



PERGAMON

Neural Networks 13 (2000) 1–9

Neural  
Networks

www.elsevier.com/locate/neunet

Contributed article

# A learning rule for dynamic recruitment and decorrelation

K.P. Körding, P. König\*

*Institute of Neuroinformatics, ETH/UNI Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*

Received 11 May 1999; received in revised form 29 September 1999; accepted 29 September 1999

## Abstract

The interest in neuronal networks originates for a good part in the option not to construct, but to train them. The mechanisms governing synaptic modifications during such training are assumed to depend on signals locally available at the synapses. In contrast, the performance of a network is suitably measured on a global scale. Here we propose a learning rule that addresses this conflict. It is inspired by recent physiological experiments and exploits the interaction of inhibitory input and backpropagating action potentials in pyramidal neurons. This mechanism makes information on the global scale available as a local signal. As a result, several desirable features can be combined: the learning rule allows fast synaptic modifications approaching one-shot learning. Nevertheless, it leads to stable representations during ongoing learning. Furthermore, the response properties of the neurons are not globally correlated, but cover the whole stimulus space. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Learning; synaptic plasticity; Inhibition; Receptive field; Backpropagating action potential; Synchronization; Oscillation

## 1. Introduction

The origin and plasticity of the highly specific connectivity of neuronal networks are central topics in neurobiology. The pattern of synaptic connections defines the processing and memory storage properties. Many studies indicate that the specification of wiring is partly genetic and partly due to epigenetic factors (Singer, 1986). The latter, i.e. learning and adaptation in the environment is implemented by synaptic plasticity: the efficacy of synapses is not constant but a function of many variables. On one hand, it seems plausible to utilize only locally available information for such a process, making the presynaptic action potential and the postsynaptic membrane potential a first choice. As a consequence, most learning rules investigated are modifications of the classic rule proposed by Hebb (1949) (Artola, Bröcher & Singer, 1990; Bienenstock, Cooper & Munro, 1982; Sejnowski, 1977; Stent, 1973). He proposed, that synaptic efficacy is increased upon correlated pre- and postsynaptic activity. On the other hand, the performance of the system is measured not on the level of individual synapses, but on the network level (Atick and Redlich, 1990; Barlow, 1989; Barlow, Kaushal & Mitchison, 1989; Comon, 1994; Dan, Atick & Reid, 1996; Deco & Obradovic, 1996). Interesting measures being the total

amount of information transmitted or the quality of the reconstruction of a stimulus. This difference in scope of the definition of a learning rule, and its evaluation creates the problem, how to make relevant information available to the local process of synaptic modifications. A number of solutions have been proposed, but as a tradeoff of interfering with the dynamics of the system and the performance of the learning rule has to be made, or strong physiological assumptions are involved, no commonly accepted solution is known.

A particularly well investigated example is the development of receptive fields in primary sensory areas. Several aspects characterize such systems. Firstly, the afferent fibers have “simple” response properties (e.g. spatially concentric on and off zones in the visual field or responding to sound within a narrow frequency band only). Secondly, the afferent fibers are topographically ordered. Thirdly, the receiving neurons have qualitatively more complex receptive fields (e.g. orientation selectivity or AM-modulation selectivity). The receptive field properties of the receiving neurons are at least in part determined by the pattern of synaptic connection with the afferent fibers. These connections presumably are formed according to some learning rule (cf. Katz & Callaway, 1992). In such a system the performance of a “local” learning rule governing the modification of these synapses has to be evaluated according to several global measures. *Firstly*, all possible stimuli should specifically activate some neurons in the network, i.e. the

\* Corresponding author. Tel.: +41-1-635-3060; fax: +41-1-635-3053.

E-mail address: peterk@ini.phys.ethz.ch (P. König).

### Nomenclature

$\omega$ , $\Delta\omega$ , $\Delta\omega_{LTP}$ , $\Delta\omega_{LTD}$ , $\Delta\omega_{Norm}$	synaptic efficacy, change of synaptic efficacy, and the its different constituents
$\alpha_{LTP}$ , $\alpha_{LTD}$	rate parameters of the change of synaptic efficacy
$\Delta t$ , $\tau$	difference of the arrival time of pre- and post-synaptic action potential and the time constant ruling such detection of coincidences

union of all receptive fields should cover the stimulus space. This implies, that the receptive fields are not globally correlated and the neurons may not all learn the same pattern. Obviously, this makes some form of interaction between the neurons necessary (Barlow, 1989). *Secondly*, rules of synaptic plasticity should allow quick learning. Performance of biological system indicates extremely fast performance, reaching one-shot learning in extreme cases. This requirement precludes extensive sampling using a long iterative procedure to determine the space of stimuli to be covered. In contrast, a decision has to be made on the spot, when a particular stimulus is presented (Buhmann, 1989; Wendemuth & Sherrington, 1993). *Thirdly*, the system should allow ongoing learning and be stable simultaneously. Within the capacity of the network later learning should not diminish old memories. Thus, repeated presentation of one

stimulus should not induce an increase of the respective representation without bounds. *Last, but not least*, a learning rule should be compatible with known physiological properties of cortical neurons. The mechanisms should not interfere with the processing of stimuli by the network as such. Each of these problems has been addressed by different studies using individually tailored learning rules. However, none seems suited to solve all of these issues.

