From: F. Moss and S. Gielen (Editors), 'Handbook of Biological Physics (Vol. 4), Neuro-informatics, Neural Modelling', Elsevier Amsterdam 2001, 771-823

Theory of Synaptic Plasticity

J. Leo van Hemmen
Physik Department der TU München
D-85747 Garching bei München, Germany

Abstract

A general theory of synaptic plasticity is developed that is based on a precise timing of pre- and postsynaptic action potentials. As for long-term plasticity, the key notions are Hebbian learning and a 'learning window' through which presynaptic input and postsynaptic output of a neuron are related to each other and, hence, to the plasticity change they induce, viz., potentiation or depression. Short-term plasticity is treated on the basis of available presynaptic resources. The paper starts by analyzing formal neurons at various levels of activity and the way in which they allow storage and retrieval of spatiotemporal patterns. Wherever possible, the underlying neurobiology is indicated.

1 Introduction

What is synaptic plasticity? If something changes, the first question is: what changes, where, and how? Apparently, we have to focus on 'synapses' [1, 2, 3] but what happens there and why? These questions, though natural and, as we will see, fascinating, do not have simple answers and they need some biophysical background to be fully understood. They are important since in general a neuronal network stores its information in the synapses. In the present chapter we first provide the necessary background, then turn to the theory of information storage, ranging from simple activity patterns of formal neurons to long-term processes involving spatiotemporal patterns of realistic neurons, and finish the argument with a detailed account of short-term plasticity. References will be provided as we go along. Since we concentrate on storage of spatio*temporal* patterns the reader has to consult the literature [4, Ch. 13][5] for other learning rules and related results such as Sejnowski's seminal work [6].

First of all, what is synaptic plasticity and where precisely does it occur? Since we are concerned with biological neural nets we will now review the essentials of neuronal anatomy as far as they are relevant to our present purpose, viz., developing learning theory; see Koch [4, Chs. 4 & 13] for biological details that

are skipped here. Stated simply, a neuron consists of three parts: the dendritic tree gathers the input, the soma with the axon hillock is the "CPU" generating action potentials or spikes (Fig. 1) on the basis of the input provided, and the axon conveys this output to other neurons. It is important to realize that spikes are the only output of a cortical neuron. A spike lasts typically about one millisecond (1 ms). Synapses are the axonal terminals on the dendritic tree of other neurons. When an axon bifurcates, the amplitude of a spike does not decrease but remains constant, about 0.1 V. This is due to an active propagation process.

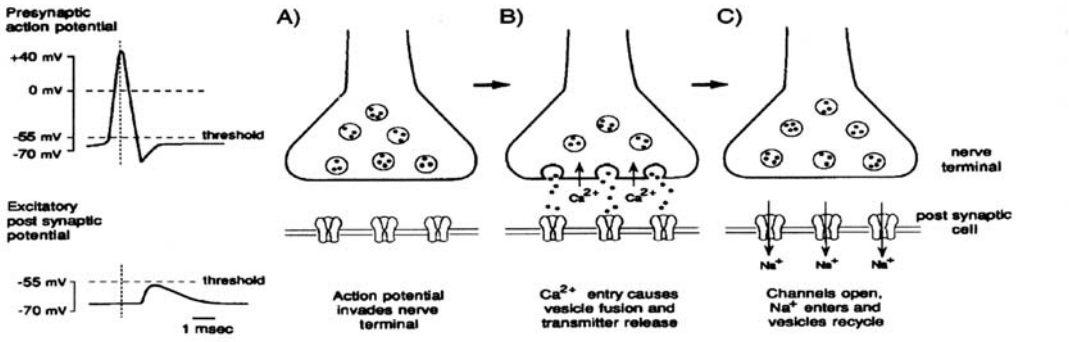

Figure 1*:* Synaptic transmission at a chemical synapse in a neuromuscular junction. An action potential (upper left) arrives at the presynaptic terminal (A), induces an influx of Ca^{2+} ions that cause the vesicles, round and filled with neurotransmitter molecules, to fuse with the membrane (B, exocytosis) and release their contents into the synaptic cleft separating the presynaptic membrane from the postsynaptic one. The neurotransmitter molecules diffuse across the cleft and bind to the receptors at the postsynaptic side. In this figure of an excitatory synapse, they induce the opening (C) of Na^+ channels, leading to an excitatory postsynaptic potential at the axon hillock of the soma (lower left); see also Fig. 2. In cortical synapses a similar process occurs, though it is more stochastic – cf. Eq. (1) – and the EPSPs have a smaller amplitude. Reprinted by permission [7].

Figure 2 is a picture of a pyramidal cell, a typical cortical neuron. An axon in general bifurcates several times but, as we have seen, spikes look the same everywhere. At the axonal ends one has synapses, which contact the dendritic trees of other neurons. It is here that neuronal information is stored. Most of the postsynaptic potentials (Fig. 1) rely on chemical transmission from a *presynaptic* axonal terminal to a *postsynaptic* dendritic spine. A cleft, 20–40 nm wide, separates the two parts. When a spike arrives, calcium ions enter the axonal terminal, where vesicles with 30–40 nm diameter and filled with a few thousand neurotransmitter molecules are waiting. As a consequence of the Ca^{2+} influx, these then move to the membrane bordering the synaptic cleft. In cortical synapses at most a few of them will fuse with the membrane at one of a small number n (say, 1, 2, or 3) of fixed release sites and emit their contents into the cleft, a process called

exocytosis [13, 14]. Because of the transmitter-mediated opening of postsynaptic ion channels this kind of synapse is called 'chemical'. In contrast, some synapses have electrical transmission; e.g. that of the Mauthner cell [15, 16, 17].

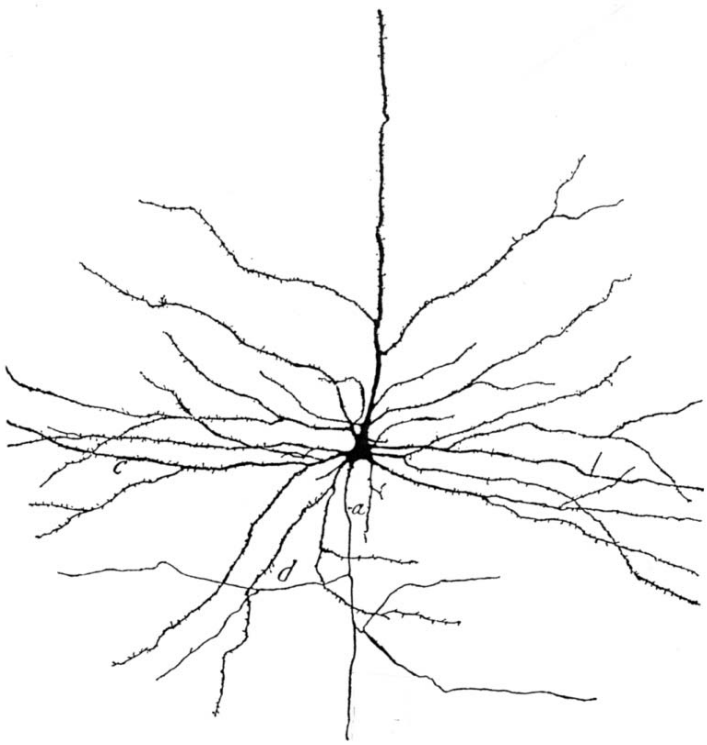

Figure 2*:* This wonderful picture of a pyramidal cell from the motor cortex of a 30-year-old man stems from Ramón y Cajal [10], who drew it more than a century ago; whole-cell staining after Golgi. (a) Axon via which action potentials (spikes) leave the "pyramidal" soma of the neuron (hence its name) to synapses on dendritic trees of other neurons, (c) dendrites that gather the postsynaptic currents from synapses, terminals of axons coming from elsewhere, (d) axonal collaterals that branch off.

The released neurotransmitter molecules diffuse across the synaptic cleft, which takes about 10μs, and dock into receptors of ion channels, which are then opened so that ions enter the postsynaptic dendritic spine, if any. If the ions are positive, e.g., Na^+, then the dendritic tree of the receiving neuron experiences a positive influx, which reappears as an excitatory postsynaptic potential (EPSP) at the soma; see Fig. 1. This is what we will concentrate on, though most of the considerations below are much more general. For the sake of simplicity we describe an EPSP stemming from a spike arriving at a synapse connecting an axon from neuron j to the dendritic tree of neuron i by $J_{ij}\varepsilon(t)$. Here $\varepsilon(t) \geq 0$ is a fixed response function with maximum 1, t denotes time, and J_{ij} is called the "strength" or *efficacy* of a synapse connecting neuron j to neuron i. In cortex we

typically have one, at most two connections [4, §4.2], [9, Sects. 20 & 33]. To simplify the notation and without loss of generality we therefore assume that for each pair $\{i, j\}$ there is at most one synaptic connection. Synaptic plasticity refers to changes of J_{ij} as time proceeds. All this looks simple and straightforward but there is a stochastic complication, which can be eliminated.

We have three ingredients of a synaptic response [4, 8]:

- n presynaptic release sites (or active zones); for cortical synapses, n is a small number near to 1.

- the probability $0 \leq p \leq 1$ that a vesicle will fuse with the membrane at a specific release site and emit a neurotransmitter 'quantum' into the synaptic cleft; there is strong evidence [8] that the probability of getting more than one vesicle is negligibly small. For a cortical synapse $p = 0.3$ is not exceptional so that with $n = 1$ the number of neurotransmitter quanta present after a spike has arrived is a mere, though fairly well-educated, guess. The release processes at different sites can be taken to be independent so that a binomial distribution governs the probability $p(n, k)$ of getting the release of $0 \leq k \leq n$ quanta,

$$p(n, k) = \binom{n}{k} p^k (1 - p)^{n-k} \ . \tag{1}$$

  One can verify that the *mean* number of quanta is

$$\langle k \rangle = \sum_{k=0}^{n} k \, p(n, k) = np \ . \tag{2}$$

  Consequently, J_{ij} as determined by (1) is a stochastic variable itself.

- A quantum induces a *post*synaptic response. A succinct notation is Q. We take Q to be the maximum of an excitatory postsynaptic potential (EPSP) generated by a single quantum, i.e., we look for the maximal response as a function of time. For inhibitory synapses the learning ability is at the moment less clear but it seems that, if they can change their efficacy, they do so in a way analogous to their excitatory counterparts so that their effect can be measured by the minimum of an inhibitory postsynaptic potential (IPSP). Throughout what follows we will focus on excitatory synapses, leaving the realization of *mutatis mutandis* for the inhibitory ones to the reader.

Altogether the mean postsynaptic response induced by a spike is $\mathcal{R} = npQ$ but this is of no help when a *specific* spike arrives since its effect is in general never the mean. Why then compute the mean response? The input of a cortical neuron is provided by many synapses; a typical number is 10^4. As for the simple

stochastic process (1), it can be taken to be independent for different synapses. At time t the potential $v_i(t)$ at the axon hillock, the "CPU" of neuron i, is to first approximation a sum of the different postsynaptic potentials,

$$v_i(t) = \sum_{j(\neq i),f} J_{ij}\varepsilon(t - t_j^f) \tag{3}$$

where f in t_j^f labels all spikes of neuron j; the postsynaptic potential $\varepsilon(t)$ is causal in that it vanishes for $t < 0$ so that future events do not influence the present. For the time being axonal delays are incorporated in the ε. Let us suppose that neuron i has N 'synaptic' neighbors j and let us consider $v_i(t)/N$ for a moment,

$$v_i(t)/N = N^{-1} \sum_{j(\neq i)} J_{ij}\varepsilon(t - t_j^f) \approx N^{-1} \sum_{j(\neq i)} \langle J_{ij}\rangle \varepsilon(t - t_j^f) \ , \tag{4}$$

where the synaptic strengths J_{ij} are independent stochastic variables and the interspike intervals are assumed to be in the millisecond range, an every-day fact. The last equality is exact in the limit $N \to \infty$ and, as such, a consequence of the strong law of large numbers [11, 12]. A sum that can be replaced by its average is called *self-averaging* – a most valuable property.

In our case N *is* large so that, by the central limit theorem in conjunction with the law of the iterated logarithm (see Appendix A), the approximate equality in (4) is an equality up to an error of order $\mathcal{O}(1/\sqrt{N})$ that is the deviation from the mean and has a Gaussian distribution; the contents of the "laws of large numbers" as referred to above have been listed in Appendix A. In fact, we can allow ε to depend on both i and j while the J_{ij} may have any distribution with finite second moment. Then we have that, for i fixed, the $f_j := J_{ij}\varepsilon_{ij}$ are independent but not identically distributed random variables. Nevertheless the strong law of large numbers and the central limit theorem still hold [11, 12]. They even do so, when the f_j are weakly dependent in the sense that, given two positions j and k, their correlation functions $\langle (f_j - \langle f_j\rangle)(f_k - \langle f_k\rangle)\rangle$ decrease fast enough as a function of $|j - k|$. For (3) the above statements hold as well, provided one multiplies everything in sight by N and realizes that the deviation from the (nonzero) mean can be estimated by the law of the iterated logarithm [11, 12]; cf. Appendix A.

At a first sight one might object that the J_{ij} govern the dynamics of the system as a whole and, therefore, in the long run the latter induces dependencies among the former. That may well be but is completely irrelevant since the local 'gambling' we are considering here is that of vesicles fusing with the cellular membrane, i.e., exocytosis, a process lasting for a millisecond or less [13, 14]. On this time scale fusion processes are independent both inside and outside a synapse.

Despite the randomness, we arrive at the pleasant result that the leading

contribution to $v_i(t)$ is deterministic and given by

$$v_i(t) = \sum_{j(\neq i)} \langle J_{ij} \rangle \varepsilon(t - t_j^f) \tag{5}$$

as long as most of the expectation values $\langle J_{ij} \rangle$ are nonzero. That is to say, $\langle J_{ij} \rangle = n_{ij} p_{ij} Q_{ij} \neq 0$ where $\{ij\}$ labels the synapse; a generic one will carry no label. If nonzero, we can stick to studying synaptic plasticity in its dependence upon npQ. If a synaptic strength vanishes, the synapse can, and will, be dropped. According to the present state of the art [4], long-term synaptic plasticity lasting for hours means that both p and Q change, whereas short-term synaptic plasticity lasting for seconds is equivalent to saying that p alone changes. In both cases, n seems to be fixed, though Bonhoeffer et al. [18] have found that in LTP n may change as well. We will present a theory of synaptic plasticity that incorporates both long- and short-term effects. As a side remark we note there is also channel noise, which is generated by random gating of voltage-gated ion channels. It is different from synaptic noise, as has been explained in detail elsewhere [20], and can in principle be handled similarly to (4).

The synaptic action considered so far is called *ionotropic* as it is governed by postsynaptic ionic channels. It is fast and implements computations underlying, for example, rapid perception and motor control. Here, then, are two classes of (glutamate) receptors determining the state (open/closed) of their underlying ionic channels: NMDA (N-methyl-D-aspartate) and non-NMDA. The name NMDA is that of an agonist absent from the brain itself but used by neurobiologists to discern them. The NMDA receptors do need a strong depolarization to become active, e.g., through a positive potential change stemming from a postsynaptic spike [21] [4, Ch. 19]. NMDA receptors are 10 times slower than their neighboring non-NMDA counterparts. Furthermore, they are important to long-term potentiation (LTP) since they allow Ca^{2+} ions to enter the cell – in addition to Na^+ and K^+. On the other hand, the non-NMDA receptors convey the fast excitatory traffic that has to pass in a few milliseconds. Their typical EPSP is that of an *alpha function*,

$$\varepsilon(t) := (t/\tau) \exp(1 - t/\tau) \tag{6}$$

having its maximum 1 at $t = \tau$, and $\tau \approx 5\,\mathrm{ms}$. Of course ε is causal in that it vanishes for $t < 0$. For an extensive discussion of this and other types of response function the reader is referred to Gerstner's chapter in this book [27].

In addition to ionotropic receptors, which open ionic channels that permit a certain type of ion to cross the postsynaptic membrane, there are also *metabotropic* receptors where binding of a neurotransmitter leads to the activation of a second messenger such as Ca^{2+} ions. The messenger molecules then have to diffuse to particular ionic channels, which is a relatively slow process. The action of metabotropic receptors can, and usually will, extend over a long distance both

in space and in time. We will not treat them here but refer to the literature [4, 22] for further details concerning both types of receptor. Instead we turn to a simplification of the mathematical description of spike generation, the Spike Response Model (SRM).

2 Spike Response Model

Spikes require a mathematically intricate description that can be summarized by the following system of equations (7) - (9). The key variable is is the membrane potential V, which can be measured,

$$I_m = C_m \frac{\mathrm{d}}{\mathrm{d}t} V + I_{\text{channel}} + I_{\text{ext}} \ . \tag{7}$$

Here I_m is the total current, C_m is a membrane capacitance, I_{ext} is an externally applied current, e.g., due to synaptic input, while currents I_ℓ through the $1 \leq \ell \leq n$ ionic channels give

$$I_{\text{channel}} = \sum_{\ell=1}^{n} I_\ell \qquad \text{with} \qquad I_\ell = g_\ell(\mathbf{m})(V - E_\ell) \ . \tag{8}$$

The conductances $g_\ell = g_\ell(\mathbf{m})$ depend on a vector $\mathbf{m}$. The heart of nearly any differential-equation model producing spikes is a set of *auxiliary variables* $\mathbf{m} = (m_\heartsuit)$, each of the components $m_\heartsuit$ satisfying an ordinary differential equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} m_\heartsuit = [m_{\heartsuit,\infty}(V) - m_\heartsuit]/\tau_\heartsuit(V) \ . \tag{9}$$

For constant V, the variable $m_\heartsuit$ relaxes to $m_{\heartsuit,\infty}(V)$ at the rate $1/\tau_\heartsuit(V)$. In general the functions $m_{\heartsuit,\infty}(V)$ and $\tau_\heartsuit(V)$, which both depend on the membrane potential V, are the result of an extensive fitting procedure, the most famous one being that of Hodgkin and Huxley (1952), who got the Nobel price for their ingenious fit [4, 23, 24] for function sets $\{g_\ell(\mathbf{m}), m_{\heartsuit,\infty}(V), \tau_\heartsuit(V)\}$ describing two active ionic channels (Na^+ and K^+) and three auxiliary variables. In their notation, $\mathbf{m} = (m, n, h)$ and $g_{\text{Na}}(\mathbf{m}) = \bar{g}_{\text{Na}} m^3 h$, $g_{\text{K}}(\mathbf{m}) = \bar{g}_{\text{K}} n^4$, where $\bar{g}_{\text{Na}}$ and $\bar{g}_{\text{K}}$ are constants, as is the third conductance describing a 'leak' current. It is good to realize that only V, I_{channel}, and I_{ext} are accessible to experiment whereas the auxiliary variables $m_\heartsuit$ are not.

It is typical to real neurons and also to all these auxiliary-variable models that they produce spikes when, to excellent approximation, the potential V exceeds a threshold ϑ. A glance at the equations suffices to convince any reader that this statement is not evident. In fact, it is the result of the careful fit alluded to in the previous paragraph. A second glance at the upshot of what a spike with a 1 ms width induces at a postsynaptic neuron, viz., an EPSP such as the

alpha function in (6) with a several ms width, may then suffice to let the beholder wonder whether the precise form of a spike is really important to what it induces. Most of the time it is not and one can stick to a simplification, the *Spike Response Model* [25, 26, 27]. This is what we now focus on.

Let us discretize time and break the continuous time axis into parts of length $\Delta t = 1\,\mathrm{ms}$. We then write $t = 1, 2, \ldots$ and specify the state of neuron i by a Boolean variable $n_i(t) \in \{0, 1\}$: $n_i(t) = 1$ when the neuron fires, $n_i(t) = 0$ when it does not. Looking at synapse $\{i, j\}$ with synaptic strength J_{ij}, we note that it induces an EPSP $J_{ij}\varepsilon(t - t_i^f - \Delta_{ij}^{\mathrm{ax}})$ at neuron i for a spike that arose at neuron j at time t_j^f and was delayed by $\Delta_{ij}^{\mathrm{ax}}$ ms, the axonal delay that occurs when the spike travels along the axon from neuron j to the synaptic terminal $\{i, j\}$. Throughout what follows we write J_{ij} instead of $\langle J_{ij} \rangle$ since the membrane potential $v_i(t)$ at the axon hillock (our "CPU") is given by (5), a sum of many terms.

