# SPIKING NEURON MODELS

## Single Neurons, Populations, Plasticity

Wulfram Gerstner
*Swiss Federal Institute of Technology, Lausanne*

Werner M. Kistler
*Erasmus University, Rotterdam*

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org



First published 2002

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Times 11/14pt. *System* LaTeX 2$_\varepsilon$ [DBD]

*A catalogue record of this book is available from the British Library*

ISBN 0 521 81384 0 hardback
ISBN 0 521 89079 9 paperback

# Contents

# 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 are

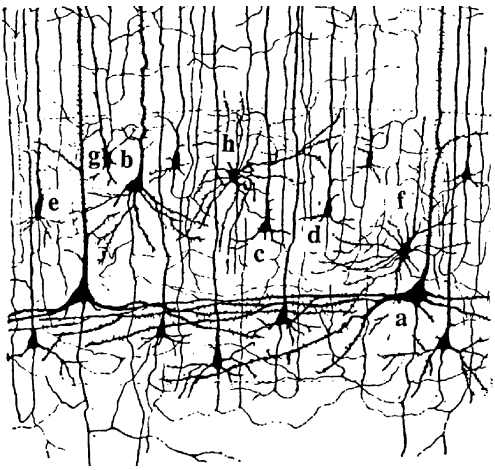


Fig. 1.1. This reproduction of a drawing of Ramón y Cajal shows a few neurons in the mammalian cortex that he observed under the microscope. Only a small portion of the neurons contained in the sample of cortical tissue have been made visible by the staining procedure; the density of neurons is in reality 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 Ramón y Cajal (1909).

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 nonlinear 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.

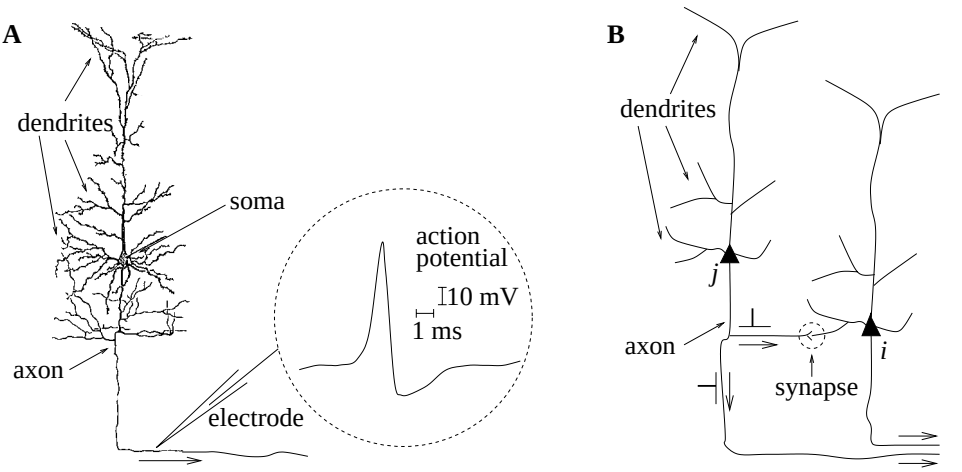


Fig. 1.2. **A**. Single neuron in a drawing by Ramón y Cajal. Dendrite, soma, and axon can be clearly distinguished. The inset shows an example of a neuronal action potential (schematic). The action potential is a short voltage pulse of 1–2 ms duration and an amplitude of about 100 mV. **B**. Signal 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).

Many of its axonal branches end in the direct neighborhood 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 membranes, called the synaptic cleft. When an action potential arrives at a synapse, it triggers a complex chain of biochemical processing steps that lead to the 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 already has 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.

