomes unstable, and inhibition becomes necessary to achieve stability at low rates [11]. Hence, a more detailed model is required.

15.3.4 Persistent neural activity in an object working memory model

As we have just demonstrated, the self-consistency equations whose solutions provide the firing rates of the different neural populations in the steady-states of the recurrent network can, in some cases, have multiple solutions for the same set of parameters and external inputs to the network. When this is the case, transient external inputs can switch the state of the network among its possible stationary solutions. Conceptually, the network can now function as a short-term or working memory system, as its state of activity is no longer uniquely specified by the 'static' variables of the system (cellular or network parameters, unspecific external inputs, etc) but also carries information about the recent history of transient inputs to the network. More generally, a recurrent network can display multi-stability, whereby a resting state of spontaneous activity coexists with multiple attractor states (stable neural firing patterns), each of which encodes a different sensory stimulus. Therefore, the identity of a transient input is encoded and stored in the level of spiking activity of a distinct neural assembly in the recurrent circuit. Such stimulus-selective persistent activity has been documented during electrophysiological experiments on behaving monkeys during working memory tasks [35, 41, 42, 43, 44, 45, 46, 47, 48, 55, 56, 73, 79, 81, 85, 86, 89, 100, 124].

We now describe in some detail an object working memory model that has been analyzed both at the mean-field level and with numerical simulations. For the model to comply with the basic phenomenology of the data from object working memory experiments, the simple bistable network presented in the previous section has to be considerably enlarged. First, local networks in association cortices are likely to be endowed with much more than two attractors. The experiments in the temporal lobe with a large number of stimuli (up to 100) [81, 85, 86, 100] suggest the following picture:

- In the absence of external stimulation, networks in the temporal lobe are in a spontaneous activity state, in which all neurons fire at low levels of several Hz;
- Upon presentation of a particular *familiar* stimulus, a small sub-population of neurons in localized areas of the temporal lobe exhibit persistent activity; this fraction of neurons can be estimated to be around 1% or a few % [81]. Thus, the representation of *familiar* stimuli in these areas is sparse;
- Representations of different stimuli have very small overlaps, since neurons typically respond to only one or a few images in the set of shown images [81].

A model for object working memory based on these observations has been built in several stages [9, 11, 12, 21]. The model of [11] is a network of randomly connected excitatory and inhibitory neurons much as the one discussed in Section 15.3.1.3. In addition, the excitatory population is divided in sub-populations, where a given sub-population is assumed to have a strong visual response to a particular stimulus.

A schematic representation of the architecture of the model is shown in Figure 15.6A. The network consists of two large pools of interacting pyramidal cells and interneurons. Both populations are fully connected with themselves and with each other. The pyramidal cell population is itself divided in several sub-populations. Since the experiments show that single cells only respond to a very small fraction of the stimulus set (approximately 1%), the model assumes that each sub-population shows selective responses to a single stimulus, and that the different sub-populations are non-overlapping. In addition to this set of 'selective' sub-populations, there is a sub-population of neurons which are not selective to any particular object. All neurons also receive unspecific excitation from outside the network. In the model of [11], synaptic transmission was assumed to be instantaneous. [25] proposed a model with more biologically plausible synapses (a full description of the spiking neuron model, as well as the complete set of mean-field equations, can be found in that paper). In the model, excitatory transmission is both AMPAR- and NMDAR-mediated, though with a dominant contribution of the NMDA component at the recurrent synapses and a dominant AMPA component on the external inputs, while inhibitory transmission is mediated by $\mathrm{GABA_A}$ receptors. Neurons belonging to the same selective sub-population are assumed to be frequently co-activated by the visual input which drives them effectively, and, therefore, the excitatory synapses connecting them are assumed to have undergone Hebbian synaptic potentiation, so that the synaptic strength at these recurrent connections is supposed to be larger than average.

As shown in Figure 15.6B, when the strength of these synapses is increased to be approximately twice the average (baseline) excitatory coupling strength, there is a sudden 'bifurcation' at which bistability emerges in each selective sub-population. Therefore, a graded difference in the coupling strength could lead to qualitatively different network behaviors (e.g., with or without persistent activity). Again the firing rate in the elevated persistent activity state is fairly low, in agreement with the data. Also, although the firing rates predicted by the mean-field model (solid curve) slightly overestimate the results from the direct simulations of the original spiking neural network (filled dots), the agreement between the two is reasonably good.

An example of the behavior of the network when the sub-populations are bistable is shown in Figure 15.6C. The upper part of the figure shows rastergrams from selected neurons from each of the sub-populations, and the lower part shows the population firing rate for each of the sub-populations. During the time interval marked as sample, the firing rate of the external inputs to the sub-population marked as number 1 is transiently elevated. The resulting increase in activity in this sub-population persists during a delay of several seconds, self-sustained by recurrent synaptic reverberations. At the end of the delay, an excitatory input to the whole network, signaling the behavioral response [29], switches the network back to the resting state. The persistent delay activity slightly increases the excitatory drive to the interneuron population, which also increases its activity. As a result, other sub-populations not selected by the transient input are more strongly inhibited during the delay.

15.3.5 Stability of the persistent activity state

The stability of the persistent state in a single population network is easily read out from a graph such as the one of Figure 15.5. In networks with several populations,

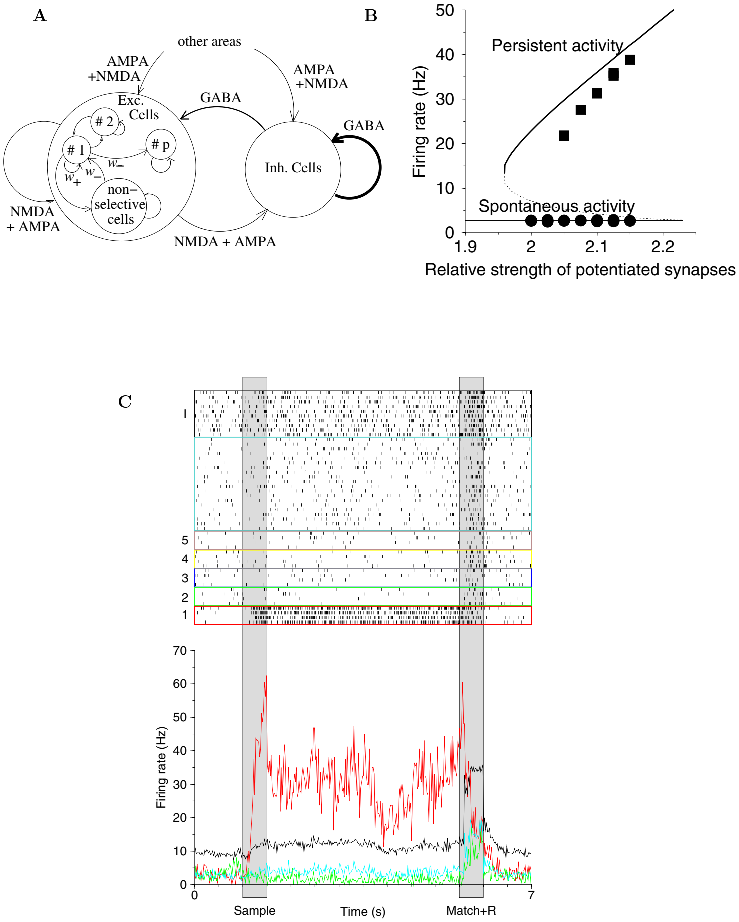

Figure 15.6

Behavior of an object working memory network. **(A)**. Schematic representation of the network. Circles represent the different sub-populations. Labels on the arrows indicate the type of synaptic connection between them. The width of the arrows qualitatively represents the strength of the corresponding synaptic connections. **(B)**. Bifurcation diagram showing the onset of bistability as a function of the strength of the connections within a selective sub-population relative to a baseline. Lines are the prediction from the mean-field version of the model, with solid (dashed) lines representing stable (unstable) steady states. Squares (spontaneous activity) and circles (elevated persistent activity state) are results from simulations of the spiking network (continued).

Figure 15.6

(C). Time course of the activity of the different sub-populations of the network in a delayed match-to-sample protocol. The network is initially in a resting state with low and uniform spontaneous firing activity. A brief stimulus to one of the neural sub-populations (indicated in red) triggers persistent activity that is self-sustained by recurrent reverberations and that is confined to that neural sub-population. This memory is erased, and the network is turned off, by another transient input during the match+response time epoch. Top: rastergrams. Bottom: firing rate histogram across the corresponding sub-populations. Red: selective sub-population receiving a transient external excitation during the sample period. Green, yellow, blue and brown: sub-populations selective to other stimuli. Cyan: sub-population of non-selective excitatory cells. Black: inhibitory interneurons. (See color insert.)

such as the excitatory-inhibitory networks, the analysis becomes more complicated, and the stationary state with the highest firing rate can destabilize through an oscillatory instability. Oscillatory synchrony can then disrupt bistability, or multi-stability. Conditions for stability of a persistent activity state in presence of synchrony can be understood from the simple following intuitive argument. Consider a network of neurons which, upon the arrival of a transient excitatory input, switch their activity from a few Hz in spontaneous activity to an elevated activity state of 20-40 Hz. Let us consider a single neuron firing at 25 Hz in this persistent activity state. This neuron fires, on average, every 40 ms. Since the tonic external input has not changed, in order for the neuron to maintain this firing rate, the recurrent network must provide enough current during the next few tens of milliseconds after each spike so that the cell will spike again. How can this be achieved? A possibility is that the network operates in an asynchronous state, where the assumed statistical independence of the firing times in different neurons is satisfied. In such a state, the fraction of neurons firing a spike across the network is, on average, constant. This property implies that, if the number of cells is large enough, the network will generate a tonic input, constant in time on average, which can sustain the firing of the single cells in a stable manner.

On the other hand, when there is some degree of synchrony, the fraction of cells firing at any given time starts to fluctuate on average, even for a very large network size. Consider the extreme case in which, at some instant of time, the whole network is perfectly synchronized, i.e., all neurons fire at the same time. Unless some mechanism exists which can keep a memory of this burst of activity and somehow delay it, turning it into an input to the cells at a *later* time, the activity of the network would decay to the resting spontaneous state [60]. A mechanism working in that direction could be implemented by the long time constant of NMDAR-mediated synaptic transmission [120]. Intuitively, according to Equation (15.29), if a synapse has time constant τ_{syn} the current into the cell resulting from a single spike at t_{spk} is still 37% of its maximum at $t_{spk} + \tau_{syn}$. This delay between the occurrence of an excitatory event, and its effect on the post-synaptic cell could, therefore, be beneficial for the stability of persistent activity in the presence of synchrony.

In addition, a long synaptic time constant for excitation might help to stabilize the asynchronous state itself. Oscillations easily occur in networks of interacting excitatory and inhibitory neurons, if the time constant of inhibition is longer than that of excitation (see e.g., [26, 120]). The intuitive reason is that, if such a network is perturbed from its steady state, the excitation will build up before the inhibition has time to suppress it. This excess of excitation will result in an increased inhibition which eventually overcomes the excitation, resulting in an overall suppresion of the excitation in the network. A decreased drive to the inhibitory cells leads to a decay of their activity, releasing the excitatory population and the rhythmic cycle starts again. When the excitation is slower, this type of oscillatory instability is prevented, as any excitatory perturbation results in an increased inhibition before the excitation has time to build up.

An example of the effect of changing the effective time constant of excitation on the stability of persistent activity is shown in Figure 15.7, for the object-working memory network described above. Remember that in this network, GABA$_\mathrm{A}$R- mediated inhibition has a time constant of 10 ms, and AMPAR- and NMDAR-mediated excitation have time constants of 2 ms and 100 ms respectively. In the Figure, the temporal course of the average activity of a sub-population of selective cells after the application of a transient excitatory input is shown, as the relative contributions of AMPA and NMDA receptors at each excitatory synapse is varied systematically, thus taking the network from a scenario in which excitation is slower than inhibition to one in which it is faster. When the AMPA:NMDA ratio of the charge entry per unitary EPSC is 0.1 (measured at $V = -55$ mV, near threshold), the average activity is fairly constant in time and shows only small amplitude fluctuations which do not destabilize the persistent state. As the ratio is progressively increased, excitation becomes faster, and the amplitude of the fluctuations grows. However, due to the delay effect mentioned above, the state of persistent elevated activity is still stable for fairly large amplitude fluctuations in the average activity (see Figure 15.7c). Indeed, in these conditions, the power spectrum of the average activity shows a clear peak near 40 Hz (see inset in Figure 15.7e). Although the issue of whether the persistent activity observed in the cortex is indeed oscillatory is controversial, a similar spectral structure has been recorded in local field potentials of area LIP [90] (but see [30]). As expected, when excitation becomes too fast, the amplitude of the oscillations becomes too large, and NMDAR-mediated excitation is unable to bridge the gap between activity bursts, with the resulting destabilization of the persistent activity state, and the network's working memory behavior is lost (Figure 15.7D).

These arguments raise the possibility that NMDAR-mediated excitation, or more generally, slow synaptic or cellular recurrent excitation could help to prevent oscillatory instability resulting from the excitation-inhibition loop, thereby contributing to the stability of persistent activity generated in recurrent cortical networks. Other factors can of course affect this stability. For instance, mutual inhibition between interneurons can in some conditions reduce the propensity of instability in an excitation-inhibition loop [62, 116]. This effect can be understood using a two-population rate model like Equations (15.89). When there is no synaptic latency, it can be shown by the linear stability analysis of a steady state that the I-to-I coupling

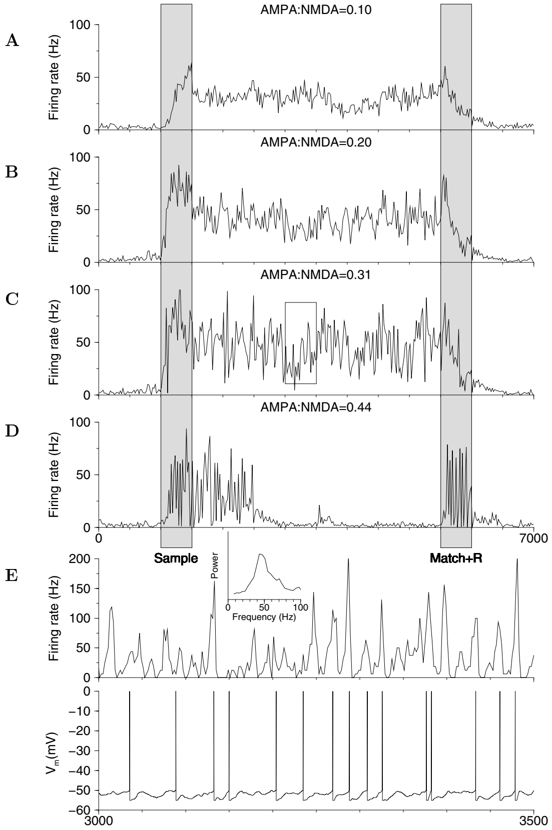

Figure 15.7

Stability of pesistent activity as a function of the AMPA:NMDA ratio. **(A-D)**. Temporal course of the average firing rate across a sub-population of selective cells in the network of Figure 15.6 after transient excitatory input, for different levels of the AMPA:NMDA ratio. This ratio is defined as that of charge entry through a unitary post-synaptic current at $V = -55$ mV (near threshold). As the ratio is increased, oscillations of a progressively larger amplitude develop, which eventually destabilize the persistent activity state. **(E)**. Snapshot of the activity of the network in **C** between 3 and 3.5 seconds. Top: Average network activity. Bottom: Intracellular voltage trace of a single neuron. **Inset**. Power spectrum of the average activity of the network, showing a peak in the gamma ($\sim$ 40 Hz) frequency range. Persistent activity is stable even in the presence of synchronous oscillations.

effectively reduces the time constant of the inhibitory population dynamics; faster inhibition thus reduces the likelihood for this type of oscillatory instability [112, 116]. However, it is important to emphasize that the oscillatory instability with reduced NMDA:AMPA ratio in our working memory models (Figure 6 and Figure 6 in [29]) was observed in the presence of strong I-to-I synaptic connections. The same result was also obtained with a spatial working memory model of Hogkin-Huxley-type conductance-based neurons [112]. Note that the models of [116] and [62] did not include synaptic latency, which has been shown to favor fast synchronous oscillations in a purely inhibitory network [21, 19, 26]. In general, stronger NMDA:AMPA ratio promotes asynchrony. Stronger I-to-I coupling without latency contributes to network asynchrony, but with synaptic/cellular latency could lead to oscillatory instability.

Other factors which might also contribute to the stability of persistent activity states include intrinsic ionic currents with long time constants [112], or bistability at the single cell level [28, 75]. Finally, heterogeneities, both in cellular and in connectivity properties, and noise, tend to desynchronize the network.

15.3.6 Multistability in balanced networks

Can bistability, or multi-stability, occur in a balanced network? Furthermore, can multistability occur between several states in which all neurons fire in a Poissonian fashion? These questions are interesting from a theoretical point of view, but are also raised directly by available data suggesting that the irregularity in the output spiking activity of cortical neurons is as high in high-rate persistent activity states as in the low-rate spontaneous activity state [30].

The balanced model of Section 15.3.1.2 seems incompatible with multistability, since the rates depend linearly on the external inputs through Equation (15.66). Indeed, unless the matrix of gain coefficients in equations (15.66) is singular (which requires a biologically unrealistic fine-tuning of parameters) these equations have a single solution for a fixed external input and are, consequently, incompatible with bistability. This is the manifestation of a general problem: any non-linear behavior in balanced networks requires a significant amount of fine-tuning in network parameters.

Several partial solutions to this problem can be suggested. First, in networks with finite connectivity such as the one of Section 15.3.4, multistability can be found even though the background state of the network is qualitatively a ‘balanced state’ (strong coupling, irregularly firing neurons). In fact, the multistability properties of such networks can be understood in the limit $C \to \infty$, $J \sim 1/\sqrt{C}$, if the relative size f of the selective sub-populations are taken to be small compared to the rest of the network, $f \sim 1/\sqrt{C}$: in this case, the leading order to the mean input in both cue and non-cue populations vanishes due to the ‘balance’ condition, but the corrections to the leading order are finite and different between the cue and non-cue populations, leading to a non-linear equation for the rates in the cue population (see [20] for a discussion of the sparse coding limit $f \to 0$). Thus, multistability is relatively easily achieved. The intuitive picture is that, while the global activity of the network is set by the 'balance' condition set by the strong global inhibition, the activity in the small selective excitatory sub-population becomes essentially uncoupled from the rest of the network. Consequently, this sub-population behaves essentially as the weakly coupled excitatory network of Section 15.3.3. However, a direct consequence of this scenario is that the CV in persistent activity must be lower than in spontaneous activity, because the mean inputs are larger in the cue population, while the variance remains unchanged (see Figure 2). Thus, there is no multistability between several 'balanced states'.

It is therefore possible for small sub-populations within a larger balanced network to be bistable in a robust way, but at the price that the small sub-populations themselves do not remain balanced in both steady-states. Is there an alternative? The idea would be to find a scenario in which the variance in the cue population also increases in a significant way from spontaneous to persistent activity, so that the increase in CV induced by the increase in variance counterbalances the decrease induced by the increase in the mean. One can even imagine a scenario in which the mean does not change, but the variance does. Such a scenario was introduced in [93, 94] for a network with finite connectivity C. It is a generalization of the model of [11] (see Figure 15.6A) in which the interneurons are also subdivided in selective sub-populations. Such a network is divided functionally in 'columns' or 'micro-columns' composed both of excitatory and inhibitory populations. Both populations are activated in a selective way when their preferred stimulus is shown. Consequently, in a persistent state, both excitatory and inhibitory populations raise their firing rates. A similar phenomenon was observed in experiments monitoring the activity of neurons in the prefrontal cortex of primates during working memory tasks. Recordings of nearby putative pyramidal cells and interneurons showed that the two sub-populations increase their firing rate during the delay period [92]. This has lead to the postulate of a micro-columnar organization of the pre-frontal cortex [92].

The 'micro-columnar' network has been studied at the mean-field level [93, 94]. In order to do a systematic investigation of the spiking variability resulting from different types of network organizations, the mean-field theory has been extended to be self-consistent both at the level of rates and CVs. In the previously discussed models, Poisson spiking statistics was an assumption, so the irregularity in the spiking activity in the pre-synaptic spike trains was 'fixed'. In [93] this assumption was relaxed by assuming that the neuronal spike trains can be described as renewal processes characterized by their mean rate and CV (a renewal point process is characterized by independent ISI intervals from an arbitrary distribution). When the statistics of the renewal spike trains are close to the Poisson case, the output rate and CV of the post-synaptic neuron can be calculated as a function of the rate and CV of its inputs, leading to steady state solutions in which both the rate and the CV are calculated self-consistently. Using this framework, multistability in the micro-columnar network described above has been studied using simple heuristic firing rate dynamics similar to Equation (15.88). The synaptic interactions between neurons depend on whether they belong to the same or to different micro-columns, and again, selective micro-columns are characterized by stronger excitatory recurrent interactions. In Figure 15.8, we show the time course of activity of a bistable micro-column in

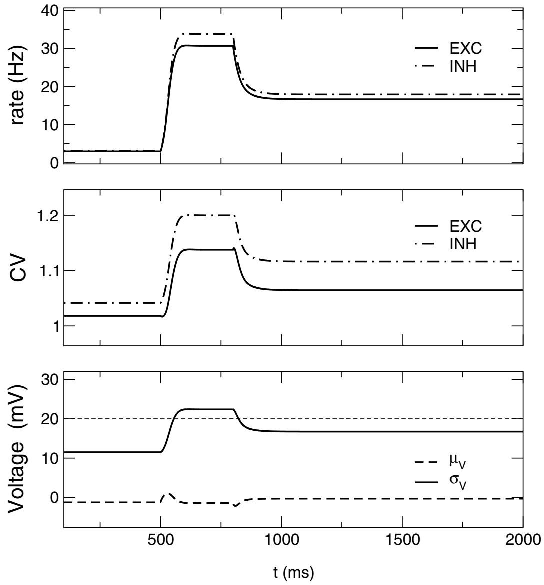

Figure 15.8

Bistability in a balanced multi-columnar cortical circuit. **(A)**. Temporal evolution of the firing rate from the excitatory and inhibitory sub-populations of a column. At $t = 500$ ms, a transient excitatory input was applied to both sub-populations. The elevated activity state in response to this input outlasts the stimulus offset. Note the elevated firing rate of the inhibitory sub-population also. **(B)**. Same as above for the CV of the two sub-populations. The CV increases with the firing rate. **(C)**. The figure shows the quantities $\mu_V = \mu_C/g_L$ and $\sigma_V = \sqrt{\tau_m}\sigma_C/C_m$ of the neurons in the excitatory sub-population. They correspond to the mean and standard deviation of the current, but are expressed in mV to facilitate comparison with the distance between V_{th} and V_L, equal to 20 mV (dashed line). The mean input current remains essentially the same for both the resting state and persistent state, regardless of their very different firing rates. The increase in firing rate in response to the stimulus is due to an increase in the amplitude of the fluctuating component of the current, hence the increase in CV above. In this network, both stable states are in the balanced regime.

response to a transient input. Panels A and B show the average rate and CV of the excitatory and inhibitory sub-populations in the micro-column. In this network, since the micro-column remains balanced in the elevated rate state, the CV of both sub-populations remains close to one in the delay period. Both the similar courses of activity of the excitatory and inhibitory populations, and the high CV during elevated persistent activity are consistent with measurements from prefrontal neurons in working memory tasks [30, 92]. The reason for this behavior is that the mean current to both sub-populations (see the lower panel for the case of the excitatory sub-population) remains approximately constant as the network switches between its two stable states. The increase in firing rate is due to an increase in the fluctuations in the current. Indeed, as a result of this, the CV actually increases in the elevated firing rate state.

This increase in CV is in contrast to the decrease in CV in models in which the network is not balanced in the elevated activity state, like the networks described in the previous two sections. This qualitative difference between relative change in spiking variability in these scenarios should, in principle, be experimentally testable, although the small difference in CV observed in Figure 15.8 would be hard to detect in experimentally recorded spike trains, due to limited sampling problems. Further experimental data are needed to resolve this issue.

As suggested at the beginning of this section, this scenario still suffers from a fine tuning problem. The range of multistability in the network with balanced persistent state is extremely small for realistic numbers of inputs per cell [93]. In fact, such multistability vanishes in the large C limit, even if the sub-populations are taken to scale as $1/\sqrt{C}$, because in that limit the difference in the fluctuations between spontaneous and persistent activity vanishes.

The fundamental problem which precludes robust bistability in balanced networks is the different scaling of the first two moments of the input current with the number of inputs and with the connection strength. While the mean scales as JC, the variance scales as J^2C. It is thus impossible to find a scaling relationship between J and C that keeps both moments finite when $C \to \infty$.

It is possible that cross-correlations in the activity of different neurons might provide a solution to the 'linearity' problem of balanced networks. The different scaling of the mean and the variance is a direct consequence of the fact that we have assumed the different inputs to the cell to be independent, so that the variance of their linear sum is the sum of their variances. If the inputs to the cell showed significant correlations, the variance would now scale as $(JC)^2$, in which case any scaling relationship between J and C would have the same effect on the mean and on the variance. It would therefore be of great interest to incorporate cross-correlations in a self-consistent manner into the picture we have been describing in this chapter.

15.4 Summary and future directions

In this chapter, we have presented analytical mean-field techniques that can be used to study the collective properties of large networks of spiking neurons. In analyzing the self-consistent steady-states of these networks, we observed that the self-consistency equations have sometimes multiple stable states. This leads quite naturally to the the interpretation of these networks as models of working memory systems. The methods discussed here help to understand in detail in which conditions multistability can be achieved in large networks of spiking neurons. The results that have been discussed are of course only the current status of a rapidly growing field. Extensions of both the mean-field techniques and of network architectures for working memory are either already done, under way, or should be done in the near future. We discuss here several of these possible extensions.

- **More realistic single neuron models.** The LIF lacks several features of real neurons. First, it lacks any sub-threshold resonance phenomena [65]. Generalizations of LIFs with several variables have been introduced that possess such sub-threshold resonance properties and can be studied analytically in stochastic contexts along the lines of Section 15.2.3 [98]. Second, it lacks an intrinsic spiking mechanism. The firing rate of neurons with intrinsic spike generation mechanism can be studied in the context of the 'quadratic integrate-and-fire' neuron [22], and even more realistic neurons can be studied analytically (Fourcaud et al. SFN 2002 abstract). Furthermore, mean-field theory can be extended to a recurrent network of Hodgkin-Huxley-type conductance-based single neurons [109]. This generalization may be important, e.g., the network stability may be different depending on whether single neurons are described by Hogkin-Huxley-type models or leaky integrate-and-fire models [26, 50].

- **More realistic synaptic dynamics.** The mean-field description of realistic synaptic interactions can be improved in at least two ways. First, synaptic fluctuations act through conductance changes, which are multiplied with the driving force $(V - E_{syn})$ to yield synaptic current. Therefore the noise is multiplicative. We have sidepassed this difficulty by replacing V with its average, so that the noise term becomes additive to the voltage equation. It would be desirable to be able to deal analytically with multiplicative noise. Second, synapses display short-term depression and facilitation [113, 131]. Mean-field models that incorporate synaptic depression have been investigated [115, 120], but the implications of short-term plasticity to recurrent networks, especially to working memory models, still await to be fully explored.

- **Extension to correlations between neurons.** In this chapter we have always assumed that the spiking activity of different cells was independent. Although the experimentally observed cross-correlations are relatively weak [14, 31, 74, 130], they might have a large impact on the input-output relationship of a neuron, since when the correlation coefficient of the inputs to a cell is not zero, the fluctuations in its total afferent synaptic are proportional to the number of inputs to the neuron, instead of to its square root, as in the models we have described. Although the analytic treatment of cross-correlations is technically complicated, a systematic characterization of their effect on the rate and variability of simple spiking neuron models is becoming available [38, 82, 101, 102]. The real challenge is to extend the framework here presented in such a way to include cross-correlations in a self-consistent way. A first step in this direction has been taken by [78], where cross-correlation functions in a recurrent fully connected recurrent network of spike response neurons [53] have been calculated.

Finally, let us end with a note on recurrent networks that display a *continuum* of stable neural firing patterns. Some working memory systems are believed to encode features of sensory stimuli that are *analog quantities* (such as spatial location, direction, eye position, frequency, etc). Such systems have been hypothesized to be implemented in the brain by recurrent neural networks endowed with a continuous family of attractor states. Heuristic firing-rate models of this kind have been widely investigated [5, 16, 28, 63, 91, 103, 125, 128]. More recently, more biophysical continuous attractor models of spiking neurons have been developed for spatial working memory [29], parametric working memory [80] and short-term memory in the oculomotor system [104]. Theoreticians have also begun to analyze mean-field models that are derived from these spiking neural network models [80, 104, 109]. Further progress in this direction will considerably advance our theoretical understanding of recurrent cortical networks, and shed insights into the cellular and network mechanisms of working memory.

Acknowledgments. A.R. thanks Ruben Moreno for helpful discussions. A.R. and X-J.W. are supported by NIMH, the Alfred Sloan Foundation and the Swartz Foundation. N.B. is supported by the C.N.R.S.

Appendix 1: The diffusion approximation

We will follow the exposition by [97]. We consider the case of a single post-synaptic neuron which receives C_E excitatory and C_I inhibitory independent Poisson inputs of rates ν_E and ν_I respectively, each delivering a charge J_E and J_I per spike though an "instantaneous" synaptic current (see above). For this discussion, we measure voltages with respect to V_L, i.e., $V_L = 0$. We will also measure the effect of each pre-synaptic spike by the size of the resulting instantaneous jump in the membrane potential $\bar{J}_{E,I} = J_{E,I}/C_m$. Since the process is Markov, it satisfies

$$\rho(V,t+\Delta t|V_0,t_0) = \int_{-\infty}^{\infty} dV' \rho(V,t+\Delta t|V',t)\rho(V',t|V_0,t_0). \qquad (15.98)$$

If Δt is sufficiently small, so that $\Delta t \ll \tau_m$, and so that the probability of receiving more than one spike in Δt is negligible, and since the pre-synaptic spikes produce discrete jumps, if follows that

$$\begin{aligned}\rho(V,t+\Delta t|V',t) = {} & [1-(C_E\nu_E+C_I\nu_I)\Delta t]\delta(V_0'-V)+ \\ & + C_E\nu_E\Delta t\delta(V_1'-V)+C_I\nu_I\Delta t\delta(V_2'-V),\end{aligned} \qquad (15.99)$$

where the first, second and third terms correspond to the probabilities of receiving no spikes, an excitatory spike or an inhibitory spike in Δt respectively, and $V_{0,1,2}'$ are the values of the depolarization at $t+\Delta t$ in these three cases, given that the depolarization was V' at t. In order to calculate $V_{0,1,2}'$, we use the fact that, since Δt is small enough, the exponential time course of V in between spikes can be approximated by a linear decay. Thus

$$\begin{aligned}V_0' &= V'(1-\frac{\Delta t}{\tau_m}) \\ V_1' &= V'(1-\frac{\Delta t_1}{\tau_m})+\bar{J}_E+[V'(1-\frac{\Delta t_1}{\tau_m})+\bar{J}_E]\frac{\Delta t_2}{\tau_m} \\ V_2' &= V'(1-\frac{\Delta t_1}{\tau_m})-\bar{J}_I+[V'(1-\frac{\Delta t_1}{\tau_m})-\bar{J}_I]\frac{\Delta t_2}{\tau_m},\end{aligned} \qquad (15.100)$$

where $\Delta t_1+\Delta t_2=\Delta t$. Using the property $\delta(f(x))=\delta(x-x')/|\partial_x f(x')|$ with x' such that $f(x')=0$, and expanding to first order in $\Delta t/\tau_m$, equation (15.98) can be expressed as

$$\begin{aligned}\rho(V,t+\Delta t|V_0,t_0) = {} & (1+\frac{\Delta t}{\tau_m})\left[(1-(C_E\nu_E+C_I\nu_I)\Delta t)\,\rho(V(1+\frac{\Delta t}{\tau_m}),t|V_0,t_0)+\right. \\ & + C_E\nu_E\Delta t\,\rho([V-\bar{J}_E](1+\frac{\Delta t}{\tau_m}),t|V_0,t_0)+ \\ & \left. + C_I\nu_I\Delta t\,\rho([V+\bar{J}_I](1+\frac{\Delta t}{\tau_m}),t|V_0,t_0)\right],\end{aligned} \qquad (15.101)$$

which, upon taking the limit $\Delta t \to 0$ becomes

$$\begin{aligned}\frac{\partial}{\partial t}\rho(V,t|V_0,t_0) = {} & \frac{\partial}{\partial V}[(\frac{V}{\tau_m})\rho(V,t|V_0,t_0)]+C_E\nu_E[\rho(V-\bar{J}_E,t|V_0,t_0) \qquad (15.102) \\ & - \rho(V,t|V_0,t_0)]+ + C_I\nu_I[\rho(V+\bar{J}_I,t|V_0,t_0)-\rho(V,t|V_0,t_0)].\end{aligned}$$

Finally, expressing the terms in square brackets as a Taylor series expansion around V, one obtains

$$\frac{\partial}{\partial t}\rho(V,t|V_0,t_0)=\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\,\frac{\partial^n}{\partial V^n}[A_n\,\rho(V,t|V_0,t_0)]. \qquad (15.103)$$

where

$$A_1(V) = -\frac{V}{\tau_m}+\bar{J}_E C_E\nu_E-\bar{J}_I C_I\nu_I \qquad (15.104)$$

$$A_n = \bar{J}_E^n C_E\nu_E+(-1)^n\bar{J}_I^n C_I\nu_I \qquad n=2,3,\ldots \qquad (15.105)$$

are called the infinitesimal moments of the process. The intuitive nature of the diffusion approximation becomes now clear: the smaller $\bar{J}_{E,I}$, the fewer the terms needed to express $\rho(V \mp \bar{J}_{E,I}, t|V_0, t_0)$ as a Taylor series expansion arround V, and the fewer the terms one has to maintain in the infinite-order Equation (15.103) to give an accurate description of the process.

Appendix 2: Stability of the steady-state solutions for $\rho_{ss}(V)$

The function $\rho_{ss}(V)$ is the solution of the stationary Fokker-Planck Equation (15.23) with the appropriate boundary conditions. To assess the dynamical stability of this solution, one has to use the general Fokker-Planck Equation (15.9) to test the effect of small perturbations on the steady-state distribution. We briefly mention the logic of this procedure.

For the sake of simplicity, we only discuss here a simple situation in which synapses are instantaneous, with a latency τ_l ms after the pre-synaptic spike time. In this case, the dynamical counterparts to Equations (15.63, 15.68) are

$$\tau_m \frac{\partial \rho}{\partial t} = \frac{\sigma_{ext}^2}{2} \frac{\partial^2 \rho}{\partial V^2} - \frac{\partial}{\partial V}\left[\left(V - \mu_{ext} - \bar{J}\tau_m \nu(t - \tau_l)\right)\rho\right], \tag{15.106}$$

in the weakly coupled, fully connected case (no noise in recurrent inputs), and

$$\begin{aligned}\tau_m \frac{\partial \rho}{\partial t} &= \frac{\left(\sigma_{ext}^2 + C\bar{J}^2\tau_m \nu(t-\delta)\right)}{2} \frac{\partial^2 \rho}{\partial V^2} \\ &\quad - \frac{\partial}{\partial V}\left[\left(V - \mu_{ext} - C\bar{J}\tau_m \nu(t - \tau_l)\right)\rho\right],\end{aligned} \tag{15.107}$$

in the strongly coupled, sparsely connected case (noise in recurrent inputs). In both cases the boundary conditions are given by Equations (15.19,15.20).

The stationary solution to Equation (15.106) (resp. 15.107) is Equation (15.63) (resp. 15.68). To study their stability, a linear stability analysis must be performed. It consists in looking for solutions to Equations (15.106, 15.107) of the form

$$\rho(V,t|V_0,t_0) = \rho_{ss}(V) + \delta\rho(V,\lambda)\exp(\lambda t) \tag{15.108}$$

$$\nu(t) = \nu_{ss} + \delta\nu(\lambda)\exp(\lambda t), \tag{15.109}$$

where ρ_{ss}, ν_{ss} correspond to the stationary solution, $\delta\rho$ and $\delta\nu$ are small deviations around the stationary solution that evolves in time with the (complex) growth rate λ. Upon inserting Equation (15.109) in Equations (15.106) or (15.107), and keeping the term first order in $\delta\rho$ and $\delta\nu$, an equation results for the possible growth rates λ. Solutions with $\mathrm{Re}(\lambda) > 0$ indicate that the stationary state is unstable. Instabilities might come about with a positive real eigenvalue: this is a mean rate instability, which typically occur in a network with strong recurrent excitation. Alternatively, an instability associated with a positive real part of a complex eigenvalue signals a Hopf bifurcation. If the bifurcation is supercritical, the network exhibits a synchronized oscillation with a frequency close to $\mathrm{Im}(\lambda)$. For more details on this approach, see [1] for the scenario with a simplified model with purely external noise, and [24] for a model with recurrent noise.

References

[1] L. F. Abbott and C. van Vreeswijk (1993), Asynchronous states in a network of pulse-coupled oscillators, *Phys. Rev. E,* **48**: 1483-1490.

[2] M. Abramowitz and I. A. Stegun (1970), *Tables of Mathematical Functions,* Dover Publications, NY.

[3] E. D. Adrian (1928), *The Basis of Sensation: The Action of the Sense Organs*, W. W. Norton: NY.

[4] B. Ahmed and J. C. Anderson and R. J. Douglas and K. A. Martin and J. C. Nelson, Polyneuronal innervation of spiny stellate neurons in cat visual cortex, *J. Comp. Neurol.,* **341**: 39-49.

[5] S. Amari (1977), Dynamics of pattern formation in lateral-inhibition type neural fields, *Biol. Cybern.,* **27**: 77-87.

[6] D. J. Amit (1995), The Hebbian paradigm reintegrated: local reverberations as internal representations, *Behav. Brain Sci.,* **18**: 617.

[7] D. J. Amit and M. V. Tsodyks (1992), Effective neurons and attractor neural networks in cortical environment, *Network*, **3**, 121-137.

[8] D. J. Amit and M. V. Tsodyks (1991), Quantitative study of attractor neural network retrieving at low spike rates I: Substrate – spikes, rates and neuronal gain, *Network*, **2**: 259-274.

[9] D. J. Amit and M. V. Tsodyks (1991), Quantitative study of attractor neural network retrieving at low spike rates II: Low-rate retrieval in symmetric networks, *Network*, **2**: 275.

[10] D. J. Amit and N. Brunel (1997), Dynamics of a recurrent network of spiking neurons before and following learning, *Network,* **8**: 373-404.

[11] D. J. Amit and N. Brunel (1997), Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex, *Cerebral Cortex*, **7**: 237-252.

[12] D. J. Amit and N. Brunel and M. V. Tsodyks (1994), Correlations of cortical

Hebbian reverberations: experiment *vs* theory, *J. Neurosci.,* **14**: 6435-6445.

[13] M. C. Angulo and J. Rossier and E. Audinat (1999), Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex, *J. Neurophysiol.,* **82**: 1295-1302.

[14] W. Bair and E. Zohary and W. T. Newsome (2001), Correlated firing in macaque visual area MT: time scales and relationship to behavior, *J. Neurosci,* **21**: 1676-1697.

[15] M. Bartos and I. Vida and M. Frotscher and J. R. P. Geiger and P. Jonas (2001), Rapid signaling at inhibitory synapses in a dendate gyrus interneuron network, *J. Neurosci.,* **21**: 2687-2698.

[16] R. Ben-Yishai and R. Lev Bar-Or and H. Sompolinsky (1995), Theory of orientation tuning in visual cortex, *Proc. Natl. Acad. Sci. USA,* **92**: 3844-3848.

[17] V. Braitenberg and A. Schutz (1991), *Anatomy of the Cortex*, Springer-Verlag.

[18] P. C. Bressloff and S. Coombes (2000), Dynamics of strongly coupled spiking neurons, *Neural Computation,* **12**: 91-129.

[19] N. Brunel (2000), Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.,* **8**: 183-208.

[20] N. Brunel (2000), Persistent activity and the single cell f-I curve in a cortical network model, *Network*, **11**: 261-280.

[21] N. Brunel and F. Chance and N. Fourcaud and L. Abbott (2001), Effects of synaptic noise and filtering on the frequency response of spiking neurons, *Phys. Rev. Lett.,* **86**: 2186-2189.

[22] N. Brunel and P. Latham (2003), Firing rate of noisy quadratic integrate-and-fire neurons, submitted manuscript.

[23] N. Brunel and S. Sergi (1998), Firing frequency of integrate-and-fire neurons with finite synaptic time constants, *J. Theor. Biol.,* **195**: 87-95.

[24] N. Brunel and V. Hakim (1999), Fast global oscillations in networks of integrate-and-fire neurons with low firing rates, *Neural Computation,* **11**: 1621-1671.

[25] N. Brunel and X. J. Wang (2001), Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition, *J. Comput. Neurosci.,* **11**: 63-85.

[26] N. Brunel and X.-J. Wang (2003), What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance, *J. Neurophysiol. in press*.

[27] M. A. Cohen and S. Grossberg (1983), Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *Transactions IEEE*, **SMC-13**: 815-826.

[28] M. Camperi and X.-J. Wang (1998), A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability, *J. Comput. Neurosci.*, **5**: 383-405.

[29] A. Compte and N. Brunel and P. S. Goldman-Rakic and X.-J. Wang (2000), Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model, *Cerebral Cortex,* **10**: 910-923.

[30] A. Compte and C. Constantinidis and J. Tegnér and S. Raghavachari and M. Chafee and P. S. Goldman-Rakic and X.-J. Wang (2002), Spectral properties of mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task, (Submitted to *J. Neurophysiol.*).

[31] C. Constantinidis and P. S. Goldman-Rakic (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex, *J. Neurophysiol.,* **88**: 3487-3497.

[32] A. Destexhe and D. Paré (1999), Impact of network activity on the integrative properties of neocortical pyramidal neurons *in vivo*, *J. Neurophysiol.,* **81**: 1531-1547.

[33] A. Destexhe and Z. F. Mainen and T. J. Sejnowski (1998), Kinetic models of synaptic transmission, in *Methods in Neuronal Modeling*, C. Koch and I. Segev (eds.), MIT Press, Cambridge, MA, 1-25.

[34] C. R. Doering and P. S. Hagan and C. D. Levermore(1987), Bistability driven by weakly colored gaussian-noise: the fokker-planck boundary layer and mean 1st-passage times, *Phys. Rev. Lett,* **59**: 2129-2132.

[35] C. A. Erickson and R. Desimone (1999), Responses of macaque perirhinal neurons during and after visual stimulus association learning, *J. Neurosci.*, **19**: 10404-10416.

[36] G. B. Ermentrout (1994), Reduction of conductance based models with slow synapses to neural nets, *Neural Computation,* **6**: 679-695.

[37] G. B. Ermentrout (1998), Neural networks as spatio-temporal pattern-forming systems, *Rep. Prog. Phys.*, **61**: 353-430.

[38] J. Feng and D. Brown (2000), Impact of correlated inputs on the output of the ingrate-and-fire model, *Neural Computation,* **12**: 671-92.

[39] D. Ferster and K. D. Miller (2000), Neural mechanisms of orientation selectivity in the visual cortex, *Annu. Rev. Neurosci.,* **23**: 441-471.

[40] N. Fourcaud and N. Brunel (2002), Dynamics of firing probability of noisy integrate-and-fire neurons, *Neural Computation,* **14**: 2057-2110.

[41] S. Funahashi and C. J. Bruce and P. S. Goldman-Rakic (1989), Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex, *J. Neurophysiol.*, **61**: 331-349.

[42] S. Funahashi and C. J. Bruce and P. S. Goldman-Rakic (1991), Neuronal activ-

ity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex, *J. Neurophysiol.*, **65**: 1464-1483.

[43] S. Funahashi and C. J. Bruce and P. S. Goldman-Rakic (1990), Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms, *J. Neurophysiol.*, **63**: 814-831.

[44] J. M. Fuster (1995), *Memory in the Cerebral Cortex,* MIT Press, Cambridge MA.

[45] J. M. Fuster (1973), Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory, *J. Neurophysiol.,* **36**: 61-78.

[46] J. M. Fuster and G. Alexander (1971), Neuron activity related to short-term memory, *Science*, *173*: 652-654.

[47] J. M. Fuster and J. P. Jervey (1981), Inferotemporal neurons distinguish and retain behaviourally relevant features of visual stimuli, *Science*, **212**: 952-955.

[48] J. M. Fuster and R. H. Bauer and J. P. Jervey (1982), Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks, *Exp. Neurol.,* **77**: 679-694.

[49] C. W. Gardiner (1986), *Handbook of Stochastic Methods for Physics Chemistry and the Natural Sciences,* Springer-Verlag.

[50] C. Geisler and N. Brunel and X.-J. Wang (2003), What determines the frequency of fast network oscillations with irregular neural discharges? II. Contributions of single cell membrane dynamics, in preparation.

[51] W. Gerstner (2000), Population dynamics of spiking neurons: fast transients, asynchronous states, and locking, *Neural Computation* **12**: 43-89.

[52] W. Gerstner (1995), Time structure of the activity in neural network models, *Phys. Rev. E ,* **51**: 738-758.

[53] W. Gerstner and W. M. Kistler (2002), *Spiking Neuron Models: Single Neurons, Populations, Plasticity,* Cambridge University Press.

[54] D. T. Gillespie (1992), *Markov Processes, an Introduction for Physical Scientists,* Academic Press.

[55] P. S. Goldman-Rakic (1995), Cellular basis of working memory, *Neuron,* **14**: 477-485.

[56] P. S. Goldman-Rakic (1987), Circuitry of primate prefrontal cortex and regulation of behavior by representational memory, in *Handbook of Physiology – The Nervous System V*, F. Plum and V. Mountcastle (eds.), Bethesda, Maryland: American Physiological Society, 373-417.

[57] J. Guckenheimer and P. Holmes (1983), *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Vields,* Springer Verlag.

[58] A. Gupta and Y. Wang and H. Markram (2000), Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex, *Science,* **287**: 273-278.

[59] M. Gur and A. Beylin and D. M. Snodderly (1997), Response variability of neurons in primary visual cortex (V1) of alert monkeys, *J. Neurosci.,* **17**: 2914-2920.

[60] B. S. Gutkin and C. R. Laing and C. L. Colby and C. C. Chow and G. B. Ermentrout (2001), Turning on and off with excitation: the role of spike time asynchrony and synchrony in sustained neural activity, *J. Comput. Neurosci.*, **11**: 121-134.

[61] P. S. Hagan and C. R. Doering and C. D. Levermore (1989), Mean exit times for particles driven by weakly colored noise, *SIAM J. Appl. Math.,* **49:** 1480-1513.

[62] D. Hansel and G. Mato (2003), Asynchronous states and the emergence of synchrony in large networks of interacting excitatory and inhibitory neurons, *Neural Comp.,* **15**: 1-56.

