Network 3 (1992) 139–164. Printed in the UK

Associative memory in a network of 'spiking' neurons

Wulfram Gerstner†§ and J Leo van Hemmen‡

† Department of Physics, University of California, Berkeley, CA 94720, USA
‡ Physik-Department der Technischen Universität München, Institut für Theoretische Physik, D-8046 Garching bei München, Federal Republic of Germany

Received 19 March 1991

Abstract. The Hopfield network provides a simple model of an associative memory in a neuronal structure. It is, however, based on highly artificial assumptions, especially the use of formal two-state neurons or graded-response neurons. In this paper we address the question of what happens if formal neurons are replaced by a model of 'spiking' neurons. We do so in two steps. First, we show how to include refractoriness and noise into a simple threshold model of neuronal spiking. The spike trains resulting from such a model reproduce the distribution of interspike intervals and gain functions found in real neurons. In a second step we connect the model neurons so as to form a large associative memory system. The spike transmission is described by a synaptic kernel which includes axonal delays, 'Hebbian' synaptic efficacies, and a realistic postsynaptic response. The collective behaviour of the system is predicted by a set of dynamical equations which are exact in the limit of a large and fully connected network that has to store a finite number of patterns. We show that in a stationary retrieval state the statistics of the spiking dynamics is completely wiped out and the system reduces to a network of graded-response neurons. In the case of an oscillatory retrieval state, however, the spiking noise and the internal time constants of the neurons become important and determine the behaviour of the system.

1. Introduction

Real biological networks are complex systems built of neurons with a variety of electrical and biochemical properties. In a mathematical model of these networks one has to neglect a lot of details and concentrate on some important properties. But what are the relevant facts? Information received at the sensory level is encoded in *spike trains* which are then transmitted to different parts of the brain where the main processing step occurs. Since all the spikes of any particular neuron look alike, the information of the spike train is obviously not contained in the exact shape of the spikes, but rather in their arrival times in relation to earlier spikes or in correlation with other neurons. A model neuron which tries to keep track of the voltage trace even during the spiking—like the Hodgkin–Huxley equations (Hodgkin and Huxley 1952) and similar models (FitzHugh 1961, Nagumo *et al* 1962)—carries therefore non-essential details if we are interested in the information of the spike train only. On the other hand, a simple two-state neuron or threshold model is too simplistic

§ Present address: Physik-Department der Technischen Universität München, Institut für Theoretische Physik, D-8046 Garching bei München, Federal Republic of Germany. E-mail: wulfram.gerstner@physik.tu-muenchen.de

0954-898X/92/020139+26$04.50

since it cannot reproduce the variety of spiking behaviour found in real neurons. The same is true for continuous or analogue neurons which disregard the stochastic nature of the spiking process completely. In this work we construct a model neuron which is intermediate between these extremes: we are not concerned with the shape of the spikes and detailed voltage traces but we want to model spike trains with realistic interval distributions and rate functions.

In the first part of the paper (section 2) we present the model neuron and give examples of its firing behaviour. Depending on the internal and external parameters, the model neuron will produce random or regular firing, oscillating, bursting or adaptive behaviour. In the second part of the paper (section 3), the model neurons are linked together via synapses. The description of the synaptic connections in section 3.1 emphasizes axonal delays, a 'Hebbian' learning rule for the synaptic efficacies, and a realistic postsynaptic response function. Thus, we construct an associative memory for a finite number of patterns. In the more technical sections 3.2–3.4 we derive exact equations of motion and give an analytical solution for the fully connected network in terms of the overlaps with the learnt patterns. We show that in the case of stationary solutions the system can be reduced to a network of graded-response neurons. In section 3.5 we discuss the static solutions and give analytical results for two typical examples. Sections 3.6 and 3.7 feature oscillatory solutions. Here we emphasize the influence of the spiking noise and the internal time constants of the model on the collective behaviour of the network. We finish with a couple of conclusions in section 4.

2. Properties of a single model neuron

2.1. Definition of the model neuron

From a neural network point of view it is often convenient to consider a neuron as a simple computational unit with no internal parameters. In this case, the neuron is described either as a 'digital' threshold unit or as a nonlinear 'analogue' element with a sigmoid gain function. If we interpret the input as the driving current of the neuron and the output as its firing rate, we achieve a simple mapping to real neurons. While such a simple model might be useful for formal considerations in abstract networks, it is hard to see how it could be modified to include other features of neuronal spiking: How can we account for the statistical properties of the spike train beyond the mean firing rate? And what about bursting, adapting or oscillating neurons? To mention but a few of the problems with real neurons.

On the other hand, a model which takes into account all the known biological details of a particular neuron should start with the microscopic dynamics of electrically or chemically gated ion channels. Thus it confronts us with the task of integrating a large number of coupled nonlinear differential equations, a task which is not only rather infeasible for larger networks but might not even be necessary since the only information relevant to the network seems to be the spiking event of the neuron.

We would therefore like to adopt an intermediate approach and go one step beyond the simple input–output unit without worrying, though, about too many microscopic details. Our description of the spiking dynamics emphasizes three basic notions of neurobiology: *threshold, refractory period* and *noise.* In particular, we describe the internal state of the neuron by a continuous variable h which depends on the synaptic contributions from other neurons as well as on the spiking history of the

neuron itself. Though h is a formal variable with no direct physical meaning, it is useful to think of it as some current through the membrane of the soma. In a simple threshold crossing process, a spike would be initiated as soon as $h(t)$ crosses the threshold θ. Due to the statistical fluctuations of the momentary ion currents around $h(t)$, however, the spiking will be a statistical event, spikes coming a bit too early or a bit too late compared to the formal threshold crossing time. This fact will be taken into account by introducing a stochastic spiking 'rate constant' ρ which depends on the difference between the membrane current h and the threshold θ. We choose the rate constant ρ in analogy with the kinetic reaction constant to have an exponential dependence upon this difference:

$$\rho(h) = \frac{1}{\tau_0} \exp[\beta(h - \theta)] \tag{1}$$

where the formal temperature β^{-1} is a measure for the noise and τ_0 is an internal time constant of the neuron characterizing the rate at threshold. For $\beta \to \infty$ the rate constant ρ changes from 0 to ∞ when h passes from values below threshold to a value above threshold. This means that a spike will be fired immediately after h has crossed the threshold. For finite β the spiking will be delayed or advanced stochastically. The rate constant ρ of stochastic spiking is the relevant parameter in a continuous-time description and should not be confused with the probability P_F of firing during one time step in a discrete-time representation. If the variable h changes only slowly during a conveniently chosen time Δt ($\mathrm{d}h/\mathrm{d}t << h/\Delta t$), we can integrate the spiking rate constant ρ over Δt yielding the probability P_F of firing during a time step of length Δt:

$$P_F(h) = 1 - \exp[-\Delta t \rho(h)] = 1 - \exp\left\{-\frac{\Delta t}{\tau_0} \exp[\beta(h - \theta)]\right\}. \tag{2}$$

This is a sigmoid function with $P_F(h) \to 1$ for $h >> \theta$ and $P_F(h) \to 0$ for $h << \theta$. It is similar to the firing probability $P_F(h) = \frac{1}{2}(1 + \tanh \beta h)$ usually assumed in statistical neural networks, but it lacks the 'antisymmetry' under inversion. With equation (2) we have an analytical procedure to go from the continuous-time model to a dynamics with discrete time-steps which we will use later on.

If a spike is initiated in a real neuron, the neuron goes through a cycle of ion influx and efflux which changes the internal currents on a fast time scale and prevents immediate firing of another spike. To model this, we reset the variable h after each spike by adding a negative *refractory field* $h^r(t)$,

$$h(t) = h^s(t) + h^r(t) \tag{3}$$

with

$$h^r(t) = \sum_i \epsilon^r(t - t_i) \tag{4}$$

where t_i is the firing time of the ith spike and $h^s(t)$ is the sum of the postsynaptic contributions due to incoming spikes from other neurons. The *refractory function* $\epsilon^r(\tau)$ is the central notion of our model. The introduction of such a function is

justified at this point purely phenomenologically. It is based on the traditional neurobiological ideas of the refractoriness of a neuron and it allows us to model neurons with different spiking characteristics. On a microscopic level we consider the refractory function as a means to describe the effects of the various ion currents. Simply speaking, the main effect of the ion currents immediately after the spike is to raise (or lower) the effective threshold value. This is exactly what is modelled by the refractory function $\epsilon^r(\tau)$.

2.2. *Examples of spiking behaviour*

Instead of integrating the unknown ion currents to get the refractory function, we adopt a phenomenological approach and guess a refractory function which will lead to a realistic spiking behaviour of the model neuron. As a first example we assume a simple refractory function with an *absolute* and a quickly decaying *relative* refractory period. A neuron with such a refractory function shows the desired sigmoid dependence of the firing frequency on the input current and at the same time realistic spiking statistics (figure 1). Indeed, the interval distribution changes from approximately Poisson for driving currents below threshold to approximately Gaussian above threshold. This can be compared with the experimental spike discharge patterns of spontaneous and stimulated activity (figure 2) that have been found in the cochlear nucleus of anesthetized cats (Pfeiffer and Kiang 1965). The general shape of the distribution and its stimulus dependence is rather similar to figure 1(*c*).

An analytical expression for the interval distribution can be found as follows. Consider an ensemble of model neurons that *all* have fired at $\tau = 0$. Due to absolute refractoriness, they cannot spike immediately afterwards, but then the effect of refractoriness as modelled by the refractory function $\epsilon^r(\tau)$ decays and more and more neurons of the ensemble will fire again. The portion of neurons that have not spiked for a time τ after the spike at $\tau = 0$ is given by the 'survival function' (Perkel *et al* 1967)

$$p(\tau) = \exp\left(-\int_0^\tau \mathrm{d}s\, \rho(s)\right) \tag{5}$$

with the firing rate constant, equation (1),

$$\rho(s) = (1/\tau_0)\exp\{\beta[h^r(s) + h^s - \theta]\}. \tag{6}$$

The *interval distribution* $D(\tau)\mathrm{d}\tau$ is the portion of neurons that have 'survived' for the time τ *and* fire between τ and $\tau + \mathrm{d}\tau$, i.e.

$$D(\tau) = p(\tau)\rho(\tau) = -\frac{\partial}{\partial\tau}p(\tau). \tag{7}$$

This is the function that has been plotted in figure 1(*c*).

