Spiking Neuron Models

Single Neurons, Populations, Plasticity

Wulfram Gerstner and Werner M. Kistler

Cambridge University Press, 2002

Preface

The task of understanding the principles of information processing in the brain poses, apart from numerous experimental questions, challenging theoretical problems on all levels from molecules to behavior. This books concentrates on modeling approaches on the level of neurons and small populations of neurons, since we think that this is an appropriate level to adress fundamental questions of neuronal coding, signal transmission, or synaptic plasticity. In this text we concentrate on theoretical concepts and phenomenological models derived from them. We think of a neuron primarily as a dynamic element that emits output pulses whenever the excitation exceeds some threshold. The resulting sequence of pulses or 'spikes' contains all the information that is transmitted from one neuron to the next. In order to understand signal transmission and signal processing in neuronal systems, we need an understanding of their basic elements, i.e., the neurons, which is the topic of part I. New phenomena emerge when several neurons are coupled. Part II introduces network concepts, in particular pattern formation, collective excitations, and rapid signal transmission between neuronal populations. Learning concepts presented in Part III are based on spike-time dependent synaptic plasticity.

We wrote this book as an introduction to spiking neuron models for advanced undergraduate or graduate students. It can be used either as the main text for a course that focuses on neuronal dynamics; or as part of a larger course in Computational Neuroscience, theoretical biology, neuronal modeling, biophysics, or neural networks. For a one-semester course on neuronal modeling, we usually teach one chapter per week focusing on the first sections of each chapter for lectures and give the remainder as reading assignment. Many of the examples can be adapted to become exercises or projects for students. While writing the book we had in mind students of physics, mathematics, or computer science with an interest in biology; but it might also be useful for students of biology who are interested in mathematical modeling. All the necessary mathematical concepts are introduced in an elementary fashion, and we have provided many illustrative figures which complement the mathematical analysis and help the reader picture what is going on. No prior knowledge beyond undergraduate mathematics should be necessary to read the book. An asterix (*) marks those sections that have a more mathematical focus. These sections can be skipped at a first reading.

We have also tried to keep the book self-contained with respect to the underlying Neurobiology. The fundamentals of neuronal excitation and synaptic signal transmission are briefly introduced in Chapter 1 together with an outlook on the principal topics of the book, viz., formal spiking neuron models and the problem of neuronal coding. In Chapter 2 we review biophysical models, such as the Hodgkin-Huxley equations of neuronal dynamics, and models of dendritic integration based on the cable equation. These are the starting point for a systematic reduction to neuron models with a reduced complexity that are open to an analytical treatment. Whereas Chapter 3 is dedicated to two-dimensional differential equations as a description of neuronal dynamics, Chapter 4 introduces formal spiking neuron models, namely the integrate-and-fire model and the Spike Response Model. These formal neuron models are the foundation for all the following chapters. Part I on "Single Neuron Models" is rounded off by Chapter 5 which

gives an overview of spike-train statistics and illustrates how noise can be implemented in spiking neuron models.

The step from single neuron models to networks of neurons is taken in Chapter 6 where equations for the macroscopic dynamics of large populations of neurons are derived. Based on these equations phenomena like signal transmission and coding (Chapter 7), oscillations and synchrony (Chapter 8), and pattern formation in spatially structured networks (Chapter 9) are investigated. Up to this point, only networks with a fixed synaptic connectivity have been discussed. The third part of the book, finally, deals with synaptic plasticity and its role for development, learning, and memory. In Chapter 10, principles of Hebbian plasticity are presented and various models of synaptic plasticity are described that are more or less directly inspired by neurbiological findings. Equations that relate the synaptic weight dynamics to statistical properties of the neuronal spike activity are derived in Chapter 11. Last but not least, Chapter 12 presents an – admittedly personal – choice of illustrative applications of spike-timing dependent synaptic plasticity to fundamental problems of neuronal coding.

While the book contains material which is now considered as standard for courses of the type mentioned earlier, it also provides a bridge to current research which has developed over the last few years. In most chapters, the reader will find some sections which either report recent results or shed new light on well-known models. The viewpoint taken in the presentation of the material is of course highly subjective and a bias towards our own research is obvious. Nevertheless, we hope that the book will find the interest of students and researchers in the field.

Werner M. Kistler
Wulfram Gerstner
Lausanne, November 2001

Acknowledgments

This book would not have been possible without the help of our friends and colleagues who made many valuable comments and spotted numerous mistakes. In particular, we would like to thank Silvio Borer, Nicolas Brunel, Stefano Fusi, Evan Haskell, Andreas Herz, Renaud Jolivet, Maurizio Mattia, Jean-Pascal Pfister, Hans-Ekkehard Plesser, Marc van Rossum, Walter Senn, and Arnaud Tonnelier for a careful reading of the text. We owe special thanks to Leo van Hemmen who encouraged us to write a book on our joint work. Even though the scope of the text has changed over the years, the original initiative is due to him. Finally, we thank all the students who attended tutorials or classes taught by us and whose critical comments helped us to organize our thoughts.

W.G. and W.M.K.

Contents

Chapter 1

Introduction

The aim of this chapter is to introduce several elementary notions of neuroscience, in particular the concepts of action potentials, postsynaptic potentials, firing thresholds, and refractoriness. Based on these notions a first phenomenological model of neuronal dynamics is built that will be used as a starting point for a discussion of neuronal coding. Due to the limitations of space we cannot – and do not want to – give a comprehensive introduction into such a complex field as neurobiology. The presentation of the biological background in this chapter is therefore highly selective and simplistic. For an in-depth discussion of neurobiology we refer the reader to the literature mentioned at the end of this chapter. Nevertheless, we try to provide the reader with a minimum of information necessary to appreciate the biological background of the theoretical work presented in this book.

1.1 Elements of Neuronal Systems

Over the past hundred years, biological research has accumulated an enormous amount of detailed knowledge about the structure and function of the brain. The elementary processing units in the central nervous system are neurons which are connected to each other in an intricate pattern. A tiny portion of such a network of neurons is sketched in Fig. 1.1 which shows a drawing by Ramón y Cajal, one of the pioneers of neuroscience around 1900. We can distinguish several neurons with triangular or circular cell bodies and long wire-like extensions. This picture can only give a glimpse of the network of neurons in the cortex. In reality, cortical neurons and their connections are packed into a dense network with more than 10^4 cell bodies and several kilometers of 'wires' per cubic millimeter. In other areas of the brain the wiring pattern may look different. In all areas, however, neurons of different sizes and shapes form the basic elements.

The cortex does not consist exclusively of neurons. Beside the various types of neuron there is a large number of 'supporter' cells, so-called glia cells, that

Fig. 1.1: This reproduction of a drawing of Ramón y Cajal shows a few neurons in the mammalian cortex that he observed under the microscope. Only a small portion of the neurons contained in the sample of cortical tissue have been made visible by the staining procedure; the density of neurons is in reality much higher. Cell b is a nice example of a pyramidal cell with a triangularly shaped cell body. Dendrites, which leave the cell laterally and upwards, can be recognized by their rough surface. The axons are recognizable as thin, smooth lines which extend downwards with a few branches to the left and right. From Ramòn y Cajal (1909).

are required for energy supply and structural stabilization of brain tissue. Since glia cells are not directly involved in information processing, we will not discuss them any further. We will also neglect a few rare subtypes of neuron, such as analog neurons in the mammalian retina. Throughout this book we concentrate on spiking neurons only.

1.1.1 The Ideal Spiking Neuron

A typical neuron can be divided into three functionally distinct parts, called dendrites, soma, and axon; see Fig. 1.2. Roughly speaking, the dendrites play the role of the 'input device' that collects signals from other neurons and transmits them to the soma. The soma is the 'central processing unit' that performs an important non-linear processing step: If the total input exceeds a certain threshold, then an output signal is generated. The output signal is taken over by the 'output device', the axon, which delivers the signal to other neurons.

The junction between two neurons is called a synapse. Let us suppose that a neuron sends a signal across a synapse. It is common to refer to the sending neuron as the presynaptic cell and to the receiving neuron as the postsynaptic cell. A single neuron in vertebrate cortex often connects to more than 10^4 postsynaptic neurons. Many of its axonal branches end in the direct neighborhood

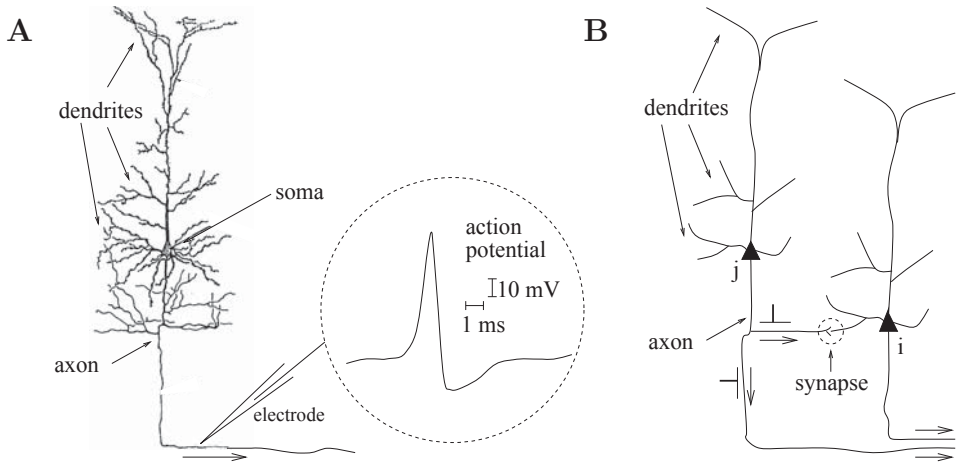

Fig. 1.2: **A**. Single neuron in a drawing by Ramón y Cajal. Dendrite, soma, and axon can be clearly distinguished. The inset shows an example of a neuronal action potential (schematic). The action potential is a short voltage pulse of 1-2 ms duration and an amplitude of about 100 mV. **B**. Signal transmission from a presynaptic neuron j to a postsynaptic neuron i. The synapse is marked by the dashed circle. The axons at the lower right end lead to other neurons (schematic figure).

of the neuron, but the axon can also stretch over several centimeters so as to reach to neurons in other areas of the brain.

1.1.2 Spike Trains

The neuronal signals consist of short electrical pulses and can be observed by placing a fine electrode close to the soma or axon of a neuron; see Fig. 1.2. The pulses, so-called action potentials or spikes, have an amplitude of about 100 mV and typically a duration of 1-2 ms. The form of the pulse does not change as the action potential propagates along the axon. A chain of action potentials emitted by a single neuron is called a spike train – a sequence of stereotyped events which occur at regular or irregular intervals. Since all spikes of a given neuron look alike, the form of the action potential does not carry any information. Rather, it is the number and the timing of spikes which matter. The action potential is the elementary unit of signal transmission.

Action potentials in a spike train are usually well separated. Even with very strong input, it is impossible to excite a second spike during or immediately after a first one. The minimal distance between two spikes defines the absolute refractory period of the neuron. The absolute refractory period is followed by a phase of relative refractoriness where it is difficult, but not impossible to excite an action potential.

1.1.3 Synapses

The site where the axon of a presynaptic neuron makes contact with the dendrite (or soma) of a postsynaptic cell is the synapse. The most common type of synapse in the vertebrate brain is a chemical synapse. At a chemical synapse, the axon terminal comes very close to the postsynaptic neuron, leaving only a tiny gap between pre- and postsynaptic cell membrane, called the synaptic cleft. When an action potential arrives at a synapse, it triggers a complex chain of bio-chemical processing steps that lead to a release of neurotransmitter from the presynaptic terminal into the synaptic cleft. As soon as transmitter molecules have reached the postsynaptic side, they will be detected by specialized receptors in the postsynaptic cell membrane and open (either directly or via a biochemical signaling chain) specific channels so that ions from the extracellular fluid flow into the cell. The ion influx, in turn, leads to a change of the membrane potential at the postsynaptic site so that, in the end, the chemical signal is translated into an electrical response. The voltage response of the postsynaptic neuron to a presynaptic action potential is called the postsynaptic potential.

Apart from chemical synapses neurons can also be coupled by electrical synapses, so-called gap junctions. Specialized membrane proteins make a direct electrical connection between the two neurons. Not very much is known about the functional aspects of gap junctions, but they are thought to be involved in the synchronization of neurons.

1.2 Elements of Neuronal Dynamics

The effect of a spike on the postsynaptic neuron can be recorded with an intracellular electrode which measures the potential difference $u(t)$ between the interior of the cell and its surroundings. This potential difference is called the membrane potential. Without any spike input, the neuron is at rest corresponding to a constant membrane potential. After the arrival of a spike, the potential changes and finally decays back to the resting potential, cf. Fig. 1.3A. If the change is positive, the synapse is said to be excitatory. If the change is negative, the synapse is inhibitory.

At rest, the cell membrane has already a strong negative polarization of about -65 mV. An input at an excitatory synapse reduces the negative polarization of the membrane and is therefore called depolarizing. An input that increases the negative polarization of the membrane even further is called hyperpolarizing.

1.2.1 Postsynaptic Potentials

Let us formalize the above observation. We study the time course $u_i(t)$ of the membrane potential of neuron i. Before the input spike has arrived, we have

$u_i(t) = u_{\rm rest}$. At $t = 0$ the presynaptic neuron j fires its spike. For $t > 0$, we see at the electrode a response of neuron i

$$u_i(t) - u_{\rm rest} =: \epsilon_{ij}(t)\,. \tag{1.1}$$

The right-hand side of Eq. (1.1) defines the postsynaptic potential (PSP). If the voltage difference $u_i(t) - u_{\rm rest}$ is positive (negative) we have an excitatory (inhibitory) postsynaptic potential or short EPSP (IPSP). In Fig. 1.3A we have sketched the EPSP caused by the arrival of a spike from neuron j at an excitatory synapse of neuron i.

1.2.2 Firing Threshold and Action Potential

Consider two presynaptic neurons $j = 1, 2$, which both send spikes to the postsynaptic neuron i. Neuron $j = 1$ fires spikes at $t_1^{(1)}, t_1^{(2)}, \ldots$, similarly neuron $j = 2$ fires at $t_2^{(1)}, t_2^{(2)}, \ldots$. Each spike evokes a postsynaptic potential ϵ_{i1} or ϵ_{i2}, respectively. As long as there are only few input spikes, the total change of the potential is approximately the sum of the individual PSPs,

$$u_i(t) = \sum_j \sum_f \epsilon_{ij}(t - t_j^{(f)}) + u_{\rm rest}\,, \tag{1.2}$$

i.e., the membrane potential responds linearly to input spikes; see Fig. 1.3B.

On the other hand, linearity breaks down if too many input spikes arrive during a short interval. As soon as the membrane potential reaches a critical value ϑ, its trajectory shows a behavior that is quite different from a simple summation of PSPs: The membrane potential exhibits a pulse-like excursion with an amplitude of about 100 mV, viz., an action potential. This action potential will propagate along the axon of neuron i to the synapses of other neurons. After the pulse the membrane potential does not directly return to the resting potential, but passes through a phase of hyperpolarization below the resting value. This hyperpolarization is called 'spike-afterpotential'.

Single EPSPs have amplitudes in the range of one millivolt. The critical value for spike initiation is about 20 to 30 mV above the resting potential. In most neurons, four spikes – as shown schematically in Fig. 1.3C – are thus not sufficient to trigger an action potential. Instead, about 20-50 presynaptic spikes have to arrive within a short time window before postsynaptic action potentials are triggered.

1.3 A Phenomenological Neuron Model

In order to build a phenomenological model of neuronal dynamics, we describe the critical voltage for spike initiation by a formal threshold ϑ. If $u_i(t)$ reaches ϑ

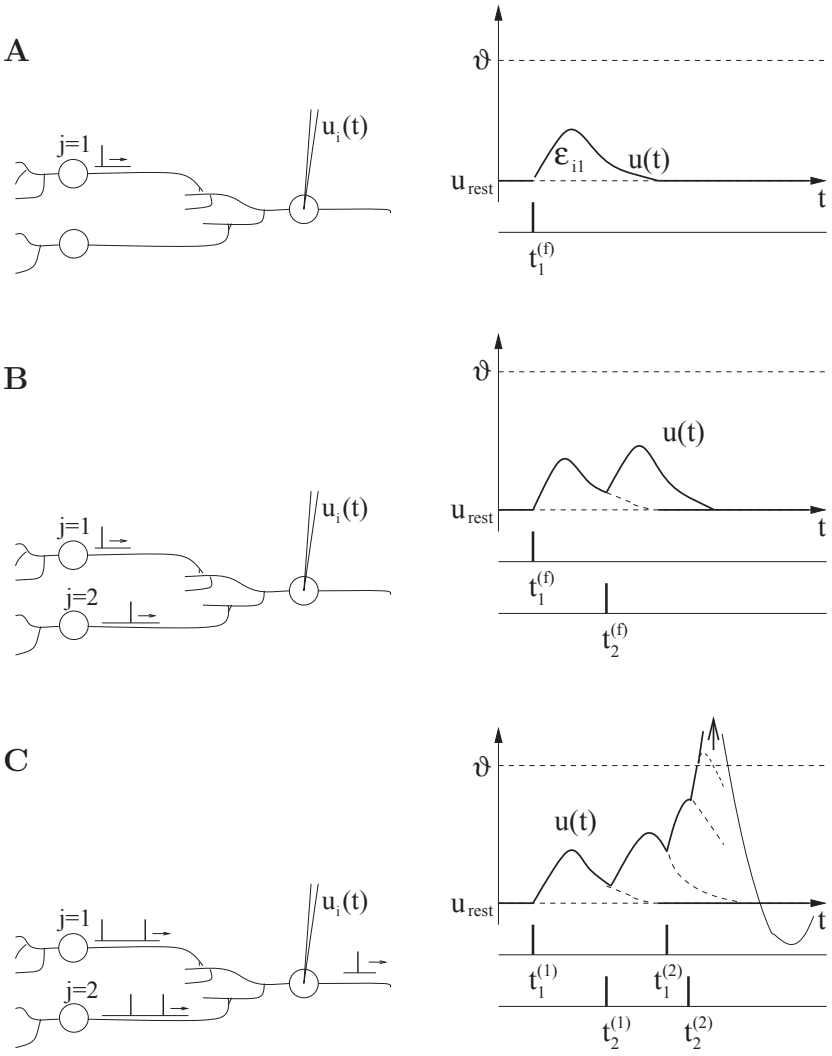

Fig. 1.3: A postsynaptic neuron i receives input from two presynaptic neurons $j = 1, 2$. **A**. Each presynaptic spike evokes an excitatory postsynaptic potential (EPSP) that can be measured with an electrode as a potential difference $u_i(t) - u_{\rm rest}$. The time course of the EPSP caused by the spike of neuron $j = 1$ is $\epsilon_{i1}(t - t_1^{(f)})$. **B**. An input spike from a second presynaptic neuron $j = 2$ that arrives shortly after the spike from neuron $j = 1$, causes a second postsynaptic potential that adds to the first one. **C**. If $u_i(t)$ reaches the threshold ϑ, an action potential is triggered. As a consequence, the membrane potential starts a large positive pulse-like excursion (arrow). On the voltage scale of the graph, the peak of the pulse is out of bounds. After the pulse the voltage returns to a value below the resting potential.

from below we say that neuron i fires a spike. The moment of threshold crossing defines the firing time $t_i^{(f)}$. The model makes use of the fact that action potentials always have roughly the same form. The trajectory of the membrane potential during a spike can hence be described by a certain standard time course denoted by $\eta(t - t_i^{(f)})$.

1.3.1 Definition of the Model SRM$_0$

Putting all elements together we have the following description of neuronal dynamics. The variable u_i describes the momentary value of the membrane potential of neuron i. It is given by

$$u_i(t) = \eta(t - \hat{t}_i) + \sum_j \sum_f \epsilon_{ij}(t - t_j^{(f)}) + u_{\rm rest} \tag{1.3}$$

where $\hat{t}_i$ is the last firing time of neuron i, i.e., $\hat{t}_i = \max\{t_i^{(f)} \,|\, t_i^{(f)} < t\}$. Firing occurs whenever u_i reaches the threshold ϑ from below,

$$u_i(t) = \vartheta \text{ and } \frac{\mathrm{d}}{\mathrm{d}t} u_i(t) > 0 \quad \Longrightarrow \quad t = t_i^{(f)} \tag{1.4}$$

The term ϵ_{ij} in (1.3) describes the response of neuron i to spikes of a presynaptic neuron j. The term η in (1.3) describes the form of the spike and the spike-afterpotential.

Note that we are only interested in the potential *difference*, viz., the distance from the resting potential. By an appropriate shift of the voltage scale, we can always set $u_{\rm rest} = 0$. The value of $u(t)$ is then directly the distance from the resting potential. This is implicitly assumed in most neuron models discussed in this book.

The model defined in equations (1.3) and (1.4) is called SRM$_0$ where SRM is short for Spike Response Model (Gerstner, 1995). The subscript zero is intended to remind the reader that it is a particularly simple 'zero order' version of the full model that will be introduced in Chapter 4. Phenomenological models of spiking neurons similar to the models SRM$_0$ have a long tradition in theoretical neuroscience (Hill, 1936; Stein, 1965; Geisler and Goldberg, 1966; Weiss, 1966). Some important limitations of the model SRM$_0$ are discussed below in Section 1.3.2. Despite the limitations, we hope to be able to show in the course of this book that spiking neuron models such as the Spike Response Model are a useful conceptual framework for the analysis of neuronal dynamics and neuronal coding.

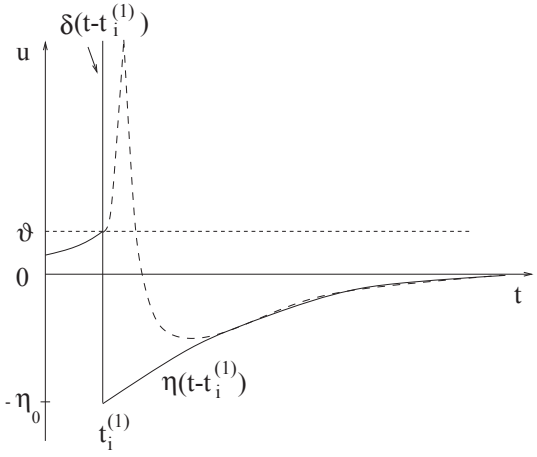

Fig. 1.4: In formal models of spiking neurons the shape of an action potential (dashed line) is usually replaced by a δ pulse (vertical line). The negative overshoot (spike-afterpotential) after the pulse is included in the kernel $\eta(t-t_i^{(1)})$ (thick line) which takes care of 'reset' and 'refractoriness'. The pulse is triggered by the threshold crossing at $t_i^{(1)}$. Note that we have set $u_{\rm rest}=0$.

Example: Formal pulses

In a simple model, we may replace the exact form of the trajectory η during an action potential by, e.g., a square pulse, followed by a negative spike-afterpotential,

$$\eta(t-t_i^{(f)}) = \begin{cases} 1/\Delta t & \text{for } 0 < t-t_i^{(f)} < \Delta t \\ -\eta_0 \exp\left(-\frac{t-t_i^{(f)}}{\tau}\right) & \text{for } \Delta t < t-t_i^{(f)} \end{cases} \tag{1.5}$$

with parameters $\eta_0, \tau, \Delta t > 0$. In the limit of $\Delta t \to 0$ the square pulse approaches a Dirac δ function; see Fig. 1.4.

The positive pulse marks the moment of spike firing. For the purpose of the model, it has no real significance, since the spikes are recorded explicitly in the set of firing times $t_i^{(1)}, t_i^{(2)}, \ldots$. The negative spike-afterpotential, however, has an important implication. It leads after the pulse to a 'reset' of the membrane potential to a value below threshold. The idea of a simple reset of the variable u_i after each spike is one of the essential components of the integrate-and-fire model that will be discussed in detail in Chapter 4.

If $\eta_0 \gg \vartheta$ then the membrane potential after the pulse is significantly lower than the resting potential. The emission of a second pulse immediately after the first one is therefore more difficult, since many input spikes are needed to reach the threshold. The negative spike-after potential in Eq. (1.5) is thus a simple model of neuronal refractoriness.

Example: Formal spike trains

Throughout this book, we will refer to the moment when a given neuron emits an action potential as the firing time of that neuron. In models, the firing time is usually defined as the moment of threshold crossing. Similarly, in experiments firing times are recorded when the membrane potential reaches some threshold value u_ϑ from below. We denote firing times of neuron i by $t_i^{(f)}$ where $f = 1, 2, \ldots$ is the label of the spike. Formally, we may denote the spike train of a neuron i as the sequence of firing times

$$S_i(t) = \sum_f \delta(t - t_i^{(f)}) \tag{1.6}$$

where $\delta(x)$ us the Dirac δ function with $\delta(x) = 0$ for $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x) \mathrm{d}x = 1$. Spikes are thus reduced to points in time.

1.3.2 Limitations of the Model

The model presented in Section 1.3.1 is highly simplified and neglects many aspects of neuronal dynamics. In particular, all postsynaptic potentials are assumed to have the same shape, independently of the state of the neuron. Furthermore, the dynamics of neuron i depends only on its most recent firing time $\hat{t}_i$. Let us list the major limitations of this approach.

(i) Adaptation, Bursting, and Inhibitory Rebound

To study neuronal dynamics experimentally, neurons can be isolated and stimulated by current injection through an intracellular electrode. In a standard experimental protocol we could, for example, impose a stimulating current that is switched at time t_0 from a value I_1 to a new value I_2. Let us suppose that $I_1 = 0$ so that the neuron is quiescent for $t < t_0$. If the current I_2 is sufficiently large, it will evoke spikes for $t > t_0$. Most neurons will respond to the current step with a spike train where intervals between spikes increase successively until a steady state of periodic firing is reached; cf. Fig. 1.5A. Neurons that show this type of adaptation are called regularly-firing neurons (Connors and Gutnick, 1990). Adaptation is a slow process that builds up over several spikes. Since the model SRM_0 takes only the most recent spike into account, it cannot capture adaptation. Detailed neuron models which will be discussed in Chapter 2 describe the slow processes that lead to adaptation explicitly. To mimic adaptation with formal spiking neuron models we would have to add up the contributions to refractoriness of several spikes back in the past; cf. Chapter 4.

A second class of neurons are fast-spiking neurons. These neurons show now adaptation and can therefore be well approximated by the model SRM_0 introduced in Section 1.3.1. Many inhibitory neurons are fast-spiking neurons. Apart from regular-spiking and fast-spiking neurons, there are also bursting neurons

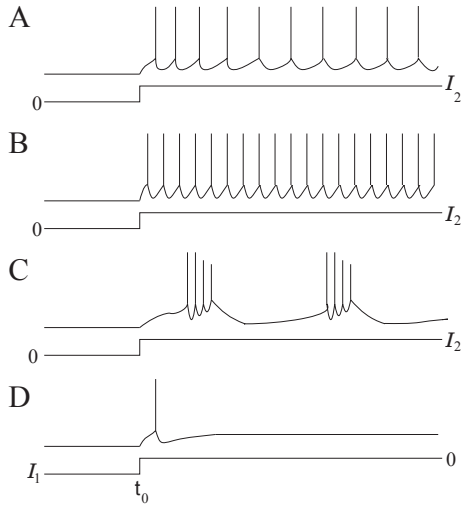

Fig. 1.5: Response to a current step. In **A** - **C**, the current is switched on at $t = t_0$ to a value $I_2 > 0$. Regular-spiking neurons (**A**) exhibit adaptation of the interspike intervals whereas fast-spiking neurons (**B**) show no adaptation. An example of a bursting neuron is shown in **C**. Many neurons emit an inhibitory rebound spike (**D**) after an inhibitory current $I_1 < 0$ is switched off. Schematic figure.

which form a separate group (Connors and Gutnick, 1990). These neurons respond to constant stimulation by a sequence of spikes that is periodically interrupted by rather long intervals; cf. Fig. 1.5C. Again, a neuron model that takes only the most recent spike into account cannot describe bursting. For a review of bursting neuron models, the reader is referred to (Izhikevich, 2000).

Another frequently observed behavior is post-inhibitory rebound. Consider a step current with $I_1 < 0$ and $I_2 = 0$, i.e., an inhibitory input that is switched off at time t_0; cf. Fig. 1.5D. Many neurons respond to such a change with one or more 'rebound spikes': Even the release of inhibition can trigger action potentials. We will return to inhibitory rebound in Chapter 2.

(ii) Saturating excitation and shunting inhibition

In the model SRM_0 introduced in Section 1.3.1, the form of a postsynaptic potential generated by a presynaptic spike at time $t_j^{(f)}$ does not depend on the state of the postsynaptic neuron i. This is of course a simplification and reality is somewhat more complicated. In Chapter 2 we will discuss detailed neuron models that describe synaptic input as a change of the membrane conductance. Here we simply summarize the major phenomena.

In Fig. 1.6 we have sketched schematically an experiment where the neuron is driven by a constant current I_0. We assume that I_0 is too weak to evoke firing so that, after some relaxation time, the membrane potential settles at a constant value u_0. At $t = t^{(f)}$ a presynaptic spike is triggered. The spike generates a current

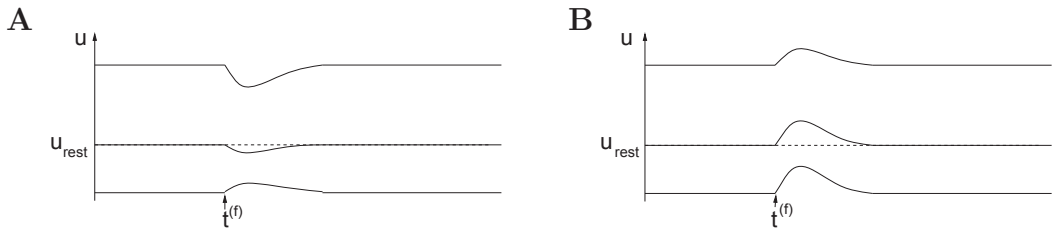

Fig. 1.6: The shape of postsynaptic potentials depends on the momentary level of depolarization. **A**. A presynaptic spike that arrives at time $t^{(f)}$ at an inhibitory synapse has hardly any effect on the membrane potential when the neuron is at rest, but a large effect if the membrane potential u is above the resting potential. If the membrane is hyperpolarized below the reversal potential of the inhibitory synapse, the response to the presynaptic input changes sign. **B**. A spike at an excitatory synapse evokes a postsynaptic potential with an amplitude that depends only slightly on the momentary voltage u. For large depolarizations the amplitude becomes smaller (saturation). Schematic figure.

pulse at the postsynaptic neuron (postsynaptic current, PSC) with amplitude

$$\text{PSC} \propto u_0 - E_{\text{syn}} \tag{1.7}$$

where u_0 is the membrane potential and E_{syn} is the 'reversal potential' of the synapse. Since the amplitude of the current input depends on u_0, the response of the postsynaptic potential does so as well. Reversal potentials are systematically introduced in Chapter 2.2; models of synaptic input are discussed in Chapter 2.4.

Example: Shunting Inhibition and Reversal Potential

The dependence of the postsynaptic response upon the momentary state of the neuron is most pronounced for inhibitory synapses. The reversal potential of inhibitory synapses E_{syn} is below, but usually close to the resting potential. Input spikes thus have hardly any effect on the membrane potential if the neuron is at rest; cf. 1.6a. However, if the membrane is depolarized, the very same input spikes evoke a nice inhibitory postsynaptic potentials. If the membrane is already hyperpolarized, the input spike can even produce a depolarizing effect. There is a intermediate value $u_0 = E_{\text{syn}}$ – the reversal potential – where the response to inhibitory input 'reverses' from hyperpolarizing to depolarizing.

Though inhibitory input usually has only a small impact on the membrane potential, the local conductivity of the cell membrane can be significantly increased. Inhibitory synapses are often located on the soma or on the shaft of the dendritic tree. Due to their strategic position a few inhibitory input spikes can 'shunt' the whole input that is gathered by the dendritic tree from hundreds of excitatory synapses. This phenomenon is called 'shunting inhibition'.

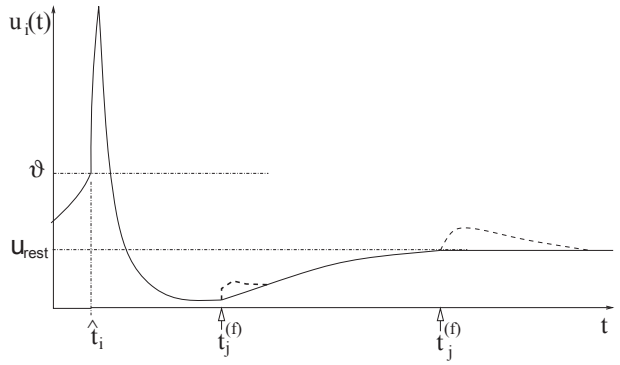

Fig. 1.7: The shape of postsynaptic potentials (dashed lines) depends on the time $t - \hat{t}_i$ that has passed since the last output spike current if neuron i. The postsynaptic spike has been triggered at time $\hat{t}_i$. A presynaptic spike that arrives at time $t_j^{(f)}$ shortly after the spike of the postsynaptic neuron has a smaller effect than a spike that arrives much later. The spike arrival time is indicated by an arrow. Schematic figure.

The reversal potential for excitatory synapses is usually significantly above the resting potential. If the membrane is depolarized $u_0 \gg u_{\rm rest}$ the amplitude of an excitatory postsynaptic potential is reduced, but the effect is not as pronounced as for inhibition. For very high levels of depolarization a saturation of the EPSPs can be observed; cf. 1.6b.

Example: Conductance Changes after a Spike

The shape of the postsynaptic potentials does not only depend on the level of depolarization but, more generally, on the internal state of the neuron, e.g., on the timing relative to previous action potentials.

Suppose that an action potential has occured at time $\hat{t}_i$ and that a presynaptic spike arrives at a time $t_j^{(f)} > \hat{t}_i$. The form of the postsynaptic potential depends now on the time $t_j^{(f)} - \hat{t}_i$; cf. Fig. 1.7. If the presynaptic spike arrives during or shortly after a postsynaptic action potential it has little effect because some of the ion channels that were involved in firing the action potential are still open. If the input spike arrives much later it generates a postsynaptic potential of the usual size. We will return to this effect in Chapter 2.2.

Example: Spatial Structure

The form of postsynaptic potentials also depends on the location of the synapse on the dendritic tree. Synapses that are located at the distal end of the dendrite are expected to evoke a smaller postsynaptic response at the soma than a synapse that is located directly on the soma; cf. Chapter 2. If several inputs occur on the

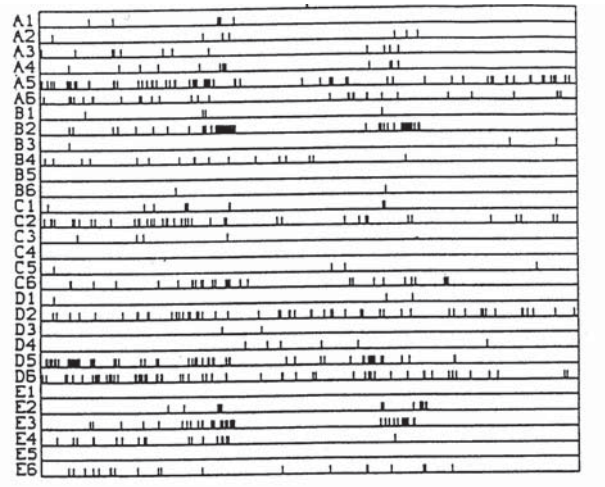

Fig. 1.8: Spatio-temporal pulse pattern. The spikes of 30 neurons (A1-E6, plotted along the vertical axes) are shown as a function of time (horizontal axis, total time is 4 000 ms). The firing times are marked by short vertical bars. From Krüger and Aiple (1988).

same dendritic branch within a few milliseconds, the first input will cause local changes of the membrane potential that influence the amplitude of the response to the input spikes that arrive slightly later. This may lead to saturation or, in the case of so-called 'active' currents, to an enhancement of the response. Such nonlinear interactions between different presynaptic spikes are neglected in the model SRM_0. A purely linear dendrite, on the other hand, can be incorporated in the model as we will see in Chapter 4.

1.4 The Problem of Neuronal Coding

The mammalian brain contains more than 10^{10} densely packed neurons that are conected to an intricate network. In every small volume of cortex, thousands of spikes are emitted each millisecond. An example of a spike train recording from thirty neurons is shown in Fig. 1.8. What is the information contained in such a spatio-temporal pattern of pulses? What is the code used by the neurons to transmit that information? How might other neurons decode the signal? As external observers, can we read the code and understand the message of the neuronal activity pattern?

The above questions point to the problem of neuronal coding, one of the fundamental issues in neuroscience. At present, a definite answer to these questions is not known. Traditionally it has been thought that most, if not all, of the relevant information was contained in the mean firing rate of the neuron. The firing rate is usually defined by a temporal average; see Fig. 1.9. The experimentalist

sets a time window of, say $T = 100\,\text{ms}$ or $T = 500\,\text{ms}$ and counts the number of spikes $n_{\text{sp}}(T)$ that occur in this time window. Division by the length of the time window gives the mean firing rate

$$\nu = \frac{n_{\text{sp}}(T)}{T} \tag{1.8}$$

usually reported in units of s^{-1} or Hz.

The concept of mean firing rates has been successfully applied during the last 80 years. It dates back to the pioneering work of Adrian (Adrian, 1926, 1928) who showed that the firing rate of stretch receptor neurons in the muscles is related to the force applied to the muscle. In the following decades, measurement of firing rates became a standard tool for describing the properties of all types of sensory or cortical neurons (Mountcastle, 1957; Hubel and Wiesel, 1959), partly due to the relative ease of measuring rates experimentally. It is clear, however, that an approach based on a temporal average neglects all the information possibly contained in the exact timing of the spikes. It is therefore no surprise that the firing rate concept has been repeatedly criticized and is subject of an ongoing debate (Abeles, 1994; Bialek et al., 1991; Hopfield, 1995; Shadlen and Newsome, 1994; Softky, 1995; Rieke et al., 1996; Oram et al., 1999).

During recent years, more and more experimental evidence has accumulated which suggests that a straightforward firing rate concept based on temporal averaging may be too simplistic to describe brain activity. One of the main arguments is that reaction times in behavioral experiments are often too short to allow long temporal averages. Humans can recognize and respond to visual scenes in less than 400 ms (Thorpe et al., 1996). Recognition and reaction involve several processing steps from the retinal input to the finger movement at the output. If, at each processing step, neurons had to wait and perform a temporal average in order to read the message of the presynaptic neurons, the reaction time would be much longer.

In experiments on a visual neuron in the fly, it was possible to 'read the neural code' and reconstruct the time-dependent stimulus based on the neuron's firing times (Bialek et al., 1991). There is evidence of precise temporal correlations between pulses of different neurons (Abeles, 1994; Lestienne, 1996) and stimulus dependent synchronization of the activity in populations of neurons (Eckhorn et al., 1988; Gray and Singer, 1989; Gray et al., 1989; Engel et al., 1991a; Singer, 1994). Most of these data are inconsistent with a naïve concept of coding by mean firing rates where the exact timing of spikes should play no role.

In the following sections, we review some potential coding schemes and ask: What exactly is a pulse code – and what is a rate code? The question of neuronal coding has important implications for modeling, because pulse codes require a more detailed description of neuronal dynamics than rate codes. Models of neurons at different levels of detail will be the topic of part I of the book.

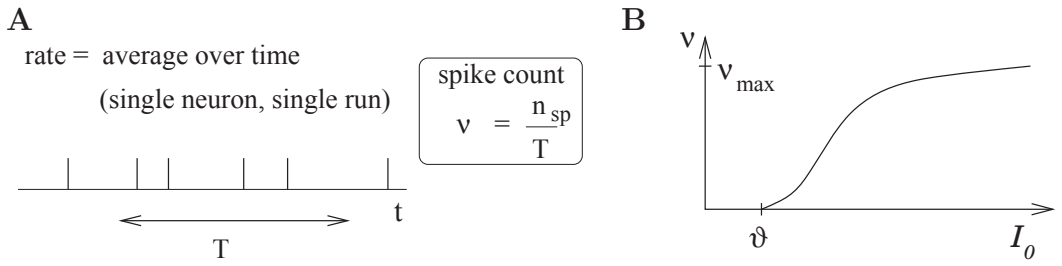

Fig. 1.9: **A**. Definition of the mean firing rate via a temporal average. **B**. Gain function, schematic. The output rate ν is given as a function of the total input I_0.

1.5 Rate Codes

A quick glance at the experimental literature reveals that there is no unique and well-defined concept of 'mean firing rate'. In fact, there are at least three different notions of rate which are often confused and used simultaneously. The three definitions refer to three different averaging procedures: either an average over time, or an average over several repetitions of the experiment, or an average over a population of neurons. The following three subsections will reconsider the three concepts. An excellent discussion of rate codes can be found in (Rieke et al., 1996).

1.5.1 Rate as a Spike Count (Average over Time)

The first and most commonly used definition of a firing rate refers to a temporal average. As discussed in the preceding section, this is essentially the spike count in an interval of duration T divided by T; see Fig. 1.9. The length T of the time window is set by the experimenter and depends on the type of neuron recorded from and the stimulus. In practice, to get sensible averages, several spikes should occur within the time window. Typical values are $T = 100$ ms or $T = 500$ ms, but the duration may also be longer or shorter.

This definition of rate has been successfully used in many preparations, particularly in experiments on sensory or motor systems. A classical example is the stretch receptor in a muscle spindle (Adrian, 1926). The number of spikes emitted by the receptor neuron increases with the force applied to the muscle. Another textbook example is the touch receptor in the leech (Kandel and Schwartz, 1991). The stronger the touch stimulus, the more spikes occur during a stimulation period of 500 ms.

These classical results show that the experimenter as an external observer can evaluate and classify neuronal firing by a spike count measure – but is this really the code used by neurons in the brain? In other words, is a neuron which receives

signals from a sensory neuron only looking at and reacting to the number of spikes it receives in a time window of, say, 500 ms? We will approach this question from a modeling point of view later on in the book. Here we discuss some critical experimental evidence.

From behavioral experiments it is known that reaction times are often rather short. A fly can react to new stimuli and change the direction of flight within 30-40 ms; see the discussion in (Rieke et al., 1996). This is not long enough for counting spikes and averaging over some long time window. The fly has to respond after a postsynaptic neuron has received one or two spikes. Humans can recognize visual scenes in just a few hundred milliseconds (Thorpe et al., 1996), even though recognition is believed to involve several processing steps. Again, this does not leave enough time to perform temporal averages on each level. In fact, humans can detect images in a sequence of unrelated pictures even if each image is shown for only 14 – 100 milliseconds (Keysers et al., 2001).

Temporal averaging can work well in cases where the stimulus is constant or slowly varying and does not require a fast reaction of the organism - and this is the situation usually encountered in experimental protocols. Real-world input, however, is hardly stationary, but often changing on a fast time scale. For example, even when viewing a static image, humans perform saccades, rapid changes of the direction of gaze. The image projected onto the retinal photo receptors changes therefore every few hundred milliseconds.

Despite its shortcomings, the concept of a firing rate code is widely used not only in experiments, but also in models of neural networks. It has led to the idea that a neuron transforms information about a single input variable (the stimulus strength) into a single continuous output variable (the firing rate); cf. Fig. 1.9B. The output rate ν increases with the stimulus strength and saturates for large input I_0 towards a maximum value $\nu^{\max}$. In experiments, a single neuron can be stimulated by injecting with an intra-cellular electrode a constant current I_0. The relation between the measured firing frequency ν and the applied input current I_0 is sometimes called the frequency-current curve of the neuron. In models, we formalize the relation between firing frequency (rate) and input current and write $\nu = g(I_0)$. We refer to g as the neuronal gain function or transfer function.

From the point of view of rate coding, spikes are just a convenient way to transmit the analog output variable ν over long distances. In fact, the best coding scheme to transmit the value of the rate ν would be by a regular spike train with intervals $1/\nu$. In this case, the rate could be reliably measured after only two spikes. From the point of view of rate coding, the irregularities encountered in real spike trains of neurons in the cortex must therefore be considered as noise. In order to get rid of the noise and arrive at a reliable estimate of the rate, the experimenter (or the postsynaptic neuron) has to average over a larger number of spikes. A critical discussion of the temporal averaging concept can be found in (Shadlen and Newsome, 1994; Softky, 1995; Rieke et al., 1996).

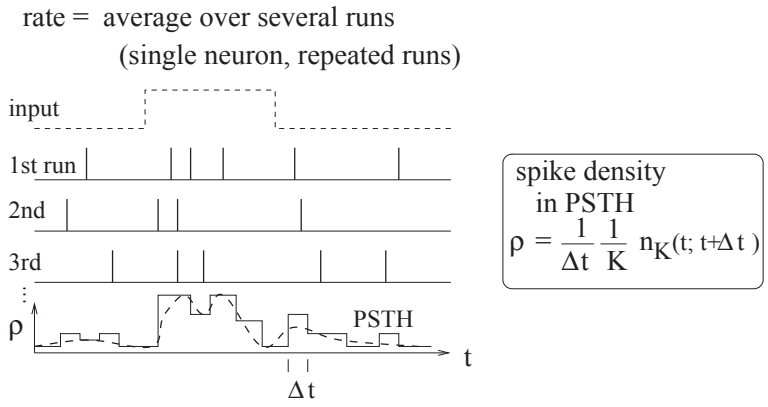

Fig. 1.10: Definition of the spike density in the Peri-Stimulus-Time Histogram (PSTH) as an average over several runs of the experiment.

1.5.2 Rate as a Spike Density (Average over Several Runs)

There is a second definition of rate which works for stationary as well as for time-dependent stimuli. The experimenter records from a neuron while stimulating with some input sequence. The same stimulation sequence is repeated several times and the neuronal response is reported in a Peri-Stimulus-Time Histogram (PSTH); see Fig. 1.10. The time t is measured with respect to the start of the stimulation sequence and Δt is typically in the range of one or a few milliseconds. The number of occurrences of spikes $n_K(t; t + \Delta t)$ summed over all repetitions of the experiment divided by the number K of repetitions is a measure of the typical activity of the neuron between time t and $t + \Delta t$. A further division by the interval length Δt yields the spike density of the PSTH

$$\rho(t) = \frac{1}{\Delta t} \frac{n_K(t; t + \Delta t)}{K} . \tag{1.9}$$

Sometimes the result is smoothed to get a continuous 'rate' variable. The spike density of the PSTH is usually reported in units of Hz and often called the (time-dependent) firing rate of the neuron.

As an experimental procedure, the spike density measure is a useful method to evaluate neuronal activity, in particular in the case of time-dependent stimuli. The obvious problem with this approach is that it can not be the decoding scheme used by neurons in the brain. Consider for example a frog which wants to catch a fly. It can not wait for the insect to fly repeatedly along exactly the same trajectory. The frog has to base its decision on a single 'run' – each fly and each trajectory is different.

Nevertheless, the experimental spike density measure can make sense, if there are large populations of independent neurons that receive the same stimulus. Instead of recording from a population of N neurons in a single run, it is experimentally easier to record from a single neuron and average over N repeated

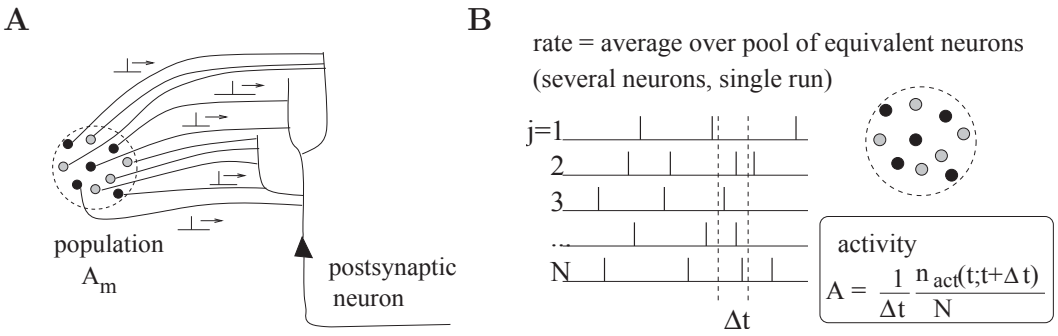

Fig. 1.11: **A**. A postsynpatic neuron receives spike input from the population m with activity A_m. **B**. The population activity is defined as the fraction of neurons that are active in a short interval $[t, t + \Delta t]$ divided by Δt.

runs. Thus, the spike density coding relies on the implicit assumption that there are always populations of neurons and therefore leads us to the third notion of a firing rate, viz., a rate defined as a population average.

1.5.3 Rate as a Population Activity (Average over Several Neurons)

The number of neurons in the brain is huge. Often many neurons have similar properties and respond to the same stimuli. For example, neurons in the primary visual cortex of cats and monkeys are arranged in columns of cells with similar properties (Hubel and Wiesel, 1962, 1977; Hubel, 1988). Let us idealize the situation and consider a population of neurons with identical properties. In particular, all neurons in the population should have the same pattern of input and output connections. The spikes of the neurons in a population m are sent off to another population n. In our idealized picture, each neuron in population n receives input from all neurons in population m. The relevant quantity, from the point of view of the receiving neuron, is the proportion of active neurons in the presynaptic population m; see Fig. 1.11A. Formally, we define the population activity

$$A(t) = \frac{1}{\Delta t}\frac{n_{\rm act}(t; t+\Delta t)}{N} = \frac{1}{\Delta t}\frac{\int_t^{t+\Delta t}\sum_j\sum_f \delta(t - t_j^{(f)})\,{\rm d}t}{N} \qquad (1.10)$$

where N is the size of the population, $n_{\rm act}(t; t+\Delta t)$ the number of spikes (summed over all neurons in the population) that occur between t and $t + \Delta t$ and Δt a small time interval; see Fig. 1.11. Eq. (1.10) defines a variable with units $\rm s^{-1}$ – in other words, a rate.

The population activity may vary rapidly and can reflect changes in the stimulus conditions nearly instantaneously (Gerstner, 2000a; Brunel et al., 2001). Thus

the population activity does not suffer from the disadvantages of a firing rate defined by temporal averaging at the single-unit level. A potential problem with the definition (1.10) is that we have formally required a homogeneous population of neurons with identical connections which is hardly realistic. Real populations will always have a certain degree of heterogeneity both in their internal parameters and in their connectivity pattern. Nevertheless, rate as a population activity (of suitably defined pools of neurons) may be a useful coding principle in many areas of the brain. For inhomogeneous populations, the definition (1.10) may be replaced by a weighted average over the population.

Example: Population vector coding

We give an example of a weighted average in an inhomogeneous population. Let us suppose that we are studying a population of neurons which respond to a stimulus $\mathbf{x}$. We may think of $\mathbf{x}$ as the location of the stimulus in input space. Neuron i responds best to stimulus $\mathbf{x}_i$, another neuron j responds best to stimulus $\mathbf{x}_j$. In other words, we may say that the spikes for a neuron i 'represent' an input vector $\mathbf{x}_i$ and those of j an input vector $\mathbf{x}_j$. In a large population, many neurons will be active simultaneously when a new stimulus $\mathbf{x}$ is represented. The location of this stimulus can then be estimated from the weighted population average

$$\mathbf{x}^{\rm est}(t) = \frac{\int_t^{t+\Delta t} \sum_j \sum_f \mathbf{x}_j \, \delta(t - t_j^{(f)}) \, dt}{\int_t^{t+\Delta t} \sum_j \sum_f \delta(t - t_j^{(f)}) \, {\rm d}t} \tag{1.11}$$

Both numerator and denominator are closely related to the population activity (1.10). The estimate (1.11) has been successfully used for an interpretation of neuronal activity in primate motor cortex (Georgopoulos et al., 1986; Wilson and McNaughton, 1993). It is, however, not completely clear whether postsynaptic neurons really evaluate the fraction (1.11). In any case, eq. (1.11) can be applied by external observers to 'decode' neuronal signals, if the spike trains of a large number of neurons are accessible.

1.6 Spike Codes

In this section, we will briefly introduce some potential coding strategies based on spike timing.

1.6.1 Time-to-First-Spike

Let us study a neuron which abruptly receives a 'new' input at time t_0. For example, a neuron might be driven by an external stimulus which is suddenly switched on at time t_0. This seems to be somewhat academic, but even in a

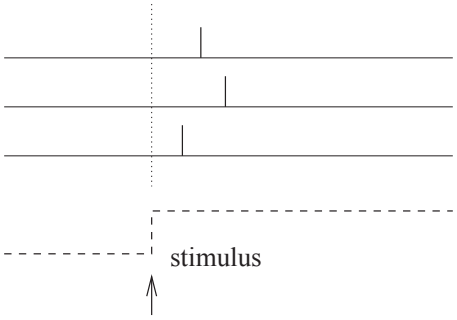

Fig. 1.12: Time-to-first spike. The spike train of three neurons are shown. The third neuron from the top is the first one to fire a spike after the stimulus onset (arrow). The dashed line indicates the time course of the stimulus.

realistic situation abrupt changes in the input are quite common. When we look at a picture, our gaze jumps from one point to the next. After each saccade, the photo receptors in the retina receive a new visual input. Information about the onset of a saccade would easily be available in the brain and could serve as an internal reference signal. We can then imagine a code where for each neuron the timing of the *first* spike after the reference signal contains all information about the new stimulus. A neuron which fires shortly after the reference signal could signal a strong stimulation, firing somewhat later would signal a weaker stimulation; see Fig. 1.12.

In a pure version of this coding scheme, each neuron only needs to fire a single spike to transmit information. (If it emits several spikes, only the first spike after the reference signal counts. All following spikes would be irrelevant.) To implement a clean version of such a coding scheme, we imagine that each neuron is shut off by inhibition as soon as it has fired a spike. Inhibition ends with the onset of the next stimulus (e.g., after the next saccade). After the release from inhibition the neuron is ready to emit its next spike that now transmits information about the new stimulus. Since each neuron in such a scenario transmits exactly one spike per stimulus, it is clear that only the timing conveys information and not the number of spikes.

A coding scheme based on the time-to-first-spike is certainly an idealization. In a slightly different context coding by first spikes has been discussed by S. Thorpe (Thorpe et al., 1996). Thorpe argues that the brain does not have time to evaluate more than one spike from each neuron per processing step. Therefore the first spike should contain most of the relevant information. Using information-theoretic measures on their experimental data, several groups have shown that most of the information about a new stimulus is indeed conveyed during the first 20 or 50 milliseconds after the onset of the neuronal response (Optican and Richmond, 1987; Kjaer et al., 1994; Tovee et al., 1993; Tovee and Rolls, 1995).

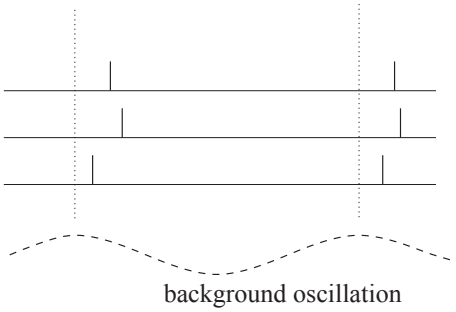

Fig. 1.13: Phase. The neurons fire at different phases with respect to the background oscillation (dashed). The phase could code relevant information.

Rapid computation during the transients after a new stimulus has also been discussed in model studies (Hopfield and Herz, 1995; Tsodyks and Sejnowski, 1995; van Vreeswijk and Sompolinsky, 1997; Treves et al., 1997). Since time-to-first spike is a highly simplified coding scheme, analytical studies are possible (Maass, 1998).

1.6.2 Phase

We can apply a code by 'time-to-first-spike' also in the situation where the reference signal is not a single event, but a periodic signal. In the hippocampus, in the olfactory system, and also in other areas of the brain, oscillations of some global variable (for example the population activity) are quite common. These oscillations could serve as an internal reference signal. Neuronal spike trains could then encode information in the phase of a pulse with respect to the background oscillation. If the input does not change between one cycle and the next, then the same pattern of phases repeats periodically; see Fig. 1.13.

The concept of coding by phases has been studied by several different groups, not only in model studies (Hopfield, 1995; Jensen and Lisman, 1996; Maass, 1996), but also experimentally (O'Keefe, 1993). There is, for example, evidence that the phase of a spike during an oscillation in the hippocampus of the rat conveys information on the spatial location of the animal which is not fully accounted for by the firing rate of the neuron (O'Keefe, 1993).

1.6.3 Correlations and Synchrony

We can also use spikes from other neurons as the reference signal for a spike code. For example, synchrony between a pair or many neurons could signify special events and convey information which is not contained in the firing rate of the neurons; see Fig. 1.14. One famous idea is that synchrony could mean 'belonging together' (Milner, 1974; von der Malsburg, 1981). Consider for example a complex

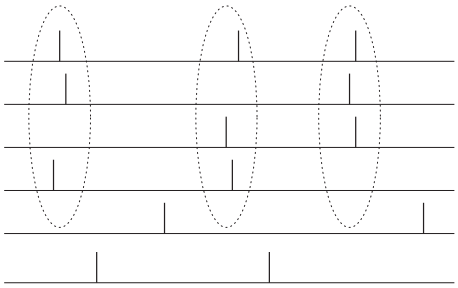

Fig. 1.14: Synchrony. The upper four neurons are nearly synchronous, two other neurons at the bottom are not synchronized with the others.

scene consisting of several objects. It is represented in the brain by the activity of a large number of neurons. Neurons which represent the same object could be 'labeled' by the fact that they fire synchronously (von der Malsburg, 1981; von der Malsburg and Buhmann, 1992; Eckhorn et al., 1988; Gray and Singer, 1989). Coding by synchrony has been studied extensively both experimentally (Eckhorn et al., 1988; Gray and Singer, 1989; Gray et al., 1989; Singer, 1994; Engel et al., 1991b,a; Kreiter and Singer, 1992) and in models (Wang et al., 1990; von der Malsburg and Buhmann, 1992; Eckhorn et al., 1990; Aertsen and Arndt, 1993; König and Schillen, 1991; Schillen and König, 1991; Gerstner et al., 1993a; Ritz et al., 1994; Terman and Wang, 1995; Wang, 1995). For a review of potential mechanism, see (Ritz and Sejnowski, 1997).