Once a neuron has fired it, so to speak, refuses to do so directly afterwards; this is the absolute refractory period. Furthermore, it takes some time to recover. Then it is rather reluctant to fire; this is the relative refractory period. It may fire but needs some extra input as compared to the original threshold ϑ. All this is taken into account by a refractory potential η that is added to v once a neuron has fired; it is taken to be causal so that $\eta(t) = 0$ for $t < 0$. If $\eta(t) < 0$ for $t > 0$, then we need, so to speak, more v to let a neuron fire. In this way we can keep the threshold ϑ fixed and arrive at the simple dynamics

$$n_i(t + \Delta t) = \Theta[v_i(t) - \vartheta] \tag{10}$$

with

$$v_i(t) = \sum_f \eta(t - t_i^f) + \sum_j J_{ij}\varepsilon(t - t_j^f - \Delta_{ij}^{\mathrm{ax}}) \tag{11}$$

Here Θ is the Heaviside step function with $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. A sum over f is always a sum over firing times t^f of whatever neuron, here neuron i and its 'neighbors' j. Neuron i 'fires' as soon as its membrane potential v_i reaches the firing threshold ϑ from below,

$$\lim_{t \nearrow t_i^f} v_i(t) = \vartheta \qquad \text{and} \qquad \lim_{t \nearrow t_i^f} \frac{\mathrm{d}v_i(t)}{\mathrm{d}t} > 0 \ . \tag{12}$$

The potential v_i being in general a continuous function, the second condition is a mathematical formulation of the fact that there is no spike appearing when v_i returns from being above ϑ. In passing we note that the dynamics (10) need not be based on discrete time. With a few, trivial, modifications it works equally well for continuous time.

Throughout what follows we will work with the above dynamics where during each time step *all* neurons are updated. This is the so-called parallel dynamics, for biological neurons the natural one. Absolute refractory behavior means η

assuming the value $-\infty$ while relative refractory behavior is equivalent to saying η is negative but finite. Though bursts can be described easily by allowing η to be positive during an appropriate period of time, we will not study this explicitly. For arbitrary axonal delays $\Delta_{ij}^{\mathrm{ax}}$ there is no hope for obtaining an exact solution of the network dynamics. As Eqs. (10) and (11) show explicitly, the Spike Response Model incorporates the three essential ingredients of neuronal spike generation, viz., a variable threshold, spikes, and their effect, the postsynaptic potentials. All three are a response to external input – hence the name of the model. It incorporates the integrate-and-fire model as a special case [27].

To finish this section, it may be well to face the question: Why does a time-discrete dynamics such as (10) make sense? After all, it treats the effect of a spike as a Kronecker delta. The answer is that, as we have already seen, the width of a postsynaptic potential in general greatly exceeds that of a spike so that we can treat the latter as an *approximate* delta function. Discretizing time we then end up with a simple Kronecker delta, as advertised.

3 Hebbian Learning in a Network of Formal Neurons

Encoding and decoding are two sides of the same medal. Encoding means asking a two-fold question: how do we represent the neurons' activity and how do we store spatiotemporal activity patterns in the synapses connecting the neurons? Hebbian learning is a prominent, and very efficient, way of information storage. Decoding means reading out stored information. We will treat both.

3.1 Representation of Neuronal Activity

Formal neurons are described by a Boolean variable indicating their activity. (A Boolean variable assumes only two values.) We have already met $n_i(t) \in \{0, 1\}$. Neuron i is active at time $t \in \mathbb{Z}\Delta t$ when $n_i(t) = 1$ and it is quiescent when $n_i(t) = 0$. Here $\mathbb{Z}$ represents the integers. It is convenient to take $\Delta t = 1\,\mathrm{ms}$ as the duration of a spike. For what follows we also introduce a pseudo-spin $S_i = 2n_i - 1 \in \{-1, 1\}$. Now $S_i(t) = 1$ when neuron i fires a spike at time t and $S_i(t) = -1$ when it does not. The above distinction leads to two ways of encoding neuronal activity, viz., the 0/1 and the ± 1 representation (coding). Each of them needs a specific context, to which we now turn.

Apparently there are at least two ways of coding a neuron's activity: through $\{0, 1\}$ and through ± 1. Their choice is dictated by the *global* activity of a network. In a theoretical analysis, an activity pattern is a set of independent, identically distributed (iid) random variables. Different patterns are also taken to be independent. There are at least three reasons for doing so. First, in this way we avoid any special assumption concerning the state of the network. Second, iid

random variables are easy to generate on a computer by means of a random number generator [19]. Third, one can use laws of large numbers from the theory of probability [11, 12] to analyze collective behavior. This is what we are going to do.

Suppose about half of the neurons in a network are active during each time step. Then a pattern μ is a set of iid random variables $\{\xi_i^\mu; 1 \leq i \leq N\}$ where $\xi_i^\mu = \pm 1$ with probability $p = 1/2$ and N is the size of the network. The two states of the neuron are practically equivalent. Let us imagine that ± 1 corresponds to black and white (you can make a choice yourself). It is, so to speak, immaterial whether we have black and white or white and black pixels in that one can interchange black and white but all these patterns "look" the same. If we now require that the spatial average of the activity as defined below in (13) be zero, then the pseudo-spin representation $S_i \in \{-1, 1\}$ is the appropriate one.

There is a convenient way of characterizing the activity of a pattern, say μ. We have, as $N \to \infty$,

$$a_\mu := \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \xi_i^\mu = p_\mu - (1 - p_\mu) = 2p_\mu - 1 \tag{13}$$

by the strong law of large numbers [11] applied to the iid random variables $\{\xi_i^\mu; 1 \leq i \leq N\}$ equipped with a probability distribution assigning a probability p_μ to $\xi_i^\mu = +1$ and $1 - p_\mu$ to $\xi_i^\mu = -1$. Hence for $p_\mu = 1/2$ the parameter a_μ is bound to vanish.

On the other hand, when the global activity in a network is low, i.e., only a few neurons are active during each time step, then $p_\mu \approx 0$ and $a_\mu \approx -1$. What, then, should be the right mathematical representation of a neuron's activity? Clearly, activity is the exception and inactivity is the rule so that $n_i(t) \in \{0, 1\}$ is just what we are looking for since the active sites with $n_i(t) = 1$ carry the information. Though inactive neurons are the majority, they carry the label $n_i(t) = 0$ and, thus, do not count – as they should. We will see that mathematically all this fits together quite nicely in decoding neuronal information.

3.2 Hebbian Learning

Donald Hebb's classic *The Organization of Behavior – A Neurophysiological Theory* [28, 29] appeared in 1949. On p. 62 of this book one can find the now famous "neurophysiological postulate": *When an axon of cell A is near enough to excite a cell B and* repeatedly *or* persistently *takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.*

One may then wonder, of course, where the above "metabolic change" might take place. Hebb directly continued by suggesting that "synaptic knobs develop" and on p. 65 he states very explicitly: "I have chosen to assume that the growth

of synaptic knobs, with or without neurobiotaxis, is the basis of the change of facilitation[1] from one cell on another, and this is not altogether implausible". No, as we now know, it is not. It is just a bit more complicated.

Sloppily formulated, Hebbian learning is learning by repetition: "Practice makes perfect". The organism repeats the very same pattern or sequence of patterns several times and in so doing trains the synapses. (A nice example is the barn owl learning to perform azimuthal sound localization, an example we will consider later on.) It has turned out that Hebbian learning is robust, faithful, and a key to understanding map formation in the cortex.

Hebb also formulated a second idea having nearly an equally big impact as his learning rule, viz., that of an *assembly* (or ensemble) of neurons. If neuronal behavior should code a synapse, then a postsynaptic neuron should fire at the right moment. To attain its firing threshold, a neuron needs well-timed input in a narrow time window from many other neurons. The 'assembly' should then fire more or less simultaneously. This is its distinguishing feature. The activity patterns we will be analyzing in learning theory are often concrete examples of Hebbian assemblies.

Hebb's postulate has been formulated in plain English – but not more than that – and the main question we are facing here is how to implement it mathematically. From a higher point of view, one might define Hebbian learning to be long-term synaptic plasticity induced by pre- and/or postsynaptic activity and *local* in space and time. Most of the information which is presented to a network, then, varies in space and time. So what is needed is a *common* representation of *both* the spatial *and* the temporal aspects. As a pattern changes, the system should be able to measure and store this change. How can it do that?

3.3 Spatiotemporal Patterns and ± 1 Coding

As in real life, a network may, but need not, learn. Suppose then it does and let us imagine a spatiotemporal pattern of duration $T_{\rm stp}$, i.e., a sequence of patterns $\{\xi_i^t; 1 \le i \le N,\ 0 < t \le T_{\rm stp}\}$; for fixed i but different times t the patterns ξ_i^t may be identical. We simply describe such a pattern by $\{S_i(t); 1 \le i \le N,\ 0 < t \le T_{\rm stp}\} := \{\mathbf{S}(t); 0 < t \le T_{\rm stp}\}$ where $\mathbf{S}$ is now a *given* vector function of time. In addition, we assume that our network is going to "learn" this pattern. The question is: how?

Without further ado we first consider the answer [30, 31, 32] that specifies how J_{ij} is to change after our pattern $\mathbf{S}$ has been taught the network during a period of time starting at $t - T_l$ and ending at t,

$$\Delta J_{ij} = \zeta_{ij}(\Delta_{ij}^{\rm ax})\,\frac{1}{T_l}\sum_{t=1}^{T_l} S_i(t+\Delta t)S_j(t-\Delta_{ij}^{\rm ax})\ , \tag{14}$$

[1] Webster's Ninth New Collegiate Dictionary says (:to those who do not belong to the in-crowd): facilitation = the increasing of the ease or intensity of response by repeated stimulation.

which is to be added to the existing synaptic efficacy. Here T_l is the learning time that may, but need not, be identical with the duration $T_{\rm stp}$ of the spatiotemporal pattern under consideration; for instance, because the pattern is repeated ("practice makes perfect"). The rationale of (14) is the following. We look at synapse $\{ij\}$ at time t. What we see there is the activity of neuron j at time $t-\Delta_{ij}^{\rm ax}$ since the axonal delay lasts $\Delta_{ij}^{\rm ax}$ ms. We correlate this with the activity of neuron i at $t+\Delta t$. It is important to realize that this is *one time step later*. With the benefit of hindsight this is perfectly reasonable since the synapse at time t should tell the postsynaptic neuron what to do next. In technical terms, the retrograde effect of a neuronal action potential as noticed by the synapse is taken to be instantaneous so that a delta function is a fair approximation [21]. The prefactor $\zeta_{ij}(\Delta_{ij}^{\rm ax}) \geq 0$ is still at our disposal. For the sake of simplicity we assume no self-interaction is present ($J_{ii} = 0$).

It seems plain that, in a ± 1 coding, Eq. (14) is not quite what Hebb [28] had in mind: if neurons i and j are active, $\Delta J_{ij} \geq 0$, but the same holds true, if i and j are *in*active. The former would be fine to Hebb, the latter somewhat weird. With hindsight this is, however, perfectly reasonable since the states 'active' and 'inactive' ($p = 1/2$) are equivalent: what's in a name? (Representation is all.) On the other hand, $\Delta J_{ij} \leq 0$ if one of the neurons is active and the other quiescent.

In (14) S_i and S_j are treated on an equal footing. Symmetry with respect to an interchange of i and j reigns, if the pattern is a stationary one, i.e., $S_i(t) := \xi_i$ for all $0 < t \leq T_l$. Apart from the high activity, a condition that will soon be relaxed, a constant firing might happen in neurons with negligible refractory behavior; in view of our present choice of $\Delta t = 1\,$ms, it is equivalent to saying that the maximal firing rate is 1000 Hz, which is not completely off.

Let us now study the effect of a stationary pattern in conjunction with $q \gg 1$ similar patterns, the Hopfield model [33]. The network has no delays ($\Delta^{\rm ax} = 0$). In practice this means that we have $(q+1)N$ independent, identically distributed random variables ξ_i^μ which assume the values ± 1 with probabilities p for $+1$, $1-p$ for -1, and mean $a = 2p - 1$. Furthermore, $\mu = 0$ corresponds to the pattern ξ we started with. Hopfield took a fully connected neural network with $p = 1/2$ so that $a = 0$ – by good reason, as we will soon see. The patterns are presented to the network one after the other and learned through (14). Altogether we obtain, putting $\zeta_{ij} = 1/N$,

$$J_{ij} = N^{-1} \sum_{\mu=0}^{q} \xi_i^\mu \xi_j^\mu \ . \tag{15}$$

The dynamics being given by

$$S_i(t + \Delta t) = \mathrm{sgn}[v_i(t) - \vartheta] \ , \tag{16}$$

the threshold is taken to vanish, i.e., $\vartheta = 0$, and so is the average $\langle J_{ij} \rangle = 0$; here sgn is the sign function. The rationale of these two requirements, which belong together, will soon become clear.

Following Hopfield, we now imagine that EPSPs are also instantaneous, i.e., delta functions, while refractory behavior lasts as long as a spike (so that $\eta \equiv 0$), and compute the potential v_i with the original pattern $\boldsymbol{\xi}$ as input to our neural network. Using (15) and (16) in conjunction with (11) we then find, as $N \to \infty$,

$$\begin{aligned} v_i &= N^{-1} \sum_{j(\neq i)} \left(\sum_\mu \xi_i^\mu \xi_j^\mu \right) \xi_j = \xi_i + N^{-1} \sum_{\substack{j(\neq i) \\ \mu(\neq 0)}} \xi_i^\mu \xi_j^\mu \xi_j \\ &= \xi_i + qa^3 + N^{-1} \sum_{\substack{j(\neq i) \\ \mu(\neq 0)}} (\xi_i^\mu \xi_j^\mu \xi_j - a^3) \ . \end{aligned} \tag{17}$$

By the central limit theorem in conjunction with the law of the iterated logarithm (see Appendix A), the last sum on the right is Gaussian (with mean zero) and of the order $\mathcal{O}(\sqrt{q/N})$. The above argument is a signal-to-noise ratio analysis with ξ_i being the signal and the last sum representing the noise.

If, then, $a \neq 0$ but $\vartheta = 0$ in (16), there is no hope for storing anything but a few patterns since qa^3 will wash out the signal ($|\xi_i| = 1$) for q large enough. That is why Hopfield took $p = 1/2$ so that $a = 0$. Then a faithful retrieval is possible only if $q/N < 0.138$; for q/N beyond this fraction no pattern can be retrieved. It is easy to see that an inequality of this kind must exist; determining the precise number 0.138 is a completely different story [34, 35, 36]. We simply return to (17) and note that the standard deviation of the sum is $\sqrt{q/N}$. When it becomes too big, it will wash out the signal, as did qa^3 in the case $a \neq 0$. In fact, according to the law of the iterated logarithm we get as an upper bound $q_{\max}/N = 0.5$; apparently it has to be less. Hopfield [33, pp. 2556/7] already found 0.15 numerically, which was surprisingly close to 0.138. Of course one could adapt the threshold, if $a \neq 0$, and take $\vartheta = -qa^3$ but what should tell a synapse that q patterns have been stored?

For spatio*temporal* patterns with $a = 0$ the learning rule (14) in combination with the dynamics (16) and $\vartheta = 0$ works well, provided the system has a *broad distribution* of delays. It is easy to see why,

$$v_i(t) = \sum_j J_{ij} S_j(t - \Delta_{ij}^{\text{ax}}) \ . \tag{18}$$

The J_{ij} look back into the near past and tell neuron i at time t what to do next – in agreement with (14) and (16). That is to say, the delay Δ_{ij}^{ax} that has been taken into account during learning plays the very same role during retrieval. If, then, a certain activity pattern keeps constant during a time δ_p, there should be delays $\Delta_{ij}^{\text{ax}} > \delta_p$ to "throw" the system out of this state into the next. A 'broad' distribution then means $\Delta_{ij}^{\text{ax}} > \delta_p$ for 'enough' j. Figure 3 illustrates the potency of learning rule (14). For biased signals with $a \neq 0$ it has to be modified; see §3.5.

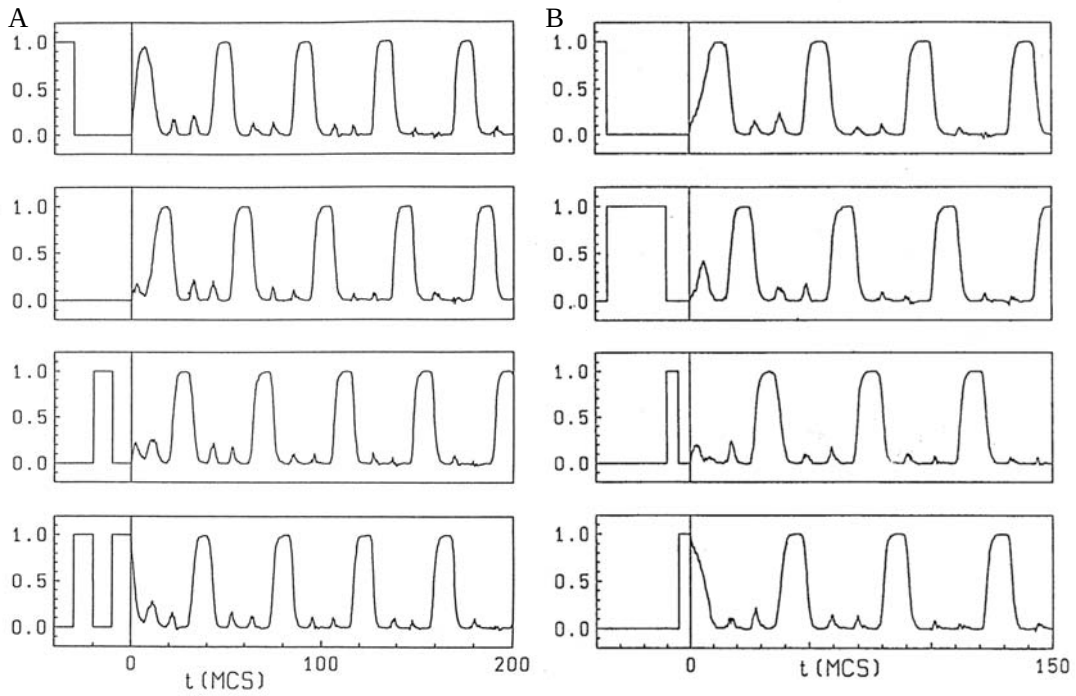

Figure 3: (A) *Space warp.* The overlaps $m_\mu(t) = N^{-1} \sum_i \xi_i^\mu S_i(t)$ with patterns $1, 2, 3$, and 4 (from top to bottom, ± 1 representation with $p_\mu \equiv 1/2$) have been plotted as a function of time. The network with maximal delay $\Delta_{\max} = 40$ has learned the cycle $1, 2, 3, 4$ (or *BACH*) where each pattern lasts 10 time units; the network size is $N = 512$. After it has been presented the faulty pattern sequence $1, \mathbf{4}, 3, 4$ (or *BHCH*) as initial condition (space warp) for $-\Delta_{\max} \leq t \leq 0$, the correct order *BACH* is spontaneously retrieved. The initial conditions appear in the boxes on the left of $t = 0$. (B) *Time warp.* The same system as in (A) has (among other things) again been taught the theme *BACH*, all 'notes' having a duration of 10 time units. The overlaps with *B*, *A*, *C*, and *H* (from top to bottom) have been plotted as a function of time. After the network has been shown a pattern sequence with the wrong timing (time warp) as initial condition for $-\Delta_{\max} \leq t \leq 0$ with $\boldsymbol{A}$ lasting *much* longer than *B*, *C*, and *H*, the correct cycle (i.e., with its correct timing) is spontaneously retrieved. As in (A), the initial conditions appear in the boxes on the left of $t = 0$. Taken from [32]

3.4 Hebbian Unlearning

In the ± 1 representation, we can define a spatial average $a(t) = N^{-1} \sum_i S_i(t)$ à la (13) for each discrete pattern but in practice there is no hope that it will vanish. If, however, it is small enough, viz., $|a| < |a_c| = 0.54$, then *unlearning* [41] removes the correlations, restores the patterns, and greatly increases the storage capacity. Though at a first sight unlearning looks a bit weird, it is a very powerful algorithm. Its motivation stems from neurobiology.

In the late seventies Hobson and McCarley [38] suggested that there exists a dream state generator in the *pons* that, during REM sleep, produces a series of pulses in the forebrain, the PGO (ponto-geniculo-occipital) bursts. These pulses provide frequent and semi-random stimuli to the cortex and might thus function as the driving force of rapid-eye-movement (REM) dreams. The Hobson-McCarley idea was taken up by Crick and Mitchison [39], who assumed that during REM sleep the cortical network, once it has been excited by a PGO burst, relaxes to a parasitic or spurious state, which is then weakened or, as they called it, 'unlearned'. The proposal of Crick and Mitchison found an immediate implementation as a three-step procedure for constant patterns in the work of Hopfield et al. [40], while the present version for general, spatiotemporal, patterns is due to van Hemmen et al. [41]:

(i) *Random shooting*, corresponding to a PGO burst in the brain and giving a random initial state.
(ii) *Relaxation* to a limit state $\boldsymbol{\chi} = (\chi_i^d(t); 1 \le i \le N)$, where we assume (in many cases this can be proven) that the limit state $\boldsymbol{\chi}$ is either stationary or a limit cycle.
(iii) *Unlearning* through

$$J_{ij} \to J_{ij} - \epsilon \Delta J_{ij} \tag{19}$$

with $0 < \epsilon \ll 1$ and ΔJ_{ij}, pure Hebbian learning of $\boldsymbol{\chi}$ after (14) but now being multiplied by $-\epsilon$. The *minus* sign in front of ϵ has led to the name *un*learning. The unlearning parameter ϵ must be small, say, two orders of magnitude smaller that the available learning parameters in sight.