The mean interval length is obviously given by

$$\overline{\tau} = \int_0^\infty \tau D(\tau)\,\mathrm{d}\tau = \int_0^\infty p(\tau)\,\mathrm{d}\tau. \tag{8}$$

The *mean firing rate* can then be defined as the inverse of the mean interval length $\overline{\tau}$, and if we plot the mean firing rate as a function of the input current h^s we get

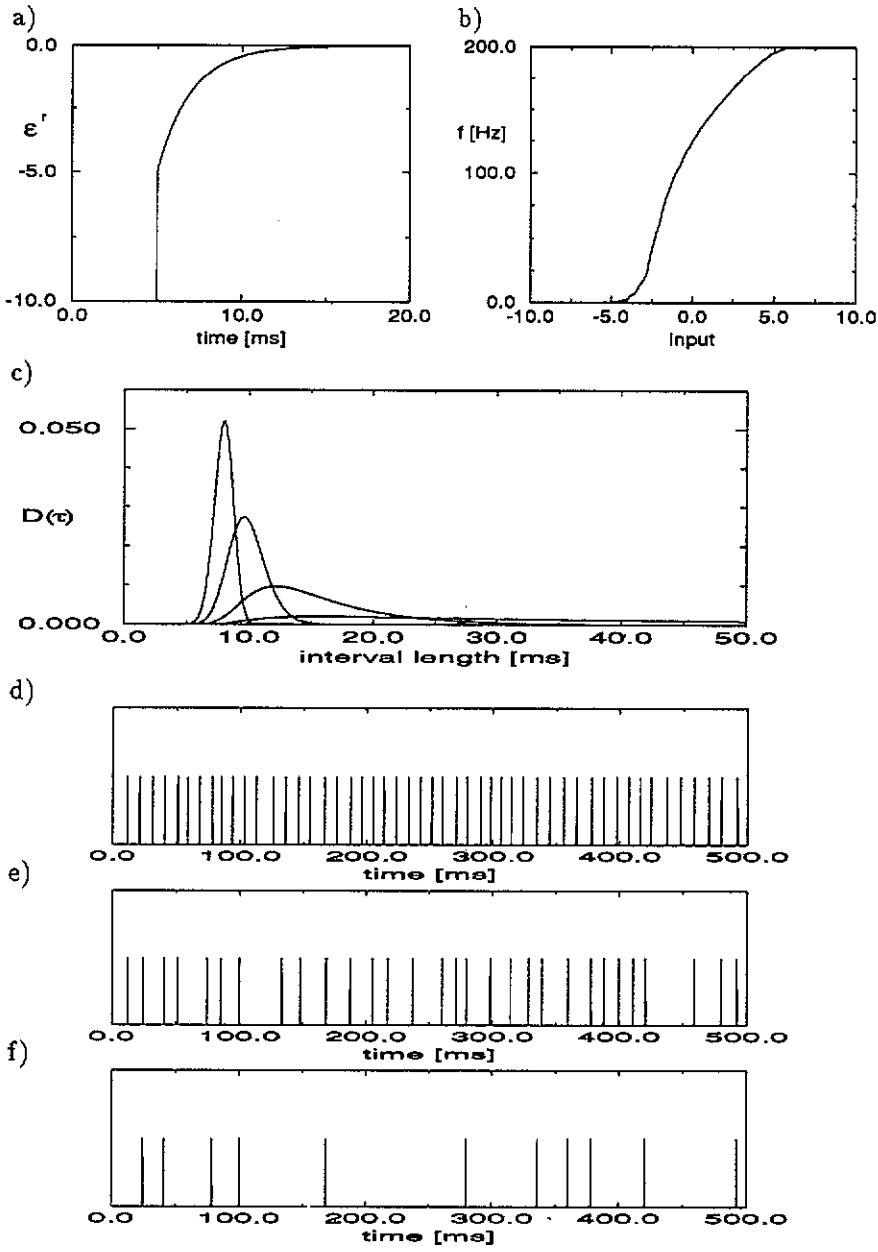

Figure 1. Standard neuron. (*a*) Refractory function $\epsilon^r(\tau)$ with an absolute refractory period of 5 ms and an exponentially decaying relative refractory period (time constant 2 ms) yielding the gain function and spiking characteristics shown in (*b*)–(*f*). (*b*) Gain function with the typical sigmoid dependence of the spiking frequency upon the input current. (*c*) Interval distribution $D(\tau)$. For large input currents the distribution is sharply peaked and of 'Gaussian' shape indicating regular firing (input $I = 0$, leftmost curve; $I = -1$, second curve). Low input yields a flat distribution with a long exponential tail as expected for stochastic firing ($I = -2$, third curve; $I = -3$, bottom curve). (*d*)–(*f*) examples of spike trains. For large input current the firing of the model neuron is regular ($I = -1$ in (*d*)), for low input currents it is stochastic ($I = -2$ in (*e*); $I = -3$ in (*f*)).

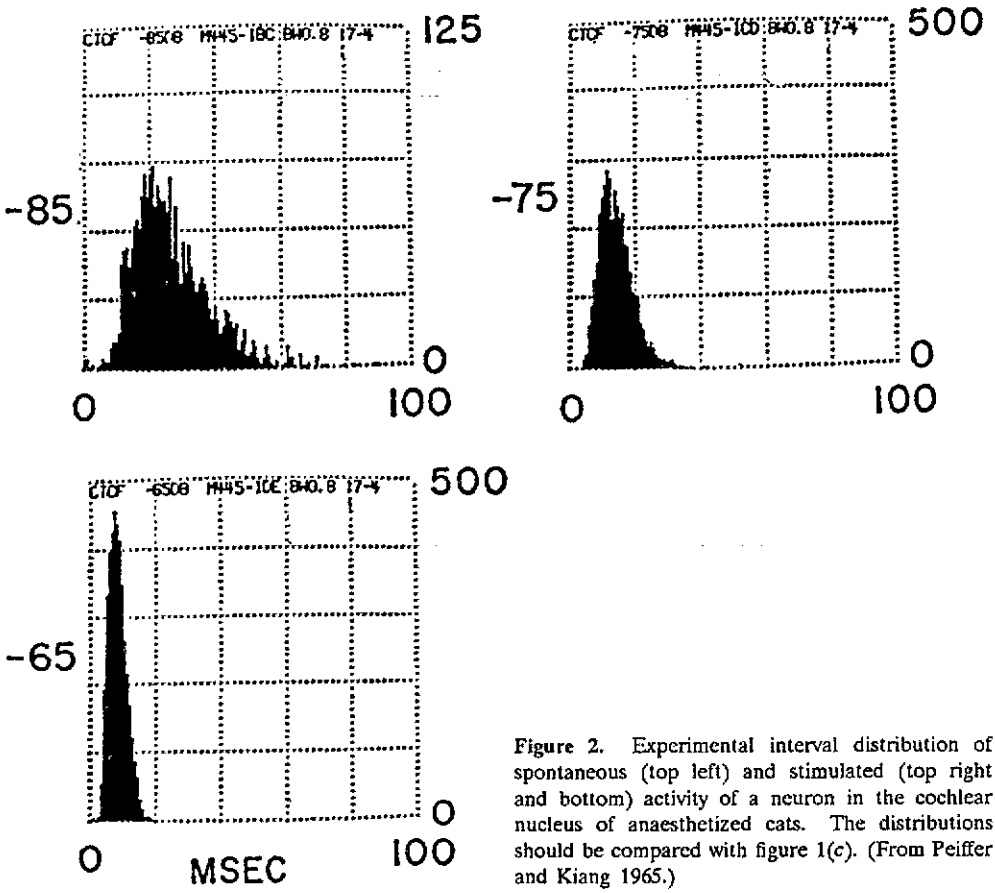

Figure 2. Experimental interval distribution of spontaneous (top left) and stimulated (top right and bottom) activity of a neuron in the cochlear nucleus of anaesthetized cats. The distributions should be compared with figure 1(*c*). (From Peiffer and Kiang 1965.)

the gain function $f(h^s)$ of the neuron (figure 1(*b*)). These notions will be again of importance in section 3 where we discuss a large network of the model neurons.

Other forms of the refractory function lead to a variety of different firing characteristics including oscillating or bursting behaviour. While it remains an open question whether bursting is a collective effect of an ensemble of neurons or else an intrinsic property of a single neuron, it is interesting to see that there exist simple models which *are* capable to produce bursting as a single neuron effect. Figure 3 shows such a 'bursting' neuron at three different input levels. At a low input level the bursts are noise induced and appear at irregular intervals (figure 3(*b*)), with higher input currents the bursts are more frequent and appear in regular intervals (figure 3(*c*)). If the input current rises even further, the 'bursting' neuron switches to a regularly oscillating behaviour (figure 3(*d*)). The suppression of the long interburst intervals leads to a marked increase in the mean firing rate (figure 3(*e*)). Please note the pronounced adaptation effect in figure 3(*d*). This is due to the slow time constant of the refractory function and can be seen in any model neuron with a long-tailed refractory function (figure 4). In these cases, the gain function reaches only slowly its saturation value (figure 4(*b*)); even at a high input level, the neuron settles to a state with a comparatively small firing rate.