More generally, not only synchrony but any precise spatio-temporal pulse pattern could be a meaningful event. For example, a spike pattern of three neurons, where neuron 1 fires at some arbitrary time t_1 followed by neuron 2 at time $t_1 + \delta_{12}$ and by neuron 3 at $t_1 + \delta_{13}$, might represent a certain stimulus condition. The same three neurons firing with different relative delays might signify a different stimulus. The relevance of precise spatio-temporal spike patterns has been studied intensively by Abeles (Abeles, 1991; Abeles et al., 1993; Abeles, 1994). Similarly, but on a somewhat coarse time scale, correlations of auditory and visual neurons are found to be stimulus dependent and might convey information beyond that contained in the firing rate alone (deCharms and Merzenich, 1996; Steinmetz et al., 2000).

1.6.4 Stimulus Reconstruction and Reverse Correlation

Let us consider a neuron which is driven by a time dependent stimulus $s(t)$. Every time a spike occurs, we note the time course of the stimulus in a time window of about 100 milliseconds immediately before the spike. Averaging the results over several spikes yields the typical time course of the stimulus just before a spike (de Boer and Kuyper, 1968). Such a procedure is called a 'reverse correlation'

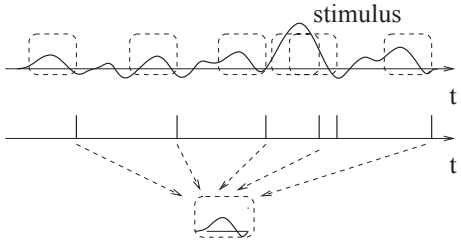

Fig. 1.15: Reverse correlation technique (schematic). The stimulus in the top trace has caused the spike train shown immediately below. The time course of the stimulus just before the spikes (dashed boxes) has been averaged to yield the typical time course (bottom).

approach; see Fig. 1.15. In contrast to the PSTH experiment sketched in Section 1.5.2 where the experimenter averages the neuron's response over several trials with the same stimulus, reverse correlation means that the experimenter averages the input under the condition of an identical response, viz., a spike. In other words, it is a spike-triggered average; see, e.g., (de Ruyter van Stevenick and Bialek, 1988; Rieke et al., 1996). The results of the reverse correlation, i.e., the typical time course of the stimulus which has triggered the spike, can be interpreted as the 'meaning' of a single spike. Reverse correlation techniques have made it possible to measure, for example, the spatio-temporal characteristics of neurons in the visual cortex (Eckhorn et al., 1993; DeAngelis et al., 1995).

With a somewhat more elaborate version of this approach, W. Bialek and his co-workers have been able to 'read' the neural code of the H1 neuron in the fly and to reconstruct a time-dependent stimulus (Bialek et al., 1991; Rieke et al., 1996). Here we give a simplified version of their argument.

Results from reverse correlation analysis suggest, that each spike signifies the time course of the stimulus preceding the spike. If this is correct, a reconstruction of the complete time course of the stimulus $s(t)$ from the set of firing times $\mathcal{F} = \{t^{(1)}, \ldots t^{(n)}\}$ should be possible; see Fig. 1.16.

As a simple test of this hypothesis, Bialek and coworkers have studied a linear reconstruction. A spike at time $t^{(f)}$ gives a contribution $\kappa(t - t^{(f)})$ to the estimation $s^{\text{est}}(t)$ of the time course of the stimulus. Here, $t^{(f)} \in \mathcal{F}$ is one of the firing times and $\kappa(t - t^{(f)})$ is a kernel which is nonzero during some time before and around $t^{(f)}$; cf. Fig. 1.16B. A linear estimate of the stimulus is

$$s^{\text{est}}(t) = \sum_{f=1}^{n} \kappa(t - t^{(f)}) \,. \tag{1.12}$$

The form of the kernel κ was determined through optimization so that the average reconstruction error $\int \mathrm{d}t[s(t) - s^{\text{est}}(t)]^2$ was minimal. The quality of the reconstruction was then tested on additional data which was not used for the

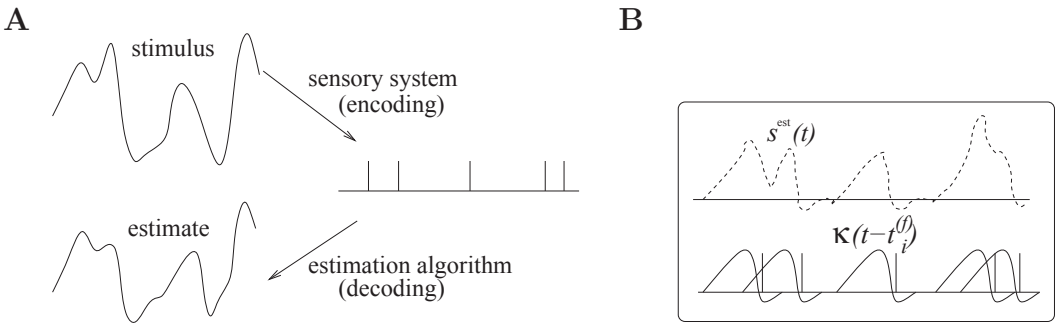

Fig. 1.16: Reconstruction of a stimulus (schematic). **A**. A stimulus evokes a spike train of a neuron. The time course of the stimulus may be estimated from the spike train; redrawn after [Rieke et al., 1996]. **B**. In the framework of linear stimulus reconstruction, the estimation $s^{\rm est}(t)$ (dashed) is the sum of the contributions κ (solid lines) of all spikes.

optimization. Surprisingly enough, the simple linear reconstruction (1.12) gave a fair estimate of the time course of the stimulus even though the stimulus varied on a time scale comparable to the typical interspike interval (Bialek et al., 1991; Bialek and Rieke, 1992; Rieke et al., 1996). This reconstruction method shows nicely that information about a time dependent input can indeed be conveyed by spike timing.

1.7 Discussion: Spikes or Rates?

The dividing line between spike codes and firing rates is not always as clearly drawn as it may seem at first sight. Some codes which were first proposed as pure examples of pulse codes have later been interpreted as variations of rate codes. For example the stimulus reconstruction (1.12) with kernels seems to be a clear example of a spike code. Nevertheless, it is also not so far from a rate code based on spike counts (Abbott, 1994; Theunissen and Miller, 1995). To see this, consider a spike count measure with a running time window $K(.)$. We can estimate the rate ν at time t by

$$\nu(t) = \frac{\int K(\tau)\, S(t-\tau) d\tau}{\int K(\tau) d\tau} \tag{1.13}$$

where $S(t) = \sum_{f=1}^{n} \delta(t - t^{(f)})$ is the spike train under consideration. The integrals run from minus to plus infinity. For a rectangular time window $K(\tau) = 1$ for $-T/2 < \tau < T/2$ and zero otherwise, (1.13) reduces exactly to our definition (1.8) of a rate as a spike count measure.

The time window in (1.13) can be made rather short so that at most a few spikes fall into the interval T. Furthermore, there is no need that the window

$K(.)$ be symmetric and rectangular. We may just as well take an asymmetric time window with smooth borders. Moreover, we can perform the integration over the δ function which yields

$$\nu(t) = c \sum_{f=1}^{n} K(t - t^{(f)}) \tag{1.14}$$

where $c = [\int K(s)ds]^{-1}$ is a constant. Except for the constant c (which sets the overall scale to units of one over time), the generalized rate formula (1.14) is now identical to the reconstruction formula (1.12). In other words, the linear reconstruction is just the firing rate measured with a cleverly optimized time window.

Similarly, a code based on the 'time-to-first-spike' is also consistent with a rate code. If, for example, the mean firing rate of a neuron is high for a given stimulus, then the first spike is expected to occur early. If the rate is low, the first spike is expected to occur later. Thus the timing of the first spike contains a lot of information about the underlying rate.

Finally, a code based on population activities introduced above as an example of a rate code may be used for very fast temporal coding schemes. As discussed later in Chapter 6, the population activity reacts quickly to any change in the stimulus. Thus rate coding in the sense of a population average is consistent with fast temporal information processing, whereas rate coding in the sense of a naïve spike count measure is not.

The discussion of whether or not to call a given code a rate code is still ongoing, even though precise definitions have been proposed (Theunissen and Miller, 1995). What is important, in our opinion, is to have a coding scheme which allows neurons to quickly respond to stimulus changes. A naïve spike count code with a long time window is unable to do this, but many of the other codes are. The name of such a code, whether it is deemed a rate code or not is of minor importance.

Example: Towards a definition of rate codes

We have seen above in Eq. (1.14) that stimulus reconstruction with a linear kernel can be seen as a special instance of a rate code. This suggests a formal definition of a rate code via the reconstruction procedure: If all information contained in a spike train can be recovered by the linear reconstruction procedure of Eq. (1.12), then the neuron is, by definition, using a rate code. Spike codes would then be codes where a linear reconstruction is not successful. Theunissen and Miller have proposed a definition of rate coding that makes the above ideas more precise (Theunissen and Miller, 1995).

To see how their definition works, we have to return to the reconstruction formula (1.12). It is, in fact, the first term of a systematic Volterra expansion for

the estimation of the stimulus from the spikes (Bialek et al., 1991)

$$s^{\text{est}}(t) = \sum_f \kappa_1(t - t^{(f)}) + \sum_{f,f'} \kappa_2(t - t^{(f)}, t - t^{(f')}) + \dots . \tag{1.15}$$

For a specific neuron, inclusion of higher-order terms $\kappa_2, \kappa_3, \dots$ may or may not improve the quality of the estimation. For most neurons where the reconstruction has been carried through it seems that the higher-order terms do not contribute a large amount of information (Rieke et al., 1996). The neurons would then be classified as rate coding.

Let us now suppose that the reconstruction procedure indicates a significant contribution of the second-order term. Does this exclude rate coding? Unfortunately this is not the case. We have to exclude two other possibilities. Firstly, we might have chosen a suboptimal stimulus. A neuron might for example encode the variable x by a rate code, so that a nearly perfect linear reconstruction of x would be possible,

$$x(t) \approx x^{\text{est}} = \sum_{f=1}^{n} \kappa_{1;\text{x}}(t - t^{(f)}) \,; \tag{1.16}$$

But if we chose a stimulus $s = x^2$ instead of x, then the reconstruction for s^{est} would involve second-order terms, even though the neuron is really using rate code.

Secondly, according to Theunissen and Miller [1995] a spike code should show a temporal structure that is more precise than the temporal structure of the stimulus. The fact that neurons show precise and reliable spike timing as such is, for them, not sufficient to classify the neuron as a temporal encoder, since the neuronal precision could just be the image of precise temporal input. Let us consider a stimulus with cut-off frequency ω. In order to exclude the possibility that the timing is induced by the stimulus, Theunissen and Miller propose to consider the Fourier spectrum of the higher-order reconstruction kernels. If the Fourier transform of the higher-order kernels contains frequencies less than ω only, then the code is a rate code. If higher-order kernels are significant and contain frequencies above ω, then the information is encoded temporally. A positive example of a spike code (or of 'temporal encoding') according to this definition would be the code by correlation and synchrony introduced above. Another example would be the phase code, in particular if the number of spikes per cycle is independent of the stimulus strength. For the exact mathematical definition of a temporal code according to Theunissen and Miller, the reader is refered to the original literature (Theunissen and Miller, 1995).

1.8 Summary

The neuronal signal consists of short voltage pulses called action potentials or spikes. These pulses travel along the axon and are distributed to several postsynaptic neurons where they evoke postsynaptic potentials. If a postsynaptic neuron receives several spikes from several presynaptic neurons within a short time window, its membrane potential may reach a critical value and an action potential is triggered. This action potential is the output signal which is, in turn, transmitted to other neurons.

The sequence of action potentials contains the information that is conveyed from one neuron to the next – but what is the code used by the neurons? Even though it is a question of fundamental importance the problem of neuronal coding is still not fully resolved. We have reviewed three concepts of rate codes, viz. spike count over some time window, spike density in a histogram, and population activity in an ensemble of neurons. All three concepts have been successfully used in experimental data analysis. All of these concepts are, however, problematic when they are interpreted as the actual code used for neuronal information transmission. A constructive criticism of rate codes may come from a presentation of potential spike codes, if their usefulness in terms of computational power or ease of implementation in biological hardware can be shown. It should be clear that modeling cannot give definite answers to the problem of neuronal coding. The final answers have to come from experiments. One task of modeling may be to discuss possible coding schemes, study their computational potential, exemplify their utility, and point out their limitations.

It is difficult to draw a clear border line between pulse and rate codes. Whatever the name of the code, it should offer a neural system the possibility to react quickly to changes in the input. This seems to be a minimum requirement if fast behavioral reaction times are to be accounted for.

If pulse coding is relevant, a description of information processing in the brain must be based on spiking neuron models. If all information is contained in the mean firing rate, then models on the level of rates suffice. Since we do not want to take any decision *a priori* about the neuronal code, we concentrate in this book on models of spiking neurons. In some cases, for example for stationary input, it will turn out that the spiking neuron models can be strictly reduced to rate models; in other cases such a reduction is not possible. By modeling on the level of spiking neurons, the question of neuronal coding can thus be kept open.

Literature

An elementary, non-technical introduction to neurons and synapses can be found in the book by Thompson (1993). At an intermediate level is "From neuron to brain" by Kuffler et al. (1984). A standard textbook on Neuroscience covering a

wealth of experimental results is "Principles of Neural Science" by Kandel and Schwartz (1991).

Phenomenological spiking neuron models similar to the model discussed in Section 1.3.1 have a long tradition in theoretical neuroscience, e.g., (Lapicque, 1907; Hill, 1936; McCulloch and Pitts, 1943; Stein, 1965; Geisler and Goldberg, 1966; Weiss, 1966; Stein, 1967b). They are reviewed in Holden (1976), Tuckwell (1988), and Maass and Bishop (1998).

An excellent discussion of the problem of neuronal coding can be found in the book 'SPIKES - Exploring the neural code' by Rieke et al. (1996). The debate of spikes versus rates is also highlighted in several papers (Abeles, 1994; Abbott, 1994; Shadlen and Newsome, 1994; Softky, 1995; Maass and Bishop, 1998; Theunissen and Miller, 1995).

Part I

Single Neuron Models

Chapter 2

Detailed Neuron Models

From a biophysical point of view, action potentials are the result of currents that pass through ion channels in the cell membrane. In an extensive series of experiments on the giant axon of the squid, Hodgkin and Huxley succeeded to measure these currents and to describe their dynamics in terms of differential equations. In Section 2.2, the Hodgkin-Huxley model is reviewed and its behavior illustrated by several examples.

The Hodgkin-Huxley equations are the starting point for detailed neuron models which account for numerous ion channels, different types of synapse, and the specific spatial geometry of an individual neuron. Ion channels, synaptic dynamics, and the spatial structure of dendrites are the topics of Sections 2.3–2.5. The Hodgkin-Huxley model is also an important reference model for the derivation of simplified neuron models in Chapters 3 and 4. Before we can turn to the Hodgkin-Huxley equations, we need to give some additional information on the equilibrium potential of ion channels.

2.1 Equilibrium potential

Neurons are just as other cells enclosed by a membrane which separates the interior of the cell from the extracellular space. Inside the cell the concentration of ions is different from that in the surrounding liquid. The difference in concentration generates an electrical potential which plays an important role in neuronal dynamics. In this section, we want to provide some background information and give an intuitive explanation of the equilibrium potential.

2.1.1 Nernst potential

From the theory of thermodynamics, it is known that the probability that a molecule takes a state of energy E is proportional to the Boltzmann factor $p(E) \propto \exp(-E/kT)$ where k is the Boltzmann constant and T the temperature. Let us

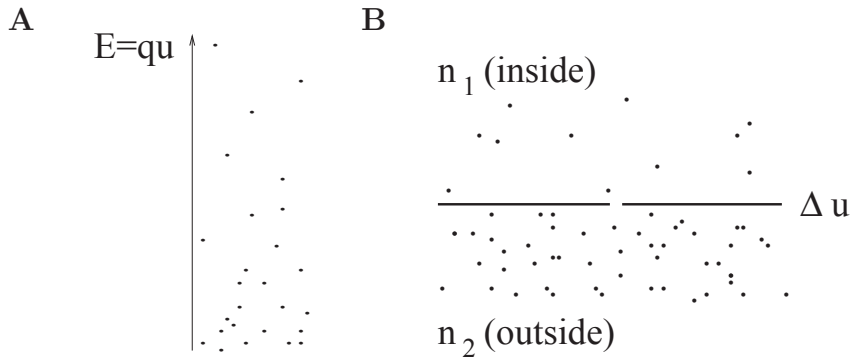

Fig. 2.1: **A**. At thermal equilibrium, positive ions in an electric field will be distributed so that less ions are in a state of high energy and more at low energy. Thus a voltage difference generates a gradient in concentration. **B**. Similarly, a difference in ion concentration generates an electrical potential. The concentration n_2 inside the neuron is different from the concentration n_1 of the surround. The resulting potential is called the Nernst-potential. The solid line indicates the cell membrane. Ions can pass through the gap.

consider positive ions with charge q in a static electrical field. Their energy at location x is $E(x) = q\,u(x)$ where $u(x)$ is the potential at x. The probability to find an ion in the region around x is therefore proportional to $\exp[-q\,u(x)/kT]$. Since the number of ions is huge, we may interpret the probability as a ion density. For ions with positive charge $q > 0$, the ion density is therefore higher in regions with low potential u. Let us write $n(x)$ for the ion density at point x. The relation between the density at point x_1 and point x_2 is

$$\frac{n(x_1)}{n(x_2)} = \exp\left[-\frac{q\,u(x_1) - q\,u(x_2)}{k\,T}\right] \tag{2.1}$$

A difference in the electrical potential $\Delta u = u(x_1) - u(x_2)$ generates therefore a difference in ion density; cf. Fig. 2.1.

Since this is a statement about an equilibrium state, the reverse must also be true. A difference in ion density generates a difference Δu in the electrical potential. We consider two regions of ions with concentration n_1 and n_2, respectively. Solving (2.1) for Δu we find that, at equilibrium, the concentration difference generates a voltage

$$\Delta u = \frac{k\,T}{q} \ln\frac{n_2}{n_1} \tag{2.2}$$

which is called the Nernst potential (Hille, 1992).

2.1.2 Reversal Potential

The cell membrane consists of a thin bilayer of lipids and is a nearly perfect electrical insulator. Embedded in the cell membrane are, however, specific proteins which act as ion gates. A first type of gate are the ion pumps, a second one are ion channels. Ion pumps actively transport ions from one side to the other. As a result, ion concentrations in the intra-cellular liquid differ from that of the surround. For example, the sodium concentration inside the cell ($\approx$ 60mM/l) is lower than that in the extracellular liquid ($\approx$ 440 mM/l). On the other hand, the potassium concentration inside is higher ($\approx$ 400 mM/l) than in the surround ($\approx$ 20 mM/l).

Let us concentrate for the moment on sodium ions. At equilibrium the difference in concentration causes a Nernst potential E_{Na} of about +50 mV. That is, at equilibrium the interior of the cell has a positive potential with respect to the surround. The interior of the cell and the surrounding liquid are in contact through ion channels where Na^+ ions can pass from one side of the membrane to the other. If the voltage difference Δu is smaller than the value of the Nernst potential E_{Na}, more Na^+ ions flow into the cell so as to decrease the concentration difference. If the voltage is larger than the Nernst potential ions would flow out the cell. Thus the direction of the current is reversed when the voltage Δu passes E_{Na}. For this reason, E_{Na} is called the reversal potential.

Example: Reversal Potential for Potassium

As mentioned above, the ion concentration of potassium is higher inside the cell ($\approx$ 400 mM/l) than in the extracellular liquid ($\approx$ 20 mM/l). Potassium ions have a single positive charge $q = 1.6 \times 10^{-19}$ C. Application of the Nernst equation with the Boltzmann constant $k = 1.4 \times 10^{-23}$ J/K yields $E_{\mathrm{K}} \approx -77\,\mathrm{mV}$ at room temperature. The reversal potential for K^+ ions is therefore negative.

Example: Resting Potential

So far we have considered either sodium or potassium. In real cells, these and other ion types are simultaneously present and contribute to the voltage across the membrane. It is found experimentally that the resting potential of the membrane is about $u_{\mathrm{rest}} \approx$ -65 mV. Since $E_{\mathrm{K}} < u_{\mathrm{rest}} < E_{\mathrm{Na}}$, potassium ions will, at the resting potential, flow out of the cell while sodium ions flow into the cell. The active ion pumps balance this flow and transport just as many ions back as pass through the channels. The value of u_{rest} is determined by the dynamic equilibrium between the ion flow through the channels (permeability of the membrane) and active ion transport (efficiency of the ion pumps).

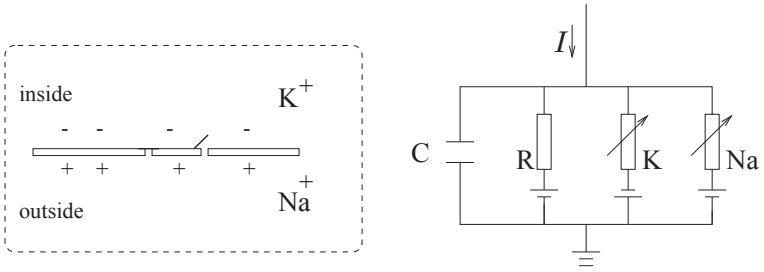

Fig. 2.2: Schematic diagram for the Hodgkin-Huxley model.

2.2 Hodgkin-Huxley Model

Hodgkin and Huxley (Hodgkin and Huxley, 1952) performed experiments on the giant axon of the squid and found three different types of ion current, viz., sodium, potassium, and a leak current that consists mainly of Cl^- ions. Specific voltage-dependent ion channels, one for sodium and another one for potassium, control the flow of those ions through the cell membrane. The leak current takes care of other channel types which are not described explicitly.

2.2.1 Definition of the model

The Hodgkin-Huxley model can be understood with the help of Fig. 2.2. The semipermeable cell membrane separates the interior of the cell from the extracellular liquid and acts as a capacitor. If an input current $I(t)$ is injected into the cell, it may add further charge on the capacitor, or leak through the channels in the cell membrane. Because of active ion transport through the cell membrane, the ion concentration inside the cell is different from that in the extracellular liquid. The Nernst potential generated by the difference in ion concentration is represented by a battery.

Let us now translate the above considerations into mathematical equations. The conservation of electric charge on a piece of membrane implies that the applied current $I(t)$ may be split in a capacitive current I_C which charges the capacitor C and further components I_k which pass through the ion channels. Thus

$$I(t) = I_C(t) + \sum_k I_k(t) \tag{2.3}$$

where the sum runs over all ion channels. In the standard Hodgkin-Huxley model there are only three types of channel: a sodium channel with index Na, a potassium channel with index K and an unspecific leakage channel with resistance R; cf. Fig. 2.2. From the definition of a capacity $C = Q/u$ where Q is a charge and u the voltage across the capacitor, we find the charging current $I_C = C\,\mathrm{d}u/\mathrm{d}t$.

Chapter 3

Two-Dimensional Neuron Models

The behavior of high-dimensional nonlinear differential equations is difficult to visualize – and even more difficult to analyze. Two-dimensional differential equations, however, can be studied in a transparent manner by means of a phase plane analysis. A reduction of the four-dimensional equation of Hodgkin and Huxley to a two-variable neuron model is thus highly desirable. In the first section of this chapter we exploit the temporal properties of the gating variables of the Hodgkin-Huxley model so as to approximate the four-dimensional differential equation by a two-dimensional one. Section 3.2 is devoted to the phase plane analysis of generic neuron models consisting of two coupled differential equations, one for the membrane potential and the other one for the so-called relaxation variable. One of the questions to which we will return repeatedly throughout this chapter is the problem of the firing threshold. Section 3.3 summarizes some results on threshold and excitability in two-dimensional models. As a first step, however, we have to go through the approximations that are necessary for a reduction of the Hodgkin-Huxley model to two dimensions.

3.1 Reduction to two dimensions

In this section we perform a systematic reduction of the four-dimensional Hodgkin-Huxley model to two dimensions. To do so, we have to eliminate two of the four variables. The essential ideas of the reduction can also be applied to detailed neuron models that may contain many different ion channels. In this case, more than two variables would have to be eliminated, but the procedure would be completely analogous (Kepler et al., 1992).

3.1.1 General approach

We focus on the Hodgkin-Huxley model discussed in Chapter 2.2 and start with two qualitative observations. First, we see from Fig. 2.3B that the time scale of

the dynamics of the gating variable m is much faster than that of the variables n, h, and u. This suggests that we may treat m as an instantaneous variable. The variable m in the ion current equation (2.5) of the Hodgkin-Huxley model can therefore be replaced by its steady-state value, $m(t) \to m_0[u(t)]$. This is what we call a *quasi steady state approximation*.

Second, we see from Fig. 2.3B that the time constants $\tau_n(u)$ and $\tau_h(u)$ are roughly the same, whatever the voltage u. Moreover, the graphs of $n_0(u)$ and $1-h_0(u)$ in Fig. 2.3A are rather similar. This suggests that we may approximate the two variables n and $(1-h)$ by a single effective variable w. To keep the formalism slightly more general we use a linear approximation $(b-h) \approx a\,n$ with some constants a, b and set $w = b - h = a\,n$. With $h = b - w$, $n = w/a$, and $m = m_0(u)$, equations (2.4) - (2.5) become

$$C\frac{\mathrm{d}u}{\mathrm{d}t} = -g_{\mathrm{Na}}[m_0(u)]^3\,(b-w)\,(u-V_{\mathrm{Na}}) - g_{\mathrm{K}}\left(\frac{w}{a}\right)^4\,(u-V_{\mathrm{K}}) - g_L\,(u-V_L) + I\,, \quad (3.1)$$

or

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{1}{\tau}\left[F(u,w) + R\,I\right]\,, \quad (3.2)$$

with $R = g_L^{-1}$, $\tau = R\,C$ and some function F. We now turn to the three equations (2.6). The m equation has disappeared since m is treated as instantaneous. Instead of the two equations (2.6) for n and h, we are left with a single effective equation

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{1}{\tau_w}G(u,w)\,, \quad (3.3)$$

where τ_w is a parameter and G a function that has to be specified. Eqs. (3.2) and (3.3) define a general two-dimensional neuron model. The mathematical details of the reduction of the four-dimensional Hodgkin-Huxley model to the two equations (3.2) and (3.3) are given below. Before we go through the mathematical step, we will present two examples of two-dimensional neuron dynamics. We will return to these examples repeatedly throughout this chapter.

Example: Morris-Lecar model

Morris and Lecar (1981) proposed a two-dimensional description of neuronal spike dynamics. A first equation describes the evolution of the membrane potential u, the second equation the evolution of a slow 'recovery' variable $\hat{w}$. In dimensionless variables, the Morris-Lecar equations read

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -g_1\,\hat{m}_0(u)\,(u-1) - g_2\,\hat{w}\,(u-V_2) - g_L\,(u-V_L) + I\,, \quad (3.4)$$

$$\frac{\mathrm{d}\hat{w}}{\mathrm{d}t} = -\frac{1}{\tau(u)}\left[\hat{w} - w_0(u)\right]\,. \quad (3.5)$$

The voltage has been scaled so that one of the reversal potentials is unity. Time is measured in units of $\tau = RC$. If we compare Eq. (3.4) with Eq. (3.1), we note that the first current term on the right-hand side of Eq. (3.1) has a factor $(b-w)$ which closes the channel for high voltage and which is absent in (3.4). Another difference is that neither $\hat{m}_0$ nor $\hat{w}$ in Eq. (3.4) have exponents. To clarify the relation between the two models, we could set $\hat{m}_0(u) = [m_0(u)]^3$ and $\hat{w} = (w/a)^4$. In the following we consider Eqs. (3.4) and (3.5) as a model in its own rights and drop the hats over m_0 and w.

The equilibrium functions shown in Fig. 2.3A typically have a sigmoidal shape. It is reasonable to approximate the voltage dependence by

$$m_0(u) = \frac{1}{2}\left[1 + \tanh\left(\frac{u-u_1}{u_2}\right)\right] \tag{3.6}$$

$$w_0(u) = \frac{1}{2}\left[1 + \tanh\left(\frac{u-u_3}{u_4}\right)\right] \tag{3.7}$$

with parameters $u_1, \ldots, u_4$, and to approximate the time constant by

$$\tau(u) = \frac{\tau_w}{\cosh\left(\frac{u-u_3}{u_4}\right)} \tag{3.8}$$

with a further parameter τ_w.

The Morris-Lecar model (3.4)–(3.8) gives a phenomenological description of action potentials. Action potentials occur, if the current I is sufficiently strong. We will see later on that the firing threshold in the Morris-Lecar model can be discussed by phase plane analysis.

Example: FitzHugh-Nagumo model

FitzHugh and Nagumo where probably the first to propose that, for a discussion of action potential generation, the four equations of Hodgkin and Huxley can be replaced by two, i.e., Eqs. (3.2) and (3.3). They obtained sharp pulse-like oscillations reminiscent of trains of action potentials by defining the functions $F(u,w)$ and $G(u,w)$ as

$$\begin{aligned} F(u,w) &= u - \frac{1}{3}u^3 - w \\ G(u,w) &= b_0 + b_1\, u - w\,, \end{aligned} \tag{3.9}$$

where u is the membrane voltage and w is a recovery variable (FitzHugh, 1961; Nagumo et al., 1962). Note that both F and G are linear in w; the sole non-linearity is the cubic term in u. The FitzHugh-Nagumo model is one of the simplest model with non-trivial behavior lending itself to a phase plane analysis, which will be discussed below in Sections 3.2 and 3.3.

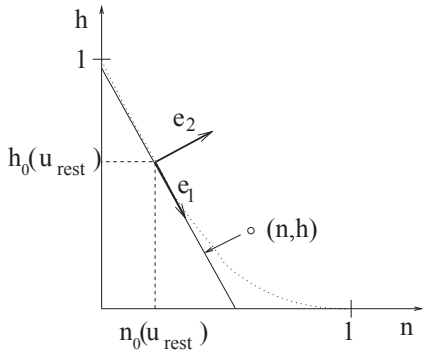

Fig. 3.1: Arbitrary points (n, h) are projected onto the line in direction of $\boldsymbol{e}_1$ and passing through the point $(n_0(u_{\text{rest}}), h_0(u_{\text{rest}}))$. The dotted line gives the curve $(n_0(u), h_0(u))$.

3.1.2 Mathematical steps (*)

The reduction of the Hodgkin-Huxley model to Eqs. (3.2) and (3.3) presented in this paragraph is inspired by the geometrical treatment of Rinzel (1985); see also the slightly more general method of Abbott and Kepler (1990) and Kepler et al. (1992).

The overall aim of the approach is to replace the variables n and h in the Hodgkin-Huxley model by a single effective variable w. At each moment of time, the values $(n(t), h(t))$ can be visualized as points in the two-dimensional plane spanned by n and h; cf. Fig. 3.1. We have argued above that the time course of the variable n is expected to be similar to that of $1 - h$. If, at each time, n was equal to $1 - h$, then all possible points (n, h) would lie on the straight line $h = 1 - n$ passing through the points $(0, 1)$ and $(1, 0)$ of the plane. To keep the model slightly more general we allow for an arbitrary line $h = b - a\,n$ which passes through $(0, b)$ and $(1, b - a)$. It would be unreasonable to expect that all points $(n(t), h(t))$ that occur during the temporal evolution of the Hodgkin-Huxley model fall exactly on that line. The reduction of the number of variables is achieved by a projection of those points onto the line. The position along the line $h = b - a\,n$ gives the new variable w; cf. Fig. 3.1. The projection is the essential approximation during the reduction.

To perform the projection, we will proceed in three steps. A minimal condition for the projection is that the approximation introduces no error while the neuron is at rest. As a first step, we therefore shift the origin of the coordinate system to the rest state and introduce new variables

$$x = n - n_0(u_{\text{rest}}) \tag{3.10}$$

$$y = h - h_0(u_{\text{rest}})\,. \tag{3.11}$$

At rest, we have $x = y = 0$.

Second, we turn the coordinate system by an angle α which is determined as follows. For a given constant voltage u, the dynamics of the gating variables n and h approaches the equilibrium values $(n_0(u), h_0(u))$. The points $(n_0(u), h_0(u))$ as a function of u define a curve in the two-dimensional plane. The slope of the curve at $u = u_{\rm rest}$ yields the turning angle α via

$$\tan\alpha = \frac{\frac{\mathrm{d}h_0}{\mathrm{d}u}\big|_{u_{\rm rest}}}{\frac{\mathrm{d}n_0}{\mathrm{d}u}\big|_{u_{\rm rest}}} \,. \tag{3.12}$$

Turning the coordinate system by α moves the abscissa $\boldsymbol{e}_1$ of the new coordinate system in a direction tangential to the curve. The coordinates (z_1, z_2) in the new system are

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} . \tag{3.13}$$

Third, we set $z_2 = 0$ and retain only the coordinate z_1 along $\boldsymbol{e}_1$. The inverse transform,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} , \tag{3.14}$$

yields $x = z_1 \cos\alpha$ and $y = z_1 \sin\alpha$ since $z_2 = 0$. Hence, after the projection, the new values of the variables n and h are

$$n' = n_0(u_{\rm rest}) + z_1 \cos\alpha \,, \tag{3.15}$$
$$h' = h_0(u_{\rm rest}) + z_1 \sin\alpha \,. \tag{3.16}$$

In principle, z_1 can directly be used as the new effective variable. From (3.13) we find the differential equation

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = \cos\alpha \frac{\mathrm{d}n}{\mathrm{d}t} + \sin\alpha \frac{\mathrm{d}h}{\mathrm{d}t} \,. \tag{3.17}$$

We use (2.7) and replace, on the right-hand side, $n(t)$ and $h(t)$ by (3.15) and (3.16). The result is

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = -\cos\alpha \, \frac{z_1 \cos\alpha + n_0(u_{\rm rest}) - n_0(u)}{\tau_n(u)} - \sin\alpha \frac{z_1 \sin\alpha + h_0(u_{\rm rest}) - h_0(u)}{\tau_h(u)} \,, \tag{3.18}$$

which is of the form $\mathrm{d}z_1/\mathrm{d}t = G(u, z_1)$, as desired.

To see the relation to Eqs. (3.1) and (3.3), it is convenient to rescale z_1 and define

$$w = -\tan\alpha \, n_0(u_{\rm rest}) - z_1 \sin\alpha \,. \tag{3.19}$$

If we introduce $a = -\tan\alpha$, we find from Eq. (3.15) $n' = w/a$ and from Eq. (3.16) $h' = b - w$ which are the approximations that we have used in (3.1). The differential equation for the variable w is of the desired form $\mathrm{d}w/\mathrm{d}t = G(u, w)$ and can

be found from Eq. (3.18). If we approximate the time constants τ_n and τ_h by a common function $\tau(u)$, the dynamics of w is

$$\frac{\mathrm{d}w}{\mathrm{d}t} = -\frac{1}{\tau(u)}\left[w - w_0(u)\right] . \tag{3.20}$$

with a new equilibrium function $w_0(u)$ that is a linear combination of the functions h_0 and n_0. From Eqs. (3.18) and (3.19) we find

$$w_0(u) = -\sin\alpha\left[\cos\alpha\, n_0(u) + \sin\alpha\, h_0(u) - c\right] \tag{3.21}$$

with a parameter c that is determined by direct calculation. In practice, both $w_0(u)$ and $\tau(u)$ are fitted by the expressions (3.7) and (3.8).

3.2 Phase plane analysis

In two-dimensional models, the temporal evolution of the variables $(u, w)^T$ can be visualized in the so-called phase plane. From a starting point $(u(t), w(t))^T$ the system will move in a time Δt to a new state $(u(t + \Delta t), w(t + \Delta t))^T$ which has to be determined by integration of the differential equations (3.2) and (3.3). For Δt sufficiently small, the displacement $(\Delta u, \Delta w)^T$ is in the direction of the flow $(\dot{u}, \dot{w})^T$, i.e.,

$$\begin{pmatrix} \Delta u \\ \Delta w \end{pmatrix} = \begin{pmatrix} \dot{u} \\ \dot{w} \end{pmatrix} \Delta t\,, \tag{3.22}$$

which can be plotted as a vector field in the phase plane. Here $\dot{u} = \mathrm{d}u/\mathrm{d}t$ is given by (3.2) and $\dot{w} = \mathrm{d}w/\mathrm{d}t$ by (3.3). The flow field is also called the phase portrait of the system. An important tool in the construction of the phase portrait are the nullclines which are introduced now.

3.2.1 Nullclines

Let us consider the set of points with $\dot{u} = 0$, called the u-nullcline. The direction of flow on the u-nullcline is in direction of $(0, \dot{w})^T$, since $\dot{u} = 0$. Hence arrows in the phase portrait are vertical on the u-nullcline. Similarly, the w-nullcline is defined by the condition $\dot{w} = 0$ and arrows are horizontal. The fixed points of the dynamics, defined by $\dot{u} = \dot{w} = 0$ are given by the intersection of the u-nullcline with the w-nullcline. In Fig. 3.2 we have three fixed points.

So far we have argued that arrows on the u-nullcline are vertical, but we do not know yet whether they point up or down. To get the extra information needed, let us return to the w-nullcline. By definition, it separates the region with $\dot{w} > 0$ from the area with $\dot{w} < 0$. Suppose we evaluate $G(u, w)$ on the right-hand side of Eq. (3.3) at a single point, e.g, at $(0, 1)$. If $G(0, 1) > 0$, then the whole area

Chapter 4

Formal Spiking Neuron Models

Detailed conductance-based neuron models can reproduce electrophysiological measurements to a high degree of accuracy, but because of their intrinsic complexity these models are difficult to analyze. For this reason, simple phenomenological spiking neuron models are highly popular for studies of neural coding, memory, and network dynamics. In this chapter we discuss formal threshold models of neuronal firing. Spikes are generated whenever the membrane potential u crosses some threshold ϑ from below. The moment of threshold crossing defines the firing time $t^{(f)}$,

$$t^{(f)}: \quad u(t^{(f)}) = \vartheta \quad \text{and} \quad \left.\frac{\mathrm{d}u(t)}{\mathrm{d}t}\right|_{t=t^{(f)}} > 0\,. \tag{4.1}$$

Since spikes are stereotyped events they are fully characterized by their firing time. We focus on models that are based on a single variable u. Some well-known instances of spiking neuron models differ in the specific way the dynamics of the variable u is defined. We start our discussion with the integrate-and-fire neuron (Section 4.1) and turn then to the Spike Response Model (Section 4.2). In Section 4.3 we illustrate the relation of spiking neuron models to conductance-based models. Section 4.4 outlines an analytical approach for a study of integrate-and-fire neurons with passive dendrites. As a first application of spiking neuron models we reconsider in Section 4.5 the problem of neuronal coding. The spiking neuron models introduced in this chapter form the basis for the analysis of network dynamics and learning in the following chapters.

4.1 Integrate-and-fire model

In this section, we give an overview of integrate-and-fire models. The leaky integrate-and-fire neuron introduced in Section 4.1.1 is probably the best-known example of a formal spiking neuron model. Generalizations of the leaky integrate-and-fire model include the nonlinear integrate-and-fire model that is discussed in

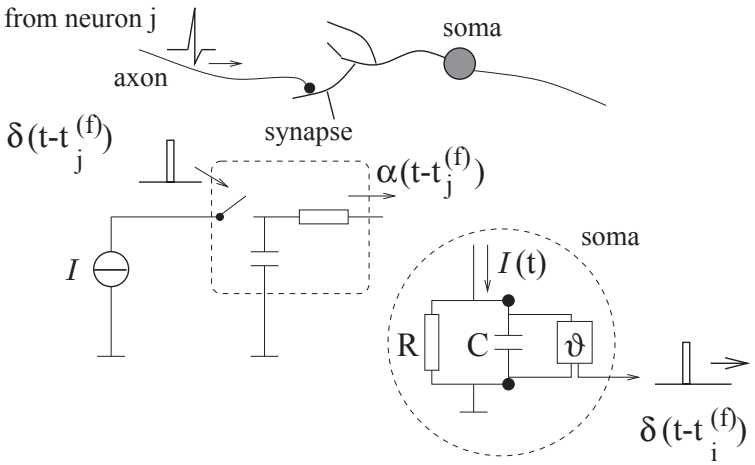

Fig. 4.1: Schematic diagram of the integrate-and-fire model. The basic circuit is the module inside the dashed circle on the right-hand side. A current $I(t)$ charges the RC circuit. The voltage $u(t)$ across the capacitance (points) is compared to a threshold ϑ. If $u(t) = \vartheta$ at time $t_i^{(f)}$ an output pulse $\delta(t - t_i^{(f)})$ is generated. Left part: A presynaptic spike $\delta(t - t_j^{(f)})$ is low-pass filtered at the synapse and generates an input current pulse $\alpha(t - t_j^{(f)})$.

Section 4.1.2. All integrate-and-fire neurons can either be stimulated by external current or by synaptic input from presynaptic neurons. Standard formulations of synaptic input are given in Section 4.1.3.

4.1.1 Leaky Integrate-and-Fire Model

The basic circuit of an integrate-and-fire model consists of a capacitor C in parallel with a resistor R driven by a current $I(t)$; see Fig. 4.1. The driving current can be split into two components, $I(t) = I_R + I_C$. The first component is the resistive current I_R which passes through the linear resistor R. It can be calculated from Ohm's law as $I_R = u/R$ where u is the voltage across the resistor. The second component I_C charges the capacitor C. From the definition of the capacity as $C = q/u$ (where q is the charge and u the voltage), we find a capacitive current $I_C = C\,\mathrm{d}u/\mathrm{d}t$. Thus

$$I(t) = \frac{u(t)}{R} + C\,\frac{\mathrm{d}u}{\mathrm{d}t}\,. \tag{4.2}$$

We multiply (4.2) by R and introduce the time constant $\tau_m = R\,C$ of the 'leaky integrator'. This yields the standard form

$$\tau_m\,\frac{\mathrm{d}u}{\mathrm{d}t} = -u(t) + R\,I(t)\,. \tag{4.3}$$

We refer to u as the membrane potential and to τ_m as the membrane time constant of the neuron.

In integrate-and-fire models the form of an action potential is not described explicitly. Spikes are formal events characterized by a 'firing time' $t^{(f)}$. The firing time $t^{(f)}$ is defined by a threshold criterion

$$t^{(f)}: \quad u(t^{(f)}) = \vartheta \,. \tag{4.4}$$

Immediately after $t^{(f)}$, the potential is reset to a new value $u_r < \vartheta$,

$$\lim_{t \to t^{(f)}; t > t^{(f)}} u(t) = u_r \,. \tag{4.5}$$

For $t > t^{(f)}$ the dynamics is again given by (4.3) until the next threshold crossing occurs. The combination of leaky integration (4.3) and reset (4.5) defines the basic integrate-and-fire model (Stein, 1967b). We note that, since the membrane potential is never above threshold, the threshold condition (4.1) reduces to the criterion (4.4), i.e., the condition on the slope $\mathrm{d}u/\mathrm{d}t$ can be dropped.

In its general version, the leaky integrate-and-fire neuron may also incorporate an absolute refractory period, in which case we proceed as follows. If u reaches the threshold at time $t = t^{(f)}$, we interrupt the dynamics (4.3) during an absolute refractory time $\Delta^{\rm abs}$ and restart the integration at time $t^{(f)} + \Delta^{\rm abs}$ with the new initial condition u_r.

Example: Constant stimulation and firing rates

Before we continue with the definition of the integrate-and-fire model and its variants, let us study a simple example. Suppose that the integrate-and-fire neuron defined by (4.3)–(4.5) is stimulated by a constant input current $I(t) = I_0$. For the sake of simplicity we take the reset potential to be $u_r = 0$.

As a first step, let us calculate the time course of the membrane potential. We assume that a spike has occurred at $t = t^{(1)}$. The trajectory of the membrane potential can be found by integrating (4.3) with the initial condition $u(t^{(1)}) = u_r = 0$. The solution is

$$u(t) = R\, I_0 \left[1 - \exp\left(-\frac{t - t^{(1)}}{\tau_m}\right)\right] . \tag{4.6}$$

The membrane potential (4.6) approaches for $t \to \infty$ the asymptotic value $u(\infty) = R\, I_0$. For $R\, I_0 < \vartheta$ no further spike can occur. For $R\, I_0 > \vartheta$, the membrane potential reaches the threshold ϑ at time $t^{(2)}$, which can be found from the threshold condition $u(t^{(2)}) = \vartheta$ or

$$\vartheta = R\, I_0 \left[1 - \exp\left(-\frac{t^{(2)} - t^{(1)}}{\tau_m}\right)\right] . \tag{4.7}$$

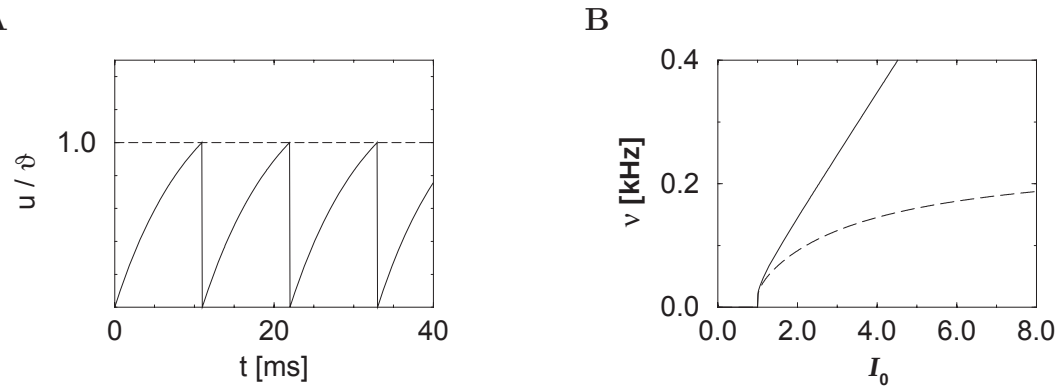

Fig. 4.2: **A**. Time course of the membrane potential of an integrate-and-fire neuron driven by constant input current $I_0 = 1.5$. The voltage $u(t)$ is normalized by the value of the threshold $\vartheta = 1$. **B**. Gain function. The firing rate ν of an integrate-and-fire neuron without (solid line) and with absolute refractoriness of $\delta_{\rm abs} = 4$ ms (dashed line) as a function of a constant driving current I_0. Current units are normalized so that the onset of repetitive firing is at $I_\theta = 1$. Other parameters are $R = 1$, $\tau_m = 10\,$ms, and $u_r = 0$.

Solving (4.7) for the time interval $T = t^{(2)} - t^{(1)}$ yields

$$T = \tau_m \ln \frac{R\,I_0}{R\,I_0 - \vartheta}\,. \tag{4.8}$$

After the spike at $t^{(2)}$ the membrane potential is again reset to $u_r = 0$ and the integration process starts again. If the stimulus I_0 remains constant, the following spike will occur after another interval of duration T. We conclude that for a constant input current I_0, the integrate-and-fire neuron fires regularly with period T given by (4.8). For a neuron with absolute refractory period the firing period T' is given by $T' = T + \Delta^{\rm abs}$ with T defined by Eq. (4.8). In other words, the interspike interval is longer by an amount $\Delta^{\rm abs}$ compared to that of a neuron without absolute refractory period.

The mean firing rate of a noiseless neuron is defined as $\nu = 1/T$. The firing rate of an integrate-and-fire model with absolute refractory period $\Delta^{\rm abs}$ stimulated by a current I_0 is therefore

$$\nu = \left[\Delta^{\rm abs} + \tau_m \ln \frac{R\,I_0}{R\,I_0 - \vartheta}\right]^{-1}\,. \tag{4.9}$$

In Fig. 4.2B the firing rate is plotted as a function of the constant input I_0 for neurons with and without absolute refractory period.

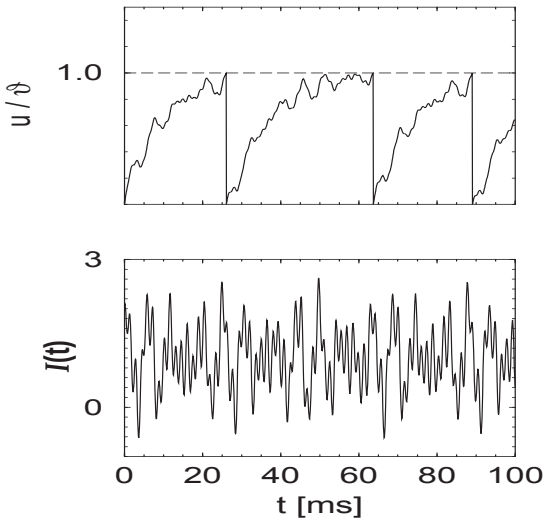

Fig. 4.3: Voltage $u(t)$ of an integrate-and-fire model (top) driven by the input current $I(t)$ shown at the bottom. The input $I(t)$ consists of a superposition of four sinusoidal components at randomly chosen frequencies plus a positive bias current $I_0 = 1.2$ which drives the membrane potential towards the threshold.

Example: Time-dependent stimulus $I(t)$

The results of the preceding example can be generalized to arbitrary stimulation conditions and an arbitrary reset value $u_r < \vartheta$. Let us suppose that a spike has occurred at $\hat{t}$. For $t > \hat{t}$ the stimulating current is $I(t)$. The value u_r will be treated as an initial condition for the integration of (4.3), i.e.,

$$u(t) = u_r \exp\left(-\frac{t-\hat{t}}{\tau_m}\right) + \frac{1}{C}\int_0^{t-\hat{t}} \exp\left(-\frac{s}{\tau_m}\right) I(t-s)\,\mathrm{d}s\,. \tag{4.10}$$

This expression describes the membrane potential for $t > \hat{t}$ and is valid up to the moment of the next threshold crossing. If $u(t) = \vartheta$, the membrane potential is reset to u_r and integration restarts; see Fig. 4.3.

4.1.2 Nonlinear integrate-and-fire model

In a general *nonlinear* integrate-and-fire model, Eq. (4.3) is replaced by

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} u = F(u) + G(u)\, I\,; \tag{4.11}$$

cf. Abbott and van Vreeswijk (1993). As before, the dynamics is stopped if u reaches the threshold ϑ and reinitialized at $u = u_r$. A comparison with Eq. (4.3)

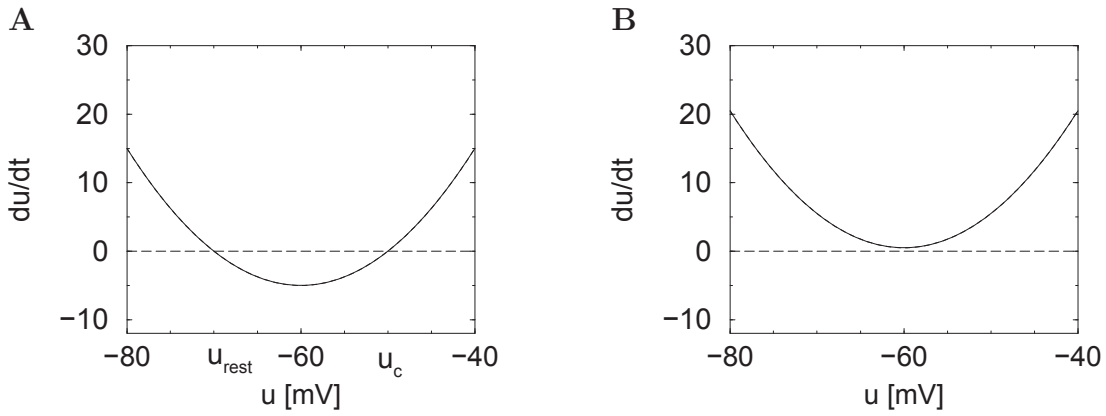

Fig. 4.4: Quadratic integrate-and-fire model. **A**. Without external current $I = 0$, the membrane potential relaxes for all initial condition $u < u_c$ to the resting potential $u_{\rm rest}$. If the membrane potential is moved above u_c, the potential increases further since $\mathrm{d}u/\mathrm{d}t > 0$. The neuron is said to fire if u reaches the threshold $\vartheta = -40\,\mathrm{mV}$. **B**. A constant super-threshold current I is characterized by the fact that $\mathrm{d}u/\mathrm{d}t > 0$ for all u. If u reaches the firing threshold of -40 mV, it is reset to -80 mV. This results in repetitive firing.

shows that $G(u)$ can be interpreted as a voltage-dependent input resistance while $-F(u)/(u-u_{\rm rest})$ corresponds to a voltage-dependent decay constant. A specific instance of a nonlinear integrate-and-fire model is the *quadratic* model (Latham et al., 2000; Hansel and Mato, 2001; Feng, 2001),

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} u = a_0 \,(u - u_{\rm rest})\,(u - u_c) + RI \;, \tag{4.12}$$

with parameters $a_0 > 0$ and $u_c > u_{\rm rest}$; cf. Fig. 4.4. For $I = 0$ and initial conditions $u < u_c$, the voltage decays to the resting potential $u_{\rm rest}$. For $u > u_c$ it increases so that an action potential is triggered. The parameter u_c can therefore be interpreted as the critical voltage for spike initiation by a short current pulse. We will see in the next example that the quadratic integrate-and-fire model is closely related to the so-called Θ-neuron, a canonical type-I neuron model (Ermentrout, 1996; Latham et al., 2000).

Rescaling and standard forms (*)

It is always possible to rescale the variables so that threshold and membrane time constant are equal to unity and that the resting potential vanishes. Furthermore, there is no need to interpret the variable u as the membrane potential. For example, starting from the nonlinear integrate-and-fire model Eq. (4.11), we can

introduce a new variable $\tilde{u}$ by the transformation

$$u(t) \longrightarrow \tilde{u}(t) = \tau \int_0^{u(t)} \frac{\mathrm{d}x}{G(x)} \tag{4.13}$$

which is possible if $G(x) \neq 0$ for all x in the integration range. In terms of $\tilde{u}$ we have a new nonlinear integrate-and-fire model of the form

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t} = \gamma(\tilde{u}) + I(t) \tag{4.14}$$

with $\gamma(\tilde{u}) = \tau F(u)/G(u)$. In other words, a general integrate-and-fire model (4.11) can always be reduced to the standard form (4.14). By a completely analogous transformation, we could eliminate the function F in Eq. (4.11) and move all the dependence into a new voltage dependent G (Abbott and van Vreeswijk, 1993).

Example: Relation to a canonical type I model (*)

In this section, we show that there is a close relation between the quadratic integrate-and-fire model (4.12) and the canonical type I phase model,

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = [1 - \cos\phi] + \Delta I\,[1 + \cos\phi]\,; \tag{4.15}$$

cf. Section 3.2.4 (Ermentrout, 1996; Ermentrout and Kopell, 1986; Strogatz, 1994; Hoppensteadt and Izhikevich, 1997; Latham et al., 2000).

Let us denote by I_θ the minimal current necessary for repetitive firing of the quadratic integrate-and-fire neuron. With a suitable shift of the voltage scale and constant current $I = I_\theta + \Delta I$ the equation of the quadratic neuron model can then be cast into the form

$$\frac{\mathrm{d}u}{\mathrm{d}t} = u^2 + \Delta I\,. \tag{4.16}$$

For $\Delta I > 0$ the voltage increases until it reaches the firing threshold $\vartheta \gg 1$ where it is reset to a value $u_r \ll -1$. Note that the firing times are insensitive to the actual values of firing threshold and reset value because the solution of Eq. (4.16) grows faster than exponentially and diverges for finite time (hyperbolic growth). The difference in the firing times for a finite threshold of, say, $\vartheta = 10$ and $\vartheta = 10\,000$ is thus negligible.

We want to show that the differential equation (4.16) can be transformed into the canonical phase model (4.15) by the transformation

$$u(t) = \tan\left(\frac{\phi(t)}{2}\right)\,. \tag{4.17}$$

To do so, we take the derivative of (4.17) and use the differential equation (4.15) of the generic phase model. With help of the trigonometric relations $\mathrm{d}\tan x/\mathrm{d}x = 1/\cos^2(x)$ and $1+\cos x = \cos^2(x/2)$ we find

$$\begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}t} &= \frac{1}{\cos^2(\phi/2)} \frac{1}{2} \frac{\mathrm{d}\phi}{\mathrm{d}t} \\ &= \tan^2(\phi/2) + \Delta I = u^2 + \Delta I \,. \end{aligned} \tag{4.18}$$

Thus Eq. (4.17) with $\phi(t)$ given by (4.15) is a solution to the differential equation of the quadratic integrate-and-fire neuron. The quadratic integrate-and-fire neuron is therefore (in the limit $\vartheta \to \infty$ and $u_r \to -\infty$) equivalent to the generic type I neuron (4.15).

4.1.3 Stimulation by Synaptic Currents

So far we have considered an isolated neuron that is stimulated by an external current $I(t)$. In a more realistic situation, the integrate-and-fire model is part of a larger network and the input current $I(t)$ is generated by the activity of presynaptic neurons.

In the framework of the integrate-and-fire model, each presynaptic spike generates a postsynaptic current pulse. More precisely, if the presynaptic neuron j has fired a spike at $t_j^{(f)}$, a postsynaptic neuron i 'feels' a current with time course $\alpha(t-t_j^{(f)})$. The total input current to neuron i is the sum over all current pulses,

$$I_i(t) = \sum_j w_{ij} \sum_f \alpha(t - t_j^{(f)}) \,. \tag{4.19}$$

The factor w_{ij} is a measure of the efficacy of the synapse from neuron j to neuron i.