[63] D. Hansel and H. Sompolinsky (1998), Modeling feature selectivity in local cortical circuits, in *Methods in Neuronal Modeling,* C. Koch and I. Segev (eds.), MIT Press, Cambridge, MA.

[64] J. J. Hopfield (1984), Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. U.S.A.*, **81**: 3088-3092.

[65] B. Hutcheon and Y. Yarom (2000), Resonance, oscillation and the intrinsic frequency preferences of neurons, *TINS,* **23**: 216-222.

[66] C. E. Jahr and C. F. Stevens (1990), Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics, *J. Neurosci.,* **10**: 3178-3182.

[67] M. M. Klosek and P. S. Hagan (1998), Colored noise and a characteristic level crossing problem, *J. Math. Phys.,* **39**: 931-953.

[68] B. W. Knight (1972), Dynamics of encoding in a population of neurons, *J. Gen. Physiol.,* **59**: 734-766.

[69] C. Koch (1999), *Biophysics of Computation: Information Processing in Single Neurons*, Oxford University Press.

[70] K. W. Koch and J. M. Fuster (1989), Unit activity in monkey parietal cortex related to haptic perception and temporary memory, *Exp. Brain Res.,* **76**: 292-306.

[71] A. A. Koulakov and S. Raghavachari and A. Kepecs and J. E. Lisman, Model for a robust neural integrator, *Nat. Neurosci.*, **5**: 775-782.

[72] U. Kraushaar and P. Jonas (2000), Efficacy and stability of quantal GABA

release at a hippocampal interneuron-principal neuron synapse, *J. Neurosci.*, **20**: 5594-5607.

[73] K. Kubota and H. Niki (1971), Prefrontal cortical unit activity and delayed alternation performance in monkeys, *J. Neurophysiol.*, **34:** 337-347.

[74] D. Lee and N. L. Port and W. Kruse and A. P. Georgopoulos (1998), Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex, *J. Neurosci.,* **18**: 1161-1170.

[75] J. E. Lisman and J.-M. Fellous and X.-J. Wang (1998), A role for NMDA-receptor channels in working memory, *Nat. Neurosci.*, **1**: 273-275.

[76] K. A. Martin (2002), Microcircuits in the visual cortex, *Curr. Opin. Neurobiol.* **12**: 418-425.

[77] M. Mattia and P. Del Giudice (2002), Population dynamics of interacting spiking neurons, *Phys. Rev. E*, **66**: 051917.

[78] C. Meyer and C. van Vreeswijk (2002), Temporal correlations in stochastic networks of spiking neurons, *Neural Computation,* **14**: 369-404.

[79] E. K. Miller and C. A. Erickson and R. Desimone (1996), Neural mechanisms of visual working memory in prefrontal cortex of the macaque, *J. Neurosci.*, **16**: 5154-5167.

[80] P. Miller and C. Brody and R. Romo and X.-J. Wang (2003), A network model of parametric working memory, Submitted to *Cerebral Cortex*.

[81] Y. Miyashita (1988), Neuronal correlate of visual associative long-term memory in the primate temporal cortex, *Nature*, **335**: 817-820.

[82] R. Moreno and J. de la Rocha and A. Renart and N. Parga (2002), Response of spiking neurons to correlated inputs, *Phys. Rev. Lett.*, **89**: 288101.

[83] R. Moreno and N. Parga (2003), Response of a leaky integrate and fire neuron to white noise input filtered by synapses with an arbitraty time constant, in preparation.

[84] V. B. Mountcastle (1997), The cortical organization of the neocortex, *Brain*, **120**: 701-722.

[85] K. Nakamura and K. Kubota (1995), Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task, *J. Neurophysiol.*, **74**: 162-178.

[86] Y. Naya and K. Sakai and Y. Miyashita (1996), Activity of primate inferotemporal neurons related to a sought target in pair-association task, *Proc. Natl. Acad. Sci. USA*, **93**: 2664-2669.

[87] L. Nowak and P. Bregestovski and P. Ascher and A. Herbet and A. Prochiantz (1984), Magnesium gates glutamate-activated channels in mouse central neurones, *Nature*, **307**: 462-5.

[88] D. Q. Nykamp and D. Tranchina (2000), A population density approach that facilitates large-scale modeling of neural networks: analysis and an application to orientation tuning, *J. Comp. Neurosci.,* **8:** 19-30.

[89] S. P. Ó Scalaidhe and F. A. W. Wilson and P. S. Goldman-Rakic (1997), Areal segregation of face-processing neurons in prefrontal cortex, *Science*, **278**: 1135-1138.

[90] B. Pesaran and J. S. Pezaris and M. Sahani and P. P. Mitra and R. A. Andersen (2002), Temporal structure in neuronal activity during working memory in macaque parietal cortex, *Nat. Neurosci.*, **5**: 805-811.

[91] A. D. Redish and A. N. Elga and D. S. Touretzky (1996), A coupled attractor model of the rodent head direction system, *Network,* **7**: 671-685.

[92] S. G. Rao and G. V. Williams and P. S. Goldman-Rakic (1999), Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC, *J. Neurophysiol.*, **81**: 1903-1916.

[93] A. Renart (2000), *Multi-Modular Memory Systems*, Universidad Autónoma de Madrid.

[94] A. Renart, J. de la Rocha, N. Parga, and E.T. Rolls (2001). A model of the IT-PF network in object working memory which includes balanced persistent activity and tuned inhibition. *Neurocomputing* **38-40**, 1525-1531.

[95] A. Renart and P. Song and X. J. Wang (2003), Homeostatic synaptic plasticity leads to robust spatial working memory without fine-tuning of cellular or network properties, Submitted to *Neuron*.

[96] A. Renart and R. Moreno and X. J. Wang and N. Parga (2003), Bistability in balanced recurrent networks, submitted.

[97] L. M. Ricciardi (1977), *Diffusion Processes and Related Topics on Biology* Springer-Verlag, Berlin.

[98] M. Richardson and N. Brunel and V. Hakim (2003), From subthreshold to firing-rate resonance, *J. Neurophysiol.,* **89**: in press.

[99] H. Risken (1984), *The Fokker-Planck Equation: Methods of Solution and Applications,* Springer-Verlag, Berlin.

[100] K. Sakai and Y. Miyashita (1991), Neural organization for the long-term memory of paired associates, *Nature*, **354**: 152-155.

[101] E. Salinas and T. J. Sejnowski (2000), Impact of correlated synaptic input on output firing rate and variability in simple neuronal models, *Journal of Neuroscience*, **20**: 6193-6209.

[102] E. Salinas and T. J. Sejnowski (2002), Integrate-and-fire neurons driven by correlated stochastic input, *Neural Computation,* **14**: 2111-2155.

[103] H. S. Seung (1996), How the brain keeps the eyes still, *Proc. Natl. Acad. Sci. USA,* **93**: 13339-13344.

[104] H. S. Seung and D. D. Lee and B. Y. Reis and D. W. Tank (2000), Stability of the memory of eye position in a recurrent network of conductance-based model neurons, *Neuron,* **26**: 259-271.

[105] M. N. Shadlen and W. T. Newsome (1994), Noise, neural codes and cortical organization, *Current opinion in Neurobiol.,* **4**: 569-579.

[106] M. N. Shadlen and W. T. Newsome (1998), The variable discharge of cortical neurons: implications for connectivity, computation, and information coding, *J. Neurosci.,* **18**: 3870-3896.

[107] M. J. Shelley and D. McLaughlin and R. Shapley and J. Wielaard (2002), States of high conductance in a large-scale model of the visual cortex, *J. Comput. Neurosci.,* **13**: 93-109.

[108] S. Shinomoto and Y. Sakai and S. Funahashi (1999), The Ornstein-Uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex, *Neural Comput.*, **11**: 935-51.

[109] O. Shriki and D. Hansel and H. Sompolinsky (2003), Rate models for conductance based cortical neuronal networks, *Neural Comput.*, **15**: in press.

[110] W. R. Softky and C. Koch (1993), The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs, *J. Neurosci.*, **13**: 334-350.

[111] H. Sompolinsky and R. Shapley (1997), New perspectives on the mechanisms for orientation selectivity, *Curr. Opin. Neurobiol.,* **7**: 514-522.

[112] J. Tegnér and A. Compte and X.-J. Wang (2002), Dynamical stability of reverberatory neural circuits, *Biol. Cybern.* **87**: 471-481.

[113] A. M. Thomson (2000), Facilitation, augmentation and potentiation at central synapses, *Trends in Neurosciences,* **23**: 305-312.

[114] A. Treves (1993), Mean-field analysis of neuronal spike dynamics, *Network,* **4**: 259-284.

[115] M. V. Tsodyks and K. Pawelzik and H. Markram (1998), Neural networks with dynamic synapses, *Neural Computation,* **10**: 821-835.

[116] M. V. Tsodyks and and W. E. Skaggs and T. J. Sejnowski and B. L. McNaughton (1997), Paradoxical effects of external modulation of inhibitory interneurons, *J. Neurosci.*, **17**: 4382-4388.

[117] H. C. Tuckwell (1988), *Introduction to Theoretical Neurobiology,* Cambridge: Cambridge University Press.

[118] C. van Vreeswijk and H. Sompolinsky (1996), Chaos in neuronal networks with balanced excitatory and inhibitory activity, *Science*, **274**: 1724-1726.

[119] C. van Vreeswijk and H. Sompolinsky (1998), Chaotic balanced state in a model of cortical circuits, *Neural Computation,* **10**: 1321-1371.

[120] X.-J. Wang (1999), Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory, *J. Neurosci.,* **19**: 9587-9603.

[121] X.-J. Wang (2001), Synaptic reverberation underlying mnemonic persistent activity, *Trends Neurosci.,* **24**: 455-463.

[122] X.-J. Wang and G. Buzsáki (1996), Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model, *J. Neurosci.,* **16**: 6402-6413.

[123] N. Wax (1954), *Selected Papers on Noise and Stochastic Processes,* Dover Publications.

[124] G. V. Williams and P. S. Goldman-Rakic (1995), Modulation of memory fields by dopamine D1 receptors in prefrontal cortex, *Nature*, **376**: 572-575.

[125] H. R. Wilson and J. D. Cowan (1972), Excitatory and inhibitory interactions in localized populations of model neurons, *Biophys. J.,* **12**: 1-24.

[126] H. R. Wilson and J. D. Cowan (1973), A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, *Kybernetik,* **13**: 55-80.

[127] Z. Xiang and J. R. Huguenard and D. A. Prince (1998), GABA$_A$ receptor mediated currents in interneurons and pyramidal cells of rat visual cortex, *J. Physiol.,* **506**: 715-730.

[128] K. Zhang (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensembles: A theory, *J. Neurosci.,* **16**: 2112-2126.

[129] F.-M Zhou and J. J. Hablitz (1998), AMPA receptor-mediated EPSCs in rat neocortical layer II/III interneurons have rapid kinetics, *Brain Research,* **780**: 166-169.

[130] E. Zohary and M. N. Shadlen and W. T. Newsome (1994), Correlated neuronal discharge rate and its implications for psychophysical performance, *Nature*, **370**: 140-3.

[131] R. S. Zucker and W. G. Regehr (2002), Short-term synaptic plasticity, *Annu. Rev. Physiol.*, **64**: 355-405.

Chapter 16

The Operation of Memory Systems in the Brain

Edmund T. Rolls
University of Oxford, Dept. of Experimental Psychology, South Parks Road, Oxford OX1 3UD, England. www.cns.ox.ac.uk

CONTENTS

16.1 Introduction

This chapter describes memory systems in the brain based on closely linked neurobiological and computational approaches. The neurobiological approaches include evidence from brain lesions which show the type of memory for which each of the brain systems considered is necessary; and analysis of neuronal activity in each of these systems to show what information is represented in them, and the changes that take place during learning. Much of the neurobiology considered is from non-human primates as well as humans, because the operation of some of the brain systems involved in memory and connected to them have undergone great development in primates. Some such brain systems include those in the temporal lobe, which develops massively in primates for vision, and which sends inputs to the hippocampus via highly developed parahippocampal regions; and the prefrontal cortex. Many memory systems in primates receive outputs from the primate inferior temporal visual cortex, and understanding the perceptual representations in this of objects, and how they are appropriate as inputs to different memory systems, helps to provide a coherent way to understand the different memory systems in the brain (see [82], which provides a more extensive treatment of the brain architectures used for perception and memory). The computational approaches are essential in order to understand how the circuitry could retrieve as well as store memories, the capacity of each memory system in the brain, the interactions between memory and perceptual systems, and the speed of operation of the memory systems in the brain.

The architecture, principles of operation, and properties of the main types of network referred to here, autoassociation or attractor networks, pattern association networks, and competitive networks, are described by [82] and [92].

16.2 Functions of the hippocampus in long-term memory

The inferior temporal visual cortex projects via the perirhinal cortex and entorhinal cortex to the hippocampus (see Figure 16.1), which is implicated in long-term memory, of, for example, where objects are located in spatial scenes, which can be thought of as an example of episodic memory. The architecture shown in Figure 16.1 indicates that the hippocampus provides a region where visual outputs from the inferior temporal visual cortex can, via the perirhinal cortex and entorhinal cortex, be brought together with outputs from the ends of other cortical processing streams. In this section, we consider how the visual input about objects is in the correct form for the types of memory implemented by the perirhinal and hippocampal systems, how the hippocampus of primates contains a representation of the visual space being viewed, how this may be similar computationally to the apparently very different representation of places that is present in the rat hippocampus, how these spatial representations are in a form that could be implemented by a continuous attractor which could be updated in the dark by idiothetic inputs, and how a unified attractor theory of hippocampal function can be formulated using the concept of mixed attractors. The visual output from the inferior temporal visual cortex may be used to provide the perirhinal and hippocampal systems with information about objects that is useful in visual recognition memory, in episodic memory of where objects are seen, and for building spatial representations of visual scenes. Before summarizing the computational approaches to these issues, we first summarize some of the empirical evidence that needs to be accounted for in computational models.

16.2.1 Effects of damage to the hippocampus and connected structures on object-place and episodic memory

Partly because of the evidence that in humans with bilateral damage to the hippocampus and nearby parts of the temporal lobe, anterograde amnesia is produced [100], there is continuing great interest in how the hippocampus and connected structures operate in memory. The effects of damage to the hippocampus indicate that the very long-term storage of at least some types of information is not in the hippocampus, at least in humans. On the other hand, the hippocampus does appear to be necessary to learn certain types of information, that have been characterized as declarative, or knowing that, as contrasted with procedural, or knowing how, which is spared in amnesia. Declarative memory includes what can be declared or brought to mind as a proposition or an image. Declarative memory includes episodic memory (memory for particular episodes), and semantic memory (memory for facts) [100].

In monkeys, damage to the hippocampus or to some of its connections such as the fornix produces deficits in learning about where objects are and where responses must be made (see [12]) and [76]. For example, macaques and humans with damage to the hippocampus or fornix are impaired in object-place memory tasks in which not only the objects seen, but where they were seen, must be remembered [28, 60, 99]. Such object-place tasks require a whole-scene or snapshot-like memory [25]. Also, fornix lesions impair conditional left-right discrimination learning, in which the visual appearance of an object specifies whether a response is to be made to the left or the right [94]. A comparable deficit is found in humans [61]. Fornix sectioned monkeys are also impaired in learning on the basis of a spatial cue which object to choose (e.g., if two objects are on the left, choose object A, but if the two objects are on the right, choose object B) [26]. Further, monkeys with fornix damage are also impaired in using information about their place in an environment. For example, [27] found learning impairments when the position of the monkey in the room determined which of two or more objects the monkey had to choose. Rats with hippocampal lesions are impaired in using environmental spatial cues to remember particular places [35, 45], and it has been argued that the necessity to utilize allocentric spatial cues [14], to utilize spatial cues or bridge delays [34, 37], or to perform relational operations on remembered material [19], may be characteristic of the deficits.

One way of relating the impairment of spatial processing to other aspects of hip-

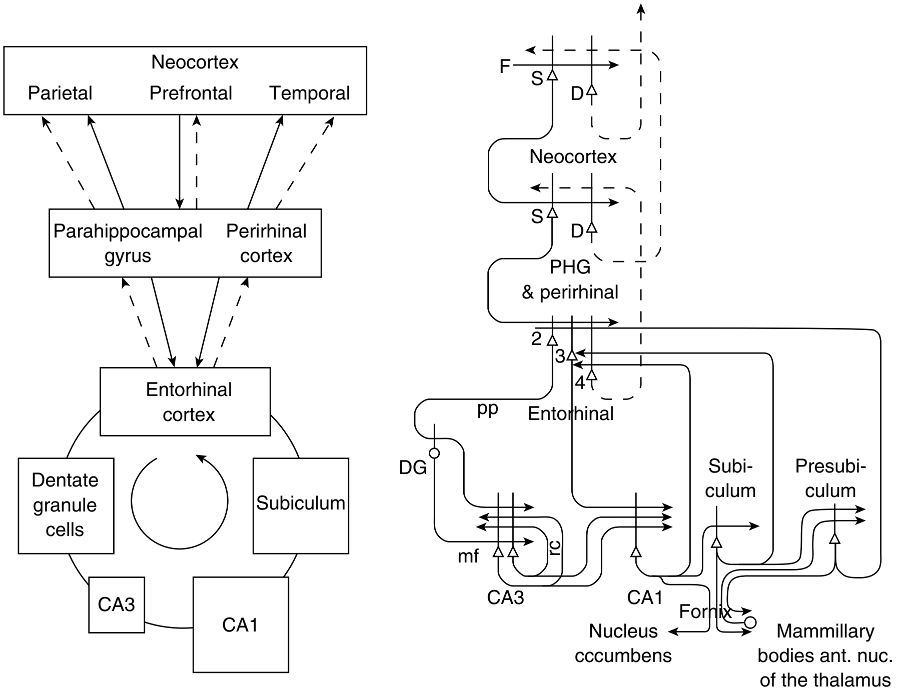

Figure 16.1

Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells; and great divergence again in the backprojections. Left: block diagram. Right: more detailed representation of some of the principal excitatory neurons in the pathways. Abbreviations: D, Deep pyramidal cells; DG, dentate granule cells; F, forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy. mf: mossy fibres; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collaterals of the CA3 hippocampal pyramidal cells; S, superficial pyramidal cells; 2, pyramidal cells in layer 2 of the entorhinal cortex; 3, pyramidal cells in layer 3 of the entorhinal cortex; 5, 6, pyramidal cells in the deep layers of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites.

pocampal function (including the memory of recent events or episodes in humans) is to note that this spatial processing involves a snapshot type of memory, in which one whole scene with its often unique set of parts or elements must be remembered. This memory may then be a special case of episodic memory, which involves an arbitrary association of a set of spatial and/or non-spatial events that describe a past episode. For example, the deficit in paired associate learning in humans (see [100]) may be especially evident when this involves arbitrary associations between words, for example, window — lake.

It appears that the deficits in 'recognition' memory (tested for example for visual stimuli seen recently in a delayed match to sample task) produced by damage to this brain region are related to damage to the perirhinal cortex [121, 122], which receives from high order association cortex and has connections to the hippocampus (see Figure 16.1) [106, 107]. The functions of the perirhinal cortex in memory are discussed by [82].

16.2.2 Neurophysiology of the hippocampus and connected areas

In the rat, many hippocampal pyramidal cells fire when the rat is in a particular place, as defined for example by the visual spatial cues in an environment such as a room [39, 53, 54]. There is information from the responses of many such cells about the place where the rat is in the environment. When a rat enters a new environment B connected to a known environment A, there is a period in the order of 10 minutes in which as the new environment is learned, some of the cells that formerly had place fields in A develop instead place fields in B. It is as if the hippocampus sets up a new spatial representation which can map both A and B, keeping the proportion of cells active at any one time approximately constant [117]. Some rat hippocampal neurons are found to be more task-related, responding for example to olfactory stimuli to which particular behavioural responses must be made [19], and some of these neurons may in different experiments show place-related responses.

It was recently discovered that in the primate hippocampus, many spatial cells have responses not related to the place where the monkey is, but instead related to the place where the monkey is looking [78, 79, 85]. These are called 'spatial view cells', an example of which is shown in Figure 16.2. These cells encode information in allocentric (world-based, as contrasted with egocentric, body-related) coordinates [29, 93]. They can in some cases respond to remembered spatial views in that they respond when the view details are obscured, and use idiothetic (self-motion) cues including eye position and head direction to trigger this memory recall operation [71]. Another idiothetic input that drives some primate hippocampal neurons is linear and axial whole body motion [58], and in addition, the primate presubiculum has been shown to contain head direction cells [72].

Part of the interest of spatial view cells is that they could provide the spatial representation required to enable primates to perform object-place memory, for example remembering where they saw a person or object, which is an example of an episodic memory, and indeed similar neurons in the hippocampus respond in object-place memory tasks [84]. Associating together such a spatial representation with a repre-

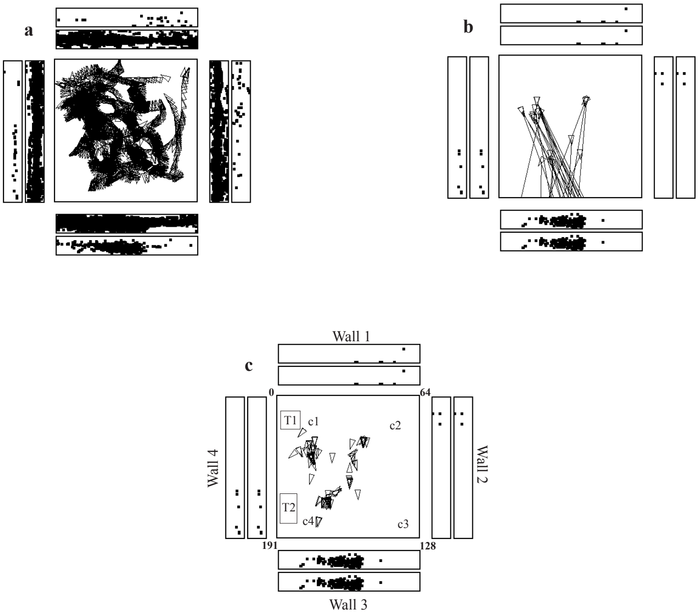

Figure 16.2

Examples of the firing of a hippocampal spatial view cell when the monkey was walking around the laboratory. a. The firing of the cell is indicated by the spots in the outer set of 4 rectangles, each of which represents one of the walls of the room. There is one spot on the outer rectangle for each action potential. The base of the wall is towards the centre of each rectangle. The positions on the walls fixated during the recording sessions are indicated by points in the inner set of 4 rectangles, each of which also represents a wall of the room. The central square is a plan view of the room, with a triangle printed every 250 ms to indicate the position of the monkey, thus showing that many different places were visited during the recording sessions. b. A similar representation of the same 3 recording sessions as in (a), but modified to indicate some of the range of monkey positions and horizontal gaze directions when the cell fired at more than 12 spikes/s. c. A similar representation of the same 3 recording sessions as in (b), but modified to indicate more fully the range of places when the cell fired. The triangle indicates the current position of the monkey, and the line projected from it shows which part of the wall is being viewed at any one time while the monkey is walking. One spot is shown for each action potential. (After Georges-François, Rolls and Robertson, 1999)

sentation of a person or object could be implemented by an autoassociation network implemented by the recurrent collateral connections of the CA3 hippocampal pyramidal cells [75, 76, 92]. Some other primate hippocampal neurons respond in the object-place memory task to a combination of spatial information and information about the object seen [84]. Further evidence for this convergence of spatial and object information in the hippocampus is that in another memory task for which the hippocampus is needed, learning where to make spatial responses conditional on which picture is shown, some primate hippocampal neurons respond to a combination of which picture is shown, and where the response must be made [13, 48].

These primate spatial view cells are thus unlike place cells found in the rat [39, 51, 53, 54, 117]. Primates, with their highly developed visual and eye movement control systems, can explore and remember information about what is present at places in the environment without having to visit those places. Such spatial view cells in primates would thus be useful as part of a memory system, in that they would provide a representation of a part of space that would not depend on exactly where the monkey or human was, and that could be associated with items that might be present in those spatial locations. An example of the utility of such a representation in humans would be remembering where a particular person had been seen. The primate spatial representations would also be useful in remembering trajectories through environments, of use for example in short-range spatial navigation [58, 79].

The representation of space in the rat hippocampus, which is of the place where the rat is, may be related to the fact that with a much less developed visual system than the primate, the rat's representation of space may be defined more by the olfactory and tactile as well as distant visual cues present, and may thus tend to reflect the place where the rat is. An interesting hypothesis on how this difference could arise from essentially the same computational process in rats and monkeys is as follows [17, 79]. The starting assumption is that in both the rat and the primate, the dentate granule cells and the CA3 and CA1 pyramidal cells respond to combinations of the inputs received. In the case of the primate, a combination of visual features in the environment will over a typical viewing angle of perhaps 10–20 degrees result in the formation of a spatial view cell, the effective trigger for which will thus be a combination of visual features within a relatively small part of space. In contrast, in the rat, given the very extensive visual field which may extend over 180–270 degrees, a combination of visual features formed over such a wide visual angle would effectively define a position in space, that is a place. The actual processes by which the hippocampal formation cells would come to respond to feature combinations could be similar in rats and monkeys, involving for example competitive learning in the dentate granule cells, autoassociation learning in CA3 pyramidal cells, and competitive learning in CA1 pyramidal cells [75, 76, 92, 115]. Thus spatial view cells in primates and place cells in rats might arise by the same computational process but be different by virtue of the fact that primates are foveate and view a small part of the visual field at any one time, whereas the rat has a very wide visual field. Although the representation of space in rats therefore may be in some ways analogous to the representation of space in the primate hippocampus, the difference does have implications for theories, and modelling, of hippocampal function.

In rats, the presence of place cells has led to theories that the rat hippocampus is a spatial cognitive map, and can perform spatial computations to implement navigation through spatial environments [10, 11, 54, 57]. The details of such navigational theories could not apply in any direct way to what is found in the primate hippocampus. Instead, what is applicable to both the primate and rat hippocampal recordings is that hippocampal neurons contain a representation of space (for the rat, primarily where the rat is, and for the primate primarily of positions ‘out there’ in space) which is a suitable representation for an episodic memory system. In primates, this would enable one to remember, for example, where an object was seen. In rats, it might enable memories to be formed of where particular objects (for example those defined by olfactory, tactile, and taste inputs) were found. Thus at least in primates, and possibly also in rats, the neuronal representation of space in the hippocampus may be appropriate for forming memories of events (which usually in these animals have a spatial component). Such memories would be useful for spatial navigation, for which according to the present hypothesis the hippocampus would implement the memory component but not the spatial computation component. Evidence that what neuronal recordings have shown is represented in the non-human primate hippocampal system may also be present in humans is that regions of the hippocampal formation can be activated when humans look at spatial views [21, 55].

16.2.3 Hippocampal models

These neuropsychological and neurophysiological analyses are complemented by neuronal network models of how the hippocampus could operate to store and retrieve large numbers of memories [73, 75, 76, 92, 114, 115]). One key hypothesis (adopted also by [46]) is that the hippocampal CA3 recurrent collateral connections which spread throughout the CA3 region provide a *single autoassociation network* that enables the firing of *any* set of CA3 neurons representing one part of a memory to be associated together with the firing of any other set of CA3 neurons representing another part of the same memory (cf. [44]). The generic architecture of an attractor network is shown in Figure 16.5. Associatively modifiable synapses in the recurrent collateral synapses allow memories to be stored, and then later retrieved from only a part, as described by [4, 32, 33, 82, 92]. The number of patterns p each representing a different memory that could be stored in the CA3 system operating as an autoassociation network would be as shown in equation 16.1 (see [82, 92], which describe extensions to the analysis developed by [33]).

$$p \approx \frac{C^{\mathrm{RC}}}{a\ln(\frac{1}{a})}k \qquad (16.1)$$

where C^{RC} is the number of synapses on the dendrites of each neuron devoted to the recurrent collaterals from other CA3 neurons in the network, a is the sparseness of the representation, and k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3. Given that C^{RC} is approximately 12,000 in the rat, the resulting storage capacity would be greater than 12,000 memories, and perhaps up to 36,000 memories if the sparseness a of the representation was as low as 0.02 [114, 115].

Another part of the hypothesis is that the very sparse (see Figure 16.3) but powerful connectivity of the mossy fibre inputs to the CA3 cells from the dentate granule

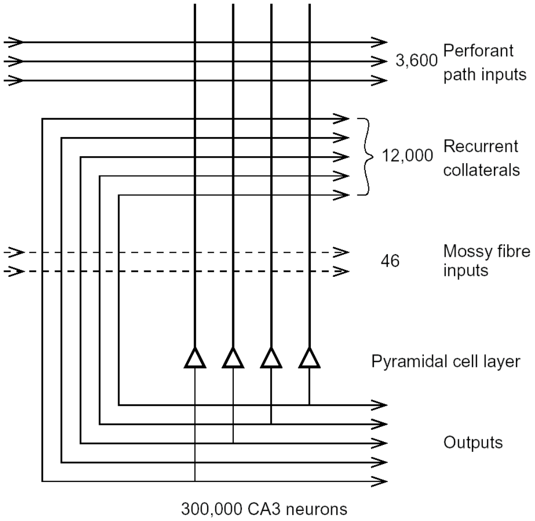

Figure 16.3

The numbers of connections from three different sources onto each CA3 cell from three different sources in the rat. (After Treves and Rolls 1992, and Rolls and Treves 1998.)

cells is important during learning (but not recall) to force a new, arbitrary, set of firing onto the CA3 cells which dominates the activity of the recurrent collaterals, so enabling a new memory represented by the firing of the CA3 cells to be stored [73, 75, 114].

The perforant path input to the CA3 cells, which is numerically much larger but at the apical end of the dendrites, would be used to initiate recall from an incomplete pattern [92, 114]. The prediction of the theory about the necessity of the mossy fibre inputs to the CA3 cells during learning but not recall has now been confirmed [42]. A way to enhance the efficacy of the mossy fibre system relative to the CA3 recurrent collateral connections during learning may be to increase the level of acetyl choline by increasing the firing of the septal cholinergic cells [31].

Another key part of the quantitative theory is that not only can retrieval of a memory by an incomplete cue be performed by the operation of the associatively modified CA3 recurrent collateral connections, but also that recall of that information to the neocortex can be performed via CA1 and the hippocampo-cortical and cortico-cortical backprojections [76, 81, 92, 115] shown in Figure 16.1. In this case, the number of memory patterns p^{BP} that can be retrieved by the backprojection system is

$$p^{\mathrm{BP}} \approx \frac{C^{\mathrm{BP}}}{a^{\mathrm{BP}} \ln(\frac{1}{a^{\mathrm{BP}}})} k^{\mathrm{BP}} \tag{16.2}$$

where C^{BP} is the number of synapses on the dendrites of each neuron devoted to backprojections from the preceding stage (dashed lines in Figure 16.1), a^{BP} is the sparseness of the representation in the backprojection pathways, and k^{BP} is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3. The insight into this quantitative analysis came from treating each layer of the backprojection hierarchy as being quantitatively equivalent to another iteration in a single recurrent attractor network [113, 115]. The need for this number of connections to implement recall, and more generally constraint satisfaction in connected networks (see [82]), provides a fundamental and quantitative reason for why there are approximately as many backprojections as forward connections between the adjacent connected cortical areas in a cortical hierarchy. This, and other computational approaches to hippocampal function, are included in a special issue of the journal *Hippocampus* (1996), 6(6).

Another aspect of the theory is that the operation of the CA3 system to implement recall, and of the backprojections to retrieve the information, would be sufficiently fast, given the fast recall in associative networks built of neurons with continuous dynamics (see [82]).

16.2.4 Continuous spatial representations, path integration, and the use of idiothetic inputs

The fact that spatial patterns, which imply continuous representations of space, are represented in the hippocampus has led to the application of continuous attractor models to help understand hippocampal function. Such models have been developed by [8, 95, 101, 102, 104, 105], (see [82]). Indeed, we have shown how a continuous attractor network could enable the head direction cell firing of presubicular cells to be maintained in the dark, and updated by idiothetic (self-motion) head rotation cell inputs [72, 101]. The continuous attractor model has been developed to understand how place cell firing in rats can be maintained and updated by idiothetic inputs in the dark [104]. The continuous attractor model has also been developed to understand how spatial view cell firing in primates can be maintained and updated by idiothetic eye movement and head direction inputs in the dark [71, 105].

The way in which path integration could be implemented in the hippocampus or related systems is described next. Single-cell recording studies have shown that some neurons represent the current position along a continuous physical dimension or space even when no inputs are available, for example in darkness. Examples include neurons that represent the positions of the eyes (i.e., eye direction with respect to the head), the place where the animal is looking in space, head direction, and the place where the animal is located. In particular, examples of such classes of cells include head direction cells in rats [50, 62, 109, 110] and primates [72], which respond maximally when the animal's head is facing in a particular preferred direction; place cells in rats [43, 47, 49, 52, 56] that fire maximally when the animal is in a particular location; and spatial view cells in primates that respond when the monkey is looking towards a particular location in space [29, 71, 85]. In the parietal cortex there are many spatial representations, in several different coordinate frames (see [6] and [82]), and they have some capability to remain active during memory periods when the stimulus is no longer present. Even more than this, the dorsolateral prefrontal cortex networks to which the parietal networks project have the capability to maintain spatial representations active for many seconds or minutes during short-term memory tasks, when the stimulus is no longer present (see below).

A class of network that can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial position, head direction, etc is a ‘Continuous Attractor’ neural network (CANN). It uses excitatory recurrent collateral connections between the neurons to reflect the distance between the neurons in the state space of the animal (e.g., head direction space). These networks can maintain the bubble of neural activity constant for long periods wherever it is started to represent the current state (head direction, position, etc) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision. Global inhibition is used to keep the number of neurons in a bubble or packet of actively firing neurons relatively constant, and to help to ensure that there is only one activity packet. Continuous attractor networks can be thought of as very similar to autoassociation or discrete attractor networks (see [82]), and have the same architecture, as illustrated in Figure 16.5. The main difference is that the patterns stored in a CANN are continuous patterns, with each neuron having broadly tuned firing which decreases with for example a Gaussian function as the distance from the optimal firing location of the cell is varied, and with different neurons having tuning that overlaps throughout the space. Such tuning is illustrated in Figure 16.4. For comparison, autoassociation networks normally have discrete (separate) patterns (each pattern implemented by the firing of a particular subset of the neurons), with no continuous distribution of the patterns throughout the space (see Figure 16.4). A consequent difference is that the CANN can maintain its firing at any location in the trained continuous space, whereas a discrete attractor or autoassociation network moves its population of active neurons towards one of the previously learned attractor states, and thus implements the recall of a particular previously learned pattern from an incomplete or noisy (distorted) version of one of the previously learned patterns. The energy landscape of a discrete attractor network (see [82]) has separate energy minima, each one of which corresponds to a learned pattern, whereas the energy landscape of a continuous attractor network is flat, so that the activity packet remains stable with continuous firing wherever it is started in the state space. (The state space refers to set of possible spatial states of the animal in its environment, e.g., the set of possible head directions.) I next describe the operation and properties of continuous attractor networks, which have been studied by for example [3], [111], and [119], and then, following [101], address four key issues about the biological application of continuous attractor network models.

One key issue in such continuous attractor neural networks is how the synaptic strengths between the neurons in the continuous attractor network could be learned in biological systems (Section 16.2.4.2).

A second key issue in such Continuous Attractor neural networks is how the bubble of neuronal firing representing one location in the continuous state space should be updated based on non-visual cues to represent a new location in state space. This is essentially the problem of path integration: how a system that represents a mem-

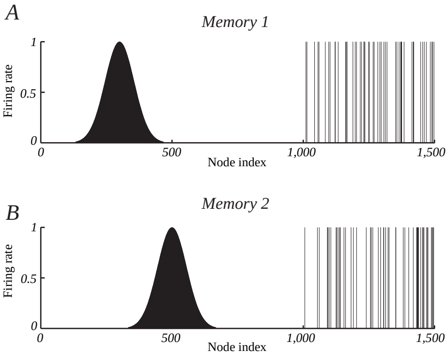

Figure 16.4

The types of firing patterns stored in continuous attractor networks are illustrated for the patterns present on neurons 1–1000 for Memory 1 (when the firing is that produced when the spatial state represented is that for location 300), and for Memory 2 (when the firing is that produced when the spatial state represented is that for location 500). The continuous nature of the spatial representation results from the fact that each neuron has a Gaussian firing rate that peaks at its optimal location. This particular mixed network also contains discrete representations that consist of discrete subsets of active binary firing rate neurons in the range 1001–1500. The firing of these latter neurons can be thought of as representing the discrete events that occur at the location. Continuous attractor networks by definition contain only continuous representations, but this particular network can store mixed continuous and discrete representations, and is illustrated to show the difference of the firing patterns normally stored in separate continuous attractor and discrete attractor networks. For this particular mixed network, during learning, Memory 1 is stored in the synaptic weights, then Memory 2, etc., and each memory contains part that is continuously distributed to represent physical space, and part that represents a discrete event or object.

ory of where the agent is in physical space could be updated based on idiothetic (self-motion) cues such as vestibular cues (which might represent a head velocity signal), or proprioceptive cues (which might update a representation of place based on movements being made in the space, during for example walking in the dark).

A third key issue is how stability in the bubble of activity representing the current location can be maintained without much drift in darkness, when it is operating as a memory system (see [82] and [101]).

A fourth key issue is considered below in which I describe networks that store both continuous patterns and discrete patterns (see Figure 16.4), which can be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present.

16.2.4.1 The generic model of a continuous attractor network

The generic model of a continuous attractor is as follows. (The model is described in the context of head direction cells, which represent the head direction of rats [50, 109] and macaques [72], and can be reset by visual inputs after gradual drift in darkness.) The model is a recurrent attractor network with global inhibition. It is different from a Hopfield attractor network [33] primarily in that there are no discrete attractors formed by associative learning of discrete patterns. Instead there is a set of neurons that are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g., head directions) represented by the neurons. Neurons that represent similar states (locations in the state space) of the agent in the physical world have strong synaptic connections, which can be set up by an associative learning rule, as described in Section 16.2.4.2. The network updates its firing rates by the following 'leaky-integrator' dynamical equations. The continuously changing activation h_i^{HD} of each head direction cell i is governed by the Equation

$$\frac{dh_i^{\mathrm{HD}}(t)}{dt} = -h_i^{\mathrm{HD}}(t) + \frac{\phi_0}{C^{\mathrm{HD}}} \sum_j (w_{ij} - w^{\mathrm{inh}}) r_j^{\mathrm{HD}}(t) + I_i^V, \qquad (16.3)$$

where r_j^{HD} is the firing rate of head direction cell j, w_{ij} is the excitatory (positive) synaptic weight from head direction cell j to cell i, w^{inh} is a global constant describing the effect of inhibitory interneurons, and τ is the time constant of the system*. The term $-h_i^{\mathrm{HD}}(t)$ indicates the amount by which the activation decays (in the leaky integrator neuron) at time t. (The network is updated in a typical simulation at much smaller timesteps than the time constant of the system, τ.) The next term in Equation (16.3) is the input from other neurons in the network r_j^{HD} weighted by the recurrent collateral synaptic connections w_{ij} (scaled by a constant ϕ_0 and C^{HD} which is the number of synaptic connections received by each head direction cell from other head direction cells in the continuous attractor). The term I_i^V represents a visual input to head direction cell i. Each term I_i^V is set to have a Gaussian response profile in most continuous attractor networks, and this sets the firing of the cells in the continuous attractor to have Gaussian response profiles as a function of where the agent is located in the state space (see e.g., Figure 16.4), but the Gaussian assumption is not crucial. (It is known that the firing rates of head direction cells in both rats [50, 109] and macaques [72] is approximately Gaussian.) When the agent is operating without visual input, in memory mode, then the term I_i^V is set to zero. The firing rate r_i^{HD} of cell i is determined from the activation h_i^{HD} and the sigmoid function

$$r_i^{\mathrm{HD}}(t) = \frac{1}{1 + e^{-2\beta(h_i^{\mathrm{HD}}(t) - \alpha)}}, \qquad (16.4)$$

where α and β are the sigmoid threshold and slope, respectively.

*Note that here I use r rather than y to refer to the firing rates of the neurons in the network, remembering that, because this is a recurrently connected network (see Figure 16.5), the output from a neuron y_i might

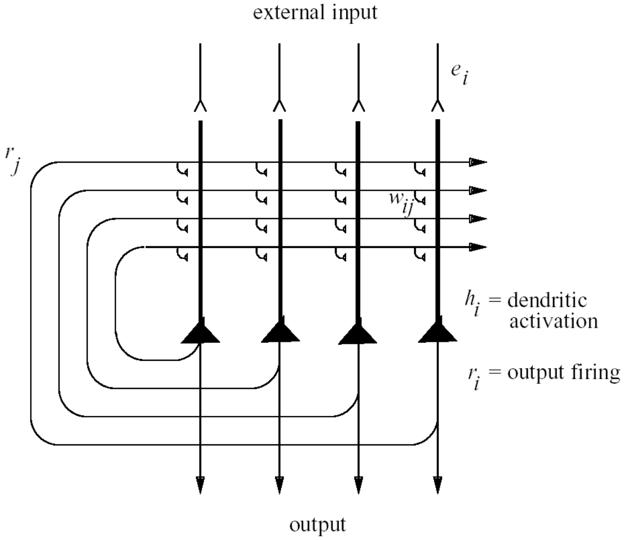

Figure 16.5

The architecture of an attractor neural network.

16.2.4.2 Learning the synaptic strengths between the neurons that implement a continuous attractor network

So far we have said that the neurons in the continuous attractor network are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g., head directions, spatial views etc.) represented by the neurons. In many simulations, the weights are set by formula to have weights with these appropriate Gaussian values. However, [101] showed how the appropriate weights could be set up by learning. They started with the fact that since the neurons have broad tuning that may be Gaussian in shape, nearby neurons in the state space will have overlapping spatial fields, and will thus be co-active to a degree that depends on the distance between them. They postulated that therefore the synaptic weights could be set up by associative learning based on the co-activity of the neurons produced by external stimuli as the animal moved in the state space. For example, head direction cells are forced to fire during learning by visual cues in the environment that produce Gaussian firing as a function of head direction from an optimal head direction for each cell. The learning rule is simply that the weights w_{ij} from head direction cell j with firing rate r_j^{HD} to head direction cell i with firing rate r_i^{HD} are updated according to an associative

be the input x_j to another neuron.

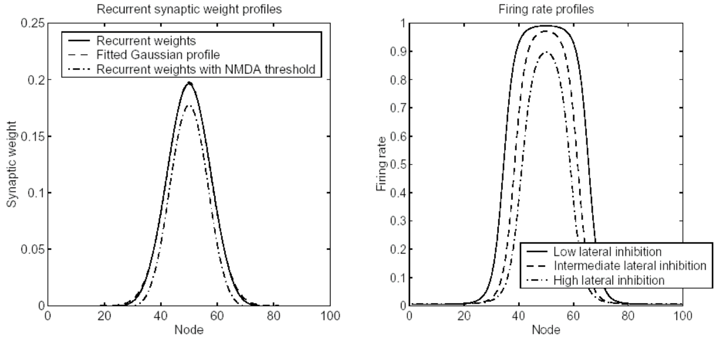

Figure 16.6

Training the weights in a continuous attractor network with an associative rule (equation 16.5). Left: The trained recurrent synaptic weights from head direction cell 50 to the other head direction cells in the network arranged in head direction space (solid curve). The dashed line shows a Gaussian curve fitted to the weights shown in the solid curve. The dash-dot curve shows the recurrent synaptic weights trained with rule equation (16.5), but with a non-linearity introduced that mimics the properties of NMDA receptors by allowing the synapses to modify only after strong postsynaptic firing is present. Right: The stable firing rate profiles forming an activity packet in the continuous attractor network during the testing phase when the training (visual) inputs are no longer present. The firing rates are shown after the network has been initially stimulated by visual input to initialize an activity packet, and then allowed to settle to a stable activity profile without visual input. The three graphs show the firing rates for low, intermediate and high values of the lateral inhibition parameter w^{inh}. For both left and right plots, the 100 head direction cells are arranged according to where they fire maximally in the head direction space of the agent when visual cues are available. After Stringer, Trappenberg, Rolls and de Araujo (2002).

(Hebb) rule

$$\delta w_{ij} = k r_i^{\mathrm{HD}} r_j^{\mathrm{HD}} \tag{16.5}$$

where δw_{ij} is the change of synaptic weight and k is the learning rate constant. During the learning phase, the firing rate r_i^{HD} of each head direction cell i might be the following Gaussian function of the displacement of the head from the optimal firing direction of the cell

$$r_i^{\mathrm{HD}} = e^{-s_{\mathrm{HD}}^2/2\sigma_{\mathrm{HD}}^2}, \tag{16.6}$$

where s_{HD} is the difference between the actual head direction x (in degrees) of the agent and the optimal head direction x_i for head direction cell i, and σ_{HD} is the standard deviation.

[101] showed that after training at all head directions, the synaptic connections develop strengths that are an almost Gaussian function of the distance between the cells in head direction space, as shown in Figure 16.6 (left). Interestingly if a non-linearity is introduced into the learning rule that mimics the properties of NMDA

receptors by allowing the synapses to modify only after strong postsynaptic firing is present, then the synaptic strengths are still close to a Gaussian function of the distance between the connected cells in head direction space (see Figure 16.6, left). They showed that after training, the continuous attractor network can support stable activity packets in the absence of visual inputs (see Figure 16.6, right) provided that global inhibition is used to prevent all the neurons becoming activated. (The exact stability conditions for such networks have been analyzed by [3]). Thus [101] demonstrated biologically plausible mechanisms for training the synaptic weights in a continuous attractor using a biologically plausible local learning rule.

So far, we have considered how spatial representations could be stored in continuous attractor networks, and how the activity can be maintained at any location in the state space in a form of short-term memory when the external (e.g., visual) input is removed. However, many networks with spatial representations in the brain can be updated by internal, self-motion (i.e., idiothetic), cues even when there is no external (e.g., visual) input. Examples are head direction cells in the presubiculum of rats and macaques, place cells in the rat hippocampus, and spatial view cells in the primate hippocampus (see Section 16.2). The major question arises about how such idiothetic inputs could drive the activity packet in a continuous attractor network, and in particular, how such a system could be set up biologically by self-organizing learning.

One approach to simulating the movement of an activity packet produced by idiothetic cues (which is a form of path integration whereby the current location is calculated from recent movements) is to employ a look-up table that stores (taking head direction cells as an example), for every possible head direction and head rotational velocity input generated by the vestibular system, the corresponding new head direction [95]. Another approach involves modulating the strengths of the recurrent synaptic weights in the continuous attractor on one but not the other side of a currently represented position, so that the stable position of the packet of activity, which requires symmetric connections in different directions from each node, is lost, and the packet moves in the direction of the temporarily increased weights, although no possible biological implementation was proposed of how the appropriate dynamic synaptic weight changes might be achieved [119]. Another mechanism (for head direction cells) [97] relies on a set of cells, termed (head) rotation cells, which are co-activated by head direction cells and vestibular cells and drive the activity of the attractor network by anatomically distinct connections for clockwise and counter-clockwise rotation cells, in what is effectively a look-up table. However, no proposal was made about how this could be achieved by a biologically plausible learning process, and this has been the case until recently for most approaches to path integration in continuous attractor networks, which rely heavily on rather artificial pre-set synaptic connectivities.