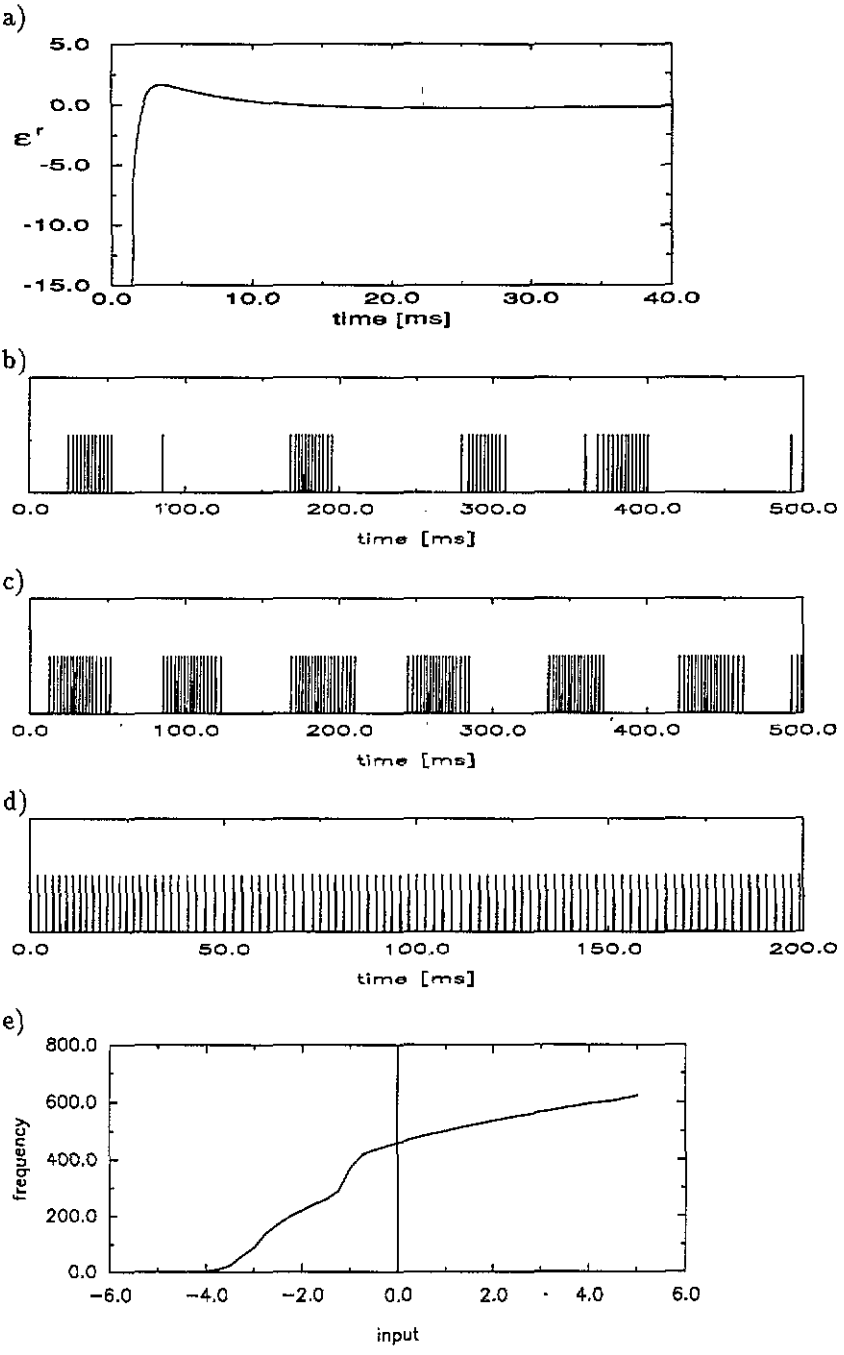

Figure 3. Bursting neuron. (*a*) Refractory function $\epsilon^r(\tau)$ with a slight overshooting at intermediate time scales leading to the bursting behaviour shown in (*b*)–(*e*). (*b*)–(*d*) Examples of spike trains. At a low input level ($I = -3$ in (*b*)) the bursts have various lengths and appear in irregular intervals. For higher input current ($I = -2$ in (*c*)) the bursting is regular. (*d*) Strong input leads to regular oscillations of the model neuron. Note the adaption at the beginning of the spike train. (*e*) Gain function of the bursting model neuron.

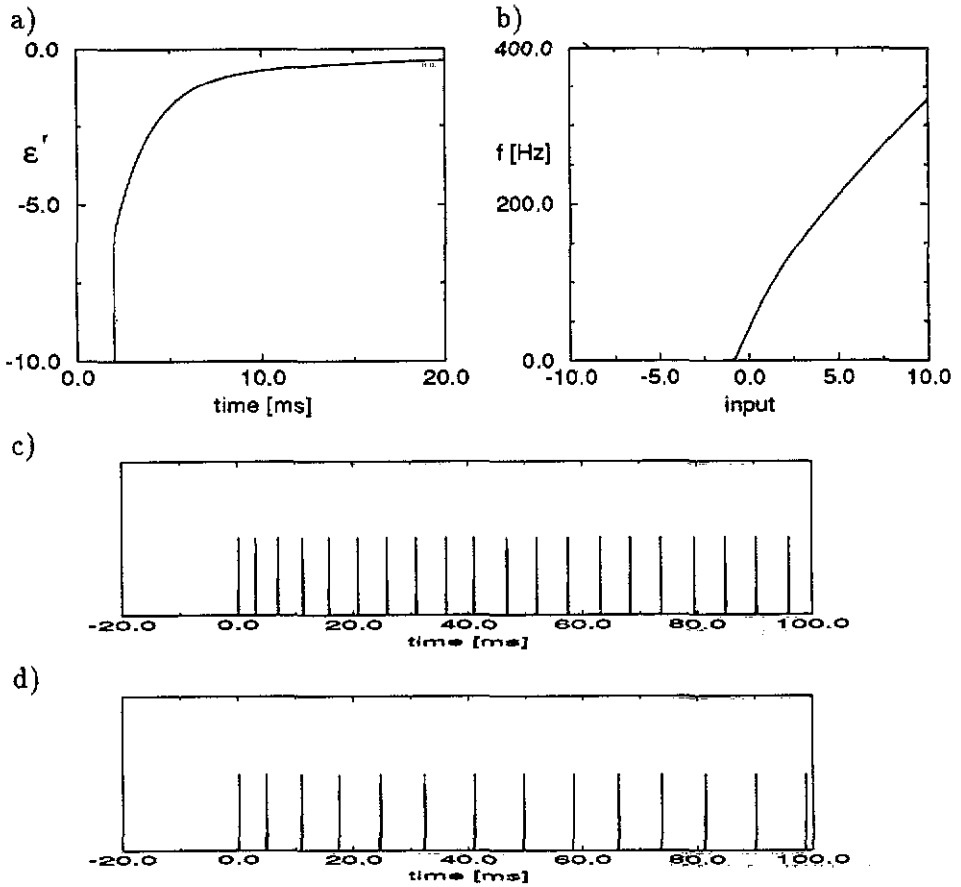

Figure 4. Adaptive neuron. A refractory function $\epsilon^r(\tau)$ with a long relative refractory period (*a*) leads to a adaptive behaviour. This is seen in the spike trains (*c*) and (*d*) which start with a couple of short intervals before they settle at a much longer spiking period. With an input of $I = 4$ (*c*) the interval length is growing from 3 to 7 ms; with $I = 2$ (*d*) intervals range between 4.8 and 8 ms. The firing rate after an adaptation period of 50 ms is plotted in (*b*).

2.3. Comparison to the Hodgkin–Huxley model

As a final application of the ideas presented here, we would like to compare our approach to the well-known Hodgkin–Huxley model (Hodgkin and Huxley 1952). Numerical integration of the Hodgkin–Huxley equations for a realistic set of parameters shows that the current flow during the spiking process and immediately afterwards is always the same, independent of the level of the driving current. This is the underlying reason why we can hope to replace the internal dynamics by a simple refractory function $\epsilon^r(\tau)$ which can be found as follows. The ideal refractory function should be designed in a way that the simplified model exactly reproduces the gain function $f(I)$ of the Hodgkin–Huxley equations. We identify now the driving current I of the Hodgkin–Huxley model with the input h^s of our model. For a fixed $h^s > \theta$, a noiseless model neuron ($\beta \to \infty$) fires regularly with a period T which can be found

from

$$\epsilon^{\mathrm{r}}(T) \approx h^{\mathrm{r}}(T) = \theta - h^{\mathrm{s}}. \tag{9}$$

The period T should equal the firing period of the Hodgkin–Huxley model at the same input

$$T = 1/f(I) \tag{10}$$

with $I = h^{\mathrm{s}}$. Solving equations (9), (10) for different driving currents we can construct the ideal $\epsilon^{\mathrm{r}}(\tau)$ which we have plotted in figure 5(*a*). To simplify it further, we can approximate this function in the relevant region of input currents by an ansatz

$$\epsilon^{\mathrm{r}}(\tau) \approx \begin{cases} -\infty & \text{for } 0 < \tau \leqslant t_0 \\ \alpha[(\tau - t_0)^{-1} - (\tau_{\mathrm{m}} - t_0)^{-1}] & \text{for } t_0 < \tau \leqslant \tau_{\mathrm{m}} \\ 0 & \text{for } \tau > \tau_{\mathrm{m}} \end{cases} \tag{11}$$

with parameters $\alpha = 48, \tau_{\mathrm{m}} = 19.1$ ms and $t_0 = 9$ ms. For the threshold θ in equations (1) and (10) we take the threshold current of the Hodgkin–Huxley model. The gain function of both the Hodgkin–Huxley model and the approximated model are plotted in figure 5(*b*). The finite range of the refractory function, i.e. $\epsilon^{\mathrm{r}}(\tau) = 0$ for $\tau > \tau_{\mathrm{m}}$, results in a discontinuity of the gain function at threshold.

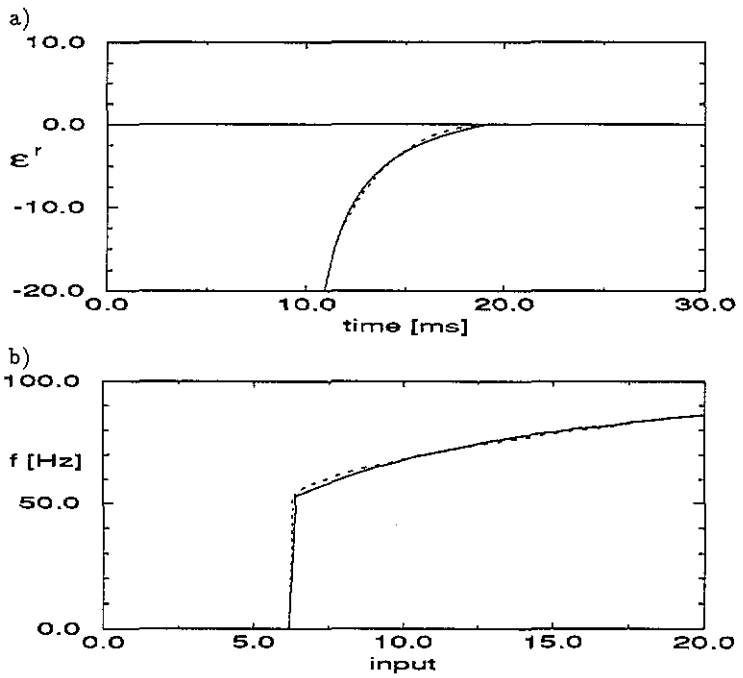

Figure 5. Hodgkin–Huxley neuron. (*a*) Refractory function $\epsilon^{\mathrm{r}}(\tau)$. Starting from the gain function of the full set of equations we derive the refractory function of this model (dotted curve, see text for details). An approximated refractory function is shown by the full curve. It vanishes for $\tau > \tau_{\mathrm{m}} = 19.1$ ms. (*b*) Gain function of both the approximated model (full curve) and the full model (dotted curve). Note the discontinuity at threshold.

The above model may be compared with the approach of Abbott and Kepler (1991) who systematically reduce the Hodgkin–Huxley equations to various model systems, in particular to a dynamic binary model. They have shown that it is possible to describe the simplified system by a continuous variable u combined with a fire-and-reset process which is taken care of by a two-state variable S. This result may give a qualitative justification of our ansatz, but the details of our description are different from their model. Abbott and Kepler concentrate on the *dynamic* description of refractoriness and recharging after firing, putting less emphasis on the driving current I. Our approach, however, can exactly reproduce the I-dependence and thus the gain function of the Hodgkin–Huxley model, at the cost of neglecting details of the underlying dynamics.

