PERGAMON

Neural Networks 14 (2001) 941–953

Neural Networks

www.elsevier.com/locate/neunet

2001 Special issue

From artificial neural networks to spiking neuron populations and back again

Marc de Kamps*, Frank van der Velde

Unit of Experimental and Theoretical Psyhology, Leiden University, Wassenaarsewag 52, 2333 AK Leiden, The Netherlands

Received 10 October 2000; revised 4 April 2001; accepted 4 April 2001

Abstract

In this paper, we investigate the relation between Artificial Neural Networks (ANNs) and networks of populations of spiking neurons. The activity of an artificial neuron is usually interpreted as the firing rate of a neuron or neuron population. Using a model of the visual cortex, we will show that this interpretation runs into serious difficulties. We propose to interpret the activity of an artificial neuron as the steady state of a cross-inhibitory circuit, in which one population codes for 'positive' artificial neuron activity and another for 'negative' activity. We will show that with this interpretation it is possible, under certain circumstances, to transform conventional ANNs (e.g. trained with 'back-propagation') into biologically plausible networks of spiking populations. However, in general, the use of biologically motivated spike response functions introduces artificial neurons that behave differently from the ones used in the classical ANN paradigm. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Negative activation; Perceptrons; Population; Rate problems; Rate coding; Spiking neurons; Visual cortex

1. Introduction

Artificial Neural Networks (ANNs) have been used widely to model higher level cognitive functions, such as object recognition (Fukushima, 1988; Riesenhuber & Poggio, 1999), visual attention (van der Velde & de Kamps, 2001), processing of spatial information (Pouget & Snyder, 2000), to name but a few. These networks are the method of choice whenever insufficient data are available to describe the model directly in terms of neurophysiology, or when such a description would simply be too complicated, e.g. when many cortical areas are involved.

A particularly important class are the so-called multi-layer feed-forward perceptron networks, which are known to be Universal Approximators, in the sense that they can approximate a very large class of continuous functions (Hornik, Stinchcombe & White, 1989). Given the importance of these networks for models of higher cognitive functions, it is an important question as to how they could be implemented by spiking neurons in the cortex.

At present, this interpretation is a controversial issue. Some simply consider multi-layer feed-forward perceptron networks to be biologically implausible (Hertz, Krogh & Palmer, 1991; Hoffman, 1998). Others suggest that the activation of a perceptron corresponds to the average firing rate of a spiking neuron or group of neurons (Rolls & Treves, 1998).

In order to investigate the issue, it is important to realize that a model of feed-forward cortical networks has to meet a number of requirements, in order to be biologically plausible. First of all, it has to account for the processing speed which has been observed experimentally in such feed-forward networks. In the visual cortex for instance, object recognition can take place within 100–150 ms (Chelazzi, Miller, Duncan & Desimone, 1993; Thorpe, Fize & Marlot, 1996). Secondly, the average firing rates in the network must be low. In vivo measurements of firing rates in visual cortex using micro-electrodes (Chelazzi et al., 1993; Motter, 1994) show that these rates are typically below 40 Hz, except for transient activity. Thirdly, one expects only base rate activity in neurons that do not have an input stimulus within their receptive field, and a perceptron network must reflect this fact.

The first two requirements exclude the interpretation of perceptron activation as an average firing rate of individual neurons. It has been pointed out (Maass, 1997) that in order to carry out object recognition within 100 ms, several cortical areas are involved, each of which only has 20–30 ms to complete its calculation. However, if firing rates are well

* Corresponding author. Tel.: +31-71-5273631; fax: +31-71-5273783.
E-mail address: kamps@fsw.leidenuniv.nl (M. de Kamps).

0893-6080/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0893-6080(01)00068-5

below 100 Hz, 20–30 ms would already be required just to sample the firing rate of the neurons in that area.

To overcome this problem, it has been suggested (Maass, 1997) that perceptron activation is coded by inter-spike intervals and it has been demonstrated that the networks using this code are able to perform calculations very rapidly. There are, however, several reasons, some of which we will discuss below, why we believe that this coding is an unlikely candidate for multi-layer feed-forward cortical networks.

Another way to overcome the seemingly contradictory requirements of low rates and fast response to input, is to consider the perceptron's activation to be a population rate, rather than the firing rate of individual neurons. Population rates are defined as the fraction of neurons in a population that fire during a time interval, divided by that interval. It has been pointed out by many (Gerstner, 2000, and references therein) that populations can respond very quickly to input indeed.

In this paper, we will adopt the (population) rate coding hypothesis, but we will demonstrate that with a naive interpretation of this hypothesis, the second and third requirements for biological plausibility, discussed above, are violated. First of all, perceptron activations would be much too high to account for realistic firing rates, as they are observed in the cortex. Also, spurious activation, i.e. activation that is not related to the input stimulus, would occur in higher layers of the network. We will show that these rate problems are related, and that in order to avoid them one must choose a squashing function for the perceptron that maps zero input on zero output.

We will introduce a simple cross-inhibitory circuit of neuron populations to implement these squashing functions in the cortex and we will use Wilson–Cowan dynamics (Wilson & Cowan, 1972) to describe the circuit. Wilson–Cowan dynamics give the firing rate of a homogeneous neuron population as a function of excitatory and inhibitory inputs in terms of a spike response function, which is undetermined, apart from a few general conditions it has to satisfy.

Every squashing function which maps zero input on zero output can be interpreted in terms of our circuit, by a suitable choice for a spike response function for the different populations. This includes squashing functions that are not strictly positive in their domain. In particular, the classical perceptron with a squashing function that has a range of $[-1, 1]$ can be understood as a version of the circuit where linear spike response functions are assumed for the input populations and non-linear spike response functions for the output populations. This shows that squashing functions which are anti-symmetric, can easily be interpreted in terms of a circuit of spiking neurons, which is an interesting result in its own right.

We will show mathematically that for a particular choice of the spike response function, namely a clipped linear response function, the steady state of the circuit corresponds exactly to the activation value of a perceptron with a clipped linear squashing function. This result is of great theoretical interest since this squashing function satisfies the conditions of Universal Approximator Theorems. The exact mathematical correspondence between perceptrons and the steady state of our proposed circuit implies that the results of universal Approximator Theorems carry over to networks of this circuit and, therefore, prove that cortical networks of spiking neurons, using rate coding, can be Universal Approximators. In our opinion, this is a very strong indication indeed, that the great computing power of feed-forward cortical networks, which for instance has been demonstrated in object recognition in the visual cortex (Chelazzi et al., 1993; Tanaka, 1996), can be explained by their resemblance to feed-forward perceptron networks.

Moreover, we will demonstrate that some aspects of the classical perceptron indeed are biologically implausible. Once a biologically plausible choice is made for the spike response function, the steady state of the circuit corresponds to a perceptron that is different from the classical perceptron. Due to the non-linearity of the spike response function, this perceptron will also be non-linear in its inputs. We will demonstrate a particularly interesting property of such perceptrons, namely that they 'saturate', i.e. lose their discriminatory power, if the input activations are very high.