Here we present a learning rule which addresses—not necessarily providing the final solution to—the issues listed above. It is inspired by physiological experiments which indicate that action potentials which are elicited in a neuron propagate not only anterogradely along the axon, but also retrogradely invade the dendritic tree (Stuart & Sakmann, 1994). If such a “backpropagating” action potential arrives at a synapse simultaneously (i.e. within a small temporal window), with an action potential in the presynaptic afferent fiber the efficacy of the respective synapse is increased (Magee, Hoffman, Colbert & Johnston, 1998; Markram, Lubke, Frotscher & Sakmann, 1997). The modifications of synaptic efficacy are dependent on the precise temporal relationship of pre- and postsynaptic activity (Markram et al., 1997) which had been predicted for theoretical reasons (Gerstner, Ritz & van Hemmen, 1993). Recent experimental results give an interesting twist to this scheme. Strong activation of inhibitory synapses located at the proximal dendritic tree may block or attenuate the retrograde propagation of the action potential in the dendritic tree (Spruston,

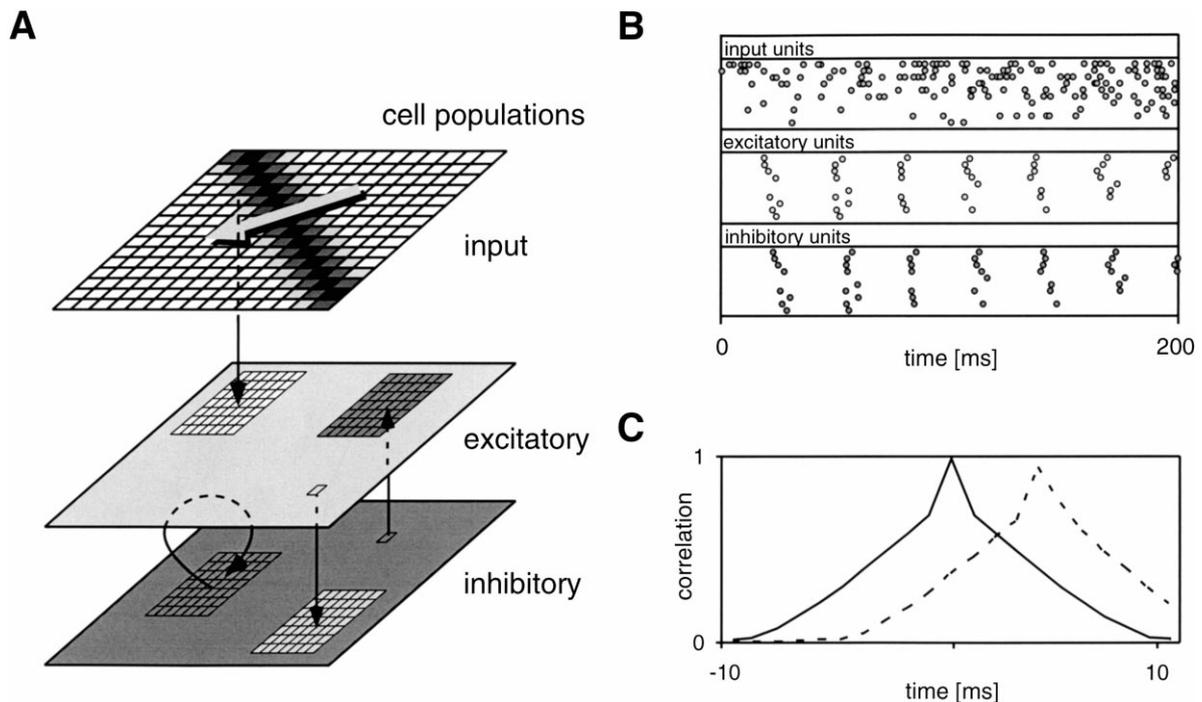


Fig. 1. The architecture of the network: (a) The neurons are arranged in a stack of three 2-dimensional layers. One stimulus is projected as a gray level picture onto the input layer (top). The large arrow indicates the direction of movement of this stimulus orthogonal to its orientation. The small arrows indicate implemented connections with the size of projection zones. (b) Time trace of the activity of 10 input units (top), excitatory units (middle) and inhibitory units (bottom) during stimulation. (c) Correlation function of the excitatory neurons and excitatory neurons (solid line) and inhibitory neurons (dashed line), respectively.

