Hierarchical Self-Organization of Minicolumnar Receptive Fields

Jörg Lücke
Institut für Neuroinformatik
Ruhr-Universität Bochum
D-44780 Bochum, Germany
Ph.: +49 234 32 27921 , Fax: +49 234 32 14210
luecke@neuroinformatik.rub.de

Acknowledgments. I thank R. P. W¨urtz for many discussions and suggestions and I thank J. D. Bouecke for his help in acquiring simulation results. Further, partial funding by the EU in the RTN MUHCI (HPRN-CT-2000-00111) and the German BMBF in the project LOKI (01 IN 504 E9) is gratefully acknowledged.

Abstract

We study self-organization of receptive fields (RFs) of cortical minicolumns. Input driven self-organization is induced by Hebbian synaptic plasticity of afferent fibers to model minicolumns based on spiking neurons and background oscillations. If input in the form of spike patterns is presented during learning, the RFs of minicolumns hierarchically specialize to increasingly small groups of similar RFs in a series of nested group subdivisions. In a number of experiments we show that the system finds clusters of similar spike patterns, that it is capable of evenly cover the input space if the input is continuously distributed, and that it extracts basic features from input consisting of superpositions of spike patterns. With a continuous version of the bars test we, furthermore, demonstrate the system's ability to evenly cover the space of extracted basic input features. The hierarchical nature and its flexibility with respect to input distinguishes the presented type of self-organization from others including similar but non-hierarchical self-organization as discussed in (L¨ucke and von der Malsburg, 2004). The capabilities of the presented system match crucial properties of the plasticity of cortical RFs and we suggest it as a model for their hierarchical formation.

Keywords: Cerebral Cortex, Self-Organization, Columnar Organization, Synaptic Plasticity, Unsupervised Learning, Cortical Maps

List of Symbols

n_i^α	activity of neuron i in minicolumn α
n_i^E	activity of input neuron i
T_{ij}^α	internal connection from neuron j to i in minicolumn α
R_{ij}^α	afferent fiber from input neuron j to neuron i of minicolumn α
$\vec{R}^\alpha$	receptive field of minicolumn α
$\vec{I}$	input vector
s	number of internal connections received by a neuron
r	number of afferent fibers received by a neuron
c	synaptic weight of an internal connection
$\tilde{c}$	synaptic weight of an afferent fiber
κ	ratio between internal and external input
k	number of minicolumns
m	number of neurons per minicolumns
N	number of input neurons
p_α	mean neural activity of minicolumn α
P	mean neural activity of the macrocolumn
$\mathcal{I}$	inhibition between minicolumns
Θ_o	the neurons' constant threshold
Θ_{noise}	neuronal threshold noise
ν	gain factor of intercolumnar inhibition
ν_c	critical value of ν
T	period length of a ν-cycle
χ	threshold for synaptic plasticity
a_χ, b_χ	scalar parameters of χ dynamics
$\langle\cdot\rangle_\tau$	time average over ν-cycles
λ	decay rate of χ
$\mathcal{E}$	synaptic growth factor
b	number of bars in the bars test
p_o	appearance probability of a bar in the bars test

1 Introduction

Self-organization is believed to play a central role in the development of the central nervous system of higher vertebrates. The most prominent example is probably the ontogenesis of fiber projection systems and especially the development of retinotectal maps (H¨aussler and von der Malsburg, 1983). Other examples of self-organization deal with structures in the barrel field of the rodent cortex or with the formation of ocular dominance and orientation columns (see Erwin et al., 1995; Swindale, 1996; von der Malsburg, 2003, for overviews). The idea that self-organization is also responsible for the formation of interconnection structures in advanced levels of cortical development has become increasingly popular in the last decades and gave birth to abstract neural network models based on input driven self-organization, e.g. (von der Malsburg, 1985; Kohonen, 1995; Fritzke, 1995). In recent years the development of intra-cortical connections could be monitored in more and more detail especially in primary sensory areas. In the primary visual cortex, e.g., interconnections between orientation columns were shown to develop hierarchically from coarsely specialized to increasingly fine tuned RFs (Callaway and Katz, 1990; Chapman et al., 1996).

In this paper we discuss a self-organization process which reflects the hierarchical nature of RF formation. The model is based on a system presented in (L¨ucke and von der Malsburg, 2004) and integrates fundamental aspects of neural information processing such as Hebbian type synaptic plasticity, columnar interconnection structure, spiking neurons, and background oscillation. In a number of experiments we investigate the system's abilities to form appropriate RFs for very different types of input. The type of self-organization, hereby, differs from other neural models by its hierarchical approach and by its high flexibility with respect to the input.

In Sec. 2 we shortly define and discuss the neural dynamics of a model macrocolumn. In Sec. 3 we define the dynamics of Hebbian plasticity of afferent fibers to the macrocolumn and discuss the emerging self-organization. In Sec. 4 the system is applied to different types of input and in Sec. 5 we discuss the system's relation to neuroscience and compare it with other unsupervised neural models.

2 Neural Dynamics of a Model Macrocolumn

Our system is based on a model of the cortical macrocolumn or segregate (Favorov and Diamond, 1990). Macrocolumns can be identified both anatomically and physiologically (Favorov and Diamond, 1990; Elston and Rosa, 2000; Lubke et al., 2000) and are shown to process stimuli from the same source such as an area of the visual field or a patch of a the body surface (Favorov and Whitsel, 1988; Favorov and Diamond, 1990). The macrocolumn consists of a number of minicolumns[1] which present the smallest neural modules and consist of several tens up to a few hundred neurons (Peters and Yilmaz, 1993) (see Mountcastle, 1997; Buxhoeveden and Casanova, 2002, for overviews and Jones, 2000, for a critical discussion).

[1] sometimes also called microcolumns

In this section we consider the dynamics of a model macrocolumn without Hebbian plasticity. This part of the dynamics is essentially the same as the one presented in (L¨ucke and von der Malsburg, 2004), except of two technically important differences: First, intermediate spike rates of input neurons (see below) can encode grey-level instead of only black-and-white input images and, second, the system can be operated with a variable number of input neurons and, hence, variable size of input images.