[101] introduced a proposal with more biological plausibility about how the synaptic connections from idiothetic inputs to a continuous attractor network can be learned by a self-organizing learning process. The essence of the hypothesis is described with Figure 16.7. The continuous attractor synaptic weights w^{RC} are set up under the influence of the external visual inputs I^{V} as described in Section 16.2.4.2. At

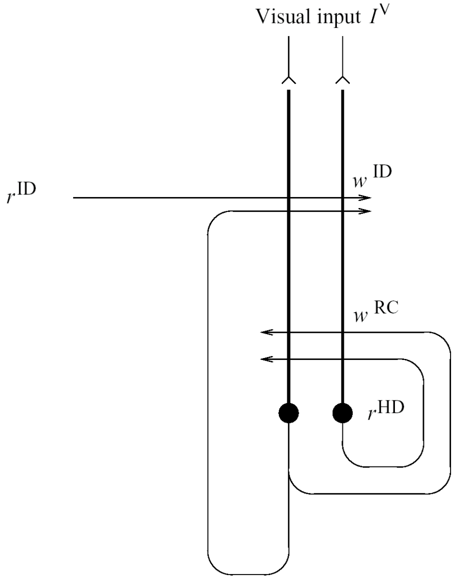

Figure 16.7

General network architecture for a one-dimensional continuous attractor model of head direction cells which can be updated by idiothetic inputs produced by head rotation cell firing r^{ID}. The head direction cell firing is r^{HD}, the continuous attractor synaptic weights are w^{RC}, the idiothetic synaptic weights are w^{ID}, and the external visual input is I^{V}.

the same time, the idiothetic synaptic weights w^{ID} (in which the ID refers to the fact that they are in this case produced by idiothetic inputs, produced by cells that fire to represent the velocity of clockwise and anticlockwise head rotation), are set up by associating the change of head direction cell firing that has just occurred (detected by a trace memory mechanism described below) with the current firing of the head rotation cells r^{ID}. For example, when the trace memory mechanism incorporated into the idiothetic synapses w^{ID} detects that the head direction cell firing is at a given location (indicated by the firing r^{HD}) and is moving clockwise (produced by the altering visual inputs I^{V}), and there is simultaneous clockwise head rotation cell firing, the synapses w^{ID} learn the association, so that when that rotation cell firing occurs later without visual input, it takes the current head direction firing in the continuous attractor into account, and moves the location of the head direction attractor in the appropriate direction.

For the learning to operate, the idiothetic synapses onto head direction cell i with firing r_i^{HD} need two inputs: the memory traced term from other head direction cells $\bar{r}_j^{\mathrm{HD}}$ (given by

$$\bar{r}^{\mathrm{HD}}(t+\delta t) = (1-\eta)r^{\mathrm{HD}}(t+\delta t) + \eta\bar{r}^{\mathrm{HD}}(t) \qquad (16.7)$$

where η is a parameter set in the interval [0,1] which determines the contribution of the current firing and the previous trace), and the head rotation cell input with firing r_k^{ID}; and the learning rule can be written

$$\delta w_{ijk}^{\text{ID}} = \tilde{k} r_i^{\text{HD}} \bar{r}_j^{\text{HD}} r_k^{\text{ID}}, \tag{16.8}$$

where $\tilde{k}$ is the learning rate associated with this type of synaptic connection. The head rotation cell firing (r_k^{ID}) could be as simple as one set of cells that fire for clockwise head rotation (for which k might be 1), and a second set of cells that fire for anticlockwise head rotation (for which k might be 2).

After learning, the firing of the head direction cells would be updated in the dark (when $I_i^V = 0$) by idiothetic head rotation cell firing r_k^{ID} as follows

$$\begin{aligned}\tau \frac{dh_i^{\text{HD}}(t)}{dt} &= -h_i^{\text{HD}}(t) + \frac{\phi_0}{C^{\text{HD}}} \sum_j (w_{ij} - w^{inh}) r_j^{\text{HD}}(t) + I_i^V \\ &+ \phi_1 \Big(\frac{1}{C^{\text{HD}\times\text{ID}}} \sum_{j,k} w_{ijk}^{\text{ID}} r_j^{\text{HD}} r_k^{\text{ID}} \Big).\end{aligned} \tag{16.9}$$

Equation 16.9 is similar to equation 16.3, except for the last term, which introduces the effects of the idiothetic synaptic weights w_{ijk}^{ID}, which effectively specify that the current firing of head direction cell i, r_i^{HD}, must be updated by the previously learned combination of the particular head rotation now occurring indicated by r_k^{ID}, and the current head direction indicated by the firings of the other head direction cells r_j^{HD} indexed through j.[†] This makes it clear that the idiothetic synapses operate using combinations of inputs, in this case of two inputs. Neurons that sum the effects of such local products are termed Sigma-Pi neurons. Although such synapses are more complicated than the two-term synapses used throughout the rest of this book, such three-term synapses appear to be useful to solve the computational problem of updating representations based on idiothetic inputs in the way described. Synapses that operate according to Sigma-Pi rules might be implemented in the brain by a number of mechanisms described by [38] (Section 21.1.1), [36], and [101], including having two inputs close together on a thin dendrite, so that local synaptic interactions would be emphasized.

Simulations demonstrating the operation of this self-organizing learning to produce movement of the location being represented in a continuous attractor network were described by [101], and one example of the operation is shown in Figure 16.8. They also showed that, after training with just one value of the head rotation cell firing, the network showed the desirable property of moving the head direction being represented in the continuous attractor by an amount that was proportional to the value of the head rotation cell firing. [101] also describe a related model of the idiothetic cell update of the location represented in a continuous attractor, in which the

[†] The term $\phi_1/C^{\text{HD}\times\text{ID}}$ is a scaling factor that reflects the number $C^{\text{HD}\times\text{ID}}$ of inputs to these synapses, and enables the overall magnitude of the idiothetic input to each head direction cell to remain approximately the same as the number of idiothetic connections received by each head direction cell is varied.

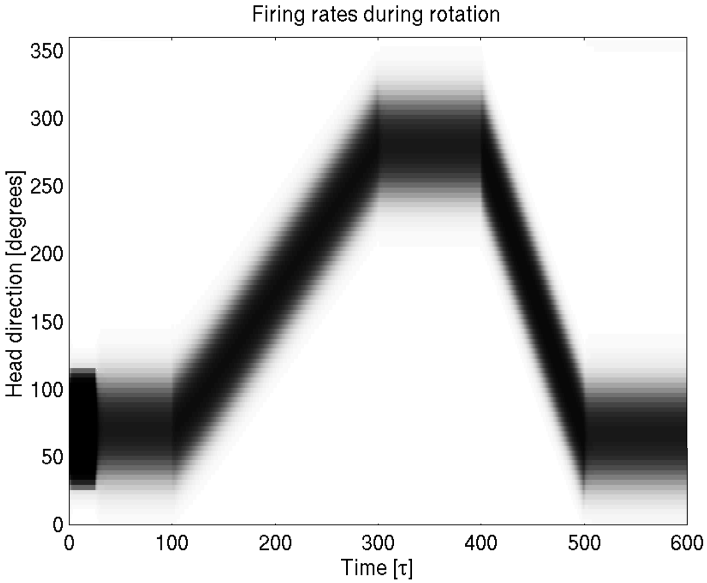

Figure 16.8

Idiothetic update of the location represented in a continuous attractor network. The firing rate of the cells with optima at different head directions (organized according to head direction on the ordinate) is shown by the blackness of the plot, as a function of time. The activity packet was initialized to a head direction of 75 degrees, and the packet was allowed to settle without visual input. For $t = 0$ to $t = 100$ there was no rotation cell input, and the activity packet in the continuous attractor remained stable at 75 degrees. For $t = 100$ to $t = 300$ the clockwise rotation cells were active with a firing rate of 0.15 to represent a moderate angular velocity, and the activity packet moved clockwise. For $t = 300$ to $t = 400$ there was no rotation cell firing, and the activity packet immediately stopped, and remained still. For $t = 400$ to $t = 500$ the anti-clockwise rotation cells had a high firing rate of 0.3 to represent a high velocity, and the activity packet moved anti-clockwise with a greater velocity. For $t = 500$ to $t = 600$ there was no rotation cell firing, and the activity packet immediately stopped.

rotation cell firing directly modulates in a multiplicative way the strength of the recurrent connections in the continuous attractor in such a way that clockwise rotation cells modulate the strength of the synaptic connections in the clockwise direction in the continuous attractor, and vice versa. It should be emphasized that although the cells are organized in Figure 16.8 according to the spatial position being represented, there is no need for cells in continuous attractors that represent nearby locations in the state space to be close together, as the distance in the state space between any two neurons is represented by the strength of the connection between them, not by where the neurons are physically located. This enables continuous attractor networks to represent spaces with arbitrary topologies, as the topology is represented in the connection strengths [101, 102, 104, 105]. Indeed, it is this that enables many different charts each with its own topology to be represented in a single continuous attractor network [8].

16.2.4.3 Continuous attractor networks in two or more dimensions

Some types of spatial representation used by the brain are of spaces that exist in two or more dimensions. Examples are the two- (or three-) dimensional space representing where one is looking at in a spatial scene. Another is the two- (or three-) dimensional space representing where one is located. It is possible to extend continuous attractor networks to operate in higher dimensional spaces than the one-dimensional spaces considered so far [111, 104]. Indeed, it is also possible to extend the analyses of how idiothetic inputs could be used to update two-dimensional state spaces, such as the locations represented by place cells in rats [104] and the location at which one is looking represented by primate spatial view cells [102, 105]. Interestingly, the number of terms in the synapses implementing idiothetic update do not need to increase beyond three (as in Sigma-Pi synapses) even when higher dimensional state spaces are being considered [104]. Also interestingly, a continuous attractor network can in fact represent the properties of very high dimensional spaces, because the properties of the spaces are captured by the connections between the neurons of the continuous attractor, and these connections are of course, as in the world of discrete attractor networks, capable of representing high dimensional spaces [104]. With these approaches, continuous attractor networks have been developed of the two-dimensional representation of rat hippocampal place cells with idiothetic update by movements in the environment [104], and of primate hippocampal spatial view cells with idiothetic update by eye and head movements [102, 105].

16.2.5 A unified theory of hippocampal memory: mixed continuous and discrete attractor networks

If the hippocampus is to store and retrieve episodic memories, it may need to associate together patterns which have continuous spatial attributes, and other patterns which represent objects, which are discrete. To address this issue, we have now shown that attractor networks can store both continuous patterns and discrete patterns, and can thus be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present (see Figure 16.4 and [88]). In this network, when events are stored that have both discrete (object) and continuous (spatial) aspects, then the whole place can be retrieved later by the object, and the object can be retrieved by using the place as a retrieval cue. Such networks are likely to be present in parts of the brain that receive and combine inputs both from systems that contain representations of continuous (physical) space, and from brain systems that contain representations of discrete objects, such as the inferior temporal visual cortex. One such brain system is the hippocampus, which appears to combine and store such representations in a mixed attractor network in the CA3 region, which thus is able to implement episodic memories which typically have a spatial component, for example where an item such as a key is located.

This network thus shows that in brain regions where the spatial and object processing streams are brought together, then a single network can represent and learn associations between both types of input. Indeed, in brain regions such as the hippocampal system, it is essential that the spatial and object processing streams are brought together in a single network, for it is only when both types of information are in the same network that spatial information can be retrieved from object information, and vice versa, which is a fundamental property of episodic memory. It may also be the case that in the prefrontal cortex, attractor networks can store both spatial and discrete (e.g., object-based) types of information in short-term memory (see below).

16.2.6 The speed of operation of memory networks: the integrate-and-fire approach

Consider for example a real network whose operation has been described by an autoassociative formal model that acquires, with learning, a given attractor structure. How does the state of the network approach, in real time during a retrieval operation, one of those attractors? How long does it take? How does the amount of information that can be read off the network's activity evolve with time? Also, which of the potential steady states is indeed a stable state that can be reached asymptotically by the net? How is the stability of different states modulated by external agents? These are examples of dynamical properties, which to be studied require the use of models endowed with some dynamics. An appropriate such model is one which incorporates integrate-and-fire neurons.

The concept that attractor (autoassociation) networks can operate very rapidly if implemented with neurons that operate dynamically in continuous time is described by [82] and [92]. The result described was that the principal factor affecting the speed of retrieval is the time constant of the synapses between the neurons that form the attractor ([7, 59, 92, 112]). This was shown analytically by [112], and described by [92] Appendix 5. If the (inactivation) time constant of AMPA synapses is taken as 10 ms, then the settling time for a single attractor network is approximately 15–17 ms [7, 59, 92]. A connected series of four such networks (representing for example four connected cortical areas) each involving recurrent (feedback) processing implemented by the recurrent collateral synaptic connections, takes approximately 4 x 17 ms to propagate from start to finish, retrieving information from each layer as the propagation proceeds [59, 82]. This speed of operation is sufficiently rapid that such attractor networks are biologically plausible [82, 92].

The way in which networks with continuous dynamics (such as networks made of real neurons in the brain, and networks modelled with integrate-and-fire neurons) can be conceptualized as settling so fast into their attractor states is that spontaneous activity in the network ensures that some neurons are close to their firing threshold when the retrieval cue is presented, so that the firing of these neurons is influenced within 1–2 ms by the retrieval cue. These neurons then influence other neurons within milliseconds (given the point that some other neurons will be close to threshold) through the modified recurrent collateral synapses that store the information. In this way, the neurons in networks with continuous dynamics can influence each other within a fraction of the synaptic time constant, and retrieval can be very rapid [82, 92].

16.3 Short-term memory systems

16.3.1 Prefrontal cortex short-term memory networks, and their relation to temporal and parietal perceptual networks

A common way that the brain uses to implement a short-term memory is to maintain the firing of neurons during a short memory period after the end of a stimulus (see [24] and [92]). In the inferior temporal cortex this firing may be maintained for a few hundred ms even when the monkey is not performing a memory task [18, 89, 90, 91]. In more ventral temporal cortical areas such as the entorhinal cortex the firing may be maintained for longer periods in delayed match to sample tasks [108], and in the prefrontal cortex for even tens of seconds [23, 24]. In the dorsolateral and inferior convexity prefrontal cortex the firing of the neurons may be related to the memory of spatial responses or objects [30, 118] or both [63], and in the principal sulcus / arcuate sulcus region to the memory of places for eye movements [22] (see [82]). The firing may be maintained by the operation of associatively modified recurrent collateral connections between nearby pyramidal cells producing attractor states in autoassociative networks (see [82]).

For the short-term memory to be maintained during periods in which new stimuli are to be perceived, there must be separate networks for the perceptual and short-term memory functions, and indeed two coupled networks, one in the inferior temporal visual cortex for perceptual functions, and another in the prefrontal cortex for maintaining the short-term memory during intervening stimuli, provide a precise model of the interaction of perceptual and short-term memory systems [67, 70] (see Figure 16.9). In particular, this model shows how a prefrontal cortex attractor (autoassociation) network could be triggered by a sample visual stimulus represented in the inferior temporal visual cortex in a delayed match to sample task, and could keep this attractor active during a memory interval in which intervening stimuli are shown. Then when the sample stimulus reappears in the task as a match stimulus, the inferior temporal cortex module showed a large response to the match stimulus, because it is activated both by the visual incoming match stimulus, and by the consistent backprojected memory of the sample stimulus still being represented in the prefrontal cortex memory module (see Figure 16.9). This computational model makes it clear that in order for ongoing perception to occur unhindered implemented by posterior cortex (parietal and temporal lobe) networks, there must be a separate set of modules that is capable of maintaining a representation over intervening stim-

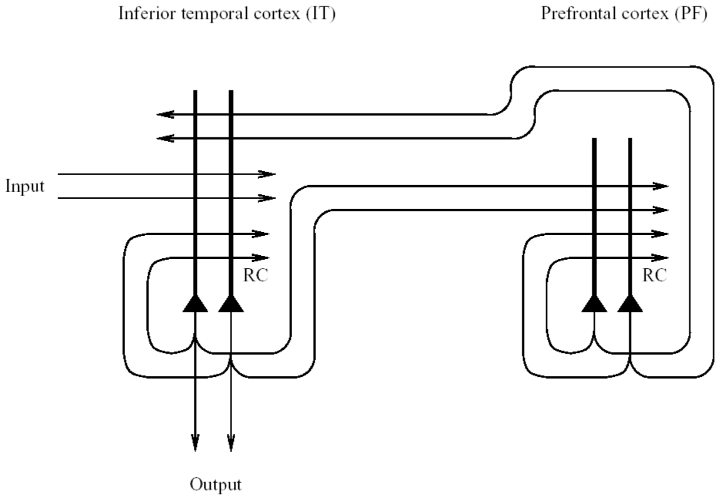

Figure 16.9

A short-term memory autoassociation network in the prefrontal cortex could hold active a working memory representation by maintaining its firing in an attractor state. The prefrontal module would be loaded with the to-be-remembered stimulus by the posterior module (in the temporal or parietal cortex) in which the incoming stimuli are represented. Backprojections from the prefrontal short-term memory module to the posterior module would enable the working memory to be unloaded, to for example influence on-going perception (see text). RC - recurrent collateral connections.

uli. This is the fundamental understanding offered for the evolution and functions of the dorsolateral prefrontal cortex, and it is this ability to provide multiple separate short-term attractor memories that provides we suggest the basis for its functions in planning. [67] and [70] performed analyses and simulations which showed that for working memory to be implemented in this way, the connections between the perceptual and the short-term memory modules (see Figure 16.9) must be relatively weak. As a starting point, they used the neurophysiological data showing that in delayed match to sample tasks with intervening stimuli, the neuronal activity in the inferior temporal visual cortex (IT) is driven by each new incoming visual stimulus [64, 66], whereas in the prefrontal cortex, neurons start to fire when the sample stimulus is shown, and continue the firing that represents the sample stimulus even when the potential match stimuli are being shown [65]. The architecture studied by [70] was as shown in Figure 16.9, with both the intramodular (recurrent collateral) and the intermodular (forward IT to PF, and backward PF to IT) connections trained on the set of patterns with an associative synaptic modification rule. A crucial parameter is the strength of the intermodular connections, *g*, which indicates the relative strength of the intermodular to the intramodular connections. This parameter measures effectively the relative strengths of the currents injected into the neurons by the intermodular relative to the intra-modular connections, and the importance of setting this parameter to relatively weak values for useful interactions between coupled attractor networks was highlighted by [68] and [69] (see [82]). The patterns themselves were sets of random numbers, and the simulation utilized a dynamical approach with neurons with continuous (hyperbolic tangent) activation functions (see Section 16.3.2 and [5, 40, 41, 96]). The external current injected into IT by the incoming visual stimuli was sufficiently strong to trigger the IT module into a state representing the incoming stimulus. When the sample was shown, the initially silent PF module was triggered into activity by the weak ($g > 0.002$) intermodular connections. The PF module remained firing to the sample stimulus even when IT was responding to potential match stimuli later in the trial, provided that g was less than 0.024, because then the intramodular recurrent connections could dominate the firing (see Figure 16.10). If g was higher than this, then the PF module was pushed out of the attractor state produced by the sample stimulus. The IT module responded to each incoming potentially matching stimulus provided that g was not greater than approximately 0.024. Moreover, this value of g was sufficiently large that a larger response of the IT module was found when the stimulus matched the sample stimulus (the match enhancement effect found neurophysiologically, and a mechanism by which the matching stimulus can be identified). This simple model thus shows that the operation of the prefrontal cortex in short-term memory tasks such as delayed match to sample with intervening stimuli, and its relation to posterior perceptual networks, can be understood by the interaction of two weakly coupled attractor networks, as shown in Figures 16.9 and 16.10.

The same network can also be used to illustrate the interaction between the prefrontal cortex short-term memory system and the posterior (IT or PP) perceptual regions in visual search tasks, as illustrated in Figure 16.11.

16.3.2 Computational details of the model of short-term memory

The model network of [67] and [70] consists of a large number of (excitatory) neurons arranged in two modules with the architecture shown in Figure 16.9. Following [5, 40], each neuron is assumed to be a dynamical element which transforms an incoming afferent current into an output spike rate according to a given transduction function. A given afferent current I_{ai} to neuron i $(i = 1, \ldots, N)$ in module a $(a = \mathbf{IT}, \mathbf{PF})$ decays with a characteristic time constant τ but increases proportionally to the spike rates of the rest of the neurons in the network (both from inside and outside its module) connected to it, the contribution of each presynaptic neuron, e.g., neuron j from module b, and in proportion to the synaptic efficacy J_{ij}^{ab} between the two.[‡] This can be expressed through the following equation

$$\frac{dI_{ai}(t)}{dt} = -\frac{I_{ai}(t)}{\tau} + \sum_{b,j} J_{ij}^{(a,b)} \nu_{bj} + h_{ai}^{(\text{ext})} \quad . \tag{16.10}$$

[‡] On this occasion we revert to the theoretical physicists' usual notation for synaptic weights or couplings, J_{ij}, from w_{ij}.

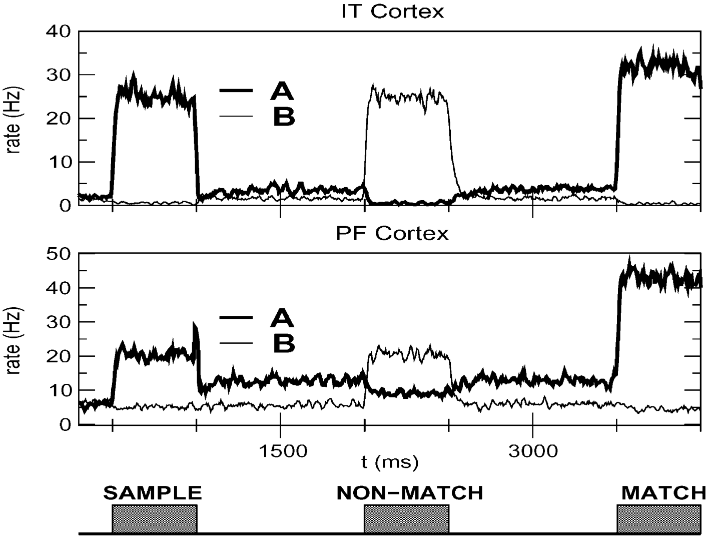

Figure 16.10

Interaction between the prefrontal cortex (PF) and the inferior temporal cortex (IT) in a delayed match to sample task with intervening stimuli with the architecture illustrated in Figure 16.9. Above: activity in the IT attractor module. Below: activity in the PF attractor module. The thick lines show the firing rates of the set of neurons with activity selective for the Sample stimulus (which is also shown as the Match stimulus, and is labelled **A**), and the thin lines the activity of the neurons with activity selective for the Non-Match stimulus, which is shown as an intervening stimulus between the Sample and Match stimulus and is labelled **B**. A trial is illustrated in which **A** is the Sample (and Match) stimulus. The prefrontal cortex module is pushed into an attractor state for the sample stimulus by the IT activity induced by the sample stimulus. Because of the weak coupling to the PF module from the IT module, the PF module remains in this Sample-related attractor state during the delay periods, and even while the IT module is responding to the non-match stimulus. The PF module remains in its Sample-related state even during the Non-Match stimulus because once a module is in an attractor state, it is relatively stable. When the Sample stimulus reappears as the Match stimulus, the PF module shows higher Sample stimulus-related firing, because the incoming input from IT is now adding to the activity in the PF attractor network. This in turn also produces a match enhancement effect in the IT neurons with Sample stimulus-related selectivity, because the backprojected activity from the PF module matches the incoming activity to the IT module. After Renart, Parga and Rolls, 2000 and Renart, Moreno, de la Rocha, Parga and Rolls, 2001.

An external current $h_{ai}^{(\mathrm{ext})}$ from outside the network, representing the stimuli, can also be imposed on every neuron. Selective stimuli are modelled as proportional to the stored patterns, i.e., $h_{ai}^{\mu(\mathrm{ext})} = h_a \eta_{ai}^{\mu}$, where h_a is the intensity of the external current to module a.

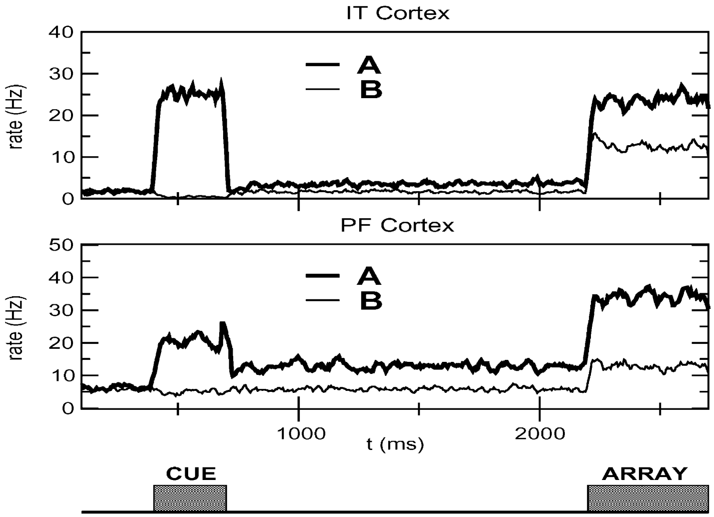

Figure 16.11

Interaction between the prefrontal cortex (PF) and the inferior temporal cortex (IT) in a visual search task with the architecture illustrated in Figure 16.9. Above: activity in the IT attractor module. Below: activity in the PF attractor module. The thick lines show the firing rates of the set of neurons with activity selective for search stimulus **A**, and the thin lines the activity of the neurons with activity selective for stimulus **B**. During the cue period either **A** or **B** is shown, to indicate to the monkey which stimulus to select when an array containing both **A** and **B** is shown after a delay period. The trial shown is for the case when **A** is the cue stimulus. When stimulus **A** is shown as a cue, then via the IT module, the PF module is pushed into an attractor state **A**, and the PF module remembers this state during the delay period. When the array **A** + **B** is shown later, there is more activity in the PF module for the neurons selective for **A**, because they have inputs both from the continuing attractor state held in the PF module and from the forward activity from the IT module which now contains both **A** and **B**. This PF firing to **A** in turn also produces greater firing of the population of IT neurons selective for **A** than in the IT neurons selective for **B**, because the IT neurons selective for **A** are receiving both **A**–related visual inputs, and **A**–related backprojected inputs from the PF module. After Renart, Parga and Rolls, 2000 and Renart, Moreno, de la Rocha, Parga and Rolls, 2001.

The transduction function of the neurons transforming currents into rates was chosen as a threshold hyperbolic tangent of gain G and threshold θ. Thus, when the current is very large the firing rates saturate to an arbitrary value of 1.

The synaptic efficacies between the neurons of each module and between the neurons in different modules are respectively

$$J_{ij}^{(a,a)} = \frac{J_0}{f(1-f)N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\,(\eta_{aj}^{\mu} - f) \qquad i \neq j \;;\; a = \mathbf{IT}, \mathbf{PF} \tag{16.11}$$

$$J_{ij}^{(a,b)} = \frac{g}{f(1-f)N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\,(\eta_{bj}^{\mu} - f) \qquad \forall\; i,j \;;\; a \neq b \;. \tag{16.12}$$

The intra-modular connections are such that a number P of sparse independent configurations of neural activity are dynamically stable, constituting the possible sustained activity states in each module. This is expressed by saying that each module has learned P binary patterns $\{\eta_{ai}^{\mu} = 0, 1,\ \mu = 1, \ldots, P\}$, each of them signalling which neurons are active in each of the sustained activity configurations. Each variable η_{ai}^{μ} is allowed to take the values 1 and 0 with probabilities f and $(1-f)$ respectively, independently across neurons and across patterns. The inter-modular connections reflect the temporal associations between the sustained activity states of each module. In this way, every stored pattern μ in the IT module has an associated pattern in the PF module which is labelled by the same index. The normalization constant $N_t = N(J_0 + g)$ was chosen so that the sum of the magnitudes of the inter- and the intra-modular connections remains constant and equal to 1 while their relative values are varied. When this constraint is imposed the strength of the connections can be expressed in terms of a single independent parameter g measuring the relative intensity of the inter- vs. the intra-modular connections (J_0 can be set equal to 1 everywhere).

Both modules implicitly include an inhibitory population of neurons receiving and sending signals to the excitatory neurons through uniform synapses. In this case the inhibitory population can be treated as a single inhibitory neuron with an activity dependent only on the mean activity of the excitatory population. We chose the transduction function of the inhibitory neuron to be linear with slope γ.

Since the number of neurons in a typical network one may be interested in is very large, e.g., $\sim 10^5 - 10^6$, the analytical treatment of the set of coupled differential equations (16.10) becomes untractable. On the other hand, when the number of neurons is large, a reliable description of the asymptotic solutions of these equations can be found using the techniques of statistical mechanics [40]. In this framework, instead of characterizing the states of the system by the state of every neuron, this characterization is performed in terms of *macroscopic* quantities called *order parameters* which measure and quantify some global properties of the network as a whole. The relevant order parameters appearing in the description of the system are the overlap of the state of each module with each of the stored patterns m_a^{μ} and the average activity of each module x_a, defined respectively as:

$$m_a^{\mu} = \frac{1}{\chi N} \ll \sum_i (\eta_{ai}^{\mu} - f)\nu_{ai} \gg_{\eta} \;;\; x_a = \frac{1}{N} \ll \sum_i \nu_{ai} \gg_{\eta} \quad , \tag{16.13}$$

where the symbol $\ll \ldots \gg_{\eta}$ stands for an average over the stored patterns.

Using the free energy per neuron of the system at zero temperature $\mathscr{F}$ (which is not written explicitly to reduce the technicalities to a minimum), [70] and [67] modelled the experiments by giving the order parameters the following dynamics:

$$\tau \frac{\partial m_a^\mu}{\partial t} = -\frac{\partial \mathscr{F}}{\partial m_a^\mu} \quad ; \quad \tau \frac{\partial x_a}{\partial t} = -\frac{\partial \mathscr{F}}{\partial x_a} \quad . \tag{16.14}$$

These dynamics ensure that the stationary solutions, corresponding to the values of the order parameters at the attractors, correspond also to minima of the free energy, and that, as the system evolves, the free energy is always minimized through its gradient. The time constant of the macroscopical dynamics was chosen to be equal to the time constant of the individual neurons, which reflects the assumption that neurons operate in parallel. Equations (16.14) were solved by a simple discretizing procedure (first order Runge-Kutta method). An appropriate value for the time interval corresponding to one computer iteration was found to be $\tau/10$ and the time constant has been given the value $\tau = 10\ ms$.

Since not all neurons in the network receive the same inputs, not all of them behave in the same way, i.e., have the same firing rates. In fact, the neurons in each of the modules can be split into different sub-populations according to their state of activity in each of the stored patterns. The mean firing rate of the neurons in each sub-population depends on the particular state realized by the network (characterized by the values of the order parameters). Associated with each pattern there are two large sub-populations denoted as foreground (all active neurons) and background (all inactive neurons) for that pattern. The overlap with a given pattern can be expressed as the difference between the mean firing rate of the neurons in its foreground and its background. The average was calculated over all other sub-populations to which each neuron in the foreground (background) belonged to, where the probability of a given sub-population is equal to the fraction of neurons in the module belonging to it (determined by the probability distribution of the stored patterns as given above). This partition of the neurons into sub-populations is appealing since, in neurophysiological experiments, cells are usually classified in terms of their response properties to a set of fixed stimuli, i.e., whether each stimulus is effective or ineffective in driving their response.

The modelling of the different experiments proceeded according to the macroscopic dynamics (16.14), where each stimulus was implemented as an extra current into free energy for a desired period of time.

Using this model, results of the type described in Section 16.3.1 were found [67, 70]. The paper by [67] extended the earlier findings of [70] to integrate-and-fire neurons, and it is results from the integrate-and-fire simulations that are shown in Figures 16.10 and 16.11.

16.3.3 Computational necessity for a separate, prefrontal cortex, short-term memory system

This approach emphasizes that in order to provide a good brain lesion test of prefrontal cortex short-term memory functions, the task set should require a short-term memory for stimuli over an interval in which other stimuli are being processed, because otherwise the posterior cortex perceptual modules could implement the short-term memory function by their own recurrent collateral connections. This approach also emphasizes that there are many at least partially independent modules for short-term memory functions in the prefrontal cortex (e.g., several modules for delayed saccades; one or more for delayed spatial (body) responses in the dorsolateral prefrontal cortex; one or more for remembering visual stimuli in the more ventral prefrontal cortex; and at least one in the left prefrontal cortex used for remembering the words produced in a verbal fluency task – see Section 10.3 of [92]).

This computational approach thus provides a clear understanding of why a separate (prefrontal) mechanism is needed for working memory functions, as elaborated in Section 16.3.1. It may also be commented that if a prefrontal cortex module is to control behaviour in a working memory task, then it must be capable of assuming some type of executive control. There may be no need to have a single central executive additional to the control that must be capable of being exerted by every short-term memory module. This is in contrast to what has traditionally been assumed for the prefrontal cortex [98].

16.3.4 Role of prefrontal cortex short-term memory systems in visual search and attention

The same model shown in Figure 16.9 can also be used to help understand the implementation of visual searchtasks in the brain [70]. In such a visual search task, the target stimulus is made known beforehand, and inferior temporal cortex neurons then respond more when the search target (as compared to a different stimulus) appears in the receptive field of the IT neuron [15, 16]. The model shows that this could be implemented by the same system of weakly coupled attractor networks in PF and IT shown in Figure 16.9 as follows. When the target stimulus is shown, it is loaded into the PF module from the IT module as described for the delayed match to sample task. Later, when the display appears with two or more stimuli present, there is an enhanced response to the target stimulus in the receptive field, because of the backprojected activity from PF to IT which adds to the firing being produced by the target stimulus itself [67, 70] (see Figure 16.11). The interacting spatial and object networks described by [82]) in Chapters 9–11, take this analysis one stage further, and show that once the PF–IT interaction has set up a greater response to the search target in IT, this enhanced response can in turn by backprojections to topologically mapped earlier cortical visual areas move the “attentional spotlight” to the place where the search target is located.

16.3.5 Synaptic modification is needed to set up but not to reuse short-term memory systems

To set up a new short-term memory attractor, synaptic modification is needed to form the new stable attractor. Once the attractor is set up, it may be used repeatedly when triggered by an appropriate cue to hold the short-term memory state active by continued neuronal firing even without any further synaptic modification (see [37] and [82]). Thus manipulations that impair the long-term potentiation of synapses (LTP)air the formation of new short-term memory states, but not the use of previously learned short-term memory states. [37] analyzed many studies of the effects of blockade of LTP in the hippocampus on spatial working memory tasks, and found evidence consistent with this prediction. Interestingly, it was found that if there was a large change in the delay interval over which the spatial information had to be remembered, then the task became susceptible, during the transition to the new delay interval, to the effects of blockade of LTP. The implication is that some new learning is required when the rat must learn the strategy of retaining information for longer periods when the retention interval is changed.

16.4 Invariant visual object recognition

[74] proposed a feature hierarchical model of ventral stream visual objecting from the primary visual cortex (V1), via V2 and V4 to the inferior temporal visual cortex which could learn to represent objects invariantly with respect to position on the retina, scale, rotation and view. The theory uses a short-term (‘trace’) memory term in an associative learning rule to help capture the fact that the natural statistics of the visual world reflect the fact that the same object is likely to be present over short-time periods, for example over 1 or 2 seconds during which an object is seen from different views. A model of the operation of the system has been implemented in a four-layer network, corresponding to brain areas V1, V2, V4 and inferior temporal visual cortex (IT), with convergence to each part of a layer from a small region of the preceding layer, and with local competition between the neurons within a layer implemented by local lateral inhibition [20, 82, 83, 116] (see Figure 16.12). During a learning phase each object is learned. This is done by training the connections between modules using a trace learning rule with the general form

$$\delta w_{ij} = \alpha \bar{y}_i{}^{\tau} x_j^{\tau} \tag{16.15}$$

where x_j^{τ} is the jth input to the neuron at time step τ, y_i is the output of the ith neuron, and w_{ij} is the jth weight on the ith neuron.

The trace $\bar{y}_i{}^{\tau}$ is updated according to

$$\bar{y}_i{}^{\tau} = (1-\eta) y_i{}^{\tau} + \eta \bar{y}_i{}^{\tau-1}. \tag{16.16}$$

The parameter $\eta \in [0,1]$ controls the relative contributions to the trace $\bar{y}_i{}^{\tau}$ from the instantaneous firing rate $y_i{}^{\tau}$ at time step τ and the trace at the previous time step $\bar{y}_i{}^{\tau-1}$.

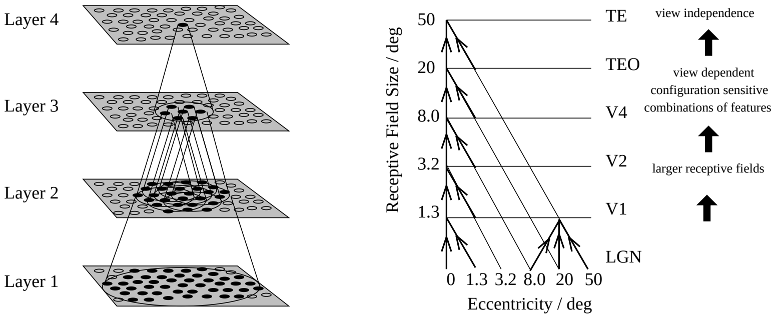

Figure 16.12

Convergence in the visual system. Right – as it occurs in the brain. V1: visual cortex area V1; TEO: posterior inferior temporal cortex; TE: inferior temporal cortex (IT). Left – as implemented in VisNet. Convergence through the network is designed to provide fourth layer neurons with information from across the entire input retina.

16.5 Visual stimulus–reward association, emotion, and motivation

Learning about which visual and other stimuli in the environment are rewarding punishing, or neutral is crucial for survival. For example, it takes just one trial to learn if a seen object is hot when we touch it, and associating that visual stimulus with the pain may help us to avoid serious injury in the future. Similarly, if we are given a new food which has an excellent taste, we can learn in one trial to associate the sight of it with its taste, so that we can select it in future. In these examples, the previously neutral visual stimuli become conditioned reinforcers by their association with a primary (unlearned) reinforcer such as taste or pain. Our examples show that learning about which stimuli are rewards and punishments is very important in the control of motivational behaviour such as feeding and drinking, and in emotional such as fear and pleasure. The type of learning involved is pattern association, between the conditioned and the unconditioned stimulus. This type of learning provides a major example of how the visual representations provided by the inferior temporal visual cortex are used by the other parts of the brain [77, 80, 82]. In this section we consider where in sensory processing this stimulus-reinforcement association learning occurs, which brain structures are involved in this type of learning, how the neuronal networks for pattern association learning may actually be implemented in these regions, and how the distributed representation about objects provided by the inferior temporal cortex output is suitable for this pattern association learning.

The crux of the answer to the last question is that the inferior temporal cortex representation is ideal for this pattern association learning because it is a transform-invariant representation of objects, and because the code can be read by a neuronal system which performs dot products using neuronal ensembles as inputs, which is precisely what pattern associators in the brain need, because they are implemented by neurons which perform as their generic computation a dot product of their inputs with their synaptic weight vectors (see [82] and [92]).

A schematic diagram summarizing some of the conclusions reached [77, 82, 92] is shown in Figure 16.13. The pathways are shown with more detail in Figure 16.14. The primate inferior temporal visual cortex provides a representation that is independent of reward or punishment, and is about objects. The utility of this is that the output of the inferior temporal visual cortex can be used for many memory and related functions (including episodic memory, short-term memory, and reward/punishment memory) independently of whether the visual stimulus is currently rewarding or not. Thus we can learn about objects, and place them in short-term memory, independently of whether they are currently wanted or not. This is a key feature of brain design. The inferior temporal cortex then projects into two structures, the amygdala and orbitofrontal cortex, that contain representations of primary (unlearned) reinforcers such as taste and pain. These two brain regions then learn associations between visual and other previously neutral stimuli, and primary reinforcers [77], using what is highly likely to be a pattern association network, as illustrated in Figure 16.13. A difference between the primate amygdala and orbitofrontal cortex may be that the orbitofrontal cortex is set up to perform reversal of these associations very rapidly, in as little as one trial. Because the amygdala and orbitofrontal cortex represent primary reinforcers, and learn associations between these and neutral stimuli, they are key brain regions in emotions which can be understood as states elicited by reinforcers, that is rewards and punishers), and in motivational states such as feeding and drinking [77].

16.6 Effects of mood on memory and visual processing

The current mood state can affect the cognitive evaluation of events or memories (see [9], [87]). An example is that when they are in a depressed mood, people tend to recall memories that were stored when they were depressed. The recall of depressing memories when depressed can have the effect of perpetuating the depression, and this may be a factor with relevance to the etiology and treatment of depression. A normal function of the effects of mood state on memory recall might be to facilitate continuity in the interpretation of the reinforcing value of events in the environment, or in the interpretation of an individual's behaviour by others, or simply to keep behaviour motivated to a particular goal. Another possibility is that the effects of mood on memory do not have adaptive value, but are a consequence of having a general cortical architecture with backprojections. According to the latter hypothesis, the selection pressure is great for leaving the general architecture operational, rather than

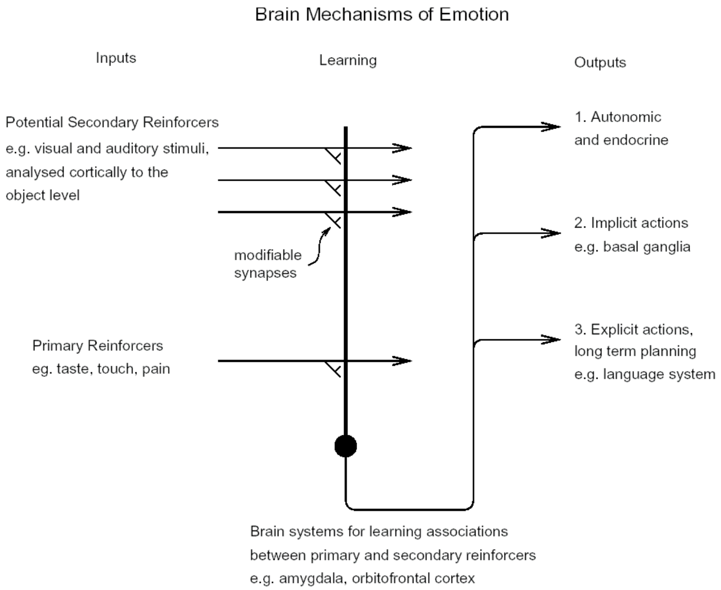
Brain Mechanisms of Emotion
Inputs
Learning
Outputs
Potential Secondary Reinforcers
e.g. visual and auditory stimuli,
analysed cortically to the
object level
modifiable
synapses
Primary Reinforcers
eg. taste, touch, pain
1. Autonomic
and endocrine
2. Implicit actions
e.g. basal ganglia
3. Explicit actions,
long term planning
e.g. language system
Brain systems for learning associations
between primary and secondary reinforcers
e.g. amygdala, orbitofrontal cortex

Figure 16.13

Schematic diagram showing the organization of brain networks involved in learning reinforcement associations of visual and auditory stimuli. The learning is implemented by pattern association networks in the amygdala and orbitofrontal cortex. The visual representation provided by the inferior temporal cortex is in an appropriate form for this pattern association learning, in that information about objects can be read from a population of IT neurons by dot-product neuronal operations.

trying to find a genetic way to switch off backprojections just for the projections of mood systems back to perceptual systems (cf. [86]).

[87] (see also [75] and [77]) have developed a theory of how the effects of mood on memory and perception could be implemented in the brain. The architecture, shown in Figure 16.15, uses the massive backprojections from parts of the brain where mood is represented, such as the orbitofrontal cortex and amygdala to the cortical areas such as the inferior temporal visual cortex and hippocampus-related areas (labelled IT in Figure 16.15) that project into these mood-representing areas [2, 1]. The model uses an attractor in the mood module (labelled amygdala in Figure 16.15), which helps the mood to be an enduring state, and also an attractor in IT. The system is treated as a system of coupled attractors (see [82]), but with an odd twist: many different perceptual states are associated with any one mood state. Overall, there is a large number of perceptual / memory states, and only a few mood states, so that there is a many-to-one relation between perceptual / memory states and the associated mood states. The network displays the properties that one would expect (provided that the coupling parameters g between the attractors are weak). These

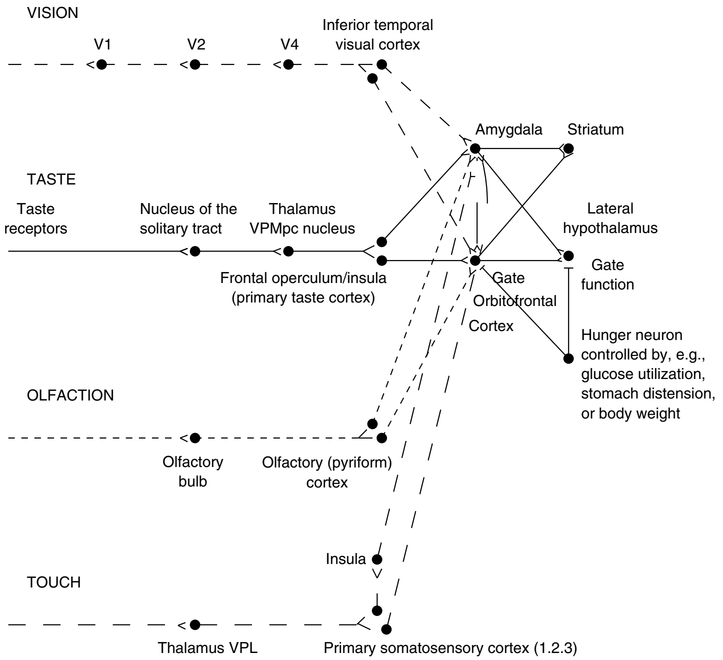
VISION
V1
V2
V4
Inferior temporal visual cortex
Amygdala
Striatum
TASTE
Taste receptors
Nucleus of the solitary tract
Thalamus VPMpc nucleus
Lateral hypothalamus
Frontal operculum/insula (primary taste cortex)
Gate
Orbitofrontal Cortex
Gate function
Hunger neuron controlled by, e.g., glucose utilization, stomach distension, or body weight
OLFACTION
Olfactory bulb
Olfactory (pyriform) cortex
Insula
TOUCH
Thalamus VPL
Primary somatosensory cortex (1.2.3)

Figure 16.14

Diagrammatic representation of some of the connections described in this chapter. V1, striate visual cortex. V2 and V4, cortical visual areas. In primates, sensory analysis proceeds in the visual system as far as the inferior temporal cortex and the primary gustatory cortex; beyond these areas, in for example the amygdala and orbitofrontal cortex, the hedonic value of the stimuli, and whether they are reinforcing or are associated with reinforcement, is represented (see text).

include the ability of a perceptual input to trigger a mood state in the ‘amygdala’ module if there is not an existing mood, but greater difficulty to induce a new mood if there is already a strong mood attractor present; and the ability of the mood to affect via the backprojections which memories are triggered.