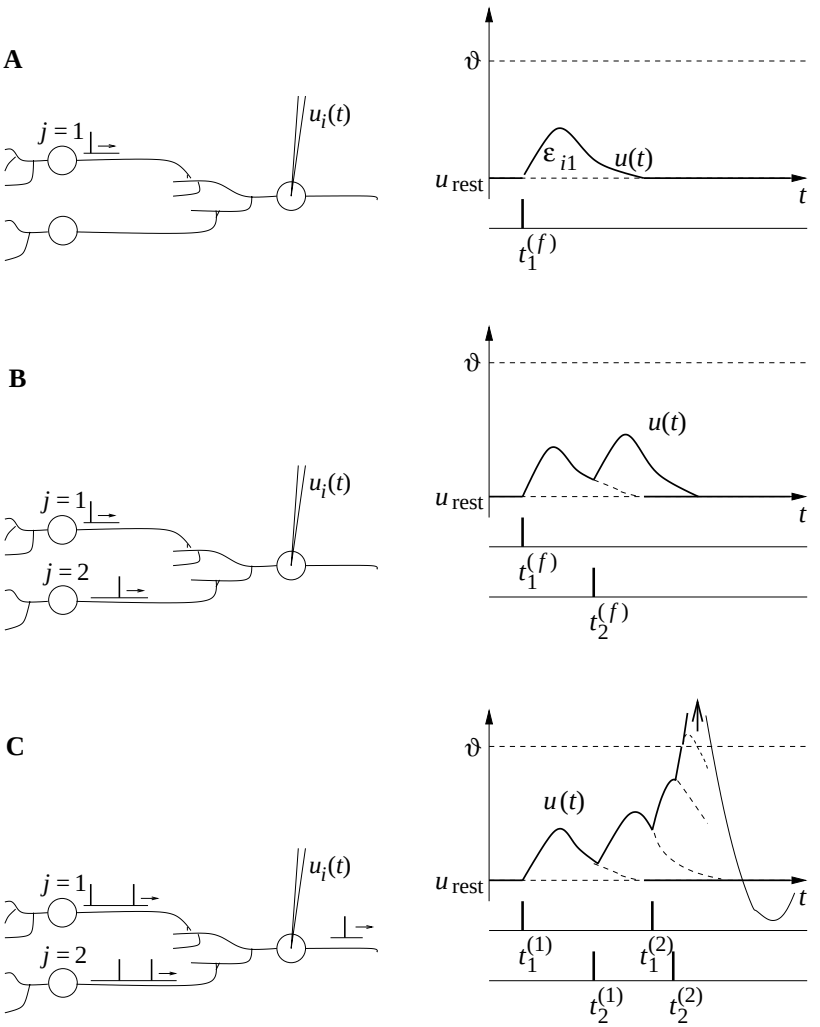


Fig. 1.3. A postsynaptic neuron $i$ receives input from two presynaptic neurons $j = 1, 2$. **A**. Each presynaptic spike evokes an excitatory postsynaptic potential (EPSP) that can be measured with an electrode as a potential difference $u_i(t) - u_{\rm rest}$. The time course of the EPSP caused by 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 $\vartheta$, 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.

### 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) PSP 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 PSP $\epsilon_{i1}$ or $\epsilon_{i2}$, 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.

However, linearity breaks down if too many input spikes arrive during a short interval. As soon as the membrane potential reaches a critical value $\vartheta$, 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 1 mV. The critical value for spike initiation is about 20–30 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 $\vartheta$. If $u_i(t)$ reaches $\vartheta$ 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 $\vartheta$ from below,

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

The term $\epsilon_{ij}$ in Eq. (1.3) describes the response of neuron $i$ to spikes of a presynaptic neuron $j$. The term $\eta$ in Eq. (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 Eqs. (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 SR Model are a useful conceptual framework for the analysis of neuronal dynamics and neuronal coding.

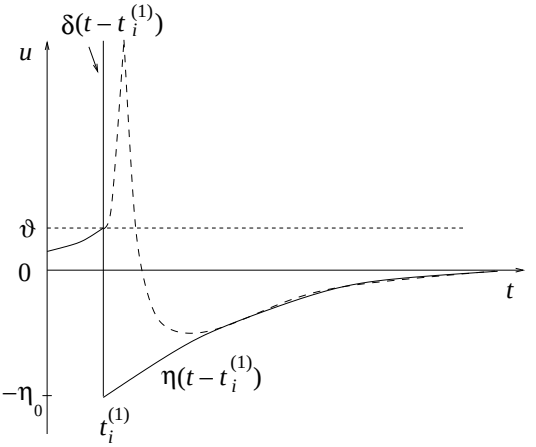


Fig. 1.4. In formal models of spiking neurons the shape of an action potential (dashed line) is usually replaced by a $\delta$ pulse (vertical line). The negative overshoot (spike-afterpotential) after the pulse is included in the kernel $\eta(t - t_i^{(1)})$ (thick line) which takes care of "reset" 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 $\eta$ during an action potential by, e.g., a square pulse, followed by a negative spike-afterpotential,