The macrocolumn is modeled as a collection of disjunct minicolumns. The minicolumns we take to consist of a collection of m excitatory neurons. As neuron model we use a McCulloch-Pitts neuron with refraction time:

$$n_i^\alpha(t+1) = \mathcal{H}(\sum_{j=1}^{m} T_{ij}^\alpha n_j^\alpha(t) \;+\; \sum_{j=1}^{N} R_{ij}^\alpha n_j^E(t) \;-\; \Theta(t)) \cdot \mathcal{H}(1 - n_i^\alpha(t))\,, \qquad (1)$$

where $\mathcal{H}(x)$ is the function, $\mathcal{H}(x) = 0$ for $x \leq 0$ and $\mathcal{H}(x) = 1$ for $x > 0$. The neuron $i \in \{1, \ldots, m\}$ of minicolumn $\alpha \in \{1, \ldots, k\}$ is active at time $(t + 1)$ if it receives enough input from neurons active at time step t and if it was not active the time step before. To activate the neuron the input has to be larger than a threshold $\Theta(t)$ which consists of a usual constant threshold Θ_o with Gaussian threshold noise Θ_{noise} and a time depending function $\mathcal{I}(t)$ which indirectly models inhibition between the minicolumns:

$$\Theta(t) \;=\; \Theta_o + \Theta_{\text{noise}} + \mathcal{I}(t) \qquad (2)$$

$$\mathcal{I}(t) \;=\; \nu \max_{\alpha=1,\ldots,k} \{p_\alpha(t)\} \qquad (3)$$

In (3) $p_\alpha(t) = \frac{1}{m}\sum_{i=1}^{m} n_i^\alpha(t)$ is the over-all activity[2] of minicolumn α. All excitatory neurons are modeled in this way and possess the same constant threshold and a threshold noise with the same variance. A neuron receives excitation from two sources: First, from other excitatory neurons of the same minicolumn, and second from neurons outside the macrocolumn, denoted n_j^E. The interconnection within a minicolumn (T_{ij}^α) we take to be fixed and random. The randomness is generated with the boundary conditions that each neuron receives a number of s synapses with synaptic weight $c = \frac{1}{s}$ and that the probability of a synapses to originate from a specific neuron of the minicolumn is constant. The afferents from external neurons (R_{ij}^α) can change in time but we require that a neuron receives r synapses of weight $\tilde{c} = \frac{w}{s}$ from external neurons. The ratio between the number of external and the number of internal connections[3],

$$\kappa \;=\; \frac{\sum_{j=1}^{N} R_{ij}^\alpha}{\sum_{j=1}^{m} T_{ij}^\alpha}\,, \qquad (4)$$

is taken to be significantly smaller than one which means that the internal input of an active minicolumn is on average significantly larger than the external input.

[2]The inhibition has to be understood as being implemented by a system of inhibitory neurons rather than by a single one. A larger number of inhibitory neurons can be expected to average over incoming activity and to distribute inhibition such that (3) is approximated.

[3]Note that due to the required boundary conditions κ is independent of i and α

The external neurons n_j^E are essentially of the same type as the neurons of the macrocolumn, i.e., they are either inactive or active, $n_j^E(t) \in \{0,1\}$, and they are refractory for one time step after they have spiked. The external neurons translate an input vector $\vec{I} \in [0,1]^N$ into spike rates. A neuron n_j^E has a probability of $\frac{1}{3}I_j$ to be active. The N dimensional vector of all afferent fibers to neuron n_i^α, $\vec{R}_i^\alpha = (R_{i1}^\alpha, \ldots, R_{iN}^\alpha)$, will be called the receptive field (RF) of the neuron and the sum of the RF vectors of all neurons of a minicolumn α will be called the RF of the minicolumn, $\vec{R}^\alpha = \sum_{i=1}^m \vec{R}_i^\alpha$ (see Fig. 1B,C). In Fig. 1A a sketch of a macrocolumn with $k = 3$ minicolumns is displayed. The input neurons are arranged in a two dimensional array because we will use two dimensional grey-level images as input vectors to conveniently display simulation results. Note, however, that the set of input neurons is not equipped with any neighborhood relationship and can therefore be arranged in an arbitrary manner.

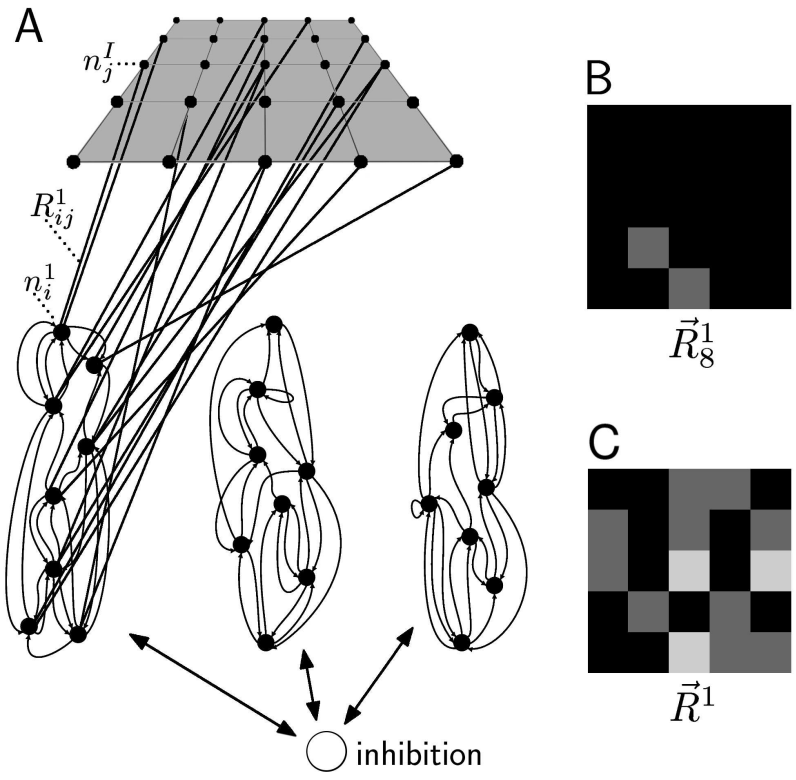

Figure 1: **A** Sketch of a macrocolumn with $k = 3$ minicolumns, parameters $m = 8$, $s = 3$, and $r = 2$, connected to $N = 25$ external neurons. The inhibition is symbolically sketched. The randomly initialized RF, $\vec{R}^1$, of minicolumn $\alpha = 1$ is fully displayed whereas RFs $\vec{R}^2$ and $\vec{R}^3$ are not. Lines within the input layer are only displayed for visualization purposes. **B** Two dimensional visualization of the RF of one neuron in minicolumn $\alpha = 1$ as grey-level image. **C** Two dimensional visualization of $\vec{R}^1$ as grey-level image.