An interesting property which was revealed by the model is that because of the many-to-few mapping of perceptual to mood states, an effect of a mood was that it tended to make all the perceptual or memory states associated with a particular mood more similar then they would otherwise have been. The implication is that the coupling parameter g for the backprojections must be quite weak, as otherwise interference increases in the perceptual / memory module (IT in Figure 16.15).

Acknowledgments: This research was supported by Medical Research Council Programme Grant PG9826105, by the MRC Interdisciplinary Research Centre for Cognitive Neuroscience, and by the Human Frontier Science Program.

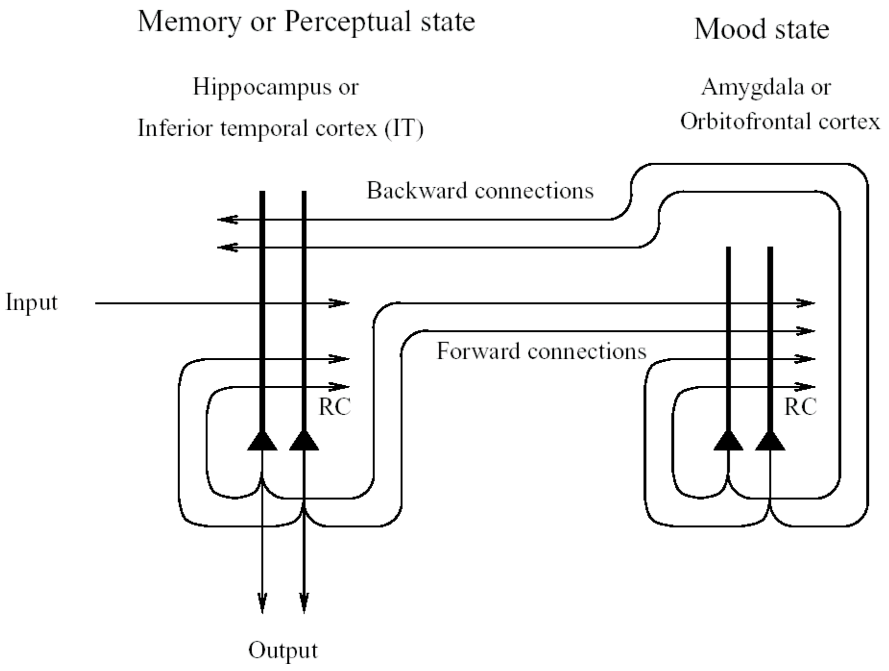

Figure 16.15

Architecture used to investigate how mood can affect perception and memory. The IT module represents brain areas such as the inferior temporal cortex involved in perception and hippocampus-related cortical areas that have forward connections to regions such as the amygdala and orbitofrontal cortex involved in mood. (After Rolls and Stringer (2001)).

References

[1] Amaral, D. G., and Price, J. L. (1984), Amygdalo-cortical projections in the monkey (*Macaca fascicularis), Journal of Comparative Neurology*, **230**, 465–496.

[2] Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992), Anatomical organization of the primate amygdaloid complex, in Aggleton, J. P. (ed.) *The Amygdala,* Wiley-Liss: New York, 1–66.

[3] Amari, S. (1977), Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics*, **27**, 77–87.

[4] Amit, D. J. (1989), *Modelling Brain Function*, Cambridge University Press: New York.

[5] Amit, D. J., and Tsodyks, M. V. (1991), Quantitative study of attractor neural network retrieving at low spike rates. I. Substrate – spikes, rates and neuronal gain, *Network*, **2**, 259–273.

[6] Andersen, R. A., Batista, A. P., Snyder, L. H., Buneo, C. A., and Cohen, Y. E. (2000), Programming to look and reach in the posterior parietal cortex, in Gazzaniga, M.S. (ed.) *The New Cognitive Neurosciences*, 2nd ed., MIT Press: Cambridge, MA, 515–524.

[7] Battaglia, F., and Treves, A. (1998), Stable and rapid recurrent processing in realistic autoassociative memories, *Neural Computation* **10**, 431–450.

[8] Battaglia, F. P., and Treves, A. (1998), Attractor neural networks storing multiple space representations: A model for hippocampal place fields, *Physical Review E*, **58**, 7738–7753.

[9] Blaney, P. H. (1986), Affect and memory: a review, *Psychological Bulletin*, **99**, 229–246.

[10] Burgess, N., and O'Keefe, J. (1996), Neuronal computations underlying the firing of place cells and their role in navigation, *Hippocampus*, **6**, 749–762.

[11] Burgess, N., Recce, M., and O'Keefe, J. (1994), A model of hippocampal function, *Neural Networks*, **7**, 1065–1081.

[12] Buckley, M. J., and Gaffan, D. (2000), The hippocampus, perirhinal cortex, and memory in the monkey, in Bolhuis, J. J. (ed.) *Brain, Perception, and Memory: Advances in Cognitive Neuroscience*, Oxford University Press: Oxford, 279–298.

[13] Cahusac, P. M. B., Rolls, E. T., Miyashita, Y., and Niki, H. (1993), Modification of the responses of hippocampal neurons in the monkey during the learning of a conditional spatial response task, *Hippocampus*, **3**, 29–42.

[14] Cassaday, H. J., and Rawlins, J. N. (1997), The hippocampus, objects, and their contexts, *Behavioural Neuroscience*, **111**, 1228–1244.

[15] Chelazzi, L., Duncan, J., Miller, E., and Desimone, R. (1998), Responses of neurons in inferior temporal cortex during memory-guuided visual search, *Journal of Neurophysiology*, **80**, 2918–2940.

[16] Chelazzi, L., Miller, E., Duncan, J., and Desimone, R. (1993), A neural basis for visual searchrior temporal cortex, *Nature* (London), **363**, 345–347.

[17] de Araujo, I. E. T., Rolls, E. T., and Stringer, S. M. (2001), A view model which accounts for the response properties of hippocampal primate spatial view cells and rat place cells, *Hippocampus*, **11**, 699–706.

[18] Desimone, R. (1996), Neural mechanisms for visual memory and their role in attention, *Proceedings of the National Academy of Sciences USA*, **93**, 13494–13499.

[19] Eichenbaum, H. (1997), Declarative memory: insights from cognitive neuro-

biology, *Annual Review of Psychology*, **48**, 547–572.

[20] Elliffe, M. C. M., Rolls, E. T., and Stringer, S. M. (2002), Invariant recognition of feature combinations in the visual system, *Biological Cybernetics*, **86**: 57-91.

[21] Epstein, R., and Kanwisher, N. (1998), A cortical representation of the local visual environment, *Nature*, **392**, 598–601.

[22] Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989), Mnemonic coding of visual space in monkey dorsolateral prefrontal cortex, *Journal of Neurophysiology*, **61**, 331–349.

[23] Fuster, J.M. (1997), *The Prefrontal Cortex*, 3rd ed., Raven Press: New York.

[24] Fuster, J.M. (2000), *Memory Systems in the Brain*, Raven Press: New York.

[25] Gaffan, D. (1994), Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection, *Journal of Cognitive Neuroscience*, **6**, 305–320.

[26] Gaffan, D., and Harrison, S. (1989), A comparison of the effects of fornix section and sulcus principalis ablation upon spatial learning by monkeys, *Behavioural Brain Research*, **31**, 207–220.

[27] Gaffan, D., and Harrison, S. (1989), Place memory and scene memory: effects of fornix transection in the monkey, *Experimental Brain Research*, **74**, 202–212.

[28] Gaffan, D., and Saunders, R. C. (1985), Running recognition of configural stimuli by fornix transected monkeys, *Quarterly Journal of Experimental Psychology*, **37B**, 61–71.

[29] Georges-Francois, P., Rolls, E. T., and Robertson, R. G. (1999), Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place, *Cerebral Cortex*, **9**, 197–212.

[30] Goldman-Rakic, P. S. (1996), The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive, *Philosophical Transactions of the Royal Society of London, Series B*, **351**, 1445–1453.

[31] Hasselmo, M. E., Schnell, E., and Barkai, E. (1995), Learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3, *Journal of Neuroscience*, **15**, 5249–5262.

[32] Hertz, J., Krogh, A., and Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Addison Wesley: Wokingham, U.K.

[33] Hopfield, J. J. (1982), Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the U.S.A.*, **79**, 2554–2558.

[34] Jackson, P. A., Kesner, R. P., and Amann, K. (1998), Memory for duration: role of hippocampus and medial prefrontal cortex, *Neurobiology of Learning and Memory*, **70**, 328–348.

[35] Jarrard, E. L. (1993), On the role of the hippocampus in learning and memory in the rat, *Behavioral and Neural Biology*, **60**, 9–26.

[36] Jonas, E. A., and Kaczmarek, L. K (1999), in Katz, P. S. (ed.), *The Inside Story: Subcellular Mechanisms of Neuromodulation*, Oxford University Press: New York 83–120.

[37] Kesner, R.P., and Rolls, E. T. (2001), Role of long-term synaptic modification in short-term memory, *Hippocampus*, **11**, 240-250.

[38] Koch, C. (1999), *Biophysics of Computation*, Oxford University Press: Oxford.

[39] Kubie, J. L., and Muller, R. U. (1991), Multiple representations in the hippocampus, *Hippocampus*, **1**, 240-242.

[40] Kuhn, R. (1990), Statistical mechanics of neural networks near saturation, in Garrido, L. (ed.), *Statistical Mechanics of Neural Networks*, Springer-Verlag: Berlin.

[41] Kuhn, R., Bos, S., and van Hemmen, J. L. (1991), Statistical mechanics for networks of graded response neurons, *Physical Review A*, **243**, 2084–2087.

[42] Lassalle, J. M., Bataille, T., and Halley, H. (2000), Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task, *Neurobiology of Learning and Memory*, **73**, 243–257.

[43] Markus, E. J., Qin, Y. L., Leonard, B. , Skaggs, W., McNaughton, B. L., and Barnes, C. A. (1995), Interactions between location and task affect the spatial and directional firing of hippocampal neurons, *Journal of Neuroscience*, **15**, 7079–7094.

[44] Marr, D. (1971), Simple memory: a theory for archicortex, *Philosophical Transactions of the Royal Society of London, Series B*, **262**, 23–81.

[45] Martin, S. J., Grimwood, P. D., and Morris, R. G. (2000), Synaptic plasticity and memory: an evaluation of the hypothesis, *Annual Review of Neuroscience*, **23**, 649–711.

[46] McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995), Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory, *Psychological Review*, **102**, 419–457.

[47] McNaughton, B. L., Barnes, C. A., and O'Keefe, J. (1983), The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats, *Experimental Brain Research*, **52**, 41–49.

[48] Miyashita, Y., Rolls, E. T., Cahusac, P. M. B. , Niki, H., and Feigenbaum, J. D.

(1989), Activity of hippocampal neurons in the monkey related to a conditional spatial response task, *Journal of Neurophysiology*, **61**, 669–678.

[49] Muller, R. U., Kubie, J. L., Bostock, E. M., Taube, J. S., and Quirk, G. J. (1991), Spatial firing correlates of neurons in the hippocampal formation of freely moving rats, in Paillard, J. (ed.), *Brain and Space*, Oxford University Press: Oxford, 296–333.

[50] Muller, R. U., Ranck, J. B., and Taube, J. S. (1996), Head direction cells: properties and functional significance, *Current Opinion in Neurobiology*, **6**, 196–206.

[51] O'Keefe, J. (1979), A review of the hippocampal place cells, *Progress in Neurobiology*, **13**, 419–439.

[52] O'Keefe, J. (1984), Spatial memory within and without the hippocampal system, in Seifert, W. (ed.) *Neurobiology of the Hippocampus*, Academic Press: London, 375–403.

[53] O'Keefe, J. (1990), A computational theory of the cognitive map, *Progress in Brain Research*, **83**, 301–312.

[54] O'Keefe, J. (1991), The hippocampal cognitive map and navigational strategies, in Paillard, J. (ed.), *Brain and Space*, Oxford University Press:Oxford, 273–295.

[55] O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., and Maguire, E. A. (1998), Place cells, navigational accuracy, and the human hippocampus, *Philosophical Transactions of the Royal Society, London [B]*, **353**, 1333–1340.

[56] O'Keefe, J., and Dostrovsky, J. (1971), The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat, *Brain Research*, **34** , 171–175.

[57] O'Keefe, J., and Nadel, L. (1978), *The Hippocampus as a Cognitive Map*, Clarendon Press: Oxford.

[58] O'Mara, S. M., Rolls, E. T., Berthoz, A., and Kesner, R. P. (1994), Neurons responding to whole-body motion in the primate hippocampus, *Journal of Neuroscience*, **14**, 6511–6523.

[59] Panzeri, S., Rolls, E. T., Battaglia, F., and Lavis, R. (2001), Speed of information retrieval in multilayer networks of integrate-and-fire neurons, *Network: Computation in Neural Systems*, **12**, 423–440.

[60] Parkinson, J. K., Murray, E. A., and Mishkin, M. (1988), A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects, *Journal of Neuroscience*, **8**, 4059–4167.

[61] Petrides, M. (1985), Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man, *Neuropsychologia*, **23**, 601–614.

[62] Ranck, Jr., J. B. (1985), Head direction cells in the deep cell layer of dorsolat-

eral presubiculum in freely moving rats, in Buzsáki, G. and Vanderwolf, C. H. (eds.) *Electrical Activity of the Archicortex*, Akadémiai Kiadó: Budapest.

[63] Rao, S.C., Rainer, G., and Miller, E.K. (1997), Integration of what and where in the primate prefrontal cortex, *Science*, **276**, 821-824.

[64] Miller, E. K., and Desimone, R. (1994), Parallel neuronal mechanisms for short-term memory, *Science*, **263**, 520–522.

[65] Miller, E. K., Erickson, C., and Desimone, R. (1996), Neural mechanism of visual working memory in prefrontal cortex of the macaque, *Journal of Neuroscience*, **16**, 5154–5167.

[66] Miller, E. K., Li, L., and Desimone, R. (1993), Activity of neurons in anterior inferior temporal cortex during a short-term memory task, *Journal of Neuroscience*, **13**, 1460–1478.

[67] Renart, A., Moreno, R., de al Rocha, J., Parga, N., and Rolls, E. T. (2001), A model of the IT–PF network in object working memory which includes balanced persistent activity and tuned inhibition, *Neurocomputing*, **38–40**, 1525–1531.

[68] Renart, A., Parga, N., and Rolls, E. T. (1999), Backprojections in the cerebral cortex: implications for memory storage, *Neural Computation*, **11**, 1349–1388.

[69] Renart, A., Parga, N., and Rolls, E. T. (1999), Associative memory properties of multiple cortical modules, *Network*, **10**, 237–255.

[70] Renart, A., Parga, N.,and Rolls, E. T. (2000), A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks, in Solla, S.A. and Leen, T.K. and Mueller, K.-R. (eds.) *Advances in Neural Information Processing Systems*, MIT Press: Cambridge Mass, **12**, 171–177.

[71] Robertson, R. G., Rolls, E. T., and Georges-François, P. (1998), Spatial view cells in the primate hippocampus: Effects of removal of view details, *Journal of Neurophysiology*, **79**, 1145–1156.

[72] Robertson, R. G., Rolls, E. T., Georges-François, P., and Panzeri, S. (1999), Head direction cells in the primate pre-subiculum, *Hippocampus*, **9**, 206–219.

[73] Rolls, E. T. (1987), Information representation, processing and storage in the brain: analysis at the single neuron level, in *The Neural and Molecular Bases of Learning*, Changeux, J.-P. and Konishi, M. (ed.), Wiley: Chichester, 503–540.

[74] Rolls, E. T. (1992), Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas, *Philosophical Transactions of the Royal Society*, **335**, 11–21.

[75] Rolls, E. T. (1989), Functions of neuronal networks in the hippocampus and neocortex in memory, in Byrne, J.H. and Berry, W.O. (eds.), *Neural Models of Plasticity: Experimental and Theoretical Approaches*, Academic Press: San

Diego, 240–265.

[76] Rolls, E. T. (1996), A theory of hippocampal function in memory, *Hippocampus*, **6**, 601–620.

[77] Rolls, E. T. (1999), *The Brain and Emotion*, Oxford University Press: Oxford.

[78] Rolls, E. T. (1999), Spatial view cells and the representation of place in the primate hippocampus, *Hippocampus*, **9**, 467–480.

[79] Rolls, E. T. (1999), The representation of space in the primate hippocampus, and its role in memory, *The Hippocampal and Parietal Foundations of Spatial Cognition*, in Burgess, N. and Jeffrey, K.J. and O'Keefe, J. (eds.), Oxford University Press: Oxford, 320–344.

[80] Rolls, E. T. (2000), Memory systems in the brain, *Annual Review of Psychology*, **51**, 599–630.

[81] Rolls, E. T. (2000), Hippocampo-cortical and cortico-cortical backprojections, *Hippocampus*, **10**, 380–388.

[82] Rolls, E. T., and Deco, G. (2002), *Computational Neuroscience of Vision*, Oxford University Press: Oxford.

[83] Rolls, E. T., and Milward, T. (2000), A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures, *Neural Computation*, **12**, 2547–2572.

[84] Rolls, E. T., Miyashita, Y., Cahusac, P. M. B., Kesner, R. P., Niki, H., Feigenbaum, J., and Bach, L. (1989), Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown, *Journal of Neuroscience*, **9**, 1835–1845.

[85] Rolls, E. T., Robertson, R. G., and Georges-François, P. (1997), Spatial view cells in the primate hippocampus, *European Journal of Neuroscience*, **9**, 1789–1794.

[86] Rolls, E. T., and Stringer, S. M. (2000), On the design of neural networks in the brain by genetic evolution, *Progress in Neurobiology*, **61**, 557–579.

[87] Rolls, E. T., and Stringer, S. M. (2001), A model of the interaction between mood and memory, *Network: Computation in Neural Systems*, **12**, 89-109.

[88] Rolls, E. T., Stringer, S. M., and Trappenberg, T. P. (2002), A unified model of spatial and episodic memory, *Proceedings of the Royal Society B*, **269**, 1087–1093.

[89] Rolls, E. T., and Tovee, M. J. (1994), Processing speed in the cerebral cortex and the neurophysiology of visual masking, *Proceedings of the Royal Society, B*, **257**, 9–15.

[90] Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., and Azzopardi, P.

(1994), The responses of neurons in the temporal cortex of primates, and face identification and detection, *Experimental Brain Research*, **101**, 474–484.

[91] Rolls, E. T., Tovee, M. J., and Panzeri, S. (1999), The neurophysiology of backward visual masking: information analysis, *Journal of Cognitive Neuroscience*, **11**, 335–346.

[92] Rolls, E. T., and Treves, A. (1998), *Neural Networks and Brain Function*, Oxford University Press: Oxford.

[93] Rolls, E. T., Treves, A., Robertson, R. G., Georges-François, P., and Panzeri, S. (1998), Information about spatial view in an ensemble of primate hippocampal cells, *Journal of Neurophysiology*, **79**, 1797–1813.

[94] Rupniak, N. M. J., and Gaffan, D. (1987), Monkey hippocampus and learning about spatially directed movements, *Journal of Neuroscience*, **7**, 2331–2337.

[95] Samsonovich, A., and McNaughton, B.L. (1997), Path integration and cognitive mapping in a continuous attractor neural network model, *Journal of Neuroscience*, **17**: 2331-2337.

[96] Shiino, M., and Fukai, T. (1990), Replica-symmetric theory of the nonlinear analogue neural networks, *Journal of Physics A: Math. Gen.*, **23**, L1009–L1017.

[97] Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995), A model of the neural basis of the rat's sense of direction, in Tesauro, G., Touretzky, D. S., and Leen, T. K. (eds.) *Advances in Neural Information Processing Systems*, vol. 7, 173–180, MIT Press: Cambridge, Massachusetts. **17**, 5900–5920.

[98] Shallice, T., and Burgess, P. (1996), The domain of supervisory processes and temporal organization of behaviour, *Philosophical Transactions of the Royal Society of London. Series B Biological Sciences*, **351**, 1405–1411.

[99] Smith, M. L., and Milner, B. (1981), The role of the right hippocampus in the recall of spatial location, *Neuropsychologia*, **19**, 781–793.

[100] Squire, L. R., and Knowlton, B. J. (2000) The medial temporal lobe, the hippocampus, and the memory systems of the brain, in *The New Cognitive Neurosciences*, Gazzaniga, M.S. (ed.), 765–779, 2nd ed., MIT Press: Cambridge, MA.

[101] Stringer, S. M., Trappenberg, T. P., Rolls, E. T., and de Araujo, I. E. T. (2002), Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells, *Network: Computation in Neural Systems*, **13**, 217–242.

[102] Stringer, S. M., and Rolls, E. T. (2002), Self-organizing continuous attractor network models of spatial view cells for an agent that is able to move freely through different locations (submitted).

[103] Stringer, S. M., and Rolls, E. T. (2002), Hierarchical dynamical models of motor function (submitted).

[104] Stringer, S. M., Rolls, E.T., Trappenberg, T. P., and de Araujo, I. E. T. (2002), Self-organizing continuous attractor networks and path integration: Two-dimensional models of place cells, *Network: Computation in Neural Systems*, **13**: 429-446.

[105] Stringer, S. M., Rolls, E. T., and Trappenberg, T. P. (2002), Self-organizing continuous attractor network models of hippocampal spatial view cells (in press).

[106] Suzuki, W. A., and Amaral, D. G. (1994), Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents, *Journal of Comparative Neurology*, **350**, 497–533.

[107] Suzuki, W. A., and Amaral, D. G. (1994), Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices, *Journal of Neuroscience*, **14**, 1856–1877.

[108] Suzuki, W. A., Miller, E. K., and Desimone, R. (1997), Object and place memory in the macaque entorhinal cortex, *Journal of Neurophysiology*, **78**, 1062–1081.

[109] Taube, J. S., Goodridge, J. P., Golob, E. G., Dudchenko, P. A., and Stackman, R. W. (1996), Processing the head direction signal: a review and commentary, *Brain Research Bulletin*, **40**, 477–486.

[110] Taube, J. S., Muller, R. U., and Ranck, Jr., J. B. (1990), Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis, *Journal of Neuroscience* , **10**, 420–435.

[111] Taylor, J. G. (1999), Neural ‘bubble’ dynamics in two dimensions: foundations, *Biological Cybernetics*, **80**, 393–409.

[112] Treves, A. (1993), Mean-field analysis of neuronal spike dynamics quantitative estimate of the information relayed by the Schaffer collaterals, *Network*, **4**, 259–284.

[113] Treves, A., and Rolls, E. T. (1991), What determines the capacity of autoassociative memories in the brain?, *Network*, **2**, 371–397.

[114] Treves, A., and Rolls, E. T. (1992), Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network, *Hippocampus*, **2**, 189–199.

[115] Treves, A., and Rolls, E. T. (1994), A computational analysis of the role of the hippocampus in memory, *Hippocampus*, **4**, 374–391.

[116] Wallis, G., and Rolls, E. T. (1997), Invariant face and object recognition in the visual system, *Progress in Neurobiology*, **51**, 167–194.

[117] Wilson, M. A., and McNaughton, B. L. (1993) Dynamics of the hippocampal ensemble code for space, *Science*, **261**, 1055-1058.

[118] Wilson, F. A. W., O'Sclaidhe, S. P., and Goldman-Rakic, P. S. (1993), Dissociation of object and spatial processing domains in primate prefrontal cortex, *Science*, **260**, 1955–1958.

[119] Zhang, K. (1996), Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory, *Journal of Neuroscience*, **16**, 2112–2126.

[120] Zhang, W. ,and Dietterich, T. G. (1996), High-performance job-shop scheduling with a time-delay TD(λ) network, in Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E. (eds.) *Advances in Neural Information Processing Systems 8,* Cambridge MA: MIT Press, 1024–1030.

[121] Zola-Morgan, S., Squire, L. R., Amaral, D. G., and Suzuki, W. A. (1989), Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment, *Journal of Neuroscience*, **9**, 4355–4370.

[122] Zola-Morgan, S., Squire, L. R., and Ramus, S. J. (1994), Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system, *Hippocampus,* **4**, 483–494.

Chapter 17

Modelling Motor Control Paradigms

Pietro G. Morasso, and Vittorio Sanguineti

University of Genova, DIST (Department of Informatics Systems and Telecommunications), Via Opera Pia 13, I-16145 Genova, Italy

CONTENTS

17.1 Introduction: the ecological nature of motor control

Modelling the way in which humans learn to coordinate their movements in daily life or in more demanding activities is an important scientific topic from many points of view, such as medical, psychological, kinesiological, cybernetic. This chapter analyses the complexity of this problem and reviews the variety of experimental and theoretical techniques that have been developed for this purpose.

With the advent of technical means for capturing motion sequences and the pioneering work of Marey [41] and Muybridge [53] in this area, the attempt of describing, modelling and understanding the organisation of movement has become a scientific topic. The fact that human movements are part of everyday life paradoxically hides their intrinsic complexity and justifies initial expectations that complete knowledge could be achieved simply by improving the measurement techniques and carrying out a few carefully designed experiments. Unfortunately, this is not the case. Each experiment is frequently the source of more questions than answers and thus the attempt to capture the complexity of purposive action and adaptive behaviour, after over a century of extensive multidisciplinary research, is far from over.

The conventional view is based on a separation of perception, movement and cognition and the segregation of perceptual, motor and cognitive processes in different parts of the brain, according to some kind of hierarchical organisation. This view is rooted in the empirical findings of neurologists of the 19th century, such as J. Hughlings Jackson, and has a surprising degree of analogy with the basic structure of a modern PC that typically consists of input and output peripherals connected to a central processor. Perhaps the analogy with modern technology justifies why this old-fashioned attitude still has its supporters, in spite of the massive empirical and conceptual challenge to this view and its inability to explain the range of skills and adaptive behaviours that characterize biological organisms.

Let us consider perception, which is the process whereby sensory stimulation is translated into organised experience. That experience, or percept, is the joint product of the stimulation and of the process itself, particularly in the perception and representation of space. An early theory of space perception put forth by the Anglican bishop G. Berkeley at the beginning of the 18th century was that the third dimension (depth) cannot be directly perceived in a visual way since the retinal image of any object is two-dimensional, as in a painting. He held that the ability to have visual experiences of depth is not inborn but can only result from logical deduction based on empirical learning through the use of other senses. The first part of the reasoning (the need of a symbolic deductive system for compensating the fallacy of the senses) is clearly wrong and the roots of such misconception can be traced back to the neoplatonic ideas of the Italian Renaissance, in general, and to the Alberti's window metaphor, in particular. Also the Cartesian dualism between body and mind is just another face of the same attitude and such Descartes' error, to quote A.R. Damasio [15], is on a par with the Berkeley's error above and is at the basis of the intellectualistic effort to explain the computational complexity of perception which characterizes a great part of the classic artificial intelligence approach. However, the latter part of Berkeley's conjecture (the emphasis on learning and intersensory integration) is surprisingly modern and agrees, on one hand, with the modern approach to neuropsychological development pioneered by J. Piaget [55] and, on another, with the so called connectionist point of view, originated in the 1980s as a computational alternative to classic artificial intelligence.

An emergent idea is also the motor theory of perception, well illustrated by A. Berthoz [8], i.e., the concept that perception is not a passive mechanism for receiving and interpreting sensory data but is the active process of anticipating the sensory consequences of an action and thereby binding the sensory and motor patterns in a coherent framework. In computational terms, this implies the existence in the brain of some kind of internal model, as a bridge between action and perception. As a matter of fact, the idea that the instructions generated by the brain for controlling a movement are utilized by the brain for interpreting the sensory consequences of the movement is already present in the pioneering work of Helmholtz and von Uexküll and its influence has resurfaced in the context of recent control models based on learning (e.g., [71]). The generally used term is corollary discharge [29] and implies an internal comparison between an out-going signal (the efferent copy) and the corresponding sensory re-afference: the coherence of the two representations is the basis for the stability of our sensorimotor world. This kind of circularity and complementarity between sensory and motor patterns is obviously incompatible with the conventional reasoning based on hierarchical structures. A similar kind of circularity is also implicit in Piaget's concept of circular reaction, which is assumed to characterize the process of sensorimotor learning, i.e., the construction of the internal maps between perceptually identified targets and the corresponding sequence of motor commands.

An additional type of circularity in the organism/environment interaction can be identified at the mechanical interface between the body and the outside world, where the mechanical properties of muscles interact with the physics of inanimate objects. This topic area has evolved from the Russian school, with the early work on the nature of reflexes by I.P. Pavlov and the subsequent critical re-examination by P.K. Anhokin and N. Bernstein [1, 7]. In particular, we owe to Bernstein the seminal observation (*the comparator model*) that motor commands alone are insufficient to determine movement but only identify some factors in a complex equation where the world dynamics has a major influence. This lead, among other things, to the identification of muscle stiffness as a relevant motor parameter and the formulation of the theory of equilibrium-point control [10, 17].

In general we may say that, in different ways, Helmholtz's corollary discharge, Piaget's circular reaction, and Bernstein's comparator model are different ways to express the ecological nature of motor control, i.e., the partnership between brain processes (including muscles) and world dynamics. This concept is graphically sketched in Figure 17.1. On the other hand, these general ideas on motor control could not provide, immediately, mathematical tools of analysis from which to build models and perform simulations. The art and science of building motor control models is

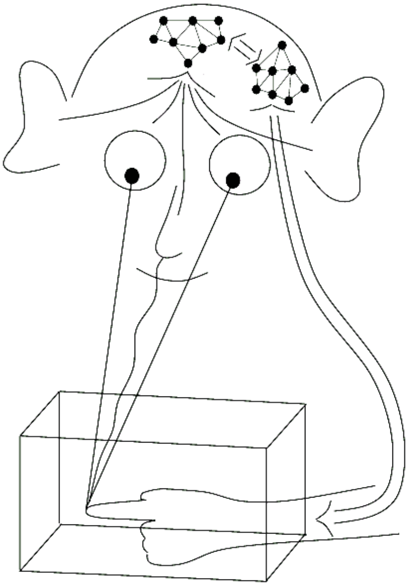

Figure 17.1
Ecological nature of the sensorimotor control system.

a later development and has been influenced by the methods designed by engineers in the field of automatic control and computer science. Most of the techniques are based on linear approximations and explicit schematisations of the phenomena. In particular, two main concepts can be singled out for their influence on the study of motor control: the concept of feedback and the concept of motor program. Their level of influence in brain theory is certainly determined by the tremendous success of these techniques in the modern technological world. However their applicability to what we may call the biological hardware is questionable for two main reasons: (i) Feedback control can only be effective and stable if the feedback delays are negligible, but this is not the case of biological feedback signals, where transduction and transmission delays add up to tens of milliseconds; (ii) The concept of motor program implies a sequential organisation, which only can be effective if the individual steps in the sequence are sufficiently fast, and this is in contrast with the parallel, distributed processing of the brain made necessary by the relative slowness of synaptic processing. In fact, the recognition of the limits of the analytic-symbolic approach has motivated, since the late 1980s, a re-evaluation of earlier approaches under the new light of connectionist thinking.

17.2 The robotic perspective

17.2.1 Logical decomposition of motor control into cascaded computational processes

The computational process that is necessary for realising a planned motor task has been the subject of a great deal of research in robotics. As sketched in Figure 17.2, five main blocks can be identified which, in the robotic approach, correspond to different procedures that, at least in principle, should be programmed independently and executed in a sequence:

1. *Planning:* this implies a detailed characterization of an intended movement by selecting the initial point, the target point, the position of obstacles, the size and orientation of the gesture, the time base, etc., independently of the specific end-effector,

   Plan:{ Collection of critical points in space and time }

2. *Trajectory formation,* which breaks down the planned trajectory into a sequence of elementary movements or strokes, identified by suitable via-points and smoothly joined together in a quasi-continuous curve

   $$x(t) = \sum_i \mathrm{stroke}_i(t - t_i)$$

   For general movements in space $x(t)$ is a 6-dimensional vector, independent of the specific end-effector.

3. *Inverse kinematics:* at this stage there must be the selection of the end-effector, with the corresponding kinematic chain, and the transformation of the time vector $x(t)$ into a vector of joint rotations $q(t)$. Geometrically this corresponds to a coordinate transformation from space coordinates (task space) to joint coordinates (joint space). If the dimensionality of the q-vector is $n = 6$, then the kinematic chain is redundant and the kinematic inversion in general admits an infinite number of solutions, described by the so-called null space of the kinematic transformation, within the workspace defined by the geometry of the arm and the joint limits. In particular, if $J(q)$ is the Jacobian matrix of the $x \rightarrow q$ transformation, then the following relation holds, also known as direct kinematic equation:

   $$\dot{x} = J(q)\dot{q} \tag{17.1}$$

   In the case of redundancy the null space is described by the following relation $0 = J(q)\dot{q}$ and its dimensionality is n-6. The solution of the inverse kinematic equations can be carried out instantaneously by using the pseudo-inverse or Moore-Penrose matrix:

   $$\dot{q} = [J(q)^T J(q)]^{-1} J(q)^T \dot{x} \tag{17.2}$$

However this solution is not numerically robust, close to the singularities of the kinematic transformation, and is *non-integrable* in the sense that, if we command a robot by means of this algorithm in relation with a desired closed trajectory in task space repeated several times (as in crank turning), we get a trajectory in joint space which is not closed and tends to drift towards the joint limits of the kinematic chain.

4. *Inverse dynamics and interaction forces:* computing inverse dynamics corresponds to the solution of the following equation

$$\tau_{\text{actuator}}(t) = I(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + J^T(q)F_{ext} \tag{17.3}$$

where $q(t)$ is the planned/desired trajectory in the joint space, $I(q)\ddot{q}$ identifies the inertial forces proportional to the acceleration, $C(q,\dot{q})\dot{q}$ the Coriolis forces quadratically dependent upon speed, $G(q)$ the gravitational forces independent of time, and F_{ext} the external disturbance/load applied to the end-effector (Figure 17.3). This equation is highly nonlinear and acts as a sort of *internal disturbance* which tends to induce deformations of the planned trajectories. In particular, the inertial terms is predominant during the initiation and termination of the movements whereas the Coriolis terms is more important in the intermediate, high-speed parts of the movements. Figure 17.4 shows a simulation involving a planar arm with two degrees of freedom. A set of 8 different trajectories were generated starting from the same initial point. For each of them Equation 17.1 was applied and the computed torque vectors were re-mapped as end-effector force vectors in the following way:

$$\begin{aligned} &\tau_{\text{actuator}}(t) = J^T(q)F_{\text{end-effector}} \\ &\Rightarrow F_{\text{end-effector}} = (J^T(q))^{-1}\tau_{\text{actuator}}(t) \end{aligned} \tag{17.4}$$

The figure shows that the patterns of end-effector forces are quite variable in relation with the movement direction. In particular, each force vector can be decomposed into a longitudinal component (oriented as the intended movement direction) and a transversal component. While the longitudinal component corresponds to the usual inertial resistance that would be present even if the degrees of freedom were controlled separately, one after the other, the transversal component is related to the non-linear dynamics determined by the interaction among degrees of freedom and, if unaccounted for, will tend to deviate laterally the planned trajectory. As can be seen from the figure, the order of magnitude of such lateral disturbances or interaction forces is the same as the main inertial components. Moreover it can be seen that the initial and final parts of the movements (characterized by low velocity but high acceleration) tend to be more affected than the intermediate high-velocity part. Obviously interaction torques only occur in multi-joint movements: if movements are broken down as sequences of single-joint rotations, then interaction forces disappear and this might explain why in several neuromotor pathologies movement segmentation is a typical adaptation to the impairment.

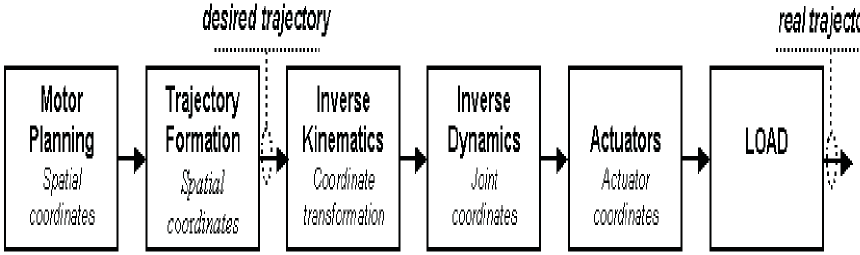

Figure 17.2
Logical decomposition of motor control into cascaded computational processes.

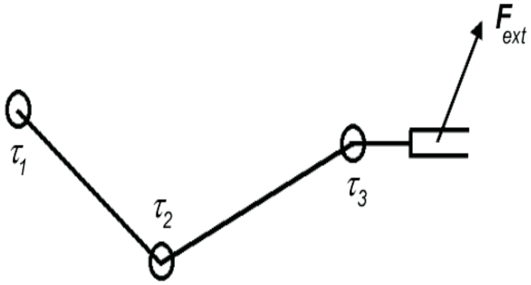

Figure 17.3
Internal and external forces in a kinematic chain. In the usual control paradigm F_{ext} is related to the load and τ_i's are the controlled actuator forces. In the passive motion paradigm F_{ext} is related to the planned trajectory and τ_i's correspond to the passive compliance of the joints to the planned pull.

5. *Actuator control:* the typical robotic actuators are force-controlled and thus the output of the inverse dynamics calculations can be directly fed to the control variables of the actuators.

17.2.2 The Achilles' heel of feedback control

In practice, the scheme of Figure 17.2 cannot be applied directly because the dynamical model is uncertain and partially unknown. Thus, on top of a computational process structured as an approximation of Figure 17.2 usually there is a layer of feedback control which can use positional and/or force feedback (Figure 17.5). The main problem of feedback control, particularly from the perspective of using it as a possible paradigm for modelling biological motor control, is that it is critically dependent

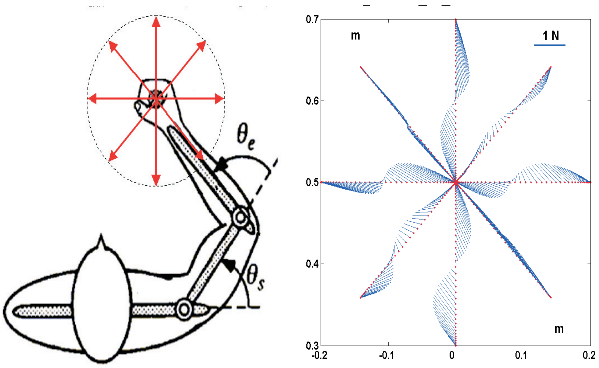

Figure 17.4

Interaction forces in reaching movements. Starting from the resting position depicted in the left panel (50 cm from the shoulder, with the elbow 90° flexed) 8 different reaching movements are considered in equally spaced directions. The nominal trajectories are generated according to the minimum-jerk criterion (duration: 0.7 s; amplitude: 20 cm). The right panel shows, for each of them, the pattern of actuator forces, mapped onto the end-effector. Parameters of the simulation: arm (mass: 2 Kg; length: 30 cm); forearm (mass: 2 Kg; length: 40 cm).

upon the delays in the control loop.

Let us consider a very simple example: a second-order load (such as a spring-mass-dashpot system), with a natural frequency of 1 Hz and a damping factor of 0.5, under the action of a standard PD (proportional + derivative) controller. If we suppose that the ratio between the proportional and derivative gains is 10:1 and compute the overall gain in such a way that the asymptotical precision of the closed-loop control in the step-response is 90%, we get the Bode diagram of Figure 17.6. In particular, the diagram shows that the phase-margin ϕ of the closed-loop control is 84^o at a frequency ω of 36 rad/s. If we now suppose that there is a delay in the feedback loop of T s and we consider that this delay introduces a frequency-dependent phase-shift equal to $\Delta\phi = \omega T$, we can compute the limit value of the delay beyond which the control becomes unstable. In the example it turns out that it is 40 ms or even less if the ratio above is less than 10:1. In robotic applications it is not difficult to limit the delays to values which are 1 or 2 order of magnitudes smaller than limit value above. This is not the case, unfortunately, of the biological motor control and for this reason a great part of the motor control mechanisms are implemented in terms of feedforward paradigms, based on suitable internal models, or on non-reflexive feedback paradigms in which internal models are used for compensating delays by means of prediction.

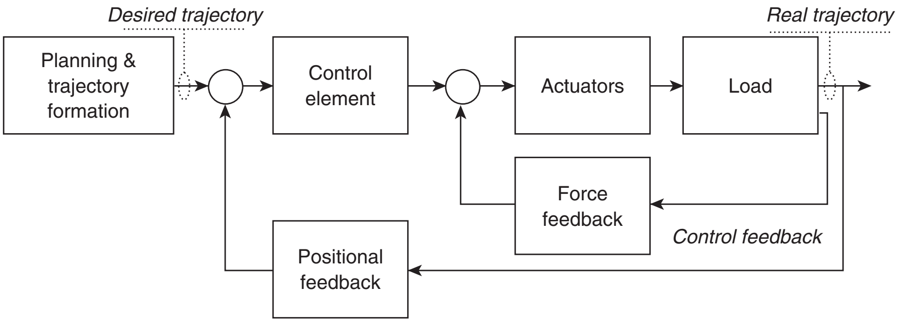

Figure 17.5

Typical block diagram of feedback control. Sensory feedback has a direct control action.

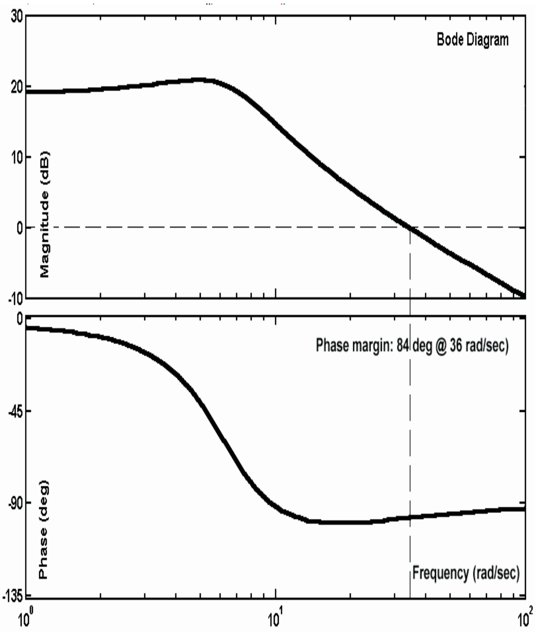

Figure 17.6

Bode diagram of a feedback controlled spring-mass model. The phase margin is computed by considering the frequency at which the log-magnitude plot intersects the 0 dB line and measuring the difference between the critical phase value (180°) and the actual phase, derived from the phase plot. Stability requires a positive phase margin and a value greater that 45° is required for suitable damping.

17.3 The biological perspective

17.3.1 Motor equivalence and spatio-temporal invariances

The experimental study of movements in humans and other mammals has shown that voluntary movements obey two important psychophysical principles, from the point of view of trajectory formation, in addition to the logarithmic dependence of choice reaction time upon the number of choices (a.k.a. Hick-Hyman law: [31]) and of movement time upon relative accuracy (a.k.a. Fitts law: [18]):

1. *The principle of motor equivalence* [7, 26]: spatio-temporal features of a planned movement are independent of the selected end-effector and thus of the specific muscles involved;

2. *The principle of kinematic invariance in the task space* (Figure 17.6): hand movements to a target are approximately straight with bell-shaped speed profiles [45], independently of the movement direction and amplitude. Applied to more complex trajectories, as in handwriting, the same principle predicts a correlation between speed and curvature [37, 46] which is consistent with the overlapped composition of subsequent primitive motor patterns similar to the reaching movements. Such invariant spatio-temporal features of normal movements can be explained by a variety of criteria of maximum smoothness, such as the minimum jerk criterion (the jerk-index is computed as the logarithm of the normalised time integral of the squared norm of the third time derivative of the hand trajectory: [20]) or the minimum torque-change criterion [68]. Kinematic invariance and smoothness are already present in the cortical patterns that precede movements [22] and thus are not a mere consequence of the *filtering action* of the peripheral motor apparatus.

As regards inverse kinematics, particularly in the case of redundant kinematic chains, a computational model was proposed by Mussa-Ivaldi et al. [52] which is based on the *Passive Motion Paradigm*. It consists of the relaxation of (an internal model of) the overall kinematic chain to the virtual pull of a force applied to the end-effector, in the direction of the intended movement. In a sense, it is the neuromotor analog of the mechanism of coordinating the motion of a wooden marionette with jointed limbs by means of attached strings. The computational merit of the model is its robustness because it achieves kinematic inversion without an explicit ill-posed inversion process, but operates with the aid of a well-posed direct computation: the passive relaxation to the virtual pull. This method does not suffer the problem of non-integrability of the Moore-Penrose method because it is *passive*. In order to make clear the meaning of passive relaxation, let us reconsider Figure 17.3, which was intended to sketch a real arm driven by a set of joint actuators capable to compensate the internal load of the Lagrangian dynamics and the external load represented by the force F_{ext}: in the Passive Motion Paradigm F_{ext} is not a real force but a virtual force vector which is generated by a motor planning process and specifies

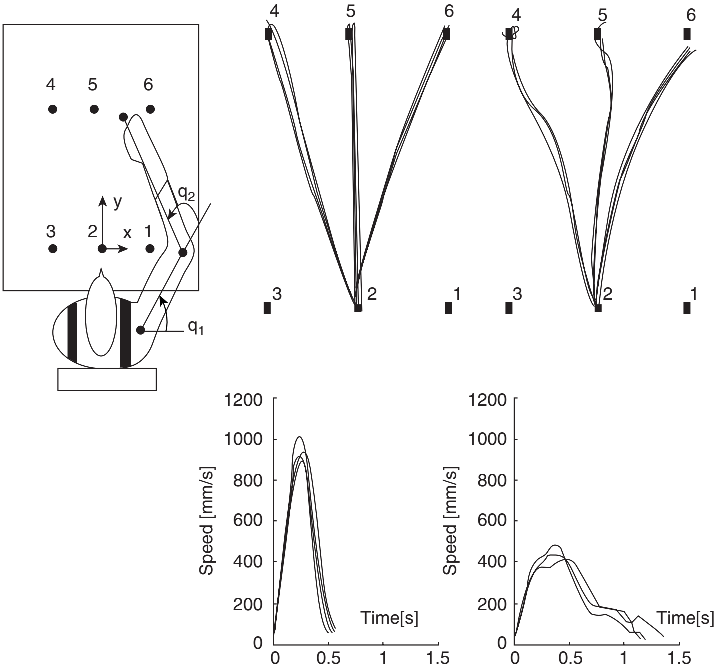

Figure 17.7

Planar reaching movements. Left panel: schematic experimental setup. Central panel: hand trajectories (top) and speed profile (bottom) in a typical control subject. Right panel: hand trajectories (top) and speed profile (bottom) in a typical cerebellar patient

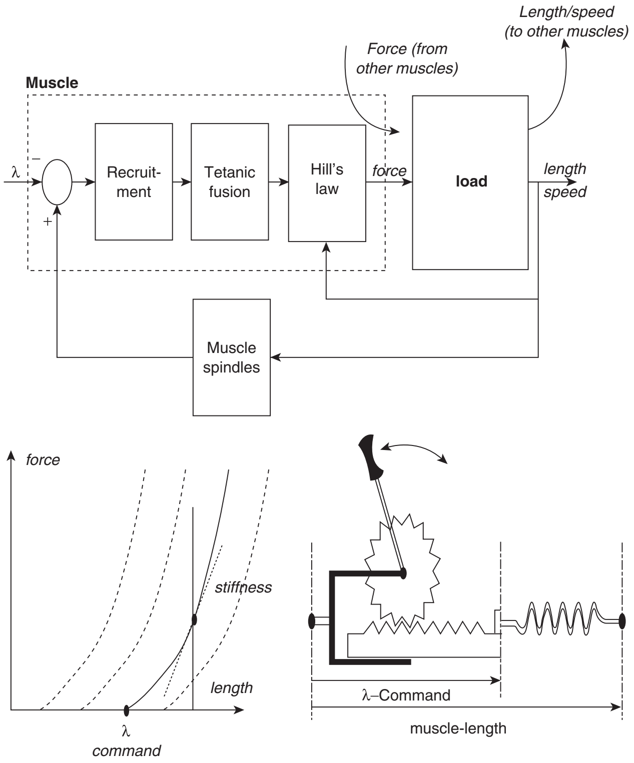

Figure 17.8

Model of muscle control based on the λ-model. λ is the threshold of the monosynaptic stretch reflex as well as the rest-length of the muscle (i.e., the value of muscle length at which muscle force vanishes). Recruitment is consistent with the size-principle and justifies the exponential shape of the length-tension curves. Tetanic fusion, which refers to the relatively slow force build-up for sudden changes of motor commands, is implemented as a low-pass filter. Hill's law is the non-linear dependence of muscle's force on the muscle's speed of contraction. Stiffness is the slope of the length-tension curve and provides an instantaneous length-feedback, which is added to the velocity-feedback related to the Hill's law. Spinal feedback (through muscle spindles) is not instantaneous and has a significant delay.