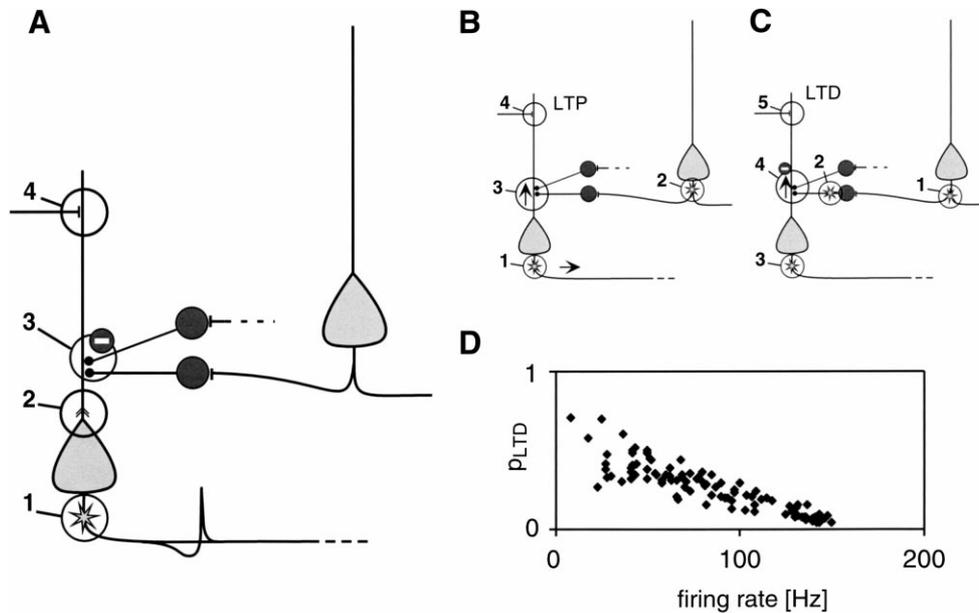


Fig. 2. The learning rule: (a) The diagram shows a small part of the network with events influencing synaptic plasticity. An action potential triggered in a pyramidal neurons at the axon hillock (1) propagates anterogradely and retrogradely (2). Synapses of inhibitory interneurons are located on the proximal dendrite (3) and affect the backpropagating action potential. The learning rule evaluates the coincidence of backpropagating and afferent action potential at the respective synapse (4). See text for details. (b) A series of events leads to an increase of synaptic efficacy in a pyramidal neuron. If an action potential is triggered in this neuron (1) before other neurons are active (2), it passes the inhibitory synapses at the proximal dendritic tree (3). Then at those locations where pre- and postsynaptic action potentials coincide (4) synaptic efficacy is increased (LTP). (c) The diagram shows a series of events leading to a decrease of synaptic efficacy in a pyramidal neuron. If other neurons in the network respond earliest to the stimulus (1), inhibitory neurons (2) are activated before or only shortly after the pyramidal neuron under consideration is activated (3). Thus, the inhibitory synapses attenuate or block the backpropagating action potential (4) and a decrease of synaptic efficacy in the distal synapses results (5). (d) The probability of inducing a decrease of synaptic efficacy is plotted against the firing frequency of the neuron.

Schiller, Stuart & Sakmann, 1995; Tsubokawa & Ross, 1996). Thus, the arrival of a retrogradely propagating action potential at the synapse gives information not only on the activation of the postsynaptic neurons, but also on the level of inhibitory activity in the network. Thus, the synaptic plasticity is contingent on locally available variables only, which, however, represent a global measure of network activity. In the following we formulate a learning rule, which captures these effects, and demonstrate how it addresses the performance criteria listed above.

## 2. Methods

### 2.1. The units and the network

We study the formation of receptive fields in a minimal model. It consists of excitatory, inhibitory and input units arranged in a 2-dimensional network (Fig. 1a). The activity of the input units follows a Poisson distribution and the instantaneous rate is set by the “visual” stimuli. Thus, they produce a noisy activity pattern with slowly varying density of spikes (Fig. 1b, upper row). The input units project to a square area in the next layer of neurons, the excitatory neurons. These connections are subject to the learning rule as described below. The excitatory units are

reciprocally coupled to the inhibitory neurons forming a negative feedback loop. Furthermore, the inhibitory neurons are coupled back onto themselves. For simulations, a conductance based leaky integrate and fire model was implemented in GENESIS. Synaptic conductance had a rise and fall time-course of 1 ms. Initial synaptic weights were chosen with a Gaussian profile and 5% noise level for the afferent connections and 10% noise level for all internal connections. All transmission delays were set to 3 ms.

### 2.2. The stimuli

The stimuli are similar to conventional “light bars” used in physiological experiments on the visual system (Hubel & Wiesel, 1962). Their luminance is constant in one direction and has a Gaussian profile in the orthogonal direction (Fig. 1a). In most simulations described below the stimuli are moving slowly across the map of input units. In some cases, where noted, static stimuli or stimuli with different geometrical properties were used.

### 2.3. The learning rule

For the learning rule we make the following assumptions: when an action potential is triggered at the axon hillock (Fig. 2a, 1) it propagates not only outwards along the axon, but additionally retrogradely through the dendrites

(Fig. 2a, 2). On its way into the dendrite, the action potential may be attenuated or throttled by inhibitory input from other neurons (Fig. 2a, 3). Changes in synaptic efficacy of the distal synapse are affected by the relative timing of pre- and postsynaptic action potentials (Fig. 2a, 4):

1. If the backpropagating action potential is coincident with the presynaptic action potential and its retrograde propagation is not impeded by inhibitory input within 3 ms (termed “winning”), then the efficacy of the respective synapse is increased (Fig. 2b).  $\Delta\omega_{LTP} = \alpha_{LTP}\tau/|\tau + \Delta t|$ , with  $\alpha_{LTP}$  a constant defining the learning rate,  $\tau = 5$  ms, and  $\Delta t$  the temporal difference of firing between pre- and postsynaptic neuron.
2. If the backpropagating action potential arrives is coincident with the presynaptic action potential, but it is attenuated by inhibitory input within 3 ms (termed “loosing”), then the efficacy of the respective synapse is decreased (Fig. 2c).  $\Delta\omega_{LTD} = \alpha_{LTD}\tau/|\tau + \Delta t|$ , with  $\alpha_{LTD}$  a constant defining the learning rate,  $\tau = 5$  ms, and  $\Delta t$  the temporal difference of firing between pre- and postsynaptic neuron.
3. Furthermore, to guarantee the stability of the system, a decay term proportional to the synaptic efficacy was included. Its effect is smaller than the primary effects of the learning rule by a factor of 20 and therefore is mostly modulatory (Horn Levy & Ruppin, 1998; Turrigiano, Leslie, Desai, Rutherford & Nelson, 1998).  $\Delta\omega_{Norm} = -\alpha_{Norm}\omega - \alpha_{Decay}$  with  $\alpha_{Norm}$  and  $\alpha_{Decay}$  constants defining the rate of weight decay.
4. For biological realism the weights were bound to be nonnegative.  $\omega = \max(\omega + \Delta\omega_{LTP/LTD} + \Delta\omega_{Norm}, 0)$ .