The difference equation system (1) together with (2) and (3) defines the neural

activity dynamics of the macrocolumn. Although analytical predictions of the behavior of this highly non-linear system seem difficult, it could be shown in (L¨ucke and von der Malsburg, 2004) that, for a large class of parameters, the global and essential properties of the system can be derived analytically by means of stability analysis and bifurcation theory[4]. We operate the system by oscillating the inhibitory gain factor[5] ν in (3):

$$\nu(t) = \begin{cases} \nu_{\text{init}} & \text{if } \tilde{t} < T_{\text{init}} \\ (\nu_{\max} - \nu_{\min})\frac{\tilde{t}-T_{\text{init}}}{T-1-T_{\text{init}}} + \nu_{\min} & \text{if } \tilde{t} \geq T_{\text{init}} \,, \end{cases} \tag{5}$$

where T_{init} is the length of an initialization phase and $\tilde{t} = t(\mathrm{mod}\, T)$ the remainder of division by the oscillation's period length T. During one saw tooth like period, which will be called *ν-cycle* from now one, the system is forced to switch off minicolumns with comparably weak input. This process, which begins anew for each ν-cycle, can be described in terms of structural instability and we refer to (L¨ucke et al., 2002) and (L¨ucke and von der Malsburg, 2004), for further details[6] .

3 Self-Organizing Receptive Fields

The afferent fibers (R^{α}_{ij}) which couple the neuron dynamics (1) to external input in the form of spike patterns we take to be subject to synaptic plasticity. The change of an afferent fiber, $\Delta R^{\alpha}_{ij}(t) = R^{\alpha}_{ij}(t) - R^{\alpha}_{ij}(t-1)$, is described by a Hebbian type dynamics with normalization:

$$\text{If } P(t) < \chi(t)\text{:} \quad \Delta R^{\alpha}_{ij}(t) \;=\; A^{\mathcal{E}}_{t}\, n^{\alpha}_{i}(t)\, n^{E}_{j}(t-1) \tag{6}$$

$$\forall i, \alpha : \frac{1}{\tilde{c}} \sum_{j=1}^{N} R^{\alpha}_{ij}(t) \;=\; r \,. \tag{7}$$

Here $P(t)$ is the over-all activity in the macrocolumn, $P(t) = \frac{1}{km}\sum_{\alpha,i} n^{\alpha}_{i}(t)$, and $\chi(t)$ is a dynamical threshold. $A^{\mathcal{E}}_{t}$ is the probability that R^{α}_{ij} is increased by $\tilde{c} = \frac{w}{s}$ if neuron n^{α}_{i} fires in the time-step proceeding the firing of input neuron n^{E}_{j}. $A^{\mathcal{E}}_{t}$ is a for each (i, j, α) triple different Bernoulli sequence with probability $W(0) = 1 - \mathcal{E}$

[4]For the analysis we first replace dynamics (1) by a dynamics which approximately describes the minicolumn activities p_{α}. The approximation is possible because of sufficiently high activity rates in active minicolumns near to bifurcation points. With a subsequent stability analysis the stationary points, which are actually observed in directly simulating (1), can be computed. Because of a small κ (4) input mediated by afferent fibers is than treated as perturbation of the internal dynamics.

[5]see, e.g., (K¨orner et al., 1999) for a discussion of a possible origin of such an oscillation

[6]Note, that the system behaves qualitatively similar also for refraction times larger than in (1). The main differences are that minicolumn activities are lower and due to a larger impact of threshold noise hysteresis is less pronounced, i.e., activation of an in the same ν-cycle already switched off minicolumn is more probable. Crucial for the functioning of the system is the non-linear deactivation of minicolumns with weak input induced by an interplay between refraction, excitation, and inhibition, which can also be expected from a system based on continuous neuron models.

and $W(\tilde{c}) = \mathcal{E}$. The parameter $\mathcal{E}$ is the synaptic growth factor. If the afferent fiber R_{ij}^{α} is increased for a given neuron n_i^{α}, one randomly chosen (non-zero) fiber $R_{ij'}^{\alpha}$ is removed in order to fulfill the normalization condition (7). The main difference to a usual Hebbian learning rule is the condition $P(t) < \chi(t)$ which has to be fulfilled in order to enable synaptic modification. If the condition is not satisfied, the afferent connections remain unchanged. The condition ensures that modification is only allowed for states of low macrocolumnar activity which are predominant in the end of a ν-cycle. Only than, after what we will call *the minicolumn selection process*, the macrocolumnar activity is reflecting properties of the input.

Equations (6) and (7) represent an abstract formulation of synaptic plasticity, which is suitable for the time-discrete neuron model. More realistic forms of synaptic plasticity would have to be based on continuous neuron models and differential equations for synaptic modification. An interplay between long-term potentiation (LTP) and long-term depression (LTD) such as spike-timing dependent plasticity (STDP) can be imagined to implement renormalization similar to (7) and a plasticity similar to (6) including an approximation of condition $P(t) < \chi(t)$. Action potentials (spikes) in minicolumns which remain active for higher inhibition have an increased probability to proceed spikes of input neurons because the incoming spikes must have positively contributed to keep the column active. In the case of low inhibition LTD can dominate because of in this case less correlated firing of minicolumn and input neurons (see Feldman, 2000, for the predominance of LTD for uncorrelated spikes). Synaptic scaling processes (see, e.g., Abbott and Nelson, 2000; Turrigiano et al., 1998) without or in combination with LTP and LTD can also be considered for implementing (6) and (7). The in our model essential requirement for synaptic plasticity to result in discriminating RFs is the predominance of LTP for low levels of macrocolumnar activity. Equations (6) and (7) with condition, $P(t) < \chi(t)$, can in this respect be regarded as an abstract and simple dynamical description satisfying this requirement.