In the discussion, we compare our results with another interpretation of perceptrons based on the population rate hypothesis (Maass & Natschläger, 2000).

2. Choosing the right squashing function

In this section, we will argue that it is essential for an ANN that models a large cortical network to have a squashing function that maps zero input on zero output. If this requirement is not satisfied, firing rates in the network will be too high to be biologically plausible and spurious activation, i.e. activation that is unrelated to the input stimulus, will occur in higher layers of the ANN. In this section, we will demonstrate this, using a simple model of form processing in the visual cortex as an example.

First, we will briefly recall the definition of a squashing function. A perceptron is determined by its output state variable o, the input state variables x_i, where i runs over all inputs, by the input weights associated with each input: w_i and by its update rule: The new output state variable o is calculated according to:

$$o = f\left(\sum_i w_i x_i - \theta\right) \tag{1}$$

Here, the sum runs over all inputs of the perceptron, x_i is the input signal, present at input i, w_i is the weight associated with input i. θ is called the perceptron's threshold and it reduces the perceptron's weighted input signal. The function $f(x)$ is the squashing function. It is a function which maps the interval $(-\infty, \infty)$ on the interval that the

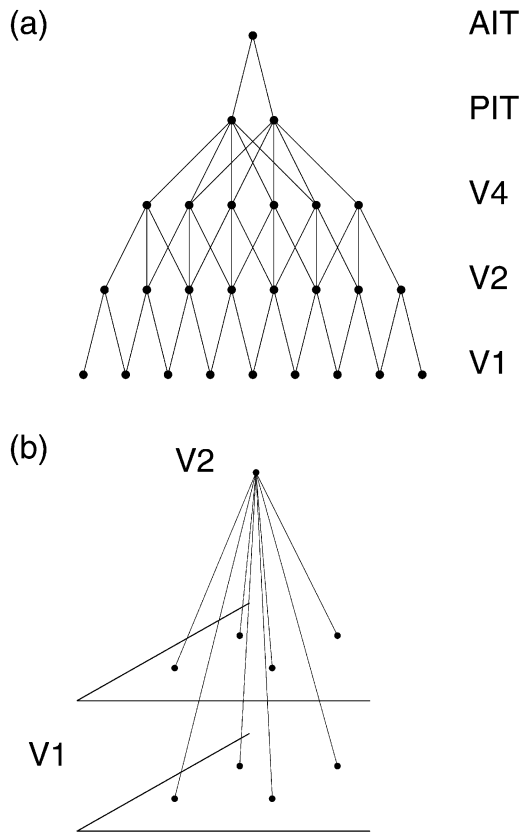

Fig. 1. The structure of a small version of our networks (A). The receptive fields are limited at each layer, except in AIT. (B), the coupling of a V2 neuron to two of the four V1 orientation layers.

perceptron's output value can take, e.g. [0, 1]. It is a non-descending function and often chosen to be continuous and differentiable, although, for instance, the Heaviside step function is also sometimes used.

We now present a simple model of form processing in the visual cortex, which takes place in the so-called 'ventral stream' (Ungerleider & Mishkin, 1982). Although the model is simple, it takes into account two essential features of the 'ventral stream': it is a multi-layer hierarchical system and its neurons have a limited receptive field, except in the highest layer.[1] Neurons that do not have an input in their receptive field only should have a baseline activity of a few spikes per second. An ANN that does not respect this fact cannot possibly be considered an accurate model of a cortical network of spiking neurons. We will show that in multi-layer feed-forward networks for squashing functions that do not map zero input on zero output this problem will occur.

One of the most widely used squashing functions, the so-called logistic equation, will provide a demonstration of this problem.

$$f(x) = \frac{1}{1 + e^{-\beta x}} \tag{2}$$

[1] This layer corresponds to anterior inferiotemporal cortex, where the receptive field contains the entire visual field.

Here, β is the noise parameter, which we will set to $\beta = 1$ throughout the paper. Most of what are nowadays known as ANNs are networks that consist of these simple units, together with an algorithm that specifies in which order the perceptrons are to be updated.

The architecture of the network we have used (van der Velde & de Kamps, 2001) is given in Fig. 1, clearly demonstrating both the hierarchical structure of the network and the limited receptive fields, which become larger higher in the network. The five layers are called 'V1', 'V2', 'V4' 'PIT' and 'AIT', respectively, after their corresponding areas in the 'ventral stream'.

Fig. 2 shows the activities in all layers of the network after one of the trained patterns has been presented to the network. The network correctly identifies the object. However, the activity in the network is also high for perceptrons that do not have a stimulus in their receptive field. It is obvious why: the perceptrons described by Eq. (2) do not map zero input on zero output.

Setting θ to a fixed value does not provide a solution for this problem. In the network shown in Fig. 3A, we set $\theta = 0$ and fixed this value during training. In this case, the activity of perceptrons that do not have a visual stimulus within their receptive field is constant, at least in V2, but the activity is also quite large.

In fact, with the use of Eqs. (1) and (2), a perceptron that receives zero input will produce an output activity of half its maximum value. Translated into biological terms, this would mean that a neuron that does not have a stimulus in its receptive field will show an activity in the order of 100 Hz or more. This is clearly not the case in the visual cortex.

Furthermore, in the higher layers, the activity does not seem to be related to the stimulus anymore. This is because the input patterns were trained at more positions than the one shown in Fig. 2. Therefore, non-constant weights are present at every position in each area and the constant activities in layer V2 will be mapped to non-constant values in layer V4, even though they are not stimulus related.

In a final attempt to use Eqs. (1) and (2), we set $\theta = 3$ and kept its value fixed during training. In this case, zero input is mapped on small output in V2 ($f(0) \approx 0.05$), as can also be seen from Fig. 3B: the average activity of the perceptrons is much smaller in this area. However, in the higher layers there are perceptrons which have a very large activity, even for neurons that do not have the stimulus in their receptive field. This follows from the fact that in order to win over the large thresholds, large weights must be developed everywhere. The activity in V2, although very small, will, therefore, result in considerable activity in higher layers, even if much of it is not stimulus related.

Although we made a very specific choice for the squashing function, using Eq. (2), the explanation of the observed spurious activity holds quite generally. In particular the last example shows that even in a network with a squashing function that maps zero on a very small output in the next