The three steps constitute a single loop, which in the present context is defined to be a "dream". It is repeated D times so that $0 < d \le D$ labels the "dreams". For an extensive study of Eq. (19), its remarkable capabilities, the dependence upon D, and appropriate references we refer to elsewhere [41]. Though the implementation is straightforward ($-\epsilon \times$ Hebbian learning), there is no detailed study yet for low-activity patterns. To obtain results as solid as the ones for ± 1 coding one has to perform very intricate programming for huge system sizes so as to get good statistics. This still has to be done.

3.5 Spatiotemporal Patterns and 0/1 Coding

As a rule, biological neural systems, i.e., neuronal networks, are characterized by a low activity, meaning that a relatively low percentage of neurons per unit of time is active. That is to say, $a \approx -1$ so that the above formalism breaks down completely. Since stationary patterns are the exception and spatiotemporal ones the rule, we are looking for an *appropriate* generalization of (14). There is an evident one: replace ξ_i^μ by a random variable that also has mean zero and variance one, $\xi_k^\mu \to (\xi_k^\mu - a)/\sqrt{1-a^2}$. This looks reasonable but it is not. Already for stationary patterns – cf. (15) – this symmetric rule is a lousy one since the storage capacity goes to zero as $a \to -1$. A way out [36, §1.6.5] is adding an extra $+1$ so as to restore the original storage capacity of the Hopfield model but there is a smarter solution that has turned out to work for spatio*temporal* patterns as well. The 'symmetric' substitution by itself did not.

Though we now work with the 0/1 representation of neuronal activity we stick to the pseudo-spin $\mathbf{S}$ for specifying our asymmetric learning rule for a synapse with axonal delay $\Delta^{\rm ax} = \Delta_{ij}^{\rm ax}$ [32],

$$\Delta J_{ij} = \zeta_{ij}(\Delta^{\rm ax})\,\frac{1}{T_l}\sum_{t=1}^{T_l} S_i(t+\Delta t)[S_j(t-\Delta^{\rm ax})-a]\ . \tag{20}$$

This is to be added to the existing synaptic efficacy. It reduces to (14) for $a = 0$. The rationale of (20) is as before, but modified since we now have $[S_j(t-\Delta^{\rm ax})-a]$ instead of $S_j(t-\Delta^{\rm ax})$, the $-a$ being crucial. In the low-activity limit with $a = -1$, $[S_j(t-\Delta^{\rm ax})-a]$ equals 2 if at time $t-\Delta^{\rm ax}$ the presynaptic neuron j is active and vanishes if the neuron is quiescent. That is perfectly reasonable since j cannot activate i if it is not active itself. After a spike has been generated by j it needs $\Delta^{\rm ax} = \Delta_{ij}^{\rm ax}$ ms to reach the synapse $\{ij\}$ at time t. Then $\Delta J_{ij} \gtrless 0$ if $S_i(t+\Delta t) = \pm 1$, corresponding to i being "told" to fire or keep quiet. In words, the presynaptic neuron is gating. Given presynaptic activity, the synaptic efficacy increases (= potentiation) when during the *next* time step the postsynaptic neuron is active whereas it decreases (= depression) when the postsynaptic neuron does *not* fire.

As Fig. 4 illustrates, the asymmetric learning rule (20) has proven to be extremely efficient for storing spatiotemporal patterns [32, 42, 43]. It was also a key to devising the 'learning window' [45] which is instrumental in describing long-term synaptic plasticity for temporally highly resolved activity patterns. Experiments of Markram et al. [48] were the first to confirm (20) as neurobiological learning rule; a more complete list will be given once we turn to explaining the notion of 'learning window' in the next section. Additional theoretical support through combinatorial and energy-saving arguments has been provided as well [49].

The dynamics is that of (10) with $v_i(t) = \sum_j J_{ij} n_j(t-\Delta_{ij}^{\rm ax})$. To see the effect

of $S_i(t+\Delta t)[S_j(t-\Delta_{ij}^{\rm ax})-a]$ as it appears in (20), we perform a simple signal-to-noise ratio analysis for a network with no delays and N as the number of neighbors each neuron is connected with while $\zeta_{ij}=1/N$. Furthermore, the network has been taught $q+1$ stationary patterns so that $J_{ij}=N^{-1}\sum_\mu \xi_i^\mu(\xi_j^\mu-a)$. The pattern $\mu=0$ with $\xi_i^0=:\xi_i$ is presented to the network while the others generate the 'noise'. We assume $a_\mu\equiv a$ for the sake of simplicity; cf. (13). Focusing on the low-activity limit $a\to -1$, we take $n_j:=(\xi_j-a)/2$ as input and find, as $N\to\infty$,

$$\begin{aligned} v_i &= N^{-1}\sum_{j(\neq i)}\left(\sum_\mu \xi_i^\mu(\xi_j^\mu-a)\right)(\xi_j-a)/2 \\ &= \xi_i\frac{(1-a^2)}{2}+\frac{1}{2N}\sum_{\substack{j(\neq i)\\ \mu(\neq 0)}}\xi_i^\mu(\xi_j^\mu-a)(\xi_j-a)=(1+a)[\xi_i+\mathcal{O}(\sqrt{q/N})]\,. \end{aligned} \quad (21)$$

Except for the common factor $(1+a)$ on the right, which we can forget about, we have the same terms as in (17). One might have argued that $(\xi_i^\mu-a)(\xi_j^\mu-a)$ (symmetry) were nicer. If so, (ξ_i-a) would replace the signal term ξ_i in (21) so that the signal for the inactive neurons, i.e., the big majority, would be strongly weakened. Because of the asymmetry the signal term ξ_i now appears alone, without $-a$ or qa^3, which greatly improves the signal-to-noise ratio. Even worse for the "symmetric" term, it does not allow spatio*temporal* low-activity patterns to evolve. For $\Delta^{\rm ax}>0$ the argument becomes more complicated but the gist does not change.

4 Time-Resolved Hebbian Learning: Looking at Synapses Through a Learning Window

The barn owl (*Tyto alba*) is able to determine the prey direction in the dark by measuring interaural time differences (ITDs) with an azimuthal accuracy of 1–2 degrees corresponding to a temporal precision of a few μs, a process of binaural sound localization. The first place in the brain where binaural signals are combined to ITDs is the laminar nucleus. A temporal precision as low as a few μs was hailed by Konishi [50] as a paradox – and rightly so since at a first sight it contradicts the slowness of the neuronal "hardware", viz., membrane time constants of the order of 200 μs. In addition, transmission delays from the ears to laminar nucleus scatter between 2 and 3 ms [44] and are thus in an interval that greatly exceeds the period of the relevant oscillations (100–500 μs). The key to the solution [45] is a Hebbian learning process – cf. §3.5 – that tunes the hardware so that only synapses and, hence, axonal connections with the right timing survive. Genetic coding is implausible because three weeks after hatching, when the head

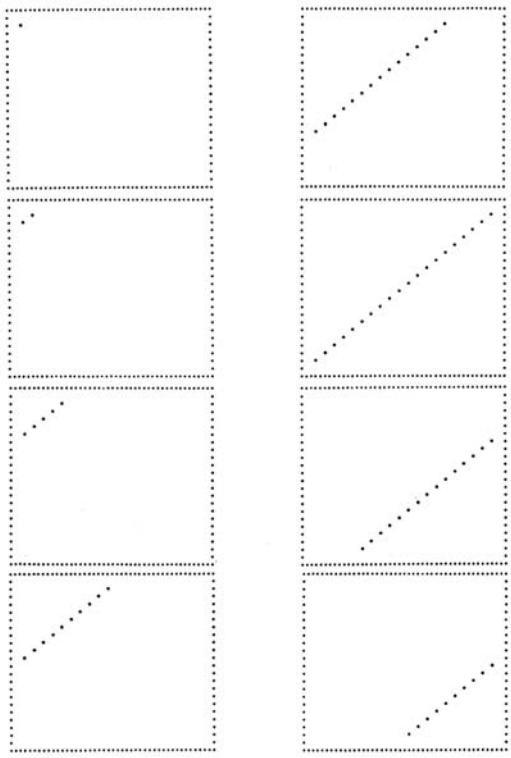

Figure 4*:* Motion of a 'phase boundary', a string of black pixels, through a 20×20 storage layer with 0/1 representation; 1 is black and 0 is white, i.e., invisible. The system starts with a single point in the upper left-hand corner and the string develops as time proceeds (top to bottom; first left, then right). During the motion, the number of black pixels varies between 1 and 20 but a in (20) does not: it is -1. Taken from [32].

is full-grown, the young barn owl cannot perform azimuthal sound localization. Three weeks later it can. So what happens in between?

The solution to the paradox involves a careful study of how synapses develop during ontogeny [45, 46, 47]. The inputs provided by many synapses decide what a neuron does but, once it has fired, the neuron determines whether each of the synaptic efficacies will in- or decrease, a process governed by the synaptic learning window, a notion that will be introduced shortly. It is a generalization of what we have seen in Eq. (20). Each of the terms below in (22) has a neurobiological origin. The process they describe is what we call *infinitesimal learning* in that synaptic increments and decrements are small. Consequently it takes quite a while before the organism has built up a 'noticeable' effect. As for the mean response $\mathcal{R} = npQ$ studied in §1, only the presynaptic probability of release p and the postsynaptic response Q can change 'continuously' whereas the number n cannot. What happens in the long run is not known yet [4, 8, 18].

For the sake of definiteness we are going to study waxing and waning of synaptic strengths associated with a *single* neuron, which therefore need not carry a label; cf. Fig. 5. The $1 \leq i \leq N$ synapses are providing their input at times t_i^f. The firing times of the neuron are denoted by t^n, it being understood that n is a label like f. Given the firing times, the change $\Delta J_i(t) := J_i(t) - J_i(t - T_l)$ of the efficacy of synapse i (synaptic strength) during a learning session of duration T_l

and ending at time t is governed by several factors,

$$\Delta J_i(t) = \eta \left[\sum_{t-T_l \le t_i^f < t} w^{\rm in} + \sum_{t-T_l \le t^n < t} w^{\rm out} + \sum_{t-T_l \le t_i^f, t^n < t} W(t_i^f - t^n) \right] . \qquad (22)$$

Here the firing times t^n of the postsynaptic neuron may, and in general will, depend on J_i. We now focus on the individual terms. The prefactor $0 < \eta \ll 1$ reminds us explicitly of learning being slow on a neuronal time scale[2]. Throughout what follows we refer to this condition as the the *adiabatic hypothesis*. It holds in numerous biological situations and has been a mainstay of computational neuroscience ever since. It may also play a beneficial role in an applied context. If it is does not hold, a numerical implementation of the learning rule (22) is straightforward, but an analytical treatment is not.

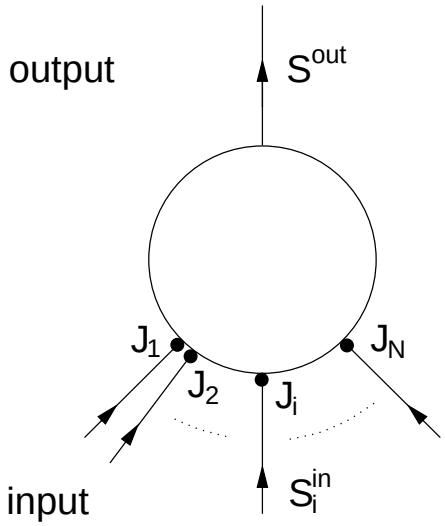

Figure 5*:* Single neuron. We study the development of synaptic weights J_i (small filled circles, $1 \le i \le N$) of a single neuron (large circle). The neuron receives input spike trains denoted by $S_i^{\rm in}$ and produces output spikes denoted by $S^{\rm out}$. Taken from [47].

Each incoming spike and each action potential of the postsynaptic neuron change the synaptic efficacy by $\eta w^{\rm in}$ and $\eta w^{\rm out}$, respectively; see the literature [51, 52, 53, 54] for experimental evidence.

The last term in (22) represents the *learning window* $W(s)$, which indicates the synaptic change in dependence upon the time difference $s = t_i^f - t^n$ between an incoming spike t_i^f and an outgoing spike t^n. When the former precedes the latter, we have $s < 0 \Leftrightarrow t_i^f < t^n$, and the result is $W(s) > 0$, implying potentiation. This seems reasonable since NMDA receptors (see §1), which are important for long-term potentiation (LTP), need a strongly positive membrane voltage to get 'accessible' by loosing the Mg^{2+} ions that block their 'gate'. A postsynaptic action potential induces a fast retrograde 'spike' doing exactly this [21]. Because the presynaptic spike arrived slightly earlier, neurotransmitter is waiting for getting

[2]Since the Greek alphabet is finite and there is no ambiguity between the present learning parameter and the refractory potential of Section 2, there is no harm in using η here as well.

access, which is allowed after the Mg^{2+} ions are gone. The result is Ca^{2+} influx. On the other hand, if the incoming spike comes "too late", then $s > 0$ and $W(s) < 0$, implying depression – in agreement with a general rule in politics, discovered a decade ago: "Those who come too late shall be punished." In neurobiological terms, there is no neurotransmitter waiting for being admitted. The learning rule (22) is a direct extension of (20), its time-discrete predecessor. There is meanwhile extensive neurobiological evidence [48, 55, 56, 57, 58, 59] in favor of this time-resolved Hebbian learning. An illustration of what a learning window does is given in Fig. 6.

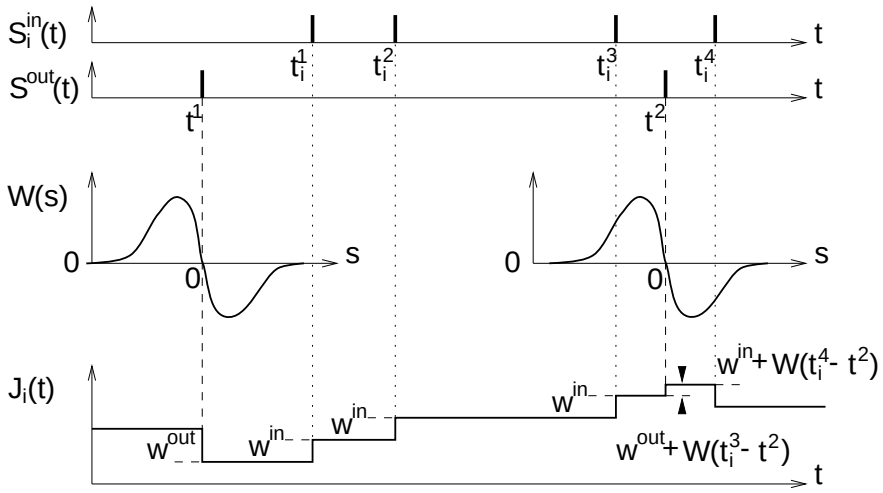

Figure 6*:* Hebbian learning and spiking neurons – schematic. In the bottom graph we show the time course of the synaptic weight $J_i(t)$ evoked through input and output spikes (upper graphs, vertical bars). An output spike, e.g., at time t^1, induces the weight J_i to change by an amount $w^{\rm out}$, which is negative here. To show the effect of correlations between input and output spikes, the learning window $W(s)$ (center graphs) has been indicated around each output spike; $s = 0$ matches the output spike times (vertical dashed lines). The three input spikes at times $t_i^f = t_i^1$, t_i^2 and t_i^3 (vertical dotted lines) increase J_i by an amount $w^{\rm in}$ each. There are no correlations between these input spikes and the output spike at time t^1. This becomes clear once we look at them "through" the learning window W centered at t^1: the input spikes are too far away in time. The next output spike at t^2, however, is close enough to the previous input spike at t_i^3. The weight J_i is changed by $w^{\rm out} < 0$ plus the contribution $W(t_i^3 - t^2) > 0$, the sum of which is positive (arrowheads). Similarly, the input spike at time t_i^4 leads to a change $w^{\rm in} + W(t_i^4 - t^2) < 0$. Taken from [47].

If other (infinitesimal) learning algorithms are discovered, one can simply adapt W accordingly. For instance, for inhibitory synapses one has found infinitesimal growth processes [60] that can be described qualitatively by putting $W := -W$ in Fig. 7; the latter shows a typical learning window for an excitatory synapse [45, 47]

$$W(s) = \eta \begin{cases} \exp(s/\tau^{\rm syn})\,[A_+(1 - s/\tilde{\tau}_+) + A_-(1 - s/\tilde{\tau}_-)] & \text{for} \quad s \le 0\,, \\ A_+ \exp(-s/\tau_+) + A_- \exp(-s/\tau_-) & \text{for} \quad s > 0\,. \end{cases} \tag{23}$$

Here, as before, $s = t_i^f - t^n$ is the time difference between presynaptic spike arrival and postsynaptic firing, η is our small learning parameter, $\tilde{\tau}_+ := \tau^{\rm syn}\tau_+/(\tau^{\rm syn} + \tau_+)$, and $\tilde{\tau}_- := \tau^{\rm syn}\tau_-/(\tau^{\rm syn} + \tau_-)$. Parameter values as used in numerical simulations [47] are $\eta = 10^{-5}$, $A_+ = 1$, $A_- = -1$, $\tau^{\rm syn} = 5\,{\rm ms}$, $\tau_+ = 1\,{\rm ms}$, and $\tau_- = 20\,{\rm ms}$.

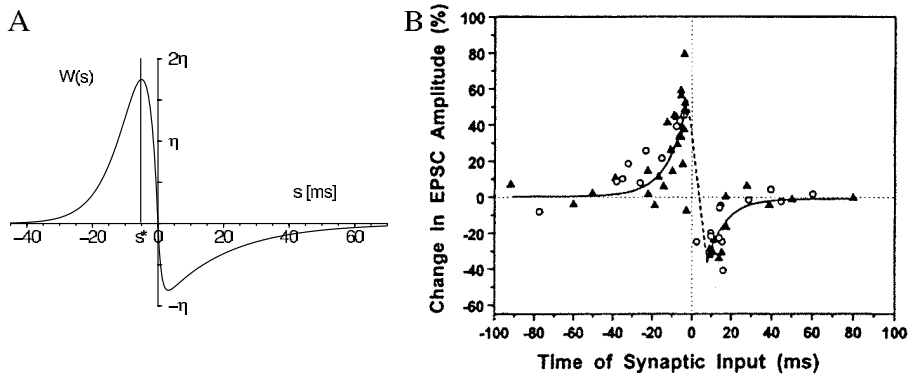

Figure 7: A. The learning window W in units of the learning parameter η as a function of the delay $s = t_i^f - t^n$ between presynaptic spike arrival at synapse i at time t_i^f and postsynaptic firing at time t^n. If $W(s)$ is positive (negative) for some s, the synaptic efficacy J_i is increased (decreased). The increase of J_i is most efficient, if a presynaptic spike arrives a few milliseconds *before* the postsynaptic neuron starts firing (vertical dashed line at $s = s^*$). For $|s| \to \infty$ we have $W(s) \to 0$. The form of the learning window and parameter values are as described in Eq. (23). Taken from [47]. B. Experimentally obtained learning window of a cell in rat hippocampus; reprinted by permission [55]. The similarity with the left figure is evident. It is important to realize that the width of the learning window is to be in agreement with other neuronal time constants. In the auditory system, for instance, these are nearly two orders of magnitude smaller so that the learning window's width scales accordingly.

Spike generation is (nearly) always a *local* process in time and so are the $1 \le i \le N$ input process generating the input spikes t_i^f. For the latter category we can, and will, take inhomogeneous Poisson processes (see Appendix B), with rate function $\lambda_i(t)$; any other local process with independent increments or short-range correlations (cf. Appendix A) would do as well. In the barn-owl case, the λ_i are periodic functions corresponding to a certain frequency. A simple explanation is as follows. The cochlea being an "inverse piano", it performs a frequency decomposition with, say, i labeling the frequencies in a cochleotopic manner. In the auditory system spike trains are therefore phase-locked to a frequency component of the acoustic stimulus. A Poisson process operating in a learning window of finite width (of a few milliseconds) emulates that input frequencies are never fixed but belong to a finite frequency range.

The time interval $[t - T_l, t)$ is taken to be big since, due to the adiabatic hypothesis, learning is so slow that we can safely assume T_l to greatly exceed neuronal times such as interspike intervals and the width of the learning window.

Nevertheless we will arrive at a relatively small change of the J_i's so that the assumption concerning T_l is self-consistent (otherwise we don't see anything). We can divide the time interval $[t - T_l, t)$ into many small intervals that are, stochastically, independent of each other – apart from a minuscule overlap at their borders. Hence the sum (22) is self-averaging; cf. the discussion following (4), a process that is incorporated here.