The comparison to the Hodgkin–Huxley model, however, should not be overestimated. One has to realize that Hodgkin and Huxley have proposed their equations to describe a rather special experimental system, i.e. the giant axon of the squid. Their equations give an accurate description of this system under well defined conditions, but they should not be taken to be general and to describe the spiking of all other types of neuron. On the other hand, our extremely simplified model which does not aim at a detailed description of the internal ion currents is flexible enough to account for various spiking features like adaptation, oscillation, 'bursting', and randomness of firing. Since the description of an individual neuron is comparatively simple it seems worthwhile to couple many of these neurons together into a large neural network.

3. Behaviour of a network of model neurons

3.1. Definition of the network

So far we have only described a dynamics that initiates the spikes in the neurons. Now we have to describe the synaptic transmission of the spikes to other neurons. To keep track of the spikes we assign to each neuron a two-state variable S_i which usually rests at -1 and flips to $+1$ only when a spike has been initiated. In the discrete-time representation that we assume in this and the following subsection the output of each neuron is described by a sequence of Ising spins $S_i(t_n)$. It is convenient to choose the length of a time step Δt to be equal to or slightly less than a spike width, say 0.5 ms. According to the considerations of section I, this should enable us to resolve *all* the information contained in a spike train with a reasonable number of discrete steps.

With the choice of the spike width as the length of a single time-step we have already selected the basic time scale of our model. The model contains, however, another time scale which is associated with the maximum spiking frequency of the neurons. Even when firing at maximum frequency the output of a neuron is still a well resolved spike train, the spikes being nicely separated from one another. Let us assume that a maximally firing neuron is quiet for $(\gamma - 1)$ time-steps before it fires again. In terms of our model this implies that the absolute refractory period has a duration of $(\gamma - 1)\Delta t$ while the maximum firing frequency is $(1/\gamma\Delta t)$.

The length of the absolute refractory period is an important parameter of our model and γ can be used to normalize the output of neuron i so as to allow a comparison with formal two-state neurons. We define

$$\tilde{S}_i = (\gamma/2)(S_i + 1). \tag{12}$$

This definition leads to a time-averaged output of $+1$, if the neuron i fires with maximum frequency, and an average output of 0, if the neuron is inactive. These limiting cases correspond therefore to a formal (1-0)-neuron σ, where $\sigma_i = +1$ has the usual interpretation of firing with maximum frequency, $\sigma_i = 0$ as inactive.

Now we want to gather these model neurons into a large network. In the spirit of the Hopfield model (Hopfield 1982) we assume a fully connected network in which we store a finite number of patterns.

If a neuron i receives a spike from another neuron j, the spike will evoke a postsynaptic potential at i, the strength of which depends on the synaptic efficacy J_{ij}. The time course of this response, however, can be taken to have a generic form independent of the strength of the synapse. If the synaptic transmission of a single spike evokes postsynaptic currents of the functional form $\epsilon^s(\tau)$, then the total postsynaptic response will be

$$h_i^s(t_n) = \sum_j J_{ij} \sum_{\tau_m=0}^{\tau_{max}} \epsilon^s(\tau_m) \tilde{S}_j(t_n - \tau_m) \tag{13}$$

where we assumed a linear superposition of the synaptic contributions. The function $\epsilon^s(\tau)$ can be the experimentally measured response function of a particular neuron or a typical model function normalized such that $1 = \int_0^\infty \epsilon^s(\tau)\, d\tau$. In our simulations we typically take

$$\epsilon^s(\tau) = \begin{cases} 0 & \text{for } \tau < \Delta_a \\ [(\tau - \Delta_a)/(\tau_s)^2] \exp[-(\tau - \Delta_a)/\tau_s] & \text{for } \tau \geqslant \Delta_a \end{cases} \tag{14}$$

where τ_s is the characteristic response time of the synapse and Δ_a is the axonal delay (figure 6). For the synaptic efficacies we assume the Hebbian matrix

$$J_{ij} = \frac{2}{N} \sum_{\mu=1}^{q} \xi_i^\mu \xi_j^\mu \qquad \text{for } i \neq j \tag{15}$$

where the variables $\xi_i^\mu \in \{\pm 1\}$ with $1 \leqslant i \leqslant N$ and $1 \leqslant \mu \leqslant q$ describe q random patterns with $\text{Prob}(\xi_i^\mu = \pm 1) = \frac{1}{2}$. The prefactor $2/N$ is useful for normalization.

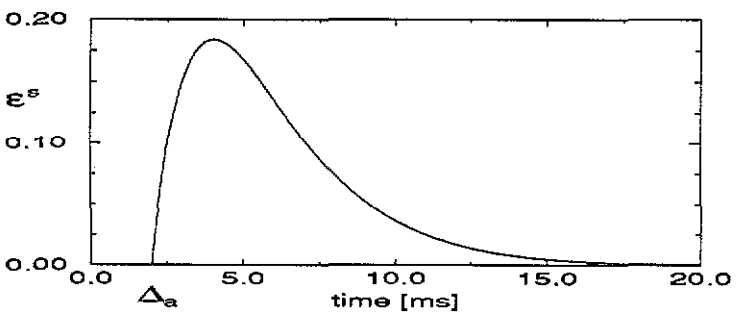

Figure 6. The synaptic response function $\epsilon^s(\tau)$ as a function of time τ after the firing of a presynaptic neuron. Due to the axonal delay, there is no response during a time Δ_a. The response function reaches its maximum after a rise time ($\tau_s \approx 2$ ms) and decays to zero afterwards.

The synaptic weights J_{ij} together with the synaptic response function $\epsilon^s(\tau)$ are the basic quantities in the linear ansatz (13) to describe the effect of the incoming spikes. After we have defined these quantities it is straightforward to incorporate the internal dynamics of the neurons as described in the preceding section. The refractory field can be introduced as the diagonal element of the synaptic connection matrix,

$$h_i^r(t_n) = \sum_{\tau_m=0}^{\tau_{max}} J_{ii}(\tau_m)[S_i(t_n - \tau_m) + 1]/2 \tag{16}$$

which is a kind of self-interaction. If all neurons are equivalent, the diagonal elements must be independent of i and $J_{ii}(\tau) = \epsilon^r(\tau)$ describes the generic response of a neuron after emission of a spike. For later calculations we note that in the case of discrete-time dynamics and a finite range of $\epsilon^r(\tau_n)$ with $\tau_n \leqslant \tau_{max}$, the refractory field takes only a finite number of values.

3.2. Derivation of the equation of motion

Through equation (2) we have defined the spiking probability $P_F(h_i)$ of neuron i as a function of the variable h_i. At every time-step all neurons are updated in parallel. We have also defined h_i in terms of the spiking history of all neurons

$$h_i(t) = h_i^s(t) + h_i^r(t). \tag{17}$$

The synaptic contribution h_i^s, equation (13) couples neuron i to all other neurons while the refractory field h_i^r gives a self-interaction of neuron i with itself. Thus we are left with a set of highly coupled stochastic equations, which we are going to solve.

We start by rewriting equation (17) in terms of the *overlaps*

$$m_\mu(t) = \frac{2}{N} \sum_{j=1}^{N} \xi_j^\mu \tilde{S}_j(t) \qquad \text{for } \mu = 1, \ldots, q \tag{18}$$

which measure the correlation of the firing state with the patterns. The overlap m_μ takes its maximum value γ only if the firing state is identical to pattern μ. It is 1, if all neurons that should be active during retrieval of pattern μ fire incoherently and with maximum rate while all other neurons stay quiescent. Using equations (13) and (16) we find

$$\begin{aligned} h_i(t_n) &= h_i^s(t_n) + h_i^r(t_n) \\ &= \sum_{\mu=1}^{q} \sum_{\tau_m=0}^{\tau_{max}} \xi_i^\mu \epsilon^s(\tau_m) m_\mu(t_n - \tau_m) + \sum_{\tau_m=0}^{\tau_{max}} \epsilon^r(\tau_m)[S_i(t_n - \tau_m) + 1]/2. \end{aligned} \tag{19}$$

Since m_μ is a global quantity independent of the index i of the neuron, all neurons which have learnt the same local information $x = (\xi_i^\mu, \mu = 1, \ldots, q)$ and which have the same momentary refractory field h^r, will experience the same momentary field $h = h(x, h^r, t_n)$.

For a finite number of patterns and finite range of the refractory field, the number of different x and h^r is finite. To be specific we assume $n + 1$ possible values for

the refractory field $h^r \in \{h^0, h^1, ..., h^n\}$ and there are 2^q different states for the vector $x \in \{\pm 1\}^q$. At each time-step we can therefore separate the N neurons into a finite number of different classes, the sublattices $I(x, h^r, t_n)$ (van Hemmen and Kühn 1986, van Hemmen *et al* 1986, 1988), each class containing neurons which experience the same field $h(x, h^r, t_n)$. The number of neurons belonging to the same class, $|I(x, h^r, t_n)|$, may be different for different classes and may change in time. In the limit of $N \to \infty$, however, each class contains an infinite number of neurons (or none at all). It follows that the *activity* A of the sublattice $I(x, h^r, t)$

$$A(x, h^r, t_n) = \frac{1}{|I(x, h^r, t_n)|} \sum_{i \in I(x, h^r, t_n)} \tilde{S}_i(t_n) \tag{20}$$

takes exactly its average value

$$\begin{aligned} A(x, h^r, t_n) &= \lim_{N \to \infty} \frac{1}{N p(x, h^r, t_n)} \sum_{i \in I(x, h^r, t_n)} \tilde{S}_i(t_n) \\ &= \gamma P_F(h(x, h^r, t_{n-1})) \end{aligned} \tag{21}$$

where $p(x, h^r, t_n)$ is the probability for a neuron to belong to the class $I(x, h^r, t_n)$ at time t_n , i.e. $Np(x, h^r, t_n) = |I(x, h^r, t_n)|$. Using these concepts (van Hemmen and Kühn 1986, van Hemmen *et al* 1986, 1988, Riedel *et al* 1988, Herz *et al* 1988, 1989) we derive an exact expression for the evolution of the overlaps

$$\begin{aligned} m_\mu(t_{n+1}) &= \lim_{N \to \infty} \frac{2}{N} \sum_{j=1}^{N} \xi_j^\mu \tilde{S}_j(t_{n+1}) \\ &= \lim_{N \to \infty} \sum_{x, h^r} \frac{2}{N} \sum_{j \in I(x, h^r, t_{n+1})} \xi_j^\mu \tilde{S}_j(t_{n+1}) \\ &= \lim_{N \to \infty} \sum_{x, h^r} 2p(x, h^r, t_n) x^\mu \sum_{j \in I(x, h^r, t_{n+1})} \frac{1}{N p(x, h^r, t_n)} \tilde{S}_j(t_{n+1}). \end{aligned} \tag{22}$$

Using (21) we obtain

$$m_\mu(t_{n+1}) = 2\gamma \sum_{x, h^r} x^\mu p(x, h^r, t_n) P_F(x, h^r, t_n) \tag{23}$$

and

$$h(x, h^r, t_n) = \sum_{\mu=1}^{q} x^\mu \sum_{\tau_m=0}^{\tau_{max}} \epsilon^s(\tau_m) m_\mu(t_n - \tau_m) + h^r. \tag{24}$$

In this implicit equation for the overlaps m_μ, the values of the $p(x, h^r, t_n)$ are still unknown. In a second step we must therefore find the equations which describe the relative probability of a neuron to belong to one specific sublattice. This, however, is not a difficult problem. For randomly drawn patterns each vector $x = (x^1, x^2, ..., x^q) \in \{\pm 1\}^q$ is equally likely. For a given x on the other hand

the time evolution of $p(x, h^r, t_n)$ is given by the master equation

$$\begin{aligned} p(x, h^0, t_{n+1}) &= p(x, h^0, t_n)[1 - P_F(h^0)] + p(x, h^1, t_n)[1 - P_F(h^1)] \\ p(x, h^k, t_{n+1}) &= p(x, h^{k+1}, t_n)[1 - P_F(h^{k+1})] \qquad 1 \leqslant k \leqslant n-1 \\ p(x, h^n, t_{n+1}) &= \sum_{k=0}^{n} p(x, h^k, t_n) P_F(h^k) \end{aligned} \tag{25}$$

which can be solved for the stationary distribution in terms of the known firing probabilities $P_F(h^k)$.