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

with parameters $\eta_0, \tau, \Delta t > 0$. In the limit of $\Delta t \to 0$ the square pulse approaches a Dirac $\delta$ 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-afterpotential 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 $\vartheta$ 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 $\delta(x)$ is the Dirac $\delta$ function with $\delta(x) = 0$ for $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x)\, dx = 1$. 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 $\text{SRM}_0$ 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.

Fast-spiking neurons form a second class of neurons. These neurons show no adaptation and can therefore be well approximated by the model $\text{SRM}_0$ introduced

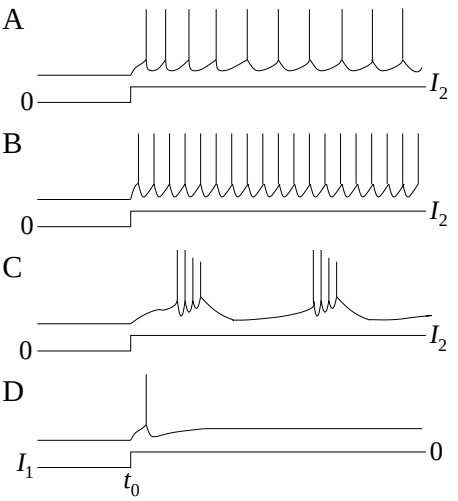


Fig. 1.5. Response to a current step. In **A**–**C**, the current is switched on at $t = t_0$ to a value $I_2 > 0$. Regular-spiking neurons (**A**) exhibit adaptation of the interspike intervals whereas fast-spiking neurons (**B**) show no adaptation. An example of a bursting neuron is shown in **C**.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

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 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 postinhibitory 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 $\text{SRM}_0$ 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.

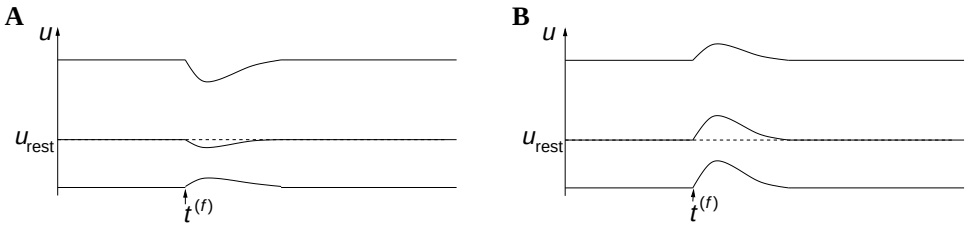


Fig. 1.6. The shape of postsynaptic potentials depends on the momentary level of depolarization. **A**. A presynaptic spike that arrives at time $t^{(f)}$ at an inhibitory synapse has hardly any effect on the membrane potential when the neuron is at rest, but a large effect if the membrane potential $u$ 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.

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 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 $E_{\text{syn}}$ 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 Section 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 $E_{\text{syn}}$ 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. Fig. 1.6A. However, if the membrane is depolarized to a value substantially above rest, the very same input spikes evoke a pronounced inhibitory potential. If the membrane is already hyperpolarized, the input spike can even produce a depolarizing effect. There is an intermediate value $u_0 = E_{\text{syn}}$ – the reversal potential – at which 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.