The above averaging was one over the randomness. We are now going to perform another one over time. To fully appreciate what is going to happen, we turn to a differential-equation problem,

$$\frac{\mathrm{d}}{\mathrm{d}t}x = \eta F(x,t) \tag{24}$$

where η is 'small' and $F(x,t)$ for fixed x is a periodic function of t, i.e, $F(x, t + T) = F(x,t)$. After one period x has hardly changed so that, *for fixed* x, we can average F over t. That is to say, instead of (24) one studies [61, 62]

$$\bar{F}(x) := \frac{1}{T}\int_{t-T}^{t} \mathrm{d}t'\, F(x,t') \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}x = \eta \bar{F}(x)\ . \tag{25}$$

Here the integral over time, viz., t', is performed *with* x, the argument of $\bar{F}$, *fixed*; the integration boundaries $t - T$ and t of the integral in (25) can be replaced by 0 and T, respectively. Hence the differential equation we arrive at is an autonomous one since $\bar{F}$ does not depend explicitly on t. It is plain that the whole argument hinges on η *being small*. In fact, under suitable conditions the 'method of averaging' [61, 62] can be generalized to nonperiodic F. Here we will simply average over a period of duration T_l and often use an overbar to indicate this.

We now return to our problem, viz., (22) averaged over the randomness, and average over time as well. This sounds quite harmless (it is) but we will soon see the effect is beneficial. To simplify the notation, we first introduce two spike flows[3],

$$S_i^{\mathrm{in}}(t) = \sum_{t_i^f \le t} \delta(t - t_i^f), \quad S^{\mathrm{out}}(t) = \sum_{t^n \le t} \delta(t - t^n), \tag{26}$$

and rewrite (22), introducing angular brackets to indicate an average over the randomness,

$$\begin{aligned} \frac{\Delta J_i(t)}{T_l} &= \eta \left\{ \frac{1}{T_l}\int_{t-T_l}^{t} \mathrm{d}t' \left[w^{\mathrm{in}} \langle S_i^{\mathrm{in}}(t') \rangle + w^{\mathrm{out}} \langle S^{\mathrm{out}}(t') \rangle \right] \right. \\ &+ \left. \frac{1}{T_l}\int_{t-T_l}^{t} \mathrm{d}t' \int_{t-T_l-t'}^{t-t'} \mathrm{d}s\, W(s) \langle S_i^{\mathrm{in}}(t'+s) S^{\mathrm{out}}(t') \rangle \right\}\ . \end{aligned} \tag{27}$$

[3]Since there is no fear of confusion we are using the same notation S for spins and spike flows.

It is evident that both types of averaging, over randomness and over time, have been taken into account. So far so good. The first term on the right, the time average $\overline{\langle S_i^{\rm in}(t)\rangle}$ of the rate function $\langle S_i^{\rm in}(t')\rangle$ for times t' in the interval $[t-T_l, t)$, is a mean which we call $\nu_i^{\rm in}(t)$. For an inhomogeneous Poisson process (see Appendix B) this is nothing but the mean intensity $\bar{\lambda}_i(t)$ where the probability of finding one spike in an interval of length Δt near t is $\lambda_i(t)\Delta t$. If λ_i is a periodic function (in the auditory system often for frequencies in the kHz range), then its time average is a constant so that the time dependence is gone and $\nu_i^{\rm in}(t) \equiv \bar{\nu}_i$. The second term, the time average of $\langle S^{\rm out}(t)\rangle$, which is to be called $\nu^{\rm out}(t)$, is harder to compute since it entails both the outgoing and all the incoming processes, the latter "deciding" together when an action potential will be generated. For later reference we summarize the above two definitions,

$$\nu_i^{\rm in}(t) := \overline{\langle S_i^{\rm in}(t')\rangle}\,, \qquad \nu^{\rm out}(t) := \overline{\langle S^{\rm out}(t)\rangle}\,. \tag{28}$$

The former refers to the input only, the latter takes the output by itself.

The truly hard nut is the double integral in (27), explicitly correlating input and output – a distinguishing property of Hebbian learning. Let us take a "typical" t', say $t' = t - T_l + xT_l$ with $0 < x < 1$. Then the lower bound of the integral over s is effectively $-xT_l$ while the upper bound is $(1-x)T_l$. The learning window W is something *local in time*; for the auditory system of the order of milliseconds, for most of the cortex seconds – anyway, much, much shorter than T_l. Hence for our "typical" t' the lower bound of the integral over s is $-\infty$ whereas the upper bound is $+\infty$ so that, up to a negligible error, we are left with

$$\begin{aligned}&\frac{1}{T_l}\int_{t-T_l}^{t} {\rm d}t' \int_{-\infty}^{\infty} {\rm d}s\, W(s)\langle S_i^{\rm in}(t'+s)S^{\rm out}(t')\rangle \\ &= \int_{-\infty}^{\infty} {\rm d}s\, W(s)\frac{1}{T_l}\int_{t-T_l}^{t} {\rm d}t'\,\langle S_i^{\rm in}(t'+s)S^{\rm out}(t')\rangle\,.\end{aligned} \tag{29}$$

Returning to (27), we note that we can transform it into a differential equation since $\Delta J_i(t) = J_i(t) - J_i(t-T_l)$ and, due to the adiabatic hypothesis, the change of J_i is so slow that $\Delta J_i(t)/T_l$ can be replaced by ${\rm d}J_i/{\rm d}t$. In other words, we choose T_l so large that it greatly exceeds all neuronal times, e.g., interspike intervals and the width of the learning window W, but on the other hand is much smaller than η^{-1} – all in all, a condition fully consistent with the Hebbian philosophy "practice makes perfect". That is to say, T_l separates neuronal and learning time scales. Then we find, using (27), (28), and (29),

$$\frac{\rm d}{{\rm d}t}J_i = \eta\left[w^{\rm in}\,\nu_i^{\rm in} + w^{\rm out}\,\nu^{\rm out} + \int_{-\infty}^{\infty} {\rm d}s\, W(s)\,\frac{1}{T_l}\int_{t-T_l}^{t} {\rm d}t'\,\langle S_i^{\rm in}(t'+s)S^{\rm out}(t')\rangle\right] \tag{30}$$

This equation is exact and describes the time evolution of *infinitesimal* synaptic plasticity for a neuron with given inputs.

It is a nice aspect of (30) that the final integral over t' is nothing but the time-averaged correlation function. The correlation function itself is $\langle S_i^{\rm in}(t'')\, S^{\rm out}(t')\rangle$. We may interpret it as the joint probability density for observing an input spike at synapse i at the time t'' and an output spike at time t'. Hence we write

$$C_i(s,t) := \frac{1}{T_l}\int_{t-T_l}^{t} {\rm d}t' \langle S_i^{\rm in}(t'+s) S^{\rm out}(t')\rangle = \overline{\langle S_i^{\rm in}(t+s) S^{\rm out}(t)\rangle}\ , \qquad (31)$$

the second equality being just a definition. Altogether we get a synaptic dynamics of appealing simplicity,

$$\boxed{\frac{\rm d}{{\rm d}t} J_i = \eta\left[w^{\rm in}\,\nu_i^{\rm in} + w^{\rm out}\,\nu^{\rm out} + \int_{-\infty}^{\infty} {\rm d}s\, W(s)\, C_i(s,t)\right]\ .} \qquad (32)$$

In this form the learning equation is easy to remember: the input rate $\nu_i^{\rm in}$ modifies the synaptic efficacy through $w^{\rm in}$, the output rate $\nu^{\rm out}$ does so through $w^{\rm out}$, and the Hebbian correlation function C_i favors or disfavors it through the learning window W.

Appearances are deceiving, however. Not only do $\nu^{\rm out}$ and C_i depend on J_i but also, through $S^{\rm out}$, on all the other J_j with $j \neq i$. Moreover, neuronal firing is intrinsically nonlinear. Hence synaptic dynamics is an intricate collective process. Figure 8 gives an illustration of what may, and often does, happen. Inspired by the barn-owl case [45], we imagine a set of axonal delay lines contacting a neuron. They exhibit a uniform distribution of delays and of synaptic strengths (> 0) to start with. The neuron is of the integrate-and-fire type and the solution as shown is numerically exact. As one sees, there is first a symmetry breaking where certain synapses "grow" faster than the rest. The fact that certain delays are favored makes that their 'auto'correlation function exceeds that of more 'distant' axonal delays. In this way they grow faster and the others deteriorate in their role of "those that come too late". This initial stage is characterized by an exponential growth (or decrease), the rationale of which will be illustrated by the Duhamel formula (49) below. The next stage where a few synapses are favored and the rest is eliminated is hard to characterize since it is governed by a nonlinear dynamics due to the integrate-and-fire neuron. The final stage is a simple saturation at an upper bound determined by finite synaptic resources or a lower bound, zero, where nothing is left. What we see is a kind of evolutionary process where only a few axonal delay lines, the "fittest", survive.

In the next section we will study an exactly soluble neuronal model that allows a disentanglement of the different inputs and, in this way, provides a more precise feeling for what is going to happen.

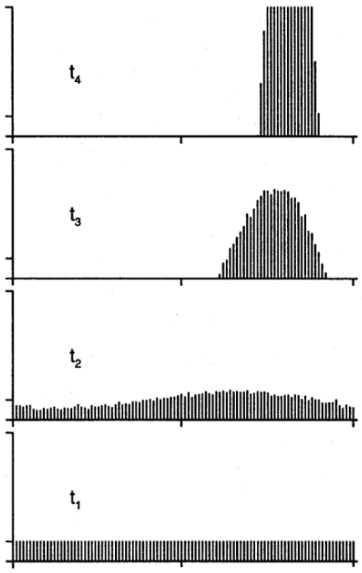

Figure 8*:* Bottom to top, selection of synapses during pattern formation of the synaptic connectivity on an integrate-and-fire neuron. As shown by the panel at time t_1, a set of axonal delay lines with a uniform distribution of synaptic strengths and delays in the interval $[0, T]$ gets Poissonian input with frequency $\omega = 2\pi/T$. The panel at time t_2 exhibits symmetry breaking and exponential growth; the latter is a simple consequence of linearizing (32). At time t_3 saturation sets in and at t_4 we have reached full saturation, a stationary state characterized by "survival of the fittest".

5 Disentangling Synaptic Inputs: the Poisson Neuron

Equations (30) and (32) tell us that "all we need" for deriving the time evolution of the synaptic efficacies is $\nu^{\rm out}$ and the function C_i correlating $S_i^{\rm in}$ and $S^{\rm out}$. In a threshold model, such as the Spike Response Model (§2), the $S_i^{\rm in}$ together determine $S^{\rm out}$ in a *non*linear way because of the threshold. Disentangling input and output and obtaining exact solutions is thus prohibitively difficult. We therefore introduce a model, the Poisson neuron, that allows for an exact solution [47] of the synaptic dynamics (32) by circumventing the threshold but keeping the firing rate.

5.1 Poisson Neuron: Definition and Properties

Spikes originate, so to speak, from the potential $v(t)$ as given by (3). We now define the *Poisson neuron* to be the inhomogeneous Poisson process (see Appendix B) with rate function, or intensity,

$$\lambda^{\rm out}(t) = \nu_0 + v(t) = \nu_0 + \sum_{i,f} J_i(t_i^f)\varepsilon(t - t_i^f) \geq 0 \qquad (33)$$

where ν_0 is a spontaneous firing rate. A Poisson process is defined by three properties: (i) the probability of finding a spike between t and $t + \Delta t$ is $\lambda^{\rm out}(t)\Delta t$, (ii) the probability of finding two or more spikes there is $o(\Delta t)$, and (iii) the process has independent increments, i.e., events in disjoint intervals are independent. When the potential $v(t)$ in (33) is high/low, the probability of getting a spike is

high/low too. The input processes (26) are taken to be Poisson as well, a reasonable, often even realistic, assumption. For those who like an explicit nonlinearity better, the *clipped* Poisson neuron with

$$\lambda^{\rm out}_{\rm clipped} = \nu_1 \Theta[v(t) - \vartheta_1] \tag{34}$$

and Θ as the Heaviside step function of (11) is a suitable substitute that also allows an exact disentanglement [63]. In fact, practically any function of v will do [64].

For the sake of convenience we require that the integral over ε (instead of ε's maximum) be one. We start by noting that $\langle S^{\rm out}\rangle = \langle \lambda^{\rm out}(t)\rangle_{\rm in}$ where the former average is over both the output process, i.e., the Poisson neuron, and over the input processes whereas the latter is over the input processes only; hence the lower index 'in' served here as a reminder. Using (33) we then get

$$\langle S^{\rm out}\rangle = \nu_0 + \sum_{i=1}^{N} J_i(t) \int_0^\infty {\rm d}s\, \varepsilon(s) \lambda_i^{\rm in}(t-s) =: \nu_0 + \sum_{i=1}^{N} J_i(t) \Lambda_i^{\rm in}(t) \ . \tag{35}$$

The first equality in (35) is as in the transition from (78) to (80) in Appendix B. In agreement with the previous section, J_i has been treated as an adiabatic variable and, thus, is taken to be constant on the time scale of T_l. In (35) it can therefore be evaluated at time t. The final equality defines $\Lambda_i^{\rm in}$ as the convolution of ε and $\lambda_i^{\rm in}$.

To compute $\langle S_i^{\rm in}(t+s) S^{\rm out}(t)\rangle$ in (30) and disentangle input and output, we exploit the properties of a Poisson process. For the moment we put $\nu_0 = 0$ and define $h_i(t) := J_i(t) \sum_f \varepsilon(t - t_i^f)$. Performing the average associated with our Poisson neuron, viz., (33), we find

$$\langle S_i^{\rm in}(t+s) S^{\rm out}(t)\rangle = \langle S_i^{\rm in}(t+s) [h_i(t) + \sum_{j(\neq i)} h_j(t)]\rangle \ . \tag{36}$$

The h_j with $j \neq i$ and $S_i^{\rm in}$ being independent, the average of their product is a simple computation as it factorizes so that we can use (35). The result is

$$\langle S_i^{\rm in}(t+s) \sum_{j(\neq i)} h_j(t)\rangle = \lambda_i^{\rm in}(t+s) \sum_{j(\neq i)} \Lambda_j^{\rm in}(t) J_j(t) \ . \tag{37}$$

It it will be recollected in (41) below.

The average $\langle S_i^{\rm in}(t+s) h_i(t)\rangle$ is over the input process (26),

$$\Big\langle \Big[\sum\nolimits_{f'} \delta(t+s-t_i^{f'})\Big] \times \Big[J_i(t) \sum\nolimits_f \varepsilon(t-t_i^f)\Big]\Big\rangle \ . \tag{38}$$

The correlations are explicitly present in the arrival times $t_i^{f'}$ and t_i^f of the spikes as they hit synapse i. The disentanglement that is to come is exactly as in the

transition from (83) to (84) in Appendix B. We approximate the delta function in (38) by the normalized indicator function $(\Delta t)^{-1}\mathbb{1}_{\{\text{spikes in } [t_l, t_l+\Delta t)\}}(\omega)$ with $\Delta t \to 0$. Here ω is sampling our probability space, i.e., the collection of random events, so that averaging means integrating over ω. Furthermore, we discretize the time axis by breaking it into intervals $[t_k, t_k + \Delta t)$ of length Δt and take $t_l := t + s$ (for $k := l$) as an end point of one of the intervals. Keeping an eye on (38) and $J_i(t)$ on ice, we have to evaluate averages of the form

$$(\Delta t)^{-1} \left\langle \mathbb{1}_{\{\text{spikes in } [t_l, t_l+\Delta t)\}}(\omega) \sum\nolimits_k \mathbb{1}_{\{\text{spikes in } [t_k, t_k+\Delta t)\}}(\omega)\, \varepsilon(t - t_k) \right\rangle \ . \tag{39}$$

As long as $\Delta t > 0$ we have to take into account that more than one spike may occur in an interval; 'no spike' is easy because it gives nothing. As $\Delta t \to 0$, the probability of getting more than one spike in an interval of length Δt is $o(\Delta t)$, in perfect agreement with neuronal refractoriness, and thus may be neglected. In the sum occurring in (39) we separate the term $k = l$ from the rest and note that for $k = l$ we get $\mathbb{1}^2 = \mathbb{1}$ whereas the events $k \neq l$ are independent so that expectation values factorize. Remembering $t_l = t + s$, we then end up with

$$\begin{aligned} & \lambda_i^{\text{in}}(t_l) \left[\varepsilon(t - t_l) + \sum_{k(\neq l)} \lambda_i^{\text{in}}(t_k) \varepsilon(t - t_k) \Delta t \right] \\ = \ & \lambda_i^{\text{in}}(t + s) \left[\varepsilon(-s) + \sum_{k(\neq l)} \lambda_i^{\text{in}}(t_k) \varepsilon(t - t_k)\, \Delta t \right] \ . \end{aligned} \tag{40}$$

In the limit $\Delta t \to 0$, the above Riemann sum converges to its integral $\int \mathrm{d}t'\, \lambda_i^{\text{in}}(t')$ $\varepsilon(t - t')$, which is nothing but $\Lambda_i^{\text{in}}(t)$. Collecting terms and reinstalling (37), $\nu_0 > 0$, and $J_i(t)$, we obtain

$$\langle S_i^{\text{in}}(t + s) S^{\text{out}}(t) \rangle = \lambda_i^{\text{in}}(t + s) \left[\nu_0 + J_i(t) \varepsilon(-s) + \sum_{j=1}^{N} J_j(t) \Lambda_j^{\text{in}}(t) \right] \tag{41}$$

where, except for $j = i$, the sum stems from (37). We note that $\varepsilon(-s) \neq 0$ only if $s < 0$. In view of causality, this makes sense since $S_i^{\text{in}}(t + s)$ can influence $S^{\text{out}}(t)$ only if $s < 0$. In fact, the term $\lambda_i^{\text{in}}(t + s) J_i(t) \varepsilon(-s)$ incorporates the way in which an input spike at synapse i at time $t + s$ influences the neuronal output at time t through $\varepsilon(-s)$ and, hence, is correlated with itself. The sum represents the influence of 'other' times ($j = i$) and other synapses ($j \neq i$).

Time-averaging (41) is trivial. We insert (41) in (31) and (32) and invoke the adiabatic hypothesis of the previous section so as to find

$$C_i(s, t) = \overline{\lambda_i^{\text{in}}(t + s)} \left[\nu_0 + J_i(t) \varepsilon(-s) \right] + \sum_{j=1}^{N} J_j(t) \overline{\lambda_i^{\text{in}}(t + s) \Lambda_j^{\text{in}}(t)} \ . \tag{42}$$

By definition, $\overline{\lambda_i^{\rm in}(t)} = \nu_i^{\rm in}(t)$. Combining (32) and (42) we obtain for $1 \leq i \leq N$,

$$\frac{\mathrm{d}}{\mathrm{d}t} J_i = \eta \left[\!\!\left[w^{\rm in}\, \nu_i^{\rm in} + w^{\rm out} \left[\nu_0 + \sum_{j=1}^N J_j \overline{\Lambda_j^{\rm in}}(t) \right] + \right.\right.$$
$$\left.\left. \int_{-\infty}^{\infty} \mathrm{d}s\, W(s) \left\{ \overline{\lambda_i^{\rm in}(t+s)}\, [\nu_0 + J_i \varepsilon(-s)] + \sum_{j=1}^N J_j \overline{\lambda_i^{\rm in}(t+s) \Lambda_j^{\rm in}(t)} \right\} \right]\!\!\right] . \quad (43)$$

Functions of time that carry no argument are to be taken at time t. A single glance suffices to convince us that (43) is a *linear* differential equation. The time averages, though slowly varying, might still depend on time. If so, the solution is standard [78, 79] but very hard, and explicit expressions can only be obtained numerically. If, on the other hand, the λ_i are periodic functions of time, as in the auditory system, the time averages are (practically) constant and an analytic approach is within reach, at least for the mean activity.

Throughout what follows we drop the prefactor η from the dynamics by rescaling time through the substitution $\eta t := t$ and redefining all functions in sight that depend on time; to this end we use (43) and bring η to the left. Alternatively, we measure everything in units of size η. We define a few quantities,

$$\begin{aligned} a_i &= w^{\rm in} \nu_i^{\rm in} + \nu_0 [w^{\rm out} + \int_{-\infty}^{\infty} \mathrm{d}s\, W(s) \overline{\lambda_i^{\rm in}(t+s)}] \ , \\ b_i &= \overline{\Lambda_i^{\rm in}(t)}\, w^{\rm out} \ , \\ c_i &= \int_{-\infty}^{\infty} \mathrm{d}s\, W(s) \varepsilon(-s) \overline{\lambda_i^{\rm in}(t+s)} \ , \\ Q_{ij}(t) &= \int_{-\infty}^{\infty} \mathrm{d}s\, W(s)\, \overline{\lambda_i^{\rm in}(t+s) \Lambda_j^{\rm in}(t)} \ , \end{aligned} \quad (44)$$

with $\Lambda_j^{\rm in}(t) = \int \mathrm{d}s\, \varepsilon(s) \lambda_i^{\rm in}(t-s)$ as in (35), and find for $1 \leq i \leq N$

$$\frac{\mathrm{d}}{\mathrm{d}t} J_i = a_i + \sum_{j=1}^N b_j J_j + c_i J_i + \sum_{j=1}^N Q_{ij} J_j \ . \quad (45)$$

We can rewrite (45) in terms of the vector $\mathbf{J} = (J_i) \in \mathbb{R}^N$ by defining the diagonal matrix $C = \mathrm{diag}\{c_1, c_2, \ldots, c_N\}$, using Dirac's bra-ket notation[4] and introducing the N-vectors $|\mathbf{1}\rangle = (1, \ldots, 1)$, $\mathbf{a} = (a_i)$, and $\mathbf{b} = (b_i)$ so that (45) reappears in the form

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{J} = \mathbf{a} + (|\mathbf{1}\rangle\langle\mathbf{b}| + C + Q)\mathbf{J} \ . \quad (46)$$

[4]Dirac's imagination has led to an appealingly simple notation. The idea is this. A Hilbert space $\mathcal{H}$ is a vector space with inner product $\mathcal{H} \times \mathcal{H} \ni \{x, y\} \mapsto \langle x|y\rangle \in \mathbb{C}$, which is taken to be *linear* in the right-hand side, viz., y. The resulting inner product looks like a bracket so that Dirac called $\langle x|$ a 'bra' and $|y\rangle$ a 'ket'. Vectors in our Hilbert space are kets and written $|y\rangle$. Then the operator $\mathbb{P} = |z\rangle\langle x|$ is projector-like and bound to operate on vectors $|y\rangle$ in such a way that $\mathbb{P}|y\rangle = |z\rangle\langle x|y\rangle \propto |z\rangle$. See the literature [88, §14.4] for additional information.