In summary, this learning rule acts as a “soft winner takes all” algorithm modulating the type of learning without directly interfering with the activity of the neurons.

#### 2.4. Data analysis

The performance of the learning rule is analyzed using several standard measures, time series diagrams (dot displays) and cross-correlation functions. Furthermore, we found it useful to look at the tuning of neurons in “stimulus space”. This allows a comprehensive representation of the tuning of all neurons. To specify any particular stimulus its orientation and position on a one-dimensional axis orthogonal to its orientation must be given. Thus, we can display the response of a neuron to all stimuli in a two dimensional graph with stimulus orientation and position as axes. Similarly, the coverage of all possible stimuli by neurons is displayed in such graphs.

### 3. Results

Here we proceed in several steps, starting with the basic dynamics of the neuronal network, and then investigating

the performance of the learning rule for sets of stimuli with increasing complexity.

#### 3.1. Network dynamics

Firstly, we characterize the basic properties of network and learning rule. Presenting a slowly moving “light bar” as a stimulus, which is typically used in physiological experiments, the activity of the input units can be characterized by a non-stationary Poisson process. The instantaneous firing frequency increases when the stimulus enters the receptive field of an input unit (Fig. 1b). Due to the negative feedback by the inhibitory units and with the parameters chosen, the network shows an oscillatory dynamics (Fig. 1b). Neurons firing at low rates show oscillatory activity at the same fundamental frequency, but miss individual cycles. Furthermore, within each oscillatory cycle their action potentials are systematically delayed with respect to the optimally stimulated neurons. The delay is of the order of few milliseconds, and thus is small compared to the period time of the oscillation. This property matches well experimental results (König, Engel, Roelfsema & Singer, 1995). In the present context it is decisive, as it translates differences in average firing rate on a time scale of hundreds of milliseconds into differences in timing of individual action potentials on a millisecond time scale. Thus, the relative strength of stimulation of different neurons is available locally in time.

The inhibitory units are driven by the excitatory units and, therefore, fire delayed. The phase shift is observed to be about 5 ms (Fig. 1c). Whether such a regular relationship between the firing of excitatory and inhibitory neurons exists in the cortex is not resolved. Nevertheless, preliminary data in favor of this hypothesis do exist (Gray, Engel, König & Singer, 1991). Within the context of the present model, the slightly delayed inhibition has the important function of separating “winning” and “loosing” neurons. The strongly stimulated neurons, firing at high rates, fire early within the oscillatory cycles, and thus, the probability of blocking the backpropagating action potential by the rising inhibition in the network is low (Fig. 2d). In contrast, in those neurons, which fire at moderate or low rates, the backpropagating action potential is blocked more often. As a result the synaptic efficacies are increased on those neurons responding best and earliest, and the respective neurons “learn” that stimulus even better. Other neurons may respond to the stimulus and contribute to its coarse coded representation (Lehky & Sejnowski, 1990), but may not further tune the pattern of their afferent synapses in favor of that stimulus. Thus, although for a given stimulus many neurons respond, the described learning rule implements a soft winner take all mechanism acting on the mechanisms of synaptic plasticity.

#### 3.2. Dynamic recruitment

Among the properties the network acquires with the

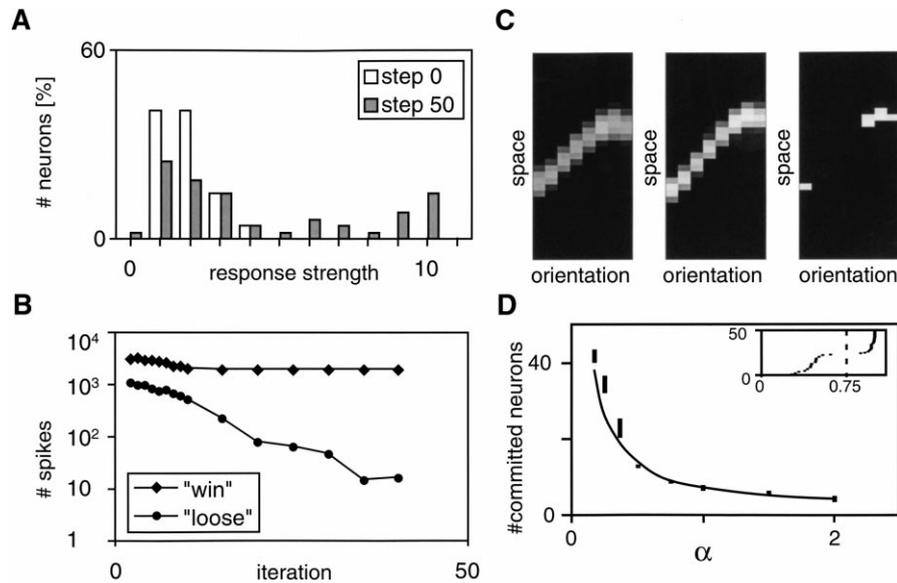


Fig. 3. Training with one stimulus: (a) The distribution of the strength of neuronal responses is shown in the initial condition (open bars) and after 50 stimulus presentations of one stimulus (gray bars). The horizontal bar indicates the range of response strength for square receptive fields with parameters as used in the simulation. (b) The number of spikes inducing LTP (diamonds) and LTD (circles) is shown for 50 stimulus presentations. (c) The induced response, averaged over all neurons, is shown on a gray scale for all stimuli. The stimuli are arranged according to their orientation and the position orthogonal to their long axis. On the left and in the middle the induced currents are shown in the initial state and after 50 stimulus presentations respectively. On the right the current has been multiplied with the I/F curve of the neurons to obtain the firing rate for each stimulus. (d) The number of committed neurons is shown for several values of  $\alpha_{LTD}$ . The mean  $\pm$  the standard error of the mean of the size of the representation is plotted as black bars. The red line shows the one-parameter fit of a hyperbola. The normalized scalar products of the receptive fields of the best tuned neuron and of all the others have a bimodal distribution as demonstrated by the cumulative distribution in the inset. This allows a robust definition of “committed” neurons. The dashed vertical line indicates the threshold of 0.75.