To train the macrocolumn network we randomly choose an input pattern $\vec{I}$ from a given database and present it in the form of input neuron spike patterns for the duration of one ν-cycle. If learning starts with randomly initialized RFs, the selection process during a ν-cycle first randomly deactivates a subset of minicolumns by spontaneous symmetry breakings. If in this way the macrocolumn activity $P(t)$ falls below the threshold $\chi(t)$, the RFs of the remaining active minicolumns specialize to a certain degree to the presented input pattern. If in a later ν-cycle a similar input pattern is shown, these minicolumns are more likely to remain active again and to further specialize. The RF specialization represents, therefore, the positive feed-back loop of the self-organization process which amplifies small fluctuations in the beginning of learning. Together with the competition of the RFs mediated by the minicolumn selection process, the RFs tend to specialize to different patterns of the input database. A crucial role is played by the threshold χ. If χ is relatively high, e.g., $\frac{1}{3}$, the minicolumnar RFs start to modify already in the middle of the selection process and large groups of minicolumns tend to form similar RFs. If χ is relatively small, the RFs avoid specializing for the same input patterns. If we start learning with large χ which is gradually decreased, the minicolumns first form

large groups with mutually similar RFs, i.e., RFs of minicolumns of the same group are more similar to each other than to RFs of other groups. If χ gets smaller, the RFs of a group specialize to different sub-patterns of the input patterns they have already been specialized to. The threshold χ can therefore be interpreted as a kind of reciprocal differentiation pressure.

χ is only updated at time steps $\tau \in \{T-1, 2T-1, \ldots\}$ at the end of each ν-cycle. For a time-depending function f let us first define the average over the last A ν-cycles:

$$\langle f \rangle_\tau \;:=\; \frac{1}{A} \sum_{a=0}^{A-1} f(\tau - aT) \tag{8}$$

At the beginning of learning, the system starts with large χ, $\chi(0) = \chi_o = \frac{1}{3}$, which is subsequently decreased by the dynamics[7]:

$$\chi(\tau+1) \;=\; \chi(\tau) - \lambda(\tau)\,(\chi(\tau) \;-\; a_\chi\, P(\tau))\,, \tag{9}$$

$$\lambda(\tau) \;=\; \lambda_o \left(\frac{\min^*_\beta \{\langle p_\beta \rangle_\tau\}}{\langle P \rangle_\tau} - b_\chi \right) . \tag{10}$$

where a_χ and b_χ are scalar parameters and $\tau \in \{T-1, 2T-1, \ldots\}$ are time-steps which always mark the end of a ν-cycle (with period T). Only at these time steps χ and λ are updated. The function $\min^*_\beta \{\ldots\}$ is an average over the least active minicolumns (see below). For small and positive λ, χ is monotonously decreased until it stabilizes around the time average of $(a_\chi\, P(\tau))$. The velocity with which the threshold χ decreases and hence the velocity the differentiation of RFs increases is controlled by λ which itself is dependent on the minicolumn activity averages at the end of the past ν-cycles. The time dependent λ slows down the differentiation process if some minicolumns are repeatedly quiescent in the minicolumn selection process and start to cease their RF specialization. For a small number of minicolumns $\min^*_\beta \{\ldots\}$ in (10) can be the usual minimum, for arbitrary k it has turned out, however, that the average over all minicolumn activities smaller than a quarter of the macrocolumn activity $P(\tau)$ gives better results:

$$L_\tau \;=\; \left\{ \beta \mid \langle p_\beta \rangle_\tau \;<\; \tfrac{1}{4}\, P(\tau) \text{ or } \langle p_\beta \rangle_\tau \;=\; \min_\gamma \{\langle p_\gamma \rangle_\tau\} \right\} \tag{11}$$

$$\min^*_\beta \{\langle p_\beta \rangle_\tau\} \;=\; \tfrac{1}{|L_\tau|} \sum_{\beta \in L_\tau} \langle p_\beta \rangle_\tau \tag{12}$$

Hebbian plasticity (6) and (7) together with time-depending and self-controlled differentiation pressure (9) and (10) results in hierarchical self-organization of minicolumnar RFs. An example is shown in Fig. 2. When synaptic modification starts the RFs are modified and fall into two groups after 10 to 50 ν-cycles. Group one, consisting of RFs $\vec{R}^1$ and $\vec{R}^2$, specializes to patterns of the upper left corner and

[7] We demand $\chi(t) \leq \chi_o$ as boundary condition in order to prevent χ from increasing during the first updates.

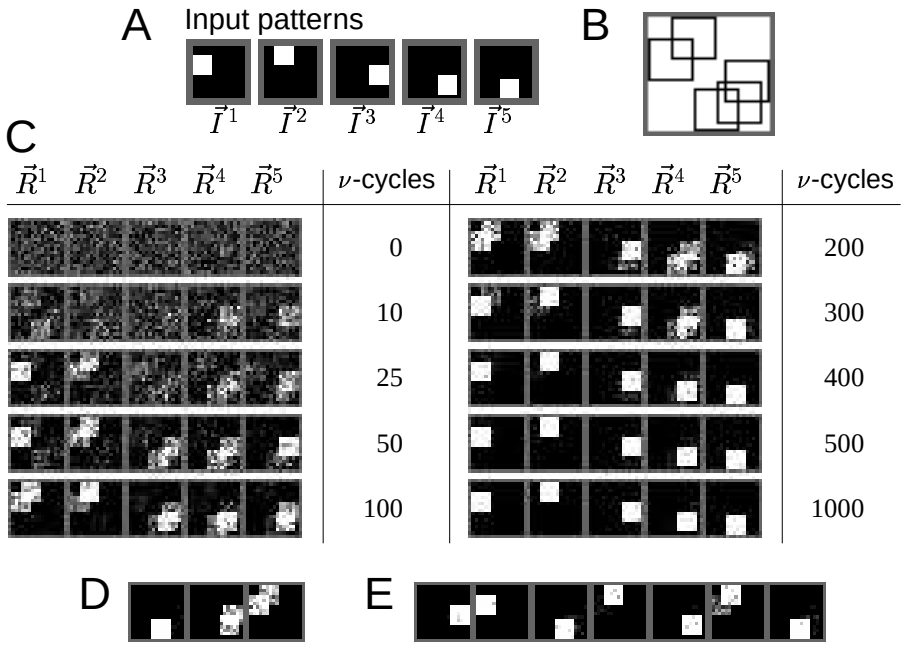

Figure 2: **A** The five different input patterns of the database. **B** Outlines of the input patterns. **C** Modification of the RFs of a macrocolumn with $k = 5$ minicolumns. **D** RFs of a macrocolumn with $k = 3$ minicolumns after 1000 ν-cycles. **E** RFs of a macrocolumn with $k = 7$ minicolumns after 1000 ν-cycles.