3.3. Continuous time limit and interpretation

For analytical considerations it is more convenient to consider the corresponding system in continuous time. We write

$$\begin{aligned} P_F &= \rho \Delta t \\ h^r &= h^r(s) \\ p(x, h^r, t_n) &= \overline{p}(x, h^r(s), t) \Delta s \\ \epsilon^s(\tau_m) &= \overline{\epsilon}^s(\tau) \Delta t \end{aligned} \tag{26}$$

and take the continuum limit $\Delta t \to 0$ while keeping the duration of the absolute refractory period $\gamma \Delta t = \gamma_r$ fixed. Then (23) and (24) turn into a set of coupled integral equations

$$m_\mu(t) = 2\gamma_r \sum_x x^\mu \int \mathrm{d}s\, \overline{p}(x, h^r(s), t) \rho(x, h^r(s), t) \tag{27}$$

and

$$h(x, h^r(s), t) = \sum_{\mu=1}^{q} x^\mu \int \mathrm{d}\tau\, \overline{\epsilon}^s(\tau) m_\mu(t - \tau) + h^r(s). \tag{28}$$

In this limit, the spike train of each neuron $\{S_i(t_1), \ldots, S_i(t_n), \ldots\}$ reduces to a sequence of δ-functions, while the master equation (25) becomes

$$\frac{\partial}{\partial t} \overline{p}(x, h^r(s), t) = -\overline{p}(x, h^r(s), t) \rho(x, h^r(s), t) + \frac{\partial}{\partial h} \overline{p}(x, h^r(s), t) \frac{\partial}{\partial s} h^r(s) \tag{29}$$

with

$$\overline{p}(x, h^r(0), t) = \int_0^\infty \mathrm{d}s\, \overline{p}(x, h^r(s), t) \rho(x, h^r(s), t) \tag{30}$$

and the normalization conditions

$$\begin{aligned} &\overline{p}(x, h^r(\infty), t) = 0 \\ &\int_0^\infty \mathrm{d}s\, \overline{p}(x, h^r(s), t) = p(x) \end{aligned} \tag{31}$$

where $p(x)$ is the probability to belong to the sublattice x.

Equations (27)–(30) are the main result of this section. Let us discuss the various quantities in these equations. Instead of equation (27) we can write $m_\mu = 2\gamma_r \sum_x x^\mu p(x) A(x,t)$ with the sublattice activity

$$A(x,t) = \frac{1}{p(x)} \int_0^\infty \mathrm{d}s\, \bar{p}(x, h^r(s), t) \rho(x, h^r(s), t). \tag{32}$$

Comparing this with equation (30) we find $p(x, h^r(0), t) = p(x)A(x,t)$. Let us analyse the meaning of equation (32). The expression $\bar{p}(x, h^r(s), t)\mathrm{d}s$ in the integrand of equation (32) is the probability of a neuron to belong to the sublattice x *and* to experience at time t a refractory field h^r with $h^r(s) \leqslant h^r \leqslant h^r(s + \mathrm{d}s)$. The factor $\rho(x, h^r(s), t)$ in the integrand is the firing rate constant, equation (1), of such a neuron, i.e.

$$\rho(x, h^r(s), t) = (1/\tau_0) \exp\{\beta[h^s(t) + h^r(s)]\}. \tag{33}$$

Here we have set $\theta = 0$. Thus the quantity $A(x,t)$ is the *ensemble averaged activity* at time t of the neurons in sublattice x independent of the refractory field $h^r(s)$ the neurons had before firing. Or, to phrase it differently, $A(x,t)\mathrm{d}t$ gives the number of neurons in sublattice x which emit a spike during the time between t and $t + \mathrm{d}t$ divided by the total number of neurons in this sublattice.

The neurons which are inactive at time t give no explicit contribution and are coupled only via the variable $h(x, h^r(s), t)$ in (28) which consists of two parts. The first one stems from the synaptic connections to other neurons and is a sum over the activities at earlier times, weighted with the response function $\bar{\epsilon}^s(\tau)$, i.e. $h^s(t) = \sum_{\mu=1}^q x^\mu \int \mathrm{d}\tau\, \bar{\epsilon}^s(\tau) m_\mu(t - \tau)$, the second is the refractory field. The refractory term h^r takes care of the spiking history of the neuron itself, while h^s includes the spiking history of all other neurons.

A major problem, of course, is to find an explicit form of the $\bar{p}(x, h^r(s), t)$. As a first step, we integrate equation (29) yielding

$$\bar{p}(x, h^r(s), t) = \bar{p}(x, h^r(0), t - s) \exp\left[\int_0^s \rho(x, h^r(s'), t - s')\, \mathrm{d}s'\right] \tag{34}$$

which suggests the following interpretation. A neuron of sublattice x with a refractory field $h^r(s)$ at time t must have spiked earlier at time $(t - s)$ (first factor of product) *and* may not have spiked since then (second factor).

3.4. Stationary solutions

The set of integral equations (27), (28), (30) and (34) provides a complete description of the dynamic evolution of a network starting from some initial conditions. An attempt to solve these equations analytically, however, seems rather infeasible. But with some additional assumptions regarding the type of behaviour that might be expected, it is possible to find special solutions.

First, let us asume that there exists a *stationary* solution to these equations. By this we mean that all macroscopic quantities, i.e. the order parameters $m_\mu(t)$ and the sublattice activities $A(x,t)$, take constant values. Since the m_μ involve an average over many neurons, such constant solutions might be possible, despite the fact that

every single neuron has an internal spiking dynamics. In fact, we expect that a solution which retrieves the (stationary) pattern $\{\xi_i^\mu\}$ has a large and constant overlap m_μ with this pattern.

If the overlaps are constant, $m_\mu(t) \equiv m_\mu$, then the local field in equation (28) reduces to

$$h(x, h^r(s), t) = \sum_{\mu=1}^{q} x^\mu m_\mu + h^r(s) = h^s + h^r(s) \tag{35}$$

with no explicit dependence upon t. The same is true for the rate constant

$$\rho(x, h^r(s), t) = (1/\tau_0)\exp\{-\beta[h^s + h^r(s)]\} = \rho(x, h^r(s)). \tag{36}$$

We can thus drop the explicit t-dependence in equations (31) and (34) which yields

$$p(x) = \overline{p}(x, h^r 0)\int_0^\infty \mathrm{d}s \exp\left[-\int_0^s \mathrm{d}s' \rho(x, h^r(s'))\right] \tag{37}$$

and with (30) and (34)

$$A(x) = \left\{\int_0^\infty \mathrm{d}s \exp\left[-\int_0^s \rho(x, h^r(s')\,\mathrm{d}s'\right]\right\}^{-1}. \tag{38}$$

This might still look rather complicated, but if we compare the final expression with equations (5) and (8) in section 2.2, we realize that the term in braces is just the mean interval length of the spike train of a neuron in sublattice x. Thus the activity $A(x)$ is equal to the *mean firing rate* of the neurons in sublattice x. At the same time $A(x)$, as defined by equation (32), is directly related to the overlaps m_μ, equation (27), which characterize the macroscopic states of the network. We can therefore express the stationary overlap m_μ in terms of the mean firing rate $f(h^s)$ at some input level $h^s(x)$. And if we know the mean firing rate as a function of the input current h^s, i.e. the gain function $f(h^s)$ of the neurons, then we can immediately find all the stationary solutions of the network equations (27)–(30). In particular, the explicit dependence on the refractory state h^r has been eliminated. This does not imply, however, that the internal state variable has disappeared altogether, but, as we have seen in section 2, the refractory function $\epsilon^r(\tau)$ is implicitly present in the shape of the gain function $f(h^s)$.

3.5. Examples of stationary solutions

In summary, we have reduced the task of finding the stationary solutions to a set of simple equations

$$\begin{aligned} m_\mu &= 2\gamma_r \sum_x p(x) x^\mu f[h(x)] \\ h(x) &= \sum_{\mu=1}^{q} x^\mu m_\mu \end{aligned} \tag{39}$$

where $f(h)$ is the gain function of the neurons and $h(x) = h^s$ is the synaptic input of neurons in sublattice x.

This result is much more general and independent of the specific model of the neuron that is used (Amit and Tsodyks 1991, Gerstner and van Hemmen 1992). It is based on the high connectivity of the network which allows application of the law of large numbers. It implies that all traces of stochastic spiking are wiped out and the gain function of the neuron is the only object that matters. Thus we are led to a description by 'graded-response neurons' which are often used as the basic units in models of neural networks (Hopfield 1984). Our derivation of equation (39) shows that it is in fact possible to justify such an ansatz for the *stationary solutions* of a highly connected network. In section 3.6 we will show that in the case of oscillatory solutions the *spiking* of the neurons becomes important—a description by mean firing rates that is a condition for graded-response models is under these circumstances not realistic. We now turn to some specific examples.