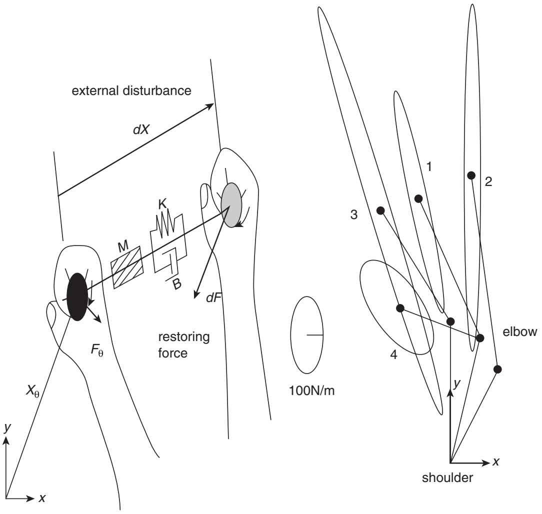

Figure 17.9

Conceptual scheme for the measurement of the hand mechanical impedance. As shown in the left panel, a grasped handle transmits to the hand small and quick disturbance vectors *dX* by means of a computer controlled robotic arm, starting from a given resting position. The restoring force vectors *dF* are measured and the response patterns, for many directions of *dX*, are fitted by a spring-mass-damper system, yielding a set of polar plots. In particular, the polar plot of the elastic coefficient (i.e., the stiffness) is shown in the right panel, for different positions of the hand in the workspace.

the intended direction and speed of the end-effector. Relaxation is a direct, not an inverse computational process and it can be implemented in a very natural way in terms of neural networks by means of interactive self-organised cortical maps [47].

17.3.2 The viscous-elastic properties of the human muscles

Perhaps the most marked difference between the robotic and the biological perspective in the field of motor control is that in contrast with the typical robotic actuators (torque motors), which are force-controlled and uni-directional (in torque motors the output torque only depends on the applied current, independent of the load reaction), the biological actuators (striated muscles) are position-controlled and bi-directional: the actual force delivered by the muscle to the load is a function of the descending motor command and of the reaction of the load. This bi-directional relationship is characterised by two main components:

- A length-dependent component, which is equivalent to an energy-storage elastic characteristic of the muscle [17, 56];

- A velocity-dependent Hill-type component, which captures the dissipative, viscous characteristic of the muscle [70].

What is important, from the motor control point of view, is that the length-tension curves of the muscles cannot be approximated in a linear way but have a characteristic exponential course, related to the progressive recruitment of motor units). This means that muscle stiffness is not constant but is a function of the particular equilibrium point. In fact, the descending motor commands determine the point of intersection λ of the exponential curves with the horizontal line: see Figure 17.8. In this model λ is the controllable parameter, or λ-command, that sets the activation threshold of the monosynaptic stretch reflex and thus determines the *rest-length* of the muscle-spring. In this sense muscles are position controlled and by exploiting the fact that each joint is activated by antagonistic groups of muscles the brain can determine independently the overall equilibrium point (via a pattern of *reciprocal* λ-commands to antagonistic muscles) and the overall stiffness (via a pattern of *coactivation* λ-commands to synergistic muscles).

Such bi-directional characteristics of the muscles in a kinematic chain, such as the arm, are combined together determining an elastic interaction between the end-effector and the load which has markedly anisotropic features (see the stiffness ellipses of Figure 17.9). This ellipse can be computed experimentally by generating small disturbances in different directions and measuring the restoring force vector. As seen in the figure, the orientation of the stiffness ellipses appear to be characterised by a polar pattern, with the long axis (where the hand appears to be stiffer) aligned along the shoulder-hand direction. The size of such ellipses is easily under voluntary control by modulation of the overall coactivation of arm muscles. On the contrary, the orientation of the ellipses does not appear to be under immediate voluntary control, with the exception of highly learned movements [14].

The bi-directional characteristics of the human muscles make them much more flexible than typical robotic actuators and the implications for the theory of motor control, only acknowledged in the late seventies, are still somehow controversial. There is no doubt that muscle stiffness can be seen as a kind of implicit (and thus instantaneous) feedback mechanism that tends to overcome the action of external and internal disturbances and loads, such as the action of gravity and the intrinsic dynamics of the body masses. The big question is concerned with the quantitative and functional relevance of this effect. For some researchers muscle stiffness is all that is needed, without any kind of internal dynamic models to compensate for internal and external disturbances. In this view, the brain is only supposed to generate smooth equilibrium-point trajectories (the reciprocal commands) and to set-up an appropriate level of coactivation. A very important feature of this model is that it assigns a computational role to the muscles, in addition to its obvious executive action. In spite of its elegance and appealing *ecological* nature, this extreme form of equilibrium-point control model would only be plausible if the empirical values of muscle stiffness, at equilibrium as well as during movement, were strong enough in relation with the usual dynamics of body motion. The problem is that this is a difficult type of measurement and the available experimental data in dynamic conditions are quite limited [23]. Yet there is a growing consensus that, although stiffness is certainly relevant as a co-factor in load-compensation, it is not sufficient by alone particularly in more demanding dynamic tasks. For example, its relative importance is likely to be much greater in miniature movements, as in the case of handwriting, which involve relatively small masses, than in the case of large sport gestures which may involve large masses, high speed and a high level of required accuracy.

17.3.3 Dynamic compensation: anticipatory feedforward and feedback control

The alternative solution to a pure *stiffness compensation* of internal and external disturbances is some combination of *anticipatory feedforward* & *feedback* control, in addition to the implicit feedback provided by muscle stiffness and the reflexive feedback provided by segmental mechanisms (Figure 17.10).

In general, a feedforward control model is based on the pre-programmed (and thus anticipatory) computation of the disturbances that will be encountered by a system when it attempts to carry out a desired plan of motion and thus it is, in a very general sense, an inverse model of the controlled plant. This computation is complex and requires learning; a good example is given the *feedback error learning* model [33] where the *feedforward controller* is trained according to the residual errors of an underlying feedback controller (Figure 17.11): in particular, the feedback error (the discrepancy between the desired and real trajectories) is used as the learning signal of the trainable feedforward model, which gradually takes over the responsibility of counterbalancing the dynamic disturbances, thus acquiring an *internal inverse model* of body dynamics. On the other hand, such learning/control scheme does not work with unstable plants/loads, like in the stabilisation of the standing posture, because cannot provide a point-attractor to a system which does not have it in the first place.

Anticipatory feedback control is responsible for the correction of the outgoing motor commands on the basis of sensory (typically visual and proprioceptive) information; in contrast with feedforward control, it requires an *internal forward model* of the body dynamics which combines a copy of the efferent command patterns with the delayed reafferent signals and thus can reconstruct, in a similar way to a Kalman filter, the actual state of the plant [16]. Also this internal model requires a process of learning but it is conceptually simpler and computationally less critical than learning the internal inverse model.

Indeed, the two control modalities seem to coexist in the motor system [59]. Experimental evidence suggests that both components may use some form of *internal model* of body dynamics [9, 25]. The view of the cerebellum as a computing machinery that has competence as regards the physics of the body [12, 32, 33, 34, 42] suggests that the cerebellar circuitry is likely to play an important role in carrying out these tasks and hints at cerebellar syndromes as crucial pathological conditions for understanding the role of the cerebellum in motor coordination.

17.4 The role of cerebellum in the coordination of multiple joints

Cerebellar syndromes, also known as ataxias, form a useful case study to improve our understanding of the mechanisms underlying sensorimotor coordination. Ataxia is the main sign of cerebellar dysfunction. According to the classic description by Holmes [27, 28], the term indicates multiple problems in the planning and execution of movements, including: (1) delay in movement initiation, (2) inaccuracy in achieving a target (dysmetria), (3) inability to perform movements of constant force and rhythm (dysdiadochokinesia), and (4) difficulty to coordinate multi-joint movements. Additional cerebellar symptoms are diminished resistance to passive limb displacement (hypotonia) and kinetic tremor (a pattern which appears at movement onset and increases in amplitude while approaching the target).

Quantitative methods for movement analysis have a long history. (In fact, the classic Holmesian description of ataxia is based on an ingenious technique for recording hand trajectories.) Methods based on kinematic and/or kinetic analysis of movements may potentially allow one to identify more subtle aspects of movement disorders as well as small changes in the degree of the involved impairments over time. For example, kinematic measurements of single-joint arm movements in patients with cerebellar ataxia [13] have provided a precise description of the alteration of the temporal structure of these movements. While patients preserved the linear relationship between peak velocity and movement amplitude which is typical of normal subjects, the speed profiles were asymmetric, with a longer deceleration phase. The asymmetry of the speed profiles was also observed in multi-joint arm movements [5, 60]: see Figure 17.7, which compares the typical patterns of normal subjects (left

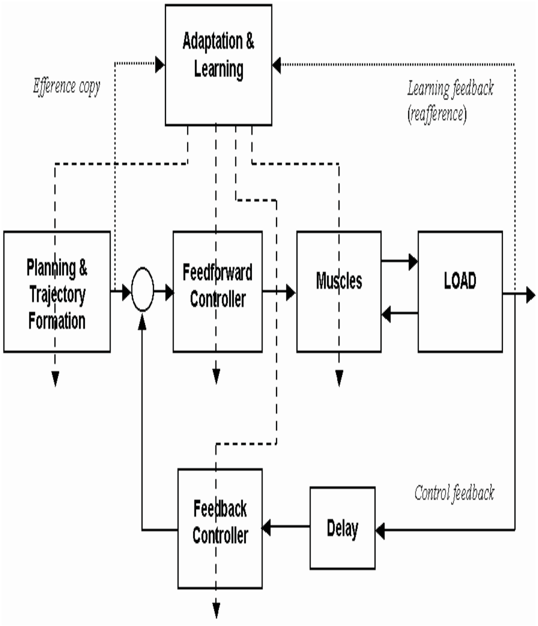

Figure 17.10

Scheme of action of anticipatory feedforward and feedback control. The feedforward controller implements an approximate inverse model of the load/plant. The model can be learned by comparing a copy of the efferent signals and the concurrent reafferent signals. The feedback controller implements an approximate direct model of the load/plant in order to recover the intrinsic delays of the afferent signals. Therefore both controllers must be trained and operate in an anticipatory manner. The feedforward controller is typically used in the compensation of self-generated disturbances, as in arm-trajectory formation, but is ineffective if the load/plant is unstable. In that case an anticipatory feedback controller is more appropriate, implemented as a sampled-data control system. Modulation of joint mechanical impedance is a synergistic process, with a typical relaxation of the level of stiffness as the trained internal model become more and more precise.

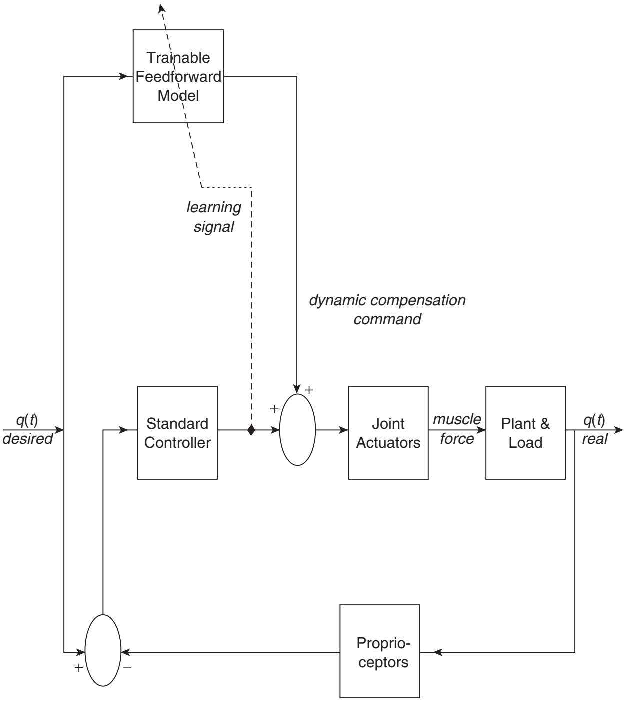

Figure 17.11

Feedback error learning. The purpose of the controller is to generate a control variable that allows the controlled variable $q(t)$ to faithfully reproduce its reference trajectory. The trainable feedforward control model operates in parallel with a standard feedback controller. At the beginning of learning the feedback controller has full responsibility for driving the system, at least in an approximate way. Its output signal is an indirect measurement of error and thus can be used as a learning error signal for the feedforward model, in association with a copy of the reference /desired signal. As learning proceeds, the error between the controlled variable and its reference value becomes less and less and thus the feedback controller diminishes its contribution to the generation of the control commands which, ultimately, will be fully generated in an anticipatory manner by the feedforward controller.

panel) and ataxic patients (right panel).

In general, multi-joint movements tend to enhance pathological features of the movements such as the un-coordination of shoulder and elbow joints, the curvature of hand paths, and dysmetria [4, 24, 35]. These authors found evidence that speed asymmetry is mainly due to an abnormal timing between the different joint rotations, because the individual joints exhibit almost normal angular speed profiles.

There is also growing evidence about the importance of *interaction torques*, which are associated with the multi-link structure of the human arm and consist of a combination of centripetal, Coriolis and inertial components (see Equation 17.3). These torque components need to be explicitly accounted for by the controller, otherwise the resulting trajectories tend to be significantly distorted, as explained in the previous sections. A related observation is that in healthy subjects the total muscle torque in each joint is highly correlated with the interaction torques whereas in cerebellar patients this correlation is absent. This finding can be interpreted as a result of a defective compensation of interaction torques, which only occur in multi-joint movements. Another consequence of the same mechanism is the decomposition of movements into separated segments [4, 65, 66], because by moving one joint at a time the interaction torques can be eliminated. In fact, interaction torques express the dynamical coupling among the joints due to the biomechanics of the arm, and uncoupling the joint motions by whatever mechanism may contribute to reduce the effects of these interaction torques. On the other hand, neutralizing interaction torques by means of movement decomposition has the obvious drawback of worsening the overall smoothness of the movement patterns. Also the recent finding by Boose [11], that in cerebellar patients there is a deficient level of phasic muscle forces, can be attributed to a defective compensation of interaction torques, particularly in the higher speed part of the movements. Moreover, it has been found [19, 43, 60] that movement inaccuracies which can be attributed to partially unaccounted interaction torques are also present in healthy subjects performing sufficiently fast movements, even though its extent is much smaller than in the ataxic patients.

17.4.1 Abnormal feedforward control in ataxic patients

In a clinical study aimed at the quantitative assessment of the movements of patients with cerebellar ataxia, Sanguineti et al. [60] recorded aiming trajectories of cerebellar patients and compared them with normal subjects, as sketched in Figure 17.7. The statistical analysis of the data shows what is already apparent in qualitative inspection: the movements of the patients are less straight, smooth and symmetric than the controls. However their performance is consistent and in order to have some understanding of the specific missing element due to the cerebellar impairment, further analysis was focused on the initial and the final part of the movement. The former analysis is based on the assumption that the early phase of a movement is exclusively under feedforward control. In particular we considered the initial *aiming error,* defined as the angle between the target and the initial movement direction. If we consider the pattern of lateral disturbances depicted in Figure 17.4 we might expect significant aiming errors if the intrinsic dynamics is not compensated for by appropriate feedforward control. In the initial part of the movement, in which acceleration is high but velocity is low, the first inertial element of Equation 17.1 dominates the dynamical behavior: $\tau_{\text{actuator}}(t) \approx I(q)\ddot{q} + J^T(q)F_{ext}$. Therefore, the apparent inertia applied by the arm to an external load is given by the following relationships, in the case of non-redundant arms:

$$J^T(q)F_{load} \approx I(q)(\ddot{q}) \Rightarrow F_{load} \approx [J^T(q))^{-1}I(q)(J(q))^{-1}]\ddot{x} = I(x)(\ddot{x})$$

where x is the vector which identifies the position/orientation of the end-effector. The end-effector inertia, represented by the matrix $I(x)$, is not isotropic, as shown by the inertia ellipse of Figure 17.12 (top). It is worth noting that the principal axis is approximately aligned with the forearm: only in this direction (and the corresponding orthogonal direction) the force and acceleration vectors are collinear. This means that, if a force vector is generated in a given direction, the corresponding acceleration vector has a sideway component which tends to deviate the movement from its intended path, with the exception of the principal directions. Figure 17.12 (bottom) displays, for each experimental subject, the relationship between the aiming error and the movement direction *relative to the principal direction of the inertia ellipse*: we can see that the error tends to vanish in the principal direction and is characterized by sideways deviations in the other directions which are consistent with the directional characteristics of the inertia. Therefore we may conclude that the pattern of aiming errors can be attributed to a defective cerebellar feedforward controller. Such biomechanical explanation in terms of unaccounted interaction forces is also consistent with the fact that both patients and controls exhibit a similar pattern of aiming errors: the difference is that the feedforward compensation in the controls is more effective than in the cerebellar patients, although it is not perfect.

In cerebellar patients, the analysis of smoothness has shown that the final part of the movements, which is typically sensitive to feedback corrections, is not specifically affected and this suggest that the cerebellar impairment is primarily related to feedforward rather than feedback control. It is also worth mentioning that in another pathological syndrome with a relevant impairment of motor coordination (Huntington's chorea), the observed brief, small amplitude, and involuntary dancelike movements cannot be interpreted as a disruption of feedforward control but rather as a selective impairment of the feedback mechanisms, active in the terminal part of the movement [62]. Since the Huntington's chorea, a genetic disease, is known to be associated with a malfunction of the caudate nuclei which are part of the basal ganglia, one might conclude that internal models for anticipatory feedforward control are likely to be localized in the cerebellum whereas internal models for anticipatory feedback control are likely to involve the basal ganglia and the cortico-thalamic loop.

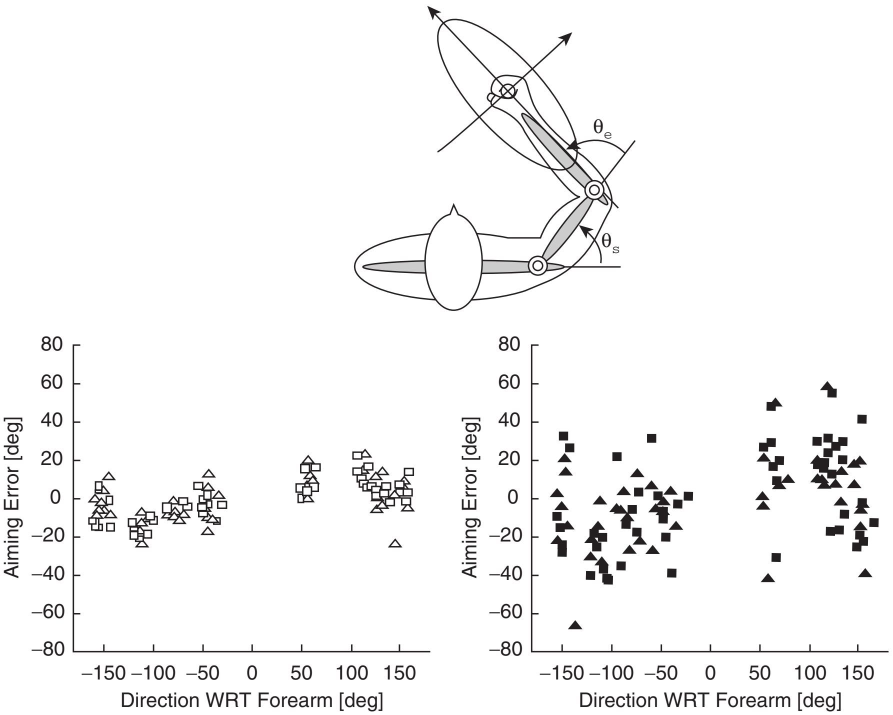

Figure 17.12

Top panel: orientation of the inertia ellipse. Bottom panel: Dependence of the aiming error on the direction of the movement with respect to the principal axis of the inertia ellipse, which is aligned along the forearm. Left graph: controls. Right graph: cerebellar patients.

17.5 Controlling unstable plants

There has been interest in recent years in the motor control of unstable tasks, particularly in two specific paradigms:

- Stabilisation of the human inverted pendulum in quiet standing [3, 38, 39, 48, 49, 69];
- Arm trajectory formation in an artificial, unstable force field [14].

In both cases the instability is associated with a potential field which has a maximum value around the reference state and feeds a diverging force field, thus having an antagonistic action with respect to the converging force field associated with muscle stiffness. In general, there are three possible mechanisms of stabilisation of an unstable load:

1. *Reflex Mechanism*, determined by a number of different sensory feedbacks; it is unfeasible in this context because it tends to worsen the instability as a consequence of the substantial delays in the control loop.
2. *Stiffness Mechanism,* related to the mechanical properties of muscles: it operates without delay, and can be modulated by means of coactivation, which is applied uniformly to all the muscles of a functional group, or by a more subtle re-distribution of activities in order to match the peculiar features of the task/load.
3. *Anticipatory feedforward/feedback mechanism,* which has an integrative central nature: it is based on two types of internal models: a) a model for multisensory fusion, thus compensating by means of *prediction* the transducton and propagation delays of sensory information, and b) a model for the generation of motor commands that anticipate the destabilising consequences of the load.

In any case, the divergent force field can be viewed as a *negative spring* and the rate of growth of the field, away from the equilibrium point, sets a *critical level of stiffness*. Above this level stiffness alone can stabilise the plant; below this level the effect of stiffness must be complemented by anticipatory control mechanisms.

17.5.1 Stabilisation of the standing posture: evidence of anticipatory compensation

In spite of its apparent simplicity, the nature of the control mechanisms that allow humans to stand up is still an object of controversy. Visual, vestibular, proprioceptive, tactile, and muscular factors clearly play a role in the stabilisation process and different authors have stressed one or the other. In particular, a model has been proposed by Winter et al. [69] that attributes to muscle stiffness alone the capability to solve the problem. According to this theory, the intervention of the CNS is limited to the selection of an appropriate tonus for the muscles of the ankle joint, in order to establish an ankle stiffness that stabilises an otherwise unstable mechanical system. Thus, in this view the stabilisation of quiet standing is a fundamentally passive process without any significant active or reactive component, except for the background setting of the stiffness parameters. The system equation of the human inverted pendulum (see Figure 17.13, left) is as follows:

$$I_p\ddot{\theta} = mgh\sin(\theta) + \tau_{ankle} \tag{17.5}$$

where θ is the sway angle, m and I_p are the mass and moment of inertia of the body, h is the distance of the COM (Center of Mass) from the ankle, g is the acceleration of gravity, and τ_{ankle} is the total ankle-torque. Under the assumption that the stabilising ankle torque is only determined by the ankle stiffness K_{ankle} and that sway angles are small, then from Equation 17.5 we can derive a critical value of the stiffness:

$$K_{critical} = mgh \tag{17.6}$$

In quiet standing the sway angle (and the horizontal shift of the COM $y =\approx h\theta$) oscillates with a frequency bandwidth below 1 Hz and such oscillations are associated with shifts (u) of the center of pressure (COP) on the support surface that have slightly bigger amplitude and substantially larger frequency band (Figure 17.13, right; [3]). It is easy to demonstrate [48] that u is proportional to the ankle torque and that u and y are linked by the following equation:

$$\ddot{y} = \frac{g}{h_e}(y-u) \tag{17.7}$$

where h_e is an *effective distance* which is quite close to the ankle-COM distance and takes into account the distribution of masses along the body axis. In this equation, which only expresses biomechanical relations and is independent of modalities of control, y is the controlled variable and u the control variable. It can also be described by saying that the COM-COP difference is proportional to the acceleration of the COM. Figure 17.13 (right) shows the relationship between the two curves. The COP, which mirrors the time-course of the ankle torque, has a higher frequency band but with good approximation the two curves are in phase (the peak of the cross-correlation occurs at a null delay). Moreover, it has been shown [21] that the emg activity of the ankle muscles anticipates the COM-COP pattern and thus it cannot be determined by segmental reflexes.

A simulation of the human inverted pendulum [50], activated by realistic muscle models compatible with experimentally measured muscle properties [30] has shown that the system is unstable. Moreover, direct estimates of the ankle stiffness have been carried out [40, 51] with different experimental methods. Both agree that the ankle stiffness during standing consistently is below the critical level mentioned above and thus is unable to stabilise the body by alone. Figure 17.14 illustrates the latter approach, based on a motorized platform mounted on top of a force platform. The motor generates small and rapid angular perturbations (1^o in less than 150 ms): the related COP shifts (see Figure 17.11, right) are proportional to the ankle stiffness.

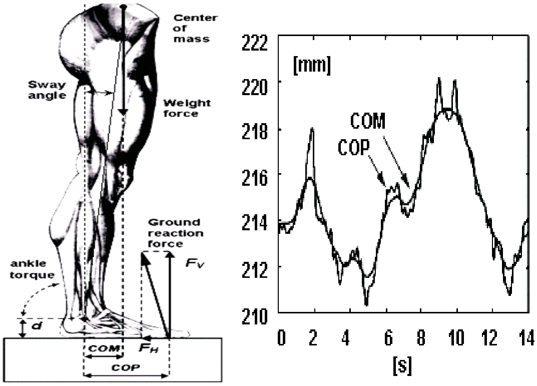

Figure 17.13

Left panel: scheme of the standing posture (COM: center of mass; COP: center of pressure); Right panel: COM and COP oscillations in the sagittal plane.

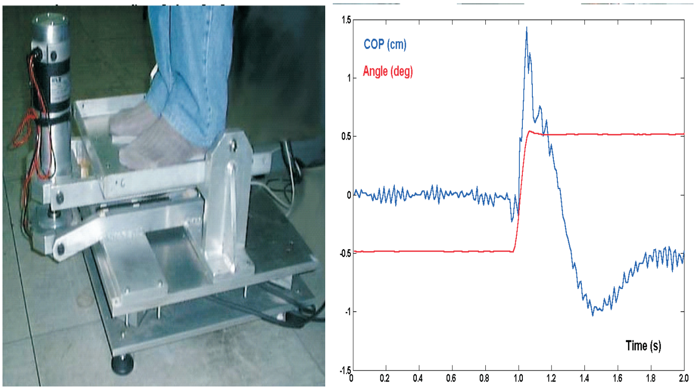

Figure 17.14

Left panel: motorized rotating platform mounted on top of a force platform. Right panel: ankle rotation $\theta(t)$ and COP displacement $y(t)$. During the quick rotation of the platform ($\triangle\theta = 1^o$, $\triangle T = 150$ms) the COM is virtually fixed and the shift of the COP ($\triangle y$) can be totally attributed to the ankle stiffness. Ankle stiffness can be estimated by means of the following equation: $mg\triangle y = K\triangle\theta$.

17.5.2 Arm trajectory in a divergent force field: evidence of stiffness modulation

In the manipulation of objects or tools people must control forces arising from interaction with the physical environment. Recent studies indicate that this compensation is achieved by learning internal models of the dynamics, that is, a neural representation of the relation between motor command and movement [34, 36, 44, 64, 71]. In these studies interactions with the physical environment are stable and a learning paradigm such as feedback error learning is capable to acquire a working internal model of inverse dynamics to be used in feedforward compensation. However, in many common tasks of everyday life (e.g., keeping a screwdriver in the slot of a screw) the interaction forces are divergent from the working point and thus represent an unstable load. In most situations, as in the case of the standing posture, we can rule out a reflexive feedback mechanism of stabilisation. Therefore the stabilisation can only be obtained by any or either of two mechanisms: 1) skillful modulation of the mechanical impedance of the arm, 2) anticipatory control based on internal models.

That the former mechanism can solve the stabilisation problem was demonstrated in a recent study [14] in which the human arm of the subject was immersed in an unstable force field generated by a robotic manipulator. The subjects performed forward/backward movements on the horizontal plane between two targets. The instability was the result of an artificial, divergent force field, perpendicular to the line of action of the nominal trajectory and proportional to the lateral displacement of the trajectory. In this way the robotic manipulator simulated a sort of inverted pendulum. The subjects were able to learn the task after a sufficiently long set of trials and the solution was achieved by a modulation and rotation of the stiffness ellipses in such a way to neutralize the divergent field. This is a more complicated strategy than the simple co-contraction of all the muscles that would increase the size of the ellipses without altering their orientation. Re-orienting the ellipses requires a subtle re-organisation of the pattern of muscle activation within a group of synergistic muscles and requires extensive training. The motivation for adopting this strategy is to optimise the mechanical impedance while keeping the metabolic cost at a minimum value.

However, it remains to be seen whether the stabilisation of unstable dynamics by means of optimal impedance matching is a general problem solving approach or is a rather special purpose paradigm limited to over-trained control patterns. The problem was addressed in a pilot study [51] in which a physical inverted pendulum could be grasped by the subjects at different heights (Figure 17.15) and the oscillations of the pendulum were measured by a pair of potentiometers, linked to the tip of the pendulum by means of a suitable articulation. The task of the subject was simply to keep the pendulum standing and the oscillations of the pendulum were recorded together with the EMG activity of two muscles of the arm: biceps and triceps (Figure 17.16). In this experimental setup it easy to demonstrate that the critical value of hand stiffness for the stabilisation of the unstable plant depends upon the point of

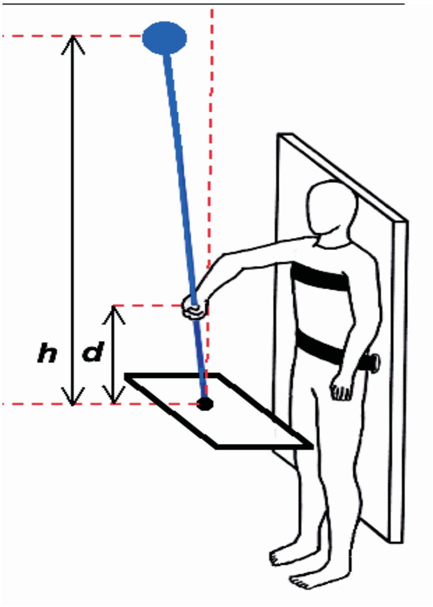

Figure 17.15
Manual stabilisation of an inverted pendulum; mass = 10 Kg; h = 1.8 m; d was varied between 0.3 and 0.8 m. In the equilibrium position the shoulder is abducted 90° and the elbow flexed 45°.

grasp (distance d) according to the following relationship:

$$K_{critical} > \frac{mgh}{d^2} \tag{17.8}$$

Therefore, it is important to compare the control patterns for values of d in which the critical hand stiffness is greater than physiological levels [67] and other values for which it is smaller. In the former, more demanding case the control system has no other choice than anticipatory control; in the latter case, on the other hand, the control system might use one solution or the other, at least in theory. In both cases, however, the results provide no evidence of stiffness modulation by means of coactivation; on the contrary, in all the cases there is a correlation between the muscle activity and the pendulum oscillation with a null delay and this is consistent with an anticipatory stabilisation mechanism.

17.5.3 Choosing between stiffness modulation and anticipatory compensation

The experimental evidence presented in the two previous sections is somehow contradictory. In one case, upright standing, physiological levels of muscle stiffness are

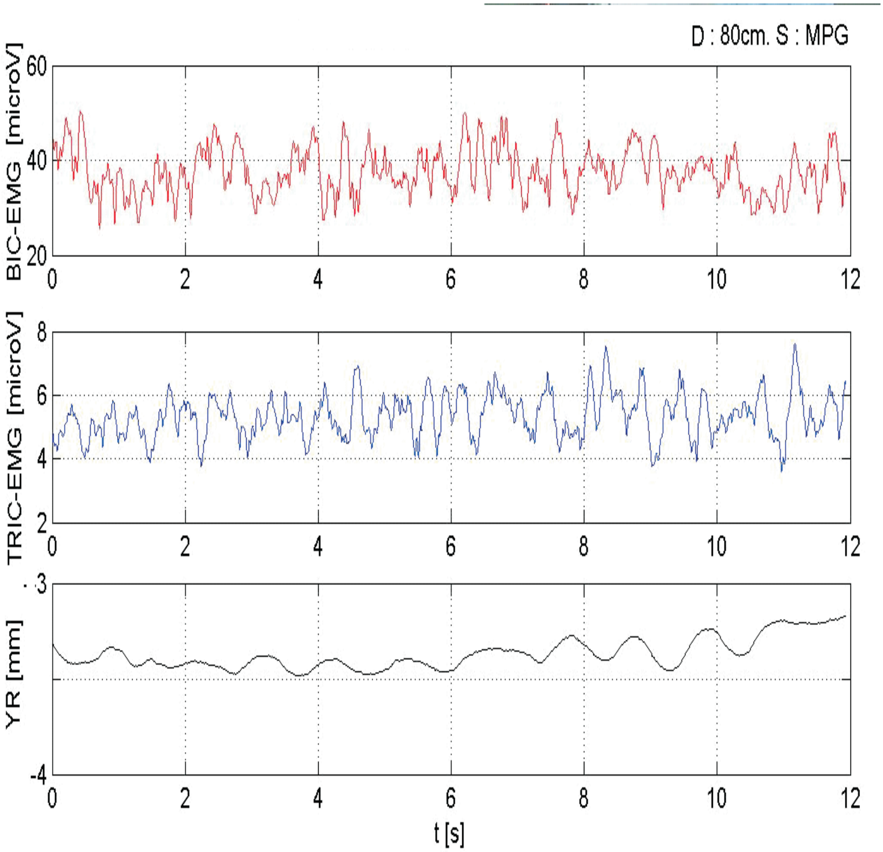

Figure 17.16

Oscillation of the pendulum in the direction perpendicular to the forearm (in the reference position the the elbow was 90° flexed): YR. TRIC and BIC show the rectified and integrated electromyographic activities of the biceps and the triceps muscles, respectively.

insufficient to stabilise the unstable plant and thus the only feasible solution is anticipatory control. In other cases, involving the upper extremity and with required levels of muscle stiffness which fall inside the physiological range of values, the solution adopted by the brain appears to depend upon the task: a) stiffness modulation in the case of the arm movements in a divergent force field, b) anticipatory compensation in the case of the manual stabilisation of an inverted pendulum. Which kind of circumstances might explain such difference of implementation?

Let us first consider, in general, the dynamics of *fall* which underlies all these paradigms. Equation 17.7 was derived in the specific case of the standing posture but, with a suitable abstraction, it is applicable to all the unstable loads characterized by an equilibrium state and a divergent force field: $\ddot{x} = \alpha x - u$, where α is a (positive) parameter depending on the structure of the unstable plant and u is the compensatory control variable. If we consider the transfer function corresponding to the equation above, we realise that the system has two real poles: $p = \pm\sqrt{\alpha}$. The positive pole is the source of instability and, in absence of an appropriate corrective action, determines the exponential *fall* from the equilibrium state ($x = 0$) with the following time constant: $T = 1/\sqrt{\alpha}$. In the case of the standing posture $\alpha = g/h_e$ and if we assume that $h_e = 1$ m we get $T = 320$ ms; in the case of the manually controlled inverted pendulum of Figure 17.15 $\alpha = g/h$, $h = 1.8$ m and thus $T = 430$ ms; in the case of the reaching movements in the divergent force field [14] $\alpha = K_{field}/M$, where K_{field} is the elastic constant of the field (about 200 N/m) and M is the apparent mass of the hand in the direction collinear with the field (about 1 Kg), thus giving $T \approx 70$ ms. Therefore we can say that in the two cases in which there is evidence of anticipatory compensation the characteristic time constant of the fall is much longer than in the case in which stiffness modulation is the adopted strategy. This result might be explained by considering that anticipatory compensation must have enough time to recover the intrinsic delays of the reafferent pathways (which come close to 100 ms) in order to generate functionally useful anticipatory commands. Therefore, anticipatory compensation in unstable tasks is only feasible if the critical time horizon before a catastrophic fall is significantly longer than 100 ms. As a matter of fact, if we consider a variety of stabilisation tasks such as standing on stilts, rope-walking, balancing a stick etc., it is easy to recognize the fact that the purpose of common tricks, like spreading out the arms or holding a long balancing rod is just to increase the natural *falling time*, thus giving time to the internal model to generate an appropriate stabilisation action.

17.5.4 Implementing anticipatory compensation

In the case of the upright posture, the underlying control process sketched in the previous section can be made evident by a particular type of analysis of the posturographic data (the time course of the COP coordinates provided by a force platform) which is based on sway density plots (SDP, [3]). A SDP is simply constructed by counting the number of consecutive samples of the COP trajectory that, for each time instant, fall inside a circle of given radius (typically 2.5 mm). It is worth noting that if such trajectories were generated by a random-walk process the sway density curve should be flat, but this is not the case, as shown in Figure 17.17; on the contrary, the SDP is characterized by a regular alternation of peaks and valleys: the peaks correspond to time intervals in which the ankle torque and the associated motor commands are relatively stable; the valleys correspond to time instants in which the ankle torque rapidly shifts from one stable value to another. Moreover, since the COM & COP oscillations are globally in phase, it follows that the sequence of peaks of the SDP must be phase-locked with the COM oscillation.

The statistical analysis of SDP data in normal subjects and in some pathological conditions [3] has shown that the peak-to-peak time interval is very stable ($T_p = 595 \pm 37$ ms) and is independent of the pathological condition. On the contrary, the amplitude of the peaks (which is equivalent to the duration of stationary motor commands) and the amplitude of the COP-shifts from one peak to the next one (which correspond to the amplitude of the anticipatory motor commands) are dependent upon the task (e.g., open eyes vs. closed eyes) and the nature of the pathological condition. To some extent, this type of organisation of the posturographic control action is functionally equivalent to the saccadic oculomotor system, alternating between two functional states:

1. acquisition and fixation of a *posturographic target*;

2. quick *saccadic jump* to the next target.

In the oculomotor system the target is typically visual, whereas in the postural stabilisation system the target must be placed on an *invisible* point, a little bit beyond the expected position of the COM. This requires a rather complex sensory processing task, certainly more complex that in the oculomotor case: its basic elements are a pair of internal models: (i) a multisensory data fusion process for estimating the position of the COM and the direction of *incipient fall* and (ii) a prediction capability for compensating the intrinsic sensory delays and producing an appropriate anticipatory postural saccade. Since the falling time constant is a biomechanical parameter and the motor controller must be time-locked with the partial falls in order to stop them, it is obvious that the T_p parameter must be stable if the subjects, even if affected by significant motor impairment as in the Parkinson's disease, are indeed able to stand without support.

On the contrary, pathological conditions are quite likely to affect the precision of the *posturographic saccades* due to a partial disruption of the internal feedforward/feedback models and thus the resulting sway will be larger and more irregular, as actually occurs.

In a similar experimental situation, Loram and Lakie [39] used an artificial task in which the subjects were required to activated their ankle musculature in order to balance a large inverted pendulum: they could easily learn such unstable task and quasi-regular sway was observed like that in quiet standing. It was observed that any resting equilibrium position of the pendulum is unstable and can be hold only temporarily; movements from a resting equilibrium position to another one could only be accomplished by a ballistic-like *throw and catch* pattern of torque which could be used for controlling both position and sway size.

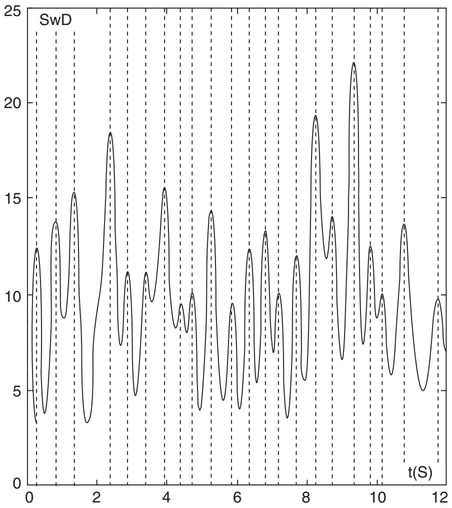

Figure 17.17

Sway density plot of a segment of postural sway (duration 12 s). The x-y components of the COP were sampled at 50 Hz. Sway density is defined as the number of consecutive samples that, for each time instant, fall inside a circle of 2.5 mm. Therefore a SWD value of 20 is equivalent to a time duration of 20/50=0.4 s. The plotted curve has been low-pass filtered at 12 Hz.

In our view, the ballistic, saccadic-like nature of postural stabilisation, which supplements the insufficient but synergistic action of ankle stiffness, is based on two essential elements: 1) an anticipatory (internal) forward model of the unstable plant which is functionally in the feedback loop and has the purpose of reconstructing the actual state and extrapolating it to the near future; 2) a sampled-data controller that fires ballistic commands capable to counteract the incipient falls. The main advantage of using such a sampled-data controller instead of a continuous one, as in the control of saccadic eye movements, is that it is less prone to instability and allows the generation of fast, stable patterns. The disadvantage is that it cannot guarantee asymptotical stability, in the Liapunov sense, with its characteristic point attractors: it can only assure a weaker condition of stability, with a persistent sway, characterised by a sort of chaotic attractor.

17.6 Motor learning paradigms

17.6.1 Learning paradigms in neural networks

At the core of the theories of neural network models is the attempt to capture general approaches for learning from experience procedures that are too complex to be expressed by means of explicit or symbolic models [2]. The mechanism of learning and memory has been an intriguing question after the establishment of the neuron theory at the turn of the 19th century [57] and the ensuing conjectures that memories are encoded at synaptic sites [26] as a consequence of a process of learning. In accordance with this prediction, synaptic plasticity was first discovered in the hippocampus and nowadays it is generally thought that LPT (long term potentiation) is the basis of cognitive learning and memory, although the specific mechanisms are still a matter of investigation.

Three main paradigms for training the parameters or *synaptic weights* of neural network models have been identified:

1. *Supervised learning,* in which a teacher or supervisor provides a detailed description of the desired response for any given stimulus and exploits the mismatch between the computed and the desired response or error signal for modifying the synaptic weights according to an iterative procedure. The mathematical technique typically used in this type of learning is known as *back propagation* and is based on a gradient-descent mechanism that attempts to minimize the average output error;

2. *Reinforcement learning*, which also assumes the presence of a *supervisor* or teacher but its intervention is only supposed to reward (or punish) the degree of success of a given control pattern, without any detailed input-output instruction. The underlying mathematical formulation is aimed at the maximization of the accumulated reward during the learning period;

3. *Unsupervised learning,* in which there is no teacher or explicit instruction and the network is only supposed to capture the statistical structure of the input stimuli in order to build a consistent but concise internal representation of the input. The typical learning strategy is called Hebbian, in recognition of the pioneering work of D.O. Hebb, and is based on a *competitive* or *self-organising* mechanism that uses the local correlation in the activity of adjacent neurons and aims at the maximization of the mutual information between stimuli and internal patterns.

17.6.2 Adaptive behaviour and motor learning

The neural machinery for learning and producing adaptive behaviours in vertebrates is sketched in Figure 17.18, which emphasizes the recurrent, non-hierarchical flow of information between the cerebral cortex, the basal ganglia/thalamus, and the cerebellum. A growing body of evidence has been accumulated in recent years that challenges the conventional view of segregated processing of perceptual, motor and cognitive information [61]. For example, it was usually considered that basal ganglia and cerebellum were specialized for motor control and different cortical areas were devoted to specific functionalities, with a clear separation of sensory, motor and cognitive areas. This is not anymore the conventional wisdom and the emerging picture is that the three main computational sites for adaptive behaviour are all concerned with processing sensorimotor patterns in a cognitive-sensitive way but are specialized as regards the learning paradigms and the types of representation:

1. The **cerebral cortex** appears to be characterized by a process of unsupervised learning that affects its basic computational modules (the micro-columns that are known to have massive recurrent connections). The function of these computations might be the representation of non-linear manifolds, such as a *body schema* in the posterior parietal cortex [47]. This view seems to be contradicted by the *fractured shape* of cortical maps [58], but the apparent contradiction may only be a side effect of the basic paradox faced by the cortical areas: how to fit higher dimensional manifolds (such as a proprioceptive body schema) onto a physically flat surface;

2. The **Cerebellum** is plausibly specialized in the kind of *supervised learning* exemplified by the feedback error learning model. Moreover, the cerebellar hardware (characterised by a large number of micro-zones, comprising mossy-fiber input, bundles of parallel fibers, output Purkinje cells, with teaching signals via climbing fibers) is well designed for the representation of time series, according to a sequence-in sequence-out type of operation [12];

3. The **Basal ganglia** are known to be involved in events of reinforcement learning that are required for the representation of goal-directed sequential behaviour [63].

17.6.3 A distributed computational architecture

Figure 17.19 summarizes some of the points outlined in the previous subsection. It must be emphasized that in spite of its apparent simplicity the underlying model is extremely complex from many points of view: (i) it is non-linear; (ii) it involves high-dimensional variables; (iii) it has a coupled dynamics, with internal and external processes; (iv) it is adaptive, with concurrent learning processes of different types. No simulation model of this complexity has been constructed so far, also because the mathematical tools for dominating its design are only partially available. However, there is a need for improving our current level of understanding in this direction because this is the only sensible way for interpreting the exponentially growing mass of data coming from new measurement techniques, such as advanced brain imaging. As a matter of fact, since the time of Marey better measurement techniques of movement analysis require better and better models of motor control, and vice versa.