Though Eq. (4.19) is a reasonable model of synaptic interaction, reality is somewhat more complicated, because the amplitude of the postsynaptic current pulse depends on the actual value of the membrane potential u_i. As we have seen in Chapter 2, each presynaptic action potential evokes a change in the *conductance* of the postsynaptic membrane with a certain time course $g(t-t^{(f)})$. The postsynaptic current generated by a spike at time $t_j^{(f)}$ is thus

$$\alpha(t - t_j^{(f)}) = -g(t - t_j^{(f)}) \, [u_i(t) - E_{\rm syn}] \,. \tag{4.20}$$

The parameter $E_{\rm syn}$ is the reversal potential of the synapse.

The level of the reversal potential depends on the type of synapse. For excitatory synapses, $E_{\rm syn}$ is much larger than the resting potential. For a voltage $u_i(t)$ close to the resting potential, we have $u_i(t) < E_{\rm syn}$. Hence the current I_i induced by a presynaptic spike at an excitatory synapse is positive and *increases*

the membrane potential [1]. The higher the voltage, the smaller the amplitude of the input current. Note that a positive voltage $u_i > u_{\rm rest}$ is itself the result of input spikes which have arrived at other excitatory synapses. Hence, there is a saturation of the postsynaptic current and the total input current is not just the sum of independent contributions. Nevertheless, since the reversal potential of excitatory synapses is usually significantly above the firing threshold, the factor $[u_i - E_{\rm syn}]$ is almost constant and saturation can be neglected.

For inhibitory synapses, the reversal potential is close to the resting potential. An action potential arriving at an inhibitory synapse pulls the membrane potential towards the reversal potential $E_{\rm syn}$. Thus, if the neuron is at rest, inhibitory input hardly has any effect on the membrane potential. If the membrane potential is instead considerably above the resting potential, then the same input has a strong inhibitory effect. This is sometimes described as silent inhibition: inhibition is only seen if the membrane potential is above the resting potential. Strong silent inhibition is also called 'shunting' inhibition, because a significantly reduced resistance of the membrane potential forms a short circuit that literally shunts excitatory input the neuron might receive from other synapses.

Example: Pulse-coupling and α-function

The time course of the postsynaptic current $\alpha(s)$ introduced in Eq. (4.19) can be defined in various ways. The simplest choice is a Dirac δ-pulse, $\alpha(s) = q\,\delta(s)$, where q is the total charge that is injected in a postsynaptic neuron via a synapse with efficacy $w_{ij} = 1$. More realistically, the postsynaptic current α should have a finite duration, e.g., as in the case of an exponential decay with time constant τ_s,

$$\alpha(s) = \frac{q}{\tau_s} \exp\left(-\frac{s}{\tau_s}\right) \Theta(s)\,. \tag{4.21}$$

As usual, Θ is the Heaviside step function with $\Theta(s) = 1$ for $s > 0$ and $\Theta(s) = 0$ else. Equation (4.21) is a simple way to account for the low-pass characteristics of synaptic transmission; cf. Fig. 4.1.

An even more sophisticated version of α includes a finite rise time τ_r of the postsynaptic current and a transmission delay $\Delta^{\rm ax}$,

$$\alpha(s) = \frac{q}{\tau_s - \tau_r}\left[\exp\left(-\frac{s-\Delta^{\rm ax}}{\tau_s}\right) - \exp\left(-\frac{s-\Delta^{\rm ax}}{\tau_r}\right)\right] \Theta(s-\Delta^{\rm ax})\,. \tag{4.22}$$

In the limit of $\tau_r \to \tau_s$, (4.22) yields

$$\alpha(s) = q\,\frac{s-\Delta^{\rm ax}}{\tau_s^2} \exp\left(-\frac{s-\Delta^{\rm ax}}{\tau_s}\right) \Theta(s-\Delta^{\rm ax})\,. \tag{4.23}$$

[1] Note that in Eq. (4.20) we consider the synaptic current as an *external* current whereas in Chapter 2 we have considered it as a membrane current and therefore used a different sign convention. In both cases, an excitatory input increases the membrane potential.

In the literature, a function of the form $x \exp(-x)$ such as (4.23) is often called an α-function. While this has motivated our choice of the symbol α for the synaptic input current, α may stand for any form of an input current pulse.

4.2 Spike response model (SRM)

The Spike Response Model (SRM) is – just like the nonlinear integrate-and-fire model – a generalization of the leaky integrate-and-fire model. The direction of the generalization is, however, somewhat different. In the nonlinear integrate-and-fire model, parameters are made *voltage* dependent whereas in the SRM they depend on the time since the last output spike. Another difference between integrate-and-fire models and the SRM concerns the formulation of the equations. While integrate-and-fire models are usually defined in terms of differential equations, the SRM expresses the membrane potential at time t as an integral over the past.

The explicit dependence of the membrane potential upon the last output spike allows us to model refractoriness as a combination of three components, viz., (i) a reduced responsiveness after an output spike; (ii) an increase in threshold after firing; and (iii) a hyperpolarizing spike after-potential. In Section 4.2.1 the Spike Response Model is introduced and its properties illustrated. Its relation to the integrate-and-fire model is the topic of Section 4.2.2. An important special case of the Spike Response Model is the simplified model SRM_0 that we have already encountered in Chapter 1.3.1. Section 4.2.3 will discuss it in more detail.

4.2.1 Definition of the SRM

In the framework of the Spike Response Model the state of a neuron i is described by a single variable u_i. In the absence of spikes, the variable u_i is at its resting value, $u_{\rm rest} = 0$. Each incoming spike will perturb u_i and it takes some time before u_i returns to zero. The function ϵ describes the time course of the response to an incoming spike. If, after the summation of the effects of several incoming spikes, u_i reaches the threshold ϑ an output spike is triggered. The form of the action potential and the after-potential is described by a function η. Let us suppose neuron i has fired its last spike at time $\hat{t}_i$. After firing the evolution of u_i is given by

$$\begin{aligned} u_i(t) &= \eta(t-\hat{t}_i) + \sum_j w_{ij} \sum_f \epsilon_{ij}(t-\hat{t}_i, t-t_j^{(f)}) \\ &\quad + \int_0^\infty \kappa(t-\hat{t}_i, s)\, I^{\rm ext}(t-s)\, {\rm d}s \end{aligned} \tag{4.24}$$

where $t_j^{(f)}$ are spikes of presynaptic neurons j and w_{ij} is the synaptic efficacy. The last term accounts for an external driving current $I^{\rm ext}$. The two sums run over all presynaptic neurons j and all firing times $t_j^{(f)} < t$ of neuron j. We emphasize that all terms depend on $t - \hat{t}_i$, i.e., the time since the last output spike.

In contrast to the integrate-and-fire neuron discussed in Section (4.1) the threshold ϑ is not fixed but may also depend on $t - \hat{t}_i$

$$\vartheta \quad \longrightarrow \quad \vartheta(t - \hat{t}_i)\,. \tag{4.25}$$

During an absolute refractory period $\Delta^{\rm abs}$, we may for example set ϑ to a large and positive value to avoid firing and let it relax back to its equilibrium value for $t > \hat{t}_i + \Delta^{\rm abs}$. Firing occurs whenever the membrane potential u_i reaches the dynamic threshold $\vartheta(t - \hat{t}_i)$ from below

$$t = t_i^{(f)} \quad \Leftrightarrow \quad u_i(t) = \vartheta(t - \hat{t}_i) \text{ and } \frac{{\rm d}u_i(t)}{{\rm d}t} > 0\,. \tag{4.26}$$

As mentioned before $\hat{t}_i$ is the *last* firing time,

$$\hat{t}_i = \max\left\{t_i^{(f)} < t\right\}\,. \tag{4.27}$$

Dynamic thresholds are a standard feature of phenomenological neuron models (Fuortes and Mantegazzini, 1962; Geisler and Goldberg, 1966; Weiss, 1966; Stein, 1967b; MacGregor and Oliver, 1974; Horn and Usher, 1989; Eckhorn et al., 1990; Abeles, 1991). Models similar to Eqs. (4.24)–(4.26) can be traced back much further; see, e.g., Hill (1936).

Interpretation

So far Eqs. (4.1) and (4.24) define a mathematical model. Can we give a biological interpretation of the terms? Let us identify the variable u_i with the membrane potential of neuron i. The functions η, κ and ϵ_{ij} are *response kernels* that describe the effect of spike emission and spike reception on the variable u_i. This interpretation has motivated the name 'Spike Response Model', SRM for short (Gerstner, 1995; Kistler et al., 1997). Let us discuss the meaning of the response kernels.

The kernel η describes the standard form of an action potential of neuron i including the negative overshoot which typically follows a spike (after-potential). Graphically speaking, a contribution η is 'pasted in' each time the membrane potential reaches the threshold ϑ; cf. Fig. 4.5. Since the form of the spike is always the same, the exact time course of the action potential carries no information. What matters is whether there is the event 'spike' or not. The event is fully characterized by the firing time $t_i^{(f)}$. In a simplified model, the *form* of the action potential may therefore be neglected as long as we keep track of the firing times

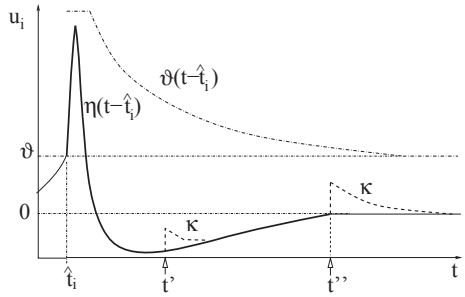

Fig. 4.5: Schematic interpretation of the Spike Response Model. The figure shows the time course $u_i(t)$ of the membrane potential of neuron i as a function of time t. A spike of neuron i has been initiated at $\hat{t}_i$. The kernel $\eta(t-\hat{t}_i)$ for $t > \hat{t}_i$ describes the form of the action potential (positive pulse) and the (negative) spike after-potential that follows the pulse (thick solid line). If an input current pulse is applied at a time t'' a long time after the firing at $\hat{t}_i$, it evokes a standard response described by the function $\kappa(\infty, t-t'')$ and indicated by the dashed line starting at t'' (arrow). An input current pulse at t' (arrow) which arrives shortly after the postsynaptic spike at $\hat{t}_i$ evokes, due to refractoriness of the neuron, a response of significantly shorter duration. Its time course is described by the response kernel $\kappa(t-\hat{t}_i, t-t')$; see the dashed line after t'. Immediately after firing at $\hat{t}_i$, the threshold is increased (dot-dashed line).

$t_i^{(f)}$. The kernel η describes then simply the 'reset' of the membrane potential to a lower value after the spike at $\hat{t}_i$ just like in the integrate-and-fire model. The leaky integrate-and-fire model is in fact a special case of the SRM as we will see below in Section 4.2.2.

The kernel $\kappa(t-\hat{t}_i, s)$ is the *linear response* of the membrane potential to an input current. It describes the time course of a deviation of the membrane potential from its resting value that is caused by a short current pulse ("impulse response"). We have already seen in Chapters 2.2 and 3 that the response depends, in general, on the time that has passed since the last output spike at $\hat{t}_i$. Immediately after $\hat{t}_i$ many ion channels are open so that the resistance of the membrane is reduced. The voltage response to an input current pulse decays therefore more rapidly back to zero than in a neuron that has been inactive. A reduced or shorter response is one of the signatures of neuronal refractoriness. This form of refractory effect is taken care of by making the kernel κ depend, via its first argument, on the time difference $t-\hat{t}_i$. We illustrate the idea in Fig. 4.5. The response to a first input pulse at t' is shorter and less pronounced than that to a second one at t'', an effect which is well-known experimentally (Fuortes and Mantegazzini, 1962; Powers and Binder, 1996; Stevens and Zador, 1998).

The kernel $\epsilon_{ij}(t-\hat{t}_i, s)$ as a function of $s = t-t_j^{(f)}$ can be interpreted as the time course of a *postsynaptic potential* evoked by the firing of a presynaptic neuron j at

time $t_j^{(f)}$. Depending on the sign of the synapse from j, to i, ϵ_{ij} models either an excitatory or inhibitory postsynaptic potential (EPSP or IPSP). Similarly as for the kernel κ, the exact shape of the postsynaptic potential depends on the time $t-\hat{t}_i$ that has passed since the last spike of the postsynaptic neuron i. In particular, if neuron i has been active immediately before the arrival of a presynaptic action potential, the postsynaptic neuron is in a state of refractoriness. In this case, the response to an input spike is smaller than that of an 'unprimed' neuron. The first argument of $\epsilon_{ij}(t-\hat{t}_i, s)$ accounts for the dependence upon the last firing time of the postsynaptic neuron.

Total Postsynaptic Potential

In order to simplify the notation for later use, it is convenient to introduce the *total postsynaptic potential*,

$$h(t|\hat{t}_i) = \sum_j w_{ij} \sum_{t_j^{(f)}} \epsilon_{ij}(t-\hat{t}_i, t-t_j^{(f)}) + \int_0^\infty \kappa(t-\hat{t}_i, s)\, I_i^{\text{ext}}(t-s)\, \mathrm{d}s\,. \quad (4.28)$$

Equation (4.24) can then be written in compact form,

$$u_i(t) = \eta(t-\hat{t}_i) + h(t|\hat{t}_i)\,. \quad (4.29)$$

Refractoriness

Refractoriness may be characterized experimentally by the observation that immediately after a first action potential it is impossible (absolute refractoriness) or more difficult (relative refractoriness) to excite a second spike (Fuortes and Mantegazzini, 1962).

Absolute refractoriness can be incorporated in the SRM by setting the dynamic threshold during a time Δ^{abs} to an extremely high value that cannot be attained.

Relative refractoriness can be mimicked in various ways; see Fig. 4.5. First, after the spike the membrane potential, and hence η, passes through a regime of hyperpolarization (after-potential) where the voltage is *below* the resting potential. During this phase, more stimulation than usual is needed to drive the membrane potential above threshold. This is equivalent to a transient increase of the firing threshold (see below). Second, ϵ and κ contribute to relative refractoriness because, immediately after an action potential, the response to incoming spikes is shorter and, possibly, of reduced amplitude (Fuortes and Mantegazzini, 1962). Thus more input spikes are needed to evoke the same depolarization of the membrane potential as in an 'unprimed' neuron. The first argument of the ϵ function (or κ function) allows us to incorporate this effect.

Removing the dynamic threshold

From a formal point of view, there is no need to interpret the variable u as the membrane potential. It is, for example, often convenient to transform the variable u so as to remove the time-dependence of the threshold. In fact, a general Spike Response Model with arbitrary time-dependent threshold $\vartheta(t-\hat{t}) = \vartheta_0 + \Delta(t-\hat{t})$, can always be transformed into a Spike Response Model with fixed threshold ϑ_0 by a change of variables

$$u(t) \longrightarrow \tilde{u}(t) = u(t) - \Delta(t-\hat{t})\,. \tag{4.30}$$

The function $\Delta(t-\hat{t})$ can easily be absorbed in the definition of the η kernel.

Example: Impulse response of the FitzHugh-Nagumo model

In Chapter 3 we have studied the FitzHugh-Nagumo model as an example of a two-dimensional neuron model. Here we want to show that the response of the FitzHugh-Nagumo model to a short input current pulse depends on the time since the last spike. Let us trigger, in a simulation of the model, an action potential at $t = 0$. This can be done by applying a short, but strong current pulse. The result is a voltage trajectory of large amplitude which we identify with the kernel $\eta(t)$. Figure 4.6 shows the hyperpolarizing spike after-potential which decays slowly back to the resting level. To test the responsiveness of the FitzHugh-Nagumo model during the recovery phase after the action potential, we apply, at a time $t^{(f)} > 0$, a second short input current pulse of low amplitude. The response to this test pulse is compared with the unperturbed trajectory. The difference between the two trajectories defines the kernel $\kappa(t-\hat{t}, t-t^{(f)})$. In Fig. 4.6 several trajectories are overlayed showing the response to stimulation at $t = 10, 15, 20, 30$ or 40. The shape and duration of the response curve depends on the time that has passed since the initiation of the action potential. Note that the time constant of the response kernel κ is always shorter than that of the hyperpolarizing spike after-potential. Analogous results for the Hodgkin-Huxley model will be discussed below in Section 4.3.1.

Example: A motoneuron model

Motoneurons exhibit a rather slow return to the resting potential after an action potential (Powers and Binder, 1996). The time constant of the decay of the hyperpolarizing spike after-potential can be in the range of 100 ms or more and is therefore much slower than the membrane time constant that characterizes the response to a short current input. On the other hand, it is found that if motoneurons are stimulated by a constant super-threshold current, their membrane potential has a roughly linear trajectory when approaching threshold. To qualitatively describe these observations, we can use a Spike Response Model with the

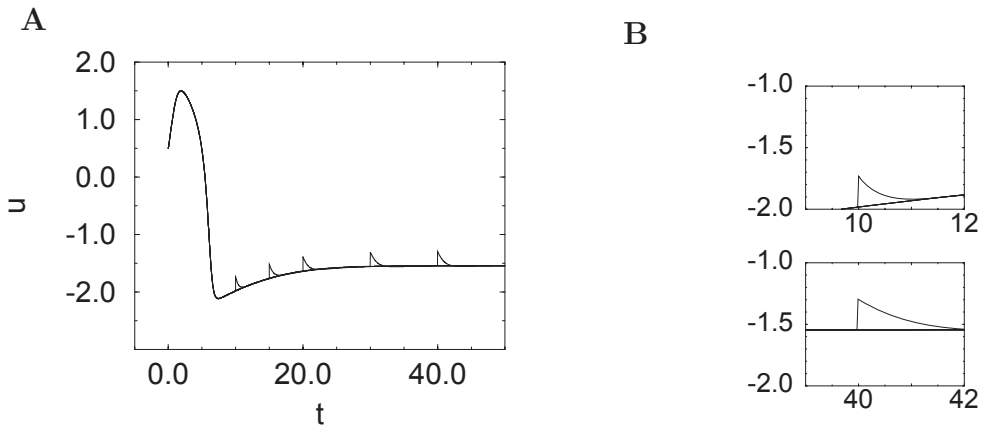

Fig. 4.6: FitzHugh-Nagumo model. An action potential has been triggered at $t = 0$. After the action potential additional pulse input occurs at $t = 10, 15, 20, 30,$ or 40 [arbitrary units]. In **A** the trajectories of all runs are plotted on top of each other. Part **B** shows a zoomed-in section of two trajectories. A pulse input at time t=10 after the onset of the action potential has a short lasting effect (top right) compared to a pulse at t=40 (bottom right). All parameters as in Fig. 3.5. There is no constant bias current.

following kernels:

$$\eta(t-\hat{t}) = -\eta_0 \, e^{-\frac{(t-\hat{t})}{\tau_{\rm refr}}} \, \Theta(t-\hat{t}) \tag{4.31}$$

$$\kappa(t-\hat{t},s) = \frac{R}{\tau_m}\left[1-e^{-\frac{(t-\hat{t})}{\tau_{\rm rec}}}\right] e^{-\frac{s}{\tau_m}} \, \Theta(s) \, \Theta(t-\hat{t}-s) \tag{4.32}$$

where τ_m is an effective passive membrane time constant, R is the input resistance, $\tau_{\rm refr}$ is the 'refractory' time constant, $\tau_{\rm rec}$ is the 'response recovery' time constant, η_0 is a scale factor for the refractory function. The passive membrane time constant τ_m and input resistance R characterize the membrane response to small current pulses. The refractory function η describes the return of the membrane potential to baseline after an action potential. It is characterized by a slow time constant $\tau_{\rm refr}$. For the κ-kernel we use a decaying exponential in s with time constant τ_m, modulated by the 'recovery' factor $\{1 - \exp[-(t-\hat{t})/\tau_{\rm rec}]\}$. This results in a spike-time dependent scaling of the amplitude of postsynaptic potentials. The recovery time $\tau_{\rm rec}$ is taken much longer than τ_m.

The effect of the modulation of the input conductance as a function of $t - \hat{t}$ is depicted in Fig. 4.7. An input current pulse shortly after the reset at time $\hat{t}$ evokes a postsynaptic potential of much lower amplitude than an input current pulse that arrives much later. Fig. 4.7 qualitatively reproduces the membrane trajectory of motoneurons when stimulated by the same input pattern (Powers and Binder, 1996; Poliakov et al., 1996).

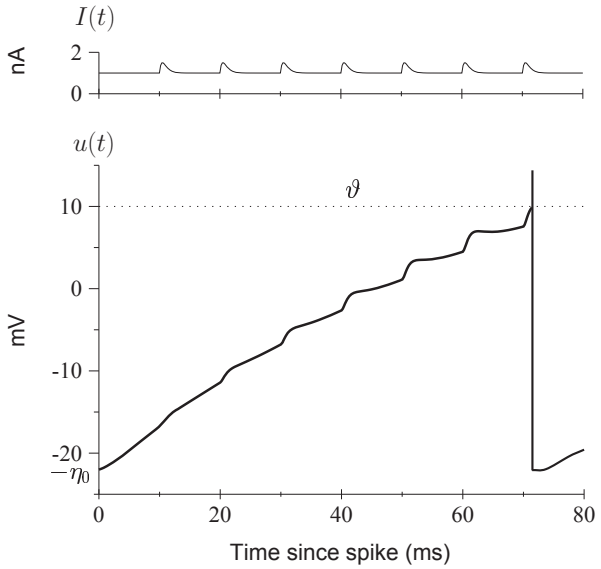

Fig. 4.7: Effect of recovery time constant $\tau_{\rm rec}$. Top: Input current consisting of a sequence of pulses superimposed on a constant bias. Bottom: The membrane potential response (thick line) to the input pulses clearly shows that the response amplitude increases as a function of the time since the last spike. Parameters: $\tau_{\rm rec} = \tau_{\rm refr} = 100\,{\rm ms}$; $\tau_m = 4\,{\rm ms}$; taken from Herrmann and Gerstner (2001b).

4.2.2 Mapping the Integrate-and-Fire Model to the SRM

In this section, we show that the leaky integrate-and-fire neuron defined in Section 4.1 is a special case of the Spike Response Model. We consider an integrate-and-fire neuron driven by external current $I^{\rm ext}$ and postsynaptic current pulses $\alpha(t - t_j^{(f)})$. The potential u_i is thus given by

$$\tau_m \frac{{\rm d}u_i}{{\rm d}t} = -u_i(t) + R \sum_j w_{ij} \sum_f \alpha(t - t_j^{(f)}) + R\, I_i^{\rm ext}(t)\,. \tag{4.33}$$

In order to construct a mapping of the integrate-and-fire model to the Spike Response Model, we integrate Eq. (4.33) with $u(\hat{t}_i) = u_r$ as its initial condition.

The result is in analogy to Eq. (4.10)

$$\begin{aligned} u(t) &= u_r \exp\left(-\frac{t-\hat{t}_i}{\tau_m}\right) \\ &\quad + \sum_j w_{ij} \sum_f \frac{1}{C} \int_0^{t-\hat{t}_i} \exp\left(-\frac{s}{\tau_m}\right) \alpha(t - t_j^{(f)} - s)\, \mathrm{d}s \\ &\quad + \frac{1}{C} \int_0^{t-\hat{t}_i} \exp\left(-\frac{s}{\tau_m}\right) I_i^{\mathrm{ext}}(t-s)\, \mathrm{d}s \\ &= \eta(t-\hat{t}_i) + \sum_j w_{ij} \sum_f \epsilon(t-\hat{t}_i, t-t_j^{(f)}) + \int_0^\infty \kappa(t-\hat{t}_i, s)\, I_i^{\mathrm{ext}}(t-s)\, \mathrm{d}s\,, \end{aligned} \tag{4.34}$$

with

$$\eta(s) = u_r \exp\left(-\frac{s}{\tau_m}\right)\,, \tag{4.35}$$

$$\epsilon(s,t) = \frac{1}{C} \int_0^s \exp\left(-\frac{t'}{\tau_m}\right) \alpha(t-t')\, \mathrm{d}t'\,, \tag{4.36}$$

$$\kappa(s,t) = \frac{1}{C} \exp\left(-\frac{t}{\tau_m}\right) \Theta(s-t)\, \Theta(t)\,. \tag{4.37}$$

As usual, $\Theta(x)$ denotes the Heaviside step function. The kernels (4.35)–(4.37) allow us to map the integrate-and-fire neuron exactly to the the Spike Response Model, as desired; cf. Eq. (4.24).

In order obtain an explicit expression for the ϵ kernel (4.36) we have to specify the time course of the postsynaptic current $\alpha(s)$. Here, we take $\alpha(s)$ as defined in (4.21), viz.,

$$\alpha(s) = \frac{q}{\tau_s} \exp\left(-s/\tau_s\right)\, \Theta(s)\,. \tag{4.38}$$

With $q = C = 1$, the integration of Eq. (4.36) yields

$$\epsilon(s,t) = \frac{\exp\left(-\frac{\max(t-s,0)}{\tau_s}\right)}{1-\frac{\tau_s}{\tau_m}} \left[\exp\left(-\frac{\min(s,t)}{\tau_m}\right) - \exp\left(-\frac{\min(s,t)}{\tau_s}\right)\right] \Theta(s)\, \Theta(t)\,; \tag{4.39}$$

cf. Fig. 4.8. If presynaptic spikes arrive before the last postsynaptic spike, then they have only a small effect on the actual value of the membrane potential because only that part of the postsynaptic current that arrives after $\hat{t}_i$ contributes to the postsynaptic potential. Spikes that arrive after $\hat{t}_i$ produce a full postsynaptic potential. Note that causality implies that the ϵ kernel has to vanishes for negative arguments.

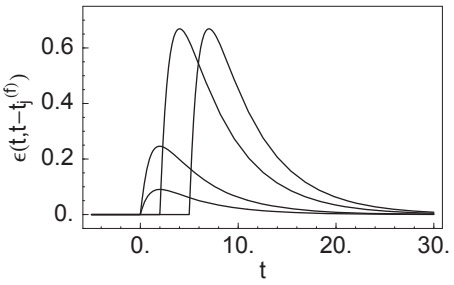

Fig. 4.8: The kernel $\epsilon(t, t - t_j^{(f)})$ as a function of t for various presynaptic firing times $t_j^{(f)} = -2$, -1, 2, 5; cf. Eq. (4.39) with $\tau_s = 1$ and $\tau_m = 5$. The last postsynaptic spike was at $\hat{t} = 0$. If presynaptic spikes arrive before the last postsynaptic spike, then they have only a small effect on the membrane potential; cf. the two small EPSPs that correspond to $t_j^{(f)} = -2$ and $t_j^{(f)} = -1$. If presynaptic spikes arrive after the last postsynaptic spike then they evoke a full-blown EPSP; cf. the two large EPSPs that correspond to $t_j^{(f)} = 2$ and $t_j^{(f)} = 5$.

Example: Spike-time dependent time constant

We have seen above that the Spike Response Model contains the integrate-and-fire model as a special case. In this example, we show in addition that even a generalization of the integrate-and-fire model that has a time dependent membrane time constant can be described within the SRM framework.

To be specific, we consider an integrate-and-fire model with spike-time dependent time constant, i.e., with a membrane time constant τ that is a function of the time since the last postsynaptic spike,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{u}{\tau(t-\hat{t})} + \frac{1}{C} I^{\mathrm{ext}}(t)\,; \tag{4.40}$$

cf. Wehmeier et al. (1989); Stevens and Zador (1998). As usual, $\hat{t}$ denotes the last firing time of the neuron. The neuron is insensitive to input during an absolute refractory period of duration Δ^{abs}. After the refractory period, the membrane potential is reset to a value u_r. Starting the integration of Eq. (4.40) at $u(\hat{t} + \Delta^{\mathrm{abs}}) = u_r$, we find for $t > \hat{t} + \Delta^{\mathrm{abs}}$

$$\begin{aligned} u(t) &= u_r \exp\left[-\int_{\hat{t}+\Delta^{\mathrm{abs}}}^{t} \frac{\mathrm{d}t'}{\tau(t'-\hat{t})}\right] \\ &\quad + \frac{1}{C}\int_0^{\infty} \Theta(t-\hat{t}-\Delta^{\mathrm{abs}}-s)\, \exp\left[-\int_{t-s}^{t} \frac{\mathrm{d}t'}{\tau(t'-\hat{t})}\right] I^{\mathrm{ext}}(t-s)\,\mathrm{d}s\,, \end{aligned} \tag{4.41}$$

which is a special case of Eq. (4.24). As we have seen above in Fig. 4.6, the effective membrane time constant of many standard neuron models is reduced

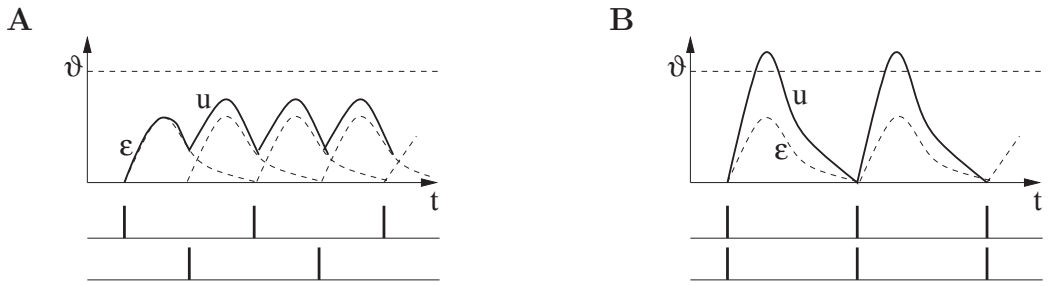

Fig. 4.28: Potential u of a postsynaptic neuron which receives input from two groups of presynaptic neurons. **A**. Spike trains of the two groups are phase shifted with respect to each other. The total potential u does not reach the threshold. There are no output spikes. **B**. Spikes from two presynaptic groups arrive synchronously. The summed EPSPs reach the threshold ϑ and cause the generation of an output spike.

where $N = 100$ is the number of presynaptic neurons. If the postsynaptic neuron has not been active in the recent past, we can neglect the refractory term η on the right-hand side of Eq. (4.112). The maximum of (4.112) occurs at $t = \tau = 10$ ms and has a value of $wNJ/e \approx 37$ mV which is above threshold. Thus the postsynaptic neuron fires before $t = 10$ ms. We conclude that the same number of input spikes can have different effects depending on their level of synchrony; cf. Fig. 4.28.

We will return to the question of coincidence detection, i.e., the distinction between synchronous and asynchronous input, in the following chapter. For a classical experimental study exploring the relevance of temporal structure in the input, see Segundo et al. (1963).

Example: Spatio-temporal summation

In neurons with a spatially extended dendritic tree the form of the postsynaptic potential depends not only on the type, but also on the location of the synapse; cf. Chapter 2. To be specific, let us consider a multi-compartment integrate-and-fire model. As we have seen above in Section 4.4, the membrane potential $u_i(t)$ can be described by the formalism of the Spike Response Model. If the last output spike $\hat{t}_i$ is long ago, we can neglect the refractory kernel η_i and the membrane potential is given by

$$u_i(t) = \sum_j w_{ij} \sum_f \epsilon_{ij}(t - t_j^{(f)}). \tag{4.113}$$

cf. Eq. (4.90). The subscript ij at the ϵ kernel takes care of the fact that the postsynaptic potential depends on the location of the synapse on the dendrite. Due to the low-pass characteristics of the dendrite, synaptic input at the tip of the dendrite causes postsynaptic potentials with a longer rise time and lower

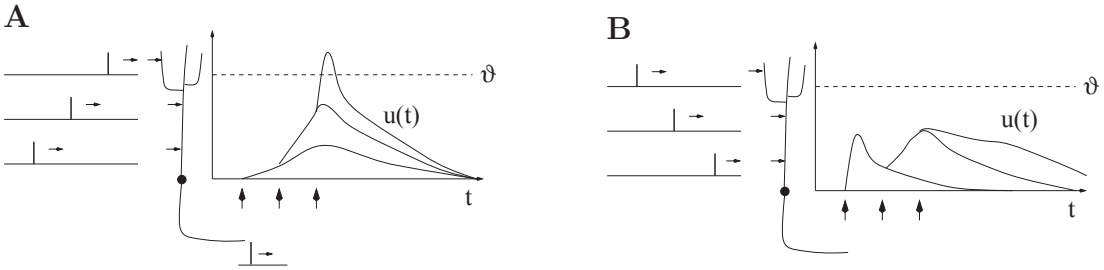

Fig. 4.29: Sensitivity to temporal order of synaptic inputs on a dendrite. **A**. A neuron is stimulated by three synaptic inputs in a sequence that starts at the distal part of the dendrite and ends with an input close to the soma. Since the EPSP caused by the distal input has a longer rise time than that generated by the proximal input, the EPSPs add up coherently and the membrane potential reaches the firing threshold ϑ. **B**. If the temporal sequence of spike inputs is reversed, the same number of input spikes does not trigger an action potential (schematic figure).

amplitude than input directly into the soma. The total potential $u_i(t)$ depends therefore on the temporal order of the stimulation of the synapses. An input sequence starting at the far end of the dendrite and approaching the soma is more effective in triggering an output spike than the same number of input spikes in reverse order; cf. Fig. 4.29.

4.6 Summary

In formal spiking neuron models, spikes are fully characterized by their firing time $t^{(f)}$ defined by a threshold criterion. Integrate-and-fire and Spike Response Model are typical examples of spiking neuron models. Leaky integrate-and-fire point neurons with current input can be mapped exactly to the Spike Response Model. Even multi-compartment integrate-and-fire models can be mapped to the Spike Response Model, if indirect effects due to previous output spikes are neglected. An integrate-and-fire model with *spike-time* dependent parameters, which is a generalization of the leaky integrate-and-fire model, can be seen as a special case of the Spike Response Model. The nonlinear integrate-and-fire model, i.e., a model where parameters are *voltage* dependent is a different generalization. The quadratic integrate-and-fire model is particularly interesting since it is a generic example for a type I neuron model.

Detailed conductance based neuron models can be approximately mapped to formal spiking neuron models. With the help of formal spiking neuron models, problems of pulse coding can be studied in a transparent graphical manner. The Spike Response Model, defined in this chapter, will be reconsidered in part II where systems of spiking neurons are analyzed.

Literature

Formal neuron models where spikes are triggered by a threshold process have been popular in the sixties (Stein, 1965, 1967b; Geisler and Goldberg, 1966; Weiss, 1966), but the ideas can be traced back much further (Lapicque, 1907; Hill, 1936). It has been recognized early that these models lend themselves for hardware implementations (French and Stein, 1970) and mathematical analysis (Stein, 1965, 1967a), and can be fitted to experimental data (Brillinger, 1988, 1992). Recent developments in computation and coding with formal spiking neurons has been reviewed in the book 'Pulsed Neural Networks' edited by Maass and Bishop (1998).

Chapter 5

Noise in Spiking Neuron Models

In vivo recordings of neuronal activity are characterized by a high degree of irregularity. The spike train of individual neurons is far from being periodic and relations between the firing patterns of several neurons seem to be random. If the electrical activity picked up by an extra-cellular electrode is made audible by a loudspeaker then we basically hear – noise. The question whether this is indeed just noise or rather a highly efficient way of coding information cannot easily be answered. Listening to a computer modem or a fax machine might also leave the impression that this is just noise. Being able to decide whether we are witnessing the neuronal activity that is underlying the composition of a poem (or the electronic transmission of a love letter) and not just meaningless noise is one of the most burning problems in Neuroscience.

Several experiments have been undertaken to tackle this problem. It seems as if neurons can react in a very reliable and reproducible manner to fluctuating currents that are injected via intracellular electrodes. As long as the same time-course of the injected current is used the action potentials occur with precisely the same timing relative to the stimulation (Mainen and Sejnowski, 1995; Bryant and Segundo, 1976). A related phenomenon can be observed by using non-stationary sensory stimulation. Spatially uniform random flicker, for example, elicits more or less the same spike train in retinal ganglion cells if the same flicker sequence is presented again (Berry et al., 1997). A similar behavior has been reported for motion-sensitive neurons of the visual system in flies (de Ruyter van Steveninck et al., 1997) and monkey cortex (Bair and Koch, 1996). On the other hand, neurons produce irregular spike trains in the absence of any temporally structured stimuli. Irregular spontaneous activity, i.e., activity that is not related in any obvious way to external stimulation, and trial-to-trial variations in neuronal responses are often considered as noise (Softky and Koch, 1993; Shadlen and Newsome, 1994).

The origin of the irregularity in the electrical activity of cortical neurons *in vivo* is poorly understood. In spiking neuron models such as the integrate-and-

fire or Spike Response Model (SRM), noise is therefore often added explicitly to neuronal dynamics so as to mimic the unpredictability of neuronal recordings. In this chapter we present three different ways to implement noise in models of neuronal networks, viz. escape noise (Section 5.3), slow noise in the parameters (Section 5.4), and diffusive noise (Section 5.5). In Section 5.6 we discuss the differences between subthreshold and superthreshold stimulation and explain its consequences for spike train variability. In the subthreshold regime, it is possible to relate the diffusive noise model to the escape noise model. Section 5.7 illustrates this relation. The noise models are finally applied to the phenomenon of stochastic resonance in Section 5.8 and compared with rate models in Section 5.9. Before we start with the discussion of the noise models, we review in Section 5.1 some experimental evidence for noise in neurons and introduce in Section 5.2 a statistical framework of spike train analysis.

5.1 Spike train variability

If neuron models such as the Hodgkin-Huxley or the integrate-and-fire model are driven by a sufficiently strong constant current, they generate a regular sequence of spikes. In neuronal models with adaptation currents[1] there might be a short transient phase at the beginning, but then all interspike intervals are constant. Spike trains of typical neurons *in vivo* show a much more irregular behavior. Whether the irregularity is the sign of noise or of a rich code is at present an open question (Softky and Koch, 1993; Shadlen and Newsome, 1994; Bair and Koch, 1996). In the first subsection we review some evidence for neuronal variability and spike train irregularity. We then discuss potential sources of noise.

5.1.1 Are neurons noisy?

Many *in vivo* experiments show noisy behavior of central neurons. The activity of neurons from the visual cortex, for example, can be recorded while a slowly moving bar is presented on a screen within the visual field of the animal (Hubel and Wiesel, 1959, 1977). As soon as the bar enters the neuron's receptive field the firing rate goes up until the bar leaves the receptive field at the opposite border. The spike train, however, varies considerably from trial to trial, if the same experiment is repeated several times. Furthermore, the very same neuron is spontaneously active even if the screen is blank and no external stimulus is applied. During spontaneous activity, the intervals between one spike and the next exhibit a large variability resulting in a broad distribution of interspike intervals; see e.g., Softky and Koch (1993).

[1] We neglect here intrinsically bursting and chaotic neurons.

Part II

Population Models

Chapter 6

Population Equations

In many areas of the brain neurons are organized in populations of units with similar properties. Prominent examples are columns in the somatosensory and visual cortex (Mountcastle, 1957; Hubel and Wiesel, 1962) and pools of motor neurons (Kandel and Schwartz, 1991). Given the large number of neurons within such a column or pool it is sensible to describe the mean activity of the neuronal population rather than the spiking of individual neurons. The idea of a population activity has already been introduced in Chapter 1.4. In a population of N neurons, we calculate the proportion of active neurons by counting the number of spikes $n_{\rm act}(t;t+\Delta t)$ in a small time interval Δt and dividing by N. Further division by Δt yields the *population activity*

$$A(t) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \frac{n_{\rm act}(t;t+\Delta t)}{N} = \frac{1}{N} \sum_{j=1}^{N} \sum_{f} \delta(t-t_j^{(f)}) \qquad (6.1)$$

where δ denotes the Dirac δ function. The double sum runs over all firing times $t_j^{(f)}$ of all neurons in the population. In other words the activity A is defined by a population average. Even though the activity has units of a rate, the population activity is quite different from a mean firing rate defined by temporal average; cf. Chapter 1.4.

Theories of population activity have a long tradition (Knight, 1972a; Wilson and Cowan, 1972, 1973; Amari, 1974; Abbott and van Vreeswijk, 1993; Gerstner and van Hemmen, 1992; Treves, 1993; Gerstner, 1995; Amit and Brunel, 1997a,b; Brunel and Hakim, 1999; Fusi and Mattia, 1999; Brunel, 2000; Gerstner, 2000b; Nykamp and Tranchina, 2000; Omurtag et al., 2000; Eggert and van Hemmen, 2001). In this chapter we study the properties of a large and homogeneous population of spiking neurons. Why do we restrict ourselves to large populations? If we repeatedly conduct the same experiment on a population of, say, one hundred potentially noisy neurons, the observed activity $A(t)$ defined in Eq. (6.1) will vary from one trial to the next. Therefore we cannot expect a population theory to

predict the activity measurements in a single trial. Rather all population activity equations that we discuss in this chapter predict the *expected* activity. For a large and homogeneous network, the observable activity is very close to the expected activity. For the sake of notational simplicity, we do not distinguish the observed activity from its expectation value and denote in the following the expected activity by $A(t)$.

After clarifying the notion of a homogeneous network in Section 6.1, we derive in Section 6.2 population density equations, i.e., partial differential equations that describe the probability that an arbitrary neuron in the population has a specific internal state. In some special cases, these density equations can be integrated and presented in the form of an integral equation. In Section 6.3 a general integral equation for the temporal evolution of the activity $A(t)$ that is exact in the limit of a large number of neurons is derived. In particular, we discuss its relation to the Wilson-Cowan equation, one of the standard models of population activity. In Section 6.4 we solve the population equation for the fixed points of the population activity and show that the neuronal gain function plays an important role. Finally, in Section 6.5 the approach is extended to multiple populations and its relation to neuronal field equations is discussed.

Most of the discussion in part II of the present book will be based upon the population equations introduced in this chapter. The population activity equations will allow us to study signal transmission and coding (cf. Chapter 7), oscillations and synchronization (cf. Chapter 8), and the formation of activity patterns in populations with a spatial structure (cf. Chapter 9). The aim of the present chapter is two-fold. Firstly, we want to provide the reader with the mathematical formalism necessary for a systematic study of spatial and temporal phenomena in large populations of neurons. Secondly, we want to show that various formulations of population dynamics that may appear quite different at a first glance, are in fact closely related. Paragraphs that are more mathematically oriented are marked by an asterix and can be omitted at a first reading.

6.1 Fully Connected Homogeneous Network

We study a large and homogeneous population of neurons; cf. Fig. 6.1. By homogeneous we mean that all neurons $1 \leq i \leq N$ are identical and receive the same external input $I_i^{\rm ext}(t) = I^{\rm ext}(t)$. Moreover, in a homogeneous population, the interaction strength between the neurons is taken to be uniform,

$$w_{ij} = \frac{J_0}{N}\,, \tag{6.2}$$

where J_0 is a parameter. For $J_0 = 0$ all neurons are independent; a value $J_0 > 0$ ($J_0 < 0$) implies excitatory (inhibitory) all-to-all coupling. The interaction

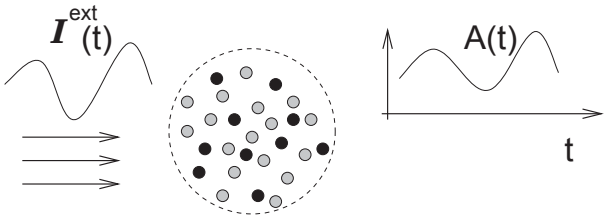

Fig. 6.1: Population of neurons (schematic). All neurons receive the same input $I^{\rm ext}(t)$ (left) which results in a time dependent population activity $A(t)$ (right).

strength scales with one over the number of neurons so that the total synaptic input to each neuron remains finite in the limit of $N \to \infty$.

Model neurons are described by formal spiking neurons as introduced in Chapter 4. In the case of leaky integrate-and-fire neurons with

$$\tau_m \frac{\mathrm{d}}{\mathrm{d}t} u_i = -u_i + R\, I_i(t) \tag{6.3}$$

a homogeneous network implies that all neurons have the same input resistance R, the same membrane time constant τ_m, as well as identical threshold and reset values. The input current I_i takes care of both the external drive and synaptic coupling

$$I_i = \sum_{j=1}^{N} \sum_f w_{ij} \alpha(t - t_j^{(f)}) + I^{\rm ext}(t)\,. \tag{6.4}$$

Here we have assumed that each input spike generates a postsynaptic current with some generic time course $\alpha(t - t_j^{(f)})$. The sum on the right-hand side of (6.4) runs over all firing times of all neurons. Because of the homogeneous all-to-all coupling, the total input current is identical for all neurons. To see this, we insert Eq. (6.2) and use the definition of the population activity, Eq. (6.1). We find a total input current,

$$I(t) = J_0 \int_0^\infty \alpha(s)\, A(t-s)\, \mathrm{d}s + I^{\rm ext}(t)\,, \tag{6.5}$$

which is independent of the neuronal index i. As an aside we note that for conductance-based synaptic input, the total input current would depend on the neuronal membrane potential which is different from one neuron to the next.

Instead of the integrate-and-fire neuron, we may also use the Spike Response Model (SRM) as the elementary unit of the population. The membrane potential of a SRM neuron is of the form

$$u_i(t) = \eta(t - \hat{t}_i) + h_{\rm PSP}(t|\hat{t}_i)\,, \tag{6.6}$$

Chapter 7

Signal Transmission and Neuronal Coding

In the preceding chapters, a theoretical description of neurons and neuronal populations has been developed. We are now ready to apply the theoretical framework to one of the fundamental problems of Neuroscience – the problem of neuronal coding and signal transmission. We will address the problem as three different questions, viz.,

(i) How does a population of neurons react to a fast change in the input? This question, which is particularly interesting in the context of reaction time experiments, is the topic of Section 7.2.

(ii) What is the response of an asynchronously firing population to an arbitrary time-dependent input current? This question points to the signal transfer properties as a function of stimulation frequency and noise. In Section 7.3 we calculate the signal transfer function for a large population as well as the signal-to-noise ratio in a finite population of, say, a few hundred neurons.

(iii) What is the 'meaning' of a single spike? If a neuronal population receives one extra input spike, how does this affect the population activity? On the other hand, if a neuron emits a spike, what do we learn about the input? These questions, which are intimately related to the problem of neural coding, are discussed in Section 7.4.

The population integral equation of Chapter 6.3 allows us to discuss these questions from a unified point of view. We focus in this chapter on a system of identical and independent neurons, i.e., a homogeneous network without lateral coupling. In this case, the behavior of the population as a whole is identical to the averaged behavior of a single neuron. Thus the signal transfer function discussed in Section 7.3 or the coding characteristics discussed in Section 7.4 can

also be interpreted as single-neuron properties. Before we dive into the main arguments we derive in Section 7.1 the *linearized* population equation that will be used throughout this chapter.

7.1 Linearized Population Equation

We consider a homogeneous population of independent neurons. All neurons receive the same current $I(t)$ fluctuating about the mean I_0. More specifically we set

$$I(t) = I_0 + \Delta I(t)\,. \tag{7.1}$$

For small fluctuations, $|\Delta I| \ll I_0$, we expect that the population activity stays close to the value A_0 that it would have for a constant current I_0, i.e.,

$$A(t) = A_0 + \Delta A(t)\,, \tag{7.2}$$

with $|\Delta A| \ll A_0$. In that case, we may expand the right-hand side of the population equation $A(t) = \int_{-\infty}^{t} P_I(t|\hat{t})\, A(\hat{t})\, \mathrm{d}\hat{t}$ into a Taylor series about A_0 to linear order in ΔA. In this section, we want to show that for spiking neuron models (either integrate-and-fire or SRM_0 neurons) the linearized population equation can be written in the form

$$\Delta A(t) = \int_{-\infty}^{t} P_0(t-\hat{t})\, \Delta A(\hat{t})\, \mathrm{d}\hat{t} + A_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_0^{\infty} \mathcal{L}(x)\, \Delta h(t-x)\, \mathrm{d}x\,, \tag{7.3}$$

where $P_0(t-\hat{t})$ is the interval distribution for constant input I_0, $\mathcal{L}(x)$ is a real-valued function that plays the role of an integral kernel, and

$$\Delta h(t) = \int_0^{\infty} \kappa(s)\, \Delta I(t-s)\, \mathrm{d}s \tag{7.4}$$

is the input potential generated by the time-dependent part of the input current. The first term of the right-hand side of Eq. (7.3) takes into account that previous perturbations $\Delta A(\hat{t})$ with $\hat{t} < t$ have an after-effect one inter-spike interval later. The second term describes the immediate response to a change in the input potential. If we want to understand the response of the population to an input current $\Delta I(t)$, we need to know the characteristics of the kernel $\mathcal{L}(x)$. The main task of this section is therefore the calculation of $\mathcal{L}(x)$.

Here we give an overview of the main results that we will obtain in the present chapter; explicit expressions for the kernel $\mathcal{L}(x)$ are presented in Tab. 7.1.

(i) In the low-noise limit, the kernel $\mathcal{L}(x)$ is a Dirac δ function. The dynamics of the population activity ΔA has therefore a term proportional to the *derivative* of the input potential; cf. Eq. (7.3). We will see that this result implies a fast response ΔA to any change in the input.

(ii) For high noise, the kernel $\mathcal{L}(x)$ depends critically on the noise model. For noise that is slow compared to the intrinsic neuronal dynamics (e.g., noise in the reset or stochastic spike arrival in combination with a slow synaptic time constant) the kernel $\mathcal{L}(x)$ is similar to that in the noise-free case. Thus the dynamics of ΔA is proportional to the *derivative* of the input potential and therefore fast.

(iii) For a large amount of 'fast' noise (e.g., escape noise), the kernel $\mathcal{L}(x)$ is broad so that the dynamics of the population activity is rather proportional to the input potential than to its derivative; cf. Eq. (7.3). As we will see, this implies that the response to a change in the input is slow.

Results for escape noise and reset noise have been derived by Gerstner (2000b) while results for diffusive noise have been presented by Brunel et al. (2001) based on a linearization of the membrane potential density equation (Brunel and Hakim, 1999). The effect of slow noise in parameters has already been discussed in Knight (1972a). Apart from the approach discussed in this section, a fast response of a population of integrate-and-fire neurons with diffusive noise can also be induced if the *variance* of the diffusive noise is changed (Lindner and Schimansky-Geier, 2001; Bethge et al., 2001).

Before we turn to the general case, we will focus in Section 7.1.1 on a noise-free population. We will see why the dynamics of $\Delta A(t)$ has a contribution proportional to the *derivative* of the input potential. In Section 7.1.2 we derive the general expression for the kernel $\mathcal{L}(x)$ and apply it to different situations. Readers not interested in the mathematical details may skip the remainder of this section and move directly to Section 7.2.

7.1.1 Noise-free Population Dynamics (*)

We start with a reduction of the population integral equation (6.75) to the noise-free case. In the limit of no noise, the input-dependent interval distribution $P_I(t \,|\, \hat{t})$ reduces to a Dirac δ function, i.e.,

$$P_I(t \,|\, \hat{t}) = \delta[t - \hat{t} - T(\hat{t})] \,. \tag{7.5}$$

where $T(\hat{t})$ is the inter-spike interval of a neuron that has fired its last spike at time $\hat{t}$. If we insert Eq. (7.5) in the integral equation of the population activity, $A(t) = \int_{-\infty}^{t} P_I(t|\hat{t})\, A(\hat{t})\, \mathrm{d}\hat{t}$, we find

$$A(t) = \int_{-\infty}^{t} \delta(t - \hat{t} - T(\hat{t}))\, A(\hat{t})\, \mathrm{d}\hat{t} \,. \tag{7.6}$$

The interval $T(\hat{t})$ of a noise-free neuron is given implicitly by the threshold condition

$$T(\hat{t}) = \min\{(t - \hat{t}) \,|\, u(t) = \vartheta; \dot{u} > 0,\, t > \hat{t}\} \,. \tag{7.7}$$

Definition	$\mathcal{L}^{\rm SRM}(x)$	$=$	$-\int_x^\infty d\xi \, \frac{\partial S(\xi\|0)}{\partial \Delta h(\xi - x)}$
	$\mathcal{L}^{\rm IF}(x)$	$=$	$\mathcal{L}^{\rm SRM}(x) + \int_0^x d\xi \, e^{-\xi/\tau} \, \frac{\partial S(x\|0)}{\partial \Delta h(\xi)}$
no noise	$\mathcal{L}_0^{\rm SRM}(x)$	$=$	$\delta(x)/\eta'$
	$\mathcal{L}_0^{\rm IF}(x)$	$=$	$\left[\delta(x) - \delta(x-T_0)\, e^{-T_0/\tau}\right]/u'$
escape noise	$\mathcal{L}^{\rm SRM}(x)$	$=$	$\int_x^\infty d\xi \, f'[u(\xi - x)]\, S_0(\xi)$
	$\mathcal{L}^{\rm IF}(x)$	$=$	$\mathcal{L}^{\rm SRM}(x) - S_0(x) \int_0^x d\xi \, e^{-\xi/\tau} \, f'[u(\xi)]$
reset noise	$\mathcal{L}^{\rm SRM}(x)$	$=$	$\delta(x)/\eta'$
	$\mathcal{L}^{\rm IF}(x)$	$=$	$\left[\delta(x) - \mathcal{G}_\sigma(x-T_0)\, e^{-T_0/\tau}\right]/u'$

Table 7.1: The kernel $\mathcal{L}(x)$ for integrate-and-fire and SRM_0 neurons (upper index IF and SRM, respectively) in the general case ('Definition'), without noise, as well as for escape and reset noise. $S_0(s)$ is the survivor function in the asynchronous state and $\mathcal{G}_\sigma$ a normalized Gaussian with width σ. Primes denote derivatives with respect to the argument.

Note that $T(\hat{t})$ is the interval *starting* at $\hat{t}$ and looking *forward* towards the next spike; cf. Fig. 7.1. The integration over the δ-function in Eq. (7.6) can be done, but since T in the argument of the δ-function depends upon $\hat{t}$, the evaluation of the integral needs some care.

We recall from the rules for δ functions that

$$\int_a^b \delta[f(x)]\, g(x)\, \mathrm{d}x = \frac{g(x_0)}{|f'(x_0)|} \tag{7.8}$$

if f has a single zero-crossing $f(x_0) = 0$ in the interval $a < x_0 < b$ with $f'(x_0) \neq 0$. The prime denotes the derivative. If there is no solution $f(x_0) = 0$ in the interval $[a, b]$, the integral vanishes. In our case, x plays the role of the variable $\hat{t}$ with $f(\hat{t}) = t - \hat{t} - T(\hat{t})$. Hence $f'(\hat{t}) = -1 - T'(\hat{t})$ and

$$A(t) = \frac{1}{1 + T'(\hat{t})}\, A(\hat{t})\,, \tag{7.9}$$

whenever a solution of $\hat{t} = t - T_b(t)$ exists. Here $T_b(t)$ is the *backward* interval of neurons that reach the threshold at time t. Eq. (7.9) allows an intuitive interpretation. The activity at time t is proportional to the number of neurons that have fired one period earlier. The proportionality constant is called *compression factor*. If the inter-spike intervals decrease ($T' < 0$), then neuronal firing times

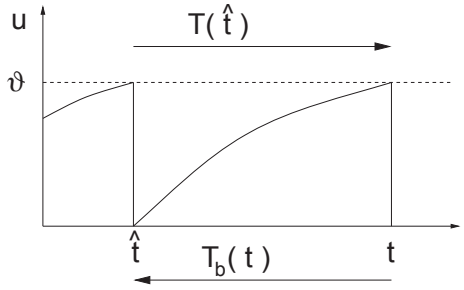

Fig. 7.1: A neuron that has fired at time $\hat{t}$ fires its next spike at $\hat{t}+T(\hat{t})$ where T is the 'forward' interval. Looking backwards we find that a neuron that fires now at time t has fired its last spike at $t - T_b(t)$ where T_b is the backward interval.

are 'compressed' and the population activity increases. If inter-spike intervals become larger ($T' > 0$), the population activity decreases; cf. Fig. 7.2.

To evaluate $T'(\hat{t})$ we use the threshold condition (7.7). From $\vartheta = u[\hat{t}+T(\hat{t})] = \eta[T(\hat{t})] + h[\hat{t}+T(\hat{t})|\hat{t}]$ we find by taking the derivative with respect to $\hat{t}$

$$0 = \eta'[T(\hat{t})]\,T'(\hat{t}) + \partial_t h[\hat{t}+T(\hat{t})|\hat{t}]\,[1+T'(\hat{t})] + \partial_{\hat{t}} h[\hat{t}+T(\hat{t})|\hat{t}]\,. \tag{7.10}$$

The prime denotes the derivative with respect to the argument. We have introduced a short-hand notation for the partial derivatives, viz., $\partial_t h(t|\hat{t}) = \partial h(t|\hat{t})/\partial t$ and $\partial_{\hat{t}} h(t|\hat{t}) = \partial h(t|\hat{t})/\partial \hat{t}$. We solve for T' and find

$$T' = -\frac{\partial_{\hat{t}} h + \partial_t h}{\eta' + \partial_t h}\,, \tag{7.11}$$

where we have suppressed the arguments for brevity. A simple algebraic transformation yields

$$\frac{1}{1+T'} = 1 + \frac{\partial_t h + \partial_{\hat{t}} h}{\eta' - \partial_{\hat{t}} h}\,, \tag{7.12}$$

which we insert into Eq. (7.9). The result is

$$A(t) = \left[1 + \frac{\partial_t h(t|\hat{t}) + \partial_{\hat{t}} h(t|\hat{t})}{\eta'(t-\hat{t}) - \partial_{\hat{t}} h(t|\hat{t})}\right] A(\hat{t})\,, \text{ with } \hat{t} = t - T_b(t)\,, \tag{7.13}$$

where $T_b(t)$ is the *backward* interval given a spike at time t. A solution $T_b(t)$ exists only if some neurons reach the threshold at time t. If this is not the case, the activity $A(t)$ vanishes. The partial derivatives in Eq. (7.13) are to be evaluated at $\hat{t} = t - T_b(t)$; the derivative $\eta' = \mathrm{d}\eta(s)/\mathrm{d}s$ is to be evaluated at $s = T_b(t)$. We may summarize Eq. (7.13) by saying that the activity at time t depends on the activity one period earlier modulated by the factor in square brackets. Note that Eq. (7.13) is still exact.