described learning rule, the number of neurons that represent one stimulus is of particular importance. Here an upper boundary to this number is determined. We start out with a network of neurons with unspecific receptive fields and study the effects of presenting the same stimulus over and over again. The trained stimulus is taken to have  $45^\circ$  orientation and is flashed on the input array along the main diagonal. Initially many neurons respond with moderate strength, as their unspecific receptive fields happen to be partly overlapping with the stimulus (Fig. 3a, open bars). After 50 stimulus presentations a limited number of neurons responds vigorously, whereas the remaining did not change the magnitude of their response (Fig. 3a, gray bars). These neurons have changed their receptive fields according to the stimulus properties. Upon each stimulus presentation they “win” and the number of spikes eliciting LTP at the synapses of the respective neurons stays constant (Fig. 3b, diamonds). After some iterations, the effective tuning does not change anymore and a steady state is reached. Indeed, also the number of neurons “loosing”, i.e. where active synapses are weakened, is reduced and approaches a steady state with very low numbers of losing neurons (Fig. 3b, circles). The development of receptive field properties of a tuned neuron is monitored in Fig. 3c. Here the response of one neuron is plotted on a gray scale for all stimuli. Initially (left panel) the membrane potential increases for stimuli of all orientation of a suitable location. After training, the response of the membrane potential (middle panel) is

more specific and spikes are triggered for a limited range of orientations only (right panel). It now covers an area in stimulus space where the stimulus selected for training is localized.

The size of the representation of a stimulus is determined by the parameters of the learning rule. In particular the ratio of the strength of LTP and LTD has a profound influence. We compare the size of the representation of a stimulus for different strength of the LTD effect, keeping the  $\alpha_{LTP}$  term constant. To determine the number of committed neurons, we plot the cumulative distribution function of the response strength. A bimodal distribution is found with a clear separation of weakly and strongly responding neurons (Fig. 3d, inset). The number of strongly responding neurons is determined in repeated runs of the simulation and the 95% confidence intervals are plotted in Fig. 3d. For small  $\alpha_{LTD}$  the effect of “loosing” is small, and thus a stimulus can support a large representation in the steady state (Fig. 3d). For larger  $\alpha_{LTD}$  the average size decreases approximately with the inverse of this parameter (Fig. 3d, red line). The dependence of the size of the representation on  $\alpha_{LTD}$  is well described by a hyperbola. In summary, it is possible to tune the desired size of the representation of a stimulus by the choice of the relative strength of LTP and LTD.

### 3.3. Decorrelation

As a next step we investigate the properties of the

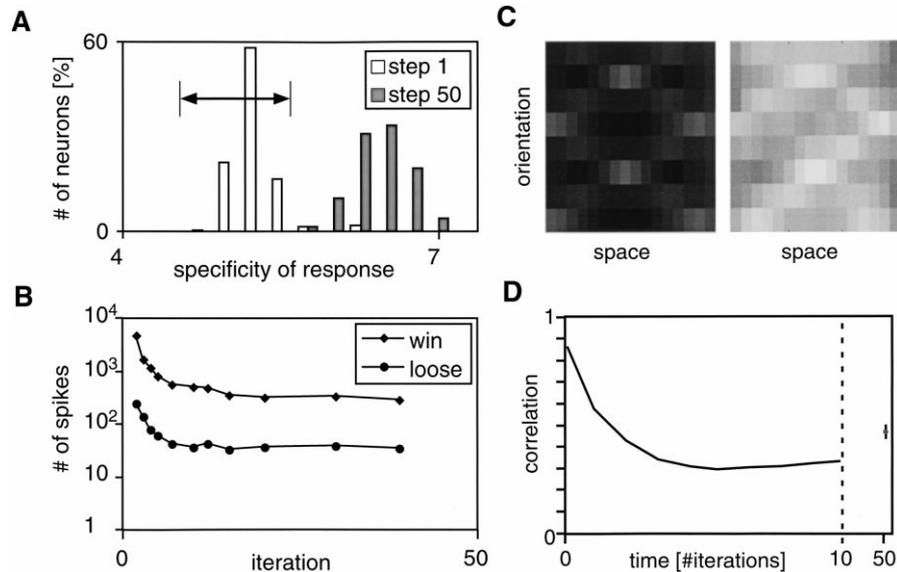


Fig. 4. Training with a set of stimuli with different orientations: (a) The distribution of the strength of neuronal responses is shown in the initial condition (open bars) and after 50 presentations of stimuli with different orientations (gray bars). The horizontal bar indicates the range of response strength for square homogeneous receptive fields. (b) The number of spikes inducing LTP (diamonds) and LTD (circles) is shown for 50 stimulus presentations. (c) The number of neurons responding specifically to a stimulus is coded on a gray scale from black to white. The initial state of the network is shown on the left. The stimuli are arranged according to their orientation and the position orthogonal to their long axis. After 50 stimulus presentations (right) all stimuli are represented by specifically tuned neurons. The residual inhomogeneities are due to finite size effects. On the border receptive fields are truncated giving rise to somewhat orientation selective receptive fields without training. On the other the number of neurons that can be recruited by a stimulus at high eccentricity is smaller than in the center. (d) The cross-correlation of receptive fields is shown over time. All receptive fields have been aligned and the pair-wise cross-correlation coefficient averaged over all pairs. The point to the right denotes the steady state value reached after about 50 iterations together with the standard deviation.