In the scheme there are many interacting processes, such as *trajectory formation* and *feedback error learning*. The latter, in particular, is obviously characterized by a supervised learning paradigm and thus we may think that its main element (the *trainable feedforward model*) is implemented in the cerebellar circuitry. The learning signal, in this case, is the discrepancy between the desired trajectory (the motor intention) and the actual trajectory, determined by the combined body-environment dynamics and measured by different proprioceptive channels. In a sense, the brain acts as its own supervisor, setting its detailed goal and measuring the corresponding performance: for this reason it is possible to speak of a *self-supervised paradigm*. The underlying behavioural strategy is an active exploration of the *space of movements* also known as *babbling*, in which the brain attempts to carry out randomly selected movements that become the teachers of themselves.

On the other hand, the trajectory formation model cannot be analysed in the same manner. It requires different *maps* for representing task-relevant variables, such as the position of the objects/obstacles in the environment, the position of the body with respect to the environment, and the relative position of the body parts. Most of these variables are not directly detectable by means of specific sensory channels but require a complex process of *sensory fusion* and *dimensionality reduction*. This kind of processing is characteristic of associative cortical areas, such as the posterior parietal cortex which is supposed to hold maps of the body schema and the external world [54] as a result of the converging information from different sensory channels. The process of cortical map formation can be modelled by competitive Hebbian learning applied both to the thalamo-cortical and cortico-cortical connections: The former connections determine the receptive fields of the cortical units whereas the latter support the formation a kind of high-dimensional grid that matches the dimensionality of the represented sensorimotor manifold. In a cortical map model sensorimotor variables are represented by means of *population codes* which change over time as a result of the map dynamics. For example, a trajectory formation process can be implemented by means of a cortical map representation of the external space that can generate a time varying population code corresponding to the *desired hand trajectory*. Another map can transform the *desired hand trajectory* into the corresponding *desired joint trajectory*, thus implementing a transformation of coordinates from the *hand space* to the *joint space*. This kind of distributed architecture is necessary for integrating multisensory redundant information into a task-relevant, lower-dimensional representation of sensorimotor spaces. On top of this computational layer, that operates in a continuous way, there is a layer of reinforcement learning that operates mostly by trial and error through two main operating modules: an *actor* that selects a sequence of actions and a *critic* that evaluates the reward and influences the action selection of the next trial. The global coherence of such multiple internal processes of adaptation, learning and control is guaranteed by an effective mechanical interface with the environment which allows a bi-directional flow of energy and information.

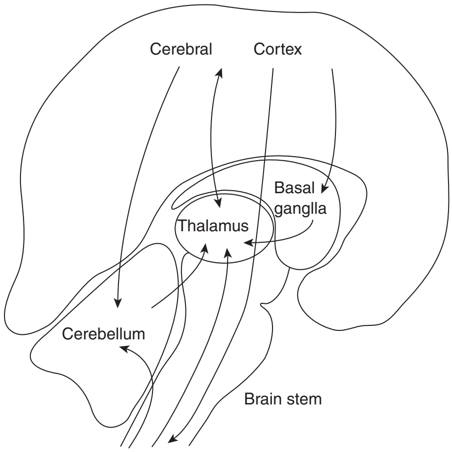

Figure 17.18
A schematic plot of the neural machinery for learning and producing adaptive behaviours in vertebrates.

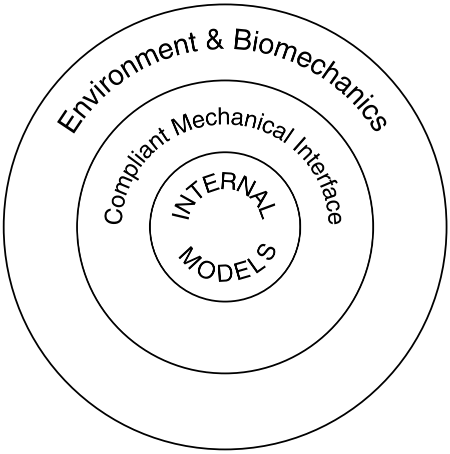

Figure 17.19
A distributed computational architecture (see Section 17.6.3 for details).

References

[1] Anokhin, P.K. (1974) *Biology and Neurophysiology of Conditioned Reflexes and Their Role in Adaptive Behaviour.* Pergamon Press.

[2] Arbib, M.A. (1995) *The Handbook of Brain Theory and Neural Networks.* MIT Press, Cambridge MA.

[3] Baratto, L., Morasso, P., Re, C., Spada, G. (2002) A new look at posturographic analysis in the clinical context. *Motor Control,* **6**, 248-273.

[4] Bastian, A.J., Martin, T.A., Keating, J.G., Thach, W.T. (1996). Cerebellar Ataxia: abnormal control of interaction torques across multiple joints. *J. Neurophysiology,* **76**, 492-509.

[5] Becker, W.J., Kunesch, E., Freund, H.J. (1990). Co-ordination of a multi-joint movement in normal humans and in patients with cerebellar dysfunction. *Canadian J. Neurological Sciences,* **17**, 264-274.

[6] Becker, W.J., Morrice, B.L., Clark, A.W., Lee, R.G. (1991). Multi-joint reaching movements and eye-hand tracking in cerebellar incoordination: investigation of a patient with complete loss of Purkinje cells. *Canadian J. Neurological Science,* **18,** 476-87.

[7] Bernstein, N.A. (1957) *The Coordination and Regulation of Movement.* Pergamon Press.

[8] Berthoz, A. (1997) *Le Sens du Mouvement*. Édition Odile Jacob, Paris, France.

[9] Bhushan, N., Shadmehr, R. (1999). Computational nature of human adaptive control during learning of reaching movements in force fields. *Biological Cybernetics,* **81**, 39-60.

[10] Bizzi, E., Hogan, N., Mussa-Ivaldi, F.A., Giszter, S.F. (1992) Does the nervous system use equilibrium-point control to guide single and multiple movements? *Behavioral and Brain Sciences,* **15**, 603-613.

[11] Boose, A., Dichgans, J., Topka, H. (1999). Deficits in phasic muscle force generation explain insufficient compensation for interaction torque in cerebellar patients. *Neuroscience Letters,* **261**, 53-56.

[12] Braitenberg, V., Heck, D., Sultan, F. (1997) The detection and generation of sequences as a key to cerebellar function: experiments and theory. *Behavioral and Brain Sciences,* **20**, 229-245.

[13] Brown, S.H., Hefter, H., Mertens, M., Freund, H.J. (1990). Disturbances in human arm trajectory due to mild cerebellar dysfunction. *J. Neurology, Neurosurgery, and Psychiatry,* **53**, 306-13.

[14] Burdet, E., Osu, R., Franklin, D.W., Milner, T.E., Kawato, M. (2001) The cen-

tral nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature,* **414**, 446-449.

[15] Damasio, A.R. (1994) *Descartes' Error. Emotion, Reason and the Human Brain.* Putnam Press, New York, U.S.

[16] Desmurget, M., Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences,* **4(11)**, 423-431.

[17] Feldman, A.G., Levin, M.F. (1995) The origin and use of positional frames of references in motor control. *Behavioral and Brain Sciences* **18**, 723-745.

[18] Fitts, P.M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *J. Experimental Psychology,* **47**, 381-391.

[19] Flanagan, J.R., Lolley, S. (2001). The inertial anisotropy of the arm is accurately predicted during movement planning. *J. Neuroscience,* **21(4)**, 1361-1369.

[20] Flash, T. and N. Hogan, N. (1985) The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neuroscience,* **7,** 1688-1703.

[21] Gatev, P., Thomas S, Kepple T, Hallet M. (1999) Feedforward ankle strategy of balance during quiet stance in adults. *J. Physiology (London)* **514**, 915-928.

[22] Georgopoulos, A., Schwartz, A., Kettner, R. (1986). Neuronal population coding of a movement direction. *Science,* **233,** 1416-1419.

[23] Gomi, H., Koike, Y., Kawato, M. (1992) Human hand stiffness during discrete point to point multijoint movements. *Proceedings IEEE EMBS*, 1628-1629.

[24] Gordon, J., Ghilardi, M.F., Cooper, S.E., Ghez, C. (1994). Accuracy of planar reaching movements. II. Systematic extent errors resulting from inertial anisotropy. *Experimental Brain Res.,* **99**, 112-130.

[25] Gribble, P.L., Ostry, D.J. (1999). Compensation for interaction torques during single and multijoint limb movement. *J. Neurophysiology,* **82(5)**, 2310-2326.

[26] Hebb, D.O (1949) *The Organization of Behavior.* J. Wiley, New York.

[27] Holmes, G. (1917). The symptoms of acute cerebellar injuries due to gunshot injuries. *Brain,* **40**, 461-535.

[28] Holmes, G. (1939). The cerebellum of man. (The Hughlings Jackson Memorial Lecture.) *Brain,* **62,** 1-30.

[29] Holst, E., von, Mittelstaedt, H. (1950) Das Reafferenzprinzip. *Wechselwirkungen zwischen Zentralnervensystem und Peripherie. Natur-wissenschaften,* **37,** 464-476.

[30] Hunter I.W., Kearney R.E. (1982) Dynamics of human ankle stiffness: variation with mean ankle torque. *J. Biomech.* **15,** 747-752.

[31] Hyman, R. (1953) Stimulus information as a determinant of reaction time. *J. Experimental Psychology,* **45**, 188-196.

[32] Ito, M. (1984). *The Cerebellum and Neural Control.* New York: Raven Press.

[33] Kawato, M., Gomi, H. (1992) A computational model of four regions of the cerebellum based on feedback error learning. *Biological Cybernetics,* **69**, 95-103.

[34] Kawato, M. (1999) Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.,* **9**, 718-727.

[35] Krakauer, J.W., Ghilardi, M.F., Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nature Neuroscience,* **2**, 1026-1031.

[36] Lackner, J.R., Dizio, P. (1994). Rapid adaptation to Coriolis force perturbations of arm trajectory. *J. Neurophysiology,* **72**, 299-313.

[37] Lacquaniti, F., Terzuolo, C., Viviani, P. (1983) The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica,* **54**, 115-130.

[38] Loram I.D., Kelly, S., Lakie, M. (2001) Human balancing of an inverted pendulum: is sway size controlled by ankle impedance? *J. Physiology,* **532**, 879-891.

[39] Loram I.D., Lakie, M. (2002a) Human balancing of an inverted pendulum: position control by small, ballistic-like, throw and catch movements. *J. Physiology,* **540**, 1111-24.

[40] Loram I.D., Lakie, M. (2002b) Direct measurement of human ankle stiffness during quiet standing: the intrinsic mechanical stiffness is insufficient for stability. *J. Physiology,* **545,** 1041-1053.

[41] Marey, E.J. (1894) *Le Mouvement.* Édition Masson, Paris, France.

[42] Marr, D. (1969). A theory of cerebellar cortex. *J. Physiology,* **202**, 437-470.

[43] Massaquoi, S., Hallett, M. (1996). Kinematics of initiating a two-joint arm movement in patients with cerebellar ataxia. *Canadian J. Neurological Science,* **23**, 3-14.

[44] McIntyre, J., Mussa-Ivaldi, F. A., Bizzi, E. (1996) The control of stable postures in the multijoint arm. *Exp. Brain Res.* **110,** 248-264.

[45] Morasso, P. (1981) Spatial control of arm movements. *Experimental Brain Research,* **42**, 223-227.

[46] Morasso, P., Mussa Ivaldi F.A. (1982) Trajectory formation and handwriting: a computational model. *Biological Cybernetics,* **45,** 131-142.

[47] Morasso, P., Sanguineti, V. (1997) *Self-Organization, Cortical Maps and Motor Control.* North Holland.

[48] Morasso, P., Schieppati M. (1999) Can muscle stiffness alone stabilize upright

standing? *J. Neurophysiology,* **83,** 1622-1626.

[49] Morasso, P., Sanguineti V. (2002) Ankle stiffness alone cannot stabilize upright standing. *J. Neurophysiology,* **88,** 2157-62.

[50] Morasso, P., Re, C., Sanguineti, V. (2002) A pilot study of the mechanisms of stabilization of the inverted pendulum. *Gait* & *Posture* **16**, suppl. 1, s209-210.

[51] Morasso, P., Casadio, M, Re, C., Sanguineti, V. (2003) *Motor Control Mechanisms in Unstable Tasks. XVII* IMEKO World Congress June 22 Dubrovnik, Croatia.

[52] Mussa-Ivaldi, F. A., Morasso, P., Zaccaria, R. (1988) Kinematic networks. A distributed model for representing and regularizing motor redundancy. *Biological Cybernetics,* **60**, 1-16.

[53] Muybridge, E. (1957) *The Human Figure in Motion.* Dover Press, New York, N.Y., U.S.

[54] Paillard, J. (1993) *Brain and Space.* Oxford University Press, Oxford, U.K.

[55] Piaget, J. (1963) *The Origin of Intelligence in Children.* Norton Press, New York, N.Y. U.S.

[56] Rack, P.M.H., Westbury, D.R. (1969) The effects of length and stimulus rate on tension in isometric cat soleus muscle. *J. Physiology,* **204**, 443-460.

[57] Ramn y Cajal, S. (1928) *Regeneration in the Vertebrate Central Nervous System.* Oxford University Press, Oxford, U.K.

[58] Rizzolatti, G., Luppino, G., Matelli, M. (1998) The organization of the cortical motor system: new concepts. *Electrencephalography and Clinical Neurophysiology,* **106,** 283-296.

[59] Sainburg, R.L., Ghez, C., Kalakanis, D. (1999). Intersegmental dynamics are controlled by sequential anticipatory, error correction, and postural mechanisms. *J. Neurophysiology,* **81**, 1045-1056.

[60] Sanguineti, V., Morasso, P., Baratto, L., Brichetto, G., Mancardi, G.L., Solaro, C. (2003). Cerebellar ataxia: Quantitative assessment and cybernetic interpretation. *Human Movement Science.* **22**, 189-195.

[61] Shepherd, G.M. (1998) *The Synaptic Organization of the Brain.* Oxford University Press, Oxford, U.K.

[62] Smith, M.A., Brandt, J., Shadmehr, R. (2000). Motor disorder in Huntington's disease begins as a dysfunction in error feedback control. *Nature,* **403**, 544-549.

[63] Sutton, R.S., Barto, A.G. (1998) *Reinforcement Learning.* MIT Press, Cambridge, MA, U.S.

[64] Thoroughman, K. A., Shadmehr R. (2000) Learning of action through adaptive combination of motor primitives. *Nature,* **407,** 742-747.

[65] Topka, H., Konczak, J., Dichgans, J. (1998a). Coordination of multi-joint arm movements in cerebellar ataxia: analysis of hand and angular kinematics. *Experimental Brain Research,* **119,** 483-492.

[66] Topka, H., Konczak, J., Schneider, K., Boose, A., Dichgans, J. (1998b). Multijoint arm movements in cerebellar ataxia: abnormal control of movement dynamics. *Experimental Brain Research,* **119,** 493-503.

[67] Tsuji, T., Morasso, P., Goto, K. and Ito, K. (1995) Human hand impedance characteristics during maintained posture. *Biological Cybernetics,* **72,** 475-485.

[68] Uno, Y., Kawato, M., Suzuki, R. (1989) Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. *Biological Cybernetics,* **61,** 89-101.

[69] Winter, D.A., Patla, A.E., Riedtyk, S., Ishac, M.G. (1998) Stiffness control of balance in quiet standing. *J. Neurophysiology,* **80**, 1211-1221.

[70] Winters, J.M. (1995) How detailed should muscle models be to understand multi-joint movement coordination. *Human Movement Science,* **14**, 401-442.

[71] Wolpert, D.M., Kawato, M. (1998) Internal models of the cerebellum. *Trends in Cogn. Sci.,* **2**, 338-347.

Chapter 18

Computational Models for Generic Cortical Microcircuits

Wolfgang Maass,[1] **Thomas Natschlaeger,**[1] **and Henry Markram** [2]

[1]*Institute for Theoretical Computer Science, Technische Universitaet Graz, A-8010 Graz, Austria, maass,tnatschl@igi.tu-graz.ac.at,;* [2]*Brain Mind Institute, EPFL, Lausanne, Switzerland, henry.markram@epfl.ch*

CONTENTS

18.1 Introduction

A key challenge for neural modeling is to explain how a continuous stream of multimodal input from a rapidly changing environment can be processed by neural microcircuits (columns, minicolumns, etc.) in the cerebral cortex whose anatomical and physiological structure is quite similar in many brain areas and species. However, a model that could explain the potentially universal computational capabilities of such microcircuits has been missing. We propose a computational model that does not require a task-dependent construction of neural circuits. Instead it is based on principles of high dimensional dynamical systems in combination with statistical learning theory, and can be implemented on generic evolved or found recurrent circuitry.

This new approach towards understanding neural computation on the micro-level also suggests new ways of modeling cognitive processing in larger neural systems. In particular it questions traditional ways of thinking about neural coding.

Common models for the organization of computations, such as for example Turing machines or attractor neural networks, are less suitable for modeling computations in cortical microcircuits, since these microcircuits carry out computations on continuous streams of inputs. Often there is no time to wait until a computation has converged, the results are needed instantly (*anytime computing*) or within a short time window (*real-time computing*). Furthermore biological data suggest that cortical microcircuits can support several real-time computational tasks in parallel, a hypothesis that is inconsistent with most modeling approaches. In addition the components of biological neural microcircuits, neurons and synapses, are highly diverse [5] and exhibit complex dynamical responses on several temporal scales. This makes them completely unsuitable as building blocks of computational models that require simple uniform components, such as virtually all models inspired by computer science, statistical physics, or artificial neural nets. Furthermore, neurons are connected by highly recurrent circuitry (*loops within loops*), which makes it particularly difficult to use such circuits for robust implementations of specific computational tasks. Finally, computations in most computational models are partitioned into discrete steps, each of which requires convergence to some stable internal state, whereas the dynamics of cortical microcircuits appears to be continuously changing. Hence, one needs a model for using continuous perturbations in inhomogeneous dynamical systems in order to carry out real-time computations on continuous input streams.

In this chapter we present a conceptual framework for the organization of computations in cortical microcircuits that is not only compatible with all these constraints, but actually requires these biologically realistic features of neural computation. Furthermore, like Turing machines, this conceptual approach is supported by theoretical results that prove the universality of the computational model, but for the biologically more relevant case of real-time computing on continuous input streams.

18.2 A conceptual framework for real-time neural computation

A computation is a process that assigns to inputs from some domain D certain outputs from some range R, thereby computing a function from D into R. Obviously any systematic discussion of computations requires a mathematical or conceptual framework, i.e., a computational model [24]. Perhaps the most well-known computational model is the Turing machine. In this case the domain D and range R are sets of finite character strings. This computational model is universal (for deterministic offline digital computation) in the sense that every deterministic digital function that is computable at all (according to a well-established mathematical definition, see

[20]) can be computed by some Turing machine. Before a Turing machine gives its output, it goes through a series of internal computation steps, the number of which depends on the specific input and the difficulty of the computational task (therefore it is called an *offline computation*). This may not be inadequate for modeling human reasoning about chess end games, but most cognitive tasks are closer related to real-time computations on continuous input streams, where online responses are needed within specific (typically very short) time windows, regardless of the complexity of the input. In this case the domain D and range R consist of time-varying functions $u(\cdot)$, $y(\cdot)$ (with analog inputs and outputs), rather than of static character strings. We propose here an alternative computational model that is more adequate for analyzing parallel real-time computations on analog input streams, such as those occurring in generic cognitive information processing tasks. Furthermore, we present a theoretical result which implies that within this framework the computational units of a powerful computational system can be quite arbitrary, provided that sufficiently diverse units are available (see the separation property and approximation property discussed in Section 18.4). It also is not necessary to *construct* circuits to achieve substantial computational power. Instead sufficiently large and complex *found* circuits tend to have already large computational power for real-time computing, provided that the reservoir from which their units are chosen is sufficiently diverse.

Our approach is based on the following observations. If one excites a sufficiently complex recurrent circuit (or other medium) with a continuous input stream $u(s)$, and looks at a later time $t > s$ at the current internal state $x(t)$ of the circuit, then $x(t)$ is likely to hold a substantial amount of information about recent inputs $u(s)$ (for the case of neural circuit models this was first demonstrated by [4]). We as human observers may not be able to understand the *code* by which this information about $u(s)$ is encoded in the current circuit state $x(t)$, but that is obviously not essential. Essential is whether a readout neuron that has to extract such information at time t for a specific task can accomplish this. But this amounts to a classical pattern recognition problem, since the temporal dynamics of the input stream $u(s)$ has been transformed by the recurrent circuit into a high dimensional spatial pattern $x(t)$. This pattern classification problem tends to be relatively easy to learn, even by a memoryless readout, provided the desired information is present in the circuit state $x(t)$. Furthermore, if the recurrent neural circuit is sufficiently large, it may support this learning task by acting like a kernel for support vector machines (see [25]), which presents a large number of nonlinear combinations of components of the preceding input stream to the readout. Such nonlinear projection of the original input stream $u(\cdot)$ into a high dimensional space tends to facilitate the extraction of information about this input stream at later times t, since it boosts the power of *linear* readouts for classification and regression tasks. Linear readouts are not only better models for the readout capabilities of a biological neuron than for example multi-layer-perceptrons, but their training is much easier and robust because it cannot get stuck in local minima of the error function (see [25] and [7]). These considerations suggest new hypotheses regarding the computational function of generic recurrent neural circuits: to serve as general-purpose temporal integrators, and simultaneously as kernels (i.e., nonlinear projections into a higher dimensional space) to facilitate subsequent linear readout of

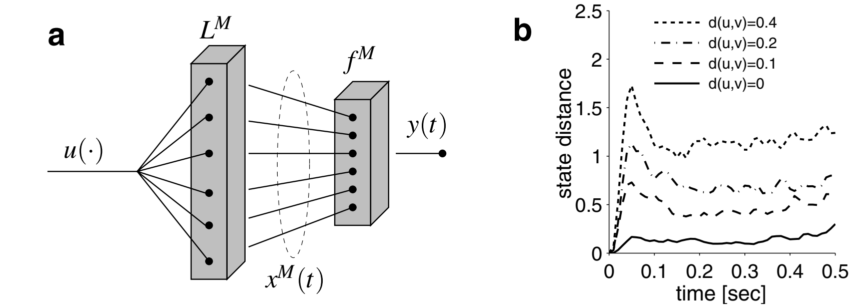

Figure 18.1

a) Structure of a Liquid State Machine (LSM). **b)** Separation property of a generic neural microcircuit. Plotted on the y-axis is the average value of $\|x_u^M(t) - x_v^M(t)\|$, where $\|\cdot\|$ denotes the Euclidean norm, and $x_u^M(t)$, $x_v^M(t)$ denote the liquid states at time t for Poisson spike trains u and v as inputs. $d(u,v)$ is defined as distance (L_2-norm) between low-pass filtered versions of u and v, see Section 18.4 for details.

information whenever it is needed. Note that in all experiments described in this article only the readouts were trained for specific tasks, whereas always a *fixed* recurrent circuit can be used for generating $x(t)$.

In order to analyze the potential capabilities of this approach, we introduce the abstract model of a Liquid State Machine (LSM), see Figure 18.1a. As the name indicates, this model has some weak resemblance to a finite state machine. But whereas the finite state set and the transition function of a finite state machine have to be custom designed for each particular computational task (since they contain its *program*), a liquid state machine might be viewed as a universal finite state machine whose *liquid* high dimensional analog state $x(t)$ changes continuously over time. Furthermore if this analog state $x(t)$ is sufficiently high dimensional and its dynamics is sufficiently complex, then the states and transition functions of many concrete finite state machines F are virtually contained in it. But fortunately it is in general not necessary to reconstruct F from the dynamics of an LSM, since the readout can be trained to recover from $x(t)$ directly the information contained in the corresponding state of a finite state machine F, even if the liquid state $x(t)$ is corrupted by some – not too large – amount of noise.

Formally, an LSM M consists of a filter L^M (i.e., a function that maps input streams $u(\cdot)$ onto streams $x(\cdot)$, where $x(t)$ may depend not just on $u(t)$, but in a quite arbitrary nonlinear fashion also on previous inputs $u(s)$; formally: $x(t) = (L^M u)(t)$), and a memoryless readout function f^M that maps at any time t the filter output $x(t)$ (i.e., the *liquid state*) into some target output $y(t)$ (only these readout functions are trained for specific tasks in the following). Altogether an LSM computes a filter that maps

$u(\cdot)$ onto $y(\cdot)$.[*]

A recurrently connected microcircuit could be viewed in a first approximation as an implementation of such general purpose filter L^M (for example some unbiased analog memory), from which different readout neurons extract and recombine diverse components of the information which was contained in the preceding input $u(\cdot)$. If a target output $y(t)$ assumes analog values, one can use instead of a single readout neuron a pool of readout neurons whose firing activity at time t represents the value $y(t)$ in space-rate-coding. In reality these readout neurons are not memoryless, but their membrane time constant is substantially shorter than the time range over which integration of information is required for most cognitive tasks. An example where the circuit input $u(\cdot)$ consists of 4 spike trains is indicated in Figure 18.2. The generic microcircuit model consisting of 270 neurons was drawn from the distribution discussed in Section 18.3. In this case 7 different linear readout neurons were trained to extract completely different types of information from the input stream $u(\cdot)$, which require different integration times stretching from 30 to 150 ms. The computations shown are for a novel input that did not occur during training, showing that each readout module has learned to execute its task for quite general circuit inputs. Since the readouts were modeled by linear neurons with a biologically realistic short time constant of just 30 ms for the integration of spikes, additional temporally integrated information had to be contained at any instance t in the current firing state $x(t)$ of the recurrent circuit (its *liquid state*), see Section 18.3 for details. Whereas the information extracted by some of the readouts can be described in terms of commonly discussed schemes for *neural codes*, it appears to be hopeless to capture the dynamics or the information content of the primary engine of the neural computation, the circuit state $x(t)$, in terms of such coding schemes. This view suggests that salient information may be encoded in the very high dimensional transient states of neural circuits in a fashion that looks like *noise* to the untrained observer, and that traditionally discussed *neural codes* might capture only specific aspects of the actually encoded information. Furthermore, the concept of *neural coding* suggests an agreement between *encoder* (the neural circuit) and *decoder* (a neural readout) which is not really needed, as long as the information is encoded in a way so that a generic neural readout can be trained to recover it.

[*] A closely related computational model was studied in [12].

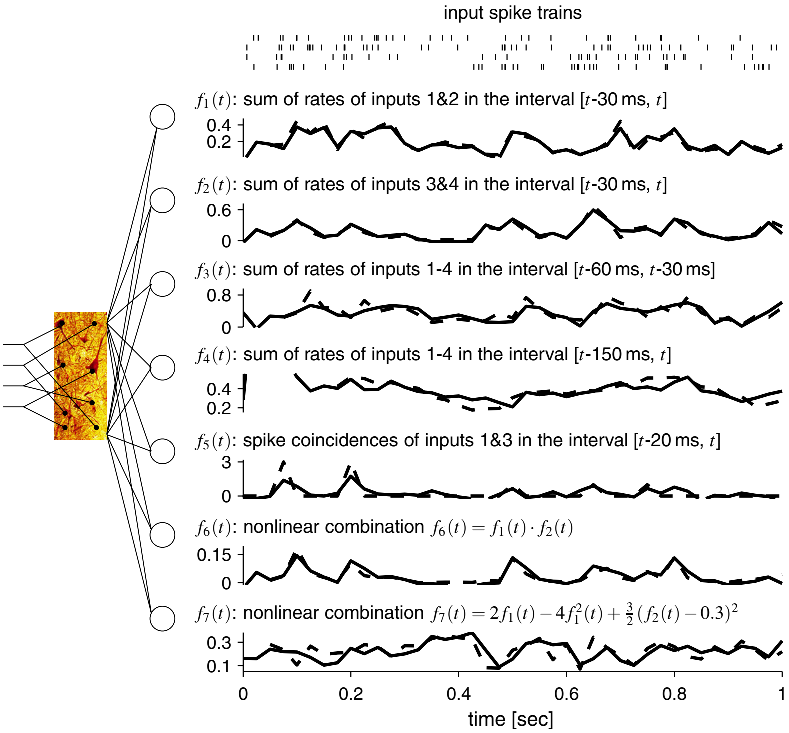

Figure 18.2

Multi-tasking in real-time. Input spike trains were randomly generated in such a way that at any time t the input contained no information about preceding input more than 30 ms ago. Firing rates $r(t)$ were randomly drawn from the uniform distribution over [0 Hz, 80 Hz] every 30 ms, and input spike trains 1 and 2 were generated for the present 30 ms time segment as independent Poisson spike trains with this firing rate $r(t)$. This process was repeated (with independent drawings of $r(t)$ and Poisson spike trains) for each 30 ms time segment. Spike trains 3 and 4 were generated in the same way, but with independent drawings of another firing rate $\tilde{r}(t)$ every 30 ms. The results shown in this figure are for test data, that were never before shown to the circuit. Below the 4 input spike trains the target (dashed curves) and actual outputs (solid curves) of 7 linear readout neurons are shown in real-time (on the same time axis). Targets were to output every 30 ms the actual firing rate (rates are normalized to a maximum rate of 80 Hz) of spike trains 1 and 2 during the preceding 30 ms (f_1), the firing rate of spike trains 3 and 4 (f_2), the sum of f_1 and f_2 in an earlier time interval [t-60 ms,t-30 ms] (f_3) and during the interval [t-150 ms, t] (f_4), spike coincidences between inputs 1 and 3 ($f_5(t)$ is defined as the number of spikes which are accompanied by a spike in the other spike train within 5 ms during the interval [t-20 ms, t]), a simple nonlinear combinations f_6 and a randomly chosen complex nonlinear combination f_7 of earlier described values. Since that all readouts were linear units, these nonlinear combinations are computed implicitly within the generic microcircuit model. Average correlation coefficients between targets and outputs for 200 test inputs of length 1s for f_1 to f_7 were $0.91, 0.92, 0.79, 0.75, 0.68, 0.87$, and 0.65.

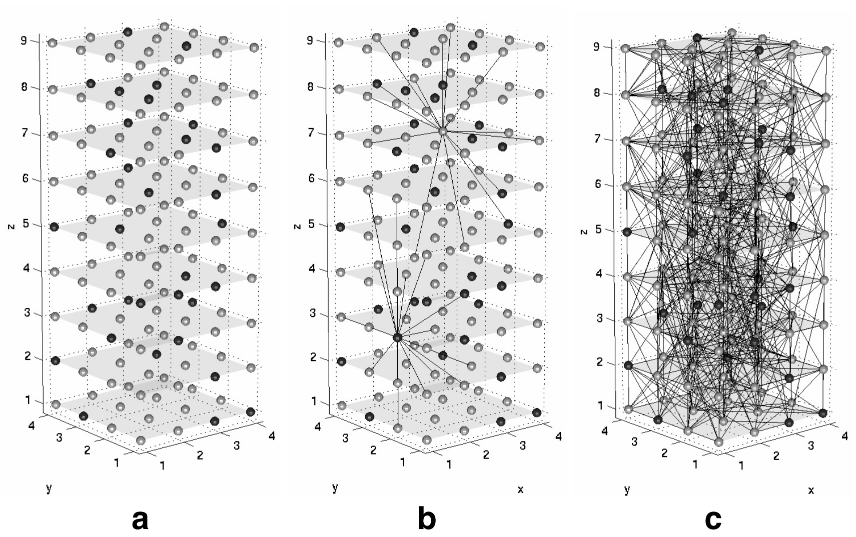

Figure 18.3

Construction of a generic neural microcircuit model, as used for all computer simulations discussed in this chapter (only the number of neurons varied). **a)** A given number of neurons is arranged on the nodes of a 3D grid. 20% of the neurons, marked in black, are randomly selected to be inhibitory. **b)** Randomly chosen postsynaptic targets are shown for two of the neurons. The underlying distribution favors local connections (see footnote on this page for details). **c)** Connectivity graph of a generic neural microcircuit model (for $\lambda = 2$, see footnote below).

Parameters of neurons and synapses were chosen to fit data from microcircuits in rat somatosensory cortex (based on [5], [15] and unpublished data from the Markram Lab). [†]

[†] *Neuron parameters*: membrane time constant 30 ms, absolute refractory period 3 ms (excitatory neurons), 2 ms (inhibitory neurons), threshold 15 mV (for a resting membrane potential assumed to be 0), reset voltage 13.5 mV, constant nonspecific background current $I_b = 13.5$ nA, input resistance 1 MΩ. *Connectivity structure*: The probability of a synaptic connection from neuron a to neuron b (as well as that of a synaptic connection from neuron b to neuron a) was defined as $C \cdot \exp(-D^2(a,b)/\lambda^2)$, where λ is a parameter which controls both the average number of connections and the average distance between neurons that are synaptically connected (we set $\lambda = 2$, see [16] for details). We assumed that the neurons were located on the integer points of a 3 dimensional grid in space, where $D(a,b)$ is the Euclidean distance between neurons a and b. Depending on whether a and b were excitatory (E) or inhibitory (I), the value of C was 0.3 (EE), 0.2 (EI), 0.4 (IE), 0.1 (II). In the case of a synaptic connection from a to b we modeled the synaptic dynamics according to the model proposed in [15], with the synaptic parameters U (use), D (time constant for depression), F (time constant for facilitation) randomly chosen from Gaussian distributions that were based on empirically found data for such connections. Depending on whether a and b were excitatory (E) or inhibitory (I), the mean values of these three parameters (with D,F expressed in seconds, s) were chosen to be .5, 1.1, .05 (EE), .05, .125, 1.2 (EI), .25, .7, .02 (IE), .32, .144, .06 (II). The SD of each parameter was chosen to be 50% of its mean. The mean of the scaling parameter A (in

18.3 The generic neural microcircuit model

We used a randomly connected circuit consisting of leaky integrate-and-fire (I&F) neurons, 20% of which were randomly chosen to be inhibitory, as generic neural microcircuit model. Best performance was achieved if the connection probability was higher for neurons with a shorter distance between their somata (see Figure 18.3). Random circuits were constructed with sparse, primarily local connectivity (see Figure 18.3), both to fit anatomical data and to avoid chaotic effects.

The *liquid state* $x(t)$ of the recurrent circuit consisting of n neurons was modeled by an n-dimensional vector consisting of the current firing activity of these n neurons. To reflect the membrane time constant of the readout neurons a low pass filter with a time constant of 30 ms was applied to the spike trains generated by the neurons in the recurrent microcircuit. The output of this low pass filter applied separately to each of the n neurons, defines the liquid state $x(t)$. Such low pass filtering of the n spike trains is necessary for the relatively small circuits that we simulate, since at many time points t no or just very few neurons in the circuit fire (see top of Figure 18.5). As readout units we used simply linear neurons, trained by linear regression (unless stated otherwise).

18.4 Towards a non-Turing theory for real-time neural computation

Whereas the famous results of Turing have shown that one can construct Turing machines that are universal for digital sequential offline computing, we propose here an alternative computational theory that is more adequate for parallel real-time computing on analog input streams. Furthermore we present a theoretical result which implies that within this framework the computational units of a powerful computational system can be quite arbitrary, provided that sufficiently diverse units are available (see the separation property and approximation property discussed below). It also is not necessary to *construct* circuits to achieve substantial computational power. Instead sufficiently large and complex *found* circuits (such as the generic

nA) was chosen to be 30 (EE), 60 (EI), -19 (IE), -19 (II). In the case of input synapses the parameter A had a value of 18 nA if projecting onto a excitatory neuron and 9 nA if projecting onto an inhibitory neuron. The SD of the A parameter was chosen to be 100% of its mean and was drawn from a gamma distribution. The postsynaptic current was modeled as an exponential decay $\exp(-t/\tau_s)$ with $\tau_s = 3$ ms ($\tau_s = 6$ ms) for excitatory (inhibitory) synapses. The transmission delays between liquid neurons were chosen uniformly to be 1.5 ms (EE), and 0.8 ms for the other connections. We have shown in [16] that without synaptic dynamics the computational power of these microcircuit models decays significantly. For each simulation, the initial conditions of each I&F neuron, i.e., the membrane voltage at time $t = 0$, were drawn randomly (uniform distribution) from the interval [13.5 mV, 15.0 mV].

circuit used as the main building block for Figure 18.2) tend to have already large computational power, provided that the reservoir from which their units are chosen is sufficiently diverse.

Consider a class $\mathscr{B}$ of basis filters B (that may for example consist of the components that are available for building filters L^M of LSMs). We say that this class $\mathscr{B}$ has the *point-wise separation property* if for any two input functions $u(\cdot), v(\cdot)$ with $u(s) \neq v(s)$ for some $s \leq t$ there exists some $B \in \mathscr{B}$ with $(Bu)(t) \neq (Bv)(t)$.[‡] There exist completely different classes $\mathscr{B}$ of filters that satisfy this point-wise separation property: $\mathscr{B}$ = {all delay lines}, $\mathscr{B}$ = {all linear filters}, and perhaps biologically more relevant $\mathscr{B}$ = {models for dynamic synapses} (see [17]).

The complementary requirement that is demanded from the class $\mathscr{F}$ of functions from which the readout maps f^M are to be picked is the well-known *universal approximation property*: for any continuous function h and any closed and bounded domain one can approximate h on this domain with any desired degree of precision by some $f \in \mathscr{F}$. Examples for such classes are $\mathscr{F}$ = {feedforward sigmoidal neural nets}, and according to [3] also $\mathscr{F}$ = {pools of spiking neurons with analog output in space rate coding}.

A rigorous mathematical theorem [16], states that for *any* class $\mathscr{B}$ of filters that satisfies the point-wise separation property and for *any* class $\mathscr{F}$ of functions that satisfies the universal approximation property one can approximate any given real-time computation on time-varying inputs with fading memory (and hence any biologically relevant real-time computation) by an LSM M whose filter L^M is composed of finitely many filters in $\mathscr{B}$, and whose readout map f^M is chosen from the class $\mathscr{F}$. This theoretical result supports the following pragmatic procedure: In order to implement a given real-time computation with fading memory it suffices to take a filter L whose dynamics is *sufficiently complex*, and train a *sufficiently flexible* readout to transform at any time t the current state $x(t) = (Lu)(t)$ into the target output $y(t)$. In principle a memoryless readout can do this, without knowledge of the current time t, provided that states $x(t)$ and $x(t')$ that require different outputs $y(t)$ and $y(t')$ are sufficiently distinct. We refer to [16] for details.

For physical implementations of LSMs it makes more sense to analyze instead of the theoretically relevant point-wise separation property the following quantitative separation property as a test for the computational capability of a filter L: How different are the liquid states $x_u(t) = (Lu)(t)$ and $x_v(t) = (Lv)(t)$ for two different input histories $u(\cdot), v(\cdot)$? This is evaluated in Figure 18.1b for the case where $u(\cdot), v(\cdot)$ are Poisson spike trains and L is a generic neural microcircuit model. It turns out that the difference between the liquid states scales roughly proportionally to the difference between the two input histories (thereby showing that the circuit dynamic is not chaotic). This appears to be desirable from the practical point of view since it implies that saliently different input histories can be distinguished more easily and in a more noise robust fashion by the readout. We propose to use such evaluation

[‡] Note that it is *not* required that there exists a single $B \in \mathscr{B}$ which achieves this separation for any two different input histories $u(\cdot)$, $v(\cdot)$.

of the separation capability of neural microcircuits as a new standard test for their computational capabilities.

18.5 A generic neural microcircuit on the computational test stand

The theoretical results sketched in the preceding section implies that there are no strong a priori limitations for the power of neural microcircuits for real-time computing with fading memory, provided they are sufficiently large and their components are sufficiently heterogeneous. In order to evaluate this somewhat surprising theoretical prediction, we tested it on several benchmark tasks.

18.5.1 Speech recognition

One well-studied computational benchmark task for which data had been made publicly available [8] is the speech recognition task considered in [9] and [10]. The dataset consists of 500 input files: the words *zero*, *one*, ..., *nine* are spoken by 5 different (female) speakers, 10 times by each speaker. The task was to construct a network of I&F neurons that could recognize each of the 10 spoken words w. Each of the 500 input files had been encoded in the form of 40 spike trains, with at most one spike per spike train[§] signaling onset, peak, or offset of activity in a particular frequency band (see top of Figure 18.4). A network was presented in [10] that could solve this task with an error[¶] of 0.15 for recognizing the pattern *one*. No better result had been achieved by any competing networks constructed during a widely publicized internet competition [9]. [‖] A particular achievement of this network (resulting from the smoothly and linearly decaying firing activity of the 800 pools of neurons) is that it is robust with regard to linear time-warping of the input spike pattern.

We tested our generic neural microcircuit model on the same task (in fact on exactly the same 500 input files). A randomly chosen subset of 300 input files was used

[§] The network constructed in [10] required that each spike train contained at most one spike.

[¶] The error (or *recognition score*) S for a particular word w was defined in [10] by $S = \frac{N_{fp}}{N_{cp}} + \frac{N_{fn}}{N_{cn}}$, where N_{fp} (N_{cp}) is the number of false (correct) positives and N_{fn} and N_{cn} are the numbers of false and correct negatives. We use the same definition of error to facilitate comparison of results. The recognition scores of the network constructed in [10] and of competing networks of other researchers can be found at [8]. For the competition the networks were allowed to be constructed especially for their task, but only one single pattern for each word could be used for setting the synaptic weights.

[‖] The network constructed in [10] transformed the 40 input spike trains into linearly decaying input currents from 800 pools, each consisting of a *large set of closely similar unsynchronized neurons* [10]. Each of the 800 currents was delivered to a separate pair of neurons consisting of an excitatory α*-neuron* and an inhibitory β*-neuron*. To accomplish the particular recognition task some of the synapses between α-neurons (β-neurons) are set to have equal weights, the others are set to zero.

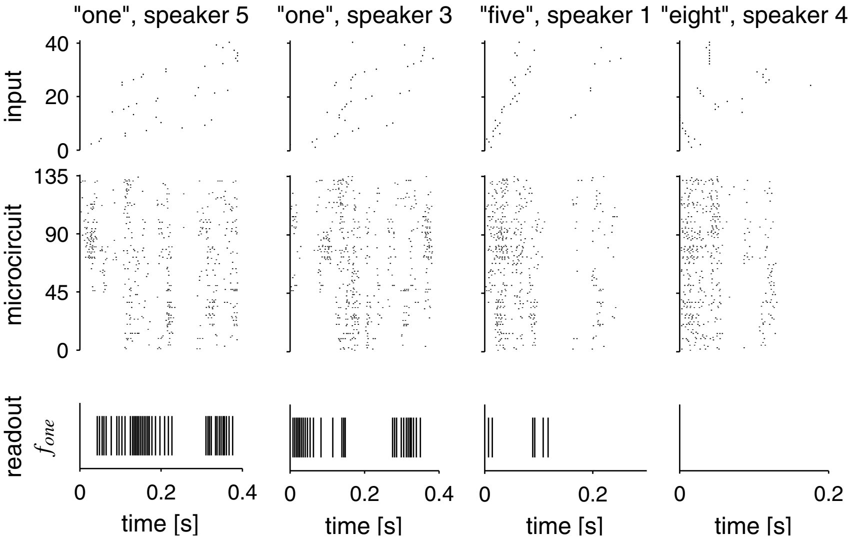

Figure 18.4

Application of our generic neural microcircuit model to the speech recognition from [10]. Top row: input spike patterns. Second row: spiking response of the 135 I&F neurons in the neural microcircuit model. Third row: output of an I&F neuron that was trained to fire as soon as possible when the word ***one*** was spoken, and as little as possible else. Although the ***liquid state*** presented to this readout neuron changes continuously, the readout neuron has learnt to view most of them as equivalent if they arise while the word ***one*** is spoken (see [16] for more material on such equivalence classes defined by readout neurons).

for training, the other 200 for testing. The generic neural microcircuit model was drawn from the distribution described in Section 18.3, hence from the same distribution as the circuit drawn for the completely different tasks discussed in Figure 18.2, with randomly connected I&F neurons located on the integer points of a $15 \times 3 \times 3$ column. The synaptic weights of 10 readout neurons f_w which received inputs from the 135 I&F neurons in the circuit were optimized (like for SVMs with linear kernels) to fire whenever the input encoded the spoken word w. Hence the whole circuit consisted of 145 I&F neurons, less than $1/30^{th}$ of the size of the network constructed in [10] for the same task.[**] Nevertheless the average error achieved after training by these randomly generated generic microcircuit models was 0.14 (measured in the same way, for the same word *one*), hence slightly better than that of the 30 times larger network custom designed for this task. The score given is the average for 50 randomly drawn generic microcircuit models. It is about the same as the error

[**] If one assumes that each of the 800 large pools of neurons in that network would consist of just 5 neurons, it contains together with the α and β-neurons 5600 neurons.

achieved by any of the networks constructed in [10] and the associated international competition.

The comparison of the two different approaches also provides a nice illustration of the difference between offline computing and real-time computing. Whereas the network of [10] implements an algorithm that needs a few hundred ms of processing time between the end of the input pattern and the answer to the classification task (450 ms in the example of Figure 2 in [10]), the readout neurons from the generic neural microcircuit were trained to provide their answer (through firing or non-firing) immediately when the input pattern ended.

We also compared the noise robustness of the generic neural microcircuit models with that of [10], which had been constructed to facilitate robustness with regard to linear time warping of the input pattern. Since no benchmark input date were available to calculate this noise robustness we constructed such data by creating as templates 10 patterns consisting each of 40 randomly drawn Poisson spike trains at 4 Hz over 0.5 s. Noisy variations of these templates were created by first multiplying their time scale with a randomly drawn factor from $[1/3,3]$) (thereby allowing for a 9 fold time warp), and subsequently dislocating each spike by an amount drawn independently from a Gaussian distribution with mean 0 and SD 32 ms. These spike patterns were given as inputs to the same generic neural microcircuit models consisting of 135 I&F neurons as discussed before. Ten readout neurons were trained (with 1000 randomly drawn training examples) to recognize which of the 10 templates had been used to generate a particular input (analogously as for the word recognition task). On 500 novel test examples (drawn from same distributions) they achieved an error of 0.09 (average performance of 30 randomly generated microcircuit models). The best one of 30 randomly generated circuits achieved an error of just 0.005. Furthermore it turned out that the generic microcircuit can just as well be trained to be robust with regard to *nonlinear* time warp of a spatio-temporal pattern (it is not known whether this could also be achieved by a constructed circuit). For the case of nonlinear (sinusoidal) time warp[††] an average (50 microcircuits) error of 0.2 is achieved (error of the best circuit: 0.02). This demonstrates that it is not really necessary to build noise robustness explicitly into the circuit. A randomly generated microcircuit model can easily be trained to have at least the same noise robustness as a circuit especially constructed to achieve that. In fact, it can also be trained to be robust with regard to types of noise that are very hard to handle with constructed circuits.