group two, consisting of RFs $\vec{R}^3$, $\vec{R}^4$, and $\vec{R}^5$, specializes to the patterns of the lower right corner (in Fig. 2C the RFs are appropriately ordered). In the interval between 50 and 200 ν-cycles RFs $\vec{R}^1$ and $\vec{R}^2$ are sensitive to different mixtures of patterns $\vec{I}^1$ and $\vec{I}^2$ and RFs $\vec{R}^3$, $\vec{R}^4$, and $\vec{R}^5$ are sensitive to different mixtures of patterns $\vec{I}^3$, $\vec{I}^4$, and $\vec{I}^5$. During this learning period the RFs are repeatedly changing between sensitivities to different mixtures of patterns of their group but never to mixtures of patterns of different groups. From about 200 ν-cycles on the RFs start to specialize to one of the patterns their group is sensitive to. $\vec{R}^1$ and $\vec{R}^2$ specialize to patterns $\vec{I}^1$ and $\vec{I}^2$, respectively, and $\vec{R}^3$ and $\vec{R}^5$ specialize to $\vec{I}^3$ and $\vec{I}^5$. RF $\vec{R}^4$ specializes somewhat slower to pattern $\vec{I}^4$ because of the overlap with patterns $\vec{I}^3$ and $\vec{I}^5$ (see Fig. 2C). From about 400 ν-cycles on, all RFs have specialized for an input pattern and their degree of specializations further increases to a final value thereafter.

If there are fewer minicolumns than input patterns, the RFs still fall into two groups and specialize to one or a superposition of two or more input patterns afterwards (see Fig. 2D). If there are more minicolumns than input patterns available, two or more RFs specialize to the same input pattern or occasionally to superpositions of two patterns (see Fig. 2E). Note, that a macrocolumn with more than five minicolumns usually needs longer than 1000 ν-cycles to fully specialize its RFs. In the experiment of Fig. 2C with $k = 5$ minicolumns, the case may occur that three RFs specialize to the upper left corner patterns such that there are only two RFs available for the lower right. A surplus of minicolumns can therefore be of advantage.

Hierarchical self-organization is also possible with small and constant λ. The RFs would take a longer time to fully differentiate in this case. However, a suitable constant λ would still result in hierarchical self-organization which is in general superior to non-hierarchical. Further, a constant λ would not require a time average over minicolumnar and macrocolumnar activities (see (10)).

Possible biological mechanisms for a change of λ and a change of the threshold for synaptic plasticity χ might be modulations in the efficiency of AMPA- and NMDA-mediated forms of glutamatergic synaptic transmission (Watt et al., 2000). As discussed in (Abbott and Nelson, 2000), neurons scale down NMDA currents in response to enhanced activity, which in turn can make it more difficult to evoke LTP and easier to induce LTD. In our model the hierarchical formation of RFs essentially requires a threshold χ which slowly decreases in time. Flexibility and learning time are significantly enhanced if the change of χ is appropriately coupled to mini- and macrocolumnar activity. Equations (9) to (12) represent one possibility to realize a well-working such coupling but it can, of course, only be regarded as an abstract and maybe more or less exact approximation of the dynamics which is really at work in cortical neurons.

4 Experiments

Patterns $\vec{I}^1$ to $\vec{I}^5$ of Fig. 2A represent very simple input. In this section we investigate the system behavior for different kinds of more complex input. In the following, all input vectors are visualized as two dimensional patterns in order to conveniently display the vectors as well as the RFs of minicolumns as two dimensional grey-level images. For all experiments we use a set of parameters which has shown to be flexible and to result in a good performance in different tests. We will refer to this set of parameters as *the standard set of parameters*. Note that the standard set of parameters was also used in the experiment of Fig. 2.

A macrocolumn with $m = 100$ neurons per minicolumn and a set of N external neurons is used. Each neuron receives $s = 15$ synapses from pre-synaptic neurons of the same minicolumn and $r = \frac{N}{m} S_d$ extra macrocolumnar synapses where $S_d = 2.73$ is the average number of afferent fibers originating from an external neuron. The ratio between external and internal neural input (4) is set to $\kappa = 0.47$ which means that the internal input to a neuron is for an active minicolumn on average more than twice as large as the external one. The synaptic strength of the afferents $\tilde{c} = \frac{w}{s}$ is determined by the above parameters, $w = \frac{sm\kappa}{NS_d}$. The values for κ and S_d are chosen to reproduce settings of earlier experiments (compare L¨ucke and von der Malsburg, 2004) for $N = 16 \times 16$. The neurons' constant threshold is $\Theta_o = \frac{1}{s}$ and the Gaussian threshold noise Θ_{noise} has a variance of $(\sigma_{\text{noise}})^2 = 0.01$.

The oscillation of the inhibitory gain factor ν is determined by $\nu_{\text{init}} = 0.1$, $\nu_{min} = 0.5$, $\nu_{max} = 1.1$, $T_{\text{init}} = 3$, and $T = 25$. For the Hebbian plasticity we use the parameters $\mathcal{E} = 0.02$, $\chi_o = \frac{1}{3}$, $\lambda_o = 0.01$, $a_\chi = 0.85$, and $b_\chi = 0.15$. The time average in (8) is computed over the last $A = 200$ ν-cycles. Note that all parameters are independent of the number of minicolumns k.

4.1 Input with Hierarchical Similarity Structure

In this section hierarchically structured input is presented, i.e., the input contains a hierarchy of groups of input patterns which members are more similar to each other than to patterns of other groups. Using artificial data in Sec. 4.1.1 and handwritten digits in Sec. 4.1.2, we show that hierarchical self-organization appropriately learns to represent the input in the case that much less RFs than input patterns exist. Furthermore, we compare hierarchical and non-hierarchical learning in this section.