3.5.1. Model with absolute refractory period. As a simple, but instructive example, let us consider a model with an absolute refractory period only, i.e. $\epsilon^r(\tau) = -\infty$ for $\tau < \gamma_r$ and 0 otherwise. Using equation (8) of section 2.2 we find the gain function

$$f(h) = \exp(\beta h)/[\tau_0 + \gamma_r \exp(\beta h)]. \tag{40}$$

Assuming that there is overlap with one pattern only, i.e. $m_\mu = m\delta_{\mu\nu}$, we find

$$m = \sinh h(\beta m)/[\alpha + \cosh(\beta m)] \tag{41}$$

where we have introduced the parameter $\alpha = \frac{1}{2}[(\gamma_r/\tau_0) + (\tau_0/\gamma_r)]$. In figure 7 we have plotted the function $G(h) = \sinh h/[\alpha + \cosh h]$ for different values of the parameter α. The intersection of $G(h)$ with a straight line of slope $1/\beta$ yields the stationary overlap m. In figure 8 we show m as a function of $T = \beta^{-1}$. For parameters $\alpha < 2$ (figure 8(*a*)), the overlap decreases continuously with temperature and reaches 0 at a critical noise level $T_c = (1 + \alpha)^{-1}$. The trivial state $m = 0$ is unstable for $T < T_c$ and stable for $T > T_c$. We thus have a continuous phase transition similar to the one that has been found in the Hopfield network (Hopfield 1982, Amit *et al* 1985) to which the system formally reduces in the limit $\alpha \to 0$.

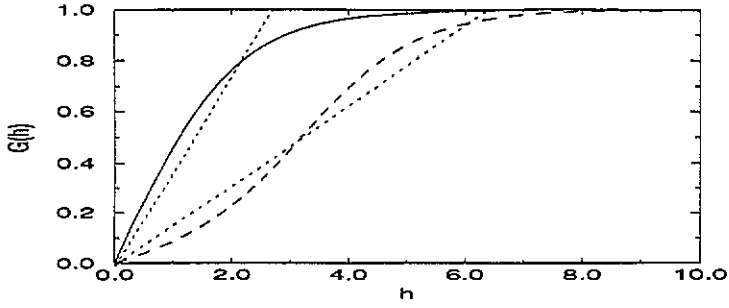

Figure 7. Model with absolute refractory period. The function $G(h) = \sinh h/[\alpha + \cosh h]$ has been plotted as a function of the field h for two different values of a α ($\alpha = 1$, full curve; $\alpha = 12$, broken curve). The intersection of $G(h)$ with a straight line of slope $1/\beta$ (dotted curves) yields the stationary overlap $|m|$. The critical temperature is lower for large α and an additional (unstable) solution appears.

The behaviour, however, is completely different for parameter values $\alpha > 2$. In this case (figure 8(*b*), (*c*)) a regime of bistability occurs in which both the trivial

state and the retrieval state are stable *stationary* states of the system. (This does not exclude that these states might be unstable with respect to oscillatory solutions, see section 3.6.) The bistability has two effects on the behaviour of the system. To discuss the first one, we keep the noise level β fixed and assume that the system starts with some initial overlap. Depending on whether the overlap is larger or smaller than the overlap m' of the unstable fixed point, the system either reconstructs the pattern and goes to the retrieval state, or it reduces the overlap and goes to the trivial state. Thus, the unstable fixed point acts as a threshold value for the retrieval of the pattern (figure 8(*c*)).

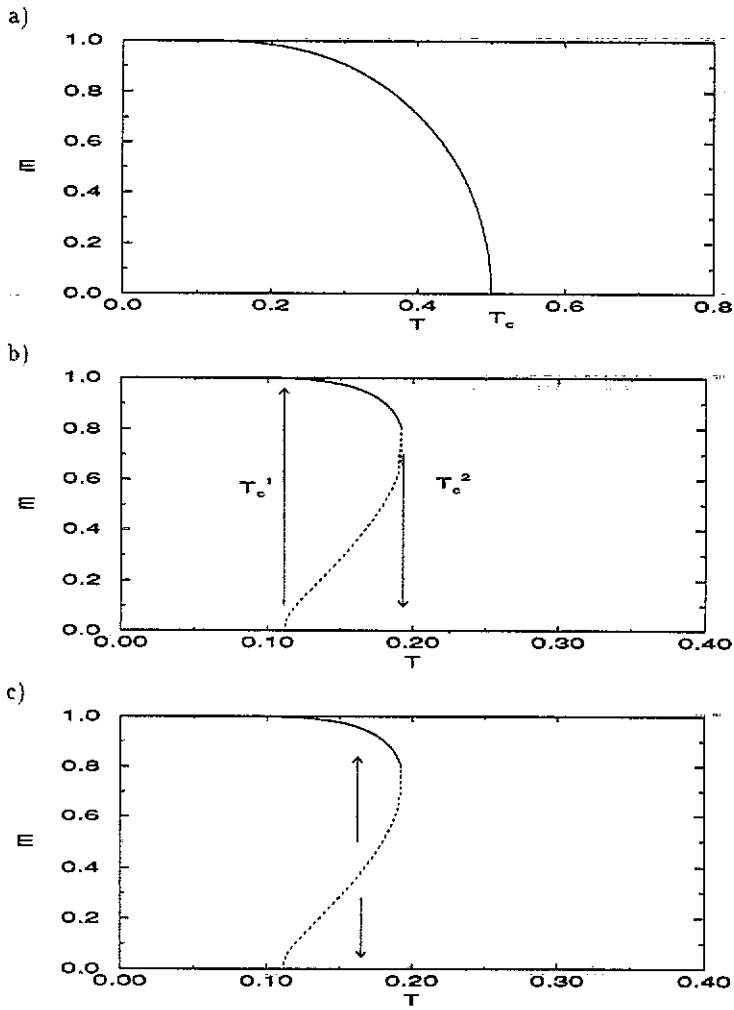

Figure 8. Overlap for a model with absolute refractory period. The overlap m is given as a function of temperature T for two values of the parameter α, $\alpha = 1$ (*a*) and $\alpha = 8$ (*b*) and (*c*), cf figure 7. For $\alpha = 1$ we have a continuous phase transition with a critical temperature $T_c = 0.5$. For $\alpha = 8$ we have a regime of bistability between $T_c^1 = 1/9$ and T_c^2 leading to a hysteresis behaviour, if the temperature is varied (*b*). At fixed T the unstable solution acts as a threshold for retrieval (*c*).

The second effect is a hysteresis behaviour as shown in figure 8(*b*). If we start at

a high noise temperature and 'cool' the system, the overlap changes discontinuously from $m = 0$ to $m \approx 1$ at the lower critical temperature $T_c^{(1)} = 1/(1+\alpha)$. On the other hand, if we 'heat' the network starting at zero temperature, the network stays in the retrieval state until the upper critical temperature $T_c^{(2)}$ is reached where it jumps discontinuously into the trivial state. In terms of thermodynamics this behaviour indicates a first-order phase transition. Thus, a network of neurons with large α, i.e. very long or extremely short refractory period compared to the intrinsic time constant τ_0, belongs to a different universality class than the Hopfield network.

Note that in the noiseless case, i.e. $T = 0$, the overlap m is equal to 1 and all neurons fire with maximum frequency. This effect, however, is an artifact of the model with absolute refractoriness which in the noiseless case has a step-like gain function. If we include a relative refractory period, then the gain function is smoother and solutions with low firing rates are possible—even in the noiseless case.

3.5.2. Model with relative refractory period. We now want to generalize the above arguments and include relative refractoriness. To keep the formulas simple we take a refractory function of the form

$$\epsilon^r(\tau) = \begin{cases} -\infty & \text{for } 0 < \tau \leqslant \gamma_r \\ -\epsilon_0/(\tau - \gamma_r) & \text{for } \tau > \gamma_r. \end{cases} \tag{42}$$

In the noiseless case, the gain function of this model is given by

$$f(h) = \begin{cases} 0 & \text{for } h < 0 \\ h/(\epsilon_0 + \gamma_r h) & \text{for } h \geqslant 0. \end{cases} \tag{43}$$

As before we assume that we have a macroscopic overlap with one pattern only, i.e. $m_\mu = m\delta_{\mu\nu}$. Using (39) and introducing the parameter $c = \epsilon_0/\gamma_r$ we find

$$|m| = |m|/(c + |m|) \tag{44}$$

with the solutions $m = 0$ and $|m| = 1 - c$.

Two things are worth noticing. First, non-trivial solutions exist only for $c \leqslant 1$. This corresponds to the well known result for graded-response neurons that the slope of the gain function must exceed one at the origin (Hopfield 1984). In terms of our model, it means that the relative refractory function must decay fast compared to the length of the absolute refractory period. Second, for $0 < c < 1$ solutions with low spiking rates are possible. This clarifies an old problem of associative neural networks. An extensive discussion of the 'problem of low spiking rates' has been presented by Amit and Tsodyks (1991) who use an argument based on the 'mean firing rate' for its solution. To see that there are solutions with low firing rates in our model, we notice that $|m|$ gives the firing rate of neurons that are active in the retrieval state in units of the maximum firing rate. Thus, adjusting c, it is possible to construct solutions with arbitrarily low spiking rates, at least in the noise-free case (figure 9(*a*)).

What happens if we add noise? In this case, the mean firing rate can be found by numerical integration of equation (8). The dependence of the mean firing rate upon the input current h yields the gain function $f(h)$ which determines via equation (39) the overlap m. In figure 9(*b*) we have plotted the function $\gamma_r[f(h) - f(-h)]$ for a neuron with $c = 0.25$ at various noise levels. The intersection with a straight line of unit slope yields the retrieval overlap. If we add more noise, the intersection point moves continuously down to zero. Thus we have a continuous phase transition and retrieval solutions with low spiking rates are possible.

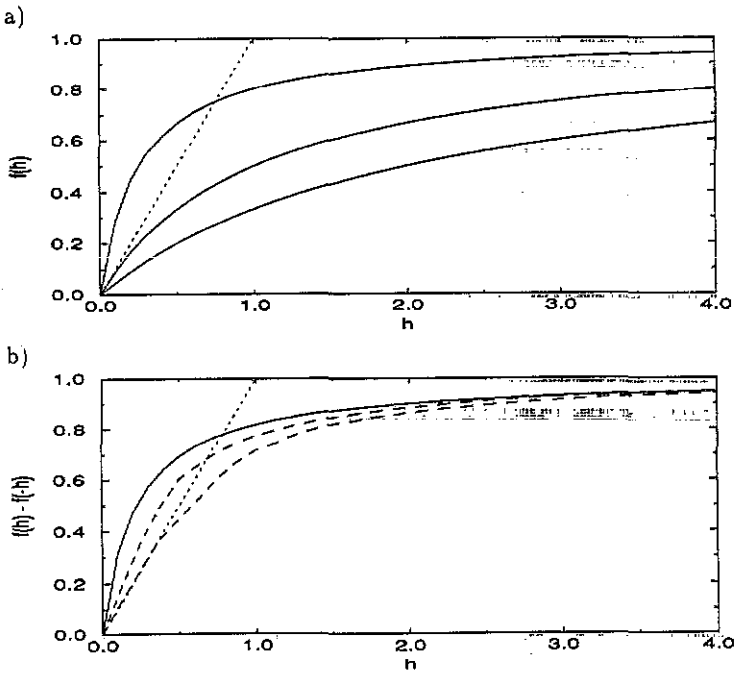

Figure 9. Model with relative refractory period. (*a*) Gain function. The intersection of the gain function $f(h)$ (solid curves) with a straight line of slope 1 (broken) yields the stationary solutions in the limit $\beta \to \infty$. For models with a long refractory period, $c > 1$, the gain function has low gain and retrieval is not possible (lower full curve, $c = 2$). For $0 \leqslant c < 1$ retrieval solution with overlap $|m| = 1 - c$ exist which allows for retrieval at low spiking rates (upper full curve, $c = 0.25$). The critical value is $c = 1$ (middle full curve). (*b*) In the case of finite β, we have to consider the function $f(h) - f(-h)$ which we have plotted for $\beta = 8$ and $\beta = 4$ (broken curves) with the parameter $c = 0.25$. The intersection of this function with a line of unit slope (dotted) yields the retrieval overlap $|m|$ which is smaller than in the noise-free case (full curve).