This test had implicitly demonstrated another point. Whereas the network of [10] was only able to classify spike patterns consisting of at most one spike per spike train, a generic neural microcircuit model can classify spike patterns without that restriction. It can for example also classify the original version of the speech data encoded into onsets, peaks, and offsets in various frequency bands, before all except

[††] A spike at time t was transformed into a spike at time $t' = g(t) := B + K \cdot (t + 1/(2\pi f) \cdot \sin(2\pi f t + \varphi))$ with $f = 2\,\text{Hz}$, K randomly drawn from [0.5,2], φ randomly drawn from $[0, 2\pi]$ and B chosen such that $g(0) = 0$.

the first events of each kind were artificially removed to fit the requirements of the network from [10].

We have also tested the generic neural microcircuit model on a much harder speech recognition task: to recognize the spoken word not only in real-time right after the word has been spoken, but even earlier when the word is still spoken.[‡‡] More precisely, each of the 10 readout neurons is trained to recognize the spoken word at any multiple of 20 ms during the 500 ms interval while the word is still spoken (*anytime speech recognition*). Obviously the network from [10] is not capable to do this. But also the trivial generic microcircuit model where the input spike trains are injected directly into the readout neurons perform poorly on this anytime speech classification task: it has an error score of 3.4 (computed as described in footnote 5, but every 20 ms). In contrast a generic neural microcircuit model consisting of 135 neurons it achieves a score of 1.4 for this anytime speech classification task (see Figure 18.4 for a sample result).

One is easily led to believe that readout neurons from a neural microcircuit can give a stable output only if the firing activity (or more abstractly: the state of the dynamical system defined by this microcircuit) has reached an attractor. But this line of reasoning underestimates the capabilities of a neural readout from high dimensional dynamical systems: even if the neural readout is just modeled by a perceptron, it can easily be trained to recognize completely different states of the dynamical system as being equivalent, and to give the same response. Indeed, Figure 18.4 showed already that the firing activity of readout neuron can become quite independent from the dynamics of the microcircuit, even though the microcircuit neurons are their only source of input. To examine the underlying mechanism for the possibility of relatively independent readout response, we re-examined the readout from Figure 18.4. Whereas the firing activity within the circuit was highly dynamic, the firing activity of the readout neurons was quite stable after training. The stability of the readout response does not simply come about because the spiking activity in the circuit becomes rather stable, thereby causing quite similar liquid states (see Figure 18.5). It also does not come about because the readout only samples a few unusual liquid neurons as shown by the distribution of synaptic weights onto a sample readout neuron (bottom of Figure 18.5). Since the synaptic weights do not change after learning, this indicates that the readout neurons have learned to define a notion of equivalence for dynamic states of the microcircuit. Indeed, equivalence classes are an inevitable consequence of collapsing the high dimensional space of microcircuit states into a single dimension, but what is surprising is that the equivalence classes are meaningful in terms of the task, allowing invariant and appropriately scaled readout responses and therefore real-time computation on *novel inputs*. Furthermore, while the input may contain salient information that is constant for a particular readout element, it may

[‡‡] It turns out that the speech classification task from [10] is in a sense too easy for a generic neural microcircuit. If one injects the input spike trains that encode the spoken word directly into the 10 readout neurons (each of which is trained to recognize one of the 10 spoken words) one also gets a classification score that is almost as good as that of the network from [10]. Therefore we consider in the following the much harder task of *anytime speech recognition*.

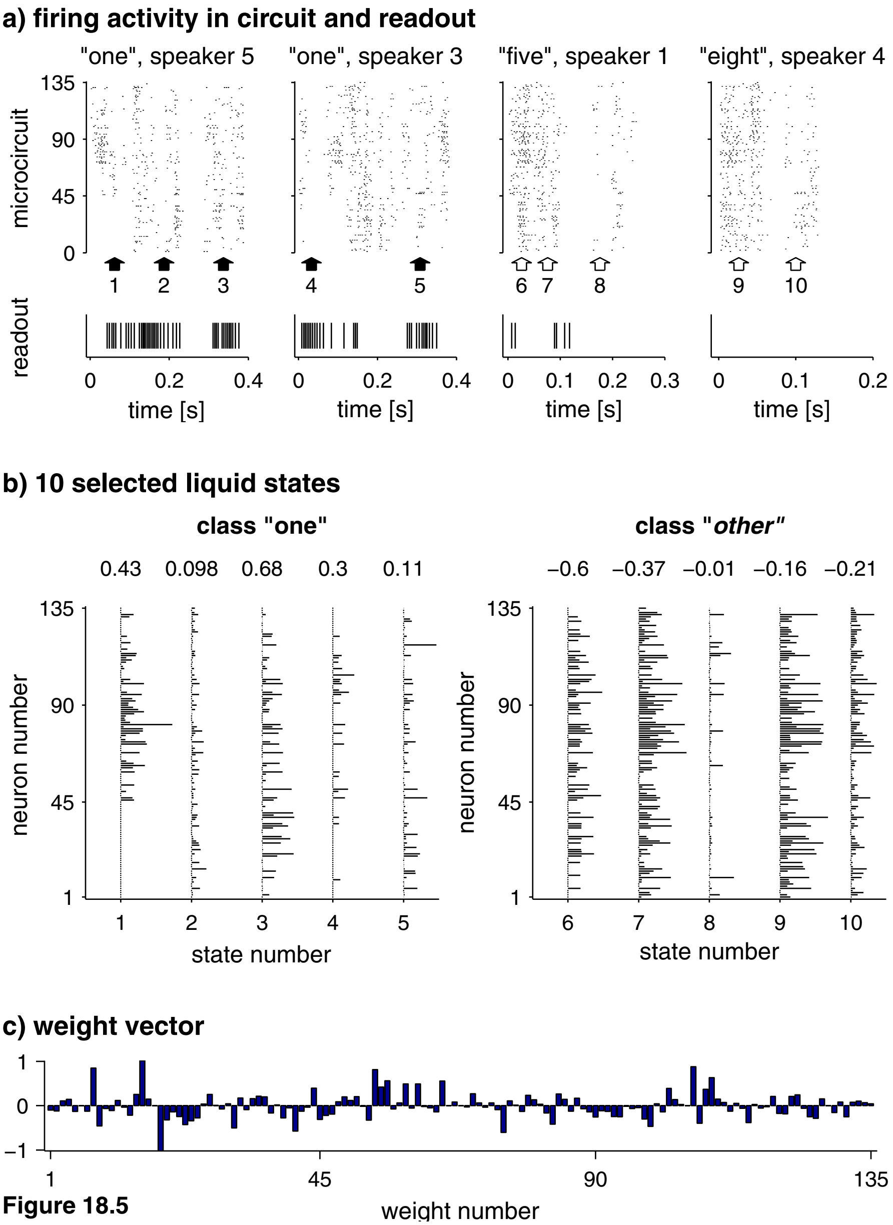

Figure 18.5

Readout defined equivalence classes of liquid states. **a)** The firing activity of the microcircuit for the speech recognition task from Figure 18.4 is reexamined. **b)** The liquid state $x(t)$ is plotted for 10 randomly chosen time points t (see arrowheads in panel a). The target output of the readout neuron is 1 for the first 5 liquid states, and 0 for the other 5 liquid states. Nevertheless the 5 liquid states in each of the 2 equivalence classes are highly diverse. But by multiplying these liquid state vectors with the weight vector of the linear readout (see panel c), the weighted sums yields the values shown above the liquid state vectors, which are separated by the threshold 0 of the readout (and by the firing threshold of the corresponding leaky integrate-and-fire neuron whose output spike trains are shown in panel a). **c)** The weight vector of the linear readout.

not be for another (see for example Figure 18.2), indicating that equivalence classes and dynamic stability exist purely from the perspective of the readout elements.

18.5.2 Predicting movements and solving the aperture problem

This section reports results of joint work with Robert Legenstein [13], [18]. The general setup of this simulated vision task is illustrated in Figure 18.6. Moving objects, a ball or a bar, are presented to an 8 x 8 array of sensors (panel a). The time course of activations of 8 randomly selected sensors, resulting from a typical movement of the ball, is shown in panel b. Corresponding functions of time, but for all 64 sensors, are projected as 64 dimensional input by a topographic map into a generic recurrent circuit of spiking neurons. This circuit with randomly chosen sparse connections had been chosen in the same way as the circuits for the preceding tasks, except that it was somewhat larger (768 neurons) to accommodate the 64 input channels. A 16 x 16 x 3 neuronal sheet was divided into 64 2 x 2 x 3 input regions. Each sensor injected input into 60 % randomly chosen neurons in the associated input region. Together they formed a topographic map for the 8 x 8 array of sensors.

The resulting firing activity of all 768 integrate-and-fire neurons in the recurrent circuit is shown in panel c. Panel d of Figure 18.6 shows the target output for 8 of the 102 readout pools. These 8 readout pools have the task to predict the output that the 8 sensors shown in panel b will produce 50 ms later. Hence their target output (dashed line) is formally the same function as shown in panel b, but shifted by 50 ms to the left. The solid lines in panel d show the actual output of the corresponding readout pools after unsupervised learning. Thus in each row of panel d the difference between the dashed and predicted line is the prediction error of the corresponding readout pool.

The diversity of object movements that are presented to the 64 sensors is indicated in Figure 18.7. Any straight line that crosses the marked horizontal or vertical line segments of length 4 in the middle of the 8 x 8 field may occur as trajectory for the center of an object. Training and test examples are drawn randomly from this – in principle infinite – set of trajectories, each with a movement speed that was drawn independently from a uniform distribution over the interval from 30 to 50 units per second (unit = side length of a unit square). Shown in Figure 18.7 are 20 trajectories that were randomly drawn from this distribution. Any such movement is carried out by an independently drawn object type (ball or bar), where bars were assumed to be oriented vertically to their direction of movement. Besides movements on straight lines one could train the same circuit just as well for predicting nonlinear movements, since nothing in the circuit was specialized for predicting linear movements.

Thirty-six readout pools were trained to predict for any such object movement the sensor activations of the 6 x 6 sensors in the interior of the 8 x 8 array 25 ms into the future. Further 36 readout pools were independently trained to predict their activation 50 ms into the future, showing that the prediction span can basically be chosen arbitrarily. At any time t (sampled every 25 ms from 0 to 400 ms) one uses for each of the 72 readout pools that predict sensory input ΔT into the future the actual activation of the corresponding sensor at time $t+\Delta T$ as target value (*correction*) for

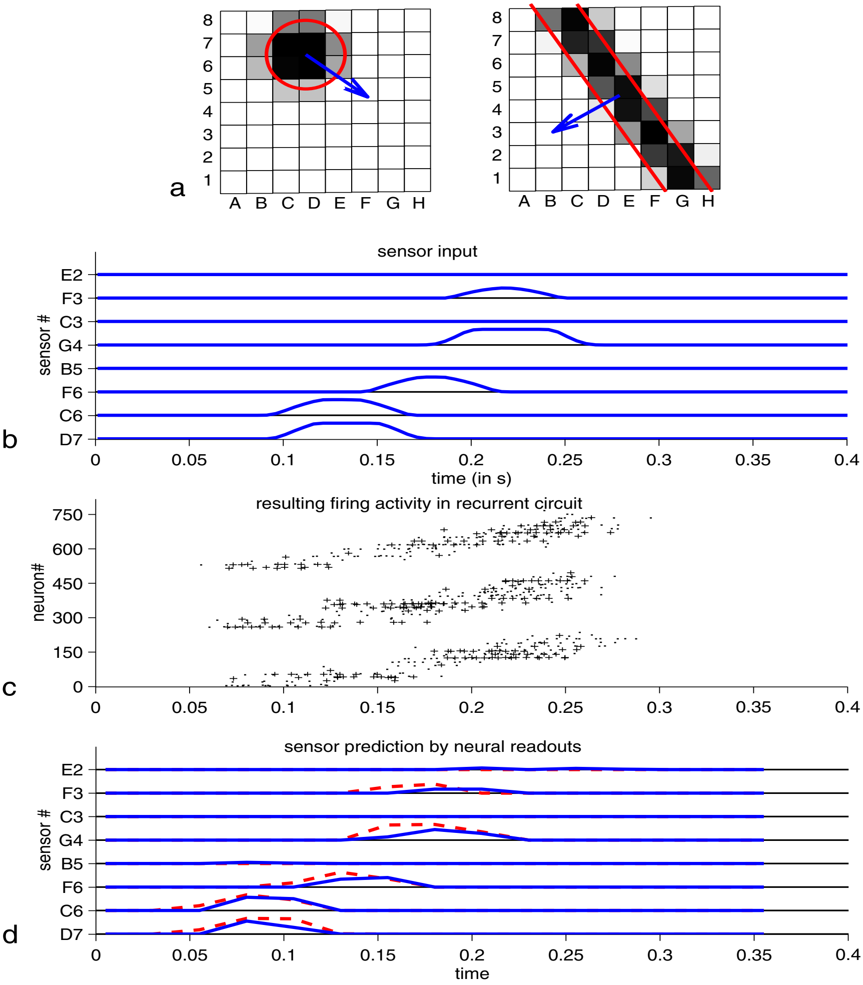

Figure 18.6

The prediction task. **a)** Typical movements of objects over a 8 x 8 sensor field. **b)** Time course of activation of 8 randomly selected sensors caused by the movement of the ball indicated on the l.h.s. of panel a. **c)** Resulting firing times of 768 integrate-and-fire neurons in the recurrent circuit of integrate-and-fire neurons (firing of inhibitory neurons marked by +). The neurons in the 16 x 16 x 3 array were numbered layer by layer. Hence the 3 clusters in the spike raster result from concurrent activity in the 3 layers of the circuit. **d)** Prediction targets (dashed lines) and actual predictions (solid lines) for the 8 sensors from panel b. (Predictions were sampled every 25 ms, solid curves result from linear interpolation.)

the learning rule. The 72 readout pools for short-term movement prediction were trained by 1500 randomly drawn examples of object movements. More precisely,

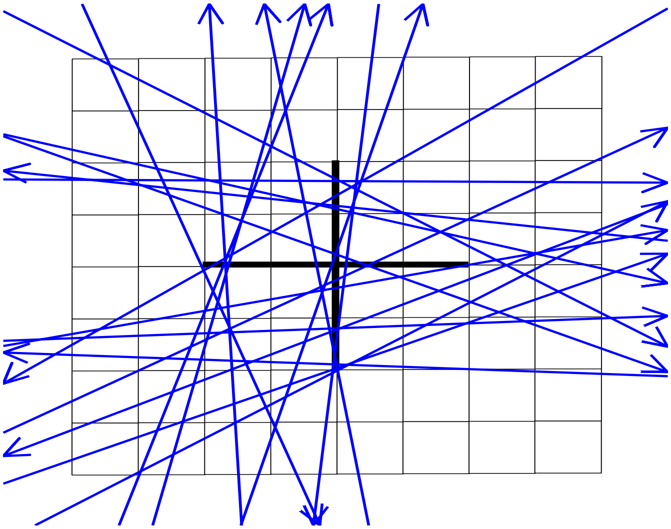

Figure 18.7

20 typical trajectories of movements of the center of an object (ball or bar).

they were trained to predict future sensor activation at any time (sampled every 25 ms) during the 400 ms time interval while the object (ball or bar) moved over the sensory field, each with another trajectory and speed.

Among the predictions of the 72 different readout pools on 300 novel test inputs there were for the 25 ms prediction 8.5% false alarms (sensory activity erroneously predicted) and 4.8% missed predictions of subsequent sensor activity. For those cases where a readout pool correctly predicted that a sensor will become active, the mean of the time period of its activation was predicted with an average error of 10.1 ms. For the 50 ms prediction there were for 300 novel test inputs 16.5% false alarms, 4.6% missed predictions of sensory activations, and an average 14.5 ms error in the prediction of the mean of the time interval of sensory activity.

One should keep in mind that movement prediction is actually a computationally quite difficult task, especially for a moving ball, since it requires context-dependent integration of information from past inputs over time and space. This computational problem is often referred to as the *aperture problem*: from the perspective of a single sensor that is currently partially activated because the moving ball is covering part of its associated unit square (i.e., its *receptive field*) it is impossible to predict whether this sensor will become more or less activated at the next movement (see [19]). In order to decide that question, one has to know whether the center of the ball is moving towards its receptive field, or is just passing it tangentially. To predict whether a sensor that is currently not activated will be activated 25 or 50 ms later, poses an even more difficult problem that requires not only information about the direction of the moving object, but also about its speed and shape. Since there exists in this experiment no preprocessor that extracts these features, which are vital for a successful prediction, each readout pool that carries out prediction for a particular sensor has to extract on its own these relevant pieces of information from the raw and unfiltered information about the recent history of sensor activities, which are still *reverberating* in the recurrent circuit.

Twenty-eight further readout pools were trained in a similar unsupervised manner (with 1000 training examples) to predict *where* the moving object is going to leave the sensor field. More precisely, they predict which of the 28 sensors on the perimeter are going to be activated by more than 50% when the moving object leaves the 8 x 8 sensor field. This requires a prediction for a context-dependent time span into the future that varies by 66% between instances of the task, due to the varying speeds of moving objects. We arranged that this prediction had to be made while the object crossed the central region of the 8 x 8 field, hence at a time when the current position of the moving object provided hardly any information about the location where it will leave the field, since all movements go through the mid area of the field. Therefore the tasks of these 28 readout neurons require the computation of the direction of movement of the object, and hence a computationally difficult disambiguation of the current sensory input. We refer to the discussion of the disambiguation problem of sequence prediction in [1] and [14]. The former article discusses difficulties of disambiguation of movement prediction that arise already if one has just pointwise objects moving at a fixed speed, and just 2 of their possible trajectories cross. Obviously the disambiguation problem is substantially more severe in our case, since a virtually unlimited number of trajectories (see Figure 18.7) of different extended objects, moving at different speeds, crosses in the mid area of the sensor field. The disambiguation is provided in our case simply through the *context* established inside the recurrent circuit through the traces (or *reverberations*) left by preceding sensor activations. Figure 18.6 shows in panel a a typical current position of the moving ball, as well as the sensors on the perimeter that are going to be active by $\geq 50\%$ when the object will finally leave the sensory field. In panel b the predictions of the corresponding 28 readout neurons (at the time when the object crosses the mid-area of the sensory field) is also indicated (striped squares). The prediction performance of these 28 readout neurons was evaluated as follows. We considered for each movement the line from that point on the opposite part of the perimeter, where the center of the ball had entered the sensory field, to the midpoint of the group of those sensors on the perimeter that were activated when the ball left the sensory field (dashed line). We compared this line with the line that started at the same point, but went to the midpoint of those sensor positions which were predicted by the 28 readout neurons to be activated when the ball left the sensory field (solid line). The angle between these two lines had an average value of 4.9 degrees for 100 randomly drawn novel test movements of the ball (each with an independently drawn trajectory and speed). Another readout pool was independently trained in a supervised manner to classify the moving object (ball or bar). It had an error of 0% on 300 test examples of moving objects. The other readout pool that was trained in a supervised manner to estimate the speed of the moving bars and balls, which ranged from 30 to 50 units per second, made an average error of 1.48 units per second on 300 test

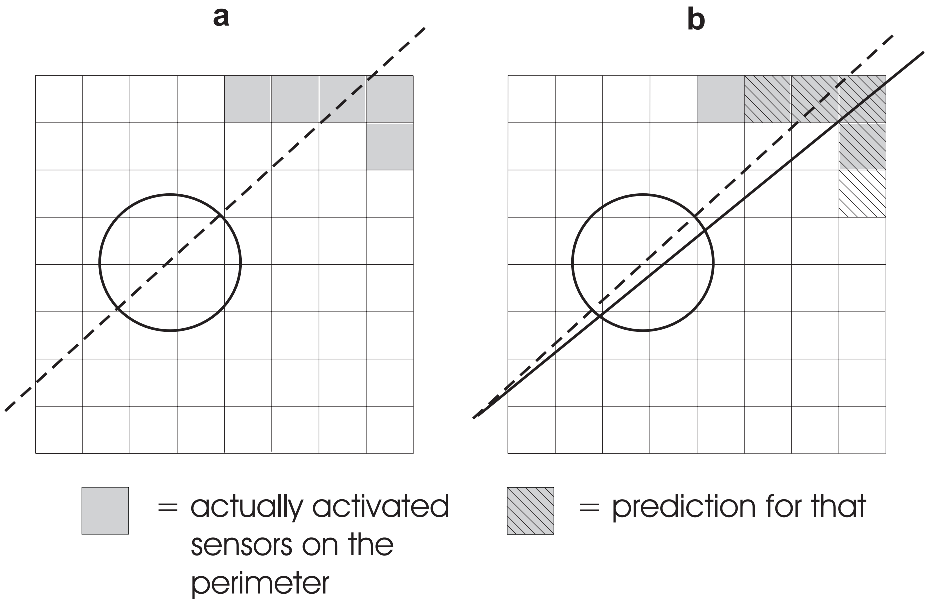

Figure 18.8

Computation of movement direction. Dashed line is the trajectory of a moving ball. Sensors on the perimeter that will be activated by $\geq 50\%$ when the moving ball leaves the sensor field are marked in panel a. Sensors marked by stripes in panel b indicate a typical prediction of sensors on the perimeter that are going to be activated by $\geq 50\%$, when the ball will leave the sensor field (time span into the future varies for this prediction between 100 and 150 ms, depending on the speed and angle of the object movement). Solid line in panel b represents the estimated direction of ball movement resulting from this prediction (its right end point is the average of sensors positions on the perimeter that are predicted to become $\geq 50\%$ activated). The angle between the dashed and solid line (average value 4.9 for test movements) is the error of this particular computation of movement direction by the simulated neural circuit.

examples. This shows that the same recurrent circuit that provides the input for the movement prediction can be used simultaneously by a basically unlimited number of other readouts, that are trained to extract completely different information about the visual input. We refer to [13] and [18] for details. Currently similar methods are applied to real-time processing of input from infra-red sensors of a mobile robot.

18.6 Temporal integration and kernel function of neural microcircuit models

In Section 18.2 we have proposed that the computational role of generic cortical microcircuits can be understood in terms of two complementary computational perspectives:

1. temporal integration of continuously entering information (*analog fading memory*)
2. creation of diverse nonlinear combinations of components of such information to enhance the capabilities of linear readouts to extract nonlinear combinations of pieces of information for diverse tasks (*kernel function*).

The results reported in the preceding section have demonstrated implicitly that both of these computational functions are supported by generic cortical microcircuit models, since all of the benchmark problems that we discussed require temporal integration of information. Furthermore, for all of these computational tasks it sufficed to train *linear* readouts to transform liquid states into target outputs (although the target function to be computed was highly nonlinear in the inputs). In this section we provide a more quantitative analysis of these two complementary computational functions.

18.6.1 Temporal integration in neural microcircuit models

In order to evaluate the temporal integration capability we considered two input distributions. These input distributions were chosen so that the mutual information (and hence also the correlation) between different segments of the input stream have value 0. Hence all temporal integration of information from earlier input segments has to be carried out by the microcircuit circuit model, since the input itself does not provide any clues about its past. We first consider a distribution of input spike trains where every 30 ms a new firing rate $r(t)$ is chosen from the uniform distribution over the interval from 0 to 80 Hz (first row in Figure 18.9). Then the spikes in each of the concurrent input spike trains are generated during each 30 ms segment by a Poisson distribution with this current rate $r(t)$ (second row in Figure 18.9). Due to random fluctuation the actual sum of firing rates $r_{measured}(t)$ (plotted as dashed line in the first row) represented by these 4 input spike trains varies around the intended firing rate $r(t)$. $r_{measured}(t)$ is calculated as the average firing frequency in the interval $[t-30\,\mathrm{ms}, t]$. Third row of Figure 18.9 shows that the autocorrelation of both $r(t)$ and $r_{measured}(t)$ vanishes after 30 ms.

Various readout neurons, that all received the same input from the microcircuit model, had been trained by linear regression to output at various times t (more precisely: at all multiples of 30 ms) the value of $r_{measured}(t)$, $r_{measured}(t-30ms)$, $r_{measured}(t-60ms)$, $r_{measured}(t-90ms)$, etc. Figure 18.10a shows (on test data not used for training) the correlation coefficients achieved between the target value and actual output value for 8 such readouts, for the case of two generic microcircuit models consisting of 135 and 900 neurons (both with the same distance-dependent connection probability with $\lambda = 2$ discussed in Section 18.3). Figure 18.10b shows that dynamic synapses are essential for this analog memory capability of the circuit, since the *memory curve* drops significantly faster if one uses instead static (*linear*) synapses for connections within the microcircuit model. Figure 18.10c shows that the intermediate *hidden* neurons in the microcircuit model are also essential for this task, since without them the memory performance also drops significantly.

It should be noted that these memory curves not only depend on the microcircuit model, but also on the diversity of input spike patterns that may have occurred in the input before, at, and after that time segment in the past from which one recalls information. Hence the recall of firing rates is particularly difficult, since there exists a huge number of diverse spike patterns that all represent the same firing rate. If one restricts the diversity of input patterns that may occur, substantially longer memory recall becomes possible, even with a fairly small circuit. In order to demonstrate this point 8 randomly generated Poisson spike trains over 250 ms, or equivalently 2 Poisson spike trains over 1000 ms partitioned into 4 segments each (see top of Figure 18.11), were chosen as template patterns. Then spike trains over 1000 ms were generated by choosing for each 250 ms segment one of the two templates for this segment, and by jittering each spike in the templates (more precisely: each spike was moved by an amount drawn from a Gaussian distribution with mean 0 and a SD that we refer to as *jitter*, see bottom of Figure 18.11). A typical spike train generated in this way is shown in the middle of Figure 18.11. Because of the noisy dislocation of spikes it was impossible to recognize a specific template from a single interspike interval (and there were no spatial cues contained in this single channel input). Instead, a pattern formed by several interspike intervals had to be recognized and classified retrospectively. The performance of 4 readout neurons trained by linear regression to recall the number of the template from which the corresponding input segment had been generated is plotted in Figure 18.12 (thin line).

For comparison the memory curve for the recall of firing rates for the same temporal segments (i.e., for inputs generated as for Figure 18.10, but with each randomly chosen target firing rate $r(t)$ held constant for 250 instead of 30 ms) is plotted as thin line in Figure 18.12, both for the same generic microcircuit model consisting of 135 neurons. Figure 18.12 shows that information about spike patterns of past inputs decays in a generic neural microcircuit model slower than information about firing rates of past inputs, even if just two possible firing rates may occur. One possible ex-

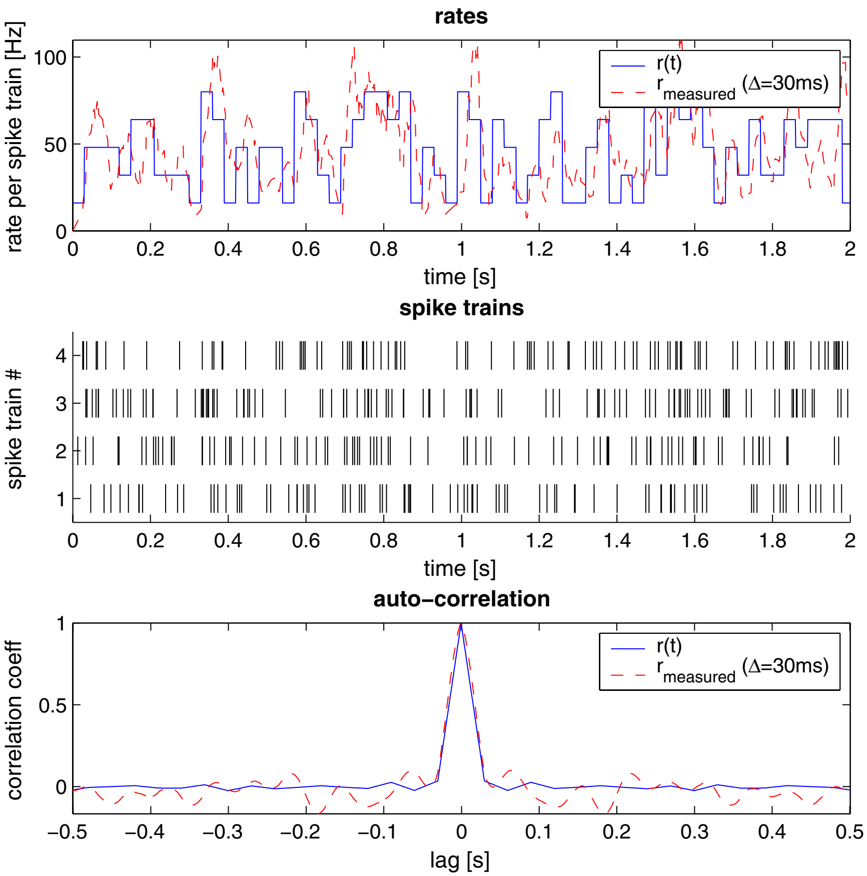

Figure 18.9

Input distribution used to determine the ***memory curves*** for firing rates. Input spike trains (second row) are generated as Poisson spike trains with a randomly drawn rate $r(t)$. The rate $r(t)$ is chosen every 30 ms from the uniform distribution over the interval from 0 to 80 Hz (first row, sold line). Due to random fluctuation the actual sum of firing rates $r_{measured}(t)$ (first row, dashed line) represented by these 4 input spike trains varies around the intended firing rate $r(t)$. $r_{measured}(t)$ is calculated as the average firing frequency in the interval $[t-30\,\mathrm{ms}, t]$. The third row shows that the autocorrelation of both $r(t)$ and $r_{measured}(t)$ vanishes after 30 ms.

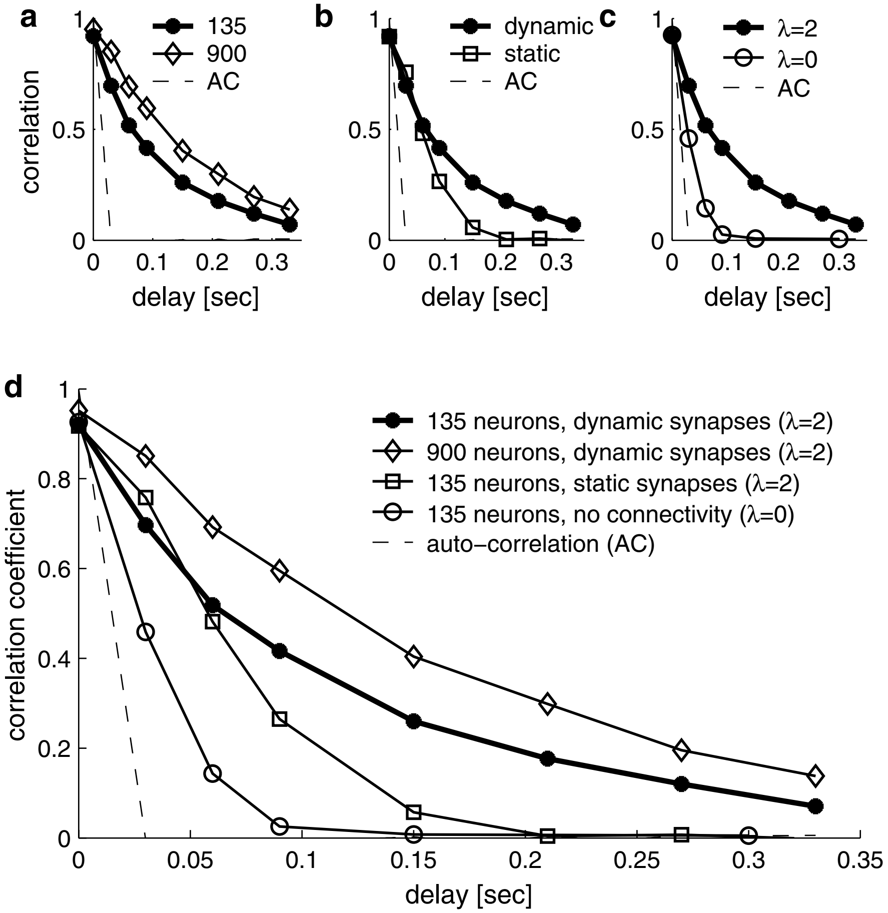

Figure 18.10

Memory curves for firing rates in a generic neural microcircuit model. **a)** Performance improves with circuit size. **b)** Dynamic synapses are essential for longer recall. **c)** Hidden neurons in a recurrent circuit improve recall performance (in the control case $\lambda = 0$ the readout receives synaptic input only from those neurons in the circuit into which one of the input spike trains is injected, hence no ***hidden*** neurons are involved). **d)** All curves from panels a to c in one diagram for better comparison. In each panel the bold solid line is for a generic neural microcircuit model (discussed in Section 18.3) consisting of 135 neurons with sparse local connectivity ($\lambda = 2$) employing dynamic synapses. All readouts were linear, trained by linear regression with 500 combinations of input spike trains (1000 in the case of the liquid with 900 neurons) of length 2 s to produce every 30 ms the desired output.

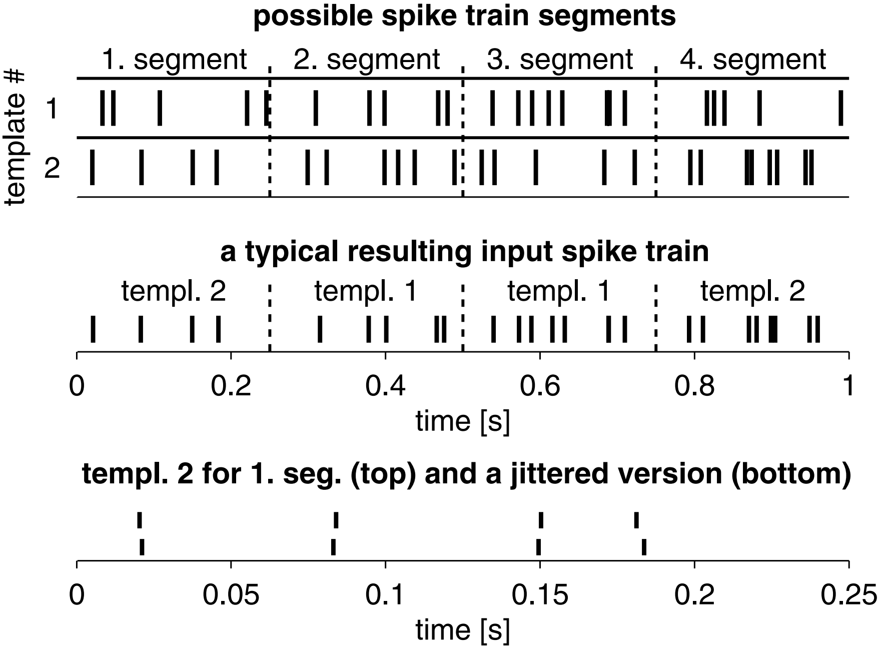

Figure 18.11

Evaluating the fading memory of a generic neural microcircuit for spike patterns. In this classification task all spike trains are of length 1000 ms and consist of 4 segments of length 250 ms each. For each segment 2 templates were generated randomly (Poisson spike train with a frequency of 20 Hz); see upper traces. The actual input spike trains of length 1000 ms used for training and testing were generated by choosing for each segment one of the two associated templates, and then generating a noisy version by moving each spike by an amount drawn from a Gaussian distribution with mean 0 and a SD that we refer to as ***jitter*** (see lower trace for a visualization of the jitter with an SD of 4 ms). The task is to output with 4 different readouts at time $t = 1000$ ms for each of the preceding 4 input segments the number of the template from which the corresponding segment of the input was generated.

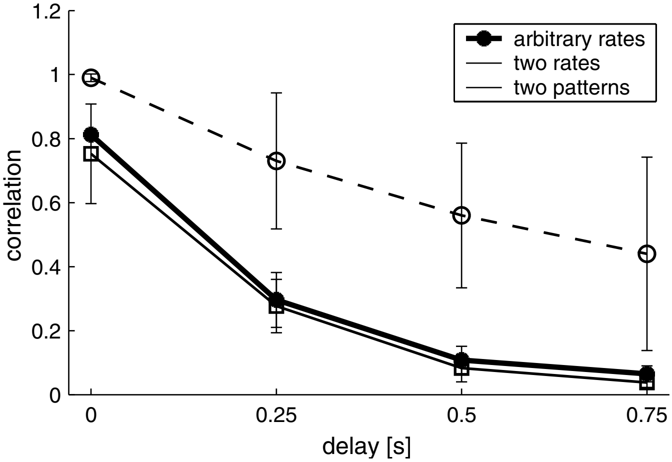

Figure 18.12

Memory curves for spike patterns and firing rates. Dashed line: correlation of trained linear readouts with the number of the templates used for generating the last input segment, and the segments that had ended 250 ms, 500 ms, and 750 ms ago (for the inputs discussed in Figure 18.11). Solid lines: correlation of trained linear readouts with the firing rates for the same time segments of length 250 ms that were used for the spike pattern classification task. Thick solid line is for the case where the ideal input firing rates can assume just 2 values (30 or 60 Hz), whereas the thin solid line is for the case where arbitrary firing rates between 0 and 80 Hz are randomly chosen. In either case the actual average input rates for the 4 time segments, which had to be recalled by the readouts, assumed of course a wider range.

planation is that the ensemble of liquid states reflecting preceding input spike trains that all represented the same firing rate forms a much more complicated equivalence class than liquid states resulting from jittered versions of a single spike pattern. This problem is amplified by the fact that information about earlier firing rates is *overwritten* with a much more diverse set of input patterns in subsequent input segments in the case of arbitrary Poisson inputs with randomly chosen rates. (The number of concurrent input spike trains that represent a given firing rate is less relevant for these memory curves; not shown.)

A theoretical analysis of memory retention in somewhat similar recurrent networks of sigmoidal neurons has been given in [11].

18.6.2 Kernel function of neural microcircuit models

It is well-known (see [22, 23, 25]) that the power of linear readouts can be boosted by two types of preprocessing:

- computation of a large number of nonlinear combinations of input components and features

- projection of the input into a very high dimensional space

In machine learning both preprocessing steps are carried out simultaneously by a so-called kernel, that uses a mathematical trick to avoid explicit computations in high-dimensional spaces. In contrast, in our model for computation in neural microcircuits both operations of a kernel are physically implemented (by the microcircuit). The high-dimensional space into which the input is projected is the state space of the neural microcircuit (a typical column consists of roughly 100 000 neurons). This implementation makes use of the fact that the precise mathematical formulas by which these nonlinear combinations and high-dimensional projections are computed are less relevant. Hence these operations can be carried out by *found* neural circuits that have not been constructed for a particular task. The fact that the generic neural microcircuit models in our simulations automatically compute an abundance of nonlinear combinations of input fragments can be seen from the fact that the target output values for the tasks considered in Figures 18.2, 18.4, 18.6, 18.8 are nonlinear in the input, but are nevertheless approximated quite well by *linear* readouts from the current state of the neural microcircuit.

The capability of neural microcircuits to boost the power of linear readouts by projecting the input into higher dimensional spaces is further underlined by joint work with Stefan Häusler [6]. There the task to recover the number of the template spike pattern used to generate the second-to-last segment of the input spike train* was carried out by generic neural microcircuit models of different sizes, ranging from 12 to 784 neurons.In each case a perceptron was trained by the Δ-rule to classify at time 0 the template that had been used to generate the input in the time segment [-500, -250 ms]. The results of the computer simulations reported in Figure 18.13 show that the performance of such (thresholded) linear readout improves drastically with the size of the microcircuit into which the spike train is injected, and therefore with the dimension of the *liquid state* that is presented to the readout.

18.7 Software for evaluating the computational capabilities of neural microcircuit models

New software for the creation, fast simulation and computational evaluation of neural microcircuit models has recently been written by Thomas Natschläger (with contributions by Christian Naeger), see [21]. This software, which has been made available for free use on WWW.LSM.TUGRAZ.AT, uses an efficient C^{++} kernel for the

*This is exactly the task of the second readout in the spike pattern classification task discussed in Figures 18.11 and 18.12.

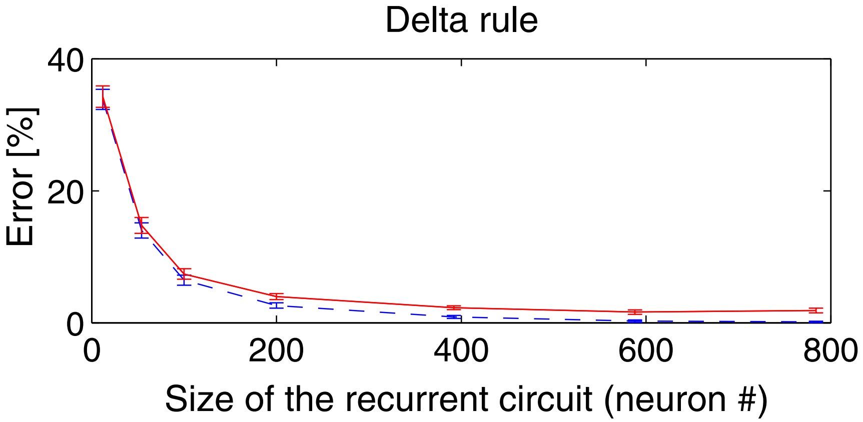

Figure 18.13

The performance of a trained readout (perceptron trained by the Δ-rule) for microcircuit models of different sizes, but each time for the same input injected into the microcircuit and the same classification task for the readout. The error decreases with growing circuit size, both on the training data (dashed line) and on new test data (solid line) generated by the same distribution.

simulation of neural microcircuits.* But the construction and evaluation of these microcircuit models can be carried out conveniently in MATLAB. In particular the website contains MATLAB scripts that can be used for validating the results reported in this chapter. The object oriented style of the software makes it easy to change the microcircuit model or the computational tasks used for these tests.

18.8 Discussion

We have presented a conceptual framework for analyzing computations in generic neural microcircuit models that satisfies the biological constraints listed in Section 18.1. Thus one can now take computer models of neural microcircuits, that can be as realistic as one wants to, and use them not just for demonstrating dynamic effects such as synchronization or oscillations, but to really carry out demanding computations with these models. The somewhat surprising result is that the inherent dynamics of cortical microcircuit models, which appears to be virtually impossible

*For example a neural microcircuit model consisting of a few hundred leaky integrate-and-fire neurons with up to 1000 dynamic synapses can be simulated in real-time on a current generation PC.

to understand in detail for a human observer, nevertheless presents information about the recent past of its input stream in such a way that a single perceptron (or linear readout in the case where an analog output is needed) can immediately extract from it the *right answer*. Traditional approaches towards producing the outputs of such complex computations in a computer usually rely on a sequential algorithm consisting of a sequence of computation steps involving elementary operations such as feature extraction, addition and multiplication of numbers, and *binding* of related pieces of information. The simulation results discussed in this chapter demonstrate that a completely different organization of such computations is possible, which does not require to implement these seemingly unavoidable elementary operations. Furthermore, this alternative computation style is supported by theoretical results (see Section 18.4), which suggest that it is in principle as powerful as von Neumann style computational models such as Turing machines, but more adequate for the type of real-time computing on analog input streams that is carried out by the nervous system.

Obviously this alternative conceptual framework relativizes some basic concepts of computational neuroscience such as receptive fields, neural coding and binding, or rather places them into a new context of computational organization. Furthermore it suggests new experimental paradigms for investigating the computational role of cortical microcircuits. Instead of experiments on highly trained animals that aim at isolating neural correlates of conjectured elementary computational operations, the approach discussed in this chapter suggests experiments on naturally behaving animals that focus on the role of cortical microcircuits as general purpose temporal integrators (analog fading memory) and simultaneously as high dimensional nonlinear kernels to facilitate linear readout. The underlying computational theory (and related experiments in machine learning) support the intuitively rather surprising finding that the precise details how these two tasks are carried out (e.g., how memories from different time windows are superimposed, or which nonlinear combinations are produced in the kernel) are less relevant for the performance of the computational model, since a linear readout from a high dimensional dynamical system can in general be trained to adjust to any particular way in which these two tasks are executed. Some evidence for temporal integration in cortical microcircuits has already been provided through experiments that demonstrate the dependence of the current dynamics of cortical areas on their initial state at the beginning of a trial, see e.g., [2]. Apparently this initial state contains information about preceding input to that cortical area. Our theoretical approach suggests further experiments that quantify the information about earlier inputs in the current state of neural microcircuits *in vivo*. It also suggests to explore in detail which of this information is read out by diverse readouts and projected to other brain areas.

The computational theory outlined in this chapter differs also in another aspect from previous theoretical work in computational neuroscience: instead of constructing hypothetical neural circuits for specific (typically simplified) computational tasks, this theory proposes to take the existing cortical circuitry *off the shelf* and examine which adaptive principles may enable them to carry out those diverse and demanding real-time computations on continuous input streams that are characteristic for the astounding computational capabilities of the cortex.

The generic microcircuit models discussed in this chapter were relatively simple insofar as they did not yet take into account more specific anatomical and neurophysiological data regarding the distribution of specific types of neurons in specific cortical layers, and known details regarding their specific connection patterns and regularization mechanisms to improve their performance (work in progress). But obviously these more detailed models can be analyzed in the same way, and it will be quite interesting to compare their computational power with that of the simpler models discussed in this chapter.

Acknowledgment: The work was partially supported by the Austrian Science Fond FWF, project # P15386.

References

[1] L. F. Abbott, and K. I. Blum (1996), Functional significance of long-term potentiation for sequence learning and prediction, *Cerebral Cortex*, **6**: 406–416.

[2] A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen (1996), Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses, *Science*, **273**: 1868–1871.

[3] P. Auer, H. Burgsteiner, and W. Maass (2002), Reducing communication for distributed learning in neural networks, in *Proc. of the International Conference on Artificial Neural Networks – ICANN 2002*, Lecture Notes in Computer Science, José R. Dorronsoro (ed.), **2415**: 123–128, Springer-Verlag. Online available as #127 from http: //www.igi.tugraz.at /maass /publications.html.

[4] D. V. Buonomano, and M. M. Merzenich (1995), Temporal information transformed into a spatial code by a neural network with realistic properties, *Science*, **267**: 1028–1030.

[5] A. Gupta, Y. Wang, and H. Markram (2000), Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex, *Science*, **287**: 273–278.

[6] S. Häusler, H. Markram, and W. Maass (2003), Perspectives of the high dimensional dynamics of neural microcircuits from the point of view of low dimensional readouts, *Complexity (Special Issue on Complex Adaptive Systems* (in press).

[7] S. Haykin (1999), *Neural Networks: A Comprehensive Foundation*, Prentice Hall.

[8] J. Hopfield, and C. Brody The Mus silicium (sonoran desert sand mouse) web page, Base: `http://moment.princeton.edu/~mus/Organism`, Data

set: Base + /Competition/digits_data.html, Scores: Base + /Docs/winners.html

[9] J. J. Hopfield, and C. D. Brody (2000), What is a moment? “Cortical” sensory integration over a brief interval, *Proc. Natl. Acad. Sci. USA*, **97**: 13919–13924.

[10] J. J. Hopfield, and C. D. Brody (2001), What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration, *Proc. Natl. Acad. Sci. USA*, **98**: 1282–1287.

[11] H. Jäger (2001), Short term memory in echo state networks, German National Research Center for Information Technology, GMD Report: 152.