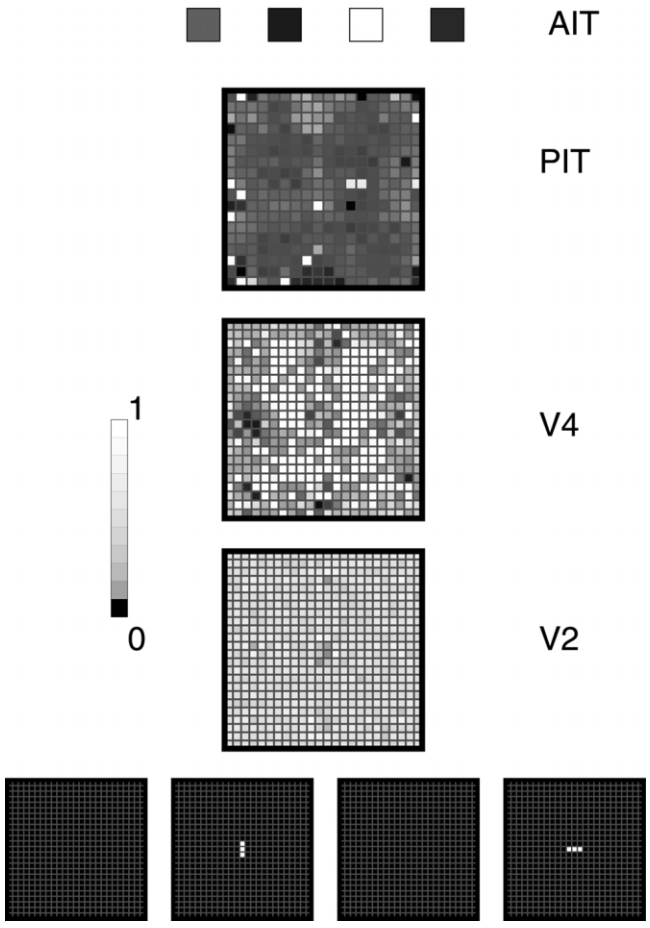

Fig. 2. Activities in the network. The standard logistic equation was used in updating the perceptrons, with values in the interval [0, 1]. θ was left free during training. A cross was presented at the four 'V1' orientation layers. The network performs recognition correctly. The activity in the network is also large in neurons that do not have a visual stimulus in their receptive field.

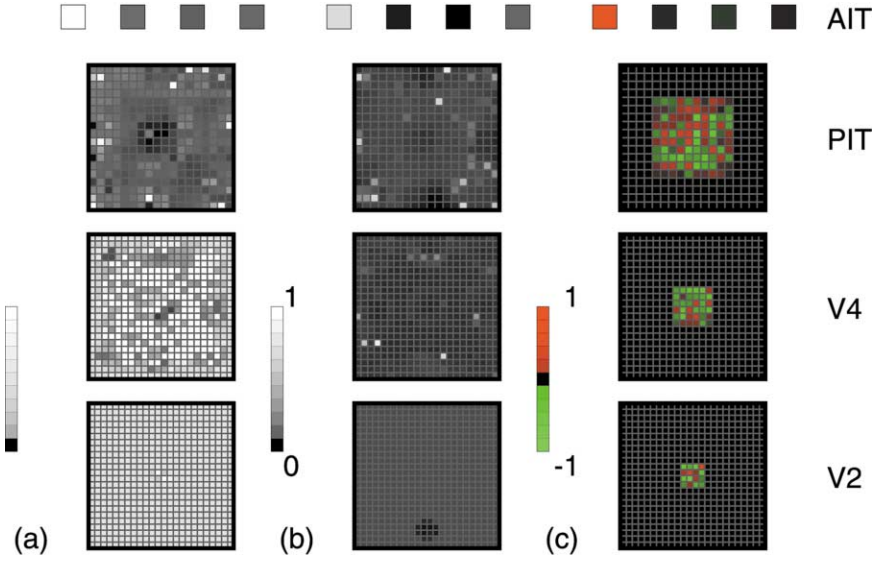

Fig. 3. Activities in the network. The same training parameters were used as in Fig. 2, but θ was fixed ($\theta = 0$ in A, $\theta = 3$ in B). In C, Eq. (3) was used and θ was fixed ($\theta = 0$). Here, only stimulus related activity is present.

higher layer, in layers the spurious activity is substantial. Therefore, to avoid spurious activation, a squashing function is needed which maps zero input on zero output.

An example of such a squashing function is:

$$f(x) = \frac{2}{1 + e^{-\beta x}} - 1 \tag{3}$$

The problem with this function is that the perceptron activation can take negative values. In order to be able to use such a squashing function, an interpretation of negative activation values in terms of firing rates, which obviously are positive, must be found. In the next section, we will see that such an interpretation can be found.

Alternatively, one might consider a squashing function which maps zero input on zero output and which is strictly positive, such as for instance:

$$f(x) = \begin{cases} x \leq 0, & 0 \\ x > 0, & \dfrac{2}{1 + e^{-\beta x}} - 1 \end{cases} \tag{4}$$

We will show, however, that these functions are contained as a special case in the formalism that we will develop. Therefore, there is no need beforehand to exclude squashing functions like (3), in favor of functions like Eq. (4).

3. The interpretation of negative activity

The fact that the perceptron can code for both negative and positive activation strongly suggest that it does not represent the firing rate of a single population. Moreover, a perceptron clearly cannot have positive and negative activation at the same time. This suggests that the simplest interpretation of activation of a perceptron using Eq. (3) is that of a circuit, such as the one in Fig. 4. It consists of two excitatory populations, P and N, two inhibitory populations I_p and I_n and two excitatory populations E_p and E_n. A current on J_p would enhance P, via E_p, and inhibit population N, via I_p. Likewise, a current on J_n would simultaneously inhibit population P, via I_n, and enhance population N, via E_n. In the next section, we will show that under quite general conditions a circuit of this kind can be conceived where:

- the circuit reaches a steady state in finite time;
- at most one of the two population is active in the steady state; and
- the larger of the two input currents J_p and J_n determines which of the two populations is active in the steady state.

We propose that a perceptron's activity in a feed-forward perceptron network corresponds to the steady state of such a circuit. Positive activity in the perceptron would indicate that one of the two populations, say P, is active, while negative activity indicates that the other population, say N, is active. The absolute value of the perceptron's activity then corresponds to the firing rate of the population that is currently active. Since the circuit is symmetric, the labels 'positive' and 'negative' have no intrinsic meaning, and both 'positive' and 'negative' activity are expressed by firing rates of excitatory populations. 'Negative' activity should not be confused with inhibition or activity of inhibitory neurons. We will clarify this point in the discussion on the biological plausibility of the circuit.

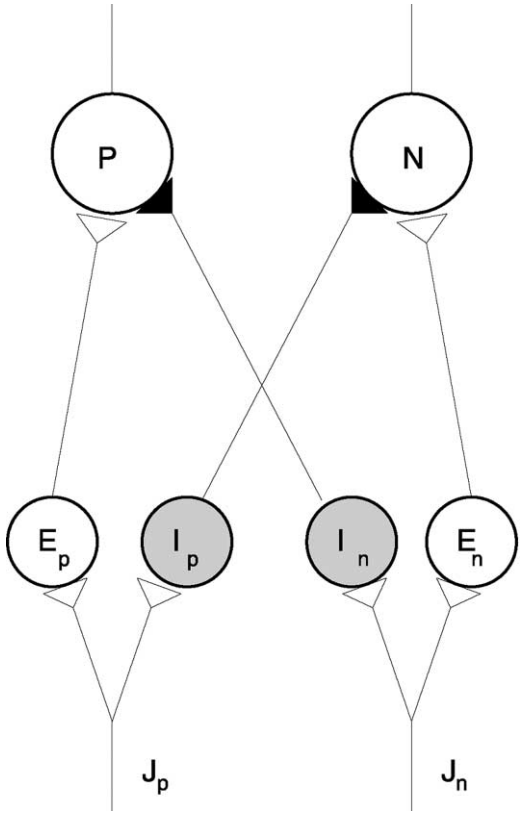

Fig. 4. A neuronal circuit that can perform a perceptron-like function. The shaded areas indicate inhibitory populations. Their action on other populations is indicated by filled triangles. The action of excitatory populations is indicated by open triangles.