As it will turn out in §5.5 that C does not dominate the asymptotic behavior of (46) we simplify it by taking $c_i \equiv c$ so that $C = c\mathbb{1}$, the spectral theory of $C+Q$ is reduced to that of Q, and we are left with

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{J} = \mathbf{a} + (|\mathbf{1}\rangle\langle\mathbf{b}| + c\mathbb{1} + Q)\mathbf{J} \ . \tag{47}$$

Input channels i and j whose delays are fixed give rise to a specific matrix element Q_{ij} that takes the delay structure into account. For input processes that are Poissonian with periodic intensity, the ν_i^{in} and ν^{out} as defined by (28) are constants when T_l is large enough, and so are the Q_{ij}. It is important to realize that this statement is true on the time scale of T_l but need *not* hold on that of the total learning time.

Finally, the learning equation (47) allows the analysis of the influence of noise on long-term synaptic plasticity. Learning results from stepwise, infinitesimally small weight changes: "Practice makes perfect". With noise, each weight performs a random walk whose expectation value is described by the ensemble-averaged equation (47). For an analysis of noise as a deviation from the mean the reader is referred to the literature [47].

5.2 Relation to Rate-Based Hebbian Learning

In neural network theory, Hebb's ideas [28] have usually been formulated as learning rules where the change of a synaptic efficacy J_i depends on the correlation between the mean firing rate ν_i^{in} of the i th presynaptic neuron and the mean firing rate ν^{out} of a postsynaptic neuron, viz.,

$$\frac{\mathrm{d}}{\mathrm{d}t}J_i := d_0 + d_1\,\nu_i^{\mathrm{in}} + d_2\,\nu^{\mathrm{out}} + d_3\,\nu_i^{\mathrm{in}}\,\nu^{\mathrm{out}} + d_4\,(\nu_i^{\mathrm{in}})^2 + d_5\,(\nu^{\mathrm{out}})^2 \ , \tag{48}$$

$d_0 < 0$ and $d_1, \ldots, d_5$ being proportionality constants. Apart from the decay term d_0 and the "Hebbian" term $\nu_i^{\mathrm{in}}\,\nu^{\mathrm{out}}$ proportional to the product of input and output rates, there are also synaptic changes which are driven by the pre- and postsynaptic rates separately. The parameters $d_0, \ldots, d_5$ may depend on J_i. Equation (48) is a general ansatz with terms up to second order *in the rates*; see, e.g., [80, 81, 82]. In case $d_0 = d_4 = d_5 = 0$ and under the (strong) assumption that S^{in} and S^{out} are independent, it is straightforward to derive (48) from (32) directly. Alternatively, one can obtain (48) from (45).

Linsker [80] has derived a mathematically equivalent equation to (47) by starting from (48) and using a linear graded-response neuron, a rate-based model. The difference between Linsker's equation and (45) is, apart from a slightly different notation, the term $c\mathbb{1}$ and the fact that (45) has been derived from underlying processes, viz., spikes, whereas Linsker's equation is the result of the rate *ansatz* (48). The present approach is far more comprehensive. Correlations between

spikes on time scales down to milliseconds or below can therefore enter the driving term Q *for structure formation*. They may be, and I expect are, essential for information processing in neuronal networks such as auditory and electro-sensory systems [83].

The mathematics of (47) has been analyzed extensively by MacKay and Miller [84] in terms of eigenvectors and eigenfunctions of the matrix $|\mathbf{1}\rangle\langle\mathbf{b}| + Q$ with $\mathbf{b} = b\mathbf{1}$ and $c = 0$. The matrix $(Q_{ij} + b + c\delta_{ij})$ in (47) contains c times the unit matrix and thus has the same eigenvectors as $(Q_{ij} + b)$ while the eigenvalues are simply shifted by c.

5.3 Synaptic Dynamics and Self-Normalization

Normalization of the average synaptic efficacy or of the mean output activity is a very desirable property for any synaptic dynamics. After all, the mean output rate should not blow up during learning but converge to a finite value in an acceptable amount of time. Standard rate-based Hebbian learning, however, can lead to unbounded growth. Several methods have been designed to control this unbounded growth, such as subtractive and multiplicative rescaling of the weights after each learning step so as to impose, e.g., $\sum_j J_j = \text{Const.}$ or $\sum_j J_j^2 = \text{Const.}$ [85]. Most of these methods make use of the J_j dependence of the parameters $d_1, \ldots, d_5$ in the learning equation (48). Mathematically they do what they ought to do but, from the point of view of biological physics, it is unclear where they come from. Hence we will derive self-normalization from scratch.

We are going to show under what conditions the arithmetic mean $J^{\text{av}} = N^{-1}\sum_i J_i$ of the synaptic efficacies, and hence the mean neuronal output, converges to a finite limit as $t \to \infty$. In other words, we focus on the question of *how* we can get self-normalization. The first result in this direction, an even more pronounced form for integrate-and-fire neurons (see §2) in the general context of (32), was found numerically by Gerstner et al. [45]. We will see that, for a realistic scenario, we need $a_i > 0$ and $b_i < 0$, whatever i. The former condition is evident once we realize that naive synaptic efficacies start at $J_i \approx 0$; after all, where else? Then (47) is nothing but $\mathrm{d}\mathbf{J}/\mathrm{d}t = \mathbf{a}$. For excitatory synapses starting at $J_i = 0$ it would be good to increase. Hence we cannot but require $a_i > 0$. We will assume the excitatory case throughout what follows; inhibition can be treated analogously.

The linear equation (47) is of the form $\mathrm{d}\mathbf{J}/\mathrm{d}t = M\mathbf{J} + \mathbf{a}$ with M being the matrix $(|\mathbf{1}\rangle\langle\mathbf{b}| + c\mathbb{1} + Q)$. As long as M is fixed, the synaptic dynamics can be solved explicitly through Duhamel's formula,

$$\mathbf{J}(t) = \exp[(t - t_0)M]\mathbf{J}(t_0) + \int_{t_0}^{t} \mathrm{d}s\, \exp[(t - s)M]\mathbf{a}\ . \tag{49}$$

The right-hand side of (49) satisfies (47) and equals $\mathbf{J}(t_0)$ at time $t = t_0$. Hence it is the solution $\mathbf{J}(t)$ we are looking for. In the asymptotic limit $t \to \infty$, the vector

a may depend on s, i.e., time, so that $\mathbf{a} = \mathbf{a}(s)$. Duhamel's formula is as valid as it was for constant **a**. To incorporate a time-dependence of M one has to replace $\exp[(t-t_0)M]$ by the corresponding solution operator $U(t,t_0)$. Since both $\mathbf{a}(s)$ and $M(t)$ in general preclude any analytic solution, we will not pursue the issue here but take both constant. Let us, then, suppose first that all eigenvalues of M, the so-called spectrum $\sigma(M)$, have strictly negative real parts. Accordingly we get $\exp(tM)\mathbf{J}(t_0) \to 0$ as $t \to \infty$ while in the very same limit the integral gives $-M^{-1}\mathbf{a}$, the fixed point of the differential equation. So it all fits, provided $-M^{-1}\mathbf{a}$ has all components nonnegative, which may well happen. We need to keep in mind, though, that the key assumption on $\sigma(M)$ remains a bit hard to verify.

We did not require M to be diagonable. In fact, except for $\sigma(M)$, we did not assume anything yet. To get sharp analytic results we now suppose (i) $b_i = b$ for all $1 \le i \le N$, which implies $|\mathbf{b}\rangle = b|\mathbf{1}\rangle$, and (ii) a specific commutator vanishes,

$$[Q, |\mathbf{1}\rangle\langle\mathbf{1}|] = 0 \ . \tag{50}$$

Thus all row and column sums of Q equal a single number qN since $|\mathbf{1}\rangle\langle\mathbf{1}|$ is the matrix whose elements all equal 1. The matrix Q being $N \times N$, q tells us how big/small a "typical" matrix element is. Phrased differently, the sum qN has the right scaling behavior as N becomes large. In passing we note that one can do with slightly less [86]; say, $[Q, |\mathbf{1}\rangle\langle\mathbf{b}|] = 0$ and variations thereof. There are at least two consequences. First, the spectral theory of $\mathbb{P} := |\mathbf{1}\rangle\langle\mathbf{1}\,|$ and Q refers to two different things that can be sorted out separately and together determine the effect of $M = b\mathbb{P} + c\mathbb{1} + Q$ in that $\exp(tM) = \exp(ct)\exp(tb\mathbb{P})\exp(tQ)$. The only eigenvector of the Hermitian $\mathbb{P}$ with nonzero eigenvalue $(= N)$ is $|\mathbf{1}\rangle = (1,\ldots,1)$; the eigenvalue N is non-degenerate so that[5] $|\mathbf{1}\rangle$ is also an eigenvector of Q. Alternatively, it is a direct outcome of (50) that the corresponding eigenvalue is qN. Second, the differential equation governing the dynamics of $J^{\rm av}$ is

$$\mathrm{d}J^{\rm av}/\mathrm{d}t = a^{\rm av} + (Nb + c + Nq)J^{\rm av} \ . \tag{51}$$

with $a^{\rm av} := N^{-1}\sum_j a_j > 0$. Let us now put $m := Nb + c + Nq$. Then the solution of (51) is again given by Duhamel's formula (with $t_0 = 0$),

$$J^{\rm av}(t) = e^{tm}J^{\rm av}(0) + m^{-1}(e^{tm} - 1)a^{\rm av} \ . \tag{52}$$

To get a finite result we simply require $m < 0$. Then $J^{\rm av}(t)$ approaches the fixed point $J^{\rm av}_* := -a^{\rm av}/m > 0$ of the differential equation. The fixed point is asymptotically stable if and only if $m < 0$. If (50) does not hold, not even approximately, things become a bit harder.

[5] Suppose two matrices A and B commute, i.e., $[A,B] := AB - BA = 0$. Let **a** be an eigenvector of A with nondegenerate eigenvalue α. That is, $A\mathbf{a} = \alpha\mathbf{a}$. Then $BA\mathbf{a} = A(B\mathbf{a}) = \alpha(B\mathbf{a})$ so that, α being nondegenerate, $B\mathbf{a} = \beta\mathbf{a}$. In other words, **a** is also an eigenvector of B. For degenerate eigenvalues of A the present argument breaks down.

5.4 Asymptotics and Structure Formation

By now there is no harm in starting with a matrix M that has a few eigenvalues with a strictly positive real part and calling those with largest and second-largest real part λ_1 and λ_2; for the sake of convenience, we also assume they are nondegenerate. We return to Duhamel's formula (49), viz., its upshot for constant $\mathbf{a}$ and constant M,

$$\mathbf{J}(t) = \exp(tM)\,\mathbf{J}(0) + M^{-1}[\exp(tM) - 1]\,\mathbf{a}\ , \tag{53}$$

which is the matrix version of (52) with $t_0 = 0$. If the matrix M is diagonable and the real part $\Re\lambda_1$ of λ_1 is appreciably bigger than $\Re\lambda_2$, we need only know the normalized eigenvectors $\mathbf{e}_1$ and $\mathbf{g}_1$ with $\langle\mathbf{g}_1|\mathbf{e}_1\rangle = 1$ belonging to the "largest" eigenvalues λ_1 and λ_1^* of M and its Hermitian conjugate (i.e., adjoint) $M^\dagger$ as they determine the leading contribution $\lambda_1|\mathbf{e}_1\rangle\langle\mathbf{g}_1|$ in the biorthogonal expansion [87, §11.23] of M in structure formation[6]. Equation (53) then tells us that, after an initial phase with $t \approx 0$, there is exponential growth along $\mathbf{e}_1$.

Since $\Re\lambda_1 > 0$ the nonzero components of $\mathbf{J}(t)$ are bound to blow up or decrease to $-\infty$ as t becomes large. This is of course unrealistic since synaptic resources are finite. For excitatory synapses we therefore assume an upper bound J^u, with $0 < J^u < \infty$, and a lower bound 0. If the efficacy of synapse i has reached J^u, it will stay there as long as its time derivative $J_i'(t)$ is positive. On the other hand, once $J_i'(t) < 0$ it may decrease. For the lower bound the argument is just the opposite. We thus see that sooner or later, with the timing depending on $\Re\lambda_1$, we get saturation of (53), i.e., of

$$\mathbf{J}(t) \approx \exp(t\lambda_1)\,[\langle\mathbf{g}_1|\mathbf{J}(0)\rangle + \lambda_1^{-1}\langle\mathbf{g}_1|\mathbf{a}\rangle]\mathbf{e}_1 - M^{-1}\mathbf{a}\ . \tag{54}$$

Once a component of $\mathbf{J}$ reaches the upper or lower bound the problem becomes nonlinear. We then take it out, fix it, and continue with the remaining problem, which is again linear; and so on. Though implausible, a 'fixed' component is allowed to return to the interior of $[0, J^u]$ once its time derivative points into

[6] If M is diagonable, then it has N independent eigenvectors $\mathbf{e}_i$ that constitute the columns of a matrix T with $T^{-1}MT = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ (i). Hence $T^\dagger M^\dagger (T^\dagger)^{-1} = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_N^*)$ (ii) and the columns $\mathbf{g}_i$ of the matrix $(T^\dagger)^{-1}$ are eigenvectors of $M^\dagger$ with eigenvalues λ_i^*. For nondegenerate eigenvalues it is a simple argument to show $\langle\mathbf{g}_j|\mathbf{e}_i\rangle = 0$ for $i \neq j$: $\langle M^\dagger\mathbf{g}_j|\mathbf{e}_i\rangle = \langle\mathbf{g}_j|M\mathbf{e}_i\rangle$ so that $\lambda_j\langle\mathbf{g}_j|\mathbf{e}_i\rangle = \langle\mathbf{g}_j|\mathbf{e}_i\rangle\lambda_i$ with $\lambda_i \neq \lambda_j$. Now $\langle\mathbf{g}_i|\mathbf{e}_i\rangle \neq 0$ because otherwise $\mathbf{g}_i = 0$, so that we can put the inner product equal to 1 and find $M = \sum_i \lambda_i|\mathbf{e}_i\rangle\langle\mathbf{g}_i|$. In addition, $\exp(tM) = \sum_i \exp(t\lambda_i)|\mathbf{e}_i\rangle\langle\mathbf{g}_i|$. For self-adjoint $M = M^\dagger$ we are back at the ordinary spectral representation. The reader may consult Merzbacher [88, §14.4] for a detailed account of Dirac's convenient bra-ket notation. In fact, the only condition on M that is needed for a biorthogonal expansion is that M be diagonable. Then (i) says $T = (\mathbf{e}_1, \ldots, \mathbf{e}_N)$ and (ii) asserts $(T^{-1})^\dagger = (\mathbf{g}_1, \ldots, \mathbf{g}_N)$. Hence biorthogonality is equivalent with $(T^{-1})T = \mathbb{1}$, which is evident. The expansion itself can be verified on a complete set of eigenvectors of M, viz., $\{\mathbf{e}_i;\, 1 \leq i \leq N\}$. QED

the 'right' direction. With the benefit of hindsight we can now formulate "self-normalization" to be a limit state where $0 < J^{\rm av} = N^{-1}\sum_i J_i < J^u$; for inhibitory synapses, or mixtures, the statement is to be modified accordingly.

5.5 Simple Example of Structure Formation

We finish this section by studying a simple, exactly soluble, case. To this end we divide the N statistically independent synapses into two groups, $\mathcal{N}_1$ and $\mathcal{N}_2$ with N_1 and N_2 synapses, respectively, and $N_1 + N_2 = N$ while $N_1,\ N_2 \gg 1$. Since each group contains many synapses, we may assume that N_1/N and N_2/N are of the same order of magnitude.

The spike input at synapses in group $\mathcal{N}_1$ is generated by a Poisson process with a constant intensity $\lambda_i^{\rm in}(t) \equiv \nu^{\rm in}$, which for $i \in \mathcal{N}_1$ is taken to be time-independent. Using the definition (44), we therefore get $Q_{ij}(t) \equiv Q_{11}$ for i and/or $j \in \mathcal{N}_1$.

The synapses $i \in \mathcal{N}_2$ are driven by some time-dependent input, $\lambda_i^{\rm in}(t) = \lambda^{\rm in}(t)$ with the same mean input rate $\overline{\lambda^{\rm in}(t)} = \nu^{\rm in}$ as in group $\mathcal{N}_1$. Without going into details about the dependence of $\lambda^{\rm in}(t)$ upon the time t we simply assume $\lambda^{\rm in}(t)$ to be such that $Q_{ij}(t) \equiv Q_{22}$ for $i, j \in \mathcal{N}_2$ and regardless of t while $Q_{ij}(t) \equiv Q_{11}$ in all other cases. Here Q_{11} and Q_{22} are constants and we have used (44). For the sake of simplicity we require in addition that $Q_{22} > Q_{11}$. In summary, we suppose in the following

$$Q_{ij}(t) = \begin{cases} Q_{22} > Q_{11} & \text{for } i, j \in \mathcal{N}_2 \\ Q_{11} & \text{otherwise} \end{cases} \tag{55}$$

We recall that Q_{ij} is a measure of the correlations in the input arriving at synapses i and j; cf. (44). Equation (55) tells us that at least some synapses receive more positively correlated input than the rest, a rather natural assumption. As one may expect in the animal kingdom, some synapses are "more equal" than others.

We now examine the evolution of the average weight in each of the two groups $\mathcal{N}_1$ and $\mathcal{N}_2$ and put

$$J_1^{\rm av} = \frac{1}{N_1}\sum_{i\in\mathcal{N}_1} J_i\ , \qquad J_2^{\rm av} = \frac{1}{N_2}\sum_{i\in\mathcal{N}_2} J_i\ . \tag{56}$$

As long as lower and upper bounds do not influence the dynamics, the corresponding rates of change are determined by (47),

$$\frac{\rm d}{{\rm d}t}\begin{pmatrix} J_1^{\rm av} \\ J_2^{\rm av} \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} bN_1 + c & bN_2 \\ bN_1 & (b+Q)N_2 + c \end{pmatrix}\begin{pmatrix} J_1^{\rm av} \\ J_2^{\rm av} \end{pmatrix} \tag{57}$$

where we have put $b := b + Q_{11}$ and $Q := Q_{22} - Q_{11} > 0$; the inequality is by assumption. Obtaining an explicit solution to the above equation is straightforward once we realize its relation to a quantum spin 1/2.

The matrix M appearing on the right in (57) is a linear combination $M = n_0 \mathbb{1} + \mathbf{n} \cdot \boldsymbol{\sigma}$ of the unit matrix $\mathbb{1}$ and the Pauli spin matrices [88, §13.6],

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{58}$$

Here $i = \sqrt{-1}$, the center dot in $\mathbf{n} \cdot \boldsymbol{\sigma}$ denotes a scalar product, and $\mathbf{n} \in \mathbb{R}^3$ if and only if the matrix M is Hermitian. If we want to use the Duhamel formula (53) – we do – we need to compute $\exp(tM)$. This is easy since simple algebra based on $\sigma_x \sigma_y = i\sigma_z$ *et cycl.* or any decent quantum mechanics book [88, §13.6] shows that $(\mathbf{n} \cdot \boldsymbol{\sigma})^{\mathbf{2}} = (\mathbf{n} \cdot \mathbf{n})\, \mathbb{1}$ and consequently

$$\exp(tM) = e^{n_0 t} \left[\cosh(\sqrt{\mathbf{n} \cdot \mathbf{n}}\, t) + \frac{\sinh(\sqrt{\mathbf{n} \cdot \mathbf{n}}\, t)}{\sqrt{\mathbf{n} \cdot \mathbf{n}}}\ \mathbf{n} \cdot \boldsymbol{\sigma} \right] . \tag{59}$$

Once M is diagonable the rest of the game is computing two eigenvalues and the corresponding eigenvectors; biorthogonality as treated in Footnote 6 is helpful.