Linearization

Let us consider a fluctuating input current that generates small perturbations in the population activity $\Delta A(t)$ and the input potential $\Delta h(t)$ as outlined at the beginning of this section. If we substitute $A(t) = A_0 + \Delta A(T)$ and $h(t|\hat{t}) = h_0 + \delta h(t|\hat{t})$ into Eq. (7.13) and linearize in ΔA and Δh we obtain an expression of the form

$$\Delta A(t) = \Delta A(t-T) + A_0\, C(t)\,, \tag{7.14}$$

where $T = 1/A_0$ is the interval for constant input I_0 and C a time-dependent factor, called compression factor. The activity at time t depends thus on the activity one inter-spike interval earlier and on the instantanuous value of the compression factor.

For SRM_0 neurons we have $h(t|\hat{t}) = h(t)$ so that the partial derivative with respect to $\hat{t}$ vanishes. The factor in square brackets in Eq. (7.13) reduces therefore to $[1 + (h'/\eta')]$. If we linearize Eq. (7.13) we find the compression factor

$$C^{\mathrm{SRM}}(t) = h'(t)/\eta'(T)\,. \tag{7.15}$$

For integrate-and-fire neurons we have a similar result. To evaluate the partial derivatives that we need in Eq. (7.13) we write $u(t) = \eta(t-\hat{t}) + h(t|\hat{t})$ with

$$\begin{aligned} \eta(t-\hat{t}) &= u_r\, e^{-\frac{t-\hat{t}}{\tau_m}} \\ h(t|\hat{t}) &= h(t) - h(\hat{t})\, e^{-\frac{t-\hat{t}}{\tau_m}}\,; \end{aligned} \tag{7.16}$$

cf. Eqs. (4.34) and (4.60). Here u_r is the reset potential of the integrate-and-fire neurons and $h(t) = \int_0^\infty \exp(-s/\tau_m)\, I(t-s)\, \mathrm{d}s$ is the input potential generated by the input current I.

Taking the derivative of η and the partial derivatives of h yields

$$\frac{\partial_t h + \partial_{\hat{t}} h}{\eta' - \partial_{\hat{t}} h} = \frac{h'(t) - h'(t-T_b)\, e^{-T_b/\tau_m}}{h'(t-T_b)\, e^{-T_b/\tau_m} - \tau_m^{-1}\,[u_r + h(t-T_b)]\, e^{-T_b/\tau_m}}\,, \tag{7.17}$$

which we now insert in Eq. (7.13). Since we are interested in the *linearized* activity equation, we replace $T_b(t)$ by the interval $T = 1/A_0$ for constant input and drop the term h' in the denominator. This yields Eq. (7.14) with a compression factor C^{IF} given by

$$C^{\mathrm{IF}}(t) = [h'(t) - h'(t-T)\, \exp(-T/\tau_m)]/u'\,. \tag{7.18}$$

Here u' is the derivative of the membrane potential for *constant* input current I_0, i.e., $u' = -\tau_m^{-1}\,[u_r + h(t-T_b)]\, e^{-T_b/\tau_m}$. The label IF is short for integrate-and-fire neurons.

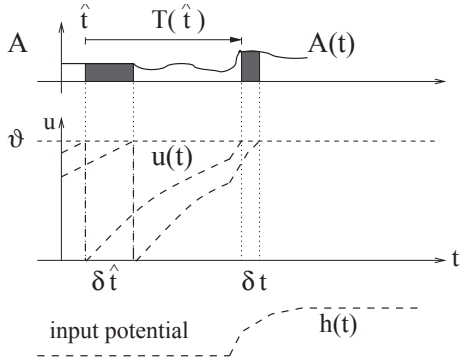

Fig. 7.2: A change in the input potential h with positive slope $h' > 0$ (dashed line, bottom) shifts neuronal firing times closer together (middle). As a result, the activity $A(t)$ (solid line, top) is higher at $t = \hat{t} + T(\hat{t})$ than it was at time $\hat{t}$ (schematic diagram); taken from (Gerstner, 2000b)

Example: Compression of firing times for SRM$_0$ neurons

In order to motivate the name 'compression factor' and to give an interpretation of Eq. (7.14) we consider SRM$_0$ neurons with an exponential refractory kernel $\eta(s) = -\eta_0 \exp(-s/\tau)$. We want to show graphically that the population activity ΔA has a contribution that is proportional to the *derivative* of the input potential.

We consider Fig. 7.2. A neuron which has fired at $\hat{t}$ will fire again at $t = \hat{t} + T(\hat{t})$. Another neuron which has fired slightly later at $\hat{t} + \delta\hat{t}$ fires its next spike at $t + \delta t$. If the input potential is constant between t and $t + \delta t$, then $\delta t = \delta\hat{t}$. If, however, h increases between t and $t + \delta t$ as is the case in Fig. 7.2, then the firing time difference is reduced. The compression of firing time differences is directly related to an increase in the activity A. To see this, we note that all neurons which fire between $\hat{t}$ and $\hat{t} + \delta\hat{t}$, must fire again between t and $t + \delta t$. This is due to the fact that the network is homogeneous and the mapping $\hat{t} \to t = \hat{t} + T(\hat{t})$ is monotonous. If firing time differences are compressed, the population activity increases.

In order to establish the relation between Fig. 7.2 and Eq. (7.15), we note that the compression faction is equal to h'/η'. For a SRM$_0$ neuron with exponential refractory kernel, $\eta'(s) > 0$ holds for all $s > 0$. An input with $h' > 0$ implies then, because of Eq. (7.14), an increase of the activity:

$$h' > 0 \Longrightarrow A(t) > A(t - T)\,. \tag{7.19}$$

7.1.2 Escape noise (*)

In this section we focus on a population of neurons with escape noise. The aim of this section is two-fold. First, we want to show how to derive the linearized population equation (7.3) that has already been stated at the beginning of Section 7.1.

Second, we will show that in the case of high noise the population activity follows the input potential $h(t)$, whereas for low noise the activity follows the derivative $h'(t)$. These results will be used in the following three sections for a discussion of signal transmission and coding properties.

In order to derive the linearized response ΔA of the population activity to a change in the input we start from the conservation law,

$$1 = \int_{-\infty}^{t} S_I(t \,|\, \hat{t})\, A(\hat{t})\, \mathrm{d}\hat{t}\,, \tag{7.20}$$

cf. (6.73). As we have seen in Chapter 6.3 the population equation (6.75) can be obtained by taking the derivative of Eq. (7.20) with respect to t, i.e.,

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{t} S_I(t \,|\, \hat{t})\, A(\hat{t})\, \mathrm{d}\hat{t}\,. \tag{7.21}$$

For constant input I_0, the population activity has a constant value A_0. We consider a small perturbation of the stationary state, $A(t) = A_0 + \Delta A(t)$, that is caused by a small change in the input current, $\Delta I(t)$. The time-dependent input generates a total postsynaptic potential,

$$h(t|\hat{t}) = h_0(t|\hat{t}) + \Delta h(t|\hat{t})\,, \tag{7.22}$$

where $h_0(t|\hat{t})$ is the postsynaptic potential for constant input I_0 and

$$\Delta h(t|\hat{t}) = \int_0^{\infty} \kappa(t-\hat{t}, s)\, \Delta I(t-s)\, \mathrm{d}s \tag{7.23}$$

is the change of the postsynaptic potential generated by ΔI. We expand Eq. (7.21) to linear order in ΔA and Δh and find

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{t} & S_0(t-\hat{t})\, \Delta A(\hat{t})\, \mathrm{d}\hat{t} \\ & + A_0 \frac{\mathrm{d}}{\mathrm{d}t} \left\{ \int_{-\infty}^{t} \mathrm{d}s \int_{-\infty}^{t} \mathrm{d}\hat{t}\, \Delta h(s|\hat{t})\, \left. \frac{\partial S_I(t \,|\, \hat{t})}{\partial \Delta h(s|\hat{t})} \right|_{\Delta h=0} \right\}. \end{aligned} \tag{7.24}$$

We have used the notation $S_0(t-\hat{t}) = S_{I_0}(t \,|\, \hat{t})$ for the survivor function of the asynchronous firing state. To take the derivative of the first term in Eq. (7.24) we use $\mathrm{d}S_0(s)/\mathrm{d}s = -P_0(s)$ and $S_0(0) = 1$. This yields

$$\begin{aligned} \Delta A(t) = \int_{-\infty}^{t} & P_0(t-\hat{t})\, \Delta A(\hat{t})\, \mathrm{d}\hat{t} \\ & - A_0 \frac{\mathrm{d}}{\mathrm{d}t} \left\{ \int_{-\infty}^{t} \mathrm{d}s \int_{-\infty}^{t} \mathrm{d}\hat{t}\, \Delta h(s|\hat{t})\, \left. \frac{\partial S_I(t \,|\, \hat{t})}{\partial \Delta h(s|\hat{t})} \right|_{\Delta h=0} \right\}. \end{aligned} \tag{7.25}$$

asynchronous firing [cf. Eqs. (6.27) and (6.28)] and $p_1(u,t)$ is a small time-dependent perturbation. The stability analysis requires a linearization of the Fokker-Planck equation (6.21) with respect to p_1 and A_1.

For short transmission delays, the asynchronous state $A(t) \equiv A_0$ can loose its stability towards an oscillation with a frequency that is much faster than the single-neuron firing rate. Brunel (2000) distinguishes two different variants of such fast oscillations. First, as in the previous example there are cluster states where the neuronal population splits into a few subgroups. Each neuron fires nearly regularly and within a cluster neurons are almost fully synchronized; cf. Section 8.2.3. Second, there are synchronous irregular states where the global activity oscillates while individual neurons have a broad distribution of interspike intervals; cf. Fig. 8.3. We will come back to synchronous irregular states in Section 8.3.

8.2 Synchronized Oscillations and Locking

We have seen in the previous section that the state of asynchronous firing can loose stability towards certain oscillatory modes that are solutions of the *linearized* population equations. We are now going to investigate oscillatory modes in more detail and check whether a large-amplitude oscillation where all neurons are firing in "lockstep" can be a stable solution of the population equations.

8.2.1 Locking in Noise-Free Populations

We consider a homogeneous population of SRM_0 or integrate-and-fire neurons which is nearly perfectly synchronized and fires almost regularly with period T. In order to analyze the existence and stability of a fully locked synchronous oscillation we approximate the population activity by a sequence of square pulses k, $k \in \{0, \pm 1, \pm 2, \dots\}$, centered around $t = kT$. Each pulse k has a certain half-width δ_k and amplitude $(2\delta_k)^{-1}$ – since all neurons are supposed to fire once in each pulse. In order to check whether the fully synchronized state is a stable solution of the population equation (6.75), we assume that the population has already fired a couple of narrow pulses for $t < 0$ with widths $\delta_k \ll T$, $k \leq 0$, and calculate the amplitude and width of subsequent pulses. If we find that the amplitude of subsequent pulses increases while their width decreases (i.e., $\lim_{k\to\infty} \delta_k = 0$), then we conclude that the fully locked state is stable.

To make the above outline more explicit, we use

$$A(t) = \sum_{k=-\infty}^{\infty} \frac{1}{2\delta_k} \mathcal{H}[t - (kT + \delta_k)]\, \mathcal{H}[(kT + \delta_k) - t] \tag{8.11}$$

as a parameterization of the population activity; cf. Fig. 8.4. Here, $\mathcal{H}(.)$ denotes

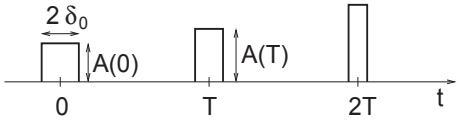

Fig. 8.4: Sequence of rectangular activity pulses. If the fully synchronized state is stable, the width δ of the pulses decreases while the amplitude increases.

the Heaviside step function with $\mathcal{H}(s) = 1$ for $s > 0$ and $\mathcal{H}(s) = 0$ for $s \leq 0$. For stability, we need to show that the amplitude $A(0), A(T), A(2T), \ldots$ of the rectangular pulses increases while the width δ_k of subsequent pulses decreases.

As we will see below, the condition for stable locking of all neurons in the population can be stated as a condition on the *slope* of the input potential h at the moment of firing. More precisely, if the last population pulse occurred at about $t = 0$ with amplitude $A(0)$ the amplitude of the population pulse at $t = T$ increases, if $h'(T) > 0$:

$$h'(T) > 0 \quad \Longleftrightarrow \quad A(T) > A(0)\,. \tag{8.12}$$

If the amplitude of subsequent pulses increases, their width decreases. In other words, we have the following *Locking Theorem.* In a spatially homogeneous network of SRM_0 or integrate-and-fire neurons, a necessary and, in the limit of a large number of presynaptic neurons ($N \to \infty$), also sufficient condition for a coherent oscillation to be asymptotically stable is that firing occurs when the postsynaptic potential arising from all previous spikes is increasing in time (Gerstner et al., 1996b).

The Locking Theorem is applicable for large populations that are already close to the fully synchronized state. A related but *global* locking argument has been presented by (Mirollo and Strogatz, 1990). The locking argument can be generalized to heterogeneous networks (Gerstner et al., 1993a; Chow, 1998) and to electrical coupling (Chow and Kopell, 2000). Synchronization in small networks has been discussed in, e.g., (Ernst et al., 1995; Chow, 1998; Hansel et al., 1995; van Vreeswijk et al., 1994; van Vreeswijk, 1996; Bose et al., 2000). For weak coupling, synchronization and locking can be systematically analyzed in the framework of phase models (Kuramoto, 1975; Ermentrout and Kopell, 1984; Kopell, 1986) or canonical neuron models (Hansel et al., 1995; Ermentrout, 1996; Hoppensteadt and Izhikevich, 1997; Izhikevich, 1999; Ermentrout et al., 2001).

Before we derive the locking condition for spiking neuron models, we illustrate the main idea by two examples.

Example: Perfect synchrony in noiseless SRM_0 neurons

In this example we will show that locking in a population of spiking neurons can be understood by simple geometrical arguments; there is no need to use the

abstract mathematical framework of the population equations. It will turn out that the results are – of course – consistent with those derived from the population equation.

We study a homogeneous network of N identical neurons which are mutually coupled with strength $w_{ij} = J_0/N$ where $J_0 > 0$ is a positive constant. In other words, the (excitatory) interaction is scaled with one over N so that the total input to a neuron i is of order one even if the number of neurons is large ($N \to \infty$). Since we are interested in synchrony we suppose that all neurons have fired simultaneously at $\hat{t} = 0$. When will the neurons fire again?

Since all neurons are identical we expect that the next firing time will also be synchronous. Let us calculate the period T between one synchronous pulse and the next. We start from the firing condition of SRM_0 neurons

$$\vartheta = u_i(t) = \eta(t - \hat{t}_i) + \sum_j w_{ij} \sum_f \epsilon(t - t_j^{(f)}) \,, \tag{8.13}$$

where $\epsilon(t)$ is the postsynaptic potential. The axonal transmission delay Δ^{ax} is included in the definition of ϵ, i.e., $\epsilon(t) = 0$ for $t < \Delta^{\text{ax}}$. Since all neurons have fired synchronously at $t = 0$, we set $\hat{t}_i = t_j^{(f)} = 0$. The result is a condition of the form

$$\vartheta - \eta(t) = J_0 \, \epsilon(t) \,, \tag{8.14}$$

since $w_{ij} = J_0/N$ for $j = 1, \ldots, N$. Note that we have neglected the postsynaptic potentials that may have been caused by earlier spikes $t_j^{(f)} < 0$ back in the past. The graphical solution of Eq. (8.14) is presented in Fig. 8.5. The first crossing point of the $\vartheta - \eta(t)$ and $J_0 \, \epsilon(t)$ defines the time T of the next synchronous pulse.

What happens if synchrony at $t = 0$ was not perfect? Let us assume that one of the neurons is slightly late compared to the others; Fig. 8.5B. It will receive the input $J_0 \, \epsilon(t)$ from the others, thus the right-hand side of (8.14) is the same. The left-hand side, however, is different since the last firing was at δ_0 instead of zero. The next firing time is at $t = T + \delta_1$ where δ_1 is found from

$$\vartheta - \eta(T + \delta_1 - \delta_0) = J_0 \, \epsilon(T + \delta_1) \,. \tag{8.15}$$

Linearization with respect to δ_0 and δ_1 yields:

$$\delta_1 < \delta_0 \quad \Longleftrightarrow \quad \epsilon'(T) > 0 \,. \tag{8.16}$$

Thus the neuron which has been late is 'pulled back' into the synchronized pulse of the others, if the postsynaptic potential ϵ is rising at the moment of firing at T. Equation (8.16) is a special case of the Locking Theorem.

We see from Fig. 8.5B that, in the case of excitatory coupling, stable locking works nicely if the transmission delay Δ^{ax} is in the range of the firing period, but slightly shorter so that firing occurs during the rise time of the EPSP.

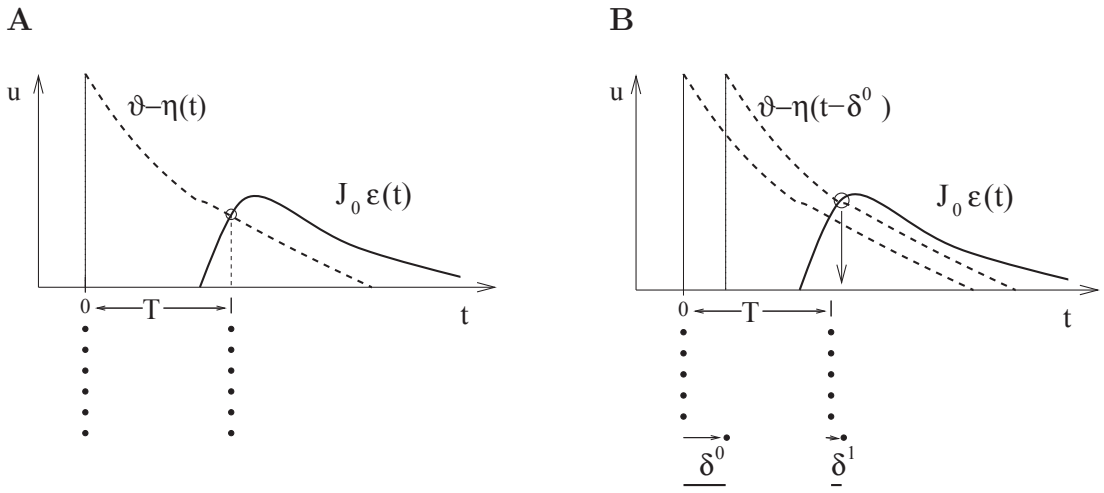

Fig. 8.5: **A**. Perfect Synchrony. All neurons have fired at $\hat{t} = 0$. The next spike occurs when the summed postsynaptic potential $J_0\,\epsilon(t)$ reaches the dynamic threshold $\vartheta - \eta(t)$. **B**. Stability of perfect synchrony. The last neuron is out of tune. The firing time difference at $t = 0$ is δ_0. One period later the firing time difference is reduced ($\delta_1 < \delta_0$), since the threshold is reached at a point where $J_0\epsilon(t)$ is rising. Adapted from Gerstner et al. (1996b).

Example: $\mathbf{SRM_0}$ neurons with inhibitory coupling

Locking can also occur in networks with purely inhibitory couplings (van Vreeswijk et al., 1994). In order to get a response at all in such a system, we need a constant stimulus I_0 or, equivalently, a negative firing threshold $\vartheta < 0$. The stability criterion, however, is equivalent to that of the previous example.

Figure 8.6 summarizes the stability arguments analogously to Fig. 8.5. In Fig. 8.6A all neurons have fired synchronously at $t = 0$ and do so again at $t = T$ when the inhibitory postsynaptic potential has decayed so that the threshold condition,

$$\vartheta - \eta(T) = J_0 \sum_k \epsilon(t - k\,T)\,, \tag{8.17}$$

is fulfilled. This state is stable if the synaptic contribution to the potential, $\sum_k \epsilon(t - k\,T)$, has positive slope at $t = T$. Figure 8.6 demonstrates that a single neuron firing at $t = \delta_0$ instead of $t = 0$ is triggered again at $t = T + \delta_1$ with $|\delta_1| < |\delta_0|$ for simple geometrical reasons.

Derivation of the locking theorem (*)

We consider a homogeneous populations of SRM neurons that are close to a periodic state of synchronized activity. We assume that the population activity in the past consists of a sequence of rectangular pulses as specified in Eq. (8.11).

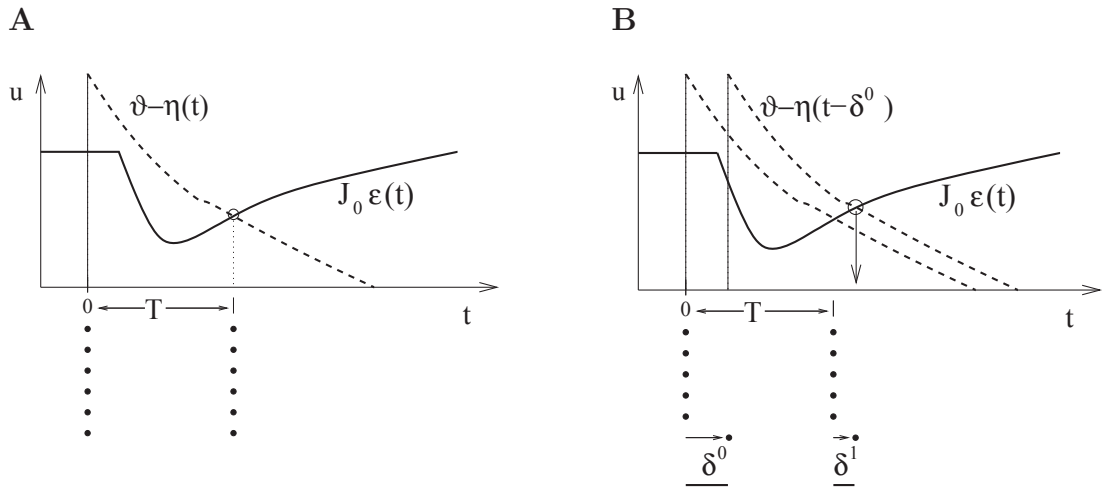

Fig. 8.6: Similar plot as in Fig. 8.5 but for purely inhibitory coupling. **A**. All neurons have fired synchronously at $\hat{t} = 0$. The next spike occurs when the summed inhibitory postsynaptic potential $J_0\,\epsilon(t)$ has decayed back to the dynamic threshold $\vartheta - \eta(t)$. **B**. Stability of perfect synchrony. The last neuron is out of tune. The firing time difference at $t = 0$ is δ_0. One period later the firing time difference is reduced ($\delta_1 < \delta_0$), since the threshold is reached at a point where $J_0\,\epsilon(t)$ is rising.

We determine the period T and the sequence of half-widths δ_k of the rectangular pulses in a self-consistent manner. In order to prove stability, we need to show that the amplitude $A(k\,T)$ increases while the halfwidth δ_k decreases as a function of k. To do so we start from the noise-free population equation (7.13) that we recall here for convenience

$$A(t) = \left[1 + \frac{\partial_t h + \partial_{\hat{t}} h}{\eta' - \partial_{\hat{t}} h}\right] A(t - T_b(t)) \tag{8.18}$$

where $\partial_t h$ and $\partial_{\hat{t}} h$ are the partial derivatives of the total postsynaptic potential $h_{\rm PSP}$ and $T_b(t)$ is the backward interval; cf. Fig. 7.1.

As a first step, we calculate the potential $h_{\rm PSP}(t|\hat{t})$. Given $h_{\rm PSP}$ we can find the period T from the threshold condition and also the derivatives $\partial_t h$ and $\partial_{\hat{t}} h$ required for Eq. (7.13). In order to obtain $h_{\rm PSP}$, we substitute Eq. (8.11) in (6.8), assume $\delta_k \ll T$, and integrate. To first order in δ_k we obtain

$$h_{\rm PSP}(t|\hat{t}) = \sum_{k=0}^{k_{\rm max}} J_0\,\epsilon(t - \hat{t}, t + k\,T) \;+\; \mathcal{O}\left[(\delta_k)^2\right] , \tag{8.19}$$

where $-\delta_0 \leq \hat{t} \leq \delta_0$ is the last firing time of the neuron under consideration. The sum runs over all pulses back in the past. Since $\epsilon(t - \hat{t}, s)$ as a function of s is

Chapter 9

Spatially Structured Networks

So far the discussion of network behavior in Chapters 6 – 8 was restricted to *homogeneous* populations of neurons. In this chapter we turn to networks that have a spatial structure. In doing so we emphasize two characteristic features of the cerebral cortex, namely the high density of neurons and its virtually two-dimensional architecture.

Each cubic millimeter of cortical tissue contains about 10^5 neurons. This impressive number suggests that a description of neuronal dynamics in terms of an averaged *population activity* is more appropriate than a description on the single-neuron level. Furthermore, the cerebral cortex is huge. More precisely, the unfolded cerebral cortex of humans covers a surface of 2200–2400 cm^2, but its thickness amounts on average to only 2.5–3.0 mm^2. If we do not look too closely, the cerebral cortex can hence be treated as a continuous two-dimensional sheet of neurons. Neurons will no longer be labeled by discrete indices but by continuous variables that give their spatial position on the sheet. The coupling of two neurons i and j is replaced by the *average* coupling strength between neurons at position x and those at position y, or, even more radically simplified, by the average coupling strength of two neurons being separated by the distance $|x - y|$. Similarly to the notion of an average coupling strength we will also introduce the *average activity* of neurons located at position x and describe the dynamics of the network in terms of these averaged quantities only. The details of how these average quantities are defined, are fairly involved and often disputable. In Sect. 9.1 we will – without a formal justification – introduce field equations for the spatial activity $A(x, t)$ in a spatially extended, but otherwise homogeneous population of neurons. These field equations are particularly interesting because they have solutions in the form of complex stationary patterns of activity, traveling waves, and rotating spirals – a phenomenology that is closely related to pattern formation in certain nonlinear systems that are collectively termed *excitable media*. Some examples of these solutions are discussed in Sect. 9.1. In Sect. 9.2 we generalize the formalism so as to account for several distinct neuronal populations, such as those formed

by excitatory and inhibitory neurons. The rest of this chapter is dedicated to models that describe neuronal activity in terms of individual action potentials. The propagation of spikes through a locally connected network of SRM neurons is considered in Section 9.3. The last section, finally, deals with the transmission of a sharp pulse packet of action potentials in a layered feed-forward structure. It turns out that there is a stable wave form of the packet so that temporal information can be faithfully transmitted through several brain areas despite the presence of noise.

9.1 Stationary patterns of neuronal activity

We start with a generic example of pattern formation in a neural network with 'Mexican-hat' shaped lateral coupling, i.e., local excitation and long-range inhibition. In order to keep the notation as simple as possible, we will use the field equation derived in Chapter 6; cf. Eq. (6.129). As we have seen in Fig. 6.8, this equation neglects rapid transients and oscillations that could be captured by the full integral equations. On the other hand, in the limit of high noise and short refractoriness the approximation of population dynamics by differential equations is good; cf. Chapter 7. Exact solutions in the low-noise limit will be discussed in Section 9.3.

Consider a single sheet of densely packed neurons. We assume that all neurons are alike and that the connectivity is homogeneous and isotropic, i.e., that the coupling strength of two neurons is a function of their distance only. We loosely define a quantity $u(x,t)$ as the average membrane potential of the group of neurons located at position x at time t. We have seen in Chapter 6 that in the stationary state the 'activity' of a population of neurons is strictly given by the single-neuron gain function $A_0(x) = g[u_0(x)]$; cf. Fig. 9.1. If we *assume* that changes of the input potential are slow enough so that the population always remains in a state of incoherent firing, then we can set

$$A(x,t) = g[u(x,t)]\,, \tag{9.1}$$

even for time-dependent situations. According to Eq. (9.1), the activity $A(x,t)$ of the population around location x is a function of the potential at that location.

The synaptic input current to a given neuron depends on the level of activity of its presynaptic neurons and on the strength of the synaptic couplings. We assume that the amplitude of the input current is simply the presynaptic activity scaled by the average coupling strength of these neurons. The total input current $I^{\rm syn}(x,t)$ to a neuron at position x is therefore

$$I^{\rm syn}(x,t) = \int \mathrm{d}y\; w\left(|x-y|\right)\; A(y,t)\,. \tag{9.2}$$

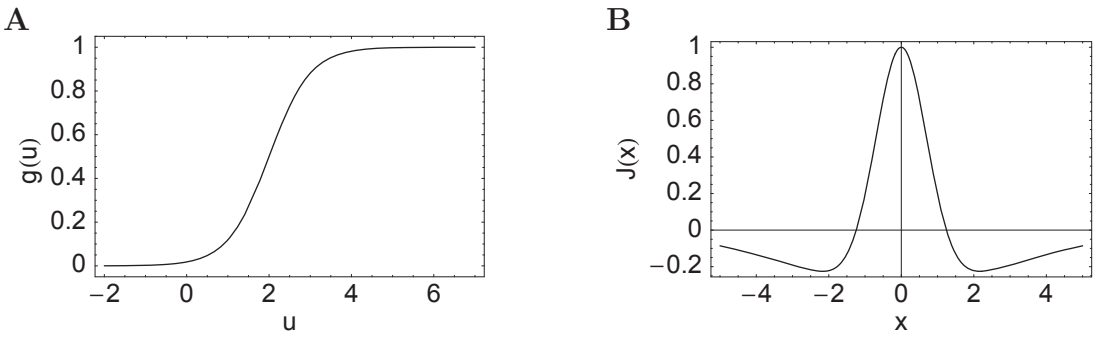

Fig. 9.1: **A**. Generic form of the sigmoidal gain function g of graded response neurons that describes the relation between the potential u and the 'activity' of the neural population. **B**. Typical 'Mexican hat'-shaped function that is used here as an ansatz for the coupling w of two neurons as a function of their distance x.

Here, w is the average coupling strength of two neurons as a function of their distance. We consider a connectivity pattern, that is excitatory for proximal neurons and predominantly inhibitory for distal neurons. Figure 9.1B shows the typical 'Mexican-hat shape' of the corresponding coupling function. Eq. (9.2) assumes that synaptic interaction is instantaneous. In a more detailed model we could include the axonal transmission delay and synaptic time constants. In that case, $A(y,t)$ on the right-hand side of Eq. (9.2) should be replaced by $\int \alpha(s)\, A(y, t-s)\, \mathrm{d}s$ where $\alpha(s)$ is the temporal interaction kernel.

In order to complete the definition of the model, we need to specify a relation between the input current and the resulting membrane potential. In order to keep things simple we treat each neuron as a leaky integrator. The input potential is thus given by a differential equation of the form

$$\tau \frac{\partial u}{\partial t} = -u + I^{\mathrm{syn}} + I^{\mathrm{ext}}\,, \tag{9.3}$$

with τ being the time constant of the integrator and I^{ext} an additional external input. If we put things together we obtain the field equation

$$\tau \frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int \mathrm{d}y\; w\left(|x-y|\right)\; g[u(y,t)] + I^{\mathrm{ext}}(x,t)\,; \tag{9.4}$$

cf. Wilson and Cowan (1973); Feldman and Cowan (1975); Amari (1977b); Kishimoto and Amari (1979). This is a nonlinear integro-differential equation for the average membrane potential $u(x,t)$.

9.1.1 Homogeneous solutions

Although we have kept the above model as simple as possible, the field equation (9.4) is complicated enough to prevent comprehensive analytical treatment. We

therefore start our investigation by looking for a special type of solution, i.e., a solution that is uniform over space, but not necessarily constant over time. We call this the homogenous solution and write $u(x,t) \equiv u(t)$. We expect that a homogenous solution exists if the external input is homogeneous as well, i.e., if $I^{\text{ext}}(x,t) \equiv I^{\text{ext}}(t)$.

Substitution of the ansatz $u(x,t) \equiv u(t)$ into Eq. (9.4) yields

$$\tau \frac{\mathrm{d}u(t)}{\mathrm{d}t} = -u(t) + \bar{w}\, g[u(t)] + I^{\text{ext}}(t)\,. \tag{9.5}$$

with $\bar{w} = \int \mathrm{d}y\; w\,(|y|)$. This is a nonlinear ordinary differential equation for the average membrane potential $u(t)$. We note that the equation for the homogeneous solution is identical to that of a single population without spatial structure; cf. Eq. (6.87) in Chapter 6.3.

The fixed points of the above equation with $I^{\text{ext}} = 0$ are of particular interest because they correspond to a resting state of the network. More generally, we search for stationary solutions for a given constant external input $I^{\text{ext}}(x,t) \equiv I^{\text{ext}}$. The fixed points of Eq. (9.5) are solutions of

$$g(u) = \frac{u - I^{\text{ext}}}{\bar{w}}\,, \tag{9.6}$$

which is represented graphically in Fig. 9.2. Depending on the strength of the external input three qualitatively different situations can be observed. For low external stimulation there is a single fixed point at a very low level of neuronal activity. This corresponds to a quiescent state where the activity of the whole network has ceased. Large stimulation results in a fixed point at an almost saturated level of activity which corresponds to a state where all neurons are firing at their maximum rate. Intermediate values of external stimulation, however, may result in a situation with more than one fixed point. Depending on the shape of the output function and the mean synaptic coupling strength $\bar{w}$ three fixed points may appear. Two of them correspond to the quiescent and the highly activated state, respectively, which are separated by the third fixed point at an intermediate level of activity.

Any potential physical relevance of fixed points clearly depends on their stability. Stability under the dynamics defined by the ordinary differential equation Eq. (9.5) is readily checked using standard analysis. Stability requires that at the intersection

$$g'(u) < \bar{w}^{-1}\,. \tag{9.7}$$

Thus all fixed points corresponding to quiescent or highly activated states are stable whereas the middle fixed point in case of multiple solutions is unstable; cf. Fig. 9.2. This, however, is only half of the truth because Eq. (9.5) only describes homogeneous solutions. Therefore, it may well be that the solutions are stable

Part III

Models of Synaptic Plasticity

Chapter 10

Hebbian Models

In the neuron models discussed so far each synapse is characterized by a single constant parameter w_{ij} that determines the amplitude of the postsynaptic response to an incoming action potential. Electrophysiological experiments, however, show that the response amplitude is not fixed but can change over time. Appropriate stimulation paradigms can systematically induce changes of the postsynaptic response that last for hours or days. If the stimulation paradigm leads to a persistent *increase* of the synaptic transmission efficacy, the effect is called long-term potentiation of synapses, or LTP for short. If the result is a *decrease* of the synaptic efficacy, it is called long-term depression (LTD). These persistent changes are thought to be the neuronal correlate of 'learning' and 'memory'.

In the formal theory of neural networks the weight w_{ij} of a connection from neuron j to i is considered as a parameter that can be adjusted so as to optimize the performance of a network for a given task. The process of parameter adaptation is called *learning* and the procedure for adjusting the weights is referred to as a *learning rule*. Here learning is meant in its widest sense. It may refer to synaptic changes during development just as well as to the specific changes necessary to memorize a visual pattern or to learn a motor task. There are many different learning rules that we cannot all cover in this book. In this chapter we consider the simplest set of rules, viz., synaptic changes that are driven by correlated activity of pre- and postsynaptic neurons. This class of learning rule can be motivated by Hebb's principle and is therefore often called 'Hebbian learning'.

10.1 Synaptic Plasticity

Over the last 30 years, a large body of experimental results on synaptic plasticity has been accumulated. Many of these experiments are inspired by Hebb's postulate (Hebb, 1949) that describes how the connection from presynaptic neuron A to a postsynaptic neuron B should be modified:

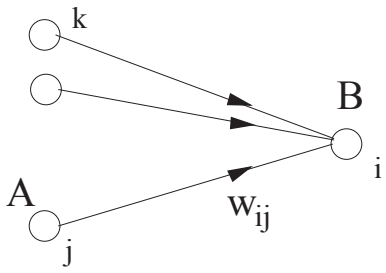

Fig. 10.1: The change at synapse w_{ij} depends on the state of the presynaptic neuron j and the postsynaptic neuron i and the present efficacy w_{ij}, but not on the state of other neurons k.

> When an axon of cell A is near enough to excite cell B or repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

Today, 50 years later, this famous postulate is often rephrased in the sense that modifications in the synaptic transmission efficacy are driven by correlations in the firing activity of pre- and postsynaptic neurons. Even though the idea of learning through correlations dates further back in the past (James, 1890), correlation-based learning is now generally called *Hebbian learning*.

Hebb formulated his principle on purely theoretical grounds. He realized that such a mechanism would help to stabilize specific neuronal activity patterns in the brain. If neuronal activity patterns correspond to behavior, then stabilization of specific patterns implies learning of specific types of behaviors (Hebb, 1949).

10.1.1 Long-Term Potentiation

When Hebb stated his principle in 1949, it was a mere postulate. More than 20 years later, long-lasting changes of synaptic efficacies were found experimentally (Bliss and Lomo, 1973; Bliss and Gardner-Medwin, 1973). These changes can be induced by the joint activity of presynaptic and postsynaptic neuron and resemble the mechanism that Hebb had in mind (Kelso et al., 1986). In this subsection we concentrate on long-term potentiation (LTP), viz., a persistent increase of synaptic efficacies. Long-term depression (LTD) is mentioned in passing.

The basic paradigm of LTP induction is, very schematically, the following (Brown et al., 1989; Bliss and Collingridge, 1993). A neuron is impaled by an intracellular electrode to record the membrane potential while presynaptic fibers are stimulated by means of a second extracellular electrode. Small pulses are applied to the presynaptic fibers in order measure the strength of the postsynaptic response (Fig. 10.2A). The amplitude of the test pulse is chosen so that the stimulation evokes a postsynaptic potential, but no action potentials.

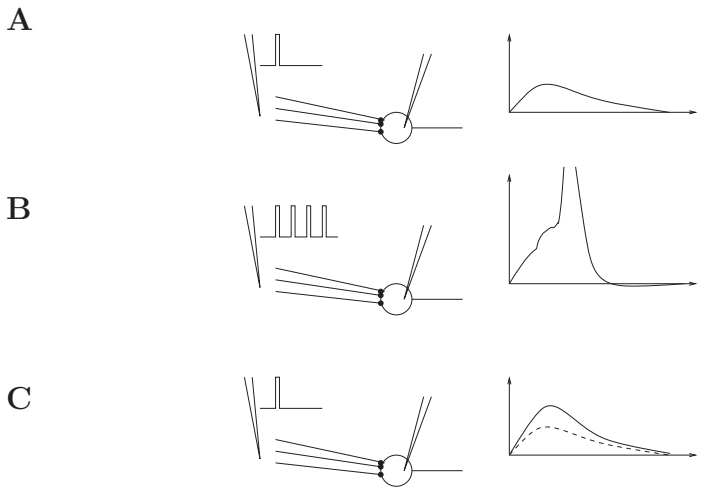

Fig. 10.2: Schematic drawing of a paradigm of LTP induction. **A**. A weak test pulse (left) evokes the postsynaptic response sketched on the right-hand side of the figure. **B**. A strong stimulation sequence (left) triggers postsynaptic firing (right, the peak of the action potential is out of bounds). **C**. A test pulse applied some time later evokes a larger postsynaptic response (right; solid line) than the initial response. The dashed line is a copy of the initial response in **A** (schematic figure).

In a second step, the input fibers are strongly stimulated by a sequence of high frequency pulses so as to evoke postsynaptic firing (Fig. 10.2B). After that the strength of the postsynaptic response to small pulses is tested again and a significantly increased amplitude of postsynaptic potentials is found (Fig. 10.2C). This change in the synaptic strength persists over many hours and is thus called *long-term potentiation*.

What can be learned from such an experiment? Obviously, the result is consistent with Hebb's postulate because the joint activity of pre- and postsynaptic units has apparently lead to a strengthening of the synaptic efficacy. On the other hand, the above experiment would also be consistent with a purely postsynaptic explanation that claims that the strengthening is solely caused by postsynaptic spike activity. In order to exclude this possibility, a more complicated experiment has to be conducted (Brown et al., 1989; Bliss and Collingridge, 1993).

In an experiment as it is sketched in Fig. 10.3 a neuron is driven by two separate input pathways labeled S (strong) and W (weak), respectively. Each pathway projects to several synapses on the postsynaptic neuron i. Stimulating of the S pathway excites postsynaptic firing but stimulation of the W channel alone does not evoke spikes. The response to the W input is monitored in order to detect changes of the synaptic efficacy. A 100 Hz input over 600 ms at the W channel evokes no LTP at the W synapses. Similarly, a 100 Hz input (over 400 ms)

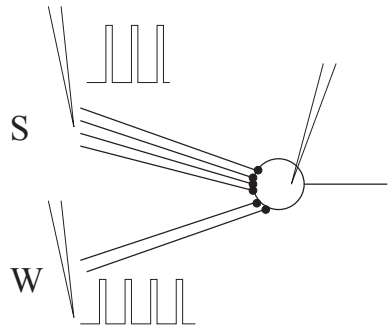

Fig. 10.3: Cooperativity in the induction of LTP. Synapses at the W channel are strengthened only if both the presynaptic site is stimulated via the W electrode and the postsynaptic neuron is active due to a simultaneous stimulation of the S pathway.

at the S channel alone does not produce LTP at the W synapses (although it may evoke a change of the S synapses). However, if both stimulations occur simultaneously, then the W synapses are strengthened. This feature of LTP induction is known as cooperativity or associativity. It is consistent with the picture that both presynaptic *and* postsynaptic activity is required to induce LTP.

Experiments as the one sketched in Figs. 10.2 and 10.3 have shown that synaptic weights change as a function of pre- and postsynaptic activity. Many other paradigms of LTP induction have been studied over the last twenty years. For example, the state of the postsynaptic neuron can be manipulated by depolarizing or hyperpolarizing currents; synaptic channels can be blocked or activated pharmacologically, etc. Nevertheless, the underlying subcellular processes that lead to LTP are still not fully understood.

10.1.2 Temporal Aspects

The essential aspect of the experiments described in the previous section is the AND condition that is at the heart of Hebb's postulate: Both pre- and postsynaptic neuron have to be active in order to induce a strengthening of the synapse. However, such a summary neglects the temporal requirements for weight changes. When are two neurons considered as being active together?

In the experiment sketched in Fig. 10.3 inputs can be switched on and off with some arbitrary timing. A large increase of the synaptic efficacy can be induced by stimulating the W and the S pathway simultaneously. If there is a certain delay in the stimulation of W and S then the synaptic efficacy is only slightly increased or even reduced. Stimulating W and S alternatively with a long interval in between does not result in any change at all (Levy and Stewart, 1983; Gustafsson et al., 1987; Debanne et al., 1994). With this setup, however, a precise mea-

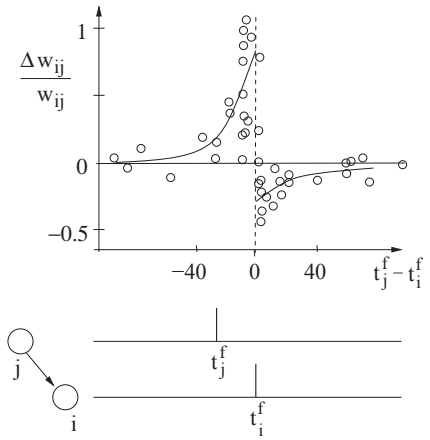

Fig. 10.4: Timing requirements between pre- and postsynaptic spikes. Synaptic changes Δw_{ij} occur only if presynaptic firing at $t_j^{(f)}$ and postsynaptic activity at $t_i^{(f)}$ occur sufficiently close to each other. Experimentally measured weight changes (circles) as a function of $t_j^{(f)} - t_i^{(f)}$ in milliseconds overlayed on a schematic two-phase learning window (solid line). A positive change (LTP) occurs if the presynaptic spike *precedes* the postsynaptic one; for a reversed timing, synaptic weights are decreased. Data points redrawn after the experiments of Bi and Poo (1998).

surement of the timing conditions for synaptic changes is difficult, because pre- and postsynaptic activity is generated by extracellular electrodes. With modern patch-clamp techniques it is possible to stimulate or record from one or several neurons intracellularly. Pairing experiments with multiple intracellular electrodes in synaptically coupled neurons have opened the possibility to study synaptic plasticity at an excellent spatial and temporal resolution (Markram et al., 1997; Zhang et al., 1998; Magee and Johnston, 1997; Debanne et al., 1998; Bi and Poo, 1998, 1999); see Bi and Poo (2001) for a review.

Figure 10.4 illustrates a pairing experiment with cultured hippocampal neurons where the presynaptic neuron (j) and the postsynaptic neuron (i) are forced to fire spikes at time $t_j^{(f)}$ and $t_i^{(f)}$, respectively (Bi and Poo, 1998). The resulting change in the synaptic efficacy Δw_{ij} after several repetitions of the experiment turns out to be a function of the spike time differences $t_j^{(f)} - t_i^{(f)}$ ('spike-time dependent synaptic plasticity'). Most notably, the direction of the change depends critically, i.e., on a millisecond time-scale, on the relative timing of pre- and postsynaptic spikes. The synapse is strengthened if the presynaptic spike occurs shortly before the postsynaptic neuron fires, but the synapse is weakened if the sequence of spikes is reversed; cf. Fig. 10.4A. This observation is indeed in agreement with Hebb's postulate because presynaptic neurons that are active slightly *before* the postsynaptic neuron are those which 'take part in firing it'

whereas those that fire later obviously did not contribute to the postsynaptic action potential. An asymmetric learning window such as the one in Fig. 10.4, is thus an implementation of the causality requirement that is implicit in Hebb's principle.

Similar results on spike-time dependent synaptic plasticity have been found in various neuronal systems (Markram et al., 1997; Zhang et al., 1998; Debanne et al., 1998; Bi and Poo, 1998, 1999; Egger et al., 1999), but there are also characteristic differences. Synapses between parallel fibers and 'Purkinje-cells' in the cerebellar-like structure of electric fish, for example, show the opposite dependence on the relative timing of presynaptic input and the (so-called 'broad') postsynaptic spike. In this case the synapse is weakened if the presynaptic input arrives shortly before the postsynaptic spike (anti-Hebbian plasticity). If the timing is the other way round then the synapse is strengthened. A change in the timing of less than 10 ms can change the effect from depression to potentiation (Bell et al., 1997b).

10.2 Rate-Based Hebbian Learning

In order to prepare the ground for a thorough analysis of spike-based learning rules in Section 10.3 we will first review the basic concepts of correlation-based learning in a firing rate formalism.

10.2.1 A Mathematical Formulation of Hebb's Rule

In order to find a mathematically formulated learning rule based on Hebb's postulate we focus on a single synapse with efficacy w_{ij} that transmits signals from a presynaptic neuron j to a postsynaptic neuron i. For the time being we content ourselves with a description in terms of mean firing rates. In the following, the activity of the presynaptic neuron is denoted by ν_j and that of the postsynaptic neuron by ν_i.

There are two aspects in Hebb's postulate that are particularly important, viz. *locality* and *cooperativity*. Locality means that the change of the synaptic efficacy can only depend on local variables, i.e., on information that is available at the site of the synapse, such as pre- and postsynaptic firing rate, and the actual value of the synaptic efficacy, but not on the activity of other neurons. Based on the locality of Hebbian plasticity we can make a rather general ansatz for the change of the synaptic efficacy,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = F(w_{ij}; \nu_i, \nu_j)\,. \tag{10.1}$$

Here, $\mathrm{d}w_{ij}/\mathrm{d}t$ is the rate of change of the synaptic coupling strength and F is a so far undetermined function (Brown et al., 1991; Sejnowski and Tesauro, 1989;

Kohonen, 1984). We may wonder whether there are other local variables (e.g., the membrane potential u_i) that should be included as additional arguments of the function F. It turns out that in standard rate models this is not necessary, since the membrane potential u_i is uniquely determined by the postsynaptic firing rate, $\nu_i = g(u_i)$, with a monotone gain function g.

The second important aspect of Hebb's postulate, cooperativity, implies that pre- and postsynaptic neuron have to be active *simultaneously* for a synaptic weight change to occur. We can use this property to learn something about the function F. If F is sufficiently well-behaved, we can expand F in a Taylor series about $\nu_i = \nu_j = 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = c_0(w_{ij}) + c_1^{\text{post}}(w_{ij})\nu_i + c_1^{\text{pre}}(w_{ij})\,\nu_j + c_2^{\text{pre}}(w_{ij})\,\nu_j^2 + c_2^{\text{post}}(w_{ij})\,\nu_i^2 + c_2^{\text{corr}}(w_{ij})\,\nu_i\,\nu_j + \mathcal{O}(\nu^3)\,. \tag{10.2}$$

The term containing c_2^{corr} on the right-hand side of (10.2) is bilinear in pre- and postsynaptic activity. This term implements the AND condition for cooperativity which makes Hebbian learning a useful concept.

The simplest choice for our function F is to fix c_2^{corr} at a positive constant and to set all other terms in the Taylor expansion to zero. The result is the prototype of Hebbian learning,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = c_2^{\text{corr}}\,\nu_i\,\nu_j\,. \tag{10.3}$$

We note in passing that a learning rule with $c_2^{\text{corr}} < 0$ is usually called anti-Hebbian because it weakens the synapse if pre- and postsynaptic neuron are active simultaneously; a behavior that is just contrary to that postulated by Hebb. A learning rule with only first-order terms gives rise to so-called non-Hebbian plasticity, because pre- or postsynaptic activity alone induces a change of the synaptic efficacy. More complicated learning rules can be constructed if higher-order terms in the expansion of Eq. (10.2), such as $\nu_i\,\nu_j^2$, $\nu_i^2\,\nu_j$, $\nu_i^2\,\nu_j^2$, etc., are included.

The dependence of F on the synaptic efficacy w_{ij} is a natural consequence of the fact that w_{ij} is bounded. If F was independent of w_{ij} then the synaptic efficacy would grow without limit if the same potentiating stimulus is applied over and over again. A saturation of synaptic weights can be achieved, for example, if the parameter c_2^{corr} in Eq. (10.2) tends to zero as w_{ij} approaches its maximum value, say $w^{\max} = 1$, e.g.,

$$c_2^{\text{corr}}(w_{ij}) = \gamma_2\,(1 - w_{ij}) \tag{10.4}$$

with a positive constant γ_2.

Hebb's original proposal does not contain a rule for a decrease of synaptic weights. In a system where synapses can only be strengthened, all efficacies

will finally saturate at their upper maximum value. An option of decreasing the weights (synaptic depression) is therefore a necessary requirement for any useful learning rule. This can, for example, be achieved by weight decay, which can be implemented in Eq. (10.2) by setting

$$c_0(w_{ij}) = -\gamma_0 \, w_{ij} \, . \tag{10.5}$$

Here, γ_0 is (small) positive constant that describes the rate by which w_{ij} decays back to zero in the absence of stimulation. Our formulation (10.2) is hence sufficiently general to allow for a combination of synaptic potentiation and depression. If we combine Eq. (10.4) and Eq. (10.5) we obtain the learning rule

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = \gamma_2 \, (1 - w_{ij}) \, \nu_i \, \nu_j - \gamma_0 \, w_{ij} \, . \tag{10.6}$$

The factors $(1 - w_{ij})$ and w_{ij} that lead to a saturation at $w_{ij} = 1$ for continued stimulation and an exponential decay to $w_{ij} = 0$ in the absence of stimulation, respectively, are one possibility to implement 'soft' bounds for the synaptic weight. In simulations, 'hard' bounds are often used to restrict the synaptic weights to a finite interval, i.e., a learning rule with weight-independent parameters is only applied as long as the weight stays within its limits.

Another interesting aspect of learning rules is *competition*. The idea is that synaptic weights can only grow at the expense of others so that if a certain subgroup of synapses is strengthened, other synapses to the same postsynaptic neuron have to be weakened. Competition is essential for any form of self-organization and pattern formation. Practically, competition can be implemented in simulations by normalizing the sum of all weights converging onto the same postsynaptic neuron (Miller and MacKay, 1994); cf. Section 11.1.3. Though this can be motivated by a limitation of common synaptic resources such a learning rule violates locality of synaptic plasticity. On the other hand, competition of synaptic weight changes can also be achieved with purely local learning rules (Oja, 1982; Kistler and van Hemmen, 2000a; Song et al., 2000; Kempter et al., 2001).

Example: Postsynaptic gating *versus* presynaptic gating

Equation (10.6) is just one possibility to specify rules for the growth and decay of synaptic weights. In the framework of Eq. (10.2), other formulations are conceivable; cf. Table 10.1. For example, we can define a learning rule of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = \gamma \, \nu_i \, [\nu_j - \nu_\theta(w_{ij})] \, , \tag{10.7}$$

where γ is a positive constant and ν_θ is some reference value that may depend on the current value of w_{ij}. A weight change occurs only if the postsynaptic neuron is active, $\nu_i > 0$. We say that weight changes are 'gated' by the postsynaptic neuron.

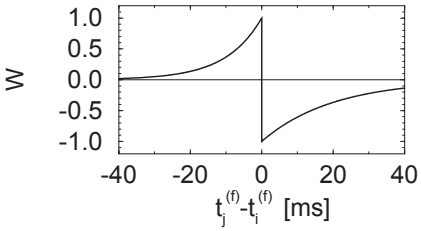

Fig. 10.7: Two-phase learning window W as a function of the time difference $s = t_j^{(f)} - t_i^{(f)}$ between presynaptic spike arrival and postsynaptic firing; cf. Eq. (10.16) with $A_+ = -A_- = 1$, $\tau_1 = 10\,\text{ms}$, and $\tau_2 = 20\,\text{ms}$ (Zhang et al., 1998).

Example: Exponential learning windows

A simple choice for the learning window – and thus for the kernels $a_2^{\text{post,pre}}$ and $a_2^{\text{pre,post}}$ – inspired by Fig. 10.4 is

$$W(s) = \begin{cases} A_+ \exp[s/\tau_1] & \text{for } s < 0\,, \\ A_- \exp[-s/\tau_2] & \text{for } s > 0\,, \end{cases} \tag{10.16}$$

with some constants $A_\pm$ and $\tau_{1,2}$. If $A_+ > 0$ and $A_- < 0$ then the synaptic efficacy is increased if presynaptic spike arrives slightly *before* the postsynaptic firing ($W(s) > 0$ for $s < 0$) and the synapse is weakened if presynaptic spikes arrive a few milliseconds *after* the output spike ($W(s) < 0$); cf. Fig. 10.7.

In order to obtain a realistic description of synaptic plasticity we have to make sure that the synaptic efficacy stays within certain bounds. Excitatory synapses, for example, should have a positive weight and must not exceed a maximum value of, say, $w_{ij} = 1$. We can implement these constraints in Eq. (10.16) by setting $A_- = w_{ij}\, a_-$ and $A_+ = (1 - w_{ij})\, a_+$. The remaining terms in Eq. (10.14) can be treated analogously[1]. For each positive term (leading to a weight increase) we assume a weight dependence $\propto (1 - w_{ij})$, while for each negative term (leading to weight decrease) we assume a weight dependence $\propto w_{ij}$. The synapse is thus no longer strengthened (or weakened) if the weight reaches its upper (lower) bound (Kistler and van Hemmen, 2000a; van Rossum et al., 2000).

10.3.2 Consolidation of Synaptic Efficacies

So far we have emphasized that the synaptic coupling strength is a dynamical variable that is subject to rapid changes dependent on pre- and postsynaptic ac-

[1]Note that w_{ij} is a step function of time with discontinuities whenever a presynaptic spike arrives or a postsynaptic action potential is triggered. In order to obtain a well-defined differential equation we specify that the amplitude of the step depends on the value of w_{ij} immediately *before* the spike. In mathematical terms, we impose the condition that $w_{ij}(t)$ is continuous from left, i.e., that $\lim_{s\to 0, s>0} w_{ij}(t^{(f)} - s) = w_{ij}(t^{(f)})$.

tivity. On the other hand, it is generally accepted that long-lasting modifications of the synaptic efficacy form the basis of learning and memory. How can fast dynamical properties be reconciled with long-lasting modifications?

Most learning theories dealing with artificial neural networks concentrate on the *induction* of weight changes. As soon as the 'learning session' is over, synaptic plasticity is 'switched off' and weights are taken as fixed parameters. In biological systems, a similar mechanism can be observed during development. There are critical periods in the early life time of an animal where certain synapses show a form of plasticity that is 'switched off' after maturation (Crepel, 1982). The majority of synapses, especially those involved in higher brain functions, however, keep their plasticity throughout their life. At first glance there is thus the risk that previously stored information is simply overwritten by new input ('palimpsest property'). Grossberg has coined the term 'stability-plasticity dilemma' for this problem (Grossberg, 1987; Carpenter and Grossberg, 1987).

To address these questions, Fusi et al. (2000) have studied the problem of the consolidation of synaptic weights. They argue that *bistability* of synaptic weights can solve the stability-plasticity dilemma. More specifically, the dynamics of synaptic efficacies is characterized by two stable fixed points at $w_{ij} = 0$ and $w_{ij} = 1$. In the absence of stimulation the synaptic weight will thus settle down at either one of these values. Pre- or postsynaptic activity can lead to a transition from one fixed point to the other, but only if the duration of the stimulus presentation or its amplitude exceeds a certain threshold. In other words, synapses can be switched on or off but this will happen only for those synapses where the learning threshold is exceeded. Learning thus affects only a few synapses so that previously stored information is mostly preserved.