The representation of a perceptron's state by two groups of populations is a natural one. An individual perceptron performs a bisection of input space by a hyperplane which is determined by the input weights. The performance of multi-layer feed-forward perceptron networks is sometimes illustrated by repeated application of this bisection, see for instance the discussion of the implementation of the XOR function in Hertz et al. (1991). Each population now labels which side of the half plane the perceptron selects, given the input.

4. Properties of the cross-inhibitory circuit

Wilson and Cowan (1972) introduced a formalism for the description of firing rates of homogeneous populations of neurons, which get contributions from other populations, both excitatory and inhibitory, as well as from external

currents. After applying a procedure that they called 'temporal coarse graining', they arrived at a differential equation for the population rate:

$$\tau\frac{dE}{dt} = -E(t) + S_E(\alpha E' + \beta I + \gamma I_{ext}) \quad (5)$$

Here, $E(t)$ is the fraction of an excitatory population, firing per unit time, at time t. E' is the total contribution of all excitatory input, I the total contribution of all inhibitory input and I_{ext} the total current from external sources. α, β and γ are constants.

τ is the time constant of the population and is not given by the formalism directly, due to the 'temporal coarse graining' procedure. This is the theoretical equivalent of applying a running time average over the population rate, which, therefore, smoothes out variations of the population rate which take place within a time step of order τ. We will treat it as a free parameter, which can be chosen to match observed data optimally.

$S_E(x)$ is the spike response function. General properties of the spike response function are that it is non-descending, tends to zero for large negative arguments and to 1 for large positive arguments. Wilson and Cowan used the logistic function as spike response function, with parameters chosen such that the population starts to respond only after a small net positive input contribution is present in the population. Later, other spike response functions have also been considered (Amit & Tsodyks, 1991), which we will discuss later on.

A similar equation holds for inhibitory populations, although the spike response function may be a little different for these. The original equation also contained a contribution involving the absolute refractive period, which we neglect. In this paper, we generally deal with situations in which rates are so low that this is a reasonable approximation.

Now, assuming that the populations that we use for our circuit are adequately described by Eq. (5), and that the circuit is symmetric, one arrives at the following set of differential equations:

$$\tau\frac{dP}{dt} = -P + S_e(AE_p - BI_n)$$

$$\tau\frac{dN}{dt} = -N + S_e(AE_n - BI_p) \qquad \tau\frac{dI_p}{dt} = -I_p + S_i(CJ_p)$$

$$\tau\frac{dI_n}{dt} = -I_n + S_i(CJ_n) \qquad \tau\frac{dE_p}{dt} = -E_p + S_e(DJ_p)$$

$$\tau\frac{dE_n}{dt} = -E_n + S_e(DJ_n) \quad (6)$$

Here, P, N, E_p, E_n, I_p and I_n are population rates and A, B, C and D are constants. $S_e(x)$ is the spike response function for excitatory populations and $S_i(x)$ the spike response functions for inhibitory populations.

To understand the behavior of the circuit, consider the situation where the input currents J_p and J_n are constant. Within a few time steps of order τ, the populations E_p, E_n, I_p and I_n reach a value close to the asymptotic values, $S_e(DJ_p)$, $S_e(DJ_n)$, $S_i(CJ_p)$ and $S_i(CJ_n)$, respectively. This means that afferents of P and N are effectively constant, and they in turn will quickly converge to values

$$\lim_{t\to\infty} P = S_e(AS_e(DJ_p) - BS_i(CJ_n)) \quad (7)$$

and

$$\lim_{t\to\infty} N = S_e(AS_e(DJ_n) - BS_i(CJ_p)) \quad (8)$$

respectively. These values define the steady state values of the circuit. For simplicity, let us assume that $A \approx B$, $C \approx D$ and $S_e \approx S_i$. It is reasonable to assume that the spike response function is zero if the inhibitory afferents dominate and only starts to become positive if the excitation exceeds the inhibition by a small threshold value a ($a \geq 0$) (Wilson & Cowan, 1972). It is clear from the equations above that if $J_p > J_n$, then $\lim_{t\to\infty} P$ is positive and $\lim_{t\to\infty} N = 0$. Since the circuit is symmetric, the reverse happens if $J_n > J_p$. This proves the claims made in the previous section for the behavior of the circuit.

Equations like Eq. (5) have been used before for neural network modeling (Almeida, 1988; Cohen & Grossberg, 1983; Pineda, 1987), but our work is different in the sense that we have applied them to a multi-population circuit.

5. From circuits to perceptrons

Looking for a neuronal substrate of a perceptron that can have both positive and negative activation, we introduced a cortical circuit which in the steady state indeed can code for both positive and negative activity. Since Eqs. (7) and (8) describe the steady state of the circuit, they also describe a perceptron-like object. It follows immediately that the behavior of this perceptron-like object is determined by the spike response function $S(x)$ (we take $S(x) = S_e(x) = S_i(x)$ from now on).

In order to interpret the circuit as a perceptron, one must first decide on how to represent weights. The connection between two perceptrons, say i and j, is characterized by a single weight, which may have a positive or a negative value. The two corresponding circuits may in principle be linked by eight connections, each characterized by its own positive strength: the P population of circuit i, may connect to the E_p, E_n, I_p and I_n populations of circuit j, as does the N population of circuit i. In order to represent these connections by a single weight, as we must for a perceptron-like interpretation of the circuit, we adopt the following convention: if the weight is positive, then its value determines the connection strength between population P of circuit i and populations E_p and I_p of circuit j and the connection strength between population N of circuit i and populations E_n and

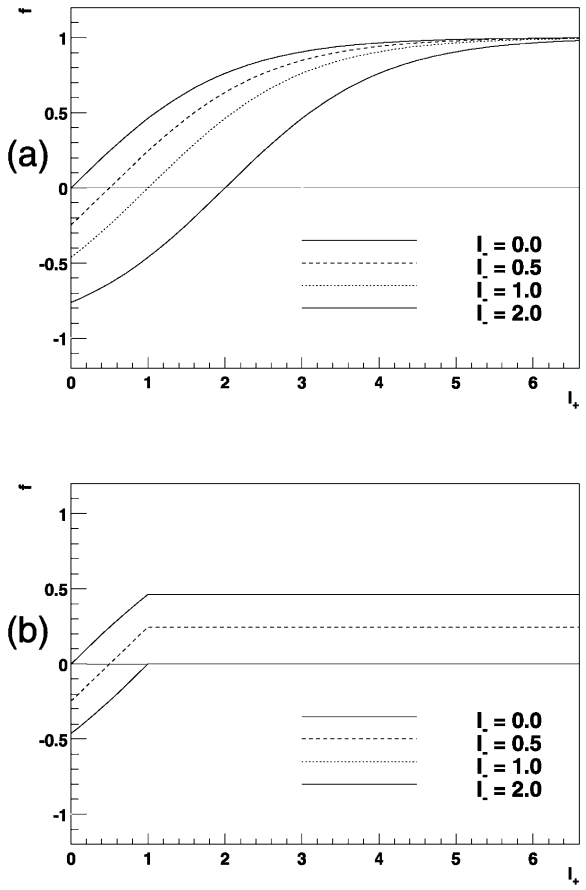

Fig. 5. The squashing behavior of a perceptron, determined by a linear spike response function for populations E_p, E_n, I_p and I_n and by a non-linear spike response function for populations P and N (A). Here, I_+ is the sum of all positive input contributions and I_- is the sum of all negative input contributions. The result is behavior identical to the classical perceptron. Introducing a bounded spike response function for populations E_p, E_n, I_p and I_n results in the 'saturation effect'. We demonstrate this in B. For $I_- > 1$, the value of I_+ the curves are on top of each other.