It is however simpler, and also more physical, to exploit the fact that both N_1/N and N_2/N are $\mathcal{O}(1)$ and take them equal, i.e., $N_1 = N_2 = N/2$. Then M is Hermitian (here real and symmetric) and $\mathbf{n} \in \mathbb{R}^3$ so that we can write $\mathbf{n} = n\hat{\mathbf{n}}$ with n being the length $\sqrt{\mathbf{n} \cdot \mathbf{n}}$ of the vector $\mathbf{n}$ and $\hat{\mathbf{n}}$ being a unit vector. In physical terms, $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}$ is the projection of the spin onto the direction $\hat{\mathbf{n}}$. Furthermore, $n_0 = (b + Q/2)N_1 + c$, $\mathbf{n} = (bN_1, 0, -QN_1/2)$, and the eigenvalues of M are $m_\pm := n_0 \pm n$. As Eq. (59) shows explicitly, the latter result also determines the asymptotics for complex $\mathbf{n}$.

Keeping (53) and (54) in mind, we can now exploit (59). The eigenstates of M are identical with those of $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}$, the projection of the 'spin' $\boldsymbol{\sigma}$ onto the direction $\hat{\mathbf{n}}$; the corresponding eigenprojections are $|\mathbf{e}_\pm\rangle\langle\mathbf{e}_\pm| = (1/2)(\mathbb{1} \pm \hat{\mathbf{n}} \cdot \boldsymbol{\sigma})$. A simple computation gives $m_\pm := n_0 \pm n = (b + Q/2)N_1 + c \pm N_1[b^2 + Q^2/4]^{1/2}$. Since $Q > 0$ by assumption, both eigenvalues $m_\pm$ are positive, if $b > 0$ and $N_1 \gg 1$ so that c is subdominant; we can use c for fine-tuning, however. If on the other hand $b < 0$, then $m_- < 0$ but $m_+ > 0$. In both cases the eigenstate $\mathbf{e}_+$ belonging to $m_+ > 0$ is dominant. As we are given $|\mathbf{e}_\pm\rangle\langle\mathbf{e}_\pm|$, the source term $\mathbf{a}$, and the initial condition $\mathbf{J}(0)$, we know what the asymptotics looks like; cf. (54).

Figure 9 shows the result of a realistic simulation, a nice academic exercise: $0 < J_i(0) = J^u$ for all i. Synapses that are "more equal" than others, win. We can now easily understand why. In the present case $\mathbf{a} = a\mathbf{1}$, $\mathbf{J}(0) = J^u\mathbf{1}$, and $J^u + a/m_+ > 0$ so that the vector $\mathbf{1}_+ := |\mathbf{e}_+\rangle\langle\mathbf{e}_+|\mathbf{1}\rangle = (1 + (N_1/n)(b - Q/2), 1 + (N_1/n)(b + Q/2))^T$ tells us what will happen; (N_1/n) does not depend on N_1. A nontrivial structure occurs only if $b < 0$ since the first component of $\mathbf{1}_+$ is then negative whereas the second is positive. This is the case in Fig. 9, where $b < 0$. For $b > 0$ both components of the vector $\mathbf{1}_+$ are positive and a trivial saturation occurs.

A careful look at the right inset of Fig. 9 reveals, however, that the agreement between theory and experiment is not perfect. Why is that? To see why, we return to (45) and adapt it to the present situation,

$$\frac{\mathrm{d}}{\mathrm{d}t}J_i = a + b\left(\sum_j J_j\right) + cJ_i + Q\delta_{i,\mathcal{N}_2}\sum_{j\in\mathcal{N}_2} J_j \ , \tag{60}$$

with $\delta_{i,\mathcal{N}_2} = 1$ if $i \in \mathcal{N}_2$ and $\delta_{i,\mathcal{N}_2} = 0$ otherwise. We started with $N_1 = N_2 = N/2$, which holds as long as none of the J_i has attained the upper or the lower bound, viz., J^u or 0. As soon as this happens, say for $i = i_0 \in \mathcal{N}_1$, we take $i = i_0$ out of (60) but *not* out of $\mathcal{N}_1$; the procedure must be repeated each time a J_i touches one of the boundaries. Furthermore, N_1 and N_2 now become dynamic variables counting the number of *active* J_i in $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, with in general $N_1 \neq N_2$. Equations (57) and (59) still hold as long as none of the active J_i hits one of the boundaries, but $\mathbf{n}$ and n_0 change continuously, as do the N_i in both (56) and (57). Moreover, $\mathbf{1}_-$ (the analog of $\mathbf{1}_+$) as well as $a \neq 0$ may, and in general will, influence the dynamics. Hence the exact dynamics given by (60) approximates but is not identical with the one given above – as advertised. In fact, the exclusion process continues until $N_1 = N_2 = 0$, beyond which nothing changes any more.

6 Short-Term Synaptic Plasticity

Despite being of 'short' duration, short-term synaptic plasticity may have profound effects on network behavior and is, in fact, closely correlated with it. We therefore start by outlining the problem and analyzing a simple model of short-term plasticity that is an adaptation of the model of Tsodyks and Markram [65, 66] to the Spike Response Model (see §2). We then specify how the synaptic efficacies change as a function of presynaptic input – and time. The resulting setup allows a full-blown study of network behavior.

6.1 The Problem

Short-term synaptic plasticity is to be contrasted with its long-term counterpart in that it refers to a change in the synaptic efficacy on a time scale of milliseconds up to seconds. It is therefore natural to inquire whether and to what extent this has functional consequences, and to elucidate the underlying mechanisms [65, 66, 67, 68, 69, 70, 71]. The experimental observation underpinning short-term synaptic plasticity is the fact [72, 89, 90, 91] that the transmission of an action potential across a synapse can have a significant influence on the amplitude of the postsynaptic potential (PSP) evoked by *subsequently* transmitted spikes. In some synapses, the height of the postsynaptic potential is increased by spikes

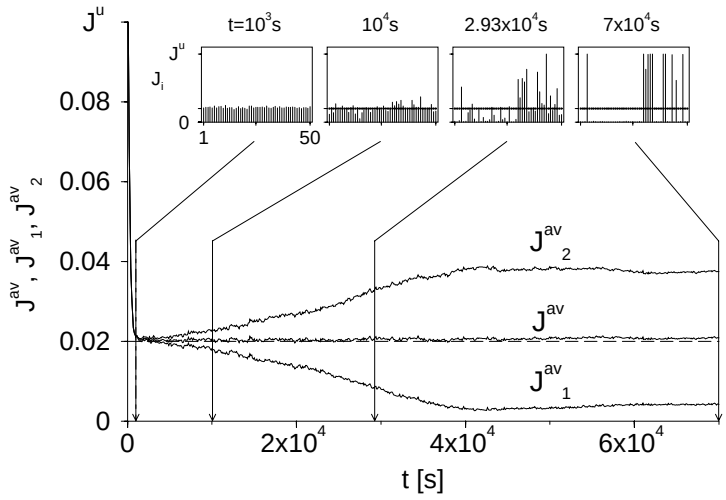

Figure 9: Temporal evolution of the average synaptic efficacies $J_1^{\rm av}$ and $J_2^{\rm av}$ as defined in (56), and $J^{\rm av} = (J_1^{\rm av} + J_2^{\rm av})/2$ as a function of the learning time t in units of 10^4 seconds. This is a fictitious time to keep the computational time finite. It is in general too fast for biology but can be adapted to it by a simple rescaling without changing the picture. The quantity $J^{\rm av}$ is the average weight of all synapses, $J_1^{\rm av}$ and $J_2^{\rm av}$ are average weights of the groups $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively. Synapses i in group $\mathcal{N}_1$, where $1 \leq i \leq 25$, receive incoherent input whereas synapses i in group $\mathcal{N}_2$, where $26 \leq i \leq 50$, are driven by a coherently modulated input intensity. Parameters are $a = 10^{-4}$, $b = -10^{-4}$, $c = 7.04 \times 10^{-5}$, and $Q = 6.84 \times 10^{-7}$, all of dimension s^{-1}; furthermore, $N_1 = N_2 = 25$. The eigenvalues of the Hermitian matrix M in (59) are $m_+ \approx c$ and $m_- = -5 \times 10^{-3}$ so that we get two different time scales. Simulations started at time $t = 0$ with a homogeneous weight distribution $J_i \equiv 0.1 = J^u$ for all i. There is a 'fast' decrease determined by $m_- < 0$ and, thus, finished within a time of order $\mathcal{O}(100\,\mathrm{s})$. All this is near the vertical axis and hardly visible. Structure formation is dominated by $m_+ > 0$ and is happening on a time scale that is two orders of magnitude slower than that of the fast relaxation. Now synaptic efficacies saturate at the upper bound ($J_2^{\rm av}$) or at the lower bound ($J_1^{\rm av}$). The insets show the weight distributions at times $t = 10^3$, 10^4, 2.93×10^4, and 7×10^4 s (arrows). Taken from [47].

that have arrived previously (short-term facilitation, STF; also called paired-pulse facilitation). In other synapses, the postsynaptic potential is decreased by previously arrived action potentials (short-term depression, STD; also called paired-pulse depression).

Short-term synaptic plasticity, or simply short-term plasticity, is different from its well-known counterpart "long-term plasticity" in at least two crucial points. First, *nomen est omen*, the time scale on which short-term plasticity operates is much shorter than that of long-term plasticity and may be well comparable to the time scale of the network dynamics. Second, short-term plasticity of a given synapse is driven by correlations in the *in*coming spike train (presynaptic correlations), whereas classical long-term plasticity is driven by correlations of both pre- and postsynaptic activity; a prominent example of the latter is Hebbian

learning as studied in the previous sections.

6.2 Modeling Short-Term Synaptic Plasticity

Modeling short-term plasticity is based on the idea that some kind of 'resources' is required to transmit an action potential across the synaptic cleft [73, 74, 66, 67, 68]. The term 'resource' can be interpreted as the available amount of neurotransmitter, some kind of ionic concentration gradient, or postsynaptic receptor availability; cf. Fig. 1. We assume that every transmission of an action potential affects the amount of available synaptic resources and, on the other hand, that the amount of available resources determines the efficiency of the transmission and therefore the maximum of the postsynaptic potential. There is meanwhile considerable evidence [4, 89, 91] that short-term plasticity is due to presynaptic effects and, hence, to presynaptic correlations only. We can think of presynaptic 'resources' as, e.g., the number of vesicles that determines the release probability [89]. The relevant notion is then the probability p as it occurs in the mean synaptic response $\mathcal{R} = npQ$; see §1.

We are going to discuss short-term plasticity in the context of the Spike Response Model (see §2). In so doing we closely follow Ref. [77]. It will turn out that the spike-response formalism is very convenient in deriving closed expressions for synaptic efficacies as a function of spike arrivals and time. The time-dependent synaptic efficacy $J_{ij}(t)$ is a function that depends both on time and on the moments of arrival of the spikes from neuron j. This function will be computed in the next subsections.

6.3 Modeling Short-Term Depression

The model of Tsodyks and Markram [66] assumes three possible states for the "resources" of a synaptic connection: effective, inactive, and recovered. Whenever an action potential arrives at a synapse a fixed portion R of the *recovered* resources becomes first effective, then inactive, and finally recovers. Transitions between these states are described via first-order kinetics using time constants $\tau_{\rm inact}$ and $\tau_{\rm rec}$. The actual postsynaptic current is proportional to the amount of *effective* resources.

In the context of the Spike Response Model the above three-state model can be simplified further since the time course of the postsynaptic current, as it is described by the transition from the effective to the inactive state, is already taken care of by the form of the postsynaptic potential (PSP) given by the response function $J\varepsilon$. Focusing on a specific synapse $\{ij\}$, we drop its label. The only relevant quantity is the maximum (minimum) J determined by the charge delivered by a single action potential. As we have seen in §1, a synaptic efficacy J can be interpreted as its mean $\langle J \rangle$. We henceforth drop the alternative 'minimum' that

takes care of an inhibitory postsynaptic potential and assume an excitatory one, the modifications for inhibition being evident.

Transitions from the effective and the inactive to the recovered state are described by linear differential equations. The maximum of a PSP only depends on the amount of resources that are actually activated by the incoming action potential. We therefore summarize the 2-step recovery of effective resources (inactive & recovered) by a single step and end up with a 2-state model of active (Z) and inactive ($\bar{Z}$) resources; see Fig. 10. Each incoming action potential instantaneously switches a proportion $0 \leq P \leq 1$ of active resources to the inactive state from where they recover to the active state with time constant τ. In mathematical terms we arrive at

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = -P\,Z\,S(t) + \tau^{-1}\,\bar{Z}\,, \qquad \bar{Z} = 1 - Z\,, \tag{61}$$

with $S(t) = \sum_f \delta(t - t_f)$ as the incoming spike train. This differential equation is well-defined, if we declare $Z(t)$ to be continuous from the left, i.e., $Z(t_f) := Z(t_f - 0)$. The solution $Z(t)$ is in the interval $[0, 1]$.

The amount of charge that is released in a single transmission and therewith the maximum of a PSP depends on the amount of resources that were switched to the inactive state, or, equivalently, on the amount of active resources immediately before the transmission. The strength of the synapse at time t is then a function of $Z(t)$ and we simply put $J(t) = J^0\,Z(t)$ where J^0 is the maximal mean synaptic efficacy.

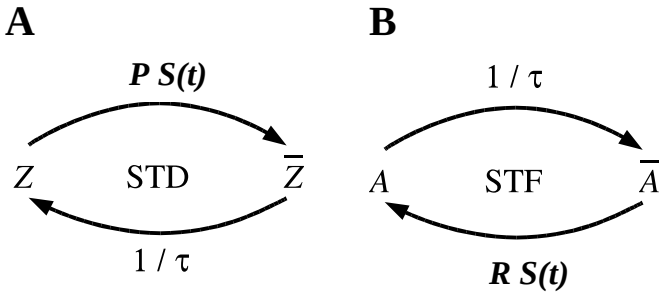

Figure 10: Schematic representation of the present model of short-term depression (A) and short-term facilitation (B). With short-term depression, every incoming action potential instantaneously switches a proportion $0 \leq P \leq 1$ of active resources Z to the inactive state $\bar{Z}$. This is described by a first-order reaction kinetics with the time-dependent rate $P\,S(t)$; here S is the incoming spike train, a sequence of delta functions. Resources relax from the inactive state to the active state with time constant τ. The model for short-term facilitation emerges from the model for short-term depression simply by inverting the directions of the arrows. $\bar{A}$ far right represents the ineffective resources, which are decimated by incoming spikes at a rate $R\,S(t)$. The active resources A relax back to the inactive state at a rate τ^{-1}. Taken from [77].

Let us now suppose that the first spike arrives at a synapse at time t_0. Immediately before the spike arrives, all resources are in their active state and $Z(t_0) = 1$.

The action potential switches a fraction P of the resources to the inactive state so that $Z(t_0 + 0) = 1 - P$. After the arrival of the action potential the inactive resources recover exponentially fast in t, and we have

$$Z(t > t_0) = 1 - P \, \exp[-(t - t_0)/\tau] \; . \tag{62}$$

At the arrival time t_1 of the subsequent spike there are only $Z(t_1)$ resources in the active state and the PSP is depressed accordingly.

To see how to proceed, we integrate (61) between $t_f - \Delta t$ and $t_f + \Delta t$ so as to obtain $Z(t_f + \Delta t) - Z(t_f - \Delta t) = -PZ(t_f) + \mathcal{O}(\Delta t)$, and take the limit $\Delta t \to 0$. Since $Z(t)$ is continuous from the left we find $Z(t_f + 0) - Z(t_f) = -PZ(t_f)$ and hence

$$Z(t_f + 0) = (1 - P)Z(t_f) \; . \tag{63}$$

Between two spikes Eq. (61) reads $\mathrm{d}Z/\mathrm{d}t = -\mathrm{d}(1 - Z)/\mathrm{d}t = \tau^{-1}(1 - Z)$, whence $(1 - Z)(t_2) = \exp[-(t_2 - t_1)/\tau](1 - Z)(t_1)$. If, now, t_1 approaches a firing time (which will also be called t_1) from above, then we get

$$\begin{aligned} Z(t_2) &= 1 - [1 - Z(t_1 + 0)] \exp[-(t_2 - t_1)/\tau] \\ &= 1 - [1 - (1 - P)Z(t_1)] \exp[-(t_2 - t_1)/\tau] \; . \end{aligned} \tag{64}$$

In the transition from the first to the second line we have exploited (63).

¿From the first few examples we can easily read off a recurrence relation that relates the amount of active resources immediately *before* the nth-spike to that of the previous spike,

$$\begin{aligned} Z(t_0) &= 1 \\ Z(t_1) &= 1 - P \, \exp\left[-(t_1 - t_0)/\tau\right] \\ Z(t_2) &= 1 - \left[1 - (1 - P) \, Z(t_1)\right] \, \exp\left[-(t_2 - t_1)/\tau\right] \\ &\vdots \\ Z(t_n) &= 1 - \left[1 - (1 - P) \, Z(t_{n-1})\right] \, \exp\left[-(t_n - t_{n-1})/\tau\right] \; . \end{aligned} \tag{65}$$

In passing we note that, instead of $Z(t_0) = 1$ we could have taken any desired initial condition $0 < Z_0 \leq 1$. The ensuing argument does not change.

The recurrence relation (65) is of the form

$$Z(t_n) = a_n + b_n \, Z(t_{n-1}) \tag{66}$$

with

$$a_n = 1 - \exp[-(t_n - t_{n-1})/\tau], \qquad b_n = (1 - P) \, \exp[-(t_n - t_{n-1})/\tau] \; . \tag{67}$$

Recursive substitution and a short calculation yield the following explicit expression for the amount of active resources,

$$\begin{aligned} Z(t_n) &= a_n + b_n\, Z(t_{n-1}) \\ &= a_n + b_n\,[a_{n-1} + b_{n-1}\, Z(t_{n-2})] \\ &= a_n + b_n\, a_{n-1} + b_n\, b_{n-1}\, a_{n-2} + \dots \\ &= \sum_{k=0}^{\infty} a_{n-k} \prod_{j=0}^{k-1} b_{n-j} \\ &= \sum_{k=0}^{\infty} a_{n-k} (1-P)^k\, \exp[-(t_n - t_{n-k})/\tau] \\ &= 1 - \frac{P}{1-P} \sum_{k=1}^{\infty} (1-P)^k\, \exp[-(t_n - t_{n-k})/\tau]\ . \end{aligned} \tag{68}$$

The synaptic efficacy at time t as a function of the spike arrival times $\cdots < t_{n-2} < t_{n-1} < t$ is given by

$$J(t; t_{n-1}, t_{n-2}, \dots) = J^0 \left\{ 1 - \frac{P}{1-P} \sum_{k=1}^{\infty} (1-P)^k\, \exp[-(t - t_{n-k})/\tau] \right\}\ . \tag{69}$$

This is a key result to what follows.

6.4 Periodic input

The synaptic efficacy J is a nonlinear function of the spike arrival times t_f. We can give a simplified expression for J in the case of a sudden onset of periodic spike input. Let $t_n = n\,T$ for $n \geq 0$ and $t_n = -\infty$ for $n < 0$. We obtain from (68) for $n > 0$,

$$\begin{aligned} Z(t_n) &= 1 - \frac{P}{1-P} \sum_{k=1}^{n} (1-P)^k\, \exp[-k\,T/\tau] \\ &= 1 - \frac{P}{e^{T/\tau} - (1-P)} \left\{ 1 - \left[(1-P)\, e^{-T/\tau} \right]^n \right\} \end{aligned} \tag{70}$$

The behavior of $Z(t_n)$ for large n can be read off easily from the above equation. Since $0 < e^{-T/\tau}\,(1-P) < 1$, the braced expression converges to unity exponentially fast and the rest, which is independent of n, gives the asymptotic value of $Z(t_n)$ as $n \to \infty$.

6.5 Modeling Short-Term Facilitation

In a similar fashion to §6.3, we can devise a model that accounts for short-term *facilitation* instead of depression. To this end, we assume that in the absence of

presynaptic spikes the fraction $A(t)$ of active synaptic resources decays with time constant τ. Each incoming spike recruits a fraction, or ratio, $0 \le R \le 1$ from the reservoir $\bar{A}$ of ineffective resources; see Fig. 10. Then the dynamics of $A(t)$ is given by

$$\frac{\mathrm{d}A}{\mathrm{d}t} = R\,\bar{A}\,S(t) - \tau^{-1}\,A\,, \qquad \bar{A} = 1 - A\,, \tag{71}$$

with $S(t) = \sum_f \delta(t - t_f)$ as the incoming spike train and $A(t)$ being continuous from the left. Magleby and Zengle [74] used a similar model to describe synaptic potentiation at a frog neuromuscular junction.