In the framework of Eq. (10.14), such a dynamics for synaptic weights can be implemented by setting

$$a_0(w_{ij}) = -\gamma\, w_{ij}\,(1 - w_{ij})\,(w_\theta - w_{ij})\,, \tag{10.17}$$

where $0 < w_\theta < 1$ and $\gamma > 0$. In the absence of stimulation, small weights ($w_{ij} < w_\theta$) decay to zero whereas large weights ($w_{ij} > w_\theta$) increase towards one. Spike activity thus has to drive the synaptic weight across the threshold w_θ before long-lasting changes take place. A combination of Eqs. (10.17) and (10.14) can therefore be considered as a model of induction and consolidation of synaptic plasticity.

10.3.3 General Framework (*)

In Eq. (10.14) weight changes occur instantaneously at the moment of presynaptic spike arrival or postsynaptic firing. In this subsection we will develop a slightly more general equation for the evolution of synaptic weights. The approach taken

in this section can be seen as a generalization of the Taylor expansion in the rate model of Section 10.2 to the case of spiking neurons.

We recall that we started our formulation of rate-based Hebbian learning from a general formula

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij} = F(w_{ij}; \nu_i, \nu_j) \tag{10.18}$$

where weight changes are given as a function of the weight w_{ij} as well as of the pre- and postsynaptic rates ν_j, ν_i; cf. Eq. (10.1). The essential assumption was that neuronal activity is characterized by firing rates that change slowly enough to be considered as stationary. Hebbian rules followed then from a Taylor expansion of Eq. (10.18). In the following, we keep the idea of an expansion, but drop the simplifications that are inherent to a description in terms of a mean firing rate.

The internal state of spiking neurons (e.g., integrate-and-fire or Spike Response Model neurons) is characterized by the membrane potential u which in turn depends on the input and the last output spike. The generalization of Eq. (10.18) to the case of spiking neurons is therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = F\left(w_{ij}(t); \{u_i^{\mathrm{post}}(t' < t)\}, \{u_j^{\mathrm{pre}}(t'' < t)\}\right) , \tag{10.19}$$

where F is now a *functional* of the time course of pre- and postsynaptic membrane potential *at the location of the synapse*. Our notation with t' and t'' is intended to indicate that the weight changes do not only depend on the momentary value of the pre- and postsynaptic potential, but also on their history $t' < t$ and $t'' < t$. The weight value w_{ij} and the local value of pre- and postsynaptic membrane potential are the essential variables that are available at the site of the synapse to control the up- and down-regulation of synaptic weights. In detailed neuron models, F would depend not only on the weight w_{ij} and membrane potentials, but also on all other variables that are locally available at the site of the synapse. In particular, there could be a dependence upon the local calcium concentration; cf. Section 10.4.

In analogy to the approach taken in Section 10.2, we now expand the right-hand side of Eq. (10.19) about the resting state $u_i^{\mathrm{post}} = u_j^{\mathrm{pre}} = u_{\mathrm{rest}}$ in a Volterra series (Volterra, 1959; Palm and Poggio, 1977; van Hemmen et al., 2000). For the sake of simplicity we shift the voltage scale so that $u_{\mathrm{rest}} = 0$. We find

$$\begin{aligned} \frac{\mathrm{d}w_{ij}}{\mathrm{d}t} &= a_0(w_{ij}) + \int_0^\infty \alpha_1^{\mathrm{pre}}(w_{ij}; s)\, u_j^{\mathrm{pre}}(t-s)\, \mathrm{d}s \\ &\quad + \int_0^\infty \alpha_1^{\mathrm{post}}(w_{ij}; s')\, u_i^{\mathrm{post}}(t-s')\, \mathrm{d}s' \\ &\quad + \int_0^\infty \int_0^\infty \alpha_2^{\mathrm{corr}}(w_{ij}; s, s')\, u_j^{\mathrm{pre}}(t-s)\, u_i^{\mathrm{post}}(t-s')\, \mathrm{d}s'\, \mathrm{d}s + \dots . \end{aligned} \tag{10.20}$$

The next terms would be quadratic in u_i^{post} or u_j^{pre} and have been neglected. Equation (10.20) provides a framework for the formulation of spike-based learning

rules and may be seen as the generalization of the general rate-based model that we have derived in Section 10.2.

In order to establish a connection with various other formulations of spike-based learning rules, we consider the time course of the pre- and postsynaptic membrane potential in more detail. At the presynaptic terminal, the membrane potential is most of the time at rest, except when an action potential arrives. Since the duration of each action potential is short, the presynaptic membrane potential can be approximated by a train of δ functions

$$u_j^{\rm pre}(t) = \sum_f \delta(t - t_j^{(f)}) \tag{10.21}$$

where $t_j^{(f)}$ denotes the spike arrival times at the presynaptic terminal.

The situation at the postsynaptic site is somewhat more complicated. For the simple spike response model SRM_0, the membrane potential can be written as

$$u_i^{\rm post}(t) = \eta(t - \hat{t}_i) + h_i(t)\,, \tag{10.22}$$

where $\hat{t}_i$ is the last postsynaptic firing time. In contrast to the usual interpretation of terms on the right-hand side of Eq. (10.22), the function η is now taken as the time course of the *back propagating* action potential at the location of the *synapse*. Similarly, $h_i(t)$ is the local postsynaptic potential at the synapse.

For a further simplification of Eq. (10.20), we need to make some approximations. Specifically we will explore two different approximation schemes. In the first scheme, we suppose that the dominating term on the right-hand side of Eq. (10.22) is the back propagating action potential, while in the second scheme we neglect η and consider h as the dominant term. Let us discuss both approximations in turn.

(i) Sharply peaked back propagating action potential

We assume that the back propagating action potential is sharply peaked, i.e., it has a large amplitude and short duration. In this case, the membrane potential of the postsynaptic neuron is dominated by the back propagating action potential and the term $h(t)$ in Eq. (10.22) can be neglected. Furthermore η can be approximated by a δ function. The membrane potential at the postsynaptic site reduces then to a train of pulses,

$$u_i^{\rm post}(t) = \sum_f \delta(t - t_i^{(f)})\,, \tag{10.23}$$

where $t_i^{(f)}$ denotes the postsynaptic firing times. Equation (10.23) is a sensible approximation for synapses that are located on or close to the soma so that the full somatic action potential is 'felt' by the postsynaptic neuron. For neurons with

active processes in the dendrite that keep the back propagating action potential well focused, Eq. (10.23) is also a reasonable approximation for synapses that are further away from the soma. A transmission delay for back propagation of the spike from the soma to the site of the synapse can be incorporated at no extra cost.

If we insert Eqs. (10.21) and (10.23) into Eq. (10.20) we find

$$\frac{\mathrm{d}w_{ij}}{\mathrm{dt}} = a_0 + \sum_f \alpha_1^{\rm pre}(t - t_j^{(f)}) + \sum_f \alpha_1^{\rm post}(t - t_i^{(f)}) + \sum_f \sum_f \alpha_2^{\rm corr}(t - t_i^{(f)}, t - t_j^{(f)}) + \dots , \quad (10.24)$$

where we have omitted the w_{ij} dependence of the right-hand side terms. In contrast to Eq. (10.14) weight changes are now continuous. A single presynaptic spike at time $t_j^{(f)}$, for example, will cause a weight change that builds up during some time after $t_j^{(f)}$. An example will be given below in Eq. (10.31).

In typical plasticity experiments, the synaptic weight is monitored every few hundred milliseconds so that the exact time course of the functions, $\alpha_1^{\rm pre}$, $\alpha_1^{\rm post}$ and $\alpha_2^{\rm corr}$ is not measured. To establish the connection to Eq. (10.14), we now assume that the weight changes are rapid compared to the time scale of weight monitoring. In other words, we make the replacement

$$\alpha_1^{\rm pre}(t - t_j^{(f)}) \quad \longrightarrow \quad a_1^{\rm pre}\, \delta(t - t_j^{(f)}) \quad (10.25)$$

$$\alpha_1^{\rm post}(t - t_i^{(f)}) \quad \longrightarrow \quad a_1^{\rm post}\, \delta(t - t_i^{(f)}) \quad (10.26)$$

where $a_1^{\rm pre} = \int_0^\infty \alpha_1^{\rm pre}(s)\, \mathrm{d}s$ and $a_1^{\rm post} = \int_0^\infty \alpha_1^{\rm post}(s)\, \mathrm{d}s$. For the correlation term we exploit the invariance with respect to time translation, i.e., the final result only depends on the time difference $t_j^{(f)} - t_i^{(f)}$. The weight update occurs at the moment of the postsynaptic spike if $t_j^{(f)} < t_i^{(f)}$ and at the moment of the presynaptic spike if $t_j^{(f)} > t_i^{(f)}$. Hence, the assumption of instantaneous update yields two terms

$$\alpha_2^{\rm corr}(t - t_i^{(f)}, t - t_j^{(f)}) \longrightarrow \begin{cases} a_2^{\rm pre,post}(t_j^{(f)} - t_i^{(f)})\, \delta(t - t_j^{(f)}) & \text{if } t_i^{(f)} < t_j^{(f)} \\ a_2^{\rm post,pre}(t_i^{(f)} - t_j^{(f)})\, \delta(t - t_i^{(f)}) & \text{if } t_i^{(f)} > t_j^{(f)} \end{cases} \quad (10.27)$$

Thus, for sharply peaked back propagating action potentials and rapid weight changes, the general framework of Eq. (10.20) leads us back to the Eq. (10.14).

(ii) No back propagating action potential

In the second approximation scheme, we assume that the membrane potential at the location of the synapse is dominated by the slowly varying potential $h_i(t)$.

This is, for example, a valid assumption in voltage-clamp experiments where the postsynaptic neuron is artificially kept at a constant membrane potential $h^{\rm post}$. This is also a good approximation for synapses that are located far away from the soma on a *passive* dendrite, so that the back propagation of somatic action potentials is negligible.

Let us consider a voltage clamp experiment where $h_i(t)$ is kept at a constant level $h^{\rm post}$. As before, we suppose that weight changes are rapid. If we insert $u_j^{\rm pre}(t) = \sum_f \delta(t - t_j^{(f)})$ and $u_i^{\rm post}(t) = h^{\rm post}$ into Eq. (10.20), we find

$$\begin{aligned} \frac{{\rm d}w_{ij}}{{\rm d}t} &= a_0 + \sum_f a_1^{\rm pre}\, \delta(t - t_j^{(f)}) \\ &\quad + a_1^{\rm post}\, h^{\rm post} + a_2^{\rm corr}\, h^{\rm post} \sum_f \delta(t - t_j^{(f)}) + \dots \end{aligned} \tag{10.28}$$

where $a_1^{\rm pre} = \int_0^\infty \alpha_1^{\rm pre}(s)\, {\rm d}s$, $a_1^{\rm post} = \int_0^\infty \alpha_1^{\rm post}(s)\, {\rm d}s$ and $a_2^{\rm corr} = \int_0^\infty \int_0^\infty a_2^{\rm corr}(s, s')\, {\rm d}s\, {\rm d}s'$. Equation (10.28) is the starting point of the theory of spike-based learning of Fusi et al. (2000). Weight changes are triggered by presynaptic spikes. The direction and value of the weight update depends on the postsynaptic membrane potential. In our framework, Eq. (10.28) is a special case of the slightly more general Eq. (10.20).

10.4 Detailed Models of Synaptic Plasticity

In the previous section we have introduced a purely phenomenological model for spike-time dependent synaptic plasticity which is at least qualitatively in agreement with experimental results. In this section we take a slightly different approach and discuss how the core idea of this model, the learning window, arises from elementary kinetic processes. We start in Section 10.4.1 with a simple mechanistic model and turn then, in Section 10.4.2 to a more detailed model with saturation. A calcium-based model is the topic of Section 10.4.3. All three models give a qualitative explanation for the learning dynamics on the level of individual spikes.

10.4.1 A Simple Mechanistic Model

The AND condition in Hebb's postulate suggests that two biochemical components are involved in the induction of LTP. We do not wish to speculate on the nature of these components, but simply call them a and b. We assume that the first component is generated by a chemical reaction chain triggered by presynaptic spike arrival. In the absence of further input, the concentration $[a]$ decays with a time constant τ_a back to its resting level $[a] = 0$. A simple way to describe

this process is

$$\frac{\mathrm{d}}{\mathrm{d}t}[a] = -\frac{[a]}{\tau_a} + d_a \sum_f \delta(t - t_j^{(f)}) \,, \tag{10.29}$$

where the sum runs over all *presynaptic* firing times $t_j^{(f)}$. Equation (10.29) states that $[a]$ is increased at each arrival of a presynaptic spike by an amount d_a. A high level of $[a]$ sets the synapse in a state where it is susceptible to changes in its weight. The variable $[a]$ by itself, however, does not yet trigger a weight change.

To generate the synaptic change, another substance b is needed. The production of b is controlled by a second process triggered by *postsynaptic* spikes,

$$\frac{\mathrm{d}}{\mathrm{d}t}[b] = -\frac{[b]}{\tau_b} + d_b \sum_f \delta(t - t_i^{(f)}) \,, \tag{10.30}$$

where τ_b is another time constant. The sum runs over all postsynaptic spikes $t_i^{(f)}$. Note that the second variable $[b]$ does not need to be a biochemical quantity; it could, for example, be the electrical potential caused by the postsynaptic spike itself.

Hebbian learning needs both 'substances' to be present at the same time, thus

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}^{\mathrm{corr}} = \gamma \, [a(t)] \, [b(t)] \,, \tag{10.31}$$

with some rate constant γ. The upper index *corr* is intended to remind us that we are dealing only with the correlation term on the right-hand side of Eq. (10.14) or Eq. (10.24).

Let us now consider the synaptic change caused by a single presynaptic spike at $t_j^{(f)} \geq 0$ and a postsynaptic spike a $t_i^{(f)} = t_j^{(f)} - s$. Integration of Eqs. (10.29) and (10.30) yields

$$\begin{aligned} [a(t)] &= d_a \, \exp[-(t - t_j^{(f)})/\tau_a] \, \Theta(t - t_j^{(f)}) \\ [b(t)] &= d_b \, \exp[-(t - t_i^{(f)})/\tau_b] \, \Theta(t - t_i^{(f)}) \,, \end{aligned} \tag{10.32}$$

where $\Theta(.)$ denotes the Heaviside step function as usual. The change caused by the pair of pulses $(t_i^{(f)}, t_j^{(f)})$, measured after a time T, is

$$\int_0^T \left(\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}^{\mathrm{corr}} \right) \mathrm{d}t = \gamma \, d_a \, d_b \int_{\max\{t_j^{(f)}, t_i^{(f)}\}}^T \exp\left[-\frac{t - t_j^{(f)}}{\tau_a} - \frac{t - t_i^{(f)}}{\tau_b} \right] \mathrm{d}t \,. \tag{10.33}$$

The integral over t can be calculated explicitely. The total weight change that is obtained for $T \gg \tau_a, \tau_b$ can be identified with the learning window. Thus we find

$$W(s) = \begin{cases} A \, \exp[s/\tau_a] & \text{for} \quad s < 0 \\ A \, \exp[-s/\tau_b] & \text{for} \quad s > 0 \end{cases} \tag{10.34}$$

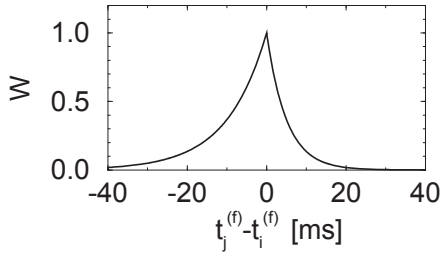

Fig. 10.8: Exponential learning window W as a function of the time difference $s = t_j^{(f)} - t_i^{(f)}$ between presynaptic spike arrival and postsynaptic firing. The time constants for exponential decay are $\tau_1 = 20\,\text{ms}$ for $s < 0$ and $\tau_2 = 10\,\text{ms}$ for $s > 0$.

with $s = t_j^{(f)} - t_i^{(f)}$ and $A = \gamma\, d_a\, d_b\, \tau_a \tau_b/(\tau_a + \tau_b)$. As expected, the change of the synaptic efficacy depends only on the time difference between pre- and postsynaptic spike (Gerstner et al., 1998); cf. Fig. 10.8.

Equation (10.34) describes the change caused by a single pair of spikes. Given a train of presynaptic input spikes and a set of postsynaptic output spikes, many combinations of firing times $(t_i^{(f)}, t_j^{(f)})$ exist. Due to the linearity of the learning equation (10.31), the total change is additive, which is consistent with Eq. (10.14).

The combination of two kinetic processes a and b thus yields an exponential learning window as in Eq. (10.16) but with $A_+ = A_-$. The learning window either describes LTP ($\gamma > 0$) or LTD ($\gamma < 0$), but not both. If we want to have an anti-symmetric learning window with LTP and LTD we need additional processes as detailed below.

Example: LTP and LTD

For a learning window incorporating both LTP and LTD, we need more microscopic variables. Let us suppose that, as before, we have variables $[a]$ and $[b]$ that contribute to LTP according to (10.31), viz.,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}^{\mathrm{LTP}} = \gamma^{\mathrm{LTP}}\, [a(t)]\, [b(t)]\,. \tag{10.35}$$

Similarly, we assume that there is a second set of variables $[c]$ and $[d]$, that initiate LTD according to

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}^{\mathrm{LTD}} = -\gamma^{\mathrm{LTD}}\, [c(t)]\, [d(t)]\,. \tag{10.36}$$

The variables $[c]$ and $[d]$ have a dynamics analogous to Eq. (10.29) and Eq. (10.30) with amplitudes d_c and d_d, and time constants τ_c and τ_d. The total weight change is the sum of both contributions,

$$w_{ij}^{\mathrm{corr}} = w_{ij}^{\mathrm{LTP}} + w_{ij}^{\mathrm{LTD}}\,, \tag{10.37}$$

and so is the learning window, i.e.,

$$W(s) = \begin{cases} A_+ \exp[s/\tau_a] - A_- \exp[s/\tau_c] & \text{for} \quad s < 0 \\ A_+ \exp[-s/\tau_b] - A_- \exp[-s/\tau_d] & \text{for} \quad s > 0 \end{cases} \tag{10.38}$$

with $A_+ = \gamma^{\rm LTP}\, d_a\, d_b\, \tau_a \tau_b/(\tau_a + \tau_b)$ and $A_- = \gamma^{\rm LTD}\, d_c\, d_d\, \tau_c \tau_d/(\tau_c + \tau_d)$ (Gerstner et al., 1996a, 1998).

We now set $d_b = 1/\tau_b$ and $d_c = 1/\tau_c$. In the limit of $\tau_b \to 0$ and $\tau_c \to 0$, we find the asymmetric two-phase learning window introduced in Eq. (10.16). Weight changes are now instantaneous. A postsynaptic spike that is triggered after a presynaptic spike arrival reads out the current value of $[a]$ and induces LTP by an amount

$$W(t_j^{(f)} - t_i^{(f)}) = \gamma^{\rm LTP}\, d_a \exp\left(-\frac{t_i^{(f)} - t_j^{(f)}}{\tau_a}\right) \qquad \text{for} \quad t_j^{(f)} < t_i^{(f)}\,. \tag{10.39}$$

A presynaptic spike $t_j^{(f)}$ that arrives *after* a postsynaptic spike reads out the current value of $[d]$ and induces LTD by an amount

$$W(t_j^{(f)} - t_i^{(f)}) = -\gamma^{\rm LTD} d_d \exp\left(-\frac{t_j^{(f)} - t_i^{(f)}}{\tau_d}\right) \qquad \text{for} \quad t_j^{(f)} > t_i^{(f)}\,. \tag{10.40}$$

10.4.2 A Kinetic Model based on NMDA Receptors

A model for LTP and LTD that is slightly more elaborate than the simplistic model discussed in the previous section has been developed by Senn et al. (1997, 2001b). This model is based on the assumption that NMDA receptors can be in one of three different states, a resting state, an 'up' and a 'down' state. Transitions between these states are triggered by presynaptic spike arrival ('rest'→ 'up') and postsynaptic firing ('rest'→ 'down'). The actual induction of LTP or LTD, however, requires another step, namely the activation of so-called second-messengers. The model assumes two types of second-messenger, one for LTP and one for LTD. If a presynaptic spike arrives before a postsynaptic spike, the up-regulation of the NMDA receptors in combination with the activitation $S_1 \to S^{\rm up}$ of the first second-messenger triggers synaptic changes that lead to LTP. On the other hand, if the presynaptic spike arrives after postsynaptic firing, the NMDA receptors are downregulated and the activation $S_2 \to S^{\rm dn}$ of the other second-messenger triggers LTD; cf. Fig. 10.9

The variables $N^{\rm up}$, $N^{\rm dn}$, and $N^{\rm rest}$ describe the portion of NMDA receptors that are in one of the three possible states ($N^{\rm up} + N^{\rm dn} + N^{\rm rest} = 1$). In the absence of pre- and postsynaptic spikes, all receptors return to the rest state,

$$\frac{\rm d}{{\rm d}t} N^{\rm rest} = \frac{N^{\rm up}}{\tau_{N^{\rm up}}} + \frac{N^{\rm dn}}{\tau_{N^{\rm dn}}}\,. \tag{10.41}$$

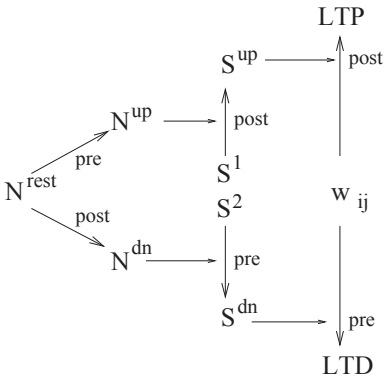

Fig. 10.9: Upper part of panel: A presynaptic spike shifts NMDA receptors from the rest state $N^{\rm rest}$ to the up-regulated state $N^{\rm up}$. If a postsynaptic spike arrives shortly afterwards, a second messenger S^1 will be activated ($S^{\rm up}$). Depending on the amount of activated second messengers $S^{\rm up}$, postsynaptic spikes lead to LTP. Lower part: Postsynaptic spikes down-regulate NMDA receptors ($N^{\rm dn}$). In the presence of $N^{\rm dn}$, presynaptic spikes activate another second messenger ($S^{\rm dn}$) leading to LTD.

$N^{\rm up}$ and $N^{\rm dn}$ decay with time constants $\tau_{N^{\rm up}}$ and $\tau_{N^{\rm dn}}$, respectively. Whenever a presynaptic spike arrives, NMDA receptors are up-regulated from rest to the 'up'-state according to

$$\frac{\mathrm{d}}{\mathrm{d}t}N^{\rm up}(t) = r^{\rm up}\, N^{\rm rest}(t) \sum_f \delta(t - t_j^{(f)}) - \frac{N^{\rm up}(t)}{\tau_{N^{\rm up}}}\,, \tag{10.42}$$

where $t_j^{(f)}$ is the arrival time of a presynaptic spike and $r^{\rm up}$ is the proportion of receptors in the 'rest' state that are up-regulated. Since presynaptic spike arrival triggers release of the neurotransmitter glutamate, which is then bound to the NMDA receptors, the 'up'-state can be identified with a state where the receptor is saturated with glutamate.

Firing of a postsynaptic spike at time $t_i^{(f)}$ leads to a down-regulation of NMDA receptors via

$$\frac{\mathrm{d}}{\mathrm{d}t}N^{\rm dn}(t) = r^{\rm dn}\, N^{\rm rest}(t) \sum_f \delta(t - t_i^{(f)}) - \frac{N^{\rm dn}(t)}{\tau_{N^{\rm dn}}}\,. \tag{10.43}$$

Senn et al. (2001b) suggest that down-regulation of the NMDA receptor is mediated by the intracellular calcium concentration that changes with each postsynaptic spike. Note that, since $N^{\rm rest} = 1 - N^{\rm up} - N^{\rm dn}$, Eqs. (10.42) and (10.43) account for saturation effects due to a limited number of NMDA-receptors.

The secondary messenger $S^{\rm up}$, which finally leads to LTP, is activated by postsynaptic spikes, but only if up-regulated NMDA channels are available. In

the absence of postsynaptic spikes the concentration of second messengers decays with time constant $\tau_{S^{\rm up}}$. Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}S^{\rm up}(t) = -\frac{S^{\rm up}(t)}{\tau_{S^{\rm up}}} + r_S\, N^{\rm up}(t)\,[1 - S^{\rm up}(t)] \sum_f \delta(t - t_i^{(f)})\,, \tag{10.44}$$

where r_S is a rate constant. Since $N^{\rm up}(t) > 0$ requires that a presynaptic spike has occurred before t, the activation of $S^{\rm up}$ effectively depends on the specific timing of pre- and postsynaptic spikes ('first pre, then post').

Similarly, the other second messenger $S^{\rm dn}$ is activated by a presynaptic spike provided that there are receptors in their down regulated state, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}S^{\rm dn}(t) = -\frac{S^{\rm dn}(t)}{\tau_{S^{\rm dn}}} + r_S\, N^{\rm dn}(t)\,[1 - S^{\rm dn}(t)] \sum_f \delta(t - t_j^{(f)})\,, \tag{10.45}$$

where $\tau_{S^{\rm dn}}$ is another decay time constant. The second messenger $S^{\rm dn}$ is therefore triggered by the sequence 'first post, then pre'. The factors $[1-S^{\rm up}]$ in Eq. (10.44) and $[1 - S^{\rm dn}]$ in Eq. (10.45) account for the limited amount of second messengers available at the synapse.

Long-term potentiation (weight increase) depends on the presence of $S^{\rm up}$, long-term depression (weight decrease) on $S^{\rm dn}$. This is described by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}w_{ij} \;&=\; \gamma_{\rm LTP}\,(1 - w_{ij})\,[S^{\rm up} - \theta^{\rm up}]_+ \sum_f \delta(t - t_i^{(f)} - \Delta) \\ &\quad -\gamma_{\rm LTD}\,w_{ij}\,[S^{\rm dn} - \theta^{\rm d}]_+ \sum_f \delta(t - t_j^{(f)} - \Delta)\end{aligned} \tag{10.46}$$

with certain parameters $\gamma_{\rm LTP/D}$ and $\theta^{\rm up/dn}$. Here, $[x]_+ = x\,\Theta(x)$ denotes a piecewise linear function with $[x]_+ = x$ for $x > 0$ and zero otherwise. The delay $0 < \Delta \ll 1$ ensures that the actual weight change occurs *after* the update of $S^{\rm up}, S^{\rm dn}$. Note that this is a third-order model. The variable $S^{\rm up} > 0$, for example, is already second-order, because it depends on presynaptic spikes followed by postsynaptic action potentials. In Eq. (10.46) the postsynaptic spike is then used again in order to trigger the weight change.

Example: Low Rates

For low pre- and postsynaptic firing rates, saturation effects can be neglected and Eq. (10.46) is equivalent to the elementary model discussed in Section 10.4.1. Let us assume that a single spike induces a small change ($r^{\rm dn}, r^{\rm up} \ll 1$) so that we can use $N^{\rm rest} \approx 1$ in Eqs. (10.42) and (10.43). The equations for $N^{\rm up}$ and $N^{\rm dn}$ are then identical to those for the 'substances' $[a]$ and $[d]$ in Eqs. (10.29), (10.35), and (10.36).

Let us furthermore assume that interspike intervals are long compared to the decay time constants $\tau_{S^{\rm up}}, \tau_{S^{\rm dn}}$ in Eqs. (10.44) and (10.45). Then $S^{\rm up}$ is negligible except during and shortly after a postsynaptic action potential. At the moment of postsynaptic firing, $S^{\rm up}$ 'reads out' the current value of $N^{\rm up}$; cf. Eq. (10.44). If this value is large than $\Theta_{\rm up}$, it triggers a positive weight change; cf. Eq. (10.46). Similarly, at the moment of presynaptic spike arrival $S^{\rm dn}$ 'reads out' the value of $N^{\rm dn}$ and triggers a weight decrease. Thus, in this limit, the model of Senn et al. corresponds to an exponential time window

$$W(s) = \begin{cases} A_+(w_{ij})\, \exp[+s/\tau_{N^{\rm up}}] & \text{for} \quad s < 0 \\ A_-(w_{ij})\, \exp[-s/\tau_{N^{\rm dn}}] & \text{for} \quad s > 0 \end{cases} \tag{10.47}$$

with $A_+(w_{ij}) = r^{\rm up}\, r_S\, (1 - w_{ij})$ and $A_-(w_{ij}) = -r^{\rm dn}\, r_S\, w_{ij}$.

Example: High Rates

If we assume that all decay time constants are much longer than typical interspike intervals then the variables $N^{\rm up/dn}$ and $S^{\rm up/dn}$ will finally reach a steady state. If we neglect correlations between pre- and postsynaptic neuron by replacing spike trains by rates, we can solve for these stationary states,

$$N^{\rm up}_\infty = \frac{\tau_{N^{\rm up}}\, r^{\rm up}\, \nu_j}{1 + \tau_{N^{\rm up}}\, r^{\rm up}\, \nu_j + \tau_{N^{\rm dn}} r^{\rm dn} \nu_i} \tag{10.48}$$

$$S^{\rm up}_\infty = \frac{\tau_{S^{\rm up}}\, r_S\, N^{\rm up}\, \nu_i}{1 + \tau_{S^{\rm up}} r_S\, N^{\rm up}\, \nu_i} \tag{10.49}$$

and similar equations for $N^{\rm dn}$ and $S^{\rm dn}$. Note that $S^{\rm up}_\infty$ is a function of ν_i and ν_j. If we put the equations for $S^{\rm up}_\infty$ and $S^{\rm dn}_\infty$ in Eq. (10.46) we get an expression of the form

$$\frac{\rm d}{{\rm d}t} w_{ij} = \gamma_{\rm LTP}\, (1 - w_{ij})\, f_{\rm LTP}(\nu_i, \nu_j)\, \nu_i - \gamma_{\rm LTD}\, w_{ij}\, f_{\rm LTD}(\nu_i, \nu_j)\, \nu_j \tag{10.50}$$

with functions $f_{\rm LTP}$ and $f_{\rm LTD}$. We linearize $f_{\rm LTP}$ with respect to ν_j about a reference value $\overline{\nu} > 0$ and evaluate $f_{\rm LTD}$ at $\nu_j = \overline{\nu}$ in order to make the right-hand side of Eq. (10.50) linear in the input ν_j. The result is

$$\frac{\rm d}{{\rm d}t} w_{ij} = \phi(w_{ij}; \nu_i)\, \nu_j \tag{10.51}$$

with $\phi(w_{ij}; 0) = \phi(w_{ij}; \nu_\theta) = 0$ for some value ν_θ and $d\phi/d\nu_i < 0$ at $\nu_i = 0$. Equation (10.51) is a generalized Bienenstock-Cooper-Monroe rule where ϕ does not only depend on the postsynaptic rate ν_i but also on the individual synaptic weight; cf. Eq. (10.12). For details see Senn et al. (2001b); Bienenstock et al. (1982).

10.4.3 A Calcium-Based Model

It has been recognized for a long time that calcium ions are an important second messenger for the induction of LTP and LTD in Hippocampus (Malenka et al., 1988; Malinow et al., 1989; Lisman, 1989) and cerebellum (Konnerth and Eilers, 1994; Lev-Ram et al., 1997). Particularly well investigated are 'NMDA synapses' in the hippocampus (Collingridge et al., 1983; Dudek and Bear, 1992; Bliss and Collingridge, 1993; Bindman et al., 1991) where calcium ions can enter the cell through channels that are controlled by a glutamate receptor subtype called NMDA (N-methyl-D-aspartic acid) receptor; cf. Section 2.4.2. These channels are involved in transmission of action potentials in glutamatergic (excitatory) synapses. If an action potential arrives at the presynaptic terminal, glutamate, the most common excitatory neurotransmitter, is released into the synaptic cleft and diffuses to the postsynaptic membrane where it binds to NMDA and AMPA receptors. The binding to AMPA receptors results in the opening of the associated ion channels and hence to a depolarization of the postsynaptic membrane. Channels controlled by NMDA receptors, however, are blocked by magnesium ions and do not open unless the membrane is sufficiently depolarized so as to remove the block. Therefore, calcium ions can enter the cell only if glutamate has been released by presynaptic activity *and* if the postsynaptic membrane is sufficiently depolarized. The calcium influx is the first step in a complex bio-chemical pathway that leads ultimately to a modification of the glutamate-sensitivity of the postsynaptic membrane.

Biophysical models of Hebbian plasticity (Lisman, 1989; Zador et al., 1990; Holmes and Levy, 1990; Gold and Bear, 1994; Schiegg et al., 1995; Shouval et al., 2001) contain two essential components, viz. a description of intracellular calcium dynamics, in particular a model of calcium entry through NMDA synapses; and a hypothesis of how the concentration of intracellular calcium influences the change of synaptic efficacy. In this section we give a simplified account of both components. We start with a model of NMDA synapses and turn then to the so-called calcium control hypothesis of Shouval et al. (2001).

NMDA receptor as a coincidence detector

We have emphasized in Sections 10.1–10.3 that all Hebbian learning rules contain a term that depends on the *correlation* between the firings of pre- and postsynaptic neurons. The signaling chain that leads to a weight change therefore has to contain a nonlinear processing step that requires that pre- and postsynaptic neurons are active within some short time window. Synaptic channels controlled by NMDA receptors are an excellent candidate for a biophysical implementation of this condition of 'coincidence' because the opening of the channel requires both the presence of glutamate which reflects presynaptic activity *and*, in order to remove the magnesium block, a depolarization of the postsynaptic membrane

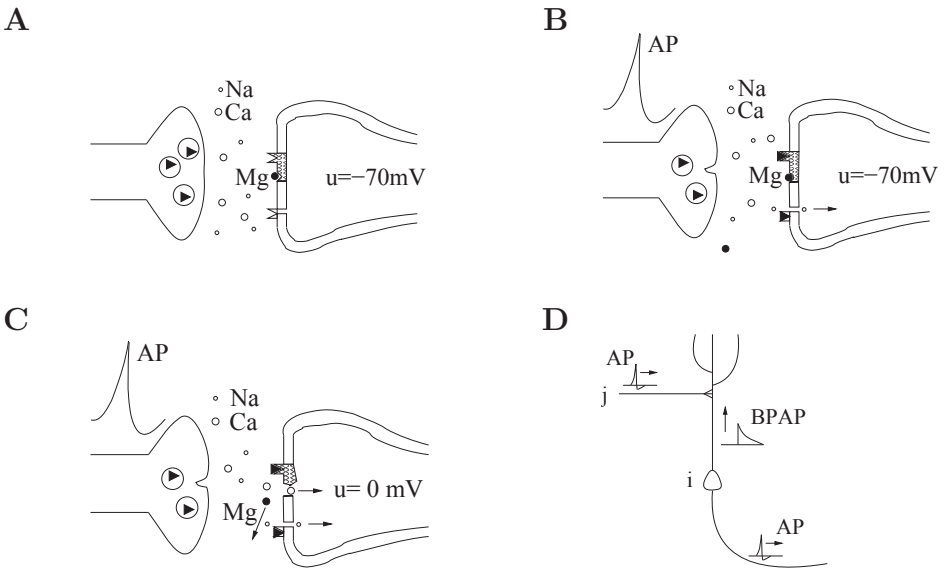

Fig. 10.10: NMDA-synapse. **A**. Vesicles in the presynaptic terminal contain glutamate as a neurotransmitter (filled triangles). At resting potential, the NMDA receptor mediated channel (hatched) is blocked by magnesium (filled circle). **B**. If an action potential (AP) arrives at the presynaptic terminal the vesicle merges with the cell membrane, glutamate diffuses into the synaptic cleft, and binds to NMDA and non-NMDA receptors on the postsynaptic membrane. At resting potential, the NMDA receptor mediated channel remains blocked by magnesium whereas the non-NMDA channel opens (bottom). **C**. If the membrane of the postsynaptic neuron is depolarized, the magnesium block is removed and calcium ions can enter into the cell. **D**. The depolarization of the postsynaptic membrane can be caused by a back propagating action potential (BPAP).

(Mayer et al., 1984; Nowak et al., 1984). A strong depolarization of the postsynaptic membrane does occur, for example, during the back propagation of an action potential into the dendritic tree (Stuart and Sakmann, 1994; Linden, 1999), which is a signature for *postsynaptic* activity.

In a simple model of NMDA-receptor controlled channels, the calcium current through the channel is described by

$$I_{\rm Ca}(t) = g_{\rm Ca}\, \alpha(t - t_j^{(f)})\, [u(t) - E_{\rm Ca}]\, B[u(t)]\,; \tag{10.52}$$

cf. Chapter 2.4. Here $g_{\rm Ca}$ is the maximal conductance of the channel and $E_{\rm Ca}$ is the reversal potential of calcium. The time course of NMDA binding at the receptors is described by $\alpha(t - t_j^{(f)})$ where $t_j^{(f)}$ is the time of spike arrival at the presynaptic terminal. The function

$$B(u) = \frac{1}{1 + 0.28\, e^{-0.062\, u}} \tag{10.53}$$

describes the unblocking of the channel at depolarized levels of membrane potential.

The time course of α is taken as a sum of two exponentials with the time constant of the slow component in the range of 100 ms. If there are several presynaptic spikes within 100 ms, calcium accumulates inside the cell. The change of the intracellular calcium concentration $[Ca^{2+}]$ can be described by

$$\frac{\mathrm{d}}{\mathrm{d}t}[Ca^{2+}](t) = I_{Ca}(t) - \frac{[Ca^{2+}](t)}{\tau_{Ca}}, \tag{10.54}$$

where $\tau_{Ca} = 125$ ms is a phenomenological time constant of decay. Without any further presynaptic stimulus, the calcium concentration returns to a resting value of zero. More sophisticated models can take calcium buffers, calcium stores, and ion pumps into account (Gamble and Koch, 1987; Zador et al., 1990; Schiegg et al., 1995).

The calcium control hypothesis

While the dynamics of NMDA synapses is fairly well understood in terms of the biophysical processes that control receptor binding and channel opening, much less is known about the complex signaling chain that is triggered by calcium and finally leads to a regulation of the synaptic efficacy (Lisman, 1989). Instead of a developing a detailed model, we adopt a phenomenological approach and assume that the change of the synaptic efficacy w_{ij} is fully determined by the intracellular calcium concentration $[Ca^{2+}]$; an assumption that has been called 'calcium control hypothesis' (Shouval et al., 2001). More specifically, we write the weight change as

$$\frac{\mathrm{d}}{\mathrm{d}t}w_{ij} = \frac{\Omega([Ca^{2+}]) - w_{ij}}{\tau([Ca^{2+}])}. \tag{10.55}$$

For constant calcium concentration, the weight w_{ij} reaches an asymptotic value $\Omega([Ca^{2+}])$ with time constant $\tau([Ca^{2+}])$.

Figure 10.11 shows the graph of the function $\Omega([Ca^{2+}])$ as it is used in the model of Shouval et al. (2001). For a calcium concentration below θ_0, the weight assumes a resting value of $w_0 = 0.5$. For calcium concentrations in the range $\theta_0 < [Ca^{2+}] < \theta_m$, the weight tends to decrease, for $[Ca^{2+}] > \theta_m$ it increases. Qualitatively, the curve $\Omega([Ca^{2+}])$ reproduces experimental results that suggest that a high level of calcium leads to an increase whereas an intermediate level of calcium leads to a decrease of synaptic weights. We will see below that the BCM rule of Eq. (10.12) is closely related to the function $\Omega[Ca^{2+}]$.

The time constant $\tau([Ca^{2+}])$ in the model equation (10.55) decreases rapidly with increasing calcium concentration; cf. Fig. 10.11B. The specific dependence has been taken as

$$\tau([Ca^{2+}]) = \frac{\tau_0}{[Ca^{2+}]^3 + 10^{-4}} \tag{10.56}$$

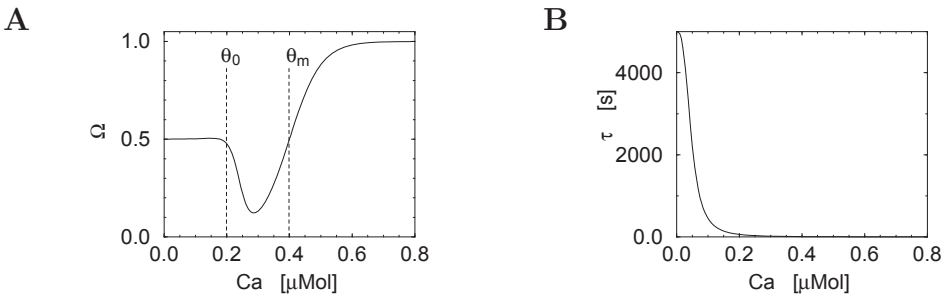

Fig. 10.11: Calcium control hypothesis. The asymptotic weight value $w_{ij} = \Omega([\mathrm{Ca}^{2+}])$ (**A**) and the time constant $\tau([\mathrm{Ca}^{2+}])$ of weight changes (**B**) as a function of the calcium concentration; cf. Eq. (10.55) [adapted from Shouval et al. (2001)].

where $\tau_0 = 500\,\mathrm{ms}$ and $[\mathrm{Ca}^{2+}]$ is the calcium concentration in μmol/l. At a low level of intracellular calcium ($[\mathrm{Ca}^{2+}] \to 0$), the response time of the weight w_{ij} is in the range of hours while for $[\mathrm{Ca}^{2+}] \to 1$ the weight changes rapidly with a time constant of 500 ms. In particular, the effective time constant for the induction of LTP is shorter than that for LTD.

Dynamics of the postsynaptic neuron

In order to complete the definition of the model, we need to introduce a description of the membrane potential u_i of the postsynaptic neuron. As in the simple spiking neuron model SRM_0 (cf. Chapter 4), the total membrane potential is described as

$$u_i(t) = \eta(t - \hat{t}_i) + \sum_f \epsilon(t - t_j^{(f)})\,. \tag{10.57}$$

Here $\epsilon(t - t_j^{(f)})$ is the time course of the postsynaptic potential generated by a presynaptic action potential at time $t_j^{(f)}$. It is modeled as a double exponential with a rise time of about 5 ms and a duration of about 50 ms. The action potential of the postsynaptic neuron is described as

$$\eta(s) = u_{\mathrm{AP}} \left(0.75\, e^{-s/\tau_{\mathrm{fast}}} + 0.25\, e^{-s/\tau_{\mathrm{slow}}}\right)\,. \tag{10.58}$$

Here $u_{\mathrm{AP}} = 100\,\mathrm{mV}$ is the amplitude of the action potential and $\hat{t}_i$ is the firing time of the last spike of the postsynaptic neuron. In contrast to the model SRM_0, η does not describe the reset of the membrane potential at the soma, but the form of the *back propagating* action potential (BPAP) at the site of the synapse. It is assumed that the BPAP has a slow component with a time constant $\tau_{\mathrm{slow}} = 35\,\mathrm{ms}$. The fast component has the same rapid time constant (about 1 ms) as the somatic action potential. The somatic action potential is not described explicitly.

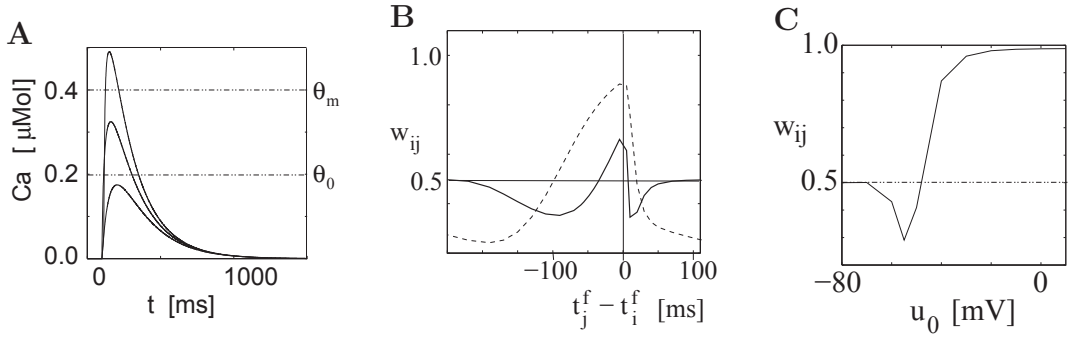

Fig. 10.12: Spike-time dependent plasticity in a calcium-based model. **A**. Calcium transient generated by a presynaptic spike in the absence of postsynaptic firing (bottom) and in the presence of a postsynaptic spike 10 ms before (middle) or 10 ms after (top) presynaptic spike arrival. Only for the sequence 'pre-before-post' the threshold θ_m for LTP can be reached. **B**. The final weights obtained after several thousands pre- and postsynaptic spikes that are generated at a rate of 1 Hz (solid line) or 3 Hz (dashed line). The weights are given as a function of the time difference between presynaptic spikes $t_j^{(f)}$ and postsynaptic spikes $t_i^{(f)}$. **C**. Pairing of presynaptic spikes with postsynaptic depolarization. The weights w_{ij} that are obtained after several hundreds of presynaptic spikes (at a rate of $\nu_j = 0.5$ Hz) as a function of the depolarization of the postsynaptic membrane. [Adapted from Shouval et al. (2001)].

Results

Given the above components of the model, we can understand intuitively how calcium influx at NMDA synapses leads to spike-time dependent plasticity. Let us analyze the behavior by comparing the calcium-based model with the elementary model of Section 10.4.1; cf. Eqs. (10.29)–(10.31). Binding of glutamate at NMDA receptors plays the role of the component a that is triggered by presynaptic firing; the back propagating action potential plays the role of the component b that is triggered by postsynaptic firing. As a result of the depolarization caused by the BPAP, the magnesium block is removed and calcium ions enter the cell. The calcium influx is proportional to the product of the NMDA-binding, i.e., the factor α in Eq. (10.52), and the unblocking, i.e., the factor $B(u)$. Finally, the increase in the calcium concentration leads to a weight change according to Eq. (10.55).

A single presynaptic spike (without a simultaneous postsynaptic action potential) leads to a calcium transient that stays below the induction threshold θ_0; cf. Fig. 10.12A. If a postsynaptic spike occurs 10 ms *before* the presynaptic spike arrival, the calcium transient has a somewhat larger amplitude that attains a level above θ_0. As a consequence, the weight w_{ij} is reduced. If, however, the postsynaptic spike occurs one or a few milliseconds *after* the presynaptic one, the calcium transient is much larger. The reason is that the blocking of the NMDA

synapse is removed during the time when the NMDA receptors are almost completely saturated by glutamate. In this case, the calcium concentration is well above θ_m so that weights increase. Since the time constant $\tau([Ca^{2+}])$ is shorter in the regime of LTP induction than in the regime of LTD induction, the positive weight change is dominant even though the calcium concentration must necessarily pass through the regime of LTD in order to reach the threshold θ_m. The resulting time window of learning is shown in Fig. 10.12B. It exhibits LTP if the presynaptic spike precedes the postsynaptic one by less than 40 ms. If the order of spiking is inverted LTD occurs. LTD can also be induced by a sequence of 'pre-before-post' if the spike time difference is larger than about 40 ms. The reason is that in this case the removal of the magnesium block (induced by the BPAP) occurs at a moment when the probability of glutamate binding is reduced; cf. the factor α in Eq. (10.52). As a consequence less calcium enters the cell – enough to surpass the threshold θ_0, but not sufficient to reach the threshold θ_1 of LTP. We emphasize that the form of the learning window is not fixed but depends on the frequency of pre- and postsynaptic spike firing; cf. Fig. 10.12B.

LTP and LTD can also be induced in the *absence* of postsynaptic spikes if the membrane potential of the postsynaptic neuron is clamped to a constant value. A pure spike-time dependent learning rule defined by a learning window $W(t_j^{(f)} - t_i^{(f)})$ is obviously not a suitable description of such an experiment. The calcium-based model of Shouval et al. (2001), however, can reproduce voltage-clamp experiments; cf. Fig. 10.12C. Presynaptic spike arrival at low frequency ($\nu_j = 0.5$ Hz) is 'paired' with a depolarization of the membrane potential of the postsynaptic neuron to a fixed value u_0. If u_0 is below -70 mV, no significant weight change occurs. For $-70\,\text{mV} < u_0 < -50\,\text{mV}$ LTD is induced, while for $u_0 > -50\,\text{mV}$ LTP is triggered. These results are a direct consequence of the removal of the magnesium block at the NMDA synapses with increasing voltage. The mean calcium concentration – and hence the asymptotic weight value – is therefore a monotonously increasing function of u_0.

Finally, we would like to emphasized the close relation between Fig. 10.12C and the function ϕ of the BCM learning rule as illustrated in Fig. 10.5. In a simple rate model, the postsynaptic firing rate ν_i is a sigmoidal function of the potential, i.e., $\nu_i = g(u_i)$. Thus, the mapping between the two figures is given by a non-linear transformation of the horizontal axis.

10.5 Summary

Correlation-based learning is, as a whole, often called Hebbian learning. The Hebb rule (10.2) is a special case of a *local* learning rule because it only depends on pre- and postsynaptic firing rates and the present state w_{ij} of the synapse, i.e., information that is easily 'available' at the location of the synapse.

Recent experiments have shown that the relative timing of pre- and postsynaptic spikes critically determines the amplitude and even the direction of changes of the synaptic efficacy. In order to account for these effects, learning rules on the level of individual spikes are formulated with a learning window that consists of two parts: If the presynaptic spike arrives before a postsynaptic output spike, the synaptic change is positive. If the timing is the other way round, the synaptic change is negative (Markram et al., 1997; Zhang et al., 1998; Debanne et al., 1998; Bi and Poo, 1998, 1999). For some synapses, the learning window is reversed (Bell et al., 1997b), for others it contains only a single component (Egger et al., 1999).

Hebbian learning is considered to be a major principle of neuronal organization during development. The first modeling studies of cortical organization development (Willshaw and von der Malsburg, 1976; Swindale, 1982) have incited a long line of research, e.g., Kohonen (1984); Linsker (1986c,b,a); MacKay and Miller (1990); Miller et al. (1989); Obermayer et al. (1992). Most of these models use in some way or another an unsupervised correlation-based learning rule similar to the general Hebb rule of Eq. (10.2); see Erwin et al. (1995) for a recent review.

Literature

Correlation-based learning can be traced back to Aristoteles [2] and has been discussed extensively by James (1890) who formulated a learning principle on the level of 'brain processes' rather than neurons:

> When two elementary brain-processes have been active together or in immediate succession, one of them, on re-occurring, tends to propagate its excitement into the other.

A chapter of James' book is reprinted in volume 1 of Anderson and Rosenfeld's collection on Neurocomputing (Anderson and Rosenfeld, 1988). More than 50 years later, Hebb's book (Hebb, 1949) of which two interesting sections are reprinted in the collection of Anderson and Rosenfeld (1988) was published. The historical context of Hebb's postulate is discussed in the review of Sejnowski (1999). In the reprint volume of Anderson and Rosenfeld (1988), articles of Grossberg (1976) and Bienenstock et al. (1982) illustrate the use of the rate-based learning rules discussed in Section 10.2. Kohonen's book gives an overview of some mathematical results for several generic rate-based learning rules (Kohonen, 1984).

[2] Aristoteles, "De memoria et reminiscentia": There is no need to consider how we remember what is distant, but only what is neighboring, for clearly the method is the same. For the changes follow each other by habit, one after another. And thus, whenever someone wishes to recollect he will do the following: He will seek to get a starting point for a change after which will be the change in question.

For reviews on (hippocampal) LTP, see the book of Byrne and Berry (1989), in particular the articles of Sejnowski and Tesauro (1989) and Brown et al. (1989). Cerebellar LTD has been reviewed by Daniel et al. (1996, 1998) and Linden and Connor (1995). Further references and a classification of different forms of LTP and LTD can be found in the nice review of Bliss and Collingridge (1993). For the relation of LTP and LTD, consult Artola and Singer (1993). A modern and highly recommendable review with a focus on recent results, in particular on spike-time dependent plasticity has been written by Bi and Poo (2001). The theoretical context of spike-time dependent plasticity has been discussed by Abbott (2000).

Chapter 11

Learning Equations

Neurons in the central nervous system form a complex network with a high degree of plasticity. In the previous chapter we have discussed synaptic plasticity from a phenomenological point of view. We now ask 'What are the consequences for the connectivity between neurons if synapses are plastic?'. To do so we consider a scenario known as unsupervised learning. We assume that some of the neurons in the network are stimulated by input with certain statistical properties. Synaptic plasticity generates changes in the connectivity pattern that reflect the statistical structure of the input. The relation between the input statistics and the synaptic weights that evolve due to Hebbian plasticity is the topic of this chapter. We start in Section 11.1 with a review of unsupervised learning in a rate-coding paradigm. The extension of the analysis to spike-time dependent synaptic plasticity is made in Section 11.2. We will see that spike-based learning naturally accounts for spatial *and* temporal correlations in the input and can overcome some of the problems of a simple rate-based learning rule.

11.1 Learning in Rate Models

We would like to understand how activity-dependent learning rules influence the formation of connections between neurons in the brain. We will see that plasticity is controlled by the statistical properties of the presynaptic input that is impinging on the postsynaptic neuron. Before we delve into the analysis of the elementary Hebb rule we therefore need to recapitulate a few results from statistics and linear algebra.

11.1.1 Correlation Matrix and Principal Components

A principal component analysis (PCA) is a standard technique to describe statistical properties of a set of high-dimensional data points and is usually performed in order to find those components of the data that show the highest variability

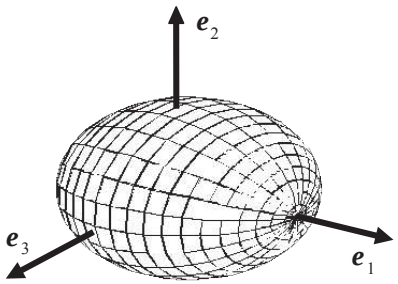

Fig. 11.1: Ellipsoid approximating the shape of a cloud of data points. The first principal component $\boldsymbol{e}_1$ corresponds to the principal axis of the ellipsoid.

within the set. If we think of the input data set as of a cloud of points in a high-dimensional vector space centered around the origin, then the first principal component is the direction of the longest axis of the ellipsoid that encompasses the cloud; cf. Fig. 11.1. If the data points consisted of, say, two separate clouds then the first principal component would give the direction of a line that connects the center points of the two clouds. A PCA can thus be used to break a large data set into separate clusters. In the following, we will quickly explain the basic idea and show that the first principal component gives the direction where the variance of the data is maximal.

Let us consider an ensemble of data points $\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^p\}$ drawn from a (high-dimensional) vector space, for example $\boldsymbol{\xi}^\mu \in \mathbb{R}^N$. For this set of data points we define the *correlation matrix* C_{ij} as

$$C_{ij} = \frac{1}{p} \sum_{\mu=1}^{p} \xi_i^\mu \, \xi_j^\mu = \left\langle \xi_i^\mu \, \xi_j^\mu \right\rangle_\mu \, . \tag{11.1}$$

Angular brackets $\langle \cdot \rangle_\mu$ denote an average over the whole set of data points. Similar to the variance of a single random variable we can also define the *covariance matrix* V_{ij} of our data set,

$$V_{ij} = \left\langle (\xi_i^\mu - \langle \xi_i^\mu \rangle_\mu) \, (\xi_j^\mu - \langle \xi_j^\mu \rangle_\mu) \right\rangle_\mu \, . \tag{11.2}$$

In the following we will assume that the coordinate system is chosen so that the center of mass of the set of data points is located at the origin, i.e., $\langle \xi_i \rangle_\mu = \langle \xi_j \rangle_\mu = 0$. In this case, correlation matrix and covariance matrix are identical.

The *principal components* of the set $\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^p\}$ are defined as the eigenvectors of the covariance matrix V. Note that V is symmetric, i.e., $V_{ij} = V_{ji}$. The eigenvalues of V are thus real-valued and different eigenvectors are orthogonal

(Horn and Johnson, 1985). Furthermore, V is positive semi-definite since

$$\boldsymbol{y}^{\mathrm{T}} V \boldsymbol{y} = \sum_{ij} y_i \left\langle \xi_i^\mu \xi_j^\mu \right\rangle_\mu y_j = \left\langle \left[\sum_i y_i \xi_i^\mu \right]^2 \right\rangle_\mu \geq 0 \tag{11.3}$$

for any vector $\boldsymbol{y} \in \mathbb{R}^N$. Therefore, all eigenvalues of V are non-negative.

We can sort the eigenvectors $\boldsymbol{e}_i$ according to the size of the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. The eigenvector with the largest eigenvalue is called the first principal component. It points in the direction where the variance of the data is maximal. To see this we calculate the variance of the projection of $\boldsymbol{\xi}^\mu$ onto an arbitrary direction $\boldsymbol{y}$ that we write as $\boldsymbol{y} = \sum_i a_i \boldsymbol{e}_i$ with $\sum_i a_i^2 = 1$ so that $\|\boldsymbol{y}\| = 1$. The variance $\sigma_{\boldsymbol{y}}^2$ along $\boldsymbol{y}$ is

$$\sigma_{\boldsymbol{y}}^2 = \left\langle [\boldsymbol{x}^\mu \cdot \boldsymbol{y}]^2 \right\rangle_\mu = \boldsymbol{y}^{\mathrm{T}} V \boldsymbol{y} = \sum_i \lambda_i a_i^2 \,. \tag{11.4}$$

The right-hand side is maximal under the constraint $\sum_i a_i^2 = 1$ if $a_1 = 1$ and $a_i = 0$ for $i = 2, 3, \ldots, N$, that is, if $\boldsymbol{y} = \boldsymbol{e}_1$.