I_n of circuit j. The other connection strengths are then set to zero. If the weight is negative, then its absolute value determines the connection strength between the population P of circuit i and populations E_n and I_n of circuit j and the connection strength between population N of circuit i and populations E_p and I_p of circuit j. Again, the others are set to zero. Since the P and N populations are excitatory, all these connection values are positive.

Now that we have defined what a weight is between two perceptrons, we can relate the input currents J_p and J_n to weights and inputs of the ANN. Introduce:

$$\sum_{j,+} w_j i_j \equiv \sum_j \theta(w_j i_j) w_j i_j \tag{9}$$

$$\sum_{k,-} w_k i_k \equiv \sum_k \theta(-w_k i_k) w_k i_k \tag{10}$$

where $\theta(x)$ is the Heaviside step function:

$$\theta(x) = \begin{cases} x < 0, & 0 \\ x \geq 0, & 1 \end{cases} \tag{11}$$

Thus, $\sum_{j,+} w_j i_j$ sums up the positive inputs $w_j i_j$ of a perceptron in an ANN and $\sum_{k,-} w_k i_k$ sums up the negative inputs $w_k i_k$ of this perceptron in an ANN. Here the w and i are the weights and inputs, respectively, of the perceptron corresponding to this circuit. We now have:

$$J_p = \sum_{j,+} w_j i_j \qquad J_n = -\sum_{k,-} w_k i_k \tag{12}$$

Now, we want to identify the steady states of the circuit with perceptron-like behavior. In order to do this we associate activity of the P population with the positive values of a squashing function and activity of the N population with the negative values of a squashing function. To make this identification formally, we define the anti-symmetric continuation $\mathscr{AC}$ of a function:

$$\mathscr{AC}[f(x)] = \begin{cases} x < 0, & -f(-x) \\ x > 0, & f(x) \end{cases} \tag{13}$$

Now, introduce a perceptron with the following update rule:

$$o = \mathscr{AC}\left[S\left(A\left(S\left(\sum_{j,+} w_j i_j \right) - S\left(-\sum_{k,-} w_k i_k \right) \right) \right) \right] \tag{14}$$

Here, A is the circuit parameter, used in Eq. (6) (note that D can be absorbed by the input weights). Substitution of Eq. (12) in Eq. (14) shows that for $J_p \geq J_n$ we have

$$o = \lim_{t \to \infty} P \tag{15}$$

and for $J_p < J_n$

$$o = -\lim_{t \to \infty} N \tag{16}$$

Eq. (14) codes for the steady state of the circuit: negative values indicate that in the steady state the N population is active with firing rate $-o$, positive values indicate that in the steady state the P population is active with firing rate o. Furthermore, Eq. (14) describes a generalized perceptron. We can show this if we allow a slight generalization of Eq. (14). If we assume that the P and N populations have spike response function $S(x)$ and the E_p, E_n, I_p and I_n populations have spike response function $s(x)$, Eq. (14) becomes

$$o = \mathscr{AC}\left[S\left(A\left(s\left(\sum_{j,+} w_j i_j \right) - s\left(-\sum_{k,-} w_k i_k \right) \right) \right) \right] \tag{17}$$

Now, assume $s(x)$ is given by:

$$s(x) = \begin{cases} x \leq 0, & 0 \\ x > 0, & x \end{cases} \tag{18}$$

The perceptron given by Eq. (17) is completely equivalent to the perceptron given by Eq. (1), if we take as a squashing

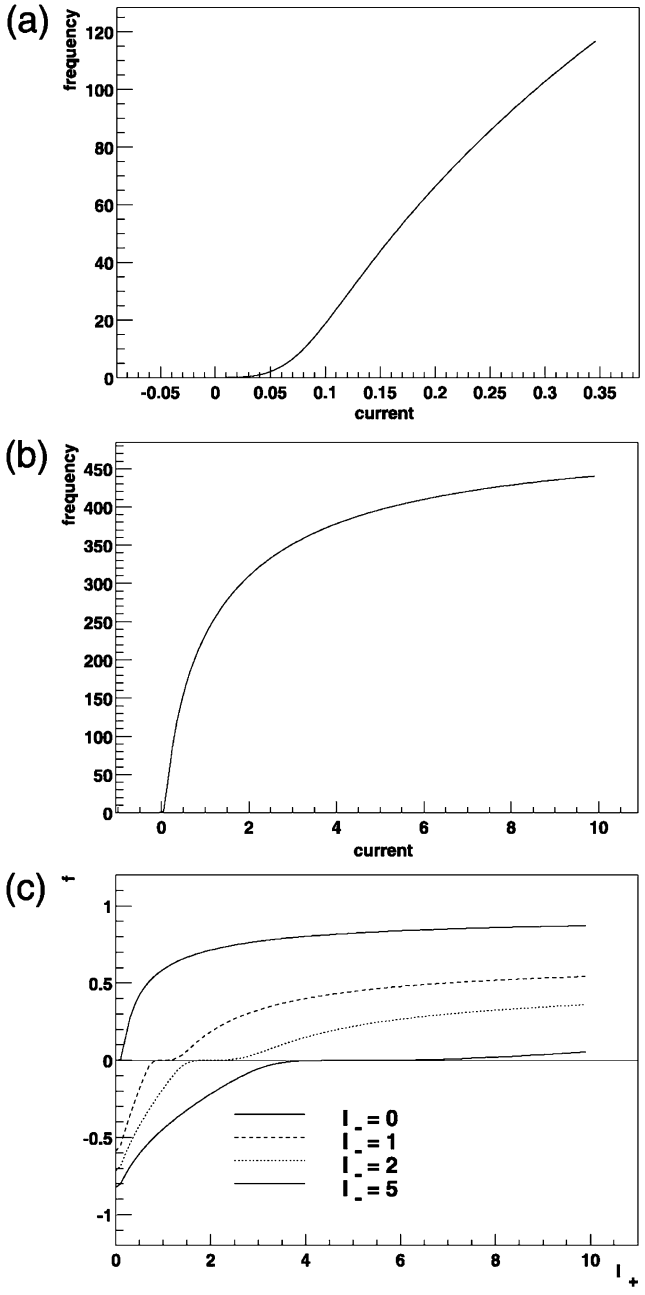

Fig. 6. The spike response function for small positive input currents (A) and large positive input currents (B). A smooth threshold for small input currents is clearly visible in (A). The resulting squashing behavior of the perceptron determined by the spike response function is given in (C) for different values of I_-, where I_- is the sum of all negative input contributions. The non-linear interaction between positive and negative inputs is clearly visible, as is the 'saturation effect'.