3.6. Oscillatory solutions

So far we have been concerned with stationary solutions only. In this case it has been possible to reduce the system to an equivalent network of graded-response neurons, equation (39), where the gain function of the neurons is the only object that matters. This is due to the fact that for constant solutions the mean firing rate which is defined by a time-averaging procedure becomes equal to the activity A which involves an ensemble average over all neurons of a given sublattice. In the case of time-dependent solutions, however, this is no longer true and we have to deal with the full set of equations for the ensemble, equations (27)–(31).

To get some insight into the type of solutions that might be possible, we consider the noise-free case, i.e. $\beta \to \infty$. In this case the rate constant ρ, cf equations (1) and (33), reduces to a strict threshold condition and the neurons fire immediately, if $h(t) = h^r(t) + h^s(t) > 0$. We could take the set of equations in the limit $\beta \to \infty$ and try to derive conditions for the stability of the stationary solutions that we have found and discussed in the preceding sections. Instabilities could lead to time-structured solutions, in particular to collective oscillations of the activity. Instead of this, we take the opposite approach. We assume that there *are* coherent oscillations and we

discuss the self-consistency of such an ansatz. To be specific, we assume that all neurons that should be active during the retrieval of pattern μ, i.e. neurons with $x^\mu = 1$, have fired coherently at times $t = 0, -T, -2T, \ldots$ while the 'off'-neurons with $x^\mu = -1$ stay quiescent. With these assumptions we can calculate the period T. For the 'on'-neurons, the activity A in equation (32) reduces to a sequence of δ-functions

$$A(x^\mu = 1, t) = \sum_{n=0}^{n=\infty} \delta(t + nT) \tag{45}$$

which yields for the overlap

$$m_\mu(t) = m(t) = \gamma_r \sum_{n=0}^{n=\infty} \delta(t + nT). \tag{46}$$

Using this in equation (28), we find the synaptic contribution

$$\begin{aligned} h^s(x^\mu, t) &= \gamma_r x^\mu \int \mathrm{d}\tau\, \bar{\epsilon}^s(\tau) \sum_{n=-\infty}^{n=\infty} \delta(t + nT - \tau) \\ &= \gamma_r x^\mu \sum_{n=0}^{\infty} \bar{\epsilon}^s(t + nT). \end{aligned} \tag{47}$$

Similarly we find for the refractory field of the 'on'-neurons (equation (4))

$$h^r(t) = \sum_{n=0}^{\infty} \epsilon^r(t + nT). \tag{48}$$

If the system approaches $t = T$, the threshold condition $h^r(T) + h^s(T) = 0$ must be fulfilled to support the coherent spiking cycle of period T. Evaluation for $m = 1$ yields for neurons with $x^\mu = 1$,

$$\sum_{n=1}^{\infty} \epsilon^r(nT) = \gamma_r \sum_{n=1}^{\infty} \bar{\epsilon}^s(nT). \tag{49}$$

Neurons with $x^\mu = -1$ should stay quiescent, which requires

$$\sum_{n=0}^{\infty} \bar{\epsilon}^s(t + nT) \geqslant 0 \qquad \text{for all } t \tag{50}$$

which is trivially fulfilled if $\bar{\epsilon}^s(\tau) \geqslant 0$ for all τ.

Equation (49) yields the period T. If furthermore $\sum_{n=2}^{\infty} \bar{\epsilon}^s(nT) \ll \bar{\epsilon}^s(T)$, and $\sum_{n=2}^{\infty} \epsilon^r(nT) \ll \epsilon^r(T)$, then T can be found by a simple graphical procedure. In figure (9) we have plotted $\gamma_r \bar{\epsilon}^s(\tau)$ and also $(-\epsilon^r(\tau))$ as a function of the time τ after the last spike. The intersection of the two functions yields the period T of a coherent oscillation. By a slightly more involved argument, it can be shown

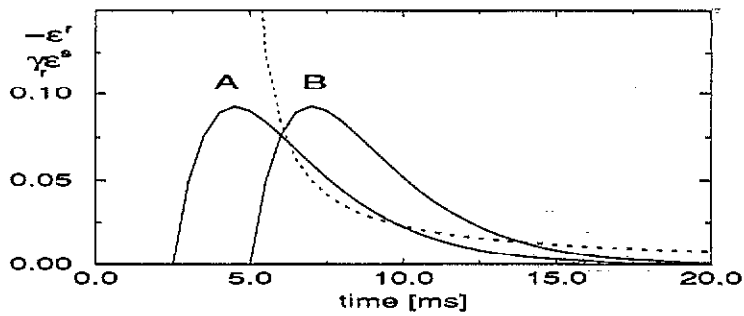

Figure 10. Network oscillations. The negative of the refractory function $-\epsilon^r(\tau)$ (dotted curve) and the synaptic response $\gamma_r\epsilon^s(\tau)$ (full curves) have been plotted as a function of the time τ after a coherent spiking event in the network at $\tau = 0$. We consider two different axonal delay times, short delays (A, $\Delta_a = 2.5$ ms) and long delays (B, $\Delta_a = 5$ ms). The intersection point $\gamma_r\epsilon^s(\tau) + \epsilon^r(\tau) = 0$ yields the oscillation period T. Note that in both models the oscillation period is approximately the same, but the oscillation is stable only if the slope of ϵ^s is positive at the intersection point, as in B.

analytically (Ritz 1991) that the collective oscillation is stable only if the slope of the synaptic field

$$\frac{\partial}{\partial t}h^s|T > 0 \tag{51}$$

which in the case of a quickly decaying refractory function reduces to

$$\frac{\partial}{\partial t}\bar{\epsilon}^s|T > 0.$$

Thus we can predict from figure (10) that in a model with short axonal delays (network A) oscillations are unstable and the system returns to a stationary state, while a long axonal delay time supports synchronized spiking (network B).

3.7. Examples of oscillatory solutions

We have simulated a network of neurons with a refractory function

$$\epsilon^r(\tau) = \begin{cases} -\infty & \text{for } 0 < \tau \leqslant \gamma_r \\ -\epsilon_0/(\tau - \gamma_r) & \text{for } \tau > \gamma_r \end{cases} \tag{52}$$

with the parameters $\gamma_r = 4.5$ ms and $\epsilon_0 = 1$ and a synaptic response function

$$\epsilon^s(\tau) = \begin{cases} 0 & \text{for } \tau < \Delta_a \\ (\tau - \Delta_a)/(\tau_s)^2]\exp[-(\tau - \Delta_a)/T_s] & \text{for } \tau \geqslant \Delta_a \end{cases} \tag{53}$$

with a response time $\tau_s = 2$ ms. The axonal delay time Δ_a is then the only free parameter and we have considered two different cases, i.e. a short delay $\Delta_a = 2.5$ ms (network A) and a longer delay $\Delta_a = 5$ ms (network B). Both functions are plotted in figure 10. The simulation results for a network of 900 neurons are shown in figure 11 where we have plotted the instantaneous overlap $m(t)$ and the postsynaptic field $h^s(t)$ as a function of time. In the simulations the time-step Δt was taken to

be 0.5 ms. In the low-noise limit, the network A with short delays approaches a stationary state, whereas the network with long delays (B) is driven into a coherent oscillation. Both the oscillation period of 6 ± 0.5 ms and the stability of the solution can be predicted from the argument of the preceding section. The magnitude of the stationary overlap $|m| = 0.77$ of network B is consistent with the calculations in section 3.5. If we add noise, the perfect synchronicity of the spiking is destroyed and the oscillation amplitude is smaller (figure 12). At high noise level ($\beta = 2$) retrieval is no longer possible. Thus noise is an important parameter that determines not only the magnitude of the retrieval overlap, but also the time structure of the solutions.

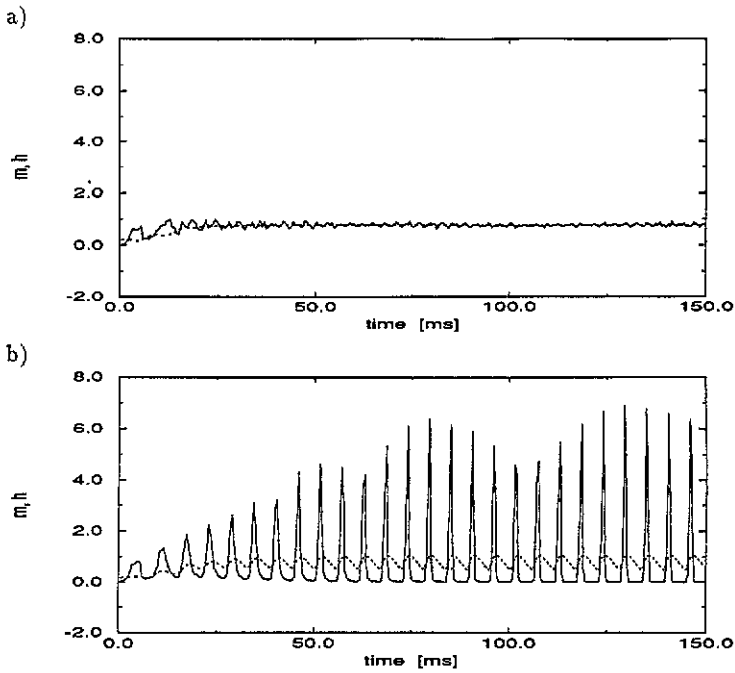

Figure 11. Dependence upon the axonal delay Δ_a. The overlap $m(t)$ (full curves) and the postsynaptic field $h^s(t)$ (dotted curves) have been plotted as a function of time. The network A with short delays ($\Delta_a = 2.5$ ms) settles into a stationary state (*a*), while system B with long delays ($\Delta_a = 5$ ms) goes into a state of synchronized spiking with period $T \approx 6$ ms (*b*). Note that the spiking in B occurs when $\partial/\partial t h^s(t) > 0$, which is a condition for stable oscillations (noise $\beta = 8$); cf figure 10.