11.1.2 Evolution of synaptic weights

In the following we analyze the evolution of synaptic weights using the Hebbian learning rules that have been described in Chapter 10. To do so, we consider a highly simplified scenario consisting of an analog neuron that receives input from N presynaptic neurons with firing rates ν_i^{pre} via synapses with weights w_i; cf. Fig. 11.2A. We think of the presynaptic neurons as 'input neurons', which, however, do not have to be sensory neurons. The input layer could, for example, consist of neurons in the lateral geniculate nucleus (LGN) that project to neurons in the visual cortex. We will see that the statistical properties of the input control the evolution of synaptic weights.

For the sake of simplicity, we model the presynaptic input as a set of static patterns. Let us suppose that we have a total of p patterns $\{\boldsymbol{\xi}^\mu; 1 < \mu < p\}$. At each time step one of the patterns $\boldsymbol{\xi}^\mu$ is selected at random and presented to the network by fixing the presynaptic rates at $\nu_i^{\mathrm{pre}} = \xi_i^\mu$. We call this the *static-pattern scenario*. The presynaptic activity drives the postsynaptic neuron and the joint activity of pre- and postsynaptic neurons triggers changes of the synaptic weights. The synaptic weights are modified according to a Hebbian learning rule, i.e., according to the correlation of pre- and postsynaptic activity; cf. Eq. (10.3). Before the next input pattern is chosen, the weights are changed by an amount

$$\Delta w_i = \gamma \, \nu^{\mathrm{post}} \, \nu_i^{\mathrm{pre}} \tag{11.5}$$

Here, $0 < \gamma \ll 1$ is a small constant called 'learning rate'. The learning rate in the static-pattern scenario is closely linked to the correlation coefficient c_2^{corr} in

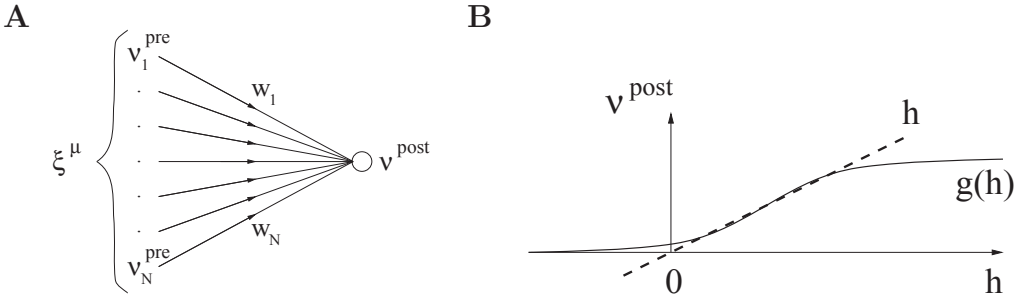

Fig. 11.2: Elementary model. **A**. Patterns $\boldsymbol{\xi}^\mu$ are applied as a set of presynaptic firing rates ν_j, i.e., $\boldsymbol{\xi}_j^\mu = \nu_j^{\rm pre}$ for $1 \leq j \leq N$. **B**. The gain function of the postsynaptic neuron is taken as linear, i.e., $\nu^{\rm post} = h$. It can be seen as a linearization of the sigmoidal gain function $g(h)$.

the continuous-time Hebb rule introduced in Eq. (10.3). In order to highlight the relation, let us assume that each pattern $\boldsymbol{\xi}^\mu$ is applied during an interval Δt. For Δt sufficiently small, we have $\gamma = c_2^{\rm corr}\, \Delta t$.

In a general rate model, the firing rate $\nu^{\rm post}$ of the postsynaptic neuron is given by a nonlinear function of the total input

$$\nu^{\rm post} = g\left(\sum_i w_i\, \nu_i^{\rm pre} \right) \; ; \tag{11.6}$$

cf. Fig. 11.2B. For the sake of simplicity, we restrict our discussion in the following to a *linear* rate model with

$$\nu^{\rm post} = \sum_i w_i\, \nu_i^{\rm pre} \, . \tag{11.7}$$

Obviously, this is a highly simplified neuron model, but it will serve our purpose of gaining some insights in the evolution of synaptic weights.

If we combine the learning rule (11.5) with the linear rate model of Eq. (11.7) we find after the presentation of pattern $\boldsymbol{\xi}^\mu$

$$\Delta w_i = \gamma \sum_j w_j\, \nu_j^{\rm pre}\, \nu_i^{\rm pre} = \gamma \sum_j w_j\, \xi_j^\mu\, \xi_i^\mu \, . \tag{11.8}$$

The evolution of the weight vector $\boldsymbol{w} = (w_1, \ldots, w_N)$ is thus determined by the iteration

$$w_i(n+1) = w_i(n) + \gamma \sum_j w_j\, \xi_j^{\mu_n}\, \xi_i^{\mu_n} \, , \tag{11.9}$$

where μ_n denotes the pattern that is presented during the nth time step.

We are interested in the long-term behavior of the synaptic weights. To this end we assume that the weight vector evolves along a more or less deterministic

trajectory with only small stochastic deviations that result from the randomness at which new input patterns are chosen. This is, for example, the case if the learning rate is small so that a large number of patterns has to be presented in order to induce a substantial weight change. In such a situation it is sensible to consider the expectation value of the weight vector, i.e., the weight vector $\langle \boldsymbol{w}(n) \rangle$ averaged over the sequence $(\boldsymbol{\xi}^{\mu_1}, \boldsymbol{\xi}^{\mu_2}, \ldots, \boldsymbol{\xi}^{\mu_n})$ of all patterns that so far have been presented to the network. From Eq. (11.9) we find

$$\begin{aligned} \langle w_i(n+1) \rangle &= \langle w_i(n) \rangle + \gamma \sum_j \left\langle w_j(n)\, \xi_j^{\mu_{n+1}}\, \xi_i^{\mu_{n+1}} \right\rangle \\ &= \langle w_i(n) \rangle + \gamma \sum_j \langle w_j(n) \rangle \left\langle \xi_j^{\mu_{n+1}}\, \xi_i^{\mu_{n+1}} \right\rangle \\ &= \langle w_i(n) \rangle + \gamma \sum_j C_{ij}\, \langle w_j(n) \rangle \; . \end{aligned} \tag{11.10}$$

The angular brackets denote an ensemble average over the whole sequence of input patterns $(\boldsymbol{\xi}^{\mu_1}, \boldsymbol{\xi}^{\mu_2}, \ldots)$. The second equality is due to the fact that input patterns are chosen *independently* in each time step, so that the average over $w_j(n)$ and $(\xi_j^{\mu_{n+1}}\, \xi_i^{\mu_{n+1}})$ can be factorized. In the final expression we have introduced the correlation matrix C_{ij},

$$C_{ij} = \frac{1}{p} \sum_{\mu=1}^{p} \xi_i^\mu\, \xi_j^\mu = \left\langle \xi_i^\mu\, \xi_j^\mu \right\rangle_\mu \; . \tag{11.11}$$

Expression (11.10) can be written in a more compact form using matrix notation,

$$\langle \boldsymbol{w}(n+1) \rangle = (\mathbb{1} + \gamma\, C)\, \langle \boldsymbol{w}(n) \rangle = (\mathbb{1} + \gamma\, C)^{n+1}\, \langle \boldsymbol{w}(0) \rangle \; , \tag{11.12}$$

where $\boldsymbol{w}(n) = (w_1(n), \ldots, w_N(n))$ is the weight vector and $\mathbb{1}$ is the identity matrix.

If we express the weight vector in terms of the eigenvectors $\boldsymbol{e}_k$ of C,

$$\langle \boldsymbol{w}(n) \rangle = \sum_k a_k(n)\, \boldsymbol{e}_k \, , \tag{11.13}$$

we obtain an explicit expression for $\langle \boldsymbol{w}(n) \rangle$ for any given initial condition $a_k(0)$, viz.,

$$\langle \boldsymbol{w}(n) \rangle = \sum_k (1 + \lambda_k)^n\, a_k(0)\, \boldsymbol{e}_k \, . \tag{11.14}$$

Since the correlation matrix is positive semi-definite all eigenvalues λ_k are real and positive. Therefore, the weight vector is growing exponentially, but the growth will soon be dominated by the eigenvector with the largest eigenvalue, i.e., the *first principal component*,

$$\langle \boldsymbol{w}(n) \rangle \xrightarrow{n \to \infty} (1 + \lambda_1)^n\, a_1(0)\, \boldsymbol{e}_1 \, ; \tag{11.15}$$

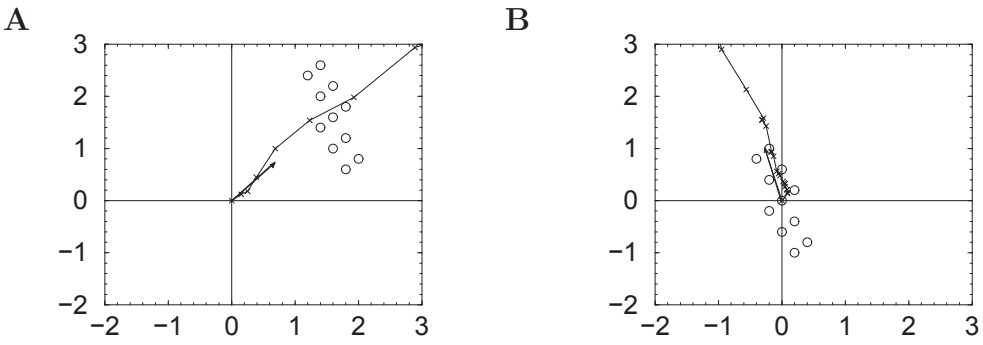

Fig. 11.3: Weight changes induced by the standard Hebb rule. Input patterns $\boldsymbol{\xi}^\mu \in \mathbb{R}^2$ are marked as circles. The sequence of weight vectors $\boldsymbol{w}(1)$, $\boldsymbol{w}(2)$, ..., is indicated by crosses connected by a solid line. **A**. The weight vector evolves in the direction of the dominant eigenvector (arrow) of the correlation matrix. **B**. If the input patterns are normalized so that their center of mass is at the origin, then the dominant eigenvector of the correlation matrix coincides with the first principal component $\boldsymbol{e}_1$ of the data set.

cf. Section 11.1.1. Recall that the output of the linear neuron model (11.7) is proportional to the projection of the current input pattern $\boldsymbol{\xi}^\mu$ on the direction $\boldsymbol{w}$. For $\boldsymbol{w} \propto \boldsymbol{e}_1$, the output is therefore proportional to the projection on the first principal component of the input distribution. A Hebbian learning rule such as Eq. (11.8) is thus able to extract the first principal component of the input data.

From a data-processing point of view, the extraction of the first principle component of the input data set by a biologically inspired learning rule seems to be very compelling. There are, however, a few drawbacks and pitfalls. First, the above statement about the Hebbian learning rule is limited to the *expectation value* of the weight vector. We will see below that, if the learning rate is sufficiently low, then the actual weight vector is in fact very close to the expected one.

Second, while the direction of the weight vector moves in the direction of the principal component, the *norm* of the weight vector grows without bounds. We will see below in Section 11.1.3 that suitable variants of Hebbian learning allow us to control the length of the weight vector without changing its direction.

Third, principal components are only meaningful if the input data is normalized, i.e., distributed around the origin. This requirement is *not* consistent with a rate interpretation because rates are usually positive. This problem, however, can be overcome by learning rules such as the covariance rule of Eq. (10.10) that are based on the deviation of the rates from a certain mean firing rate. We will see in Section 11.2.4 that a spike-based learning rule can be devised that is sensitive only to deviations from the mean firing rate and can thus find the first principal component even if the input is not properly normalized.

Self-averaging (*)

So far, we have derived the behavior of the *expected* weight vector, $\langle \boldsymbol{w} \rangle$. Here we show that explicit averaging is not necessary provided that learning is slow enough. In this case, the weight vector is the sum of a large number of small changes. The weight dynamics is thus *'self-averaging'* and the weight vector $\boldsymbol{w}$ can be well approximated by its expectation value $\langle \boldsymbol{w} \rangle$.

We start from the formulation of Hebbian plasticity in continuous time,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_i = c_2^{\mathrm{corr}} \, \nu^{\mathrm{post}} \, \nu_i^{\mathrm{pre}} ; \tag{11.16}$$

cf. Eq. (10.3). Each pattern $\boldsymbol{\xi}^\mu$ is presented for a short period of duration Δt. We assume that the weights change during the presentation by a small amount only, i.e., $\int_t^{t+\Delta t} [\mathrm{d}w_j(t')/\mathrm{d}t'] \, \mathrm{d}t' \ll w_j(t)$. This condition can be met either by a short presentation time Δt or by a small learning coefficient c_2^{corr}. Under this condition, we can take the postsynaptic firing rate $\nu^{\mathrm{post}}(t) = \sum_j w_j(t) \, \nu_i^{\mathrm{pre}}$ as constant for the duration of one presentation. The total weight change induced by the presentation of pattern $\boldsymbol{\xi}^\mu$ to first order in Δt is thus

$$\Delta w_i(t) = w_i(t + \Delta t) - w_i(t) = \gamma \sum_j w_j(t) \, \xi_j^\mu \, \xi_i^\mu + \mathcal{O}(\Delta t^2) \, . \tag{11.17}$$

with $\gamma = c_2^{\mathrm{corr}} \, \Delta t$; cf. Eq. (11.8).

In the next time step a new pattern $\boldsymbol{\xi}^\nu$ is presented so that the weight is changed to

$$w_i(t + 2\Delta t) = w_i(t + \Delta t) + c_2^{\mathrm{corr}} \, \Delta t \sum_j w_j(t + \Delta t) \; \xi_j^\nu \, \xi_i^\nu + \mathcal{O}(\Delta t^2) \, . \tag{11.18}$$

Since we keep only terms to first order in Δt, we may set $w_j(t + \Delta t) = w_j(t)$ in the sum on the right-hand side of Eq. (11.18). Let us suppose that in the interval $[t, t + p \, \Delta t]$ each of the p patterns has been applied exactly once. Then, to first order in Δt,

$$w_i(t + p \, \Delta t) - w_i(t) = c_2^{\mathrm{corr}} \, \Delta t \sum_j w_j(t) \sum_{\mu=1}^{p} \xi_i^\mu \, \xi_j^\mu + \mathcal{O}(\Delta t^2) \, . \tag{11.19}$$

For $c_2^{\mathrm{corr}} \, \Delta t \ll 1$, all higher-order terms can be neglected. Division by $p \, \Delta t$ yields

$$\frac{w_i(t + p \, \Delta t) - w_i(t)}{p \, \Delta t} = c_2^{\mathrm{corr}} \sum_j w_j(t) \, C_{ij} \, . \tag{11.20}$$

The left-hand side can be approximated by a differential operator $\mathrm{d}w/\mathrm{d}t$,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_i(t) = c_2^{\mathrm{corr}} \sum_j w_j(t) \, C_{ij} \, . \tag{11.21}$$

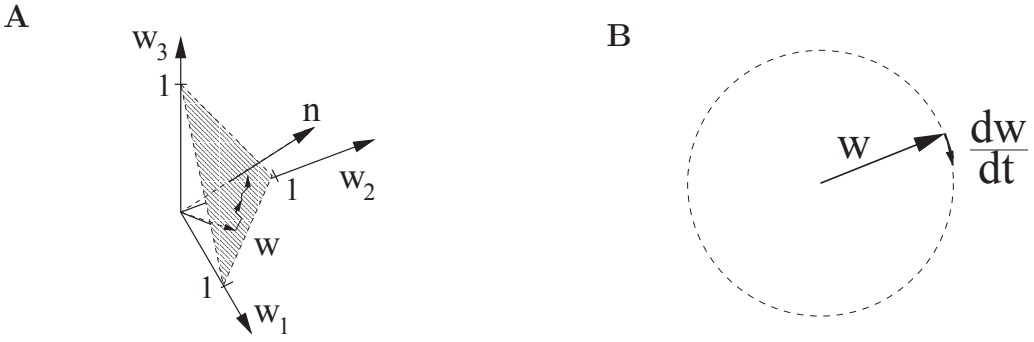

Fig. 11.4: Weight vector normalization. **A**. Normalization of the summed weights $\sum_j w_{ij} = 1$ constrains the weight vector $\boldsymbol{w}$ to a hyper-plane perpendicular to the diagonal vector $\boldsymbol{n} = (1, 1, \ldots, 1)^T$. Hard bounds $0 \leq w_{ij} \leq 1$ force the the weight vector to stay inside the shaded region. **B**. Normalization of the quadratic norm $\|\boldsymbol{w}\|^2 = 1$. The weight change $\Delta\boldsymbol{w}(n)$ is perpendicular to the current weight vector $\boldsymbol{w}(n)$ so that the length of $\boldsymbol{w}$ remains constant (Oja's learning rule).

We thus recover our previous result that weights are driven by the correlations in the input but with the additional vantage that no explicit averaging step is necessary (Sanders and Verhulst, 1985).

11.1.3 Weight Normalization

We have seen in Section 11.1.2 that the simple learning rule (10.3) leads to exponentially growing weights. Since this is biologically not plausible, we must use a modified Hebbian learning rule that includes weight decrease and saturation; cf. Chapter 10.2. Particularly interesting are learning rules that lead to a normalized weight vector. Normalization is a desirable property since it leads to a competition between synaptic weights w_{ij} that converge on the same postsynaptic neuron i. Competition means that if a synaptic efficacy increases, it does so at the expense of other synapses that must decrease.

For a discussion of weight vector normalization two aspects are important, namely *what* is normalized and *how* the normalization is achieved. Learning rules can be designed to normalize either the *sum* of weights, $\sum_j w_{ij}$, or the quadratic *norm*, $\|\boldsymbol{w}\|^2 = \sum_j w_{ij}^2$ (or any other norm on $\mathbb{R}^N$). In the first case, the weight vector is constrained to a plane perpendicular to the diagonal vector $\boldsymbol{n} = (1, \ldots, 1)$; in the second case it is constrained to a hyper-sphere; cf. Fig. 11.4.

Second, the normalization of the weight vector can either be multiplicative or subtractive. In the former case all weights are multiplied by a common factor so that large weights w_{ij} are corrected by a larger amount than smaller ones. In the latter case a common constant is subtracted from each weight. Usually, subtractive normalization is combined with hard bounds $0 \leq w_{ij} \leq w^{\max}$ in order

to avoid runaway of individual weights. Finally, learning rules may or may not fall into the class of *local* learning rules that we have considered in Chapter 10.2.

A systematic classification of various learning rules according to the above three criteria has been proposed by Miller and MacKay (1994). Here we restrict ourselves to two instances of learning with normalization properties which we illustrate in the examples below. We start with the subtractive normalization of the summed weights $\sum_j w_{ij}$ and turn then to a discussion of Oja's rule as an instance of a multiplicative normalization of $\sum_j w_{ij}^2$.

Example: Subtractive Normalization of $\sum_i w_i$

In a subtractive normalization scheme the sum over all weights, $\sum_i w_i$, can be kept constant by subtracting the average total weight change, $N^{-1} \sum_i \Delta\tilde{w}_i$, from each synapse after the weights have been updated according to a Hebbian learning rule with $\Delta\tilde{w}_i = \gamma \sum_j w_j \, \xi_j^\mu \, \xi_i^\mu$. Altogether, the learning rule is of the form

$$\begin{aligned} \Delta w_i &= \Delta\tilde{w}_i - N^{-1} \sum_j \Delta\tilde{w}_j \\ &= \gamma \Big(\sum_j w_j \, \xi_j^\mu \, \xi_i^\mu - N^{-1} \sum_k \sum_j w_j \, \xi_j^\mu \, \xi_k^\mu \Big) \, , \end{aligned} \tag{11.22}$$

where $\Delta\tilde{w}_i$ denotes the weight change that is due to the pure Hebbian learning rule without the normalization. It can easily be verified that $\sum_i \Delta w_i = 0$ so that $\sum_i w_i = \text{const}$. The temporal evolution of the weight vector $\boldsymbol{w}$ is thus restricted to a hyperplane perpendicular to $(1, \ldots, 1) \in \mathbb{R}^N$. Note that this learning rule is non-local because the change of weight depends on the activity of all presynaptic neurons.

In a similar way as in the previous section, we calculate the expectation of the weight vector $\langle \boldsymbol{w}(n) \rangle$, averaged over the sequence of input patterns $(\boldsymbol{\xi}^{\mu_1}, \boldsymbol{\xi}^{\mu_2}, \ldots)$,

$$\langle w_i(n+1) \rangle = \langle w_i(n) \rangle + \gamma \Big(\sum_j C_{ij} \, \langle w_j(n) \rangle - N^{-1} \sum_k \sum_j C_{kj} \, \langle w_j(n) \rangle \Big) \, , \tag{11.23}$$

or explicitly, using matrix notation

$$\langle \boldsymbol{w}(n) \rangle = [1 + \gamma \, (C - \bar{C})]^n \, \langle \boldsymbol{w}(0) \rangle \, , \tag{11.24}$$

with $\bar{C}_{ij} = N^{-1} \sum_k C_{kj}$. The evolution of the weight vector is thus determined by eigenvectors of the matrix $(C - \bar{C})$ that are in general different from those of the correlation matrix C. Hebbian learning with subtractive normalization is driven by the correlations of the input in the subspace orthogonal to the diagonal vector $(1, \ldots, 1)$. Though the sum of the weights stays constant individual weights keep growing. It is thus necessary to adopt an additional criterion to stop the learning

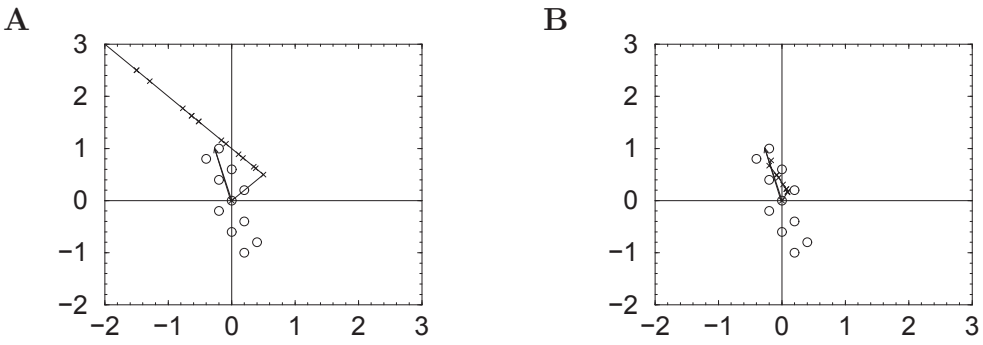

Fig. 11.5: Similar plots as in Fig. 11.3 but with weight vector normalization. **A**. With subtractive normalization, the weight vector evolves along a line that is perpendicular to the diagonal vector $(1, 1)$. Without additional constraints, the length of the weight vector grows without bounds. **B**. Oja's learning rule results in a quick convergence of the weight vector to the first principal component (arrow) of data set.

process and to prevent that some components of the weight vector grow beyond all bounds. A subtractive weight normalization is usually combined with hard boundaries for the weights; cf. Section 10.2.1. With these constraints, the weight vector converges to a final state where (almost) all weights are saturated at the upper or lower bound (Miller and MacKay, 1994); cf. Fig. 11.5A.

Example: Multiplicative Normalization of $\|\boldsymbol{w}\|$

Normalization of the sum of the weights, $\sum_i w_i$, needs an additional criterion to prevent individual weights from perpetual growth. A more elegant way is to require that the sum of the squared weights, i.e., the length of the weight vector, $\sum_i w_i^2$, remains constant. This restricts the evolution of the weight vector to a sphere in the N dimensional weight space. In addition, we can employ a multiplicative normalization scheme where all weights all multiplied by a common factor instead of subtracting a common constant. The advantage of multiplicative compared to subtractive normalization is that small weights will not change their sign during the normalization step.

In order to formalize the above idea we first calculate the 'naïve' weight change $\tilde{\boldsymbol{w}}(n)$ in time step n according to the common Hebbian learning rule,

$$\Delta\tilde{\boldsymbol{w}}(n) = \gamma\,[\boldsymbol{w}(n)\cdot\boldsymbol{\xi}^\mu]\,\boldsymbol{\xi}^\mu\,. \tag{11.25}$$

The update of the weights is accompanied by a normalization of the norm of the weight vector to unity, i.e.,

$$\boldsymbol{w}(n+1) = \frac{\boldsymbol{w}(n) + \Delta\tilde{\boldsymbol{w}}(n)}{\|\boldsymbol{w}(n) + \Delta\tilde{\boldsymbol{w}}(n)\|} \tag{11.26}$$

If we assume that the weights are changed only by a very small amount during each step ($\gamma \ll 1$), we can calculate the new weights $\boldsymbol{w}(n+1)$ to first order in γ,

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \Delta\tilde{\boldsymbol{w}}(n) - \boldsymbol{w}(n)\,[\boldsymbol{w}(n)\cdot\Delta\tilde{\boldsymbol{w}}(n)] + \mathcal{O}(\gamma^2)\,. \tag{11.27}$$

The 'effective' weight change $\Delta\boldsymbol{w}(n)$ including normalization to leading order in γ is thus

$$\Delta\boldsymbol{w}(n) = \Delta\tilde{\boldsymbol{w}}(n) - \boldsymbol{w}(n)\,[\boldsymbol{w}(n)\cdot\Delta\tilde{\boldsymbol{w}}(n)]\,, \tag{11.28}$$

which corresponds to the vector component of $\Delta\tilde{\boldsymbol{w}}$ that is *orthogonal* to the current weight vector $\boldsymbol{w}$. This is exactly what we would have expected because the length of the weight vector must stay constant; cf. Fig. 11.4B.

We may wonder whether Eq. (11.28) is a *local* learning rule. In order to answer this question, we recall that the 'naïve' weight change $\Delta\tilde{w}_j = \gamma\,\nu^{\rm post}\,\nu_j^{\rm pre}$ uses only pre- and postsynaptic information. Hence, we can rewrite Eq. (11.28) in terms of the firing rates,

$$\Delta w_j = \gamma\,\nu^{\rm post}\,\nu_j - \gamma\,w_j(n)\,\left(\nu^{\rm post}\right)^2\,. \tag{11.29}$$

In the second term on the right-hand side we have made use of the linear neuron model, i.e., $\nu^{\rm post} = \sum_k w_k\,\nu_k^{\rm pre}$. Since the weight change depends only on pre- and postsynaptic rates, Eq. (11.29), which is known as Oja's learning rule (Oja, 1982), is indeed local; cf. Eq. (10.11).

In order to see that Oja's learning rule selects the first principal component we show that the eigenvectors $\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_N\}$ of C are fixed points of the dynamics but that only the eigenvector $\boldsymbol{e}_1$ with the largest eigenvalue is stable. For any fixed weight vector $\boldsymbol{w}$ we can calculate the expectation of the weight change in the next time step by averaging over the whole ensemble of input patterns $\{\boldsymbol{\xi}^1,\boldsymbol{\xi}^2,\dots\}$. With $\langle\Delta\tilde{\boldsymbol{w}}(n)\rangle = \gamma\,C\,\boldsymbol{w}$ we find from Eq. (11.28)

$$\langle\Delta\boldsymbol{w}\rangle = \gamma\,C\,\boldsymbol{w} - \gamma\,\boldsymbol{w}\,[\boldsymbol{w}\cdot C\,\boldsymbol{w}]\,, \tag{11.30}$$

We claim that any eigenvector $\boldsymbol{e}_i$ of the correlation matrix C is a fixed point of Eq. (11.30). Indeed, if we substitute $\boldsymbol{w} = \boldsymbol{e}_i$ in the above equation we find that $\langle\Delta\boldsymbol{w}\rangle = 0$. In order to investigate the stability of this fixed point we consider a small perturbation $\boldsymbol{w} = \boldsymbol{e}_i + c\,\boldsymbol{e}_j$ in the direction of $\boldsymbol{e}_j$. Here, $|c| \ll 1$ is the amplitude of the perturbation. If we substitute $\boldsymbol{w} = \boldsymbol{e}_i + c\,\boldsymbol{e}_j$ into Eq. (11.30) we find

$$\langle\Delta\boldsymbol{w}\rangle = c\,\gamma\,(\lambda_j - \lambda_i)\,\boldsymbol{e}_j + \mathcal{O}(c^2)\,. \tag{11.31}$$

The weight vector will thus evolve in the direction of the perturbation $\boldsymbol{e}_j$ if $\lambda_j > \lambda_i$ so that initial perturbation will increase. In this case, $\boldsymbol{e}_i$ is unstable. On the other hand, if $\lambda_j < \lambda_i$ the averaged weight change tends to decrease the perturbation and $\boldsymbol{e}_i$ is stable. Consequently, the eigenvector of C with the largest eigenvalue, viz., the first principle component, is the sole stable fixed point of the dynamics generated by the learning rule of Eq. (11.26). Figure 11.5B shows a simple example.

11.1.4 Receptive Field Development

Most neurons of the visual system respond only to stimulation from a narrow region within the visual field. This region is called the *receptive field* of that neuron. Depending on the precise position of a narrow bright spot within the receptive field the corresponding neuron can either show an increase or a decrease of the firing rate relative to its spontaneous activity at rest. The receptive field is subdivided accordingly into 'ON' and 'OFF' regions in order to further characterize neuronal response properties. Bright spots in an ON region increase the firing rate whereas bright spots in an OFF region inhibit the neuron.

Different neurons have different receptive fields, but as a general rule, neighboring neurons have receptive fields that 'look' at about the same region of the visual field. This is what is usually called the *retinotopic organization* of the neuronal projections – neighboring points in the visual field are mapped to neighboring neurons of the visual system.

The visual system forms a complicated hierarchy of interconnected cortical areas where neurons show increasingly complex response properties from one layer to the next. Neurons from the lateral geniculate nucleus (LGN), which is the first neuronal relay of visual information after the retina, are characterized by so-called center-surround receptive fields. These are receptive fields that consist of two concentric parts, an ON region and an OFF region. LGN neurons come in two flavors, as ON-center and OFF-center cells. ON-center cells have a ON-region in the center of their receptive field that is surrounded by a circular OFF-region. In OFF-center cells the arrangement is the other way round; a central OFF-region is surrounded by an ON-region; cf. Fig. 11.6.

Neurons from the LGN project to the primary visual cortex (V1), which is the first cortical area involved in the processing of visual information. In this area neurons can be divided into 'simple cells' and 'complex cells'. In contrast to LGN neurons, simple cells have *asymmetric* receptive fields which results in a selectivity with respect to the orientation of a visual stimulus. The optimal stimulus for a neuron with a receptive field such as that shown in Fig. 11.6D, for example, is a light bar tilted by about 45 degrees. Any other orientation would also stimulate the OFF region of the receptive field leading to a reduction of the neuronal response. Complex cells have even more intriguing properties and show responses that are, for example, selective for movements with a certain velocity and direction (Hubel, 1995).

It is still a matter of debate how the response properties of simple cells arise. The original proposal by Hubel and Wiesel (1962) was that orientation selectivity is a consequence of the specific wiring between LGN and V1. Several center-surround cells with slightly shifted receptive fields should converge on a single V1 neuron so as to produce the asymmetric receptive field of simple cells. Alternatively (or additionally), the intra-cortical dynamics can generate orienta-

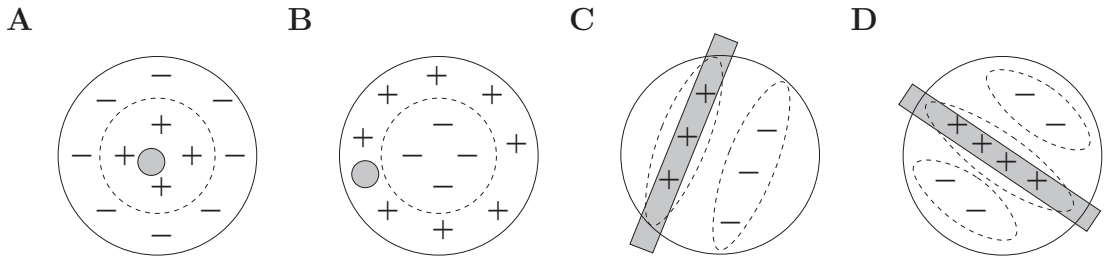

Fig. 11.6: Receptive fields (schematic). **A, B**. Circularly symmetric receptive field as typical for neurons in the LGN. ON-center cells (A) are excited by light spots (gray) falling into the center of the receptive field. In OFF-center cells (B) the arrangement of excitatory and inhibitory regions in the receptive field is reversed. **C, D**. Two examples of asymmetric receptive fields of simple cells in the primary visual cortex. The cells are best stimulated by a light bar oriented as indicated by the grey rectangle.

tion selectivity by enhancing small asymmetries in neuronal responses; cf. Section 9.1.3. In the following, we pursue the first possibility and try to understand how activity-dependent processes during development can lead to the required fine-tuning of the synaptic organization of projections from the LGN to the primary visual cortex (Linsker, 1986c,b,a; Miller et al., 1989; MacKay and Miller, 1990; Miller, 1994, 1995; Wimbauer et al., 1997a,b).

Model architecture

We are studying a model that consists of a two-dimensional layer of cortical neurons (V1 cells) and two layers of LGN neurons, namely one layer of ON-center cells and one layer of OFF-center cells; cf. Fig. 11.7A. In each layer, neurons are labeled by their position and projections between the neurons are given as a function of their positions. Intra-cortical projections, i.e., projections between cortical neurons, are denoted by $w_{\mathrm{V1,V1}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the position of the pre- and the postsynaptic neuron, respectively. Projections from ON-center and OFF-center LGN neurons to the cortex are denoted by $w_{\mathrm{V1,ON}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and $w_{\mathrm{V1,OFF}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, respectively.

In the following we are interested in the evolution of the weight distribution of projections from the LGN to the primary visual cortex. We thus take $w_{\mathrm{V1,ON}}(\boldsymbol{x}, \boldsymbol{x}')$ and $w_{\mathrm{V1,OFF}}(\boldsymbol{x}, \boldsymbol{x}')$ as the dynamic variables of the model. Intra-cortical projections are supposed be constant and dominated by short-range excitation, e.g.,

$$w_{\mathrm{V1,V1}}(\boldsymbol{x}_1, \boldsymbol{x}_2) \propto \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}{\sigma^2_{\mathrm{V1,V1}}}\right) . \tag{11.32}$$

As in the previous section we consider – for the sake of simplicity – neurons

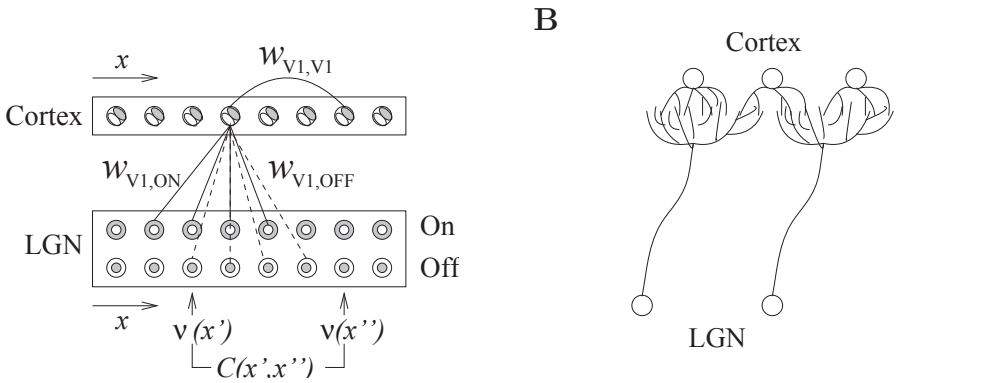

Fig. 11.7: **A**. Wiring diagram between LGN and cortex (schematic). **B**. Axons from LGN cells project only to a small region of cortex. Synaptic contacts are therefore limited to a localized cluster of cortical neurons.

with a linear gain function. The firing rate $\nu_{\rm V1}(\boldsymbol{x})$ of a cortical neuron at position $\boldsymbol{x}$ is thus given by

$$\nu_{\rm V1}(\boldsymbol{x}) = \sum_{\boldsymbol{x}'} w_{\rm V1,ON}(\boldsymbol{x},\boldsymbol{x}')\,\nu_{\rm ON}(\boldsymbol{x}') + \sum_{\boldsymbol{x}'} w_{\rm V1,OFF}(\boldsymbol{x},\boldsymbol{x}')\,\nu_{\rm OFF}(\boldsymbol{x}') + \sum_{\substack{\boldsymbol{x}' \\ (\boldsymbol{x}'\neq\boldsymbol{x})}} w_{\rm V1,V1}(\boldsymbol{x},\boldsymbol{x}')\,\nu_{\rm V1}(\boldsymbol{x}')\,, \quad (11.33)$$

where $\nu_{\rm ON/OFF}(\boldsymbol{x}')$ is the firing rate of a neuron in the ON/OFF layer of the LGN.

Due to the intra-cortical interaction the cortical activity $\nu_{\rm V1}$ shows up on both sides of the equation. Since this is a linear equation it can easily be solved for $\nu_{\rm V1}$. To do so we write $\nu_{\rm V1}(\boldsymbol{x}) = \sum_{\boldsymbol{x}'} \delta_{\boldsymbol{x},\boldsymbol{x}'}\nu_{\rm V1}(\boldsymbol{x}')$, where $\delta_{\boldsymbol{x},\boldsymbol{x}'}$ is the Kronecker δ that is one for $\boldsymbol{x} = \boldsymbol{x}'$ and vanishes otherwise. Equation (11.33) can thus be rewritten as

$$\sum_{\boldsymbol{x}'} [\delta_{\boldsymbol{x},\boldsymbol{x}'} - w_{\rm V1,V1}(\boldsymbol{x},\boldsymbol{x}')]\,\nu_{\rm V1}(\boldsymbol{x}') = \sum_{\boldsymbol{x}'} w_{\rm V1,ON}(\boldsymbol{x},\boldsymbol{x}')\,\nu_{\rm ON}(\boldsymbol{x}') + \sum_{\boldsymbol{x}'} w_{\rm V1,OFF}(\boldsymbol{x},\boldsymbol{x}')\,\nu_{\rm OFF}(\boldsymbol{x}')\,. \quad (11.34)$$

If we read the left-hand side as a multiplication of the matrix $M(\boldsymbol{x},\boldsymbol{x}') \equiv [\delta_{\boldsymbol{x},\boldsymbol{x}'} - w_{\rm V1,V1}(\boldsymbol{x},\boldsymbol{x}')]$ and the vector $\nu_{\rm V1}(\boldsymbol{x}')$ we can define the inverse I of M by

$$\sum_{\boldsymbol{x}} I(\boldsymbol{x}'',\boldsymbol{x})\,M(\boldsymbol{x},\boldsymbol{x}') = \delta_{\boldsymbol{x}'',\boldsymbol{x}'} \quad (11.35)$$

and solve Eq. (11.34) for $\nu_{\rm V1}(\boldsymbol{x}')$. We find

$$\nu_{\rm V1}(\boldsymbol{x}'') = \sum_{\boldsymbol{x}'} \bar{w}_{\rm V1,ON}(\boldsymbol{x}'',\boldsymbol{x}')\,\nu_{\rm ON}(\boldsymbol{x}') + \sum_{\boldsymbol{x}'} \bar{w}_{\rm V1,OFF}(\boldsymbol{x}'',\boldsymbol{x}')\,\nu_{\rm OFF}(\boldsymbol{x}')\,, \quad (11.36)$$

which relates the input $\nu_{\mathrm{ON/OFF}}$ to the output via the 'effective' weights

$$\bar{w}_{\mathrm{V1,ON/OFF}}(\boldsymbol{x}'', \boldsymbol{x}') \equiv \sum_{\boldsymbol{x}} I(\boldsymbol{x}'', \boldsymbol{x})\, w_{\mathrm{V1,ON/OFF}}(\boldsymbol{x}, \boldsymbol{x}')\,. \tag{11.37}$$

Plasticity

We expect that the formation of synapses between LGN and V1 is driven by correlations in the input. In the present case, these correlations are due to the retinotopic organization of projections from the retina to the LGN. Neighboring LGN neurons receiving stimulation from similar regions of the visual field are thus correlated to a higher degree than neurons that are more separated. If we assume that the activity of individual photoreceptors on the retina is uncorrelated and that each LGN neuron integrates the input from many of these receptors then the correlation of two LGN neurons can be calculated from the form of their receptive fields. For center-surround cells the correlation is a Mexican hat-shaped function of their distance (Miller, 1994; Wimbauer et al., 1997a), e.g.,

$$\begin{aligned} C_{\mathrm{ON,ON}}(\boldsymbol{x}, \boldsymbol{x}') &= C_{\mathrm{ON,ON}}(\|\boldsymbol{x} - \boldsymbol{x}'\|) \\ &= \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{\sigma^2}\right) - \frac{1}{c^2}\exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{c^2\,\sigma^2}\right), \end{aligned} \tag{11.38}$$

where c is a form factor that describes the depth of the modulation. $C_{\mathrm{ON,ON}}$ is the correlation between two ON-center type LGN neurons. For the sake of simplicity we assume that OFF-center cells have the same correlation, $C_{\mathrm{OFF,OFF}} = C_{\mathrm{ON,ON}}$. Correlations between ON-center and OFF-center cells, however, have the opposite sign, $C_{\mathrm{ON,OFF}} = C_{\mathrm{OFF,ON}} = -C_{\mathrm{ON,ON}}$.

In the present formulation of the model each LGN cell can contact every neuron in the primary visual cortex. In reality, each LGN cell sends one axon to the cortex. Though this axon may split into several branches its synaptic contacts are restricted to small region of the cortex; cf. Fig. 11.7B. We take this limitation into account by defining an *arborization function* $A(\boldsymbol{x}, \boldsymbol{x}')$ that gives the *a priori* probability that a connection between a LGN cell at location $\boldsymbol{x}$ and a cortical cell at $\boldsymbol{x}'$ is formed (Miller et al., 1989). The arborization is a rapidly decaying function of the distance, e.g.,

$$A(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{\sigma^2_{\mathrm{V1,LGN}}}\right). \tag{11.39}$$

To describe the dynamics of the weight distribution we adopt a modified form of Hebb's learning rule that is completed by the arborization function,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{\mathrm{V1,ON/OFF}}(\boldsymbol{x}, \boldsymbol{x}') = \gamma\, A(\boldsymbol{x}, \boldsymbol{x}')\, \nu_{\mathrm{V1}}(\boldsymbol{x})\, \nu_{\mathrm{ON/OFF}}(\boldsymbol{x}')\,. \tag{11.40}$$

If we use Eq. (11.34) and assume that learning is slow enough so that we can rely on the correlation functions to describe the evolution of the weights, we find

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{\mathrm{V1,ON}}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \gamma\, A(\boldsymbol{x}_1, \boldsymbol{x}_2) \sum_{\boldsymbol{x}', \boldsymbol{x}''} I(\boldsymbol{x}_1, \boldsymbol{x}') \times \left[w_{\mathrm{V1,ON}}(\boldsymbol{x}', \boldsymbol{x}'') - w_{\mathrm{V1,OFF}}(\boldsymbol{x}', \boldsymbol{x}'') \right] C_{\mathrm{ON,ON}}(\boldsymbol{x}'', \boldsymbol{x}_2) \quad (11.41)$$

and a similar equation for $w_{\mathrm{V1,OFF}}$.

Expression (11.41) is still a linear equation for the weights and nothing exciting can be expected. A prerequisite for pattern formation is competition between the synaptic weights. Therefore, the above learning rule is extended by a term $w_{\mathrm{V1,ON/OFF}}(\boldsymbol{x}, \boldsymbol{x}')\, \nu_{\mathrm{V1}}(\boldsymbol{x})^2$ that leads to weight vector normalization and competition; cf. Oja's rule, Eq. (10.11).

Simulation results

Many of the standard techniques for nonlinear systems that we have already encountered in the context of neuronal pattern formation in Chapter 9 can also be applied to the present model (MacKay and Miller, 1990; Wimbauer et al., 1998). Here, however, we will just summarize some results from a computer simulation consisting of an array of 8×8 cortical neurons and two times 20×20 LGN neurons. Figure 11.8 shows a typical outcome of such a simulation. Each of the small rectangles shows the receptive field of the corresponding cortical neuron. A bright color means that the neuron responds with an increased firing rate to a bright spot at that particular position within its receptive field; dark colors indicate inhibition.

There are two interesting aspects. First, the evolution of the synaptic weights has lead to asymmetric receptive fields, which give rise to orientation selectivity. Second, the structure of the receptive fields of neighboring cortical neurons are similar; neuronal response properties thus vary continuously across the cortex. The neurons are said to form a *map* for, e.g., orientation.

The first observation, the breaking of the symmetry of LGN receptive fields, is characteristic for all pattern formation phenomena. It results from the instability of the homogeneous initial state and the competition between individual synaptic weights. The second observation, the smooth variation of the receptive fields across the cortex, is a consequence of the excitatory intra-cortical couplings. During the development, neighboring cortical neurons tend to be either simultaneously active or quiescent and due to the activity dependent learning rule similar receptive fields are formed.

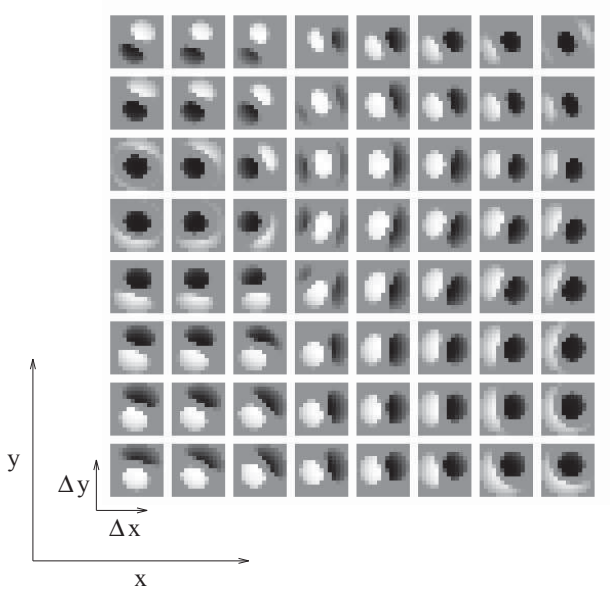

Fig. 11.8: Receptive fields (small squares) of 64 cortical neurons (large grid). Each small square shows the distribution of weights $w_{\mathrm{V1,ON}}(\boldsymbol{x}, \boldsymbol{x} + \Delta\boldsymbol{x}) - w_{\mathrm{V1,OFF}}(\boldsymbol{x}, \boldsymbol{x} + \Delta\boldsymbol{x})$, where $\boldsymbol{x}$ is the position of the cortical neuron and Δx the position of the white or black spot within the small rectangle [adapted from Wimbauer et al. (1998)].

11.2 Learning in Spiking Models

In the previous section we have seen that the evolution of synaptic weights under a rate-based learning rule depends on correlations in the input. What happens, if the rate-based learning rule is replaced by a spike-time dependent one?

In Section 11.2.1 we will derive an equation that relates the expectation value of the weight vector to statistical properties of pre- and postsynaptic spike trains. We will see that spike-time dependent plasticity is sensitive to spatial and *temporal* correlations in the input. In certain particularly simple cases spike-spike correlations can be calculated explicitly. This is demonstrated in Section 11.2.2 in the context of a linear Poisson neuron. This neuron model is also used in Section 11.2.3 for a comparison of spike-based and rate-based learning rules as well as in Section 11.2.4 where we revisit the static-pattern scenario of Section 11.1.2. Finally, in Section 11.2.5, we discuss the impact of stochastic spike arrival on the synaptic weights and derive a Fokker-Planck equation that describes the temporal evolution of the weight distribution.

11.2.1 Learning Equation

We will generalize the analysis of Hebbian learning that has been developed in Section 11.1 to spike-based learning rules based on the phenomenological model of Section 10.3.1. In this model the synaptic weight $w_{ij}(t)$ is a piecewise continuous function of time with steps whenever a presynaptic spike arrives or when a postsynaptic action potential is triggered, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = a_0 + a_1^{\rm pre}\, S_j(t) + a_1^{\rm post}\, S_i(t) + S_j(t) \int_0^\infty W(s)\, S_i(t-s)\, \mathrm{d}s + S_i(t) \int_0^\infty W(-s)\, S_j(t-s)\, \mathrm{d}s\,, \quad (11.42)$$

cf. Eqs. (10.14)–(10.15). As before we want to relate the synaptic weight change to the statistical properties of the input. Given the increased level of complexity, a few remarks about the underlying statistical ensemble are in order.

In the previous section we have considered presynaptic firing rates ν_j as random variables drawn from an ensemble of input patterns ξ_j^μ. The output rate, however, was a deterministic function of the neuronal input. In the context of spike-time dependent plasticity, we consider the set of presynaptic spike arrival times $(t_j^1, t_j^2, \dots)$ as a random variable. The underlying 'randomness' may have several reasons. For example, different stimulation paradigms may be selected one by one in very much the same way as we have selected a new input pattern in the previous section. In contrast to the rate model, we do not want to restrict ourselves to deterministic neuron models. Hence, the randomness can also be produced by a stochastic neuron model that is used in order account for noise;

cf. Chapter 5. In this case, the output spike train can be a random variable even if the input spike trains are fixed. A simple example is the Poisson neuron model that generates output spikes via an inhomogeneous Poisson process with an intensity that is a function of the membrane potential. In any case, we consider the set of spike trains $(S_1, \ldots, S_i, S_j, \ldots, S_N)$, i.e., pre- *and* postsynaptic trains, to be drawn from a stochastic ensemble. The specific properties of the chosen neuron model are thus implicitly described by the association of pre- and postsynaptic trains within the ensemble. Note that this formalism includes deterministic models as a special case, if the ensemble contains only a single postsynaptic spike train for any given set of presynaptic spike trains. In the following, all averages denoted by $\langle\cdot\rangle_E$ are to be taken relative to this ensemble.

For the time being we are interested only in the long-term behavior of the synaptic weights and not in the fluctuations that are caused by individual spikes. As in Section 11.1.2 we therefore calculate the expectation value of the weight change over a certain interval of time,

$$\langle w_{ij}(t+T) - w_{ij}(t)\rangle_E = \left\langle \int_t^{t+T} \frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t')\mathrm{d}t' \right\rangle_E . \tag{11.43}$$

With the abbreviation

$$\langle f(t)\rangle_T \equiv T^{-1} \int_t^{t+T} f(t')\,\mathrm{d}t' \tag{11.44}$$

we obtain from Eq. (11.42)

$$\begin{aligned} \frac{\langle w_{ij}(t+T) - w_{ij}(t)\rangle_E}{T} &= a_0 + a_1^{\mathrm{pre}} \left\langle \langle S_j(t)\rangle_T \right\rangle_E + a_1^{\mathrm{post}} \left\langle \langle S_i(t)\rangle_T \right\rangle_E \\ &\quad + \int_0^\infty W(s) \left\langle \langle S_i(t-s)\, S_j(t)\rangle_T \right\rangle_E \mathrm{d}s \\ &\quad + \int_{-\infty}^0 W(s) \left\langle \langle S_i(t)\, S_j(t+s)\rangle_T \right\rangle_E \mathrm{d}s\,. \end{aligned} \tag{11.45}$$

If the time interval T is long as compared to typical interspike intervals then the time average is taken over many pre- or postsynaptic spikes. We can thus assume that the average $\left\langle \langle S_i(t)\, S_j(t+s)\rangle_T \right\rangle_E$ does not change if we replace t by $t-s$ as long as $s \ll T$. Furthermore, if $W(s)$ decays to zero sufficiently fast as $|s| \to \infty$ then the integration over s in the last term of Eq. (11.45) can be restricted to a finite interval determined by the width of the learning window W . In this case it is possible to replace $\left\langle \langle S_i(t)\, S_j(t+s)\rangle_T \right\rangle_E$ by $\left\langle \langle S_i(t-s)\, S_j(t)\rangle_T \right\rangle_E$ and to collect the last two terms of Eq. (11.45) into a single integral, provided that the width

of learning window is small as compared to T. With this approximation we find

$$\frac{\langle w_{ij}(t+T) - w_{ij}(t)\rangle_E}{T} = a_0 + a_1^{\text{pre}} \left\langle \left\langle S_j(t)\right\rangle_T \right\rangle_E + a_1^{\text{post}} \left\langle \left\langle S_i(t)\right\rangle_T \right\rangle_E + \int_{-\infty}^{\infty} W(s) \left\langle \left\langle S_i(t-s)\, S_j(t)\right\rangle_T \right\rangle_E \mathrm{d}s\,. \tag{11.46}$$

The instantaneous firing rate $\nu_i(t)$ of neuron i is the ensemble average of its spike train,

$$\nu_i(t) \equiv \left\langle S_i(t)\right\rangle_E . \tag{11.47}$$

Similarly, we define the joint firing rate ν_{ij} of neuron i and j as

$$\nu_{ij}(t,t') \equiv \left\langle S_i(t)\, S_j(t')\right\rangle_E , \tag{11.48}$$

which is the joint probability density to find both a spike at time t and at time t' in neuron i and j, respectively. Note that $\nu_{ij}(t,t')$ is a probability density both in t and t' and thus has units of one over time squared.

Since averaging is a linear operations we can exchange ensemble average and time average. We obtain the following expression for the expected weight change in the interval from t to $t+T$ as a function of the statistical properties of the spike trains,

$$\frac{\langle w_{ij}(t+T) - w_{ij}(t)\rangle_E}{T} = a_0 + a_1^{\text{pre}} \left\langle \nu_j(t)\right\rangle_T + a_1^{\text{post}} \left\langle \nu_i(t)\right\rangle_T + \int_{-\infty}^{\infty} W(s) \left\langle \nu_{ij}(t-s,t)\right\rangle_T \mathrm{d}s\,. \tag{11.49}$$

The time average $\langle \nu_{ij}(t-s,t)\rangle_T$ is the correlation function of pre- and post-synaptic spike train on the interval $[t, t+T]$. This function clearly depends on the actual value of the weight vector. In deriving Eq. (11.49) we already had to assume that the correlations are a slowly varying function of time. For the sake of consistency we thus have the requirement that the weight vector itself is a slowly varying function of time. If this is the case then we can exploit the self-averaging property of the weight vector and argue that fluctuations around the expectation value are negligible and that Eq. (11.49) is a good approximation for the actual value of the weight vector. We thus drop the ensemble average on the left-hand side of Eq. (11.49) and find for the time-averaged change of the synaptic weight the following learning equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle w_{ij}(t)\right\rangle_T = a_0 + a_1^{\text{pre}} \left\langle \nu_j(t)\right\rangle_T + a_1^{\text{post}} \left\langle \nu_i(t)\right\rangle_T + \int_{-\infty}^{\infty} W(s) \left\langle \nu_{ij}(t-s,t)\right\rangle_T \mathrm{d}s\,; \tag{11.50}$$

cf. (Kempter et al., 1999; Kistler and van Hemmen, 2000a). As expected, the long-term dynamics of the synaptic weights depends on the correlation of pre- and postsynaptic spike train on the time scale of the learning window. In the following we will always use the smooth time-averaged weight vector $\langle w_{ij}(t)\rangle_T$, but for the sake of brevity we shall drop the angular brackets.

11.2.2 Spike-Spike Correlations

It is tempting to rewrite the correlation term $\langle \nu_{ij}(t-s,t)\rangle_T$ that appears on the right-hand side of Eq. (11.50) in terms of the instantaneous firing rates $\langle \nu_i(t-s)\,\nu_j(t)\rangle_T$. This, however, is only allowed, if the spike trains of neuron i and j were independent, i.e., if $\langle S_i(t-s)\,S_j(t)\rangle_E = \langle S_i(t-s)\rangle_E\,\langle S_j(t)\rangle_E$. Such an approach would therefore neglect the specific spike-spike correlations that are induced by presynaptic action potentials.

Correlations between pre- and postsynaptic spike trains do not only depend on the input statistics but also on the dynamics of the neuron model and the way new output spikes are generated. The influence of a single presynaptic spike on the postsynaptic activity can be measured by a peri-stimulus time histogram (PSTH) triggered on the time of presynaptic spike arrival; cf. Section 7.4.1. The form of the PSTH characterizes the spike-spike correlations between presynaptic spike arrival and postsynaptic action potential. For high noise, the spike-spike correlations contain a term that is proportional to the time-course of the postsynaptic potential ϵ, while for low noise this term is proportional to its derivative ϵ'; cf. Figs. 7.12.

In the following, we will calculate the spike-spike correlations in a particularly simple case, the linear Poisson neuron model. As we will see, the spike-spike correlations contain in this case a term proportional to the postsynaptic potential ϵ. The linear Poisson neuron model can therefore be considered as a reasonable approximation to spiking neuron models in the high-noise limit.

Example: Linear Poisson neuron model

As a generalization of the analog neuron with linear gain function discussed in Section 11.1.2 we consider here a linear Poisson neuron. The input to the neuron consists of N Poisson spike trains with time-dependent intensities $\nu_j(t)$. Similar to the SRM_0 neuron the membrane potential u_i of neuron i is a superposition of postsynaptic potentials ϵ with $\int_0^\infty \epsilon(s)\,\mathrm{d}s = 1$,

$$u_i(t) = \sum_j w_{ij} \int_0^\infty \epsilon(s)\,S_j(t-s)\,\mathrm{d}s\,. \tag{11.51}$$

In contrast to Section 4.2.3 we neglect refractoriness and external input.

Postsynaptic spikes are generated by an inhomogeneous Poisson process with an intensity $\nu_i^{\rm post}$ that is a (semi-)*linear* function of the membrane potential,

$$\nu_i^{\rm post}(t|u) = [u_i(t)]_+ \,. \tag{11.52}$$

Here, $[\cdot]_+$ denotes the positive part of the argument in order to avoid negative rates. In the following, however, we will always assume that $u_i(t) \geq 0$. The notation $\nu_i^{\rm post}(t|u)$ indicates that the output rate depends on the actual value of the membrane potential.