function:

$$f(x) = \begin{cases} x \leq 0, & -S(-x) \\ x > 0, & S(x) \end{cases} \tag{19}$$

In Fig. 5A, we show the value of this perceptron with two inputs, one with weight 1 and one with weight −1. We set $A = 1$ and vary the input signal on the input with the positive weight, I_+, and keep the signal on the input on the negative weight, I_-, fixed for various values. For $S(x)$ we used:

$$S(x) = \begin{cases} x \leq 0, & 0 \\ x > 0, & \dfrac{2}{1 + \mathrm{e}^{-\beta x}} - 1 \end{cases} \tag{20}$$

It is clear that this perceptron behaves like the classical perceptron, described by Eqs. (1) and (3).

To arrive at the classical perceptron, we had to make some assumptions which seem unrealistic from a biological

point of view. First of all, we had to assume that the spike response function $s(x)$ is unbounded. If we take a linear, but bounded spike response function, the behavior of the perceptron is quite different. We show this in Fig. 5B, for which we used:

$$s(x) = \begin{cases} x \leq 0, & 0 \\ 0 < x \leq 1, & x \\ x > 1, & 1 \end{cases} \tag{21}$$

In Fig. 5B, again we show the perceptron's response as a function of I_+, for fixed values of I_-. The first obvious difference with Fig. 5A is that all response values are lower. This is because population E_p for a bounded spike response function will be driven to maximum firing rates if $I_+ > 1$. The input current J_p will be clipped by population E_p, before it is fed into population P and the perceptron as a whole will not respond to a further increase of I_+. The argument of the function $S(x)$ cannot, therefore, reach large positive or negative values and the maximum value that can be reached is $S(1) \approx 0.46$, for $A = 1$.

As a consequence of the saturation of population E_p, for increasing fixed values of I_-, the sensitivity of the perceptron with respect to the difference between I_+ and I_- diminishes. For $I_- = 0.5$, the maximum perceptron value that can be reached is $S(0.5)$, and for $I_- \geq 1$ there is no way that any I_+ can cause the perceptron to respond with a positive value. We call this the 'saturation effect'.

The second assumption which seems very implausible from a biological point of view is the choice of widely different response functions for populations P and N on the one hand, and for E_P, E_N, I_P and I_N on the other hand. Let us consider a single spike response function for all populations, which is reasonable from a biological point of view.

In Fig. 6A and B we show a spike response function that was used by Amit and Tsodyks (1991). In the Appendix we give the formula. This spike response function shows a threshold behavior, which is softened by the presence of noise. In Fig. 6C we show the equivalent of Fig. 5A and B; the resulting perceptron behavior using this spike response function. The saturation effect is clearly visible and also the effects of the non-linearities in the spike response function, both at threshold and at high input current.

The most important difference between the classical artificial neuron of Eq. (1) and the one described by Eq. (14) is that in the classical artificial neuron all input currents add linearly and that in the generalized perceptron positive and negative contributions each are added separately, and then are combined non-linearly. At high input currents, this non-linearity causes the saturation effect.

In general, the circuit may respond in a more complex way, we have chosen parameters so that the circuit resembles a perceptron in steady state. A particularly interesting option is to leave out the N population. Following the reasoning in this section, it is easy to see that such a circuit would be able to implement a function like the one given by Eq. (4). Such functions satisfy the requirements of the Universal Approximator Theorems, and have, therefore, the same computational power as the circuit discussed above.

6. A special case: linear spike response functions

In general, the perceptron described by Eq. (14) does not behave like the classical perceptron, described by Eq. (1), because the former is non-linear in its inputs. For a special case, namely clipped linear spike response functions, however, a complete equivalence between the two can be established. This is important for two reasons. First of all, clipped linear spike response functions can be considered crude approximations of 'sigmoid' spike response functions. Since for the linear case there is a direct mathematical correspondence between the perceptrons of Eq. (14), which represent steady state values of the cortical circuit we propose, and classical perceptrons, it immediately shows the relevance of classical multi-layer feed-forward perceptron networks for the situation where the clipped linear spike response function is an acceptable approximation.

Secondly, since multi-layer feed-forward are Universal Approximators and the clipped linear squashing function satisfies the conditions, imposed by Universal Approximator Theorems, this directly shows that feed-forward networks of cortical circuits are Universal Approximators, for the clipped linear spike response function. It opens the possibility that this is true for other spike response functions as well. We will now first demonstrate equivalence between classical ANNs and networks of cortical circuits, for the linear case.

The clipped linear spike response function is given by Eq. (21): it leads, using Eq. (14), to the clipped linear squashing function

$$f(x) = \begin{cases} x < -1, & -1 \\ -1 \leq x < 1, & x \\ x > 1, & 1 \end{cases} \tag{22}$$

provided that the sum of the positive inputs and the sum of the negative inputs at each perceptron are each smaller than one. Otherwise, the 'saturation effect' will occur. That means that each ANN using Eqs. (1) and (22) for its perceptrons can be mapped on a network of circuits like the one in Fig. 4, using clipped linear threshold spike functions. This network of circuits has the same functionality as the original ANN, provided the 'saturation effect' does not occur.

But the 'saturation effect' can be removed in this case. Let $\alpha > I_{max}$, where $I_{max} > 1$ is the largest input current in the network,[2] that is, the largest J_p or J_n for any circuit. Now,

[2] If $I_{max} < 1$, the procedure is clearly unnecessary.