4. Discussion and conclusions

The present paper leads to four conclusions. First, we have presented a general model of a neuron based on the notions of *firing threshold, refractory period* and *noise*. This class of model neurons contains a variety of different neuron types including randomly or regularly firing neurons, oscillating or bursting neurons, all characterized by their specific refractory function. The simplest refractory function is the absolute refractory period, a model most convenient for formal considerations.

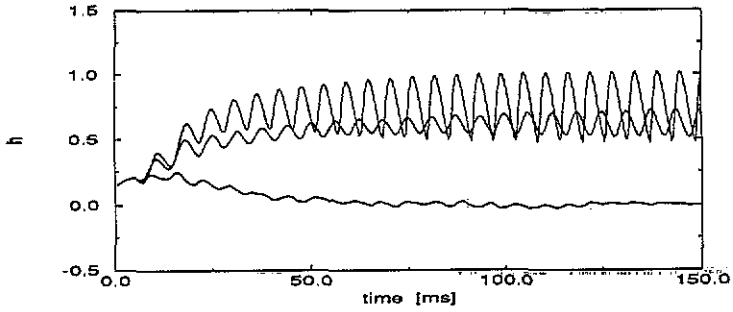

Figure 12. Influence of noise. The postsynaptic field $h^s(t)$ of the oscillatory system B (cf figure 11) has been plotted as a function of time at different noise levels ($\beta = 6$, top; $\beta = 4$, middle; $\beta = 2$, bottom). Noise destroys the synchronicity of spiking and thus lowers the oscillation amplitude of the synaptic field. At high noise level ($\beta = 2$) retrieval is no longer possible.

Our model neurons are biological in the sense that they produce *spike trains* comparable to those in real neurons, i.e. realistic spiking rates as well as realistic interval distributions. It is our assumption that all the information of neuronal signals is contained in these spike trains. A more detailed model of a neuron is thus unnecessary, if the problem is posed in the context of information storage and processing in neural systems. The main ideas of our model are certainly not new; see, for example, Buhmann and Schulten (1986), Choi (1988), Amit (1989) and Horn and Usher (1989), who all include refractoriness and neuronal spiking into formal association networks. Our unified approach, however, helps to understand the variety of neural spiking phenomena in terms of a single refractory function, thus making neuron spiking easy to program and amenable to mathematical analysis.

Second, we have constructed a network consisting of these 'spiking' neurons, each pair of neurons being connected by synapses with a realistically delayed response function. The network had to learn a finite number of random patterns using a Hebbian learning rule. We have derived exact equations of motion in terms of overlaps with the learnt patterns. For *stationary* states we have shown that the model can be reduced to a network of graded-response neurons and the usual analysis can then be applied to find the retrieval solutions. Thus, in a stationary state, the *mean firing rate* is the only object that matters. The statistics of the firing noise and the internal dynamics of the neurons is completely wiped out.

The third conclusion is, however, that this is not true in an *oscillatory* state. In the case of collective oscillations, the internal spiking dynamics of the neurons as modelled by the refractory function $\epsilon^r(\tau)$ becomes an important parameter of the problem, and so does the synaptic transmission which is characterized by the axonal delay Δ_a and the synaptic response time τ_s. The relation between these quantities determines the existence and stability of oscillatory retrieval solutions. Furthermore, the randomness introduced by the spiking noise β may not only lower the retrieval overlap but also destroy the synchronicity of the spiking. Thus, a more detailed description of the spiking dynamics and synaptic connections allows for additional effects which are beyond the scope of an approach based on formal two-state or graded-response neurons.

Finally, of course there still remain a lot of unbiological features in the model. Most of these, however, can be eliminated—at least in principle. One of the most

strikingly unbiological properties is the assumption of full connectivity. This assumption can be dropped and we could consider a randomly diluted high-connectivity network instead. In passing we note that we could also have taken an asymmetric low-connectivity network (Derrida *et al* 1987). The main difference to the present work is that the right-hand side of Equations (27), (28) has to be averaged over a Gaussian distribution. The nature of the solutions remains the same. Another objection concerns the fact that we have neglected the effect of the inhibitory neurons. Inhibition can, however, be added in a straightforward manner (Ritz 1991, van Hemmen *et al* 1992). In this case, collective oscillations with a frequency of approximately 50 Hz appear which are in agreement with experimental observations in the visual cortex of the cat (Eckhorn *et al* 1988, Gray and Singer 1989) and which may be important in the context of feature linking and pattern segmentation (von der Malsburg and Schneider 1986, Eckhorn *et al* 1988, Wang *et al* 1990).

The use of random patterns is a further unbiological feature of our model, but there are standard ways to overcome these problems (Amit *et al* 1987, Vicente and Amit 1988, Buhmann *et al* 1989, for an overview see Amit 1989). Another set of questions is connected with the use of the specific synaptic connection matrix and the 'Hebbian' learning rule. But many of the questions concerning synaptic learning itself are unsolved. The importance and mechanisms of long-term potentiation (LTP) and long-term depression (LTD) are still open problems, experimental results are scarce and their interpretation is difficult (Kelso *et al* 1986, Brown *et al* 1989, Lisman 1989, Stanton and Sejnowski 1989). Thus it is often not even clear so far, what the 'biological' features really are.

Acknowledgments

WG would like to thank William Bialek and his students at Berkeley for their generous hospitality and numerous stimulating discussions. Thanks are also due to Andreas Herz for many helpful comments and advice and to Raphael Ritz for calculations on the Hodgkin–Huxley model. We are grateful to the referees for their comments on the paper which helped to improve it substantially. W G acknowledges the financial support of the German Academic Exchange Service (DAAD) for making possible the stay at Berkeley.

References

Abbott L F and Kepler T B 1990 Model neurons: from Hodgkin–Huxley to Hopfield *Statistical Mechanics of Neural Networks (Lecture Notes in Physics 368)* ed L Garrido (Berlin: Springer) pp 5–18

Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)

Amit D J, Gutfreund H and Sompolinsky H 1985 Spin-glass models of neural networks *Phys. Rev.* A **32** 1007–32

—— 1987 Information storage in networks with low levels of activity *Phys. Rev.* A **35** 2293–303

Amit D J and Tsodyks M V 1991 Quantitative study of attractor neural networks retrieving at low spike rates: I. Substrate—spikes rates and neuronal gain *Network* **2** 259–73

Brown T H, Ganong A H, Kairiss, E W, Keenan C L and Kelso S R 1989 Long-term potentation in two synaptic systems of the hippocampal brain slice *Neural Models of Plasticity* ed J H Byrne and W O Berry (San Diego, CA: Academic) pp 266–306

Buhmann J and Schulten K 1986 Associative recognition and storage in a model network with physiological neurons *Biol. Cybern.* **54** 319–35

Buhmann J, Divko R and Schulten K 1989 Associative memory with high information content *Phys. Rev.* A **39** 2689–92

Choi M Y 1988 Dynamic model of neural networks *Phys. Rev. Lett.* **61** 2809–12

Derrida B, Gardner E and Zippelius A 1987 An exactly soluble asymmetric neural network model *Europhys. Lett.* **4** 167

Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M and Reitboeck H J 1988 Coherent oscillations: a mechanism of feature linking in the visual cortex? *Biol. Cybern.* **60** 121–30

FitzHugh R 1961 Impulses and physiological states in theoretical models of nerve membranes *Biophys. J.* **1** 445

Gerstner W and van Hemmen J L 1992 Universality in neural networks: the importance of the mean firing rate *Biol. Cybern.* **67** in press

Gray C M and Singer W 1989 Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex *Proc. Natl Acad. Sci. USA* **86** 1698–702

Herz A, Sulzer B, Kühn R and van Hemmen J L 1988 The Hebb rule: representation of static and dynamic objects in neural nets *Europhys. Lett.* **7** 663–9

—— 1989 Hebbian learning reconsidered: representation of static and dynamic objects in associative neural nets *Biol. Cybern.* **60** 457–67

Hodgkin A L and Huxley A F 1952 A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes *J. Physiol.* **117** 500–44

Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8

—— 1984 Neurons with graded response have computaional properties like those of two-state neurons *Proc. Natl Acad. Sci. USA* **81** 3088–92

Horn D and Usher M 1989 Neural networks with dynamical thresholds *Phys. Rev.* A **40** 1036–40

Kelso S R, Ganong A H and Brown T H 1986 Hebbian synapses in hippocampus *Proc. Natl Acad. Sci. USA* **83** 5326–30

Lisman J 1989 A mechanism for Hebb and anti-Hebb processes underlying learning and memory *Proc. Natl Acad. Sci. USA* **86** 9574–8

Nagumo J, Arimoto S and Yoshizawa S 1962 An active pulse transmission line simulating nerve axon *Proc. IRE* **50** 2061–70

Perkel D H, Gerstein G L and Moore G P 1967 Neuronal spike trains and stochastic point processes I. The single spike train *Biophys. J.* **7** 391–418

Pfeiffer R R and Kiang Y S 1965 Spike discharge patterns of spontaneous and continuously stimulated activity in the cochlear nucleus of anesthetized cats *Biophys. J.* **5** 301–16

Riedel U, Kühn R and van Hemmen J L 1988 Temporal sequences and chaos in neural nets *Phys. Rev.* A **38** 1105–8

Ritz R 1991 Kollektive Oszillationen in Neuronalen Netzwerken *Diplomarbeit* Technischen Universität München

Stanton P K and Sejnowski T J 1989 Associative long-term depression in the hippocampus induced by Hebbian covariance *Nature* **339** 215–8

van Hemmen J L and Kühn R 1986 Nonlinear neural networks *Phys. Rev. Lett.* **57** 913–6

van Hemmen J L, Gerstner W and Ritz R 1992 A microscopic model of collective oscillations in the cortex *Complex Dynamics in Neural Networks* ed J S Taylor (Berlin: Springer) in press

van Hemmen J L, Grensing D, Huber A and Kühn R 1986 Elementary solution of classical spin glass models *Z. Phys.* B **65** 53–63

—— 1988 Nonlinear neural networks I,II *J. Stat. Phys.* **50** 231–93

von der Malsburg C and Schneider W 1986 A neural cocktail-party processor *Biol. Cybern.* **54** 29–40

Vicente C J P and Amit D J 1988 Optimised network for sparsely coded patterns *J. Phys. A: Math. Gen.* **22** 559–69

Wang D, Buhmann J and von der Malsburg C 1990 Pattern segmentation in associative memory *Neural Comp.* **2** 94–106