For a discrete set of spike arrival times $t_f = t_0, t_1, \ldots$ the amount of effective synaptic resources immediately *before* the nth spike as a function of that before the previous spike is

$$A(t_n) = a_n + b_n\,A(t_{n-1}) \tag{72}$$

where

$$a_n = R\,\exp[-(t_n - t_{n-1})/\tau], \qquad b_n = (1-R)\,\exp[-(t_n - t_{n-1})/\tau] \tag{73}$$

are nearly identical with their companions in (67).

In a similar way to (68), we obtain an explicit expression for the amount of effective resources. We adopt a simple linear dependence of the synaptic efficacy J upon the amount of effective resources A of the form $J = J^0\,[A_0 + (1 - A_0)\,A]$, $0 \le A_0 \le 1$, with J_0 being the maximal mean synaptic efficacy and $J^0\,A_0$ its minimal strength. Altogether we obtain

$$J(t; t_{n-1}, t_{n-2}, \ldots) = J^0\,\left\{A_0 + (1 - A_0)\,\frac{R}{1-R}\,\sum_{k=1}^{\infty}(1-R)^k\,\exp\left[-(t - t_{n-k})/\tau\right]\right\} \tag{74}$$

In the case of periodic input with $t_n = n\,T$ for $n \ge 0$ and $t_n = -\infty$ for $n < 0$ the above equation reduces to the facilitation analog of (70),

$$J(t_n)/J^0 = A_0 + (1 - A_0)\,\frac{R}{e^{T/\tau} - (1-R)}\,\left\{1 - \left[(1-R)\,e^{-T/\tau}\right]^n\right\}\,. \tag{75}$$

This implies that as $n \to \infty$ the synaptic efficacy converges exponentially fast from below to the asymptotic value

$$J^{\mathrm{STF}}_{\infty} = J^0\,\left[A_0 + \frac{(1 - A_0)\,R}{e^{T/\tau} - (1-R)}\right]\,. \tag{76}$$

Short-term plasticity introduces a second time scale into the dynamics of a neuronal network. An analysis of its implications for a homogeneous network of excitatory neurons, the simplest possible case, and simulations showing intricate network behavior despite the apparent structural simplicity, can be found elsewhere [77]. In a similar vein, Buonomano [92] presents numerical evidence suggesting that short-term plasticity plays a role in neuronal decoding of temporal information.

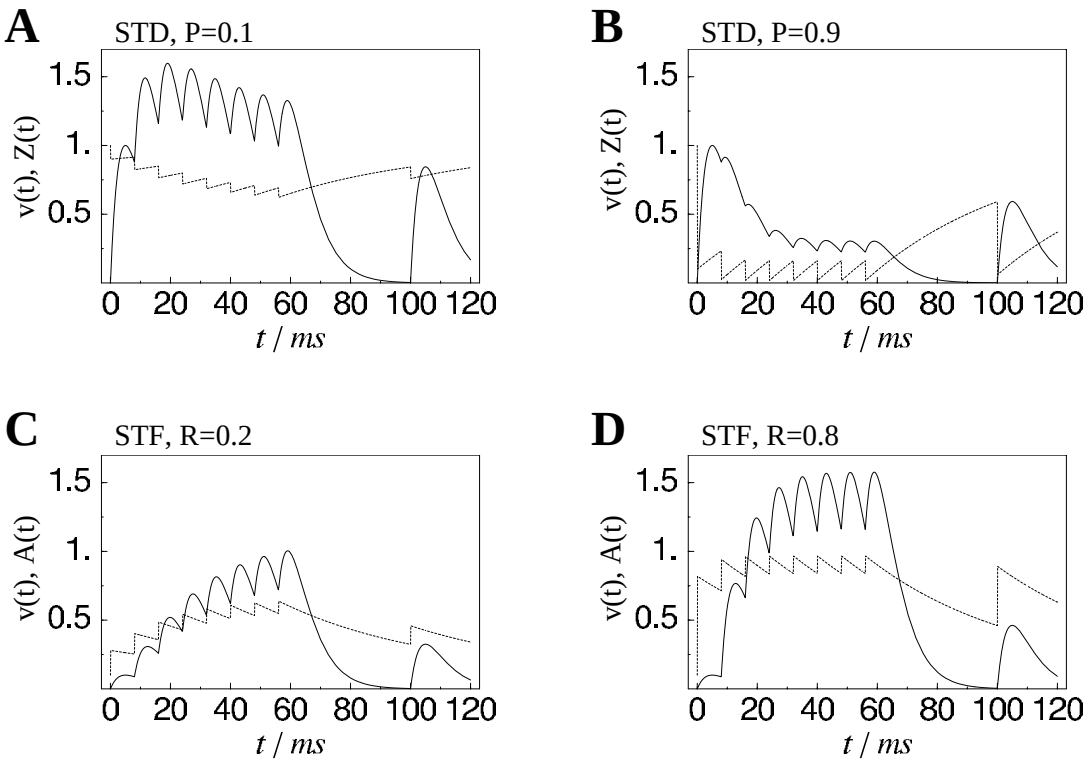

Figure 11*:* Membrane potential v (solid line) and synaptic resources Z or A (dashed line) as a function of time in case of short-term depression (**A** and **B**) and facilitation (**C** and **D**). Spikes arrive at $t = 0, 8, 16, \ldots, 56$ ms, and finally at $t = 100$ ms. In all figures, the time constant of the synaptic recovery is $\tau = 50$ ms and the rise time of the EPSP equals 5 ms; cf. (61), (71), and (6). Both Z and A are numbers between 0 and 1; Z starts at 1.0, A at $A_0 = 0.1$. In **A**, only a small portion $P = 0.1$ of all available resources is used during a single transmission so that the synapse is only slightly affected by transmitter depletion. In **B**, the parameter P is increased to $P = 0.9$. This results in a pronounced short-term depression of the synaptic strength. Short-term facilitation is illustrated in the lower two diagrams for $R = 0.2$ (**C**) and $R = 0.8$ (**D**). Taken from [77].

7 Conclusion and Open Problems

We began this chapter with a provocative question, "What is synaptic plasticity?". We end it with a more practical one, "What *induces* synaptic plasticity?", knowing that time scale and order of neuronal events may, and often do, play a key role. Hebbian learning is like a game involving three contestants: a presynaptic neuron, a postsynaptic neuron and a synapse between them. The two neurons interact via synaptic events induced by all-or-none depolarizations of their membranes, beginning in the presynaptic cell and propagating along an axon to synaptic sites on the postsynaptic neuron. This takes a finite amount of time, a delay in the millisecond range. Eventually, the postsynaptic neuron reaches firing threshold, producing an action potential that backpropagates [21]

from the site of spike initiation into the synapse.

Thus, pre- and postsynaptic cells interact at the synapse through a *learning window* that relates their spike timing to an increase or decrease in synaptic efficacy. If precise timing is not important, the learning window is broad, and rate coding is a simple consequence. Axonal delays naturally appear as elements of the learning window. The notion of learning window was first conceived in theory [45] but has since been extensively confirmed by experiment; see for instance Fig. 7. This experimental evidence emphasizes milli- and sub-millisecond timing as being of fundamental importance in the three-partner game.

Conceptually, we can often think of learning as something that occurs in infinitesimal steps, incremental increases and decreases in synaptic efficacy that, as we have shown, lend themselves to a rigorous mathematical treatment culminating in the learning equation (32). Though single synapses may behave quite erratically, we have seen that an ensemble of them provides a neuron with a practically deterministic input due to the strong law of large numbers, a mainstay of stochastic analysis. The *ensemble* of synapses decides whether the postsynaptic neuron will generate an action potential and, subsequently, a backpropagating spike. Since spike generation is highly nonlinear, we have introduced the notion of 'Poisson neuron' to linearize the dynamics of the ensemble of synaptic efficacies. By applying the central limit theorem to the presynaptic input, one can replace the Poisson neuron by a Poissonian counterpart of arbitrary nonlinearity [64] and none the less solve the dynamics nearly exactly; cf. the Berry-Esseen estimate in Appendix B.

The key to unraveling the synaptic dynamics of infinitesimal learning is recognizing that, under normal circumstances, Eq. (32) is self-averaging. Nevertheless, it remains a challenge to use a deterministic, hence nonlinear, neuron model to solve the learning equation (32). It may well be that neuronal dynamics *in conjunction with* synaptic dynamics is an insoluble problem, but the prospect of discovering a full theoretical understanding of both long- and short-term synaptic plasticity is a prize worthy of the attempt.

Appendix A: Laws of Large Numbers

The textbook by Durrett [98] is a general, though advanced, background for various formulations of the laws of large numbers listed below. To begin with, let us suppose that the f_i are independent, identically distributed random variables with mean zero. If the mean $\langle f \rangle$ is nonzero, we subtract it and consider $f_i := f_i - \langle f \rangle$ instead. There is no harm in taking the f_i to be real variables. Furthermore, we require the second moment $\langle f^2 \rangle$ to be finite. By Cauchy-Schwarz, $\langle |f| \rangle \leq$

$\langle f^2 \rangle^{1/2} < \infty$, and the variance $\sigma^2 := \langle (f - \langle f \rangle)^2 \rangle$ is finite too. Let

$$S_n = \sum_{i=1}^{n} f_i$$

be the sum of the random variables f_i. Then the following three theorems hold.

- Strong law of large numbers: $\lim_{n\to\infty} n^{-1} S_n = 0$ with probability 1. Since the f_i are sampled from a probability distribution, this means that, as $n \to \infty$, the configurations where the above equality does not hold have probability zero. In plain English, they do not occur. One also says that the above equality holds 'almost surely' (a.s.). All that is needed is $\langle |f| \rangle < \infty$.

- Central limit theorem: As $n \to \infty$, $n^{-1/2} S_n$ has a Gaussian distribution with mean zero and variance σ^2.

- Law of the iterated logarithm:

$$\limsup_{n\to\infty} \frac{|S_n|}{\sigma\sqrt{2n \ln\ln n}} = 1 \qquad \text{a.s.}$$

Etemadi [99] has given an "elementary" proof of the strong law of large numbers for *pairwise* independent, identically distributed random variables under the minimal condition $\langle |f| \rangle < \infty$. Slick proofs (occasionally with some extra conditions, say, finite fourth moment) have been given by Lamperti [11]. Breiman [12] treats the first two theorems in their full generality. The law of the iterated logarithm is an extension of the central limit theorem. Its proof is tricky.

All three theorems also hold for independent, not necessarily identically distributed random variables [11, 12, 101]. The first two even allow a weak dependence. For example, let $R_{ij} := \langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle$, and suppose the f_i do not have too wide a distribution, e.g., $\sup_i |R_{ii}| < \infty$. Then the strong law of large numbers holds [96, p. 265] [102], provided $R_{ij} \to 0$ as $|i - j| \to \infty$; that is to say, the correlations between f_i and f_j should not have too long a range. For the central limit theorem to hold, trickier conditions are required, e.g., stationarity of the sequence $f_1, f_2, \ldots$ and some kind of mixing [98, Ch. 7.7c] so that, say, $\sum_j |R_{ij}| < \infty$. Then the variance of the Gaussian limit distribution is given by

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{ij} \langle f_i f_k \rangle = \langle f_1^2 \rangle + 2 \sum_{k=2}^{\infty} \langle f_1 f_k \rangle \ .$$

Dropping stationarity, the reader may consult Scott [103] for an advanced account.

A generalization of the law of the iterated logarithm to independent but not necessarily identically distributed random variables is this [100, p. 241]. Let σ_k^2

be the variance of f_k, $B_n^2 := \sum_{k \leq n} \sigma_k^2$, and $f_n/B_n = o(1/\sqrt{\ln \ln B_n^2})$. Then we have

$$\limsup_{n \to \infty} \frac{|S_n|}{B_n \sqrt{2 \ln \ln B_n^2}} = 1 \qquad \text{a.s.}$$

Appendix B: Inhomogeneous Poisson Processes

In this appendix, which is identical with Appendix A of Kempter et al. [46] and reproduced here for convenience of the reader, we define and analyze the *inhomogeneous* Poisson process. This notion has been touched upon by Tuckwell [93, pp. 218–220] and others, e.g., Ash and Gardner [94, pp. 28–29], but neither of them explains the formalism itself or the way of computing expectation values. Since both are used extensively, we treat them here, despite the fact that the issue is considered by Snyder and Miller [95, §2.1-2.3]. Our starting assumptions in handling this problem are the same as those of Gnedenko [96, §51] for the homogeneous (uniform) Poisson process but the mathematics is different. Neither does our method resemble the Snyder and Miller approach, which starts from the other end, viz., Eq. (87). In the context of theoretical neurobiology an analysis such as the present one, focusing on the *local* behavior of a process, seems far more natural. We proceed by evaluating the mean and the variance and finish by estimating a third moment that is needed for the Berry-Esseen inequality, which tells us how good a Gaussian approximation to a *finite* sum of independent random variables is.

Definitions

Let us suppose that a certain event, in our case a spike, occurs at random instances of time. Let $N(t)$ be the number of occurrences of this event up to time t. We suppose that $N(0) = 0$, that the probability of getting a *single* event during the interval $[t, t + \Delta t)$ with $\Delta t \to 0$ is

$$\Pr\{N(t + \Delta t) - N(t) = 1\} = \lambda(t)\Delta t, \qquad \lambda \geq 0 \ , \tag{77}$$

and that the probability of getting two or more events is $o(\Delta t)$. Finally, the process has independent increments, i.e., events in disjoint intervals are independent. The stochastic process obeying the above conditions is an inhomogeneous Poisson process.

Under conditions on λ to be specified below, there are only finitely many events in a finite interval. Hence the process lives on a space Ω of monotonically non-decreasing, piece-wise constant functions on the positive real axis, having finitely many unit jumps in any finite interval. The expectation value corresponding to this inhomogeneous Poisson process is simply an integral with respect to a probability measure μ on Ω, a function space whose existence is guaranteed

by the Kolmogorov extension theorem [97, §4.4.3]. A specific realization of the process, a function on the positive real axis, is a 'point' ω in Ω. The discrete events corresponding to ω are denoted by $t_f(\omega)$ with f labeling them.

As we have seen in Eq. (3), spikes generate postsynaptic potentials ε. We now compute the average, denoted by angular brackets, of the postsynaptic potentials generated by a specific neuron during the time interval $[t_0, t)$,

$$\left\langle \sum\nolimits_f \varepsilon(t - t_f(\omega)) \right\rangle \ . \tag{78}$$

Here it is understood that $t_f = t_f(\omega)$ depends on the realization ω and $t_0 \leq t_f(\omega) < t$. We divide the interval $[t_0, t)$ into L subintervals $[t_l, t_{l+1})$ of length Δt so that at the end $\Delta t \to 0$ and $L \to \infty$ while $L\Delta t = t - t_0$. We now evaluate the integral (78) exploiting the fact that ε is a continuous function.

Let $\#\{t_l \leq t_f(\omega) < t_{l+1}\}$ denote the number of events (spikes) occurring at times $t_f(\omega)$ in the interval $[t_l, t_{l+1})$ of length Δt. In the limit $\Delta t \to 0$ the expectation value (78) can be written

$$\int_\Omega \mathrm{d}\mu(\omega) \left[\sum\nolimits_l \varepsilon(t - t_l) \, \#\{t_l \leq t_f(\omega) < t_{l+1}\} \right] \tag{79}$$

so that we are left with the Riemann integral

$$\int_{t_0}^{t} \mathrm{d}s \, \lambda(s) \, \varepsilon(t - s) \ . \tag{80}$$

We spell out why. The function $\mathbb{1}_{\{\ldots\}}$ is to be the indicator function of the set $\{\ldots\}$ in Ω; that is, $\mathbb{1}_{\{\ldots\}}(\omega) = 1$, if $\omega \in \{\ldots\}$ and it vanishes, if ω does not belong to $\{\ldots\}$. So it 'indicates' where the set $\{\ldots\}$ lives. With the benefit of hindsight we single out mutually independent sets in Ω with indicators $\mathbb{1}_{\{t_l \leq t_f(\omega) < t_{l+1}\}}$ and write the expectation value (78) in the form

$$\int_\Omega \mathrm{d}\mu(\omega) \sum_l \mathbb{1}_{\{t_l \leq t_f(\omega) < t_{l+1}\}} \varepsilon(t - t_l) \, \#\{t_l \leq t_f(\omega) < t_{l+1}\} \ . \tag{81}$$

Each indicator function in the sum equals

$$\mathbb{1}_{\{t_l \leq t_i(\omega) < t_{l+1}\}} = \mathbb{1}_{\{N(t_{l+1}) - N(t_l) = 0\}} + \mathbb{1}_{\{N(t_{l+1}) - N(t_l) = 1\}} + \mathbb{1}_{\{N(t_{l+1}) - N(t_l) \geq 2\}} \ . \tag{82}$$

In view of (78) and (81) we multiply this by $\varepsilon(t - t_l) \, \#\{t_l \leq t_f(\omega) < t_{l+1}\}$, interchange integration and summation in (81), and integrate with respect to μ. The first term on the right contributes nothing, the second gives $\varepsilon(t - t_l)\lambda(t_l)\Delta t$ and thus produces a term in the Riemann sum leading to (80), and the last term can be neglected since it is of order $o(\Delta t)$. So the eating of the pudding is that only a *single* event in the interval $[t_l, t_{l+1})$ counts as $\Delta t \to 0$. Since $\varepsilon(t)$ is a function which decreases at least exponentially fast as $t \to \infty$ there is no harm in taking $t_0 = -\infty$.

Second Moment and Variance

It is time to compute the second moment

$$\left\langle [\sum_{t_f<t} \varepsilon(t-t_f)]^2 \right\rangle \ . \tag{83}$$

In a similar vein as before we obtain, in the limit $\Delta \to 0$,

$$\begin{aligned}
& \langle \sum_{t_f,t'_f<t} \varepsilon(t-t_f)\varepsilon(t-t'_f)\rangle \\
= & \int_\Omega \mathrm{d}\mu(\omega) \sum_{l,m} \mathbb{1}_{\{t_l\le t_f(\omega)<t_{l+1}\}} \mathbb{1}_{\{t_m\le t'_f(\omega)<t_{m+1}\}} \varepsilon(t-t_f(\omega))\varepsilon(t-t'_f(\omega)) \\
= & \sum_{l\neq m} [\lambda(t_l)\Delta t\, \lambda(t_m)\Delta t]\, \varepsilon(t-t_l)\varepsilon(t-t_m) + \\
& \qquad \int_\Omega \mathrm{d}\mu(\omega) \sum_l \mathbb{1}^2_{\{t_l\le t_f(\omega)<t_{l+1}\}} \varepsilon^2(t-t_l) \\
= & \int_{t_0}^t \int_{t_0}^t \mathrm{d}t_1 \mathrm{d}t_2\ \lambda(t_1)\lambda(t_2)\varepsilon(t-t_1)\varepsilon(t-t_2) + \int_{t_0}^t \mathrm{d}s\ \lambda(s)\varepsilon^2(t-s) \\
= & \left[\int_{t_0}^t \mathrm{d}s\ \lambda(s)\varepsilon(t-s)\right]^2 + \int_{t_0}^t \mathrm{d}s\ \lambda(s)\varepsilon^2(t-s)\ .
\end{aligned} \tag{84}$$

Hence the variance is the last term on the right in (84). It is a simple exercise to verify that, when $\lambda(t) \equiv \lambda$ and $\varepsilon(t) \equiv 1$ in (80) and (84), we regain the mean and variance of the usual Poisson distribution [96, §51].

We finish the argument by computing the probability of getting k events in the interval $[t_0, t)$. For the usual, homogeneous Poisson process it is

$$\Pr\{N(t) - N(t_0) = k\} = \exp[-\lambda(t-t_0)] \cdot \frac{[\lambda(t-t_0)]^k}{k!}\ . \tag{85}$$

We now break up the interval $[t_0, t)$ into many subintervals $[\tau_l, \tau_{l+1})$ of length Δt and condition with respect to the first, second, ... arrival. The arrivals come one after the other and the probability of a *specific* sequence of events in $[t_1, t_1 + \Delta t), [t_2, t_2 + \Delta t), \ldots, [t_k, t_k + \Delta t)$ is made up of elementary events such as

$$\begin{aligned}
& \Pr\{\text{first spike in } [t_1, t_1+\Delta t)\} = \\
& \Pr\{\text{no spike in } [t_0, t_1)\}\, \Pr\{\text{spike in } [t_1, t_1+\Delta t)\} = \\
& [1-\lambda(\tau_1)\Delta t][1-\lambda(\tau_2)\Delta t]\ldots[1-\lambda(t_1-\Delta t)\Delta t]\, \lambda(t_1)\Delta t = \\
& \exp\left[-\int_{t_0}^{t_1} \mathrm{d}\tau\, \lambda(\tau)\right] \lambda(t_1)\Delta t\ .
\end{aligned} \tag{86}$$

Here we have exploited the independent-increments property and taken the limit $\Delta t \to 0$ to obtain the last equality. Repeating the above argument for the following events, including the no-event tail in $[t_k + \Delta t, t)$, multiplying the probabilities, and summing over all possible realizations we find

$$\begin{aligned} & \Pr\{N(t) - N(t_0) = k\} \\ = \;& \exp\left[-\int_{t_0}^{t} \mathrm{d}\tau\, \lambda(\tau)\right] \int_{t_0}^{t} \mathrm{d}t_k\, \lambda(t_k) \ldots \int_{t_0}^{t_3} \mathrm{d}t_2\, \lambda(t_2) \int_{t_0}^{t_2} \mathrm{d}t_1\, \lambda(t_1) \\ = \;& \exp\left[-\int_{t_0}^{t} \mathrm{d}\tau\, \lambda(\tau)\right] \cdot \frac{1}{k!} \left[\int_{t_0}^{t} \mathrm{d}s\, \lambda(s)\right]^k . \end{aligned} \tag{87}$$

In other words, $N(t) - N(t_0)$ has a Poisson distribution with parameter $\int_{t_0}^{t} \mathrm{d}s\, \lambda(s)$. If $\lambda(s) \equiv \lambda$, one regains (85). We now see two things. First, the appropriate condition on λ is that it be locally integrable. Then $\Pr\{N(t) - N(t_0) < \infty\} = 1$ as the sum of (87) over all finite k adds up to one. Furthermore, $N(t) - N(t')$ with $t_0 < t' < t$ has a Poisson distribution with parameter $\int_{t'}^{t} \mathrm{d}s\, \lambda(s)$. Second, by rescaling time through $t := \int^t \mathrm{d}s\, \lambda(s)$ one obtains [93, 94] a homogeneous Poisson process with parameter $\lambda = 1$. This also follows more directly from (77). It is of no practical help, though. For instance, in the case of the barn owl, $\lambda(t)$ is taken to be a periodic function of t, with the period determined by external sound input. The cochlea produces a whole range of frequency inputs whereas time can be rescaled only once.