proposed learning rule when many different stimuli are presented. Stimuli with randomly chosen orientations are presented to the same network as above. Initially, due to the nearly isotropic structure of the receptive fields the specificity of the neuronal responses is low (Fig. 4a, open bars). As a consequence, few stimuli are represented by

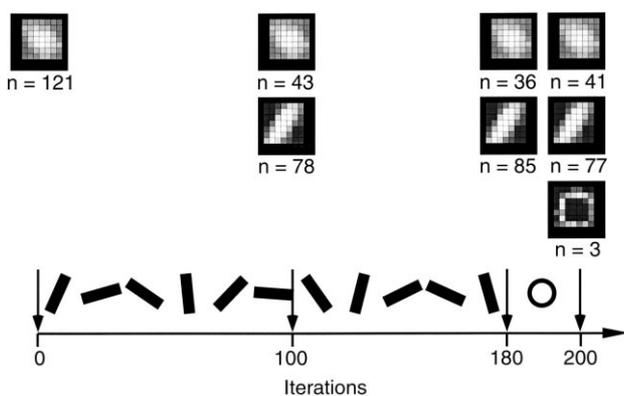


Fig. 5. Training with an inhomogeneous set of stimuli. On the horizontal axis examples of stimuli used are shown. During the first 180 iterations moving bars of random orientations were used. Then a circular stimulus is presented 20 times. The properties of the neuronal receptive fields are investigated at the start of the simulation and after 100, 180 and 200 iterations (arrows). At each time they are classified as either unspecific (upper row), or as committed to bars (middle row) or the ring (lower row). For each class a typical receptive field is shown together with the number of neurons in that class.

specific responses (Fig. 4c, left). These few specific responses are due to edge effects, where receptive fields are truncated and thus acquire orientation selective response. With increasing network size these effects diminish and become irrelevant. During the first few stimulus presentations the number of spikes generating LTP and LTD events is high. This implies that each stimulus shown has a large effect on the shaping of the receptive fields. After about 10 stimuli, however, the system settles into a steady state and the number of spikes inducing a strengthening or weakening of synapses is constant on a low level. The specificity of the neurons is significantly increased (Fig. 4b, gray bars). As a next step we investigate the tuning of these specific neurons (Fig. 4c, right). Their receptive fields cover the whole stimulus space rather homogeneously. This implies that all possible stimuli will elicit a response of a comparable number of specifically tuned neurons. As the stimuli shown are only weakly correlated (the correlation is not zero because many stimuli share the central part) this property is transferred to the distribution of receptive fields. In Fig. 4d the temporal development of the correlation of receptive fields is shown. Initially all neurons have circular receptive fields and the correlation of their receptive fields is therefore high. After only 20 stimulus presentations the correlation has drop significantly. The residual correlation is due to the center of the receptive field which is common to all stimuli, and thus to all receptive fields.

### 3.4. Combined effects

Here we show that the effects of dynamic recruitment and decorrelation of receptive fields may coexist in a neuronal network. First randomly oriented moving bars are used as stimuli as above. After presentation of 100 stimuli, a set of neurons has acquired specific receptive fields (Fig. 5). For the choice of  $\alpha_{LTP}$  and  $\alpha_{LTD}$  a representation of each stimulus by about four neurons is expected. Given that the orientation of the stimulus may take any value in the interval between 0 and 180° a direct comparison is not possible. However, taking into account that the orientation tuning of the neurons is about 20° and the retinal area is about twice the size of a receptive field the 78 specific neurons are representing about  $(180/20) \times 2 = 18$  “different” stimuli. Thus, the data observed  $78/18 = 4.33$  fit the expectation of four neurons per stimulus well enough. Another set of neurons remains unspecific, i.e. is not orientation tuned. Presenting oriented bars 80 more times has little influence on the tuning properties of the neurons. Introducing a new stimulus, a ring, a number of neurons, which had previously unspecific receptive fields, is recruited and develops receptive fields matching the ring shaped stimulus. Again, the recruitment of three neurons is compatible with the expectations given by the choice of the relative strength of LTP and LTD. These studies demonstrate that decorrelation and dynamic recruitment are properties of the proposed learning rule, which can be exploited simultaneously in the development of a neuronal network.

## 4. Discussion

In this study we investigate the properties of a learning rule which is inspired by physiological results. Firstly, these studies demonstrate that when an action potential is triggered at the axon hillock (Fig. 2a, 1) it propagates not only anterogradely along the axon, but additionally retrogradely through the dendrites (Stuart & Sakmann, 1994). Secondly, on its way into the dendrite the action potential may be attenuated or throttled by inhibitory input from other neurons (Tsubokawa & Ross, 1996). Recently, the direct interaction of the backpropagating action potential, inhibitory input and the calcium dynamics in the apical dendritic tree has been demonstrated (Larkum et al., 1999). And finally, they suggest that changes in synaptic efficacies are affected by the relative timing of pre- and postsynaptic action potentials (Gerstner et al., 1993; Markram et al., 1997). The learning rule that formalizes these results leads to several desirable properties. First, it leads to a decorrelation of receptive field properties. Second, new stimuli are learned within a few iterations. Third, it leads to stable representations during ongoing learning. It might be viewed as an instance, where the incorporation of physiological results into a neural network simulation leads to interesting new insights.