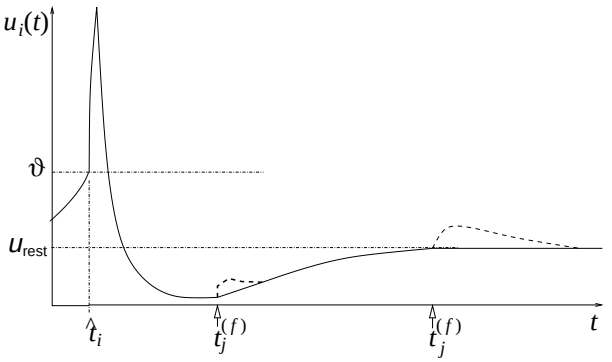


Fig. 1.7. The shape of postsynaptic potentials (dashed lines) depends on the time $t - \hat{t}_i$ that has passed since the last output spike current of neuron $i$. The postsynaptic spike has been triggered at time $\hat{t}_i$. A presynaptic spike that arrives at time $t_j^{(f)}$ shortly after the spike 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.

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".

The reversal potential for excitatory synapses is usually significantly above the resting potential. If the membrane is depolarized $u_0 \gg u_{\text{rest}}$ 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. Fig. 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 occurred 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 Section 2.2.

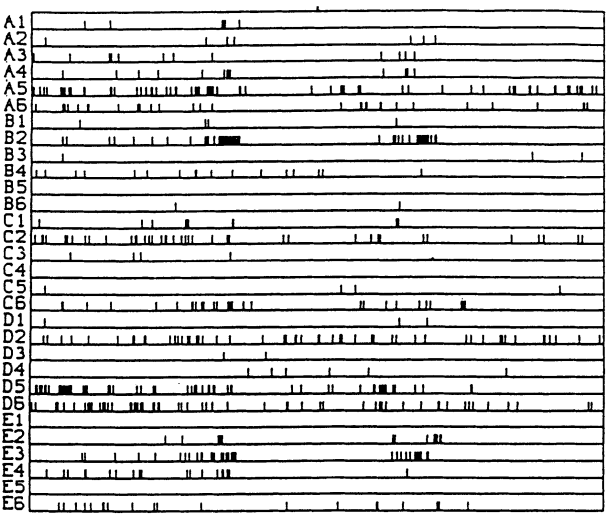


Fig. 1.8. Spatio-temporal pulse pattern. The spikes of 30 neurons (A1–E6, plotted along the vertical axes) are shown as a function of time (horizontal axis, total time is 4000 ms). The firing times are marked by short vertical bars. From Krüger and Aiple (1988).

*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 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 $\mathrm{SRM}_0$. 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 connected 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 30 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, $T = 100$ ms or $T = 500$ ms and counts the number of spikes $n_{\rm sp}(T)$ that occur in this time window. Division by the length of the time window gives the mean firing rate

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

usually reported in units of $\rm s^{-1}$ 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 the subject of an ongoing debate (Bialek et al., 1991; Abeles, 1994; Shadlen and Newsome, 1994; Hopfield, 1995; 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-

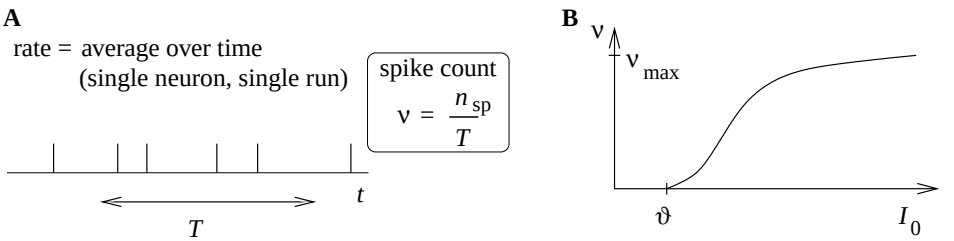


Fig. 1.9. **A**. Definition of the mean firing rate via a temporal average. **B**. Gain function, schematic. The output rate $\nu$ is given as a function of the total input $I_0$.

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.