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

[13] R. A. Legenstein, H. Markram, and W. Maass (2003), Input prediction and autonomous movement analysis in recurrent circuits of spiking neurons, *Reviews in the Neurosciences (Special Issue on Neural and Artificial Computation* (in press), Online available as #140 from http: //www.igi.tugraz.at /maass /publications.html.

[14] W. B. Levy (1996), A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks, *Hippocampus*, **6**: 579–590.

[15] Markram, H., Wang, Y., and Tsodyks, M. (1998), Differential signaling via the same axon of neocortical pyramidal neurons., *Proc. Natl. Acad. Sci.*, **95**: 5323–5328.

[16] W. Maass, T. Natschläger, and H. Markram (2002), Real-time computing without stable states: a new framework for neural computation based on perturbations, *Neural Computation*, **14**: 2531–2560. Online available as #130 from http: //www.igi.tugraz.at /maass /publications.html.

[17] W. Maass, and E. D. Sontag (1999), Neural systems as nonlinear filters, *Neural Computation*, **12**: 1743–1772. Online available as #107 from http: //www.igi.tugraz.at /maass/publications.html.

[18] W. Maass, R. A. Legenstein, and H. Markram (2002), A new approach towards vision suggested by biologically realistic neural microcircuit models, in H.H. Buelthoff, S.W. Lee, T.A. Poggio, and C. Wallraven (eds.) *Biologically Motivated Computer Vision, Proc. of the 2nd Workshop on Biologically Motivated Computer Vision*, *Lecture Notes in Computer Science* **2525**, Springer, 282-293. in press, Online available as #146 from http: //www.igi.tugraz.at /maass /publications.html.

[19] H. A. Mallot (2000), *Computational Vision*, MIT-Press, Cambridge, MA.

[20] R. I. Soare (1987), *Recursively Enumerable Sets and Degrees: A Study of Computable Functions and Computably Enumerable Sets*, Springer Verlag.

[21] T. Natschläger, H. Markram, and W. Maass (2003), Computer models and analysis tools for neural microcircuits, in *Neuroscience Databases. A Practical Guide*, R. Kötter (ed.), Kluwer Academic Publishers (Boston), Chapter 9, 123–138. Online available as #144 from http: //www.igi.tugraz.at /maass /publications.html.

[22] J. F. Rosenblatt (1962), *Principles of Neurodynamics*, Spartan Books (New York).

[23] B. Schölkopf, and A. J. Smola (2002), *Learning with Kernels*, MIT Press, Cambridge.

[24] J. E. Savage (1998), *Models of Computation: Exploring the Power of Computing*, Addison-Wesley, Reading, MA, USA.

[25] V. N. Vapnik (1998), *Statistical Learning Theory*, John Wiley, New York.

Chapter 19

Modelling Primate Visual Attention

Laurent Itti
University of Southern California, Hedco Neuroscience Building HNB-30A, Los Angeles, CA 90089-2520, U.S.

CONTENTS

19.1 Introduction

Selective visual attention is the mechanism by which we can rapidly direct our gaze towards objects of interest in our visual environment [2, 4, 5, 18, 26, 35, 52, 53]. From an evolutionary viewpoint, this rapid orienting capability is critical in allowing living systems to quickly become aware of possible prey, mates or predators in their cluttered visual world. It has become clear that attention guides where to look next based on both bottom-up (image-based) and top-down (task-dependent) cues [26]. As such, attention implements an information processing bottleneck, only allowing a small part of the incoming sensory information to reach short-term memory and visual awareness [9, 15]. That is, instead of attempting to fully process the massive sensory input in parallel, nature has devised a serial strategy to achieve near real-time performance despite limited computational capacity: Attention allows us to break down the problem of scene understanding into rapid series of computationally less demanding, localized visual analysis problems.

These orienting and scene analysis functions of attention are complemented by a feedback modulation of neural activity at the location and for the visual attributes of the desired or selected targets. This feedback is believed to be essential for binding the different visual attributes of an object, such as color and form, into a unitary percept [22, 42, 52]. That is, attention not only serves to select a location of interest, but also enhances the cortical representation at that location. As such, focal visual attention is often compared to a rapidly shiftable spotlight [11, 58], which scans our visual environment both overtly (with accompanying eye movements) or covertly (with the eyes fixed).

Finally, attention is involved in triggering behavior, and consequently is intimately related to recognition, planning and motor control [32]. Of course, not all of vision is attentional, as we can derive coarse understanding from presentations of visual scenes that are too brief for attention to explore the scene. Vision thus relies on sophisticated interactions between coarse, massively parallel, full-field pre-attentive analysis systems and the more detailed, circumscribed and sequential attentional analysis system.

In what follows, we focus on several critical aspects of selective visual attention: First, the brain area involved in its control and deployment; second, the mechanisms by which attention is attracted in a bottom-up or image-based manner towards conspicuous or salient locations in our visual environment; third, the mechanisms by which attention modulates the early sensory representation of attended stimuli; fourth, the mechanisms for top-down or voluntary deployment of attention; and fifth, the interaction between attention, object recognition and scene understanding.

19.2 Brain areas

The control of focal visual attention involves an intricate network of brain areas, spanning from primary visual cortex to prefrontal cortex. In a first approximation, selecting where to attend next is carried out, to a large extent, by distinct brain structures from recognizing what is being attended to. This suggests that a cooperation between "two visual systems" is used by normal vision [16]: Selecting where to attend next is primarily controlled by the dorsal visual processing stream (or "where/how" stream) which comprises cortical areas in posterior parietal cortex, whereas the ventral visual processing stream (or "what" stream), comprising cortical areas in inferotemporal cortex, is primarily concerned with localized object recognition [57]. It is

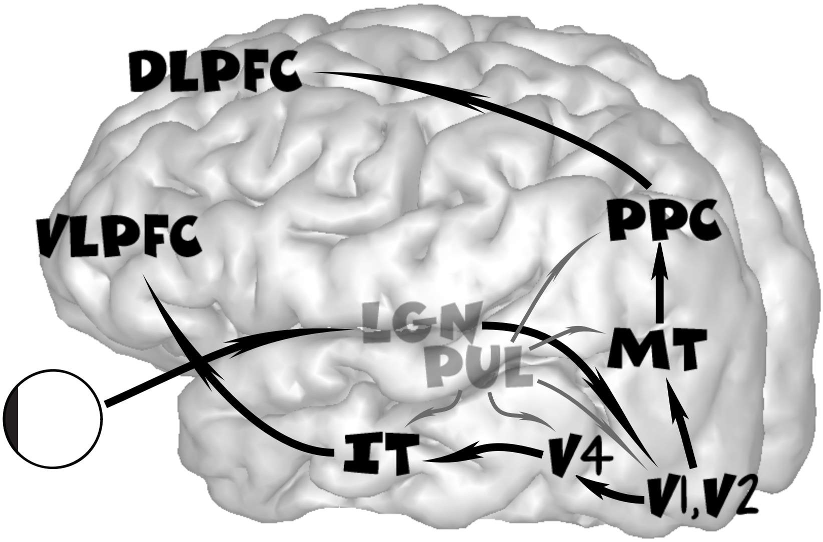

Figure 19.1

Major brain areas involved in the deployment of selective visual attention. Although single-ended arrows are shown to suggest global information flow (from the eyes to prefrontal cortex), anatomical studies suggest reciprocal connections, with the number of feedback fibers often exceeding that of feedforward fibers (except between retina and LGN). Cortical areas may be grouped into two main visual pathways: the dorsal "where/how" pathway (from V1 to DLPFC via PPC) is mostly concerned with spatial deployment of attention and localization of attended stimuli, while the ventral "what" pathway (from V1 to VLPFC via IT) is mostly concerned with pattern recognition and identification of the attended stimuli. In addition to these cortical areas, several subcortical areas including LGN and Pul play important roles in controlling where attention is to be deployed. **Key to abbreviations:** LGN: lateral geniculate nucleus; Pul: Pulvinar nucleus; V1, V2, V4: early cortical visual areas; MT: Medial temporal area; PPC: posterior parietal cortex; DLPFC: dorsolateral prefrontal cortex; IT: inferotemporal cortex; VLPFC: ventrolateral prefrontal cortex.

important to note, however, that object recognition in the ventral stream can bias the next attentional shift, for example via top-down control when an object is recognized that suggests where the next interesting object may be located. Similarly, we will see how attention strongly modulates activity in the object recognition system.

Among the brain regions participating to the deployment of visual attention include most of the early visual processing areas and the dorsal processing stream (Figure 19.1). These include the lateral geniculate nucleus of the thalamus (LGN) and cortical areas V1 (primary visual cortex) through the parietal cortex along the dorsal stream [51]. In addition, overt and covert attention have been shown to be closely related, as revealed by psychophysical [19, 28, 47, 48], physiological [1, 6, 29, 45], and imaging [7, 37] studies. Directing covert attention thus involves a number of subcortical structures that are also instrumental in producing directed eye movements. These include the deeper parts of the superior colliculus; parts of the pulvinar; the frontal eye fields in the macaque and its homologue in humans; the precentral gyrus; and areas in the intraparietal sulcus in the macaque and around the intraparietal and postcentral sulci and adjacent gyri in humans.

19.3 Bottom-up control

One important mode of operation of attention is largely unconscious and driven by the specific attributes of the stimuli present in our visual environment. This so-called bottom-up control of visual attention can easily be studied using simple visual search tasks as described below. Based on these experimental results, several computational theories and models have been developed for how attention may be attracted towards a particular object in the scene rather than another.

19.3.1 Visual search and pop-out

One of the most effective demonstrations of bottom-up attentional guidance uses simple visual search experiments, in which an odd target stimulus to be located by the observer is embedded within an array of distracting visual stimuli [52]. Originally, these experiments suggested a dichotomy between situations where the target stimulus would visually pop-out from the array and be found immediately, and situations where extensive scanning and inspection of the various stimuli in the display was necessary before the target stimulus could be located (Figure 19.2). The pop-out cases suggest that the target can be effortlessly located by relying on preattentive visual processing over the entire visual scene. In contrast, the conjunctive search cases suggest that attending to the target is a necessary precondition to being able to identify it as being the unique target, thus requiring that the search array be extensively scanned until the target becomes the object of attentional selection.

Further experimentation has revealed that the original dichotomy between the fast, parallel search observed with pop-out displays and slower, serial search observed with conjunctive displays represent the two extremes of a continuum of search difficulty [60]. Nevertheless, these experiments clearly demonstrate that if a target differs significantly from its surround (in ways which can be characterized in terms of visual attributes of the target and distractors), it will immediately draw attention towards itself. Thus, these experiments evidence how the composition of the visual scene alone is a potentially very strong component of attentional control, guiding attention from the bottom of the visual processing hierarchy up.

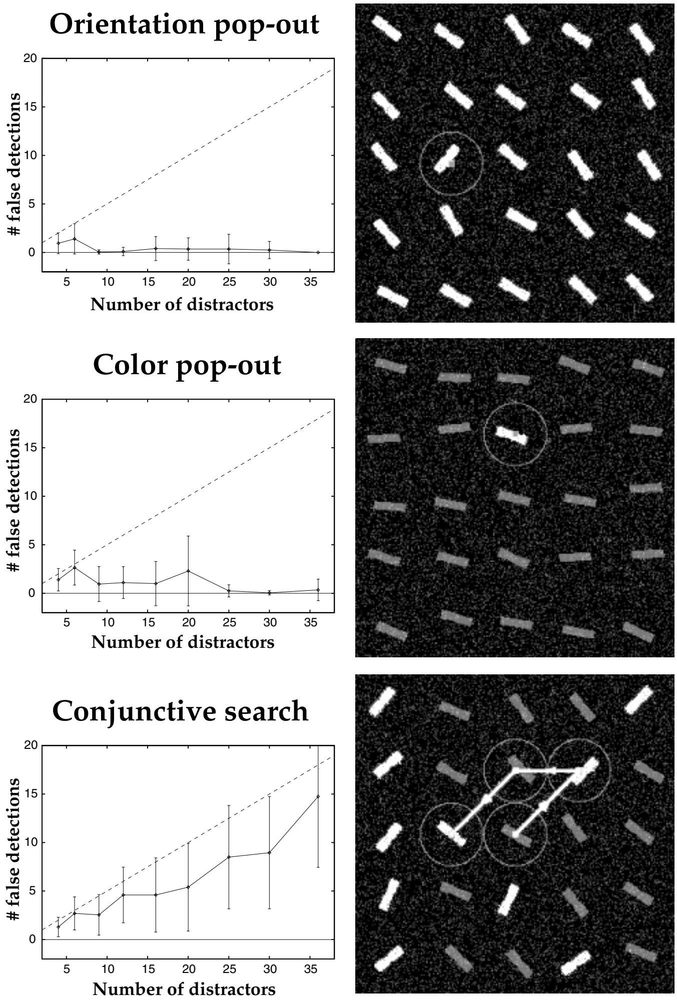

Figure 19.2

Search array experiments of the type pioneered by Treisman and colleagues. The top two panels are examples of pop-out cases where search time (here shown as the number of locations fixated before the target if found) is small and independent of the number of elements in the display. The bottom panel demonstrates a conjunctive search (the target is the only element that is dark *and* oriented like the brighter elements); in this case, a serial search is initiated, which will require more time as the number of elements in the display is increased.

19.3.2 Computational models and the saliency map

The feature integration theory of Treisman and colleagues [52] that was derived from visual search experiments has served as a basis for many computational models of bottom-up attentional deployment. This theory proposed that only fairly simple visual features are computed in a massively parallel manner over the entire incoming visual scene, in early visual processing areas including primary visual cortex. Attention is then necessary to bind those early features into a more sophisticated object representation, and the selected bound representation is (to a first approximation) the only part of the visual world which passes though the attentional bottleneck for further processing.

The first explicit neurally-plausible computational architecture of a system for the bottom-up guidance of attention was proposed by Koch and Ullman [27], and is closely related to the feature integration theory. Their model is centered around a saliency map, that is, an explicit two-dimensional topographic map that encodes for stimulus conspicuity, or salience, at every location in the visual scene. The saliency map receives inputs from early visual processing, and provides an efficient control strategy by which the focus of attention simply scans the saliency map in order of decreasing saliency.

This general architecture has been further developed and implemented, yielding the computational model depicted in Figure 19.3 [23]. In this model, the early stages of visual processing decompose the incoming visual input through an ensemble of feature-selective filtering processes endowed with contextual modulatory effects. In order to control a single attentional focus based on this multiplicity in the representation of the incoming sensory signals, it is assumed that all feature maps provide input to the saliency map, which topographically represents visual salience, irrespectively of the feature dimension by which a given location was salient. Biasing attention to focus onto the most salient location is then reduced to drawing attention towards the locus of highest activity in the saliency map. This is achieved using a winner-take-all neural network, which implements a neurally distributed maximum detector. In order to prevent attention from permanently focusing onto the most active (winner) location in the saliency map, the currently attended location is transiently inhibited in the saliency map by an inhibition-of-return mechanism. After the most salient location is thus suppressed, the winner-take-all network naturally converges towards the next most salient location, and repeating this process generates attentional scanpaths [23, 27].

Many successful models for the bottom-up control of attention are architectured around a saliency map. What differentiates the models, then, is the strategy employed to prune the incoming sensory input and extract salience. In an influential model mostly aimed at explaining visual search experiments, Wolfe [59] hypothesized that the selection of relevant features for a given search task could be performed top-down, through spatially-defined and feature-dependent weighting of the various feature maps. Although limited to cases where attributes of the target are known in advance, this view has recently received experimental support from studies of top-down attentional modulation (see below).

Tsotsos and colleagues [56] implemented attentional selection using a combination of a feedforward bottom-up feature extraction hierarchy and a feedback selective tuning of these feature extraction mechanisms. In this model, the target of attention is selected at the top level of the processing hierarchy (the equivalent of a saliency

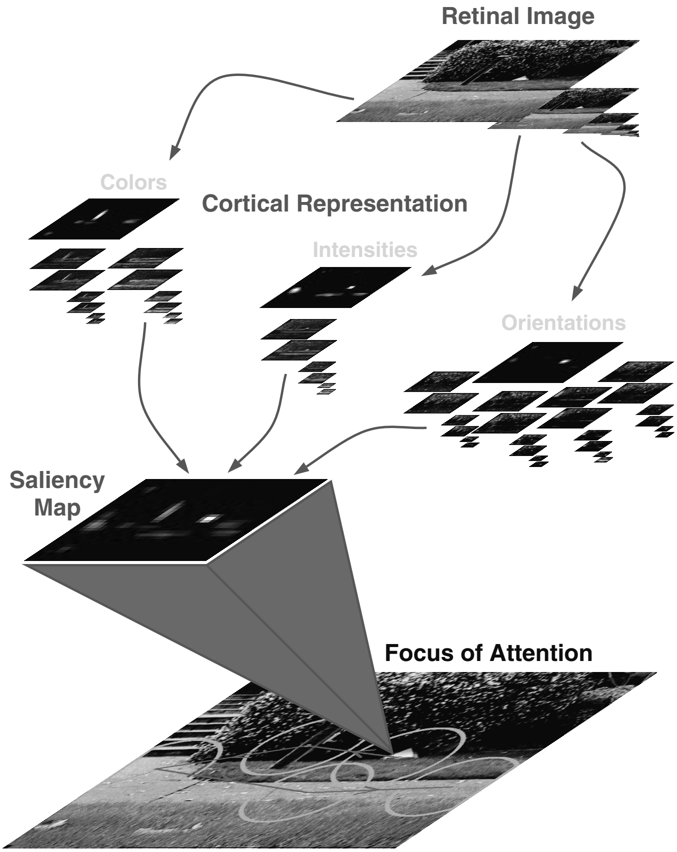

Figure 19.3

Typical architecture for a model of bottom-up visual attention based on a saliency map. The input image is analyzed by a number of early visual filters, sensitive to stimulus properties such as color, intensity and orientation, at several spatial scales. After spatial competition for salience within each of the resulting feature maps, input is provided to a single saliency map from all of the feature maps. The aximum activity in the saliency map is the next attended location. Transient inhibition of this location in the saliency map allows the system to shift towards the next most salient location.

map), based on feedforward activation and on possible additional top-down biasing for certain locations or features. That location is then propagated back through the feature extraction hierarchy, through the activation of a cascade of winner-take-all networks embedded within the bottom-up processing pyramid. Spatial competition for salience is thus refined at each level of processing, as the feedforward paths not contributing to the winning location are pruned (resulting in the feedback propagation of an “inhibitory beam” around the selected target).

Itti et al. [23, 24, 25] recently proposed a purely bottom-up model, in which spatial competition for salience is directly modelled after non-classical surround modulation effects. The model employs an iterative scheme with early termination. At each iteration, a feature map receives additional inputs from the convolution of itself by a large difference-of-Gaussians filter. The result is half-wave rectified, with a net effect similar to a winner-take-all with limited inhibitory spread, which allows only a sparse population of locations to remain active. After competition, all feature maps are simply summed to yield the scalar saliency map. Because it includes a complete biological front-end, this model has been widely applied to the analysis of natural color scenes [25]. The non-linear interactions implemented in this model strongly illustrate how, perceptually, whether a given stimulus is salient or not cannot be decided without knowledge of the context within which the stimulus is presented.

Many other models have been proposed, which typically share some of the components of the three models just described. In view of the affluence of models based on a saliency map, it is important to note that postulating centralized control based on such map is not the only computational alternative for the bottom-up guidance of attention. In particular, Desimone and Duncan [15] argued that salience is not explicitly represented by specific neurons, but instead is implicitly coded in a distributed modulatory manner across the various feature maps. Attentional selection is then performed based on top-down weighting of the bottom-up feature maps that are relevant to a target of interest. This top-down biasing (also used in Wolfe's Guided Search model [59]) requires that a specific search task be performed for the model to yield useful predictions.

19.4 Top-down modulation of early vision

The general architecture for the bottom-up control of attention presented above opens two important questions on the nature of the attentional bottleneck. First, is it the only means through which incoming visual information may reach higher levels of processing? Second, does it only involve one-way processing of information from the bottom-up, or is attention a two-way process, also feeding back from higher centers to early processing stages?

19.4.1 Are we blind outside of the focus of attention?

Recent experiments have shown how fairly dramatic changes applied to a visual scene being inspected may go unnoticed by human observers, unless those changes occur at the location currently being attended to. These change blindness experiments [39, 41] can take several forms, yielding essentially the same conclusions. One implementation consists of alternatively flashing two versions of a same scene separated by a blank screen, with the two versions differing very obviously at one location (for example, a scene in which a jet airplane is present and one of its reactors has been erased from one of the two photographs to be compared). Although the alteration is obvious when one directly attends to it, it takes naive observers several tens of seconds to locate it. Not unexpectedly, instances of this experiment which are the most difficult for observers involve a change at a location that is of little interest in terms of understanding and interpreting the scene (for example, the aforementioned scene with an airplane also contains many people, who tend to be inspected in priority).

These experiments demonstrate the crucial role of attention in conscious vision: unless we attend to an object, we are unlikely to consciously perceive it in any detail and detect when it is altered. However, as we will see below, this does not necessarily mean that there is no vision other than through the attention bottleneck.

19.4.2 Attentional modulation of early vision

A number of psychophysical end electrophysiological studies indicate that we are not entirely blind outside the focus of attention. At the early stages of processing, responses are still observed even if the animal is attending away from the receptive field at the site of recording [54], or is anesthetized [21]. Behaviorally, we can also perform fairly specific spatial judgments on objects not being attended to [4, 14], though those judgments are less accurate than in the presence of attention [31, 61]. This is in particular demonstrated by dual-task psychophysical experiments in which observers are able to simultaneously discriminate two visual stimuli presented at two distant locations in the visual field [31].

While attention thus appears not to be mandatory for early vision, it has recently become clear that it can vigorously modulate, top-down, early visual processing, both in a spatially-defined and in a non-spatial but feature-specific manner [10, 34, 55]. This modulatory effect of attention has been described as enhanced gain [54], biased [33, 36] or intensified [31] competition, enhanced spatial resolution [61], or as modulated background activity [12], effective stimulus strength [43] or noise [17].

Of particular interest in a computational perspective, a recent study by Lee et al. [31] measured psychophysical thresholds for five simple pattern discrimination tasks (contrast, orientation and spatial frequency discriminations, and two spatial masking tasks; 32 thresholds in total). They employed a dual-task paradigm to measure thresholds either when attention was fully available to the task of interest, or when it was poorly available because engaged elsewhere by a concurrent attention-demanding task. The mixed pattern of attentional modulation observed in the thresholds (up to 3-fold improvement in orientation discrimination with attention, but only 20% improvement in contrast discrimination) was quantitatively accounted for by a computational model. It predicted that attention strengthens a winner-take-all competition among neurons tuned to different orientations and spatial frequencies within one cortical hypercolumn [31], a proposition which has recently received additional experimental support.

These results indicate that attention does not implement a feed-forward, bottom-up information processing bottleneck. Rather, attention also enhances, through feedback, early visual processing for both the location and visual features being attended to.

19.5 Top-down deployment of attention

The precise mechanisms by which voluntary shifts of attention are elicited remain elusive, although several studies have narrowed down the brain areas primarily involved [8, 20, 23]. Here we focus on two types of experiments that clearly demonstrate how, first, attention may be deployed on a purely voluntary basis onto one of several identical stimuli (so that none of the stimuli is more salient than the others), and, second, how eye movements recorded from observers inspecting a visual scene with the goal of answering a question about that scene are dramatically influenced by the question being answered.

19.5.1 Attentional facilitation and cuing

Introspection easily reveals that we are able to voluntarily shift attention towards any location in our visual field, no matter how inconspicuous that location may be. More formally, psychophysical experiments may be used to demonstrate top-down shifts of attention. A typical experiment involves cueing an observer towards one of several possible identical stimuli presented on a computer screen. The cue indicates to the observer where to focus on, but only at a high cognitive level (e.g., verbal cue), so that nothing in the display would directly attract attention bottom-up towards the desired stimulus. Detection or discrimination of the stimulus at the attended location are significantly better (e.g., lower reaction time or lower psychophysical thresholds) than at uncued locations. These experiments hence suggest that voluntarily shifting attention towards a stimulus improves the perception of that stimulus.

Similarly, experiments involving decision uncertainty demonstrate that if a stimulus is to be discriminated by a specific attribute that is known in advance (e.g., discriminate the spatial frequency of a grating), performance is significantly improved compared to situations where one randomly chosen of several possible stimulus attributes are to be discriminated (e.g., discriminate the spatial frequency, contrast or orientation of a grating). Thus, we appear to also be able to voluntarily select not only where to attend to, but also the specific features of a stimulus to be attended. These results are closely related to and consistent with the spatial and featural nature of attentional modulation mentioned in the previous section.

19.5.2 Influence of task

Recording eye movements from human observers while they inspect a visual scene has revealed a profound influence of task demands on the pattern of eye movements generated by the observers [62]. In a typical experiment, different observers examine a same photograph while their eye movements are being tracked, but are asked to answer different questions about the scene (for example, estimate the age of the people in the scene, or determine the country in which the photograph was taken). Although all observers are presented with an identical visual stimulus, the patterns of eye movements recorded differ dramatically depending on the question being addressed by each observer. These experiments clearly demonstrate that task demands play a critical role in determining where attention is to be focused next.

Building in part on eye tracking experiments, Stark and colleagues [38] have proposed the scanpath theory of attention, according to which eye movements are generated almost exclusively under top-down control. The theory proposes that what we see is only remotely related to the patterns of activation of our retinas; rather, a cognitive model of what we expect to see is at the basis of our percept. The sequence of eye movements which we make to analyze a scene, then, is mostly controlled top-down by our cognitive model and serves the goal of obtaining specific details about the particular scene instance being observed, to embellish the more generic internal model. This theory has had a number of successful applications to robotics control, in which an internal model of a robot's working environment was used to restrict the analysis of incoming video sequences to a small number of circumscribed regions important for a given task.

19.6 Attention and scene understanding

We have seen how attention is deployed onto our visual environment through a cooperation between bottom-up and top-down driving influences. One difficulty which then arises is the generation of proper top-down biasing signals when exploring a novel scene; indeed, if the scene has not been analyzed and understood yet using thorough attentional scanning, how can it be used to direct attention top-down? Below we explore two dimensions of this problem: First, we show how already from a very brief presentation of a scene we are able to extract its gist, basic layout, and a number of other characteristics. This suggests that another part of our visual system, which operates much faster than attention, might be responsible for this coarse analysis; the results of this analysis may then be used to guide attention top-down. Second, we explore how several computer vision models have used a collaboration between the where and what subsystems to yield sophisticated scene recognition algorithms. Finally, we cast these results into a more global view of our visual system and the function of attention in vision.

19.6.1 Is scene understanding purely attentional?

Psychophysical experiments pioneered by Biederman and colleagues [3] have demonstrated how we can derive coarse understanding of a visual scene from a single presentation that is so brief (80 ms or less) that it precludes any attentional scanning or eye movement. A particularly striking example of such experiments consists of presenting to an observer a rapid succession of unrelated photographs of natural scenes at a high frame rate (over 10 scenes/s). After presentation of the stimuli for several tens of seconds, observers are asked whether a particular scene, for example an outdoors market scene, was present among the several hundred photographs shown. Although the observers are not made aware in advance of the question, they are able to provide a correct answer with an overall performance well over chance (Biederman, personal communication). Furthermore, observers are able to recall a number of coarse details about the scene of interest, such as whether it contained humans, or whether it was highly colorful or rather dull.

These and many related experiments clearly demonstrate that scene understanding does not exclusively rely on attentional analysis. Rather, a very fast visual subsystem which operates in parallel with attention allows us to rapidly derive the gist and coarse layout of a novel visual scene. This rapid subsystem certainly is one of the key components by which attention may be guided top-down towards specific visual locations.

19.6.2 Cooperation between where and what

Several computer vision models have been proposed for extended object and scene analysis that rely on a cooperation between an attentional (where) and localized recognition (what) subsystems.

A very interesting instance was recently provided by Schill et al. [46]. Their model aims at performing scene (or object) recognition, using attention (or eye movements) to focus on those parts of the scene being analyzed which are most informative in disambiguating its identity. To this end, a hierarchical knowledge tree is trained into the model, in which leaves represent identified objects, intermediary nodes represent more general object classes, and links between nodes contain sensorimotor information used for discrimination between possible objects (i.e., bottom-up feature responses to be expected for particular points in the object, and eye movement vectors targeted at those points). During the iterative recognition of an object, the system programs its next fixation towards the location which will maximally increase information gain about the object being recognized, and thus will best allow the model to discriminate between the various candidate object classes.

Several related models have been proposed [13, 23, 44, 49, 50], in which scanpaths (containing motor control directives stored in a "where" memory and locally expected bottom-up features stored in a "what" memory) are learned for each scene or object to be recognized. The difference between the various models comes from the algorithm used to match the sequences of where/what information to the visual scene. These include using a deterministic matching algorithm (i.e., focus next onto the next location stored in the sequence being tested against the new scene), hidden Markov models (where sequences are stored as transition probabilities between locations augmented by the visual features expected at those locations), or evidential reasoning (similar to the model of Schill and colleagues). These models typically demonstrate strong ability to recognize complex grayscale scenes and faces, in a translation, rotation and scale independent manner, but cannot account for non-linear image transformations (e.g., three-dimensional viewpoint change).

While these models provide very interesting examples of cooperation between a fast attentional cueing system and a slower localized feature analysis system, their relationship to biology has not been emphasized beyond the general architectural level. Teasing apart the brain mechanisms by which attention, localized recognition, and rapid computation of scene gist and layout collaborate in normal vision remains one of the most exciting challenges for modern visual neuroscience [40].

19.6.3 Attention as a component of vision

In this section, we have seen how vision relies not only on the attentional subsystem, but more broadly on a cooperation between crude preattentive subsystems for the computation of gist, layout and for bottom-up attentional control, and the attentive subsystem coupled with the localized object recognition subsystem to obtain fine details at various locations in the scene (Figure 19.4).

This view on the visual system raises a number of questions which remain fairly controversial. These are issues of the internal representation of scenes and objects (e.g., view-based versus three-dimensional models, or a cooperation between both), and of the level of detail with which scenes are stored in memory for later recall and comparison to new scenes (e.g., snapshots versus crude structural models). Many of these issues extend well beyond the scope of the present discussion of selective visual attention. Nevertheless, it is important to think of attention within the broader framework of vision and scene understanding, as this allows us to delegate some of the visual functions to non-attentional subsystems.

19.7 Discussion

We have reviewed some of the key aspects of selective visual attention, and how these contribute more broadly to our visual experience and unique ability to rapidly comprehend complex visual scenes.

Looking at the evidence accumulated so far on the brain areas involved with the control of attention has revealed a complex interconnected network, which spans from the earliest stages of visual processing up to prefrontal cortical areas. To a large extent, this network serves not only the function of guiding attention, but is shared with other subsystems, including the guidance of eye movements, the computation

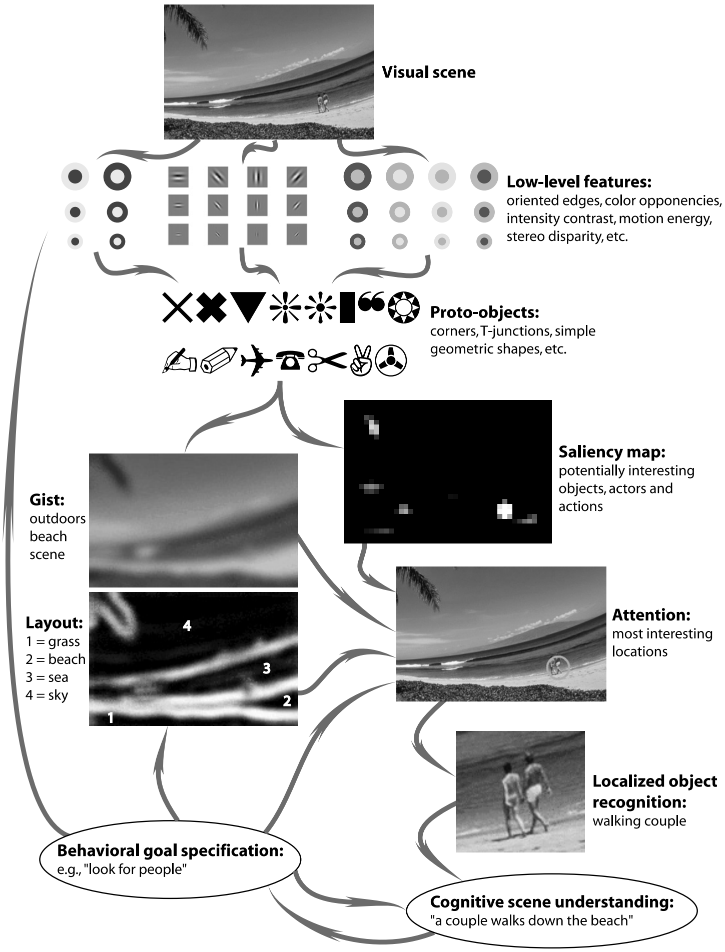
Visual scene
Low-level features:
oriented edges, color opponencies, intensity contrast, motion energy, stereo disparity, etc.
Proto-objects:
corners, T-junctions, simple geometric shapes, etc.
Saliency map:
potentially interesting objects, actors and actions
Gist:
outdoors beach scene
Layout:
1 = grass
2 = beach
3 = sea
4 = sky
4
3
2
1
Attention:
most interesting locations
Localized object recognition:
walking couple
Behavioral goal specification:
e.g., "look for people"
Cognitive scene understanding:
"a couple walks down the beach"

Figure 19.4

Simplified architecture for the understanding of visual scenes, extended from Rensink's (2000) triadic model. The incoming visual scene is analyzed by low-level visual processes (top) in a massively-parallel, full-field and pre-attentive manner up to a fairly simple "proto-object" representation. Building on this representation, gist and layout of the scene are computed in a fast, probably feedforward and non-iterative manner (left). Also building on this representation, the saliency map describes potentially interesting locations in the scene (right). Guided by saliency, gist, layout, and behavioral goal specifications, focal attention selects a region of the scene to be analyzed in further details. The result of this localized object recognition is used to incrementally refine the cognitive understanding of the contents of the scene. This understanding as well as the goal specification bias the low-level vision through feedback pathways.

of early visual features, the recognition of objects and the planning of actions.

Attention is guided towards particular locations in our visual world under a combination of competing constraints, which include bottom-up signals derived from the visual input, and top-down contraints derived from task priority and scene understanding. The bottom-up control of attention is clearly evidenced by simple visual search experiments, in which our attention is automatically drawn towards targets that pop-out from surrounding distractors. This bottom-up guidance is certainly the best understood component of attention, and many computational models have been proposed which replicate some of the human performance at exploring visual search stimuli. Most models have embraced the idea that a single topographic saliency map may be an efficient centralized representation for guiding attention. Several of these models have been applied to photographs of natural scenes, yielding remarkably plausible results. One of the important theoretical results derived from bottom-up modelling is the critical role of cortical interactions in pruning the massive sensory input such as to extract only those elements of the scene that are conspicuous.

In part guided by bottom-up cues, attention thus implements an information processing bottleneck, which allows only select elements in the scene to reach higher levels of processing. But not all vision is attentional, and even though we may easily appear blind to image details outside the focus of attention, there is still substantial residual vision of unattended objects. That is, the attentional bottleneck is not strict, and some elements in the visual scene may reach our conscious perception if they are sufficiently salient, even though attention might be engaged elsewhere in the visual environment.

In addition, attentional selection appears to be a two-way process, in which not only selected scene elements are propagated up the visual hierarchy, but the representation of these elements is also enhanced down to the earliest levels of the hierarchy through feedback signals. Thus attention not only serves the function of selecting a subset of the current scene, but also profoundly alters the cortical representation of this subset. Computationally, one mechanism for this enhancement which enjoys broad validity across a variety of visual discrimination tasks is that attention may activate a winner-take-all competition among visual neurons representing different aspects of a same visual location, thus making more explicit what the dominant characteristic of that location is. Top-down attentional modulation can be triggered not only on the basis of location, but also towards specific visual features.

Introspection easily makes evident that attention is not exclusively controlled bottom-up. Indeed, we can with little effort focus attention onto any region of our visual field, no matter how inconspicuous. Volitional shifts of attention are further evidenced by psychophysical experiments in which improved performance is observed when subjects know in advance where or what to look for, and hence presumably use a volitional shift of attention (across space or feature dimensions) in preparation for performing a visual judgement. The exact mechanisms by which volitional attention shifts are elicited remain rather elusive, but it has been widely demonstrated that high-level task specifications, such as a question asked about a visual scene, have dramatic effects on the deployment of attention and eye movements onto the scene.

Finally, it is important to consider attention not as a visual subsystem of its own that would have little interaction with other aspects of vision. Indeed, we have seen that it is highly unlikely, or impossible under conditions of very brief presentation, that we analyze and understand complex scenes only through attentional scanning. Rather, attention, object recognition, and rapid mechanisms for the extraction of scene gist and layout cooperate in a remarkable multi-threaded analysis which exploits different time scales and levels of details within interacting processing streams. Although tremendous progress has been made over the past century of the scientific study of attention, starting with William James, many of the key components of this complex interacting system remain poorly understood and elusive, thus posing ever renewed challenges for future neuroscience research.

References

[1] Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., and Fogassi, L. (1990), Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque, *J. Neurosci.*, **10** :1176-96

[2] Bergen, J., and Julesz, B. (1983), Parallel versus serial processing in rapid pattern discrimination, *Nature* (London), **303**: 696-698.

[3] Biederman I. (1972), Perceiving real-world scenes, *Science*, **177**: 77-80

[4] Braun, J., and Sagi, D.(1990), Vision outside the focus of attention, *Percept. Psychophys.*, **48**: 45-58

[5] Braun, J., and Julesz, B. (1998), Withdrawing attention at little or no cost: detection and discrimination tasks, *Percept. Psychophys.*, **60**: 1-23.

[6] Colby, C. L., and Goldberg, M. E. (1999), Space and attention in parietal cortex, *Annu. Rev. Neurosci.*, **22**: 319-49.

[7] Corbetta, M. (1998), Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems?, *Proc. Natl. Acad. Sci. USA*, **95**: 831-8.

[8] Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000), Voluntary orienting is dissociated from target detection in human posterior parietal cortex [published erratum appears in Nat Neurosci 2000 May; 3(5): 521], *Nat. Neurosci.*, **3**: 292-297.

[9] Crick, F., and Koch, C. (1998), Constraints on cortical and thalamic projections: the no-strong-loops hypothesis, *Nature*, **391**: 245-50.

[10] Barcelo, F., Suwazono, S., and Knight, R. T. (2000), Prefrontal modulation of visual processing in humans, *Nat. Neurosci.*, **3**: 399-403.

[11] Brefczynski, J. A., and DeYoe, E. A. (1999), A physiological correlate of the

'spotlight' of visual attention, *Nat. Neurosci.*, **2**: 370-374.

[12] Chawla, D., Rees, G., and Friston, K. J. (1999), The physiological basis of attentional modulation in extrastriate visual areas, *Nat. Neurosci.*, **2**: 671-676.

[13] Deco, G., and Zihl, J. (2001), A neurodynamical model of visual attention: feedback enhancement of spatial resolution in a hierarchical system, *Journal of Computational Neuroscience*, **10**: 231-151.

[14] DeSchepper, B., and Treisman, A. (1997), Visual memory for novel shapes: implicit coding without attention, *J. Exp. Psychol. Learn. Mem. Cogn.*, **22**: 27-47.

[15] Desimone, R., and Duncan, J. (1995), Neural mechanisms of selective visual attention, *Annu. Rev. Neurosci.*, **18**: 193-222.

[16] Didday, R. L. and Arbib, M. A. (1975), Eye movements and visual perception: a *two visual system* model, *Int. J. Man-Machine Studies*, **7**: 547-569.

[17] Dosher, B. A., and Lu, Z. L. (2000), Mechanisms of perceptual attention in precuing of location, *Vision Res.*, **40**: 1269-1292

[18] Hikosaka, O., Miyauchi, S., and Shimojo, S.(1996), Orienting a spatial attention-its reflexive, compensatory, and voluntary mechanisms, *Brain Res. Cogn.*, **5**: 1-9.

[19] Hoffman, J. E., and Subramaniam, B. (1995) The role of visual attention in saccadic eye movements, *Percept. Psychophys.*, **57**: 787-795.

[20] Hopfinger, J. B., Buonocore, M. H., and Mangun, G. R. (2000), The neural mechanisms of top-down attentional control, *Nat. Neurosci.*, **3**: 284-291

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

[22] Hummel, J. E., and Biederman, I. (1992), Dynamic binding in a neural network for shape recognition, *Psychol. Rev.*, **99**: 480-517.

[23] Itti L., and Koch C. (2001), Computational modeling of visual attention, *Nature Reviews Neuroscience*, **2**: 194-203.

[24] Itti L., Koch C., and Niebur E. (1998), A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**: 1254-1259.

[25] Itti L., and Koch C. (2000), A saliency-based search mechanism for overt and covert shifts of visual attention *Vision Research*, **40**: 1489-1506.

[26] James W. (1890/1981) *The Principles of Psychology* Harvard University Press, Cambridge, MA.

[27] Koch, C., and Ullman, S. (1985), Shifts in selective visual attention: towards the underlying neural circuitry, *Hum. Neurobiol.*, **4**: 219-27.

[28] Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995), The role of attention in the programming of saccades. *Vision Res.*, **35**: 1897-916.

[29] Kustov, A. A., and Robinson, D. L. (1996), Shared neural control of attentional shifts and eye movements, *Nature*, **384**: 74-7.

[30] Lee, D. K., Koch, C., and Braun, J.(1997), Visual attention is undifferentiated also for less demanding tasks, *Invest. Ophth. Vis. Sci.*, **40**: 281B241.

[31] Lee D. K., Koch C., and Braun J. (1999), Attentional capacity is undifferentiated: concurrent discrimination of form, color, and motion, *Percept. Psychophys.*, **61**: 1241-1255.

[32] Miller, E. K. (2000) The prefrontal cortex and cognitive control, *Nat. Reviews Neurosci.*, **1**: 59-65.

[33] Moran, J., and Desimone, R. (1985), Selective attention gates visual processing in the extrastriate cortex, *Science*, **229**: 782-4.

[34] Motter, B. C. (1994), Neural correlates of attentive selection for color or luminance in extrastriate area V4, *J. Neurosci.*, **14**: 2178-89.

[35] Nakayama, K., and Mackeben, M. (1989), Sustained and transient components of focal visual attention, *Vision Res.*, **29**: 1631-1647.

[36] Niebur, E., Koch., C. and Rosin, C. (1993), An oscillation-based model for the neuronal basis of attention, *Vision Res.*, **33**: 2789-802.

[37] Nobre, A. C., Gitelman, D. R., Dias, E. C., and Mesulam, M. M. (2000), Covert visual spatial orienting and saccades: overlapping neural systems, *Neuroimage*, **11**: 210-216.

[38] Noton, D., and Stark, L. (1971), Scanpaths in eye movements during pattern perception, *Science*, **171**: 308-11.

[39] O'Regan, J. K., Rensink, R. A., and Clark, J. J. (1999), Change-blindness as a result of mudsplashes, *Nature*, **398**: 34.

[40] Rensink R.A. (2000), The dynamic representation of scenes, *Vis. Cogn.*, **7**: 17-42

[41] Rensink R.A. (2002), Change detection, *Annu. Rev. Psychol.*, **53**: 245-277.

[42] Reynolds, J. H., and Desimone, R. (1999), The role of neural mechanisms of attention in solving the binding problem, *Neuron*, **24**: 19-29, 111-25.

[43] Reynolds, J. H., Pasternak, T., and Desimone, R. (2000), Attention increases sensitivity of V4 neurons [see comments], *Neuron*, **26**: 703-714.

[44] Rybak, I. A., Gusakova, V. I., Golovan, A. V., Podladchikova, L. N., and Shevtsova, N. A. (1998), A model of attention-guided visual perception and recognition, *Vision Res.*, **38**: 2387-2400.

[45] Schall, J. D., Hanes, D. P., and Taylor, T. L. (2000), Neural control of behavior:

countermanding eye movements, *Psychol. Res.*, **63**: 299-307.

[46] Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., and Zetzsche, C. (2001), Scene analysis with saccadic eye movements: top-down and bottom-up modeling, *J. Electronic Imaging*, **10**: 152-160.

[47] Shepherd, M., Findlay, J. M., and Hockey, R. J. (1986), The relationship between eye movements and spatial attention, *Q. J. Exp. Psychol.*, **38**: 475-491.

[48] Sheliga, B. M., Riggio, L., and Rizzolatti, G. (1994) Orienting of attention and eye movements *Exp. Brain Res.*, **98**: 507-22.

[49] Stark, L. W., and Choi, Y. S. (1996), Experimental methaphysics: the scanpath as an epistemological mechanism, 3-69, in Zangemeister, W. H. and Stiehl, H. S. and Freska, C. (eds.), *Visual Attention and Cognition* Elsevier Science B.V.

[50] Stark, L. W., Privitera, C. M., Yang, H., Azzariti, M., Ho, Y. F., Blackmon, T., and Chernyak, D.(2001), Representation of human vision in the brain: how does human perception recognize images?, *Journal of Electronic Imaging*, **10**: 127-146.

[51] Suder, K., and Worgotter, F. (2000), The control of low-level information flow in the visual system, *Rev. Neurosci.*, **11**: 127-146.

[52] Treisman, A. M., and Gelade, G. (1980), A feature-integration theory of attention, *Cognit. Psychol.*, **12**: 97-136,

[53] Treisman, A. (1988), Features and objects: the fourteenth Bartlett memorial lecture, *Q. J. Exp. Psychol. [A]*, **40**: 201-37.

[54] Treue, S., and Maunsell, J. H. (1996), Attentional modulation of visual motion processing in cortical areas MT and MST, *Nature*, **382**: 539-41.

[55] Treue, S., and Martinez Trujillo, J. C.(1999), Feature-based attention influences motion processing gain in macaque visual cortex, *Nature*, **399**: 575-579.

[56] Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., and Nuflo, F. (1995), Modeling visual-attention via selective tuning, *Artificial Intelligence*, **78**: 507-45.

[57] Ungerleider, L. G. and Mishkin, M. (1982), Two cortical visual systems, 549-586, in Ingle, D. G. and Goodale, M. A. A. and Mansfield, R. J. W. (eds.), *Analysis of Visual Behavior* MIT Press: Cambridge, MA.

[58] Weichselgartner, E., and Sperling, G. (1987), Dynamics of automatic and controlled visual attention, *Science*, **238**: 778-780.

[59] Wolfe, J. M. (1994), Visual search in continuous, naturalistic stimuli, *Vision Res.*, **34**: 1187-95.

[60] Wolfe, J. (1996), Visual search: a review, in Pashler, H (ed.), *Attention.* University College London Press, London, U.K.

[61] Yeshurun, Y. and Carrasco, M. (1998), Attention improves or impairs visual

performance by enhancing spatial resolution, *Nature*, **396**: 72-75.

[62] Yarbus, A. (1967), *Eye Movements and Vision*, Plenum Press: New York.