4.1.1 Artificial Data

Let us consider a generalized experiment of the one in Fig. 2. As input on a layer of 16×16 input neurons we use a 7×7 square at 24 different positions. The squares are located at all possible positions within the grey region in Fig. 3B. In Fig. 3C RF self-organization is shown for afferent fibers which are subject to Hebbian plasticity defined by equations (6) to (12). As in Fig. 2C the RFs first specialize coarsely and subsequently refine to a final specialization degree. In the experiment of Fig. 3 there are more input patterns than RFs, however. The RFs, therefore, specialize such that they nevertheless can, up to a high-degree, distinguish between the different inputs.

In contrast to Fig. 3C non-hierarchical RF self-organization is shown in Fig. 3D. Dynamics and parameters of the simulation in Fig. 3D are identical to the ones of the simulation in Fig. 3C except that the dynamical $\chi(t)$ in (6) is replaced by a constant one, $\chi = 0.55\frac{1}{k}$. It can be seen that some RFs very quickly specialize to specific input patterns but that the majority of RFs remains unspecialized. Only after some time a formerly unspecialized RF specializes to some pattern and competes with similar RFs thereafter. Even after 2×10^5 or 2×10^6 ν-cycles the representation is far from being as accurate as after 2000 or 20000 ν-cycles in the hierarchical case.

The principle advantage of the hierarchical dynamics is that it, first, induces competition between large groups of RFs and only afterwards increases competition within specialized groups. This ensures that RFs compete with other similar RFs whereas in the non-hierarchical dynamics unspecialized RFs compete with specialized RFs (see Fig. 3D), which makes differentiation qualitatively worse and causes a much longer learning time.

For the simulation of Fig. 3D we chose the value of χ which was used in (L¨ucke and von der Malsburg, 2004). For smaller values of constant χ RFs specialize in the above experiment even slower than in Fig. 3D. For a larger value of χ the RFs specialize faster but the system's ability to discriminate between correlated input patterns decreases. Even if we adjust the value of the constant χ to the above experiment in order to receive the best possible compromise between velocity of RF specialization and discrimination strength, we can neither qualitatively nor quantitatively reproduce the results of hierarchical self-organization.

The differences between hierarchical and non-hierarchical learning become less pregnant if the input patterns cover a larger fraction of the input layer or if the number of minicolumns is smaller. With larger numbers of minicolumns the superiority of hierarchical self-organization further increases, however.

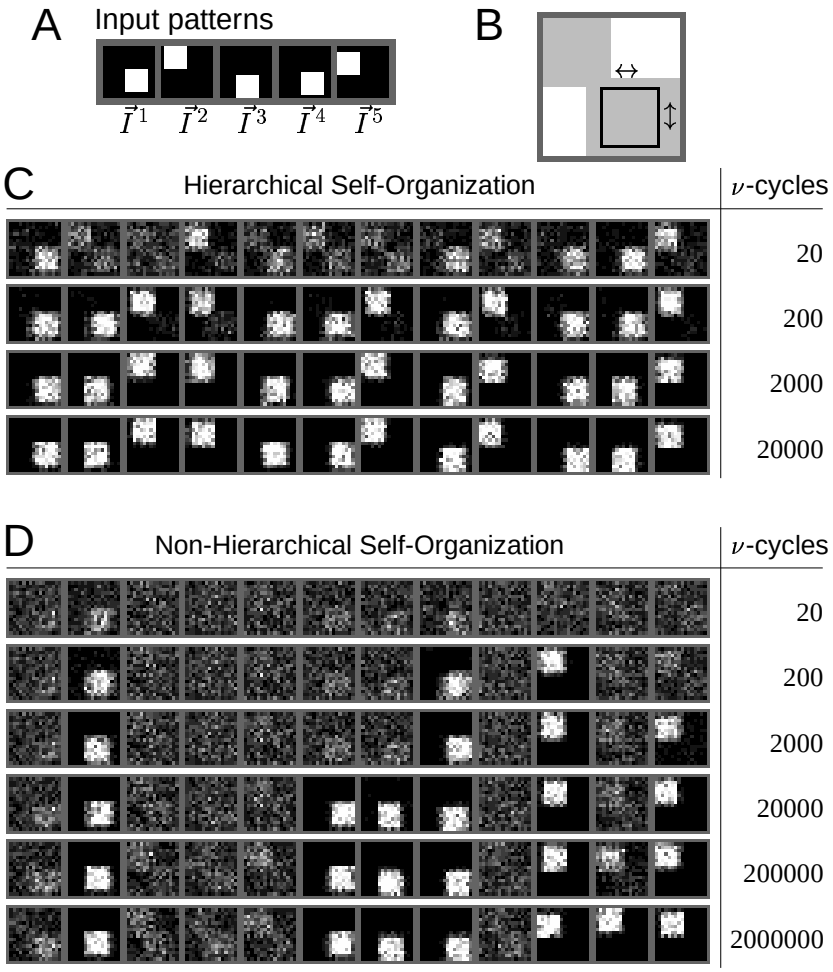

Figure 3: **A** Five of the 24 possible input patterns. **B** Illustration of possible inputs. The 7×7 square can be at any position within the grey region. **C** Hierarchical self-organization of $k = 12$ minicolumnar RFs. **D** Non-hierarchical self-organization of $k = 12$ minicolumnar RFs.

4.1.2 Hand-Written Digits

We now investigate RF self-organization for a database of input patterns $\vec{I}$ consisting of the hand-written digits zero, one, and seven. The database used is the freely available subset MNIST[8] of the NIST database. It consists of 60000 28×28 large grey-level images. For the experiment we only used the digits zero, one, and seven (see Fig. 4A for some examples). We have chosen this data because it, first, consists of two-dimensional patterns and is therefore easily displayable, second, it is commonly accessible, and, third, it can be expected to contain a hierarchy of subclasses of mutually similar patterns. In Fig. 4B RF self-organization is displayed for a set of $N = 28 \times 28$ input neurons and a macrocolumn with standard set of parameters and