Nevertheless, for the sake of clarity and ease of implementation several simplifications were used in the present work.

Most notably, we neglected in this paper the effect of the sequence of pre- and postsynaptic potentials (Markram et al., 1997). In these in vitro experiments the synaptic efficacy was increased only if the presynaptic action potential preceded the action potential in the postsynaptic neuron. Here we use a symmetric window and the synapse is strengthened as long as the two action potentials co-occur within a small temporal window. Thus, the sequence of pre and postsynaptic action potentials is not relevant. It should, however, be noted that the activity of the presynaptic neurons is described by a Poisson process. Thus, there is no information in the precise temporal structure of the afferent spike train and we did not observe any effect of including an asymmetric temporal window (data not shown). Furthermore, this phenomenon can not replace the mechanism proposed here. In all neurons, which are activated by a stimulus, whether strongly or moderately so, some afferent action potentials have to precede the postsynaptic action potential, and thus, will be learned. Therefore, an asymmetric window for the coincidence of pre- and postsynaptic action potentials cannot replace the information whether the neuron under consideration is among those responding best and earliest to the stimulus or not. Interestingly, the mechanism analyzed here is closely related to the rank-order coding scheme used by (Van Rullen et al., 1998). In this simulation the spike timing carries the information on the visual stimulus and allows rapid processing in a feed-forward network. Extending the learning rule analyzed here with an asymmetric window, i.e. making synaptic plasticity dependent on the sequence of pre- and postsynaptic action potentials, might lead to interesting applications in such a network architecture.

We are working with a set of parameters leading to oscillatory activity in the network. This establishes a reference frame making the distinction between early and late activity unambiguous (Hopfield, 1995). However, it does not seem to be a necessary condition. In the visual cortex of the cat a systematic lag of suboptimally driven neurons with respect to optimally driven neurons was found which was not tied to an oscillatory process (König et al., 1995). For a more extensive discussion see also Wennekers and Palm (1999).

In its original formulation a Hebb-type learning rule leads to a positive feedback with a run-away increase of synaptic efficacy. To solve this problem several mechanisms have been suggested: the average activity of the postsynaptic neuron defines a sliding threshold separating regions of LTP and LTD (Bienenstock et al., 1982). This leads to a normalization of each neuron’s average firing rate. In recent years experimental support for this hypothesis has become available (Bear, 1997; Kirkwood, Rioult & Bear, 1996). An alternative is to normalize the total afferent synaptic weight of each neuron (Horn et al., 1998), a mechanism which has found support by in vitro experiments (Turrigiano et al.,

1998). Here, we have chosen the second mechanisms for ease of implementation. Furthermore, these mechanisms are not exclusive and we would be surprised if not both as well as other options are used by the biological system.

Given that one of the major reasons to investigate the properties of neuronal networks is that they may be trained, the wealth of work on different types of learning rules is not surprising. Furthermore, as the features listed in the introduction have been individually addressed by previous work, why do we need yet another learning rule?

One of the major problems of existing learning rules is that different neurons tend to develop identical properties. Thus their activity becomes correlated and progressively less information is transferred to successive structures (Nadal, Brunel & Parga, 1998). A decorrelation of the receptive fields by competitive mechanisms between different neurons is therefore often used in artificial neural networks (Barlow, 1989; Hemmen 1990). However, no commonly accepted scheme to achieve that purpose using a physiologically plausible learning rule exists. Neurons may interact by lateral connections which are subject to an anti-Hebbian learning rule (Barlow & Földiak, 1989; Rubner & Schulten, 1990). The inhibitory effects of these connections forces neurons to develop decorrelated patterns of afferent synapses, which are learned according to a Hebb-type learning rule. In the present work the tangential connections are themselves not subject to synaptic plasticity, but they influence modifications of the afferent synapses. Thus, a comparable effect is achieved with a physiologically plausible mechanism.

Another aspect of interest is the spatial arrangement of neurons with different response properties. For many years it is known that neurons with similar response properties are clustered in cortex. With modern imaging methods topographic structure like bands/blobs/stripes/pinwheels have been demonstrated and a number of studies addressed these phenomena (Goodhill, 1997). Many of the different learning rules proposed are actually derived from similar principles (Erwin, Obermayer & Schulten, 1995). For example, the competition for a substrate together with the use of local correlation (Miller, Keller & Stryker, 1989) leads to patterns as those observed in the mammalian visual cortex. Nevertheless, a number of problems remain, e.g. the relation of patterns determined by different features such as orientation vs. ocular dominance (Erwin et al., 1995), and the global arrangement of these structures within a cortical area (Wolf, Bauer, Pawelzik & Geisel, 1996). In the present work, we did not address these problems and the relation of the proposed learning rule to other mechanism governing the topographic order in cortical areas is left for future work.

In the study of artificial neuronal networks “network growing methods” have been found useful. They may control hidden layer capacity to achieve optimal fitting and generalization. An interesting closely related line of research is the Adaptive-Resonance-Theory by Grossberg and his coworkers (Carpenter, Cohen & Grossberg, 1987;

Grossberg, 1975). Here a comparable mechanism is implemented on an abstract level, which allows to check whether a stimulus is already presented (Resonance) and if not, to establish a new representation upon a single presentation. In comparison, the learning rule proposed here is formulated in more physiological terms and does not “create” new cells on the fly, however it dynamically recruits previously unspecific neurons for the representation of new stimuli. As such, it might establish a physiological basis for the Adaptive-Resonance-Theory or related “network growing” methods.