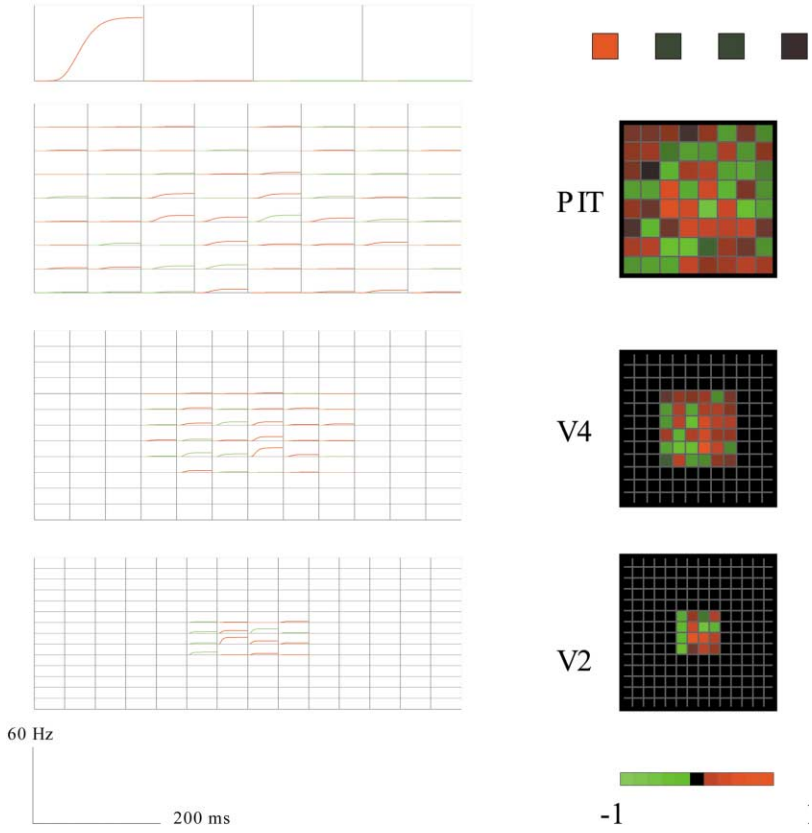

Fig. 7. The time evolution of a network of circuits (A), with clipped linear threshold spike response functions. Only the most active population is shown at each position. If this is the P population, the graph line is red, otherwise green. In (B), we show the activation values of the corresponding ordinary ANN. We took $\tau = 10$ ms for the population constant.

replace all connection weights w_{ij} in the network by $w'_{ij} \equiv w_{ij}/\alpha$ and replace A in Eq. (14) by $A' \equiv \alpha A$. The 'saturation effect' is now removed and every circuit will converge to a steady state that is exactly the same as those of the corresponding perceptron in the ANN.

We illustrate the equivalence between networks of cortical circuits and ANNs in Fig. 7. The right side shows an ANN with a structure similar to the one in Fig. 1, trained with the 'backpropagation with momentum' rule. Eq. (3) was used as a squashing function. Values for input and output pattern values different from zero were taken to be 0.1. This reflects the fact that in steady state rates in visual cortex are typically in the order of 30–70 Hz, which is about 10% of their maximum firing rate. The network clearly shows the benefit of using a squashing function like Eq. (3) in connection with the rate problems discussed above: in the higher layers there is no spurious activity and the rates in intermediate layers are quite low. In fact, the rates are so low that the squashing function can be approximated by a linear squashing function, which immediately shows the relevance of linear squashing functions. Compare this to the networks in Fig. 2, where many neurons are driven to their maximum firing rate, a situation that does not occur in the visual cortex.

We have used the weights of this ANN to simulate a network of cortical circuits that has the same architecture as the ANN. We simulated the dynamical evolution of this network, by numeric integration of Eq. (6) for all circuits. This is an explicit demonstration of a network of cortical circuits that:

- performs calculations in a way very similar to classical ANNs;
- respects important rate considerations;
- gives a meaningful interpretation for negative perceptron values in terms of the (positive) firing rates of excitatory populations of neurons; and
- allows the introduction of population dynamics in feed-forward networks and, therefore, introduces the possibility to study network dynamics in networks that were trained using connectionist methods.

7. Discussion

In this paper, we have shown that a naive interpretation of the population firing rate hypothesis leads to severe rate problems. These rate problems can be avoided if a squashing function is chosen that maps zero input on zero output. A large and important class of squashing functions

are anti-symmetric in their input and, therefore, can lead to either positive or negative activation values. Perceptrons, defined using such squashing functions, cannot be interpreted in terms of a single firing rate. Instead, we propose that the state of a perceptron can be mapped on a simple cross-inhibitory circuit. We have shown that for a suitable, but biologically implausible, choice of spike response functions, networks of such circuits are mathematically equivalent to multi-layer feed-forward perceptron networks. Using the same circuit, but introducing biologically inspired spike response functions for all populations, we have shown that the steady state of this circuit still corresponds to a perceptron-like object. This perceptron is non-linear in some of its inputs, however. One important consequence of this that the perceptron loses discriminatory power if the inputs are large, the 'saturation effect'.

Negative activation is not strictly necessary for perceptrons that map zero input on zero output. There are functions like Eq. (4), which satisfy the Universal Approximator Theorems, which do not have negative values. There are a number of reasons to consider negative activation, however. First of all, it is an interesting and important point in itself that negative perceptron activations can be interpreted in terms of spiking neurons. It implies that the well-studied logistic function makes sense as a squashing function, if it is used in the form of Eq. (3). Secondly, squashing functions like Eq. (4) can be considered a special case of the more general one where negative activation is present.

The circuit we described is the simplest circuit that guarantees that for a given input either the P or the N population becomes active, but not both. One alternative that we considered was direct lateral inhibition between the P and N population, via two populations of interneurons. Here, it turns out to be difficult to ensure that only one population is active in the steady state and for some parameter choices a steady state is not reached because of oscillations. This means that the circuit we chose is the simplest that has a clear correspondence to a perceptron.

Our arguments show that ANNs or functional equivalents can be implemented by networks of spiking neurons in the cortex, using rate coding. We have given an explicit example of an ANN, trained by conventional methods, using a conventional squashing function, that solves a very simple form recognition problem. The firing rates that result in the populations of the circuit are very low. The rates in Eq. (6) are population rates and populations may react quickly to changes in the input. Indeed, it is argued that the τ in Eq. (6) should be of the order of the duration of a synaptic pulse (Gerstner, 2000). The value we used in Fig. 7 (10 ms) is conservative, in that respect.

There are a few points to be made with respect to the biological implementation of the circuit. First of all, for anti-symmetric squashing functions, negative weights cannot be interpreted as inhibitory weights. A strong positive input signal will result in a strong negative output signal, and a strong negative input will lead to a positive output signal. If a negative weight were to have an inhibitory function, it should drive the perceptron to zero, which obviously does not happen. The discussion in Section 5 on the interpretation of the weights also makes this clear: P and N populations of a circuit are excitatory populations, characterized by positive firing rates. They are connected to other circuits by positive connection strengths. The sign of a perceptron weight merely labels which connection between populations of two circuits is indicated.

The circuit interpretation has an additional advantage: it respects Dale's law (Abeles, 1991). This states that if a neuron is excitatory, it is excitatory on all its output neurons and if it is inhibitory, then it is inhibitory on all of its output neurons. However, in perceptron networks there is usually no constraint on the sign of the weights that fan out from a given neuron. Hence, an interpretation of negative weights in terms of inhibition would immediately render the network unbiological, since it would violate Dale's law. In the circuit interpretation, Dale's law is not violated at all, because all perceptron weights refer to positive connections between excitatory populations.

Within the circuit, inhibition plays an important role, which is local, however. The long range, inter-area connections are performed by excitatory populations. This is in line with another observation on cortical neurons that inhibitory cells do not seem to make long-range connections (Abeles, 1991).

We believe that our circuit would be implemented inside a cortical column. All connections that our circuit uses are present inside a cortical column, see for instance Fig. 1.9 in Shepherd (1990). A canonical microcircuit has been proposed as a basic circuit that performs many of the different functions that have been observed in the visual cortex (Douglas, Martin & Whitteridge, 1989). Our circuit could easily be embedded in this circuit, provided there is a substructure in the GABA cell population.