Berry-Esseen Estimate

Equation (3) tells us that the neuronal input is a sum of independent, though not necessarily identically distributed, random variables corresponding to 'neighboring' neurons j. Neither independence nor a common distribution is necessary but both are quite convenient. The point is that, according to the central limit theorem (cf. Appendix A), a sum of N independent random variables[7] has a Gaussian distribution as $N \to \infty$. In our case N is definitely finite, so the question is: How good is the Gaussian approximation? The answer is provided by a classical, and remarkable, result of Berry and Esseen [11, §15].

We first formulate the Berry-Esseen result. Let $X_1, X_2, \ldots$ be independent random variables with a common distribution having variance σ^2 and finite third moment. Furthermore, let $S_N = \sum_{j=1}^{N} (X_j - \langle X_j \rangle)$ be the total input, the X_j stemming from neighboring neurons j as given by the right-hand side of (5) with N as the number of synapses, and let Y_σ be a Gaussian with mean 0 and variance σ^2. Then there is a constant $(2\pi)^{-1/2} \leq C < 0.8$ such that, whatever

[7] This N directly corresponds with the number of synapses that provide the neuronal input. There is no need to confuse it with the stochastic variable $N(t)$ of the previous subsection.

the distribution of the X_j and whatever x,

$$\left| \Pr\left\{ \frac{S_N}{\sqrt{N}} \le x \right\} - \Pr\{Y_\sigma \le x\} \right| \le \frac{C\langle |X_1 - \langle X_1 \rangle|^3 \rangle}{\sigma^3 \sqrt{N}} \ . \tag{88}$$

In the present case, σ^2 directly follows from (84). Computing $\langle |X_1 - \langle X_1 \rangle|^3 \rangle$ is a bit nasty but it is simpler, and also providing more insight, to directly estimate the third moment by Cauchy-Schwartz so as to get rid of the absolute value,

$$\langle |X_1 - \langle X_1 \rangle|^3 \rangle \le \langle (X_1 - \langle X_1 \rangle)^2 \rangle^{1/2} \langle (X_1 - \langle X_1 \rangle)^4 \rangle^{1/2} \ . \tag{89}$$

The first term on the right equals σ, the second is given by

$$\langle (X_1 - \langle X_1 \rangle)^4 \rangle = \int_{t_0}^{t} \mathrm{d}s \, \lambda(s) \, \varepsilon^4(t-s) + 3\sigma^4 \ . \tag{90}$$

where $\sigma^2 = \int_{t_0}^{t} \mathrm{d}\, s \, \lambda(s) \varepsilon^2(t-s)$. Collecting terms we can estimate the right-hand side of (88), the precision of the Gaussian approximation being determined by $1/\sqrt{N}$ as N becomes large.

Acknowledgments

It is a great pleasure to thank Wulfram Gerstner, Richard Kempter, and Werner Kistler for an enjoyable collaboration over the years and for all I have learned from them. I also thank Philip Brownell, Moritz Franosch, Richard Kempter, Christian Leibold, and Heather Read for a critical reading of parts of the manuscript and helpful suggestions; it is the author who is to be blamed for the remaining errors. Finally, I gratefully acknowledge constructive discussions with Frank den Hollander and Reinhard Lang concerning simple but optimal formulations of laws of large numbers.

References

[1] Sherrington, C.S. (1897) in: *Textbook of Physiology* (Foster, M., ed.) p. 60

[2] Shepherd, G.M. and Erulkar, S.D., Centenary of the synapse: from Sherrington to the molecular biology of the synapse and beyond. Trends Neurosci. **20**, 385–392

[3] Shepherd, G.M., editor (1998) *The Synaptic Organization of the Brain*, 4th ed., Oxford University Press, New York

[4] Koch, C. (1999) *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, New York

[5] Abbott, L. and Sejnowski, T.J., Eds. (1999) *Neural Codes and Distributed Representations*. MIT Press, Cambridge, MA

[6] Sejnowski, T.J. (1977) Storing covariance with nonlinearly interacting neurons. J. Math. Biol. **4**, 303–321; Statistical constraints on synaptic plasticity. J. Theor. Biol. **69**, 385–389

[7] Jessell, T.M. and Kandel, E.R. (1993) Synaptic transmission: a bidirectional and self-modifiable form of cell-cell communication. Neuron **10** (Suppl.), 1–30

[8] Faber, D.S., Korn, H., Redman, S.J., Thompson, S.M. and Altman, J.S., Eds. (1998) *Central Synapses: Quantal Mechanisms and Plasticity*. HFSP, Strasbourg

[9] Braitenberg, V. and Schütz, A. (1991) *Anatomy of the Cortex*, Springer, Berlin

[10] Ramón y Cajal, S. (1911) *Histologie du Système Nerveux de l'Homme et des Vertébrés*. Maloine, Paris. This is L. Azoulay's definitive French translation of the Spanish original. There is meanwhile also an English translation by N. & L.W. Swanson (Oxford University Press, 1995), based on the French one.

[11] Lamperti, J. (1966) *Probability*, Benjamin, New York; a 2nd edition (1996) appeared with Wiley, New York

[12] Breiman, L. (1968) *Probability*, Addison-Wesley, Reading, MA; a 2nd edition appeared with SIAM, Philadelphia, PA (1996). Lamperti and Breiman are both classics, the latter being much more advanced.

[13] Levitan, I.B. and Kaczmarek, L.K. (1997) *The Neuron: Cell and Molecular Biology*, 2nd edition. Oxford University Press

[14] Kandel, E.R., Schwartz, J.H. and Jessell, T.M. (2000) *Principles of Neural Science*, 4th edition. McGraw-Hill, New York; see, in particular, Ch. 14 on neurotransmitter release.

[15] Faber, D. S. and Korn, H., Eds. (1978) *Neurobiology of the Mauthner Cell*, Raven Press, New York

[16] Korn, H. and Faber, D.S. (1996) Escape behavior – brainstem and spinal cord circuitry and function. Curr. Opinion Neurobiol. **6**, 826–832

[17] Kruk, P.J., Korn, H. and Faber, D.S. (1997) The effects of geometrical parameters on synaptic transmission: A Monte Carlo simulation study. Biophys. J. **73**, 2874–2890, and references quoted therein.

[18] Engert, F. and Bonhoeffer, T. (1999) Dendritic spine changes associated with hippocampal long-term synaptic plasticity. Nature **399**, 66–70; Bonhoeffer, T., private communication.

[19] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) *Numerical Recipes*. Cambridge University Press. There are numerous editions, also for different programming languages.

[20] White, J.A., Rubinstein, J.T. and Kay, A.R. (2000) Channel noise in neurons. Trends Neurosci. **23**, 131–137

[21] Stuart, G., Spruston, N., Sakmann, B., and Häusser, M. (1997) Action potential initiation and backpropagation in neurons of the mammalian CNS. Trends Neurosci. **20**, 125–131

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

[23] Keener, J. and Sneyd, J. (1998) *Mathematical Physiology*, Springer, New York

[24] Cronin, J. (1987) *Mathematical Aspects of Hodgkin-Huxley Neural Theory*, Cambridge University Press

[25] Gerstner, W. and van Hemmen, J.L. (1992) Associative memory in a network of 'spiking' neurons. Network: Comput. Neural Syst. **3**, 139–164

[26] Kistler, W., Gerstner, W. and van Hemmen, J.L. (1997) Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. Neural Comput. **9**, 1015–1045

[27] Gerstner, W. (2000) A Framework for Spiking Neuron Models: The Spike Response Model, Chapter ? in this book.

[28] Hebb, D.O. (1949) *The Organization of Behavior*, Wiley, New York

[29] Orbach, J. (1998) *The Neuropsychological Theories of Lashley and Hebb*, University Press of America, Lanham, MD. This book gives a fascinating account of the indirect interaction between Hebb and K.S. Lashley, his thesis advisor.

[30] Herz, A.V.M, Sulzer, B., Kühn, R. and van Hemmen, J.L. (1988) The Hebb rule: Storing static and dynamic objects in an associative neural network. Europhys. Lett. **7**, 663–669

[31] 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

[32] Van Hemmen, J.L., Gerstner, W., Herz, A.V.M., Kühn, R. and Vaas, M. (1990) Encoding and decoding of patterns which are correlated in space and time. In: *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, G. Dorffner, ed., Springer, Berlin, pp. 153–162

[33] Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA **79**, 2554–2558

[34] Amit, D.J., Gutfreund, H., and Sompolinsky, H. (1985) Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys. Rev. Lett. **55**, 1530–1533

[35] Amit, D.J., Gutfreund, H., and Sompolinsky, H. (1987) Statistical mechanics of neural networks near saturation. Ann. Phys. (N.Y.) **173**, 30–67

[36] Van Hemmen, J.L. and Kühn, R. (1991) "Collective phenomena in neural networks". In: *Models of Neural Networks I*, Domany, E., van Hemmen, J.L. and Schulten, K., eds., Springer, Berlin, pp. 1–105; 2nd ed. (1995) pp. 1–113

[37] Kühn, R. and van Hemmen, J.L. (1991) "Temporal association". In: *Models of Neural Networks I*, Domany, E., van Hemmen, J.L. and Schulten, K., eds., Springer, Berlin, pp. 213–280; 2nd ed. (1995) pp. 221–288

[38] Hobson J.A. and McCarley R.W. (1977) The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. Am. J. Psychiatry **134**, 1335–1348.

[39] Crick, F. and Mitchison, G. (1983) The function of dream sleep. Nature **304**, 111–114

[40] Hopfield, J.J., Feinstein, D.I. and Palmer, R.G. (1983) 'Unlearning' has a stabilizing effect in collective memories. Nature **304**, 158–159

[41] Van Hemmen, J.L. (1987) Hebbian learning, its correlation catastrophe, and unlearning. Network: Comput. Neural Syst. **8**, V1–V17, and **9** (1998) 153

[42] Gerstner, W., Ritz, R. and van Hemmen, J.L. (1993) Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns, Biol. Cybern. **69**, 503–515

[43] Ritz, R., Gerstner, W., Fuentes, U. 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

[44] 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

[45] Gerstner, W., Kempter, R., van Hemmen, J.L. and Wagner, H. (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature **383**, 76–78

[46] Kempter, R., Gerstner, W., van Hemmen, J.L., and Wagner, H. (1998) Extracting oscillations: Neuronal coincidence detection with noisy periodic spike input. Neural Comput. **10**, 1987-2017

[47] Kempter, R., Gerstner, W. and van Hemmen, J.L. (1999) Hebbian learning and spiking neurons. Phys. Rev. E **59**, 4498–4514. Several considerations in §4 and §5 stem from this paper.

[48] 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

[49] Heerema, M. and van Leeuwen, W.A. (1999) Derivation of Hebb's rule, J. Phys. A: Math. Gen. **32**, 263–286

[50] Konishi, M. (1993) Listening with two ears. Sci. Am. **268**/4, 34–41

[51] Buonomano, D.V. and Merzenich, M.M. (1998) Cortical plasticity: from synapses to maps. Annu. Rev. Neurosci. **21**, 149–186

[52] Fitzsimonds, R.M. and Poo, M.-m. (1998) Retrograde signaling in the development and modification of synapses. Physiol. Rev. **78**, 143–170

[53] Urban, N.N. and Barrionuevo, G. (1996) Induction of Hebbian and non-Hebbian mossy fiber long-term potentiation by distinct patterns of high-frequency stimulation. J. Neurosci. **16**, 4293–4299

[54] Christofi, G., Nowicky, A.V., Bolsover, S.R. and Bindman, L.J. (1993) The postsynaptic induction of nonassociative long-term depression of excitatory synaptic transmission in rat hippocampal slices. J. Neurophysiol. **69**, 219–229

[55] 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

[56] Bi, G.-q. and Poo, M.-m. (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J. Neurosci. **18**, 10464–10472

[57] 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

[58] Bi, G.-q. and Poo, M.-m. (1999) Distributed synaptic modification in neural networks induced by patterned stimulation. Nature, **401**, 792–796

[59] Linden, D.J. (1999) The return of the spike: postsynaptic action potentials and the induction of LTP and LTD. Neuron **22**, 661–666. This is a nice review of the neurobiology underlying the present theory of time-resolved Hebbian learning.

[60] Caillard, O., Ben-Ari, Y. and Gaiarsa, J.-L. (1999) Mechanisms of Induction and Expression of Long-Term Depression at GABAergic Synapses in the Neonatal Rat Hippocampus. J. Neurosci. **19**, 7568–7577

[61] Sanders, J.A. and Verhulst, F. (1985) *Averaging Methods in Nonlinear Dynamical Systems*. Springer, New York

[62] Verhulst, F. (1996) *Nonlinear Differential Equations and Dynamical Systems*, 2nd ed., Springer, Berlin. Chapter 11 presents an excellent introduction to the method of averaging.

[63] Kistler, W.M. and van Hemmen, J.L. (2000) "Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials", Neural Comput. **12**, 385–405

[64] Leibold, C., Kempter, R. and van Hemmen, J.L. (2000) How synapses of spiking neurons learn a temporal-feature map. Preprint TUM, Garching

[65] Markram, H. and Tsodyks, M. (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons Nature **382**, 807–810.

[66] Tsodyks, M. and Markram, H. (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc. Natl. Acad. Sci. USA **94**, 719–723

[67] Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B. (1997) Synaptic depression and cortical gain control. Science **275**, 220-224

[68] Varela, J.A., Sen, K., Gibson, J., Fost, J., Abbott, L.F., and Nelson, S.B. (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. J. Neurosci. **17**, 7926–7940

[69] Senn, W., Segev, I. and Tsodyks, M. (1998) Reading neuronal synchrony with depressing synapses. Neural Comput. **10**, 815–819

[70] Senn, W. ,Tsodyks, M. and Markram, H. (2000) An algorithm for modifying neurotransmitter release probability based on pre-and post-synaptic spike timing. Neural Comput. **12**, to appear

[71] Tabak, J., Senn, W., O'Donovan, M.J. and Rinzel, J. (2000) Modeling of spontaneous activity in developing spinal cord using activity-dependent depression in an excitatory network. J. Neurosci. to appear

[72] Zucker, R.S. (1989) Short-term synaptic plasticity. Annu. Rev. Neurosci. **12**, 13–31

[73] Liley, A.W. and North, K.A.K. (1953) An electrical investigation of effects of repetitive stimulation on mammalian neuromuscular junctions. J. Neurophysiol. **16**, 509–527

[74] Magleby, K.L. and Zengel, J.E. (1975) A quantitative description of tetanic and post-tetanic potentiation of transmitter release at the frog neuromuscular junction. J. Neurophysiol. **245**, 183–208

[75] 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

[76] Gerstner, W., van Hemmen, J.L. and Cowan, J.D. (1993) What matters in neuronal locking? Neural Comput. **8**, 1689–1712

[77] Kistler, W.M. and van Hemmen, J.L. (1999) Short-term synaptic plasticity and network behavior. Neural Comput. **11**, 1579–1594

[78] Hale, J.K. and Koçak, H. (1991) *Dynamics and Bifurcations*, Springer, Berlin; see §8.6 for a simple introduction.

[79] Hale, J.K. (1963) *Oscillations in Nonlinear Systems*, McGraw-Hill, New York; republished by Dover, Mineola, NY, 1992.

[80] Linsker, R. (1986) From basic network principles to neural architecture: emergence of spatial-opponent cells. Proc. Natl. Acad. Sci. USA **83**, 7508–7512

[81] Sejnowski, T.J. and Tesauro, G. (1989) The Hebb rule for synaptic plasticity: algorithms and implementations. In: Byrne, J.H. and Berry, W.O., Eds., *Neural Models of Plasticity: Experimental and Theoretical Approaches*. Academic Press, San Diego, CA; Chap. 6, pp. 94–103

[82] Kohonen, T. (1984) *Self-Organization and Associative Memory*, Springer, Berlin

[83] Carr, C.E. (1993) Processing of Temporal Information in the Brain. Annu. Rev. Neurosci. **16**, 223–243. This wonderful review has been written with great didactic care. It is strongly recommended.

[84] MacKay, D.J.C. and Miller, K.D. (1990) Analysis of Linsker's application of Hebbian rules to linear networks. Network: Comput. Neural Syst. **1**, 257–297

[85] Miller, K.D. and MacKay, D.J.C. (1994) The role of constraints in Hebbian learning. Neural Comput. **6**, 100–126

[86] Kempter, R., Gerstner, W. and van Hemmen, J.L. (2000) Intrinsic rate normalization in spike-based and rate-based Hebbian learning. Submitted to Neural Comput.

[87] Bellman, R. (1970) *Introduction to Matrix Analysis*, 2nd. ed., McGraw-Hill, New York

[88] Merzbacher, E. (1961) *Quantum Mechanics*, Wiley, New York. The chapter on the spin, where one can find everything on Pauli spin matrices, is strongly recommended reading.

[89] Dobrunz, L.E. and Stevens, C.F. (1997) Heterogeneity of release probability, facilitation, and depletion at central synapses. Neuron **18**, 995–1008. This paper contains a nice appendix on underlying stochastic notions such as the relation between the readily releasable pool size and the release probability.

[90] Dobrunz, L.E., Huang, E.P. and Stevens, C.F. (1997) Very short-term plasticity in hippocampal synapses. Proc. Natl. Acad. Sci. USA **94**, 14843–14847

[91] Dobrunz, L.E. and Stevens, C.F. (1999) Response of hippocampal synapses to natural stimulus patterns. Neuron **22**, 157–166

[92] Buonomano, B.V. (2000) Decoding temporal information: A model based on short-term synaptic plasticity. J. Neurosci. **20**, 1129–1141

[93] Tuckwell, H.C. (1988) *Introduction to Theoretical Neurobiology*, Vol. 2, Cambridge University Press, New York

[94] Ash, R.B. and Gardner, M.F. (1975) *Topics in Stochastic Processes*, Academic, New York

[95] Snyder, D.L. and Miller, M.I. (1991) *Random Point Processes in Time and Space*, 2nd ed., Springer, New York

[96] Gnedenko, B.V. (1968) *The Theory of Probability*, 4th ed., Chelsea, New York; p. 265 (Bernstein's theorem, for App. A) and §51 (for App. B).

[97] Ash, R.B. (1972) *Real Analysis and Probability*. Academic, New York

[98] Durrett, R. (1996) *Probability: Theory and Examples*, 2nd ed., Duxbury Press, Belmont, MA

[99] Etemadi, N. (1981) An elementary proof of the strong law of large numbers. Z. Wahrscheinlichkeitstheorie verw. Geb. **55**, 119–122

[100] Prohorov, Yu.V. and Rozanov, Yu.A. (1969) *Probability Theory*, Springer, Berlin

[101] Gnedenko, B.V. and Kolmogorov, A.N. (1968) *Limit Distributions for Sums of Independent Random Variables*, 2nd ed., Addison-Wesley, Reading, MA

[102] For stationary processes there is a huge literature under the name "ergodic theorem". The theorem asserts that the limit $\lim_{n\to\infty} n^{-1}S_n$ exists with probability one. Only if the process is ergodic do we get the answer 0 a.e. See Halmos, P.R. (1956) *Ergodic Theory*, Chelsea, New York, for a succinct and elegant introduction, and Khinchin, A.I. (1949) Dover, New York, for excellent physical background information.

[103] Scott, D.J. (1973) Central limit theorems for martingales and for processes with stationary independent increments using the Skorohod representation approach. Adv. Appl. Probab. **5**, 119–137