In conclusion, a model is presented in which a single learning rule provides a solution to several central problems of learning in neural networks architectures. When applied to the afferent connections it leads to decorrelation of receptive field and dynamic recruitment of neurons by new stimuli, resulting in efficient representation. Thus, the usual tradeoff between speed of learning and memory capacity does not apply, and both stability and one shot learning are combined.

## Acknowledgements

It is a pleasure to thank Horace Barlow, Michele Giugliano, Richard Hahnloser, Christoph Rasche, Walter Senn, Leo van Hemmen, Paul Verschure and Adrian Whatley for comments on a previous version of this manuscript. This work has been supported by the Swiss National Science Foundation, the Boehringer Ingelheim Fund and SPP Neuroinformatics.

## References

- Artola, A., Bröcher, S., & Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, *347*, 69–72.
- Atick, J., & Redlich, A. (1990). What does the retina know about natural scenes? *Neural Computing*, *4*, 196–210.
- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1*, 295–311.
- Barlow, H. B., & Földiak, P. (1989). Adaptation and decorrelation in the cortex. In R. Durbin & C. Miall & G. Mitchison (Eds.), *The computing neuron*, New York: Addison-Wesley.
- Barlow, H., Kaushal, T., & Mitchison, G. (1989). Finding minimum entropy codes. *Neural Computation*, *1*, 412–423.
- Bear, M. F. (1997). How do memories leave their mark? *Nature*, *385*, 481–482.
- Bienenstock, E., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*, 32–48.
- Buhmann, J. (1989). Oscillations and low firing rates in associative memory neural networks. *Physical Review A*, *40*, 4145–4148.
- Carpenter, G. A., Cohen, M. A., & Grossberg, S. (1987). Computing with neural networks. *Science*, *235*, 1226–1227.
- Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*, 287–314.
- Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, *16*, 3351–3362.

- Deco, G., & Obradovic, D. (1996). *An information-theoretic approach to neural computing*. New York: Springer.
- Erwin, E., Obermayer, K., & Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Computation*, 7, 425–468.
- Gerstner, W., Ritz, R., & van Hemmen, J. L. (1993). Why spikkes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biological Cybernetics*, 69, 503–515.
- Goodhill, G. J. (1997). Stimulating issues in cortical map development. *Trends in Neuroscience*, 20, 375–376.
- Gray, C. M., Engel, A. K., König, P., & Singer, W. (1991). Synchronous neuronal oscillations in the cat visual cortex: functional implications. In A. Gorea & Y. Fregnac & Z. Kapoula (Eds.), *Trends and tacit assumptions in vision research*, (pp. 83–96). Cambridge: Cambridge University Press.
- Grossberg, S. (1975). A neural model of attention, reinforcement and discrimination learning. *International Review of Neurobiology*, 18, 263–327.
- Hebb, D. O. (1949). *The organization of behaviour*. New York: Wiley.
- Hemmen, J. L. V. (1990). Hebbian learning and unlearning. In W. K. Theuman (Ed.), *Neural networks and spin glasses*, (pp. 91–114). Singapore: World Scientific.
- Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376, 33–36.
- Horn, D., Levy, N., & Ruppin, E. (1998). Memory maintenance via neuronal regulation. *Neural Computation*, 10, 1–18.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Katz, L. C., & Callaway, E. M. (1992). Development of local circuits in mammalian visual-cortex. *Annual Review of Neurosciences*, 15, 31–56.
- Kirkwood, A., Rioult, M. G., & Bear, M. F. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, 381, 526–528.
- König, P., Engel, A. K., Roelfsema, P. R., & Singer, W. (1995). How precise is neuronal synchronization? *Neural Computation*, 7 (3), 469–485.
- Lehky, S. R., & Sejnowski, T. J. (1990). Neural model of stereoactivity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neurosciences*, 10, 2281–2299.
- Magee, J., Hoffman, D., Colbert, C., & Johnston, D. (1998). Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. *Annual Review of Physiology*, 60, 327–346.
- Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275, 213–215.
- Miller, K. D., Keller, J. B., & Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science*, 245, 605–615.
- Nadal, J. P., Brunel, N., & Parga, N. (1989). Nonlinear feedforward networks with stochastic outputs: infomax implies redundancy reduction. *Network*, 9 (2), 207–217.
- Rubner, J., & Schulten, K. (1990). Development of feature detectors by self-organization. *Biological Cybernetics*, 62, 193–199.
- Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4 (4), 303–321.
- Singer, W. (1986). The brain as a self-organizing system. *European Archives in Psychiatry and Neurological Sciences*, 236, 4–9.
- Spruston, N., Schiller, Y., Stuart, G., & Sakmann, B. (1995). Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science*, 268 (5208), 297–300.
- Stent, G. S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proceedings of National Academy of Sciences (USA)*, 70, 997–1001.
- Stuart, G. J., & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367, 69–72.
- Tsubokawa, H., & Ross, W. N. (1996). IPSPs modulate spike backpropagation and associated  $[Ca^{2+}]_i$  changes in the dendrites of hippocampal CA1 pyramidal neurons. *Journal of Neurophysiology*, 76 (5), 2896–2906.
- Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391, 892–896 see comments.
- Wendemuth, A., & Sherrington, D. (1993). Fast learning of biased patterns in neural networks. *International Journal of Neural Systems*, 4 (3), 223–230.
- Wennekers, T., & Palm, G. (1999). How imprecise is neuronal synchronization? *Neurocomputing*.
- Wolf, F., Bauer, H. U., Pawelzik, K., & Geisel, T. (1996). Organization of the visual cortex. *Nature*, 382, 306.