We thus have a doubly stochastic process (Cox, 1955; Bartlett, 1963) in the sense that in a first step, a set of input spike trains is drawn from an ensemble characterized by Poisson rates $\nu_j^{\rm pre}$. This realization of input spike trains then determines the membrane potential which produces in a second step a specific realization of the output spike train according to $\nu_i^{\rm post}(t|u)$. It can be shown that, because of the finite duration of the postsynaptic potential ϵ, the output spike trains generated by this composite process are no longer Poisson spike trains; their expectation value $\langle S_i(t)\rangle_E \equiv \nu_i^{\rm post}(t)$, however, is simply equivalent to the expectation value of the output rate, $\nu_i^{\rm post}(t) = \langle \nu_i^{\rm post}(t|u)\rangle_E$ (Kistler and van Hemmen, 2000a). Due to the linearity of the neuron model the output rate is given by a convolution of the input rates with the response kernel ϵ,

$$\nu_i^{\rm post}(t) = \sum_j w_{ij} \int_0^\infty \epsilon(s)\, \nu_j^{\rm pre}(t-s)\, {\rm d}s \,. \tag{11.53}$$

The joint firing rate $\nu_{ij}^{\rm post,pre}(t,t') = \langle S_i(t)\, S_j(t')\rangle_E$ of pre- and postsynaptic neuron is the joint probability density to find an input spike at synapse j at time t' and an output spike of neuron i at time t. According to Bayes' Theorem this probability equals the probability of observing an input spike at time t' times the conditional probability of observing an output spike at time t given the input spike at time t', i.e.,

$$\nu_{ij}^{\rm post,pre}(t,t') = \langle S_i(t)|\text{input spike at } t'\rangle_E \, \langle S_j(t')\rangle_E \,. \tag{11.54}$$

In the framework of a *linear* Poisson neuron, the term $\langle S_i(t)|\text{input spike at } t'\rangle_E$ equals the sum of the expected output rate (11.53) and the specific contribution $w_{ij}\,\epsilon(t-t')$ of a single (additional) input spike at time t'. Altogether we obtain

$$\nu_{ij}^{\rm post,pre}(t,t') = \nu_i^{\rm post}(t)\, \nu_j^{\rm pre}(t') + w_{ij}\, \epsilon(t-t')\, \nu_j^{\rm pre}(t') \,. \tag{11.55}$$

The first term on the right-hand side is the 'chance level' to find two spikes at t and t', respectively, if the neurons were firing independently at rates $\nu_i^{\rm post}(t)$ and $\nu_j^{\rm pre}(t')$. The second term describes the correlation that is due to synaptic coupling. If the presynaptic neuron has fired a spike at t' then the chance for the postsynaptic neuron to fire an spike at time $t > t'$ is increased by $w_{ij}\,\epsilon(t-t')$. Note that this expression respects causality: The probability to find first a postsynaptic spike and then a presynaptic spike is just chance level because $\epsilon(t-t') = 0$ for $t < t'$.

Example: Learning equation for a linear Poisson neuron

If we use the result from Eq. (11.55) in the learning equation (11.50) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}w_{ij}(t) = a_0 + a_1^{\mathrm{pre}}\,\langle\nu_j^{\mathrm{pre}}(t)\rangle_T + a_1^{\mathrm{post}}\,\langle\nu_i^{\mathrm{post}}(t)\rangle_T + \int_{-\infty}^{\infty} W(s)\,\langle\nu_i^{\mathrm{post}}(t-s)\,\nu_j^{\mathrm{pre}}(t)\rangle_T\,\mathrm{d}s + w_{ij}(t)\,\langle\nu_j^{\mathrm{pre}}(t)\rangle_T\,W_-\,, \quad (11.56)$$

with $W_- = \int_0^\infty W(-s)\,\epsilon(s)\mathrm{d}s$.

In linear Poisson neurons, the correlation between pre- and postsynaptic activity that drives synaptic weight changes consists of two contributions. The integral over the learning window in Eq. (11.56) describes correlations in the instantaneous firing rate. The last term on the right-hand side of Eq. (11.56) finally accounts for spike-spike correlations of pre- and postsynaptic neuron.

If we express the instantaneous firing rates $\nu_i(t)$ in terms of their fluctuations $\Delta\nu_i(t)$ around the mean $\langle\nu_i(t)\rangle_T$,

$$\nu_i(t) = \Delta\nu_i(t) + \langle\nu_i(t)\rangle_T\,, \quad (11.57)$$

then we can rewrite Eq. (11.56) together with Eq. (11.53) as

$$\frac{\mathrm{d}}{\mathrm{d}t}w_{ij}(t) = a_0 + a_1^{\mathrm{pre}}\,\langle\nu_j^{\mathrm{pre}}(t)\rangle_T + a_1^{\mathrm{post}}\,\langle\nu_i^{\mathrm{post}}(t)\rangle_T + \bar{W}\langle\nu_j^{\mathrm{pre}}(t)\rangle_T\,\langle\nu_i^{\mathrm{post}}(t)\rangle_T + \sum_k w_{ik}\,Q_{kj}(t) + w_{ij}(t)\,\langle\nu_j^{\mathrm{pre}}(t)\rangle_T\,W_- \quad (11.58)$$

with

$$Q_{kj}(t) = \int_{-\infty}^{\infty} W(s)\int_0^\infty \epsilon(s')\,\langle\Delta\nu_k^{\mathrm{pre}}(t-s-s')\,\Delta\nu_j^{\mathrm{pre}}(t)\rangle_T\,\mathrm{d}s'\,\mathrm{d}s\,. \quad (11.59)$$

Here we have implicitly assumed that the temporal averaging interval T is much longer than the length of the learning window, the duration of a postsynaptic potential, or a typical interspike interval, so that $\langle\nu_i^{\mathrm{post}}(t-s)\rangle_T \approx \langle\nu_i^{\mathrm{post}}(t)\rangle_T$ and $\langle\nu_j^{\mathrm{pre}}(t-s')\rangle_T \approx \langle\nu_j^{\mathrm{pre}}(t)\rangle_T$.

The term containing $Q_{kj}(t)$ on the right-hand side of Eq. (11.58) shows how spatio-temporal correlations $\langle\Delta\nu_k^{\mathrm{post}}(t')\,\Delta\nu_j^{\mathrm{pre}}(t)\rangle_T$ in the input influence the evolution of synaptic weights. What matters are correlations on the time scale of the learning window and the postsynaptic potential.

11.2.3 Relation of spike-based to rate-based learning

In Section 11.1.2 we have investigated the weight dynamics in the context of an analog neuron where the postsynaptic firing rate is an instantaneous function of

the input rates. We have seen that learning is driven by (spatial) correlations within the set of input patterns. The learning equation (11.56) goes one step further in the sense that it explicitly includes time. Consequently, learning is driven by *spatio-temporal* correlations in the input.

In order to compare the rate-based learning paradigm of Section 11.1.2 with the spike-based formulation of Eq. (11.56) we thus have to disregard temporal correlations for the time being. We thus consider a linear Poisson neuron with *stationary* input rates, $\langle \nu_j(t) \rangle_T = \nu_j(t) = \nu_j$, and assume that the synaptic weight is changing slowly as compared to the width of the learning window and the postsynaptic potential. The weight dynamics is given by Eq. (11.56),

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = a_0 + a_1^{\rm pre}\, \nu_j + a_1^{\rm post}\, \nu_i + \bar{W}\, \nu_i\, \nu_j \; + W_-\, w_{ij}(t)\, \nu_j \;, \tag{11.60}$$

with $\bar{W} = \int_{-\infty}^{\infty} W(s)\, \mathrm{d}s$ and $W_- = \int_0^{\infty} W(-s)\, \epsilon(s) \mathrm{d}s$. If we identify

$$c_0(w_{ij}) = a_0 \,, \quad c_1^{\rm pre}(w_{ij}) = a_1^{\rm pre} + w_{ij}(t)\, W_- \,, \quad c_1^{\rm post}(w_{ij}) = a_1^{\rm post} \,, \tag{11.61}$$

and

$$c_2^{\rm corr}(w_{ij}) = \bar{W} \,, \tag{11.62}$$

we recover the general expression for synaptic plasticity based on the rate description given in Eq. (10.2). The total area under the learning window thus plays the role of the correlation parameter $c_2^{\rm corr}$ that is responsible for Hebbian or anti-Hebbian plasticity in a rate formulation. The spike-spike correlations simply give rise to an additional weight-dependent term $w_{ij}(t)\, W_-$ in the parameter $c_1^{\rm pre}(w_{ij})$ that describes presynaptically triggered weight changes.

We may wonder what happens if we relax the requirement of strictly stationary rates. In the linear Poisson model, the output rate depends via Eq. (11.53) on the input rates and changes in the input rate translate into changes in the output rate. If the rate of change is small, we can expand the output rate

$$\nu_i^{\rm post}(t-s) \approx \nu_i^{\rm post}(t) - s\, \frac{\mathrm{d}}{\mathrm{d}t} \nu_i^{\rm post}(t) + \mathcal{O}(s^2) \tag{11.63}$$

on the right-hand side of Eq. (11.56),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) \;\; &= \;\; a_0 + a_1^{\rm pre}\, \nu_j^{\rm pre}(t) + a_1^{\rm post}\, \nu_i^{\rm post}(t) + \bar{W}\, \nu_i^{\rm post}(t)\, \nu_j^{\rm pre}(t) \\ &\quad + W_-\, w_{ij}(t)\, \nu_j^{\rm pre}(t) \; - \nu_j^{\rm pre}(t)\, \frac{\mathrm{d}}{\mathrm{d}t} \nu_i^{\rm post}(t) \int_{-\infty}^{\infty} s\, W(s)\, \mathrm{d}s \,. \end{aligned} \tag{11.64}$$

Here, we have dropped the temporal averages because rates are assumed to change slowly relative to T.

As compared to Eq. (11.60) we encounter an additional term that is proportional to the first moment $\int s\, W(s)\, \mathrm{d}s$ of the learning window. This term has been termed *differential*-Hebbian (Roberts, 1999; Xie and Seung, 2000) and plays a certain role in the context of conditioning and reinforcement learning (Montague et al., 1995; Rao and Sejnowski, 2001).

Stabilization of Postsynaptic Rates

Another interesting property of a learning rule of the form (10.2) or (11.60) is that it can lead to a normalization of the postsynaptic firing rate and hence to a normalization of the sum of the synaptic weights. This can be achieved even without including higher order terms in the learning equation or postulating a dependence of the parameters a_0, $a_1^{\mathrm{pre/post}}$, etc., on the actual value of the synaptic efficacy.

Consider a linear Poisson neuron that receives input from N presynaptic neurons with spike activity described by independent Poisson processes with rate ν^{pre}. The postsynaptic neuron is thus firing at a rate $\nu_i^{\mathrm{post}}(t) = \nu^{\mathrm{pre}} \sum_{j=1}^{N} w_{ij}(t)$. From Eq. (11.56) we obtain the corresponding dynamics for the synaptic weights, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = a_0 + a_1^{\mathrm{pre}}\, \nu^{\mathrm{pre}} + a_1^{\mathrm{post}}\, \nu^{\mathrm{pre}} \sum_{k=1}^{N} w_{ik}(t) + (\nu^{\mathrm{pre}})^2\, \bar{W} \sum_{k=1}^{N} w_{ik}(t) + w_{ij}(t)\, \nu^{\mathrm{pre}}\, W_- \,, \tag{11.65}$$

with $\bar{W} = \int_{-\infty}^{\infty} W(s)\, \mathrm{d}s$ and $W_- = \int_0^{\infty} \epsilon(s)\, W(-s)\, \mathrm{d}s$. In this particularly simple case the weight dynamics is characterized by a fixed point for the sum of the synaptic weights, $\sum_j w_{ij}$, and, hence, for the postsynaptic firing rate, $\nu_i^{\mathrm{post}} = \nu_{\mathrm{FP}}$,

$$\nu_{\mathrm{FP}} = -\frac{a_0 + a_1^{\mathrm{pre}}\, \nu^{\mathrm{pre}}}{a_1^{\mathrm{post}} + \nu^{\mathrm{pre}}\, \bar{W} + N^{-1}\, W_-} \,. \tag{11.66}$$

This fixed point is attractive if the denominator is negative. Since ν_i^{post} is a firing rate we have the additional requirement that $\nu_{\mathrm{FP}} \geq 0$. Altogether we thus have two conditions for the parameters of the learning rule, i.e., $a_1^{\mathrm{post}} + \nu^{\mathrm{pre}}\, \bar{W} + N^{-1}\, W_- < 0$ and $a_0 + a_1^{\mathrm{pre}}\, \nu^{\mathrm{pre}} \geq 0$. Note that we would obtain a – apart from the term $(N^{-1}\, W_-)$ – completely analogous result from the rate formulation in Eq. (10.2) if we identify $c_2^{\mathrm{corr}} = \bar{W}$; cf. Eq. (11.62). Note further, that the linearity is not essential for the stabilization of the postsynaptic rate. Any model where the output rate is a monotonous function of the sum of the synaptic weights yields qualitatively the same result.

11.2.4 Static-Pattern Scenario

In order to illustrate the above results with a concrete example we revisit the static-pattern scenario that we have already studied in the context of analog neurons in Section 11.1.2. We consider a set of static patterns $\{\boldsymbol{\xi}^\mu; 1 < \mu < p\}$ that are presented to the network in a random sequence $(\mu_1, \mu_2, \dots)$ during time steps of length Δt. Presynaptic spike trains are described by an inhomogeneous Poisson process with a firing intensity that is determined by the pattern that is currently presented. Hence, the instantaneous presynaptic firing rates are piecewise constant functions of time,

$$\nu_j^{\text{pre}}(t) = \sum_k \xi_j^{\mu_k} \, \Theta[t - (k-1)\,\Delta t] \, \Theta[k\,\Delta t - t] \,. \tag{11.67}$$

Due to the randomness by which the patterns are presented the input does not contain any no non-trivial temporal correlations. We thus expect to obtain the very same result as in Section 11.1.2, i.e., that the evolution of synaptic weights is determined by the correlation matrix of the input pattern set.

For linear Poisson neurons the joint firing rate of pre- and postsynaptic neuron is given by Eq. (11.55),

$$\nu_{ij}(t-s,t) = \nu_i^{\text{post}}(t-s)\,\nu_j^{\text{pre}}(t) + w_{ij}(t)\,\epsilon(-s)\,\nu_j^{\text{pre}}(t) \,. \tag{11.68}$$

The postsynaptic firing rate is

$$\begin{aligned} \nu_i^{\text{post}}(t) &= \sum_j \int_0^\infty w_{ij}(t-s)\,\epsilon(s)\,\nu_j^{\text{pre}}(t-s)\,\mathrm{d}s \\ &\approx \sum_j w_{ij}(t) \int_0^\infty \epsilon(s)\,\nu_j^{\text{pre}}(t-s)\,\mathrm{d}s \,, \end{aligned} \tag{11.69}$$

where we have assumed implicitly that the synaptic weights are approximately constant on the time scale defined by the duration of the postsynaptic potential ϵ so that we can pull w_{ij} in front of the integral.

As usual, we are interested in the long-term behavior of the synaptic weights given by Eq. (11.56). We thus need the time-average of $\nu_i(t-s)\,\nu_j(t)$ over the interval T,

$$\langle \nu_i^{\text{post}}(t-s)\,\nu_j^{\text{pre}}(t)\rangle_T = \sum_k w_{ik}(t) \int_0^\infty \epsilon(s') \left\langle \nu_k^{\text{pre}}(t-s-s')\,\nu_j^{\text{pre}}(t)\right\rangle_T \mathrm{d}s' \,. \tag{11.70}$$

Due to the linearity of the neuron model, the correlation of input and output is a linear combination of the correlations $\langle \nu_k^{\text{pre}}(t-s)\,\nu_j^{\text{pre}}(t)\rangle_T$ in the input firing rates,

which are independent from the specific neuron model. We assume that all patterns are presented once during the time interval T that defines the time scale on which we are investigating the weight dynamics. For $s = 0$ the time average corresponds to an ensemble average over the input patterns and the input correlation functions equals the correlation of the input pattern, $\langle \nu_k^{\rm pre}(t)\, \nu_j^{\rm pre}(t)\rangle_T = \langle \xi_k^\mu\, \xi_j^\mu\rangle_\mu$. Here, $\langle\cdot\rangle_\mu$ denotes an ensemble average over the set of input patterns. Since we have assumed that input patterns are presented randomly for time steps of length Δt the correlation $\langle \nu_k^{\rm pre}(t-s)\, \nu_j^{\rm pre}(t)\rangle_T$ will be computed from two independent input patterns if $|s| > \Delta t$, i.e., $\langle \nu_k^{\rm pre}(t-s)\, \nu_j^{\rm pre}(t)\rangle_T = \langle \xi_k^\mu\rangle_\mu\, \langle \xi_j^\mu\rangle_\mu$. For $0 < s < \Delta t$ the input correlation is a linear function of s. Altogether we obtain

$$\left\langle \nu_k^{\rm pre}(t-s)\, \nu_j^{\rm pre}(t)\right\rangle_T = \langle \xi_k^\mu\rangle_\mu\, \langle \xi_j^\mu\rangle_\mu + \left(\langle \xi_k^\mu\, \xi_j^\mu\rangle_\mu - \langle \xi_k^\mu\rangle_\mu\, \langle \xi_j^\mu\rangle_\mu\right)\, \Lambda(s/\Delta t)\,. \quad (11.71)$$

Here, Λ is the triangular function

$$\Lambda(s) = (1-|s|)\,\Theta(1-|s|)\,; \quad (11.72)$$

cf. Fig. 11.9A. If we use this result in the learning equation (11.56) we find

$$\frac{\rm d}{{\rm d}t}\bar{w}_{ij}(t) = \sum_k w_{ik}(t)\, \langle \xi_k^\mu\rangle_\mu\, \langle \xi_j^\mu\rangle_\mu\, \bar{W} + \sum_k w_{ik}(t)\, Q_{kj} + w_{ij}(t)\, \langle \xi_j^\mu\rangle_\mu\, W_-\,, \quad (11.73)$$

with $\bar{W} = \int_{-\infty}^{\infty} W(s)\,{\rm d}s$, $W_- = \int_0^\infty \epsilon(s)\, W(-s)\,{\rm d}s$, and

$$Q_{kj} = \left(\langle \xi_k^\mu\, \xi_j^\mu\rangle_\mu - \langle \xi_k^\mu\rangle_\mu\, \langle \xi_j^\mu\rangle_\mu\right) \int_{-\infty}^{\infty} W(s) \int_0^\infty \epsilon(s')\, \Lambda\left(\frac{s+s'}{\Delta t}\right)\, {\rm d}s'\, {\rm d}s\,. \quad (11.74)$$

Here we have used $\int_0^\infty \epsilon(s){\rm d}s = 1$ and dropped all non-Hebbian terms ($a_0 = a_1^{\rm pre} = a_1^{\rm post} = 0$).

In order to understand this result let us first consider the case where both the width of the learning window and the postsynaptic potential is small as compared to the duration Δt of one pattern presentation. The integral over s' in the definition of the matrix Q_{kj} is the convolution of ϵ with a triangular function centered around $s = 0$ that has a maximum value of unity. Since ϵ is normalized, the convolution yields a smoothed version of the originally triangular function that is approximately equal to unity in a neighborhood of $s = 0$; cf. Fig. 11.9B. If the learning window is different from zero only in this neighborhood, then the integral over s in Eq. (11.74) is just $\bar{W}$, the area under the learning window. We can thus collect the first two terms on the right-hand side of Eq. (11.73) and obtain

$$\frac{\rm d}{{\rm d}t} w_{ij}(t) = \sum_k w_{ik}(t)\, \langle \xi_k^\mu\, \xi_j^\mu\rangle_\mu \bar{W} + w_{ij}(t)\, \langle \xi_j^\mu\rangle_\mu\, W_-\,. \quad (11.75)$$

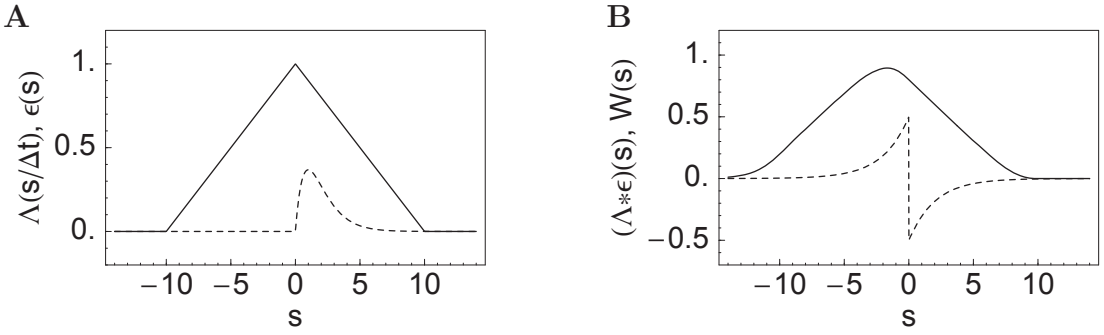

Fig. 11.9: Static-pattern scenario. **A**. Temporal correlations in the firing rate of presynaptic neurons have a triangular shape $\Lambda(s/\Delta t)$ (solid line). The correlation between pre- and postsynaptic neurons involves a convolution with the response kernel $\epsilon(s)$ (dashed line). **B**. The definition of the matrix Q_{kj} in Eq. (11.74) contains the overlap of the learning window $W(s)$ (dashed line) and the convolution $\int \epsilon(s')\,\Lambda[(s-s')/\Delta t]\,\mathrm{d}s'$ (solid line). If the duration of one presentation is long as compared to the width of the learning window and the response kernel ϵ the the overlap equals approximately the area below the learning window $\bar{W}$. If the presentation is short, as shown here, then the overlap may be different from zero, even if $\bar{W} = 0$.

Apart from the non-Hebbian term $w_{ij}(t)\,\langle \xi_j^\mu \rangle_\mu\, W_-$ the weight dynamics is determined by the correlation matrix $\langle \xi_k^\mu\, \xi_j^\mu \rangle_\mu$ of the (unnormalized) input patterns. This is exactly what we would have expected from the comparison of rate-based and spike based learning; cf. Eq. (11.60).

More interesting is the case where the time scale of the learning window is of the same order of magnitude as the presentation of an input pattern. In this case, the integral over s in Eq. (11.74) is different from $\bar{W}$ and we can choose a time window with $\bar{W} = 0$ so that the first term on the right-hand side of Eq. (11.73) vanishes. In this case, the weight dynamics is no longer determined by $\langle \xi_k^\mu\, \xi_j^\mu \rangle_\mu$ but by the matrix Q_{jk},

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = \sum_k w_{ik}(t)\, Q_{kj} + w_{ij}(t)\,\langle \xi_j^\mu \rangle_\mu\, W_- \,, \tag{11.76}$$

which is proportional to the *properly normalized* covariance matrix of the input patterns,

$$Q_{kj} \propto \langle \xi_k^\mu\, \xi_j^\mu \rangle_\mu - \langle \xi_k^\mu \rangle_\mu\, \langle \xi_j^\mu \rangle_\mu = \langle (\xi_k^\mu - \langle \xi_k^\mu \rangle_\mu)\,(\xi_j^\mu - \langle \xi_j^\mu \rangle_\mu) \rangle_\mu \,. \tag{11.77}$$

If we assume that all presynaptic neurons have the same mean activity, $\langle \xi_k^\mu \rangle_\mu = \langle \xi_j^\mu \rangle_\mu \equiv \langle \xi^\mu \rangle_\mu$ then we can rewrite Eq. (11.76) as

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = \sum_k w_{ik}(t)\,[Q_{kj} + \delta_{kj}\,\langle \xi^\mu \rangle_\mu\, W_-] \,. \tag{11.78}$$

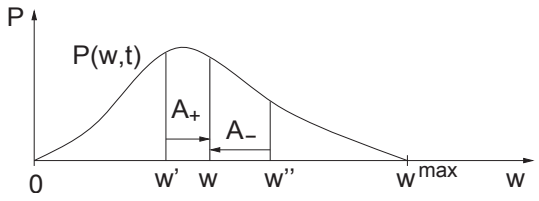

Fig. 11.10: Transitions of weight values due to synaptic plasticity. The probability density $P(w,t)$ increases if small weights increase, $w' \longrightarrow w' + A_+(w)$, or if large weights decrease, $w'' \longrightarrow w'' - A_-(w)$.

The eigenvectors and the eigenvalues of the matrix in square brackets are (apart from a common additive constant $\langle \xi^\mu \rangle_\mu W_-$ for the eigenvalues) the same as those of the matrix Q. We have already seen that this matrix is proportional to the properly normalized covariance matrix of the input patterns. If the proportionality constant is positive, i.e., if the integral over s in Eq. (11.74) is positive, then the dynamics of the weight vector is determined by the principal component of the set of input patterns.

11.2.5 Distribution of Synaptic Weights

If spike arrival times are described as a stochastic process, the weight vector itself is also a random variable that evolves along a fluctuating trajectory. In Section 11.2.1, we have analyzed the expectation value of the synaptic weights smoothed over a certain interval of time. In the limit where the synaptic weights evolve much slower than typical pre- or postsynaptic interspike intervals, an approximation of the weight vector by its expectation values is justified. However, if the synaptic efficacy can be changed substantially by only a few pre- or postsynaptic spikes then the fluctuations of the weights have to be taken into account. Here, we are investigate the resulting distribution of synaptic weights in the framework of a Fokker-Planck equation (van Rossum et al., 2000; Rubin et al., 2001).

We consider a single neuron i that receives input from several hundreds of presynaptic neurons. All presynaptic neurons fire independently at a common constant rate $\nu^{\rm pre}$. We are interested in the probability density $P(w,t)$ for the synaptic weight of a given synapse. We assume that all weights are restricted to the interval $[0, w^{\max}]$ so that the normalization $\int_0^{w^{\max}} P(w,t)\,{\rm d}w = 1$ holds. Weight changes due to potentiation or depression of synapses induce changes in the density function $P(w,t)$. The Fokker-Planck equation that we will derive below describes the evolution of the distribution $P(w,t)$ as a function of time; cf. Fig. 11.10.

For the sake of simplicity, we adopt a learning window with two rectangular

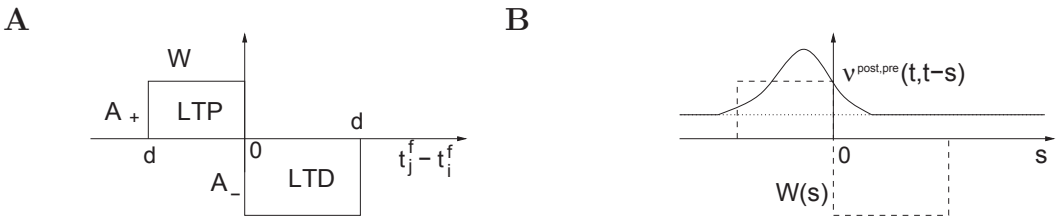

Fig. 11.11: **A**. Rectangular learning window $W(t_j^{(f)} - t_i^{(f)})$. LTP occurs if the presynaptic spike arrives before the postsynaptic one whereas LTD occurs if the order of timing is reversed. **B**. Whether LTP or LTD is dominant depends on the overlap between the learning window $W(s)$ (dashed line) and the correlations (solid line) between pre- and postsynaptic spike firing. The correlations consist of a constant bias term and a time-dependent term with a peak at negative values of s; cf. Eq. (11.71).

phases, i.e.,

$$W(s) = \begin{cases} A_+(w_{ij}) & \text{for } -d < s < 0 \\ A_-(w_{ij}) & \text{for } 0 < s < d \\ 0 & \text{else} \end{cases} \tag{11.79}$$

cf. Fig. 11.11A. Synapses are potentiated if the presynaptic spike shortly precedes the postsynaptic one. If the order of spike firing is reversed, the synapse is depressed.

There are basically two possibilities to restrict the synaptic weights to the interval $[0, w^{\max}]$; we can either impose hard or soft bounds to the weight dynamics; cf. Section 10.2.1. Hard bounds means that the weights are simply no longer increased (decreased) if the upper (lower) bound is reached. Soft bounds, on the other hand, gradually slow down the evolution if the weight approaches one of its bounds. A simple way to implement soft bounds in our formalism is to define (Kistler and van Hemmen, 2000a)

$$A_+(w_{ij}) = (w^{\max} - w_{ij})\, a_+ , \tag{11.80}$$
$$A_-(w_{ij}) = -w_{ij}\, a_- , \tag{11.81}$$

with constants a_+ and a_-. The choice of how the bounds are implemented turns out to have an important influence on the weight distribution $P(w,t)$ (van Rossum et al., 2000; Rubin et al., 2001).

In order to derive the evolution of the distribution $P(w,t)$ we consider transitions in the 'weight space' induced by pre- and postsynaptic spike firing. The

evolution is described by a master equation of the form

$$\frac{\partial}{\partial t}P(w,t) = -p_+(w)\,P(w,t) - p_-(w)\,P(w,t) + \int_0^{w^{\max}} \delta[w - w' - A_+(w')]\,p_+(w',t)\,P(w',t)\,\mathrm{d}w' + \int_0^{w^{\max}} \delta[w - w' + A_-(w')]\,p_-(w',t)\,P(w',t)\,\mathrm{d}w'\,; \tag{11.82}$$

cf. Fig. 11.10. Here p_+ (or p_-) is the probability that a presynaptic spike falls in the positive (or negative) phase of the learning window. Using the definition of the joint firing rate of pre- and postsynaptic neuron

$$\nu^{\mathrm{post,pre}}(t,t') = \langle S^{\mathrm{post}}(t)\,S^{\mathrm{pre}}(t')\rangle_E \tag{11.83}$$

we have

$$p_+(w,t) = \int_{-d}^{0} \nu^{\mathrm{post,pre}}(t,t-s)\,\mathrm{d}s \tag{11.84}$$

$$p_-(w,t) = \int_{0}^{d} \nu^{\mathrm{post,pre}}(t,t-s)\,\mathrm{d}s\,; \tag{11.85}$$

cf. Fig. 11.11B.

Equation (11.82) can be rewritten in the form of a Fokker-Planck equation if we expand the right-hand side to second order in the transition amplitudes A_+ and A_- (van Kampen, 1992),

$$\frac{\partial}{\partial t}P(w,t) = -\frac{\partial}{\partial w}[A(w)\,P(w,t)] + \frac{\partial^2}{\partial w^2}[B(w)\,P(w,t)] \tag{11.86}$$

with

$$A(w,t) = p_+(w,t)\,A_+(w) - p_-(w,t)\,A_-(w)\,, \tag{11.87}$$

$$B(w,t) = p_+(w,t)\,A_+^2(w) - p_-(w,t)\,A_-^2(w)\,. \tag{11.88}$$

The Fokker-Planck equation (11.86) can be solved numerically to find stationary solutions. It turns out that the qualitative form of the distribution depends critically on how the bounds for the weights are implemented; cf. van Rossum et al. (2000); Rubin et al. (2001) for details. With soft bounds the distribution is unimodal whereas with hard bounds it peaks at both borders of the interval; cf. Fig. 11.12. Experimental data suggests a unimodal distribution, consistent with soft bounds (van Rossum et al., 2000).

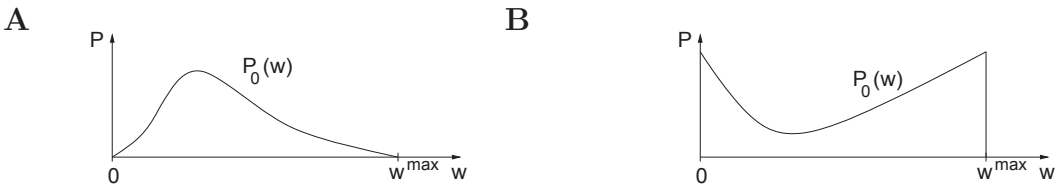

Fig. 11.12: Stationary distribution of synaptic weights. **A**. With soft bounds, the distribution of weights $P_0(w)$ has a single peak. **B**. With hard bounds, the distribution peaks at the two boundaries $w = 0$ and $w = w^{\max}$; (schematic figure)

11.3 Summary

The synaptic weight dynamics can be studied analytically if weights are changing slowly as compared to the time scale of the neuronal activity. We have seen that weight changes are driven by correlations between pre- and postsynaptic activity. More specifically, simple Hebbian learning rules can find the first principal component of a normalized input data set. If non-Hebbian terms are included then both spike-based and rate-based learning rules can be constructed that are characterized by a stable fixed point for the sum of the synaptic weights. This fixed point leads to an intrinsic normalization of the output firing rate.

The interesting aspect of spike-time dependent plasticity is that it naturally accounts for *temporal* correlations in the input by means of a learning window. Explicit expressions for temporal spike-spike correlations can be obtained for certain simple types of neuron model such as the linear Poisson model. In this case, correlations between pre- and postsynaptic neurons can be formulated in terms of the correlations in the input. It can be shown that, under certain circumstances, the weight vector evolves in the direction of the principal component of the input pattern set, even if the input is not normalized.

Spike-based and rate-based rules of plasticity are equivalent as long as *temporal* correlations are disregarded. The integral over the learning window $\int_{-\infty}^{\infty} W(s)\,\mathrm{d}s$ plays the role of the Hebbian correlation term c_2^{corr}. If rates vary rapidly, i.e. on the time scale of the learning window, then spike-time dependent plasticity is distinct from a rate-based formulation.

In addition to an analysis of the expectation value of the synaptic weight vector the distribution of weights can be described by means of a Fokker-Planck equation. The stationary distribution depends on the details of the learning rule.

References

More on the theory of unsupervised learning and principal component analysis can be found in the classical book by Hertz et al. (1991). Models of the development of receptive fields and cortical maps have a long tradition in the field

of computational neuroscience; see, e.g., von der Malsburg (1973); Sejnowski (1977); Kohonen (1984); Linsker (1986c); Sejnowski and Tesauro (1989); Miller et al. (1989); MacKay and Miller (1990); Shouval and Perrone (1995); Miller (1994); for a review see, e.g., Erwin et al. (1995); Wiskott and Sejnowski (1998). The linear rate model discussed in Section 11.1 is reviewed in Miller (1995). The essential aspects of the weight dynamics in linear networks are discussed in Oja (1982); MacKay and Miller (1990); Miller and MacKay (1994).

The theory of spike-based Hebbian learning has been developed by Gerstner et al. (1996a); Häfliger et al. (1997); Ruf and Schmitt (1997); Eurich et al. (1999); Kempter et al. (1999); Roberts (1999, 2000); Roberts and Bell (2000); Kistler and van Hemmen (2000b); van Rossum et al. (2000); Xie and Seung (2000); Song et al. (2000); Senn et al. (2001b); Rubin et al. (2001); and others. Spike-based learning rules are closely related to rules for sequence learning (Herz et al., 1988, 1989; van Hemmen et al., 1990; Gerstner et al., 1993b; Minai and Levy, 1993; Abbott and Blum, 1996; Gerstner and Abbott, 1997), where the idea of an asymmetric learning windows is exploited.

Chapter 12

Plasticity and Coding

In Chapters 10 and 11 we have explored the principle of Hebbian synaptic plasticity. In this final chapter we would like to close the chain of arguments that we have followed across the book and establish a link between synaptic plasticity and the problems of neuronal coding and signal transmission. We will start the chapter with the question of rapid and reliable signal transmission, a question that we have encountered on several occasions in this book. In Section 12.1 we will see that an asymmetric spike-time dependent learning rule is capable of detecting early events that may serve as predictors for others. Such a mechanism can speed up signal processing and, hence, the reaction time. In Section 12.2 we show that spike-time dependent plasticity can enhance signal transmission by selectively stregthening synaptic connections that transmit precisely timed spikes at the expense of those synapses that transmit poorly timed spikes. In Section 12.3 we turn to sequence learning and explore whether spike-time dependent plasticity can support coding schemes that are based on spatio-temporal spike patterns with a millisecond resolution. The last two sections study coding properties of specific neuronal systems. In Section 12.4 we will illustrate the role of an inverted (or anti-)Hebb rule for the subtraction of expectations - which has been hypothesized as an important component of signal processing in electric fish. Finally, in Section 12.5 we will see that a spike-time dependent Hebbian rule can play an important role in the developmental tuning of signal transmission in the auditory system of barn owls.

12.1 Learning to be Fast

In many real-world situations we must react rapidly to the earliest signs that could warn us about harmful stimuli. If an obstacle blocks our way, we want to avoid it before a painful contact occurs. If we ride a bicycle, we should make correcting steering movements already at small inclination angles of the bicycle, well before we fall down. Spike-time dependent learning rules with a temporally

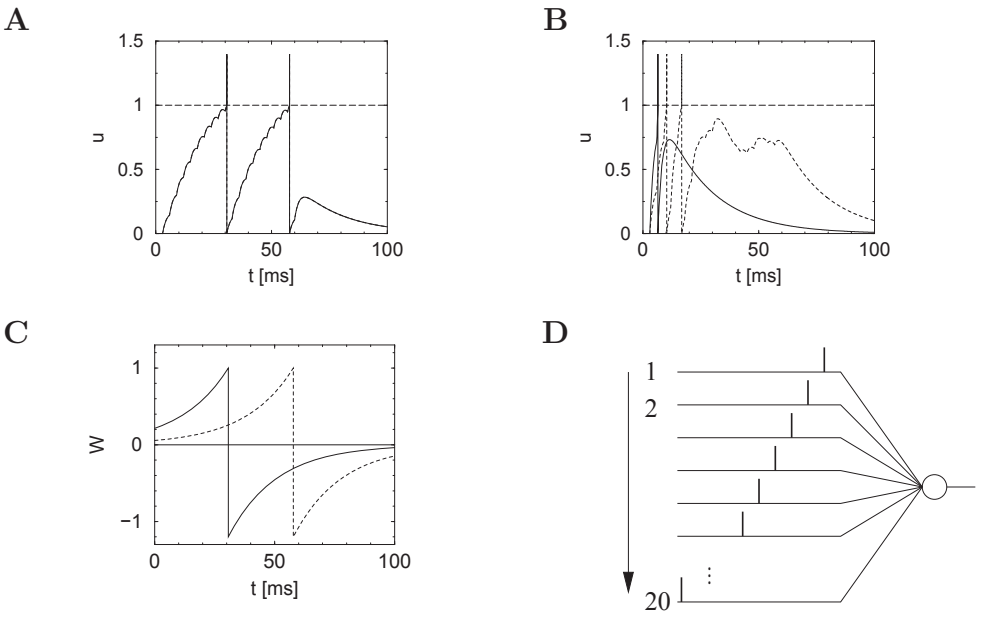

Fig. 12.1: **A**. A postsynaptic neuron receives inputs from twenty presynaptic cells at intervals of 3 ms. All synapses have the same weight. The neuron emits two output spikes at about 31 and 58 ms after stimulus onset. **B**. After 5 repetitions of the same stimulus the neuron fires after 10 ms (dashed line); after a total of 100 repetitions the neuron fires already after about 5 ms (solid line). **C**. The reason is that synapses that have been active slightly before the postsynaptic spikes are strengthened while others are depressed. To illustrate this point, the learning window $W(t_j^{(f)} - t_i^{(f)})$ is shown twice, each time centered at the postsynaptic firing time $t_i^{(f)}$ of the first trial (shown in part A). **D**. The sequence of presynaptic spikes could be generated by a stimulus that moves from top to bottom.

asymmetric learning window provide a hint of how a simple predictive coding could be implemented on the neuronal level.

Let us consider a single neuron that receives inputs from, say, twenty presynaptic cells which are stimulated one after the other; cf. Fig. 12.1. Initially, all synapses w_{ij} have the same weight w_0. The postsynaptic neuron fires two spikes; cf. Fig. 12.1A. All synapses that have been activated *before* the postsynaptic spike are strengthened while synapses that have been activated immediately afterwards are depressed; cf. Fig. 12.1C. In subsequent trials the threshold is therefore reached earlier; cf. Fig. 12.1B. After many trials, those presynaptic neurons that fire first have developed strong connections while other connections are depressed. Thus a temporally asymmetric learning rule favors connections that can serve as 'earliest predictors' of other spike events (Mehta et al., 2000; Song et al., 2000).

This theoretical observation predicts a shift in so-called place fields of hip-

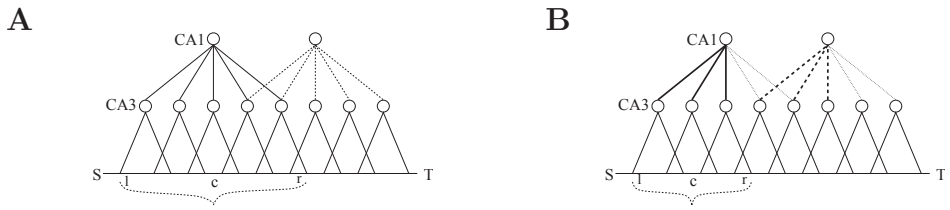

Fig. 12.2: **A** The place fields of neurons in regions CA3 of hippocampus are indicated as cones along the track that extends from S to T. The place field of the neuron in CA1 shown with solid lines has its center at c and extends from l to r. **B**. After the rat has made several movements from left to right, some connections are increased (thick lines) others decreased (dotted lines). As a result, the place field center c has moved to the left.

pocampal neurons that seems to be in agreement with experiment observations (Mehta et al., 2000). More generally, early predictors play a central role in the theory of conditioning and reinforcement learning (Rescorla and Wagner, 1972; Sutton and Barto, 1981; Schultz et al., 1997; Montague et al., 1995).

Example: Hippocampal place fields

Place cells are neurons in rodent hippocampus that are sensitive to the spatial location of the animal in an environment. The sensitive area is called the place field of the cell. If, for example, a rat runs on a linear track from a starting point S to a target point T, this movement would first activate cells with a place fields close to S, then those with a place field in the middle of the track, and finally those with a place field close to T; cf. Fig. 12.2. In a simple feedforward model of the hippocampus (Mehta et al., 2000), a first set of place cells is identified with neurons in region CA3 of rat hippocampus. A cell further down the processing stream (i.e., a cell in hippocampal region CA1) receives input from several cells in CA1. If we assume that initially all connections have the same weight, the place field of a CA1 cell is therefore broader than that of a CA3 cell.

During the experiment, the rat moves repeatedly from left to right. During each movement, the same sequence of CA3 cells is activated. This has consequences for the connections from CA3 cells to CA1 cells. Hebbian plasticity with an asymmetric learning window strengthens those connections where the presynaptic neuron fires early in the sequence. Connections from neurons that fire later in the sequence are weakened. As a result the center of the place field of a cell in CA3 is shifted to the left; cf. Fig. 12.2B. The shift of place fields predicted by asymmetric Hebbian learning has been confirmed experimentally (Mehta et al., 2000).

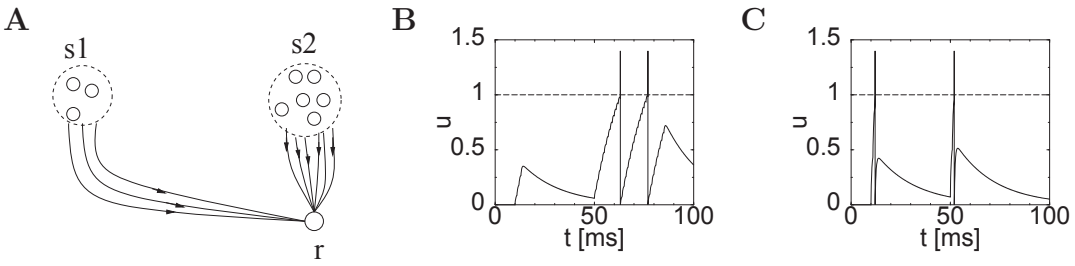

Fig. 12.3: **A**. Conditioning paradigm. A response neuron r can receive input from two neuronal populations, representing the stimuli $s1$ and $s2$. **B**. Membrane potential of the postsynaptic neuron. Before learning, stimulation of the presynaptic population $s1$ which occurs at about $t = 10\,\text{ms}$ leads to subthreshold excitation of the postsynaptic neuron whereas stimulation of group $s2$ 40 milliseconds later evokes postsynaptic firing. **C**. After learning postsynaptic firing is already triggered by the stimulus $s1$.

Example: Conditioning

The shift of responses towards early predictors plays a central role in conditioning. The basic idea is best explained by the paradigm of Pavlovian conditioning (Pavlov, 1927). Tasting or smelling food (stimulus $s2$) evokes an immediate response r. During the conditioning experiment, a bell (stimulus $s1$) rings always at a fixed time interval ΔT *before* the food stimulus. After several repetitions of the experiment, it is found that the response now occurs already after the first stimulus ($s1$). Thus the reaction has moved from stimulus $s2$ to stimulus $s1$ which reliably predicts $s2$.

Spike-time dependent plasticity with an asymmetric learning window allows to replicate this result, if the time difference ΔT between the two stimuli is less than the width of the learning window. The mechanism is identical to that of the previous example with the only difference that the input spikes are now clustered into two groups corresponding to the stimuli $s1$ and $s2$; cf. Fig. 12.3.

In behavioral experiments with monkeys, conditioning is possible with time intervals that span several seconds (Schultz et al., 1997) whereas typical learning windows extend over 50-100 milliseconds (Markram et al., 1997; Zhang et al., 1998; Magee and Johnston, 1997; Debanne et al., 1998; Bi and Poo, 1998, 1999). In order to explain conditioning with time windows longer than 100 milliseconds, additional assumptions regarding neuronal architecture and dynamics have to be made; see, e.g., Fiala et al. (1996); Brown et al. (1999); Suri and Schutz (2001). A potential solution could be provided by delayed reverberating loops; cf. Chapter 8.3. As an aside we note that, traditionally, conditioning experiments have been discussed on the level of rate coding. For slowly changing firing rates, spike-time dependent rules learning rules with an asymmetric learning window yield a differential Hebbian term [cf. Eq. (11.64)] that is proportional to the derivative

Bibliography

Abbott, L. F. (1991). Realistic synaptic inputs for model neural networks. *Network*, 2:245–258. 68

Abbott, L. F. (1994). Decoding neuronal firing and modeling neural networks. *Quart. Rev. Biophys.*, 27:291–331. 34, 38

Abbott, L. F. (2000). Synaptic plastictiy - taming the beast. *Nature Neurosci.*, 3:1178–1183. 394

Abbott, L. F. and Blum, K. I. (1996). Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex*, 6:406–416. 427

Abbott, L. F., Fahri, E., and Gutmann, S. (1991). The path integral for dendritic trees. *Biol. Cybern.*, 66:49–60. 71, 143

Abbott, L. F. and Kepler, T. B. (1990). Model neurons: from Hodgkin-Huxley to Hopfield. In Garrido, L., editor, *Statistical Mechanics of Neural Networks*. Springer, Berlin. 80, 99

Abbott, L. F. and van Vreeswijk, C. (1993). Asynchronous states in a network of pulse-coupled oscillators. *Phys. Rev. E*, 48:1483–1490. 105, 107, 213, 217, 229, 231, 256, 295

Abeles, M. (1982). *Local cortical circuits*. Springer-Verlag, Berlin Heidelberg New York. 176, 177

Abeles, M. (1991). *Corticonics*. Cambridge Univ. Press, Cambridge. 32, 111, 194, 195, 291, 351, 357, 441

Abeles, M. (1994). Firing rates and well-timed events. In Domany, E., Schulten, K., and van Hemmen, J. L., editors, *Models of Neural Networks II*, chapter 3. Springer, New York. 24, 32, 38

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. 32

Adrian, E. D. (1926). The impulses produced by sensory nerve endings. *J. Physiol. (Lond.)*, 61:49–72. 24, 25

Adrian, E. D. (1928). *The basis of sensation*. W. W. Norton, New York. 24

Aertsen, A. and Arndt, M. (1993). Response synchronization in the visual cortex. *Curr. Opin. Neurobiol.*, 3:586–594. 32

Aizenman, C. D. and Linden, D. J. (1999). Regulation of the rebound depolarization and spontaneous firing patterns of deep nuclear neurons in slices of rat cerebellum. *J. Neurophysiol.*, 82:1697–1709. 56

Amari, S. (1972). Characteristics of random nets of analog neuron-like elements. *IEEE trans. syst. man cybern.*, 2:643–657. 246, 248, 256

Amari, S. (1974). A method of statistical neurodynamics. *Kybernetik*, 14:201–215. 213, 246, 256

Amari, S. (1977a). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27:77–87. 254, 256

Amari, S. I. (1977b). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27:77–87. 246, 256, 325, 334, 341

Amit, D. J. and Brunel, N. (1997a). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*, 7:237–252. 188, 213, 246, 248, 250, 256

Amit, D. J. and Brunel, N. (1997b). A model of spontaneous activity and local delay activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7:237–252. 213, 246, 248, 250

Anderson, J. A. and Rosenfeld, E., editors (1988). *Neurocomputing: Foundations of research*. MIT-Press, Cambridge Mass. 209, 393

Artola, A., Bröcher, S., and Singer, W. (1990). Different voltage dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347:69–72. 371

Artola, A. and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.*, 16(11):480–487. 371, 394

Ash, R. (1990). *Information Theory*. Dover, New York. 320

Bair, W. and Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput.*, 8:1185–1202. 157, 158, 159, 209

Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of MT neurons in the behaving monkey. *J. Neurosci.*, 14:2870–2892. 170, 172, 173

Bartlett, M. S. (1963). The spectral analysis of point processes. *J. R. Statist. Soc. B*, 25:264–296. 416, 435

Bauer, H. U. and Pawelzik, K. (1993). Alternating oscillatory and stochastic dynamics in a model for a neuronal assembly. *Physica D*, 69:380–393. 217

Bell, C., Bodznick, D., Montgomery, J., and Bastian, J. (1997a). The generation and subtraction of sensory expectations within cerebellar-like structures. *Brain. Beh. Evol.*, 50:17–31. suppl. I. 446, 448

Bell, C., Han, V., Sugawara, Y., and Grant, K. (1997b). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, 387:278–281. 366, 393, 446, 447, 449

Bell, C. C. and Kawasaki, T. (1972). Relations among climbing fiber responses of nearby Purkinje cells. *J. Neurophysiol.*, 35:155–169. 311

Ben Arous, G. and Guionnet, A. (1995). Large deviations for langevin spin glass dynamics. *Probability Theory and Related Fields*, 102:455–509. 248

Ben-Yishai, R., Lev Bar-Or, R., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA*, 92:3844–3848. 334

Berry, M. J. and Meister, M. (1998). Refractoriness and neural precision. *J. Neurosci.*, 18:2200–2211. 169, 179

Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. USA*, 94:5411–5416. 157, 159

Bethge, M., Pawelzik, K., Rothenstein, R., and Tsodyks, M. (2001). Noise as a signal for neuronal populations. *Phys. Rev. Lett.*, xx:xx. 259

Bi, G. and Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, 18:10464–10472. 365, 366, 393, 432, 449

Bi, G. and Poo, M. (2001). Synaptic modification of correlated activity: Hebb's postulate revisited. *Ann. Rev. Neurosci.*, 24:139–166. 365, 369, 394

Bi, G. Q. and Poo, M. M. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, 401:792–796. 365, 366, 393, 432

Bialek, W. and Rieke, F. (1992). Reliability and information transmission in spiking neurons. *Trends Neurosci.*, 15(11):428–433. 34

Bialek, W., Rieke, F., de Ruyter van Stevenick, R. R., and Warland, D. (1991). Reading a neural code. *Science*, 252:1854–1857. 24, 33, 34, 36, 159

Bienenstock, E. L., Cooper, L. N., and Munroe, P. W. (1982). Theory of the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, 2:32–48. reprinted in Anderson and Rosenfeld, 1990. 371, 386, 393

Billock, V. A. (1997). Very short-term visual memory via reverberation: a role for the cortico-thalamic excitatory circuit in temporal filling-in during blinks and saccades? *Vision Res.*, 37:949–953. 312, 320

Bindman, L., Christofi, G., Murphy, K., and Nowicky, A. (1991). Long-term potentiation (ltp) and depression (ltd) in the neocortex and hippocampus: an overview. In TW, S., editor, *Aspects of synaptic transmission.*, volume 1, pages 3–25, London. Taylor Francis. 387

Bliss, T. V. P. and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361:31–39. 362, 363, 387, 394

Bliss, T. V. P. and Gardner-Medwin, A. R. (1973). Long-lasting potentation of synaptic transmission in the dendate area of unanaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, 232:357–374. 362

Bliss, T. V. P. and Lomo, T. (1973). Long-lasting potentation of synaptic transmission in the dendate area of anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, 232:551–356. 362

Bose, A., Kopell, N., and Terman, D. (2000). Almost-synchronous solutions for mutually coupled excitatory neurons. *Physica D*, 140:69–94. 301

Bower, J. M. and Beeman, D. (1995). *The Book of GENESIS: exploring realistic neural models with the GEneral NEural SImulation System.* Springer, New York. 71, 72, 76

Bressloff, P. C. (1999). Synaptically generated wave propagation in excitable neural media. *Phys. Rev. Lett.*, 82:2979–2982. 345, 347

Bressloff, P. C. and Taylor, J. G. (1994). Dynamics of compartmental model neurons. *Neural Networks*, 7:1153–1165. 143, 146

Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.*, 59:189–200. 119, 155

Brillinger, D. R. (1992). Nerve cell spike train data analysis: a progressiion of thechniques. *J. Am. Statist. Assoc.*, 87:260–271. 119, 155

Brillinger, D. R. and Segundo, J. P. (1979). Empirical examination of the threshold model of neuronal firing. *Biol. Cybern.*, 35:213–220. 119

Brown, J., Bullock, D., and Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.*, 19:10502–11. 432

Brown, T. H., Ganong, A. H., Kairiss, E. W., Keenan, C. L., and Kelso, S. R. (1989). Long-term potentiation in two synaptic systems of the hippocampal brain slice. In JH, B. and WO, B., editors, *Neural models of plasticity.*, pages 266–306. Academic Press, San Diego. 362, 363, 394

Brown, T. H., Zador, A. M., Mainen, Z. F., and Claiborne, B. J. (1991). Hebbian modifications in hippocampal neurons. In Baudry, M. and Davis, J. L., editors, *Long-term potentiation.*, pages 357–389. MIT Press, Cambridge, London. 366, 369

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, 8:183–208. 213, 246, 248, 250, 256, 296, 299, 300

Brunel, N., Chance, F., Fourcaud, N., and Abbott, L. F. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys. Rev. Lett.*, 86:2186–2189. 28, 259, 273, 276, 277, 279, 280, 290, 291

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. 160, 188, 213, 217, 222, 246, 250, 254, 255, 256, 259, 296, 299

Bryant, H. L. and Segundo, J. P. (1976). Spike initiation by transmembrane current: a white-noise analysis. *J. Physiol. (Lond.)*, 260:279–314. 157, 159, 287

Buck, J. and Buck, E. (1976). Synchronous fireflies. *Sci. Am.*, 234:74–85. 322

Bugmann, G., Christodoulou, C., and Taylor, J. G. (1997). Role of temporal integration and fluctuation detection in the highly irregular firing of leaky integrator neuron model with partial reset. *Neural Comput.*, 9:985–1000. 195

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. 194

Byrne, J. H. and Berry, W. O. (1989). *Neural Models of Plasticity.* Academic Press, San Diego. 394

Calvin, W. and Stevens, C. F. (1968). Synaptic noise and other sources of randomness in motoneuron interspike intervals. *J. Neurophysiol.*, 31:574–587. 222

Capocelli, R. M. and Ricciardi, L. M. (1971). Diffusion approximation and first passage time problem for a neuron model. *Kybernetik*, 8:214–223. 188

Carpenter, G. and Grossberg, S. (1987). Art 2: Self-organization of stable category recognition codes for analog input patterns. *Appl. Optics*, 26:4919–4930. 376

Carr, C. E. (1993). Processing of temporal information in the brain. *Annu. Rev. Neurosci.*, 16:223–243. 372, 450, 455, 456, 462

Carr, C. E. (1995). The development of nucleus laminaris in the barn owl. In Manley, G. A., Klump, G. M., Köppl, C., Fastl, H., and Oeckinghaus, H., editors, *Advances in Hearing Research*, pages 24–30, Singapure. World Scientific. 450, 457, 461

Carr, C. E. and Konishi, M. (1988). Axonal delay lines for measurement in the owl's brainstem. *Proc. Natl. Acad. Sci. USA*, 85:8311–8315. 450

Carr, C. E. and Konishi, M. (1990). A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.*, 10:3227–3246. 372, 450, 452, 455, 456, 457

Cessac, B., Doyon, B., Quoy, M., and Samuelides, M. (1994). Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. *Physica D*, 74:24–44. 246, 248, 256

Chow, C. C. (1998). Phase-locking in weakly heterogeneous neuronal networks. *Physica D*, 118:343–370. 254, 301, 309

Chow, C. C. and Kopell, N. (2000). Dynamics of spiking neurons with electrical coupling. *Neural Comput.*, 12:1643–1678. 301

Chow, C. C. and White, J. (1996). Spontaneous action potential fluctuations due to channel fluctuations. *Bioph. J.*, 71:3013–3021. 160

Collingridge, G. L., Kehl, S. J., and McLennan, H. (1983). Excitatory amino acids in synaptic transmission in the schaffer collateral-commissural pathway of the rat hippocampus. *J. Physiol.*, 334:33–46. 387

Collins, J. J., Chow, C. C., Capela, A. C., and Imhoff, T. T. (1996). Aperiodic stochastic resonance. *Phys. Rev. E*, 54:5575–5584. 201

Connor, J. A., Walter, D., and McKown, R. (1977). Neural repetitive firing - modifications of the hodgkin-huxley axon suggested by experimental results from crustacean axons. *Biophys. J.*, 18:81–102. 54

Connors, B. W. and Gutnick, M. J. (1990). Intrinsic firing patterns of diverse cortical neurons. *Trends Neurosci.*, 13:99–104. 19, 20, 255

Cordo, P., Inglis, J. T., Verschueren, S., andD. M. Merfeld, J. J. C., and Rosenblum, S. (1996). Noise in human muscle spindels. *Nature*, 383:769–770. 201

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley, New York. 320

Cox, D. R. (1955). Some statistical methods connected with series of events. *J. R. statist. Soc. B*, 17:129–164. 416, 435

Cox, D. R. (1962). *Renewal theory*. Mathuen, London. 161, 164, 255

Cox, D. R. and Lewis, P. A. W. (1966). *The statistical analysis of series of events*. Methuen, London. 164, 172, 209

Crepel, F. (1982). Regression of functional synapses in the immature mammalian cerebellum. *Trends Neurosci.*, 5:266–269. 376

Crisanti, A. and Sompolinsky, H. (1988). Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics. *Phys. Rev. A*, 37:4865–4874. 246, 314

Cronin, J. (1987). *Mathematical aspects of Hodgkin-Huxley theory*. Cambridge University Press, Cambridge.