## 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: an average over time, 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 is given elsewhere (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 classic 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 classic 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 that 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 ms (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 photoreceptors 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 $\nu$ 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 intracellular electrode a constant current $I_0$. The relation between the measured firing frequency $\nu$ and the applied input current $I_0$ is sometimes called the frequency–current curve of the neuron. In models, we

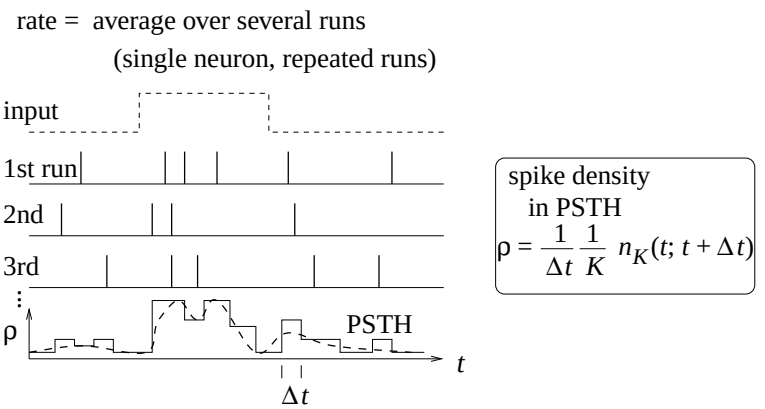


Fig. 1.10. Definition of the spike density in the peri-stimulus-time histogram (PSTH) as an average over several runs of the experiment. Taken from Gerstner (1998) with permission.

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 $\nu$ over long distances. In fact, the best coding scheme to transmit the value of the rate $\nu$ 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 elsewhere (Shadlen and Newsome, 1994; Softky, 1995; Rieke et al., 1996).

### *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 $\Delta t$ is typically in the range of one or a few milliseconds. The number of occurences 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 $\Delta t$

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 for evaluating neuronal activity, in particular in the case of time-dependent stimuli. The obvious problem with this approach is that it cannot be the decoding scheme used by neurons in the brain. Consider for example a frog which wants to catch a fly. It cannot 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 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

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

where $N$ is the size of the population, $n_{\text{act}}(t; t+\Delta t)$ the number of spikes (summed over all neurons in the population) that occur between $t$ and $t + \Delta t$, and $\Delta t$ a small

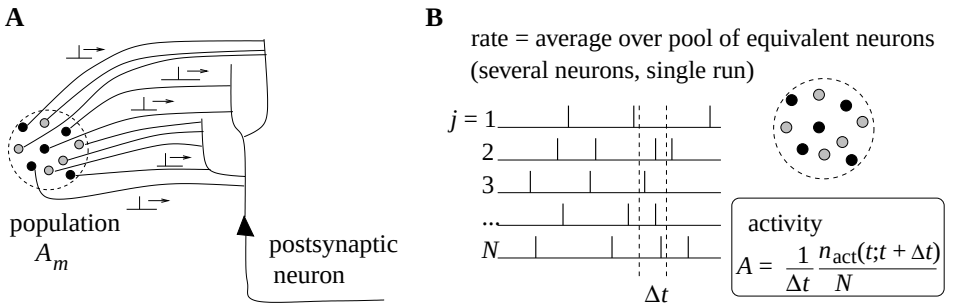


Fig. 1.11. **A**. A postsynpatic neuron receives spike input from the population $m$ with activity $A_m$. **B**. The population activity is defined as the fraction of neurons that are active in a short interval $[t, t + \Delta t]$ divided by $\Delta t$.

time interval; see Fig. 1.11. Eq. (1.10) defines a variable with units $\mathrm{s}^{-1}$ – in other words, a rate.

The population activity may vary rapidly and can reflect changes in the stimulus conditions nearly instantaneously (Gerstner, 2000; 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

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

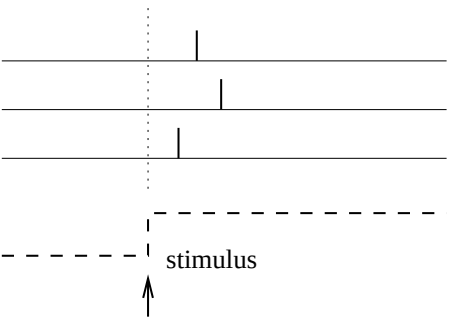


Fig. 1.12. Time-to-first spike. The spike train of three neurons are shown. The third neuron from the top is the first one to fire a spike after the stimulus onset (arrow). The dashed line indicates the time course of the stimulus.

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 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 photoreceptors 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.