[8]MNIST database of hand-written digits, NEC, yann.lecun.com/exdb/mnist/

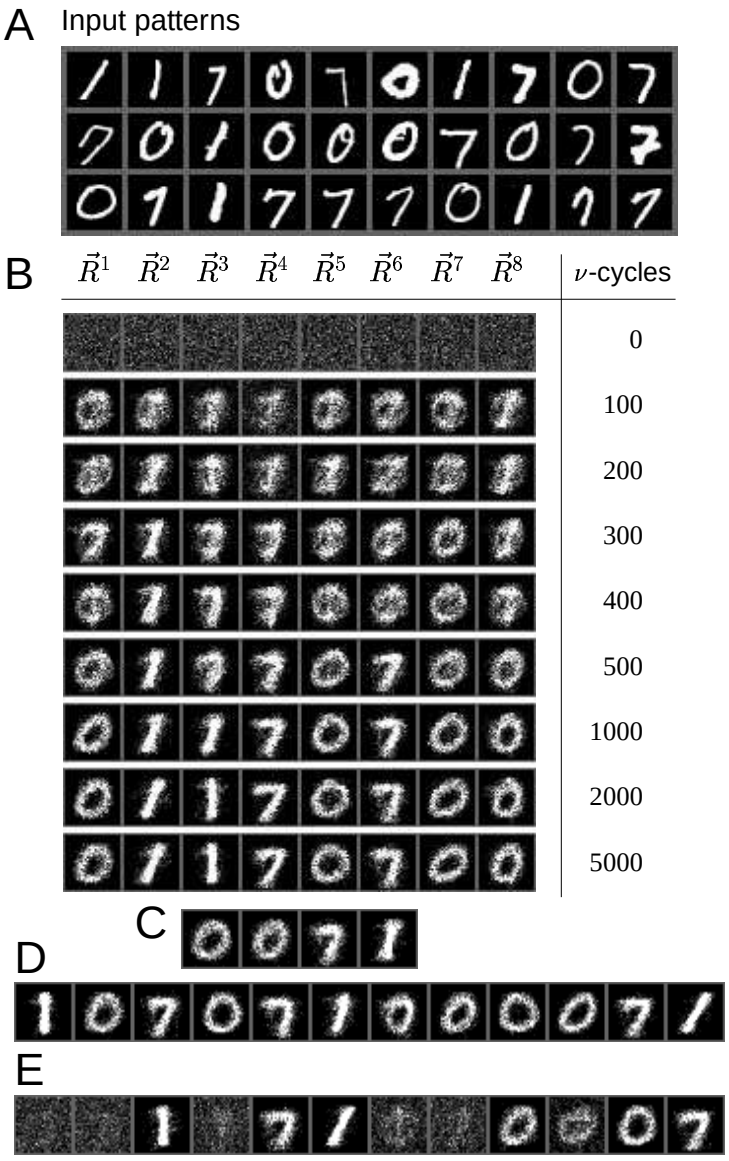

Figure 4: **A** Subset of input patterns of the hand-written digits zero, one, and seven of the MNIST database. **B** RF modification of a macrocolumn with $k = 8$ minicolumns if input of the form as displayed in **A** is presented. **C** RFs of a macrocolumn with $k = 4$ minicolumns after 5000 ν-cycles. **D** RFs of a macrocolumn with $k = 12$ minicolumns after 5000 ν-cycles. **E** RFs of a macrocolumn with $k = 12$ minicolumns and non-hierarchical dynamics with $\chi = 0.55\frac{1}{k}$ after 5000 ν-cycles.

$k = 8$ minicolumns. From 0 to 100 ν-cycles the randomly initialized RFs specialize to the region where the input neurons are most active. Some RFs coarsely show preferences to stimuli in the form of vertical bars, e.g., $\vec{R}^3$ and $\vec{R}^4$, and others are rather sensitive to circular stimuli, e.g., $\vec{R}^1$. During the first 100 ν-cycles the RFs are rather unstable and can frequently change between various kinds of input pattern mixtures. In the period from 100 to 500 ν-cycles the RF change is still high but steadily decreases until, after about 500 ν-cycles, two groups of RFs have formed. Group 1, consisting of $\vec{R}^1$, $\vec{R}^5$, $\vec{R}^7$, and $\vec{R}^8$, is sensitive to zero-type patterns whereas group 2 is sensitive to one and seven-type inputs or mixtures of them. The groups remain stable in the sense that no RF of one group will change such that it becomes sensitive to a pattern of the other group. From 500 to 1000 ν-cycles the RFs of each group further specialize to sub-patterns of their input type. This is especially obvious for group 2 where $\vec{R}^2$ and $\vec{R}^3$ form a subgroup specialized to one-type patterns and $\vec{R}^4$ and $\vec{R}^6$ form a subgroup specialized to seven-type patterns. From 1000 to 5000 ν-cycles the specialization continues and the formed subgroups decay into subgroups themselves. After 5000 ν-cycles each RF is sensitive to a different characteristic type of input patterns, e.g., RFs $\vec{R}^2$ and $\vec{R}^3$ are both sensitive to one-type input but $\vec{R}^3$ is rather sensitive to the subtype of vertical lines whereas $\vec{R}^2$ is sensitive to diagonal patterns of hand-written ones.

Note, that the course of self-organization is only sensitive to the overlap of input patterns and that there is no neighborhood relationship between the input pixels. The experiment would produce the same results if all input pixels were permuted. Induced by the overlap, a neighborhood relationship only exists between the patterns themselves and as they are two-dimensional in nature, hierarchical self-organization is reflecting this relationship.

If we apply a macrocolumn with less minicolumns to the data, the final subdivision is not as fine, see Fig. 4C for $k = 4$, and if there are more minicolumns available, the subdivision is finer, see Fig. 4D for $k = 12$. For $k = 8$ the system most often specialized four, two, and two RFs to patterns of type zero, one, and seven, respectively. Sometimes it specialized three RFs to seven-type input but usually only two for one-type input.

For comparison, the RFs of non-hierarchical self-organization with constant $\chi = 0.55\frac{1}{k}$ are shown after 5000 ν-cycles in Fig. 4E. After 5×10^4 or 5×10^5 ν-cycles the number of specialized RFs increases but the representation does not reach the accuracy as with hierarchical self-organization after 5000 ν-cycles.

We have seen that hierarchical RF self-organization works well even for strongly varying patterns. The RFs form classes of patterns by hierarchically decaying into increasingly smaller groups. Although the groups solely form on the basis of pattern overlap they represent abstractions of typical hand-written instances of the different digits[9]. The overlap similarity relationship is independent of any n-dimensional space structure such that the system can during learning be expected to become sensitive to significant classes of more general three or n-dimensional input or of input with much more complex topological structure.