We believe that population rate coding has several advantages over time coding as proposed by Maass (1997). Rate coding is robust with respect to noise and can even be used if most of the population's input is not related to the stimulus at all (Amit & Brunel, 1997a). Amit and Brunel (1997a) point out that this is, in fact, a common situation in the cortex. Given such a barrage of input spikes, which are not related to the input stimulus, and which in fact have high variability (Amit & Brunel, 1997b), it seems that an accurate coding of the input signal is difficult, if not impossible. Furthermore, it is not immediately obvious how the formalism of time coding generalizes to populations of neurons (the experimental results of Thorpe et al., 1996, are ERP measurements, which clearly shows that many neurons are involved). Given the fact that the interpretation of these matters for population rate coding is obvious and can also meet the speed requirement for cortical processing, we strongly favor this interpretation.

A recent model also using rate coding is presented in

Maass and Natschläger (2000). Maass and Natschläger estimate the firing activity y in pool V in terms of firing activities $x_1, \ldots, x_n$ in n presynaptic pools U_i and show that, under some circumstances, y can be approximated by a function of a weighted sum of x_i. They derive this result under the assumption that the release of neurotransmitter from a vesicle is a stochastic parameter and they show that the function of the weighted sum of inputs is essentially a sigmoid.

This is an intriguing result, because it may imply that functions like Eq. (2) may play a role in cortical computations. The analysis is far more detailed than previous justifications of linear weighted sigmoids (Amit, 1989). There is one problem, however, which already has been identified by the authors. In order to show rigorously that the squashing function they derive is not dominated by noise, Maass and Natschläger have to assume that a synaptic connection is determined by a single vesicle. Although there is some robustness in their work with respect to this assumption (their simulations show a relatively noise free squashing function for a synaptic connection with five release sites), this assumption is unrealistic. A synaptic connection of a motor neuron, for instance, typically has 300 release sites (Zucker, Kullmann & Bennett, 1999). This work would gain considerably in impact if it could be shown that this assumption is not essential for obtaining a relatively noise-free sigmoid.

We also propose to consider the sigmoid of Maass and Natschläger to be a spike response function (in the sense of Section 5), rather than a squashing function. If the sigmoid would be considered to be a squashing function, the rate problems we discussed in Section 2 could occur. The sigmoid in Fig. 1 of Maass and Natschläger (2000) bears strong resemblance to Eq. (2) with positive θ. In Section 2, we argued that networks using this squashing function will experience rate problems and we believe that the arguments we presented there also apply here.

The interpretation as a spike response function would also take care of another problem, that is not addressed by Maass and Natschläger. In a feed-forward network, each population U_i is the population V of the previous layer, which means that the connections between the U_i and V populations are interarea connections. Negative weights cannot be implemented directly by inhibitory U_i populations, because inhibitory connections are local to the cortical area. So, the V population has to be partitioned into subpopulations of excitatory neurons and inhibitory interneurons. To implement negative weights, an interaction between those subpopulations would have to occur. This is true, even if one insists on interpreting the sigmoid as a squashing function that is strictly positive in its domain. To accommodate for the interactions between these subpopulations one would then be forced to come up with a circuit that is quite similar to ours (which can accommodate strictly positive squashing functions).

Maass and Natschläger's work is partly motivated by the observation that some of the assumptions in the work of Wilson and Cowan (1972) are now seen as unrealistic, such as, e.g. a strong variation in the firing thresholds of the neurons. Although we used Wilson and Cowan dynamics to model our circuit, our results do not depend on this formalism. The most important assumption that we need is that the steady state of a population can be formulated in terms of spike response functions, which leads to Eqs. (7) and (8). Details of the dynamics may then differ from Eq. (6), but this affects none of our results.

Appendix A

The spike response function, given in Amit and Tsodyks (1991) is given by

$$\phi(I) = \left\{1 + \frac{\tau}{\sigma}\int_0^{\theta} \mathrm{d}z \exp\left(\frac{(z-I)^2}{\sigma^2}\right)\left[erf\left(\frac{z-I}{\sigma}\right) + 1\right]\right\}^{-1} \tag{23}$$

In this formula, I is an input current, which consists of a constant contribution and a stochastic contribution of mean μ and standard deviation σ. θ is the threshold value of the membrane potential. τ is the time constant of the neuron's membrane. This formula applies for constant μ and σ. The number of afferent spikes during one time step of length τ must be large. The parameters we used were $\tau = 5$ ms, $\sigma = 0.05$, $\theta = 0.25$ and $\mu = 0.10$. The context in which this spike response function will be valid is after the initial transient, when visual input is constant during fixation. For instance, in Amit and Brunel (1997a) this formula is applied to model short-term memory.

References

Abeles, M. (1991). *Corticonics*, Cambridge: Cambridge University Press.

Almeida, L. B. (1988). Backpropagation in perceptrons with feedback. In R. Eckmiller & Ch. von der Malsburg, *Neural computers* (pp. 199–208). Berlin: Springer-Verlag.

Amit, D. J. (1989). *Modeling brain functions*, Cambridge: Cambridge University Press.

Amit, D. J., & Brunel, N. (1997a). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, *7*, 237–252.

Amit, D. J., & Brunel, N. (1997b). Dynamics of a recurrent network of spiking neurons before and following learning. *Network*, *8*, 373–404.

Amit, D. J., & Tsodyks, M. V. (1991). Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate-spikes, rates and neuronal gain. *Network*, *2*, 259–273.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347.

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 815–826.

Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1989). *Neural Computaiton*, *1*, 480–488.

Fukishima, K. (1988). Neocognitron—a hierarchical neural network capable of visual-pattern recognition. *Neural Networks*, *1*, 119–130.

Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Computation*, *12*, 43–89.

Hertz, J. A., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*, Redwood City: Addison-Wesley.

Hoffmann, A. (1998). *Paradigms in artificial intelligence*, Berlin: Springer-Verlag.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359–366.

Maass, W. (1997). Fast sigmoidal networks via spiking neurons. *Neural Computation*, *9*, 279–304.

Maass, W., & Natschläger, T. (2000). A model for fast analog computation based on unreliable synapses. *Neural Computation*, *12*, 1679–1704.

Motter, B. C. (1994). Neural correlates of attentive selection for color and luminance in extrastriate area V4. *Journal of Neuroscience*, *14*, 2178–2189.

Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, *59*, 2229–2232.

Pouget, A., & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience supplement*, *3*, 1192–1198.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain functions*, Oxford: Oxford University Press.

Shepherd, G. M. (1990). *The synaptic organization of the brain*, New York, Oxford: Oxford University Press.

Tanaka, K. (1996). Representation of visual features of objects in the inferotemporal cortex. *Neural Networks*, *9*, 1459–1476.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield, *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: MIT Press.

van der Velde, F., & de Kamps, M. (2001). From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neurosciences*, *13* (4).

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

Zucker, R. S., Kullmann, D. M., & Bennett, M. (1999). Release of neurotransmitters. In M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts & L. R. Squire, *Fundamental neuroscience*, San Diego: Academic Press.