Daniel, H., Blond, O., Jaillard, D., and Crepel, F. (1996). Synaptic plasticity in the cerebellum. In Fazeli, M. S. and Collingridge, G. L., editors, *Cortical Plasticity, LTP and LTD*, chapter 10. BIOS Scientific Publishers Ltd., Oxford. 394

Daniel, H., Levenes, C., and Crépel, F. (1998). Cellular mechanisms of cerebellar LTD. *Trends Neurosci.*, 21:401–407. 394

de Boer, E. and Kuyper, P. (1968). Triggered correlation. *IEEE Trans. Biomed. Eng.*, 15:169–179. 32

de Ruyter van Stevenick, R. R. and Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. B*, 234:379–414. 33

de Ruyter van Steveninck, R. R., Lowen, G. D., Strong, S. P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275:1805. 157, 159

De Zeeuw, C. I., Simpson, J. I., Hoogenraad, C. C., Galjart, N., Koekkoek, S. K. E., and Ruigrok, T. J. H. (1998). Microcircuitry and function of the inferior olive. *Trends Neurosci.*, 21:391–400. 311

DeAngelis, G. C., Ohzwaw, I., and Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends Neurosci.*, 18:451–458. 33

Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1994). Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat Hippocampus in vitro. *Proc. Natl. Acad. Sci. USA*, 91:1148–1152. 364

Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1998). Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol.*, 507:237–247. 365, 366, 393, 432, 449

deCharms, R. C. and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, 381:610–613. 32

Derrida, B., Gardner, E., and Zippelius, A. (1987). An exactly solvable asymmetric neural network model. *Europhys. Lett.*, 2:167–173. 247, 314

Desmaison, D., Vincent, J.-D., and Lledo, P.-M. (1999). Control of action potential timing by intrinsic subthreshold oscillations in olfactory bulb output neurons. *J. Neurosci.*, 19:10727–10737. 293

Destexhe, A. and Pare, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol.*, 81:1531–1547. 222

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402:529–533. 351, 353, 356

Douglass, J. K., Wilkens, L., Pantazelou, E., and Moss, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365:337–340. 201

Dudek, S. M. and Bear, M. F. (1992). Homosynaptic long-term depression in area ca1 of hippocampus and effects of n-methyl-d-aspartate receptor blockade. *Proc. Natl. Acad. Sci. USA*, 89:4363–4367. 387

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., and Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biol. Cybern.*, 60:121–130. 24, 32, 293

Eckhorn, R. and Brosch, M. (1993). Synchronous oscillatory activities between areas 17 and 18 in the cat's visual cortex. *J. Neurophysiol.* in press. 293

Eckhorn, R., Krause, F., and Nelson, J. L. (1993). The rf-cinematogram: a cross-correlation technique for mapping several visual fields at once. *Biol. Cybern.*, 69:37–55. 33

Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comput.*, 2:293–307. 32, 111

Edwards, B. E. and Wakefield, G. H. (1993). The spectral shaping of neural discharges by refractory effects. *J. Acoust. Soc. Am.*, 93:3553–3564. 170, 172, 173

Egger, V., Feldmeyer, D., and Sakmann, B. (1999). Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in barrel cortex. *Nature Neurosci.*, 2:1098–1105. 366, 393

Eggert, J. and van Hemmen, J. L. (2001). Modeling neuronal assemblies: theory and implementation. *Neural Computation*, 13:1923–1974. 213

Ekeberg, O., Wallen, P., Lansner, A., Traven, H., Brodin, L., and Grillner, S. (1991). A computer based model for realistic simulations of neural networks. *Biol. Cybern.*, 65:81–90. 71

Ellias, S. A. and Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biol. Cybern.*, 20:69–98. 254, 256

Engel, A. K., König, P., Kreiter, A. K., and Singer, W. (1991a). Interhemispheric synchronization of oscillatory neural responses in cat visual cortex. *Science*, 252:1177–1179. 24, 32

Engel, A. K., König, P., and Singer, W. (1991b). Direct physiological evidence for scene segmentation by temporal coding. *Proc. Natl. Acad. Sci. USA*, 88:9136–9140. 32

Erisir, A., Lau, D., Rudy, B., and Leonard, C. S. (1999). Specific k+ channels are required to sustain high frequency firing in fast-spiking neocortical interneurons. *J. Neurophysiol.*, 82:2476–2489. 54, 132

Ermentrout, G. B. (1996). Type i membranes, phase resetting curves, and synchrony. *Neural Comput.*, 8:979–1001. 91, 99, 106, 107, 132, 301

Ermentrout, G. B. and Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. *Biol. Cybern.*, 34:137–150. 357

Ermentrout, G. B. and Kopell, N. (1984). Frequency plateaus in a chain of weakly coupled oscillators. *SIAM J. Math. Anal.*, 15:215–237. 301

Ermentrout, G. B. and Kopell, N. (1986). Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM J. Appl. Math.*, 46:233–253. 107

Ermentrout, G. B., Pascal, M., and Gutkin, B. (2001). The effects of spike frequency adaptation and negative feedback on the synchronization of neuronal oscillators. *Neural Comput.*, 13:1285–1310. 301

Ernst, U., Pawelzik, K., and Geisel, T. (1995). Synchronization induced by temporal delays in pulse-coupled oscillators. *Phys. Rev. Lett.*, 74:1570–1573. 296, 301, 310

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: a critcal comparison. *Neural Comput.*, 7:425–468. 393, 427

Eurich, C. W., Pawelzik, K., Ernst, U., Cowan, J. D., and Milton, J. G. (1999). Dynamics of self-organized delay adaption. *Phys. Rev. Lett.*, 82:1594–1597. 427, 450, 457

Feldman, J. L. and Cowan, J. D. (1975). Large-scale activity in neural nets i: Theory with application to motoneuron pool responses. *Biol. Cybern.*, 17:29–38. 254, 325

Feng, J. (2001). Is the integrate-and-fire model good enough - a review. *Neural Comput.*, 13:xx. 106, 195

Fetz, E. E. and Gustafsson, B. (1983). Relation between shapes of post-synaptic potentials and changes in firing probability of cat motoneurones. *J. Physiol.*, 341:387–410. 281, 282, 284, 285, 291

Fiala, J., Grossberg, S., and Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar purkinje cells as a substrate for adaptive timing of the classically conditioned eye-blink reflex. *J. Neurosci.*, 16:3760–3774. 432

FitzHugh, R. (1961). Impulses and physiological states in models of nerve membrane. *Biophys. J.*, 1:445–466. 79

French, A. S. and Stein, R. B. (1970). A flexible neural analog using integrated circuits. *IEEE trans. biomed. eng.*, 17(3):248–253. 155

Fuortes, M. G. F. and Mantegazzini, F. (1962). Interpretation of the repetitive firing of nerve cells. *J. Gen. Physiol.*, 45:1163–1179. 50, 111, 112, 113

Fusi, S., Del Giudice, P., and Amit, D. J. (2000). Neurophysiology of a vlsi spiking neural network: Lann21. In *Proceedings of IJCNN2000*. 160, 376, 380

Fusi, S. and Mattia, M. (1999). Collective behavior of networks with linear (vlsi) integrate and fire neurons. *Neural Comput.*, 11:633–652. 213, 217

Gabbiani, F. and Koch, C. (1998). Principles of spike train analysis. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling*, chapter 9, pages 312–360. MIT press, 2nd edition. 209

Gabbiani, F., Midtgaard, J., and Knoepfl, T. (1994). Synaptic integration in a model of cerebellar granule cells. *J. Neurophysiol.*, 72:999–1009. Corrigenda have been published in J. Neurophysiol. (1996) **75**(6), without covering, however, all typing errors. 62, 63, 72

Gamble, E. and Koch, C. (1987). The dynamics of free calcium in dendritic spines in response to repetitive synaptic input. *Science*, 236:1311–1315. 389

Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F. (1998). Stochastic resonance. *Rev. Mod. Phys.*, 70:223–287. 201, 209

Geisler, C. D. and Goldberg, J. M. (1966). A stochastic model of repetitive activity of neurons. *Biophys. J.*, 6:53–69. 17, 38, 111, 155, 177, 209

Georgopoulos, A. P., Schwartz, A., and Kettner, R. E. (1986). Neuronal populations coding of movement direction. *Science*, 233:1416–1419. 29

Gerstein, G. L. and Perkel, D. H. (1972). Mutual temporal relations among neuronal spike trains. *Biophys. J.*, 12:453–473. 164, 209

Gerstner, W. (1995). Time structure of the activity in neural network models. *Phys. Rev. E*, 51:738–758. 17, 111, 161, 213, 232, 238, 255, 296

Gerstner, W. (1998). Spiking neurons. In Maass, W. and Bishop, C. M., editors, *Pulsed Neural Networks*, chapter 1, pages 3–53. MIT-Press. 150, 151

Gerstner, W. (2000a). Population dynamics of spiking neurons: fast transients, asynchronous states and locking. *Neural Comput.*, 12:43–89. 28

Gerstner, W. (2000b). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Comput.*, 12:43–89. 124, 161, 182, 213, 233, 240, 255, 259, 263, 270, 277, 278, 279, 290, 291, 296, 298

Gerstner, W. and Abbott, L. F. (1997). Learning navigational maps through potentiation and modulation of hippocampal place cells. *J. Comput. Neurosci.*, 4:79–94. 427

Gerstner, W., Kempter, R., and van Hemmen, J. L. (1998). Hebbian learning of pulse timing in the barn owl auditory system. In Maass, W. and Bishop, C. M., editors, *Pulsed Neural Networks*, chapter 14, pages 353–377. MIT-Press. 382, 383, 451, 454

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996a). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 386:76–78. 383, 427, 442, 450, 456, 457, 458, 459

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1997). A developmental learning rule for coincidence tuning in the barn owl auditory system. In Bower, J., editor, *Computational Neuroscience: trends in research 1997*, pages 665–669. Plenum Press, New York. 450, 460

Gerstner, W., Ritz, R., and van Hemmen, J. L. (1993a). A biologically motivated and analytically soluble model of collective oscillations in the cortex: I. theory of weak locking. *Biol. Cybern.*, 68:363–374. 32, 301

Gerstner, W., Ritz, R., and van Hemmen, J. L. (1993b). Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biol. Cybern.*, 69:503–515. 427, 441, 442, 444

Gerstner, W. and van Hemmen, J. L. (1992). Associative memory in a network of 'spiking' neurons. *Network*, 3:139–164. 213, 217, 226, 232

Gerstner, W. and van Hemmen, J. L. (1993). Coherence and incoherence in a globally coupled ensemble of pulse emitting units. *Phys. Rev. Lett.*, 71:312–315. 296

Gerstner, W. and van Hemmen, J. L. (1994). Coding and information processing in neural networks. In Domany, E., van Hemmen, J. L., and Schulten, K., editors, *Models of Neural Networks II*. Springer, New York. 217

Gerstner, W., van Hemmen, J. L., and Cowan, J. D. (1996b). What matters in neuronal locking. *Neural Comput.*, 8:1653–1676. 119, 123, 301, 303, 305

Gestri, G. (1978). Dynamics of a model for the variability of the interspike intervals in a retinal neuron. *Biol. Cybern.*, 31:97–98. 182

Gewaltig, M.-O. (2000). *Evolution of synchronous spike volleys in cortical networks – Network simulations and continuous probabilistic models*. Shaker Verlag, Aachen, Germany. PhD thesis. 351, 353

Giorno, V., Nobile, A. G., and M.Ricciardi, L. (1992). Instantaneous return processes and neuronal firings. In Trappl, R., editor, *Cybernetics and Systems Research, Vol 1.*, pages 829–236. World Scientific Press. 222

Gluss, B. (1967). A model of neuron firing with exponential decay of potential resulting in diffusion equations for the probability density. *Bull. Math. Biophys.*, 29:233–243. 188, 209

Gold, J. I. and Bear, M. F. (1994). A model of dendritic spike Ca^{2+} concentration exploring possible basis for sliding synaptic modification threshold. *Proc. Natl. Acad. Sci. USA*, 91:3941–3945. 387

Goldberg, J. M., Adrian, H. O., and Smith, F. D. (1964). Response of neurons of the superior olivary complex of the cat to acoustic stimuli of long duration. *J. Neurophysiol.*, pages 706–749. 165, 167, 169

Golomb, D. and Ermentrout, G. B. (2001). Bistability in pulse propagation in networks of excitatory and inhibitory populations. *Physical Review Letters*, 86:4179–4182. 345

Golomb, D., Hansel, D., Shraiman, B., and Sompolinsky, H. (1992). Clustering in globally coupled phase oscillators. *Phys. Rev. A*, 45:3516–3530. 296

Golomb, D. and Rinzel, J. (1994). Clustering in globally coupled inhibitory neurons. *Physica D*, 72:259–282. 296

Gray, C. M., König, 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. 24, 32, 293

Gray, C. M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. USA*, 86:1698–1702. 24, 32

Grossberg, S. (1976). Adaptive pattern classification and universal recoding i: Parallel development and coding of neuronal feature detectors. *Biol. Cybern.*, 23:121–134. reprinted in Anderson and Rosenfeld, 1990. 369, 393

Grossberg, S. (1987). *The adaptive brain I.* Elsevier. 376

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. *J. Neurosci.*, 7:774–780. 364

Häfliger, P., Mahowald, M., and Watts, L. (1997). A spike based learning neuron in analog vlsi. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Advances in Neural Information Processing Systems*, volume 9, pages 692–698. The MIT Press. 427

Hale, J. and Koçak, H. (1991). *Dynamics and Bifurcations.* Springer, New York. 85, 93, 99

Hansel, D. and Mato, G. (2001). Existence and stability of persistent states in large neuronal networks. *Phys. Rev. Lett.*, 86:4175–4178. 106

Hansel, D., Mato, G., and Meunier, C. (1995). Synchrony in excitatory neural networks. *Neural Comput.*, 7:307–337. 301

Hansel, D. and Sompolinsky, H. (1998). Modeling feature selectivity in local cortical circuits. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling.* MIT Press, Cambridge MA. 334

Haykin, S. (1994). *Neural Networks.* Prentice Hall, Upper Saddle River, NJ. 205, 209

Hebb, D. O. (1949). *The organization of behavior.* Wiley, New York. 361, 362, 393

Heiligenberg, W. (1991). *Neural Nets in Electric Fish.* MIT Press, Cambridge. 372, 462

Herrmann, A. and Gerstner, W. (2001a). Noise and the psth response to current transients: I. General theory and application to the integrate-and-fire neuron. *J. Comput. Neurosci.*, 11:to appear. 200, 283, 285

Herrmann, A. and Gerstner, W. (2001b). Noise and the psth response to current transients II. Iintegrate-and-fire model with slow recovery and application to motoneuron data. *J. Comput. Neurosci.*, 11:to appear. 116, 286

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Addison-Wesley, Redwood City CA. 205, 209, 426

Hertz, J. and Prugel-Bennet, A. (1996). Learning short synfire chains by self-organization. *Network*, 7:357–363. 441

Herz, A. V. M., Sulzer, B., Kühn, R., and van Hemmen, J. L. (1988). The Hebb rule: Representation of static and dynamic objects in neural nets. *Europhys. Lett.*, 7:663–669. 427

Herz, A. V. M., Sulzer, B., Kühn, R., and van Hemmen, J. L. (1989). Hebbian learning reconsidered: Representation of static and dynamic objects in associative neural nets. *Biol. Cybern.*, 60:457–467. 427, 441, 442

Hessler, N. A., Shirke, A. M., and Malinow, R. (1993). The probability of transmitter release at a mammalian central synapse. *Nature*, 366:569–572. 160

Hill, A. V. (1936). Excitation and accomodation in nerve. *Proc. R. Soc. B*, 119:305–355. 17, 38, 111, 155

Hille, B. (1992). *Ionic channels of excitable membranes*. Sinauer Associates, Sunderland, Mass., 2nd edition. 42, 51, 55, 62, 63, 73, 76

Ho, N. and Destexhe, A. (2000). Synaptic background activity enhances the responsiveness of neocortical pyramidal neurons. *J. Neurophysiol.*, 84:1488–1496. 222

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol. (Lond.)*, 117:500–544. 44, 45, 46, 51

Holden, A. V. (1976). *Models of the stochastic activity of neurons*, volume 12 of *Lecture notes in Biomathematics*. Springer, Berlin Heidelberg New York. 38

Holmes, W. R. and Levy, W. B. (1990). Insights into associative long-term potentiation from computational models of nmda receptor-mediated calcium influx and intracellular calcium concentration changes. *J. Neurophysiol.*, 63:1148–1168. 387

Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376:33–36. 24, 31

Hopfield, J. J. and Herz, A. V. M. (1995). Rapid local synchronization of action potentials: towards computation with coupled integrate-and-fire networks. *Proc. Natl. Acad. Sci. USA*, 92:6655–6662. 31

Hoppensteadt, F. C. and Izhikevich, E. M. (1997). *Weakly connected neural networks*. Springer. 99, 107, 301

Horn, D. and Usher, M. (1989). Neural networks with dynamical thresholds. *Phys. Rev. A*, 40:1036–1040. 111

Horn, R. A. and Johnson, C. R. (1985). *Matrix analysis*. Cambridge University Press, Cambridge, UK. 397

Hubel, D. H. (1988). *Eye, brain, and vision*. W. H. Freeman, New York. 28

Hubel, D. H. (1995). *Eye, Brain, and Vision*. Scientific American Library. W. H. Freeman, ??? 334, 406

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J. Physiol. (Lond.)*, 148:574–591. 24, 158

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)*, 160:106–154. 28, 213, 406

Hubel, D. H. and Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. B*, 198:1–59. 28, 158

Huguenard, J. R. and McCormick, D. A. (1992). Simulation of the currents involved in rhythmic oscillations in thalamic relay neurons. *J. Neurophysiol.*, 68:1373–1383. 53

Hüning, Glünder, and Palm, G. (1998). Synaptic delay learning in pulse-coupled neurons. *Neural Comput.*, 10:555–565. 450, 457

Ito, M. (1984). *The Cerebellum and Neural Control*. Raven Press, New York. 72

Izhikevich, E. (1999). Class 1 neural excitability, conventional synapses, weakly connected networks, and mathematical foundations of pulse-coupled models. *IEEE Trans. Neural Netw.*, 10:499–507. 301

Izhikevich, E. (2000). Neural excitability, spiking, and bursting. *Int. J. Bifurcat. Chaos*, 10:1171–1266. 20

Izhikevich, E. (2001). Resonate-and-fire neurons. *Neural Networks*, 14:883–894. 127

Jackson, J. D. (1962). *Classical Electrodynamics*. Wiley. 68

Jahnsen, H. (1986). Electrophysiological characteristics of neurones in the guinea-pig deep cerebellar nuclei in vitro. *J. Physiol. (Lond.)*, 372:129–147. 56

James, W. (1890). *Psychology (Briefer Course), ch. 16*. Holt, New York. 362, 393

Jeffress, L. A. (1948). A place theory of sound localisation. *J. Comp. Physiol. Psychol.*, 41:35–39. 450

Jensen, O. and Lisman, J. E. (1996). Hippocampal ca3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn. Mem.*, 3:279–287. 31

Johannesma, P. I. M. (1968). Diffusion models of the stochastic acticity of neurons. In *Neural Networks*, pages 116–144, Berlin. Springer. 188, 193, 209

Kandel, E. C. and Schwartz, J. H. (1991). *Principles of Neural Science*. Elsevier, New York, 3rd edition. 25, 38, 213

Kass, R. E. and Ventura, V. (2001). A spike-train probability model. *Neural Comput.*, 13:1713–1720. 161

Keener, J. and Sneyd, J. (1998). *Mathematical Physiology*, volume 8 of *Interdisciplinary Applied Mathematics*. Springer, New York. 357

Kelso, S. R., Ganong, A. H., and Brown, T. H. (1986). Hebbian synapses in hippocampus. *Proc. Natl. Acad. Sci. USA*, 83:5326–5330. 362

Kempter, R. (1997). *Hebbsches Lernen zeitlicher Codierung: Theorie der Schallortung im Hörsystem der Schleiereule*, volume 17. Naturwissenschaftliche Reihe, Darmstadt. 459

Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E*, 59:4498–4514. 415, 427, 450

Kempter, R., Gerstner, W., and van Hemmen, J. L. (2001). Intrinsic stabilization of output rates by spike-based hebbian learning. *Neural Comput.*, submitted. 368

Kepler, T. B., Abbott, L. F., and Marder, E. (1992). Reduction of conductance-based neuron models. *Biol. Cybern.*, 66:381–387. 77, 80, 99

Keysers, C., Xiao, D. K., Foldiak, P., and Perrett, D. I. (2001). The speed of sight. *J. Cognitive Neuroscience*, 13:90–101. 26

Kirkpatrick, S. and Sherrington, D. (1978). Infinite-ranged models of spin-glasses. *Phys. Rev. B*, 17:4384–4403. 314, 319

Kirkwood, P. A. and Sears, P. A. (1978). The synaptic connexions to intercostal motoneurones as revealed by the average common excitation potential. *J. Physiology*, 275:103–134. 282, 285

Kishimoto, K. and Amari, S. (1979). Existence and stability of local excitations in homogeneous neural fields. *J. Math. Biology*, 7:303–318. 325

Kistler, W. M. (2000). Stability properties of solitary waves and periodic wave trains in a two-dimensional network of spiking neurons. *Phys. Rev. E*, 62(6):8834–8837. 119, 345, 348, 350

Kistler, W. M. and De Zeeuw, C. I. (2002). Dynamical working memory and timed responses: The role of reverberating loops in the olivo-cerebellar system. *Neural Comput. (submitted)*. 160, 312, 320, 445

Kistler, W. M. and Gerstner, W. (2001). Stable propagation of activity pulses in populations of spiking neurons. *Neural Comput.*, in press. 351, 355

Kistler, W. M., Gerstner, W., and van Hemmen, J. L. (1997). Reduction of Hodgkin-Huxley equations to a single-variable threshold model. *Neural Comput.*, 9:1015–1045. 111, 125, 128, 129, 130, 131

Kistler, W. M., Seitz, R., and van Hemmen, J. L. (1998). Modeling collective excitations in cortical tissue. *Physica D*, 114:273–295. 119, 344, 345

Kistler, W. M. and van Hemmen, J. L. (1999). Short-term synaptic plasticity and network behavior. *Neural Comput.*, 11:1579–1594. 296, 309

Kistler, W. M. and van Hemmen, J. L. (2000a). Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials. *Neural Comput.*, 12:385–405. 206, 368, 373, 375, 415, 416, 424, 433, 436, 437, 438, 440

Kistler, W. M. and van Hemmen, J. L. (2000b). Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic potentials. *Neural Comput.*, 12:385–405. 427

Kistler, W. M., van Hemmen, J. L., and De Zeeuw, C. I. (2000). Time window control: A model for cerebellar function based on synchronization, reverberation, and time slicing. In Gerrits, N. M., Ruigrok, T. J. H., and De Zeeuw, C. I., editors, *Cerebellar Modules: Molecules, Morphology, and Function*, Progress in Brain Research, pages 275–297. Elsevier, Amsterdam. 52, 293, 320

Kjaer, T. W., Hertz, J. A., and Richmond, B. J. (1994). Decoding cortical neuronal signals: network models, information estimation and spatial tuning. *J. Comput. Neurosci.*, 1:109–139. 30

Knight, B. W. (1972a). Dynamics of encoding in a population of neurons. *J. Gen. Physiol.*, 59:734–766. 182, 183, 213, 226, 232, 256, 259, 277, 278, 279, 290, 291

Knight, B. W. (1972b). The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. *J. Gen. Physiol.*, 59:767–778. 279, 291

Knight, B. W. (2000). Dynamics of encoding in neuron populations: some general mathematical features. *Neural Comput.*, 12:473–518. 217, 229, 255

Knox, C. K. (1974). Cross-corrlation functions for a neuronal model. *Biophys. J.*, 14:567–582. 282, 285, 291

Knudsen, E. I., Blasdel, G. G., and Konishi, M. (1979). Sound localization by the barn owl (tyto alba) measured with the search coil technique. *J. Comp. Physiol.*, 133:1–11. 450

Koch, C. (1999). *Biophysics of Computation*. Oxford University Press, New York. 51, 76

Koch, C., Ö. Bernander, and Douglas, R. J. (1995). Do neurons have a voltage or a current threshold for action potential initiation? *J. Comput. Neurosci.*, 2:63–82. 48, 127

Koch, C. and Segev, I. (2000). The role of single neurons in information processing. *Nat. Neurosci.*, 3(Supp):1160–1211. 75

Kohonen, T. (1984). *Self-organization and associative memory*. Springer, Berlin Heidelberg New York. 367, 369, 393, 427

Konishi, M. (1986). Centrally synthesized maps of sensory space. *Trends Neurosci.*, 9:163–168. 372, 450, 452, 462

Konishi, M. (1993). Listening with two ears. *Sci. Am.*, pages 34–41. 372, 452, 462

Konnerth, A. and Eilers, J. (1994). Synaptic plasticity and calcium dynamics in cerebellar Purkinje neurons. *Biomed. Res.*, 15:73–77. Supplement 1. 387

Kopell, N. (1986). Symmetry and phase locking in chains of weakly coupled oscillators. *Commun. Pure Appl. Math.*, 39:623–660. 301

König, P., Engel, A. K., and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.*, 19(4):130–137. 194, 195, 209

König, P. and Schillen, T. B. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. synchronization. *Neural Comput.*, 3:155–166. 32, 293

Kree, R. and Zippelius, A. (1991). Asymmetrically diluted neural networks. In Domany, E., van Hemmen, J. L., and Schulten, K., editors, *Models of Neural Networks*. Springer, Berlin. 246, 247, 314, 315

Kreiter, A. K. and Singer, W. (1992). Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *Eur. J. Neurosci.*, 4:369–375. 32

Krüger, J. and Aiple, F. (1988). Multimicroelectrode investigation of monkey striate cortex: spike train correlations in the infragranular layers. *J. Neurophysiol.*, 60:798–828. 23

Kuffler, S. W., Nicholls, J. G., and Martin, A. R. (1984). *From neuron to brain.* Sinauer, Sunderland Mass., 2nd edition. 37

Kuramoto, Y. (1975). Self-entrainment of a population of coupled nonlinear oscillators. In Araki, H., editor, *International symposium on mathematical problems in theoretical physics*, pages 420–422, Berlin Heidelberg New York. Springer. 301

Laing, C. R. and Chow, C. C. (2001). Stationary bumps in networks of spiking neurons. *Neural Comput.*, 13:1473–1494. 337

Lansky, P. (1984). On approximations of Stein's neuronal model. *J. Theor. Biol.*, 107:631–647. 190

Lansky, P. (1997). Sources of periodical force in noisy integrate-and-fire models of neuronal dynamics. *Phys. Rev. E*, 55:2040–2043. 190

Lansky, P. and Smith, C. (1989). The effect of random initial value in neural first-passage-time models. *Mathematical Biosciences*, 93:191–215. 182, 183

Lapicque, L. (1907). Recherches quantitatives sur l'excitation electrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.*, 9:620–635. 38, 155

Latham, P. E., Richmond, B. J., Nelson, P. G., and Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. *J. Neurophysiol.*, 83:808–827. 106, 107

Laurent, G. (1996). Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends Neurosci.*, 19:489–496. 293, 322

Lestienne, R. (1996). Determination of the precision of spike timing in the visual cortex of anaesthetised cats. *Biol. Cybern.*, 74:55–61. 24

Lev-Ram, V., Jiang, T., Wood, J., Lawrence, D. S., and Tsien, R. Y. (1997). Synergies and coincidence requirements between NO, cGMP, and Ca^{++} in the induction of cerebellar long-term depression. *Neuron*, 18:1025–1038. 387

Levin, J. E. and Miller, J. P. (1996). Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, 380:165–168. 201

Levy, W. B. and Stewart, D. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in hippocampus. *Neurosci,*, 8:791–797. 364

Lewis, T. J. and Gerstner, W. (2001). Comparison of integrate-and-fire neurons: a case study. *Internal report*, xx:xx. 132

Linden, D. J. (1999). The return of the spike: postsynaptic action potentials and the induction of LTP and LTD. *Neuron*, 22:661–666. 388

Linden, D. J. and Connor, J. A. (1995). Long-term synaptic depression. *Annu. Rev. Neurosci.*, 18:319–357. 369, 394

Lindner, B. and Schimansky-Geier, L. (2001). Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys. Rev. Lett.*, 86:2934–2937. 259

Linsker, R. (1986a). From basic network principles to neural architecture: emergence of orientation columns. *Proc. Natl. Acad. Sci. USA*, 83:8779–8783. 393, 407

Linsker, R. (1986b). From basic network principles to neural architecture: emergence of orientation selective cells. *Proc. Natl. Acad. Sci. USA*, 83:8390–8394. 393, 407

Linsker, R. (1986c). From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. USA*, 83:7508–7512. 393, 407, 427

Lisman, J. (1989). A mechanism for Hebb and anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. USA*, 86:9574–9578. 387, 389

Llinás, R. (1988). The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function. *Science*, 242:1654–1664. 51

Llinás, R. (1991). The noncontinuous nature of movement execution. In Humphrey, D. R. and Freund, H.-J., editors, *Motor Control: Concepts and Issues*, pages 223–242. Wiley, New York. 293

Llinás, R. and Mühlethaler, M. (1988). Electrophysiology of guinea-pig cerebellar nuclear cells in the in vitro brain stem-cerebellar preparation. *J. Physiol. (Lond.)*, 404:241–258. 56

Llinás, R. and Yarom, Y. (1986). Oscillatory properties of guinea-pig inferior olivary neurones and their pharmacological modulation: An *in vitro* study. *J. Physiol. (Lond.)*, 376:163–182. 311

Longtin, A. (1993). Stochastic resonance in neuron models. *J. Stat. Phys.*, 70:309–327. 201

Maass, W. (1996). Lower bounds for the computational power of spiking neurons. *Neural Comput.*, 8:1–40. 31, 119

Maass, W. (1998). Computing with spiking neurons. In Maass, W. and Bishop, C. M., editors, *Pulsed Neural Networks*, chapter 2, pages 55–85. MIT-Press. 31, 119

Maass, W. and Bishop, C. (1998). *Pulsed Neural Networks*. MIT-Press. 38, 155

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *preprint*. 445

MacGregor, R. J. and Oliver, R. M. (1974). A model for repetitive firing in neurons. *Kybernetik*, 16:53–64. 111

MacKay, D. J. C. and Miller, K. D. (1990). Analysis of linsker's application of Hebbian rules to linear networks. *Network*, 1:257–297. 393, 407, 410, 427

Magee, J. C. and Johnston, D. (1997). A synaptically controlled associative signal for hebbian plastiticy in hippocampal neurons. *Science*, 275:209–213. 365, 432

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268:1503–1506. 157, 159

Malenka, R. C., Kauer, J. A., Zucker, R. S., and Nicoll, R. A. (1988). Postsynaptic calcium is sufficient for potentiation of hippocampal synaptic transmission. *Science*, 242:81–84. 387

Malinow, R., Schulman, H., and Tsien, R. W. (1989). Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of ltp. *Science*, 245:862–866. 387

Manis, P. B. and Marx, S. O. (1991). Outward currents in isolated ventral cochlear nucleus neurons. *J. Neurosci.*, 11:2865–2800. 454

Manwani, A. and Koch, C. (1999). Detecting and estimating signals in noisy cable structures, I: Neuronal noise sources. *Neural Comput.*, 11:1797–1829. 159

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215. 365, 366, 393, 432, 449

Markram, H. and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382:807–810. 160

Maršálek, P., Koch, C., and Maunsell, J. (1997). On the relationship between synaptic input and spike output jitter in individual neurons. *Proc. Natl. Acad. Sci. USA*, 94:735–740. 159, 271, 351

Mascaro, M. and Amit, D. J. (1999). Effective neural response function for collective population states. *Network*, 10:351–373. 248

Mattia, M. and Del Giudice, P. (2001). On the population dynamics of interacting spiking neurons. *Preprint*, xx:xx. 296

Mayer, M. L., Westbrook, G. L., and Guthrie, P. B. (1984). Voltage-dependent block by Mg^{2+} of nmda responses in spinal cord neurones. *Nature*, 309:261–263. 388

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133. 38, 209, 313, 314

McNamara, B. and Wiesenfeld, K. (1989). Theory of stochastic resonance. *Phys. Rev. A*, 39:4854–4869. 201

Mehta, M. R., Quirk, M., and Wilson, M. (2000). Experience-dependent asymmetric shap of hippocampal receptive fields. *Neuron*, 25:707–715. 430, 431

Meinhardt, H. (1995). *Algorithmic Beauty of Sea Shells*. Springer, New York. 357

Mel, B. W. (1994). Information processing in dendritic trees. *Neural Comput.*, 6(1031–1085). 72

Meyer, C. and van Vreeswijk, C. (2001). Temporal correlations in stochastic networks of spiking neurons. *Neural Comput.*, xx:xx. 235, 255

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity dependent competition between ON- and OFF-center inputs. *J. Neurosci.*, 14:409–441. 407, 409, 427

Miller, K. D. (1995). Receptive fields and maps in the visual cortex: Models of ocular dominance and orientation columns. In Domany, E., van Hemmen, J. L., and Schulten, K., editors, *Models of neural networks III*, pages 55–78. Springer, New York. 407, 427

Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science*, 245:605–615. 393, 407, 409, 427

Miller, K. D. and MacKay, D. J. C. (1994). The role of constraints in hebbian learning. *Neural Comput.*, 6:100–126. 368, 403, 404, 427

Miller, M. I. and Mark, K. E. (1992). A statistical study of cochlear nerve discharge patterns in reponse to complex speech stimuli. *J. Acoust. Soc. Am.*, 92:202–209. 176

Milner, P. M. (1974). A model for visual shape recognition. *Psychol. Rev.*, 81:521–535. 31

Minai, A. A. and Levy, W. B. (1993). Sequence learning in a single trial. In *INNS World Congress on Neural Networks II*, pages 505–508. Intern. Neural Network Soc. 427, 441

Mirollo, R. E. and Strogatz, S. H. (1990). Synchronization of pulse coupled biological oscillators. *SIAM J. Appl. Math.*, 50:1645–1662. 301, 322

Moiseff, A. and Konishi, M. (1981). xxx. *J. Comp. Physiol.*, 144:299–304. 450

Montague, P. R., Dayan, P., Person, C., and Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature*, 377:725–728. 419, 431, 433

Moore, G. P., Segundo, J. P., Perkel, D. H., and Levitan, H. (1970). Statistical signs of synaptic interaction in neurons. *Biophys. J.*, 10:876–900. 282, 285

Morris, C. and Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.*, pages 193–213. 78

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatosensory cortex. *J. Neurophysiol.*, 20:408–434. 24, 213

Murray, J. D. (1993). *Mathematical Biology*. Springer, Berlin. 99, 357

Nagumo, J. S., Arimoto, S., and Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proc. IRE*, 50:2061–2070. 79

Nelson, M. and Rinzel, J. (1995). The Hodgkin-Huxley model. In Bower, J. M. and Beeman, editors, *The book of Genesis*, chapter 4, pages 27–51. Springer, New York.

Neltner, L., Hansel, D., Mato, G., and Meunier, C. (2000). Synchrony in heterogeneous networks of spiking neurons. *Neural Comput.*, 12:1607–1641. 296

Ngezahayo, A., Schachner, M., and Artola, A. (2000). Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *J. Neurosci.*, 20:2451–2458. 371, 373

Nowak, L., Bregestovski, P., Asher, P., Herbet, A., and Prochiantz, A. (1984). Magnesium gates glutamate-activiated channels in mouse central neurons. *Nature*, 307:462–465. 388

Nützel, K. (1991). The length of attractors in asymmetric random neural networks with deterministic dynamics. *J. Phys. A: Math. Gen.*, 24:L151–L157. 246, 314, 319

Nützel, K., Kien, J., Bauer, K., Altman, J. S., and Krey, U. (1994). Dynamics of diluted attractor neural networks with delays. *Biol. Cybern.*, 70:553–561. 160, 312

Nykamp, D. and Tranchina, D. (2000). A population density approach that facilitates large-scale modeling of neural networks: Analysis and application to orientation tuning. *J. Comput. Neurosci.*, 8:19–50. 213, 217, 220, 229, 256

Obermayer, K., Blasder, G. G., and Schulten, K. (1992). Statistical-mechanics analysis of self-organization and pattern formation during the development of visual maps. *Phys. Rev. E*, 45:7568–7589. 393

Oertel, D. (1983). Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus. *J. Neurosci.*, 3(10):2043–2053. 454

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.*, 15:267–273. 368, 370, 405, 427

O'Keefe, J. (1993). Hippocampus, theta, and spatial memory. *Curr. Opin. Neurobiol.*, 3:917–924. 31, 293

O'Keefe, J. (1993). Hippocampus, theta, and spatial memory. *Curr. Opin. Neurobiol.*, 3:917–924. 322

Omurtag, A., Knight, B. W., and Sirovich, L. (2000). On the simulation of a large population of neurons. *J. Comput. Neurosci.*, 8:51–63. 213, 217, 220, 229

Optican, L. M. and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. 3. Information theoretic analysis. *J. Neurophysiol.*, 57:162–178. 30

Oram, M. W., Wiener, M. C., Lestienne, R., and Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J. Neurophysiol.*, 81:3021–3033. 24, 351, 445

Palm, G. and Poggio, T. (1977). The Volterra representation and the Wiener expansion: Validity and pitfalls. *SIAM J. Appl. Math.*, 33:195–216. 377

Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. McGraw-Hill, New York. 161

Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford Univ. Press. 432

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967a). Neuronal spike trains and stochastic point processes I. the single spike train. *Biophys. J.*, 7:391–418. 164, 169, 209, 255

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967b). Neuronal spike trains and stochastic point processes II. simultaneous spike trains. *Biophys. J.*, 7:419–440. 164, 169

Pham, J., Pakdaman, K., Champagnat, J., and Vibert, J.-F. (1998). Activity in sparsely connected excitatory neural networks: effect of connectivity. *Neural Networks*, 11:415–434. 254

Pinto, D. J., Brumberg, J. C., Simons, D. J., and Ermentrout, G. B. (1996). A quantitative population model of whiskers barrels: re-esamining the Wilson-Cowan equations. *J. Comput. Neurosci.*, 3:247–264. 238

Plesser, H. E. (1999). *Aspects of Signal Processing in Noisy Neurons*. PhD thesis, Georg-August-Universität, Göttingen. 190, 202

Plesser, H. E. (2000). The ModUhl software collection. Technical report, MPI für Strömungsforschung, Göttingen. http://www.chaos.gwgd.de/ plesser/ModUhl.htm. 194

Plesser, H. E. and Gerstner, W. (2000). Noise in integrate-and-fire neurons: from stochastic input to escape rates. *Neural Comput.*, 12:367–384. 194, 198, 199, 200, 201, 202, 276

Plesser, H. E. and Tanaka, S. (1997). Stochastic resonance in a model neuron with reset. *Phys. Lett. A*, 225:228–234. 194

Poliakov, A. V., Powers, R. K., and Binder, M. C. (1997). Functional identification of input-output transforms of motoneurons in cat. *J. Physiol.*, 504:401–424. 281, 282, 284, 291

Poliakov, A. V., Powers, R. K., Sawczuk, A., and Binder, M. C. (1996). Effects of background noise on the response of rat and cat motoneurones to excitatory current transients. *J. Physiol.*, 495:143–157. 115, 284, 286

Powers, R. K. D. B. and Binder, M. D. (1996). Experimental evaluation of input-output models of motoneuron discharges. *J. Neurophysiol.*, 75:367–379. 112, 114, 115

Rall, W. (1989). Cable theory for dendritic neurons. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling*, pages 9–62, Cambridge. MIT Press. 66

Ramòn y Cajal, S. (1909). *Histologie du système nerveux de l'homme et des vertébré*. A. Maloine, Paris. 12

Rao, R. P. N. and Sejnowski, T. J. (2001). Spike-timing dependent Hebbian plasticity as temporal difference learning. *Neural Comput.*, 13:2221–2237. 419

Reich, D. S., Victor, J. D., and Knight, B. W. (1998). The power ratio and the interval map: spiking models and extracellular recordings. *J. Neurosci.*, 18(23):10090–10104. 161

Rescorla, R. A. and Wagner, A. R. (1972). A theory of pavlovian conditioning: variations in the effectiveness of reinforecement and nonreinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical Conditioning II: current research and theory*, pages 64–99. Appleton Century Crofts, New York. 431

Reyes, A. D., Rubel, E. W., and Spain, W. J. (1994). Membrane properties underlying the firing of neurons in the avian cochlear nucleus. *J. Neurosci.*, 14(9):5352–5364. 453, 454

Reyes, A. D., Rubel, E. W., and Spain, W. J. (1996). submitted. *preprint*, xx:xx. 453

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1996). *Spikes - Exploring the neural code*. MIT Press, Cambridge, MA. 24, 25, 26, 33, 34, 36, 38, 209, 372

Rinzel, J. (1985). Excitation dynamics: insights from simplified membrane models. *Federation Proc.*, 44:2944–2946. 80

Rinzel, J. and Ermentrout, G. B. (1989). Analysis of neuronal excitability and oscillations. In Koch, C. and Segev, I., editors, *Methods in neuronal modeling*. MIT Press, Cambridge, MA. 48, 99, 127

Rinzel, J., Terman, D., Wang, X., and Ermentrout, B. (1998). Propagating activity patterns in large-scale inhibitory neuronal networks. *Science*, 279:1351–1355. 345

Risken, H. (1984). *The Fokker Planck equation: methods of solution and applications*. Springer-Verlag, Berlin. 256

Ritz, R., Gerstner, W., and van Hemmen, J. L. (1994). A biologically motivated and analytically soluble model of collective oscillations in the cortex: II. Application to binding and pattern segmentation. *Biol. Cybern.*, 71:349–358. 32

Ritz, R. and Sejnowski, T. J. (1997). Synchronous oscillatory activity in sensory systems: new vistas on mechanisms. *Curr. Opin. Neurobiol.*, 7:536–546. 32, 322

Roberts, P. D. (1999). Computational consequences of temporally asymmetric learning rules: I. Differential Hebbian learning. *J. Comput. Neurosci.*, 7:235–246. 419, 427

Roberts, P. D. (2000). Dynamics of temporal learning rules. *Phys. Rev. E*, 62:4077. 427

Roberts, P. D. and Bell, C. C. (2000). Computational consequences of temporally asymmetric learning rules: II. Sensory image cancellation. *J. Comput. Neurosci.*, 9:67–83. 427, 447, 448, 449

Rospars, J. P. and Lansky, P. (1993). Stochastic model neuron without resetting of dendritic potential: application to the olfactory system. *Biol. Cybern.*, 69:283–294. 145

Rubin, J., Lee, D. D., and Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric Hebbian plasticity. *Phys. Rev. Lett.*, 86:364–367. 423, 424, 425, 427

Ruf, B. and Schmitt, M. (1997). Unsupervised learning in networks of spiking neurons using temporal coding. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Artificial neural networks - ICANN'97*. Springer-Verlag, Heidelberg. 427

Sanders, J. A. and Verhulst, F. (1985). *Averaging methods in nonlinear dynamics systems*. Springer-Verlag, New York. 402

Schiegg, A., Gerstner, W., and van Hemmen, J. L. (1995). Intracellular Ca^{2+} stores can account for the time course of LTP induction: A model of Ca^{2+} dynamics in dendritic spines. *J. Neurophysiol.*, 74:1046–1055. 387, 389

Schillen, T. B. and König, P. (1991). Stimulus-dependent assembly formation of oscillatory responses: Ii. desynchronization. *Neural Comput.*, 3:167–178. 32

Schneidman, E., Freedman, B., and Segev, I. (1998). Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Comput.*, 10:1679–1703. 159, 160

Schrödinger, E. (1915). Zur Theorie der Fall- und Steigversuche and Teilchen mit Brownscher Bewegung. *Physikal. Zeitschrift*, 16:289–295. 194

Schultz, W., Dayan, P., and Montague, R. R. (1997). A neural substrate for prediction and reward. *Science*, 275:1593–1599. 431, 432, 433

Segundo, J. P., Moore, G. P., Stensaas, L. J., and Bullock, T. H. (1963). Sensitivity of neurons in Aplysia to temporal patterns of arriving impulses. *J. Exp. Biol.*, 40:643–667. 153

Sejnowski, T. (1977). Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.*, 4:303–321. 427

Sejnowski, T. J. (1999). The book of hebb. *Neuron*, 24:773–776. 393

Sejnowski, T. J. and Tesauro, G. (1989). The hebb rule for synaptic plasticity: algorithms and implementations. In Byrne, J. H. and Berry, W. O., editors, *Neural Models of Plasticity*, chapter 6, pages 94–103. Academic Press. 366, 369, 394, 427

Senn, W., Schneider, M., and Ruf, B. (2001a). Activity-dependent development of axonal and dendritic delays or, why synaptic transmission should be unreliable. *Neural Comput.*, xx:xx. 442, 450, 457

Senn, W., Tsodyks, M., and Markram, H. (1997). An algorithm for synaptic modification based on exact timing of pre- and post-synaptic action potentials. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Artificial Neural Networks – ICANN '97*, pages 121–126. Springer. 383

Senn, W., Tsodyks, M., and Markram, H. (2001b). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Comput.*, 13:35–67. 383, 384, 386, 427

Senn, W., Wyler, K., Streit, J., Larkum, M., Lüscher, H.-R., Merz, F., Mey, H., Müller, L., Steinhauser, D., Vogt, K., and Wannier, T. (1996). Dynamics of a random neural network with synaptic depression. *Neural Networks*, 9:575–588. 254

Shadlen, M. N. and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.*, 4:569–579. 24, 26, 38, 157, 158, 188, 194, 195, 209

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. 188

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423. reprinted in Shannon and Weaver 1949. 320

Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2001). A unified model of calcium dependent synaptic plasticity. *submitted*, xx:xx. 387, 389, 390, 391, 392

Shouval, H. Z. and Perrone, M. P. (1995). Post-Hebbian learning rules. In Arbib, M. A., editor, *The handbook of brain theory and neural networks*, pages 645–748. MIT-Press. 427

Siebert, W. M. and Gray, P. R. (1963). Random process model for the firing pattern of single auditory nerve fibers. *Q. Prog. Rep. Lab. Elec. MIT*, 71:241. 176

Singer, W. (1994). The role of synchrony in neocortical processing and synaptic plasticity. In Domany, E., van Hemmen, J. L., and Schulten, K., editors, *Models of Neural Networks II*, chapter 4. Springer. 24, 32, 293, 322

Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, 18:555–586. 322

Softky, W. and Koch, C. (1993). The highly irregular firing pattern of cortical cells is inconsistent with temporal integration of random epsps. *J . Neurosci.*, 13:334–350. 157, 158

Softky, W. R. (1995). Simple codes versus efficient codes. *Curr. Opin. Neurobiol.*, 5:239–247. 24, 26, 38, 209

Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1988). Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262. 248, 256

Sompolinsky, H. and Kanter, I. (1986). Temporal association in asymmetric neural networks. *Phys. Rev. Lett.*, 57:2861–2864. 441

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive hebbian learning through spike-time-dependent synaptic plasticity. *Nature Neurosci.*, 3:919–926. 368, 427, 430

Sotelo, C., Llinás, R., and Baker, R. (1974). Structural study of inferior olivary nucleus of the cat: Morphological correlations of electrotonic coupling. *J. Neurophysiol.*, 37:541–559. 311

Spiridon, M., Chow, C., and Gerstner, W. (1998). Frequency spectrum of coupled stochstic neurons with refractoriness. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *ICANN'98*, pages 337–342. Springer. 255

Spiridon, M. and Gerstner, W. (1999). Noise spectrum and signal transmission trough a population of spiking neurons. *Network: Comput. Neural Syst.*, pages 257–272. 235

Spiridon, M. and Gerstner, W. (2001). Effect of lateral connections on the accuracy of the population code for a network of spiking neurons. *preprint*. 337

Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophys. J.*, 5:173–194. 17, 38, 155, 185, 209

Stein, R. B. (1967a). The information capacity of nerve cells using a frequency code. *Biophys. J.*, 7:797–826. 155

Stein, R. B. (1967b). Some models of neuronal variability. *Biophys. J.*, 7:37–68. 38, 103, 111, 155, 185, 209, 255

Steinmetz, P. N. ., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K., and Niebur, E. (2000). Attention modultaes synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404:187–190. 32

Stemmler, M. (1996). A single spike suffices: the simplest form of stochastic resonance in model neurons. *Network*, 7:687–716. 201

Stevens, C. F. and Zador, A. M. (1998). Novel integrate-and-fire like model of repetitive firing in cortical neurons. In *Proc. of the 5th Joint Symposium on Neural Computation*, page Report. can be downloaded from http://www.sloan.salk.edu/ zador/publications.html. 112, 118

Strogatz, S. H. (1994). *Nonlinear dynamical systems and chaos*. Addison-Wesley, Reading MA. 99, 107

Stuart, G. J. and Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367:69–72. 388

Sullivan, W. E. and Konishi, M. (1984). Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl. *J. Neurosci.*, 4(7):1787–1799. 452

Suri, R. E. and Schutz, W. (2001). Temporal difference model reproduces anticipatory neural activity. *Neural Comput.*, 13:841–862. 432

Sutton, R. S. and Barto, A. G. (1981). Towards a modern theory of adaptive networks: expectation and prediction. *Psychol. Review*, 88:135–171. 431

Swindale, N. V. (1982). A model for the formation of orientation columns. *Proc. Royal Soc. London B*, 215:211–230. 393

Tanabe, M., Gäwiler, B. H., and Gerber, U. (1998). L-type Ca^{2+} channels mediate the slow Ca^{2+}-dependent afterhyperpolarization current in rat ca3 pyramidal cells in vitro. *J. Neurophysiol.*, 80:2268–2273. 59, 222

Terman, D. and Wang, D. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, 81:148–176. 32

Theunissen, F. and Miller, J. P. (1995). Temporal encoding in nervous systems: a rigorous definition. *J. Comput. Neurosci,*, 2:149–162. 34, 35, 36, 38

Thompson, R. F. (1993). *The brain.* W. H. Freeman and Company, New York, 2nd edition. 37

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520–522. 24, 26, 30, 269

Tovee, M. J. and Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Vis. Cognit.*, 2(1):35–58. 30

Tovee, M. J., Rolls, E. T., Treves, A., and Belles, R. P. (1993). Information encoding and the responses of single neurons in the primate visual cortex. *J. Neurophysiol.*, 70:640–654. 30

Traub, R. D., Wong, R. K. S., Miles, R., and Michelson, H. (1991). A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *J. Neurophysiol.*, 66:635–650. 59

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network*, 4:259–284. 213, 277, 290, 296

Treves, A., Rolls, E. T., and Simmen, M. (1997). Time for retrieval in recurrent associative memories. *Physica D*, 107:392–400. 31

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. 194, 195

Tsodyks, M., Mitkov, I., and Sompolinsky, H. (1993). Patterns of synchrony in inhomogeneous networks of oscillators with pulse interaction. *Phys. Rev. Lett.*, 71:1281–1283. 254, 296

Tsodyks, M. V. and Sejnowski, T. (1995). Rapid state switching in balanced cortical networks. *Network*, 6:111–124. 31, 188

Tuckwell, H. C. (1988). *Introduction to theoretic neurobiology*. Cambridge Univ. Press, Cambridge. 38, 146, 193, 196, 209

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Phys. Rev*, 36:823–841. 190

van Hemmen, J. L., Gerstner, W., Herz, A. V. M., Kühn, R., Sulzer, B., and Vaas, M. (1990). Encoding and decoding of patterns which are correlated in space and time. In Dorffner, G., editor, *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, pages 153–162, Berlin Heidelberg New York. Springer. 427

van Hemmen, J. L., Kistler, W. M., and Thomas, E. G. F. (2000). Calculation of Volterra kernels for solutions of nonlinear differential equations. *SIAM J. Appl. Math.*, 61:1–21. 377

van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam, 2nd edition. 174, 184, 189, 193, 199, 209, 425

van Rossum, M. C. W., Bi, G. Q., and Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.*, 20:8812–8821. 375, 423, 424, 425, 427

van Vreeswijk, C. (1996). Partially synchronized states in networks of pulse-coupled neurons. *Phys. Rev. E*, 54:5522–5537. 301, 310

van Vreeswijk, C., Abbott, L. F., and Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. *J. Comput. Neurosci.*, 1:313–321. 301, 303

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726. 160, 188, 246, 248, 256, 314

van Vreeswijk, C. and Sompolinsky, H. (1997). Irregular firing in cortical circuits with inhibition/excitation balance. In Bower, J., editor, *Computational Neuroscience: Trends in Reserach, 1997*, pages 209–213. Plenum Press, New York. 31

van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.*, 10:1321–1371. 246, 256

Verhulst, F. (1996). *Nonlinear Differential Equations and Dynamical Systems*. Springer, Berlin. 85, 93, 99

Volterra, V. (1959). *Theory of functionals and of integral and integro-differential equations*. Dover, New York. 377

von der Malsburg, C. (1973). Self-organization of orientation selective cells in the striate cortex. *Kybernetik*, 14:85–100. 427

von der Malsburg, C. (1981). The correlation theory of brain function. Internal Report 81–2, MPI für Biophysikalische Chemie, Göttingen. Reprinted in Models of Neural Networks II, Domany et al. (Eds.), Springer, 1994. 31, 32

von der Malsburg, C. and Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biol. Cybern.*, 67:233–242. 32

Vreeswijk, C. A. (2000). Stability of the asynchronous state in networks of nonlinear oscillators. *Phys. Rev. Lett.*, 84:5110–5113. 296

Wang, D. (1995). Emergent synchrony in locally coupled neural oscillators. *IEEE Trans. Neural Netw.*, 6:941–948. 32

Wang, D., Buhmann, J., and von der Malsburg, C. (1990). Pattern segmentation in associative memory. *Neural Comput.*, 2:94–106. 32, 293

Wehmeier, U., Dong, D., Koch, C., and van Essen, D. (1989). Modeling the mammalian visual system. In *Methods in Neuronal Modeling*, pages 335–359. MIT Press, Cambridge. 118

Weiss, T. F. (1966). A model of the peripheral auditory system. *Kybernetik*, 3:153–175. 17, 38, 111, 155, 177, 209

Welsh, J. P., Lang, E. J., Sugihara, I., and Llinás, R. (1995). Dynamic organization of motor control within the olivocerebellar system. *Nature*, 374:453–457. 293

White, J. A., Rubinstein, J. T., and Kay, A. R. (2000). Channel noise in neurons. *Trends Neurosci.*, 23:131–137. 159

Wiesenfeld, K. and Jaramillo, F. (1998). Minireview of stochastic resonance. *Chaos*, 8:539–548. 201, 209

Willshaw, D. J. and von der Malsburg, C. (1976). How patterned neuronal connections can be set up by self-organization. *Proc. R. Soc. (London) Ser. B*, 194:431–445. 393

Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12:1–24. 176, 177, 183, 207, 213, 217, 226, 232, 236, 238, 240, 256

Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55–80. 213, 254, 256, 325

Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261:1055–1058. 29

Wimbauer, S., Gerstner, W., and van Hemmen, J. L. (1998). Analysis of a correlation-based model for the development of orientation-selective receptive fields in the visual cortex. *Network*, 9:449–466. 410, 411

Wimbauer, S., Wenisch, O. G., Miller, K. D., and van Hemmen, J. L. (1997a). Development of spatiotemporal receptive fields of simple cells: I. model formulation. *Biol. Cybern.*, 77:453–461. 407, 409

Wimbauer, S., Wenisch, O. G., van Hemmen, J. L., and Miller, K. D. (1997b). Development of spatiotemporal receptive fields of simple cells: Ii. simulation and analysis. *Biol. Cybern.*, 77:463–477. 407

Wiskott, L. and Sejnowski, T. (1998). Constraint optimization for neural map formation: a nuifying framework for weight growth and normalization. *Neural Comput.*, 10:671–716. 427

Xie, X. and Seung, S. (2000). Spike-based learning rules and stabilization of persistent neural activity. In Solla, S. A., Leen, T. K., and Müller, K.-R., editors, *Advances in Neural Information Processing Systems 12*, page xx. MIT-Press, Cambridge. 419, 427

Yamada, W. M., Koch, C., and Adams, P. R. (1989). Multiple channels and calcium dynamics. In Koch, C. and Segev, I., editors, *Methods in neuronal modeling: From synapses to networks*, chapter 4. MIT Press, Cambridge, MA. 71

Zador, A., Koch, C., and Brown, T. H. (1990). Biophysical model of a hebbian synapse. *Proc. Natl. Acad. Sci. USA*, 87:6718–6722. 387, 389

Zador, A. M. (1998). Impact of synaptic unreliability on the information transmitted by spiking neuron. *J. Neurophysiol.*, 79:1219–1229. 159

Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M.-M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395:37–44. 365, 366, 375, 393, 432, 449