[9]Digit recognition systems are generally using more invariant input representations of hand-written digits, of course.

4.2 The Continuous Bars Test

In Sec. 4.1 hierarchical RF self-organization found an accurate representation for the input patterns although many more input patterns existed than RFs were available. The RFs specialized to patterns which are, in the sense of pattern overlap, located, e.g., at about equidistant positions of the subspace of the input space in which the patterns occur (compare Fig. 3C). In this section we will present a class of input patterns which can only be accurately represented if the system extracts basic features from the input and uses distributed neural coding for pattern representation.

4.2.1 Overlapping Vertical Bars

Let us first more systematically study the ability of the system to accurately represent the subspace of training patterns. On a set of $N = 16 \times 16$ input neurons we take an input vector to consist of exactly one vertical four pixel wide bar (see Fig. 5A). We use cyclic boundary conditions such that there are 16 different such inputs. Note that a similar experiment without cyclic boundary conditions was discussed in (Spratling, 1999). In Fig. 6 RF self-organization for the same input database is visualized in

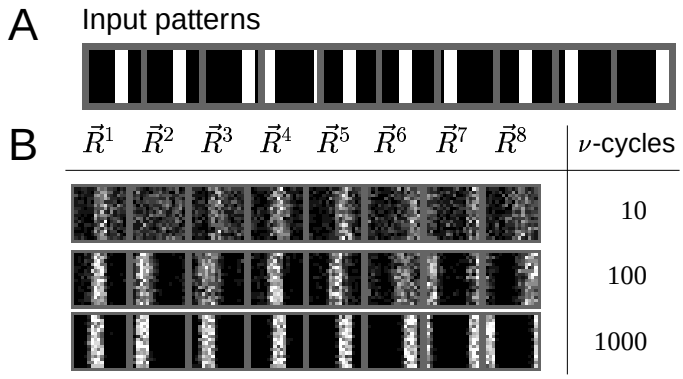

Figure 5: **A** 10 randomly chosen input patterns of vertical four pixel wide bars. **B** RF modification of a macrocolumn with $k = 8$ minicolumns if input of the form as displayed in **A** is presented.

another way. The sensitivity of a RF is displayed as an arrow whose orientation corresponds to the mean x-axis pixel position of bars the RF is sensitive to, $\phi = \frac{x}{16}360^o$, and the length of the arrow corresponds to the degree of specialization[10]. Both visualizations show that the RFs rapidly cover the input pattern space within the first 50 ν-cycles. After 50 ν-cycles the input space is rather unevenly covered, however, and there are RFs specialized for the same region. From 50 to 500 ν-cycles the specialization degree increases until the input pattern space is evenly covered by the RFs, a property which is especially well observable in Fig. 6. For a number of minicolumns smaller or larger than $k = 8$ RF self-organization is similar and the space of input patterns is, respectively, more sparsely or more densely covered.

[10] 0 means no bar and 1 means exactly one bar is preferred.

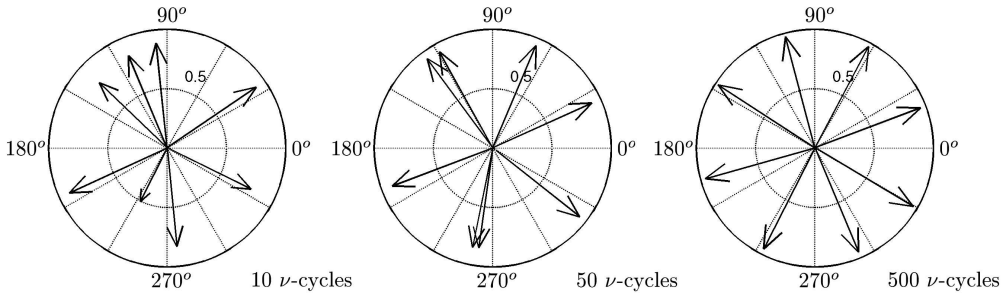

Figure 6: Visualization of RF specialization for a macrocolumn with $k = 8$ minicolumns if input of the type as in Fig. 5A is presented. Each arrow corresponds to a RF. An arrow's angle corresponds to the mean x-axis position of the vertical RF, $\phi = \frac{x}{16}360^o$. An arrow's length is zero if no bar is preferred and one (the circles radii) if exactly one bar is preferred. The RFs of one macrocolumn during one simulation are displayed after 10, 50, and 500 ν-cycles.

The experiment shows that even if the input does not contain any prominent classes of mutually similar patterns, the system organizes its RFs such that the input patterns can be processed appropriately. The RFs organize in a one-dimensional fashion, i.e., each RF has exactly two nearest neighbors with respect to RF similarity. The RFs can therefore be displayed like in Fig. 6. Note that the one-dimensional nature of the RF self-organization is solely induced by the one-dimensional overlap relation between the bars. The overlap interrelations of the input patterns themselves induce the dimensionality or, more general, the topology of the RF self-organization much like the distance interrelation of a metric space gives rise to its topology. If the input of a 16×16 array of input neurons does not consist of all possible vertical bars but of all possible squares of a certain size, e.g. 4×4, the RFs evenly cover a two dimensional space[11] after learning.

4.2.2 The Continuous Bars Test

We now combine the so called *bars test* (F¨oldiák, 1990) and the above demonstrated ability to represent continuous bars to what we will call the *continuous bars test*. The input patterns we use consist of superpositions of horizontal and vertical bars of a width of four pixels. On an input array of 16×16 neurons with cyclic boundary conditions a total of $b = 32$ different bars can occur. Each bar we take to appear in an input image with probability p_o (see Fig. 7A for some examples with $p_o = \frac{1}{b}$). In Fig. 7B RF self-organization is shown for $b = 32$ bars with $p_o = \frac{1}{b}$ and $k = 10$ minicolumns. In the time from 0 to 500 ν-cycles the RFs are not very stable and their degree of specialization is very low. During learning they become more and more stable, however, and from about 500 ν-cycles on their specialization degree increases until a final RF configuration is established somewhen between 2000 and

[11] a two dimensional torus for cyclic boundary conditions

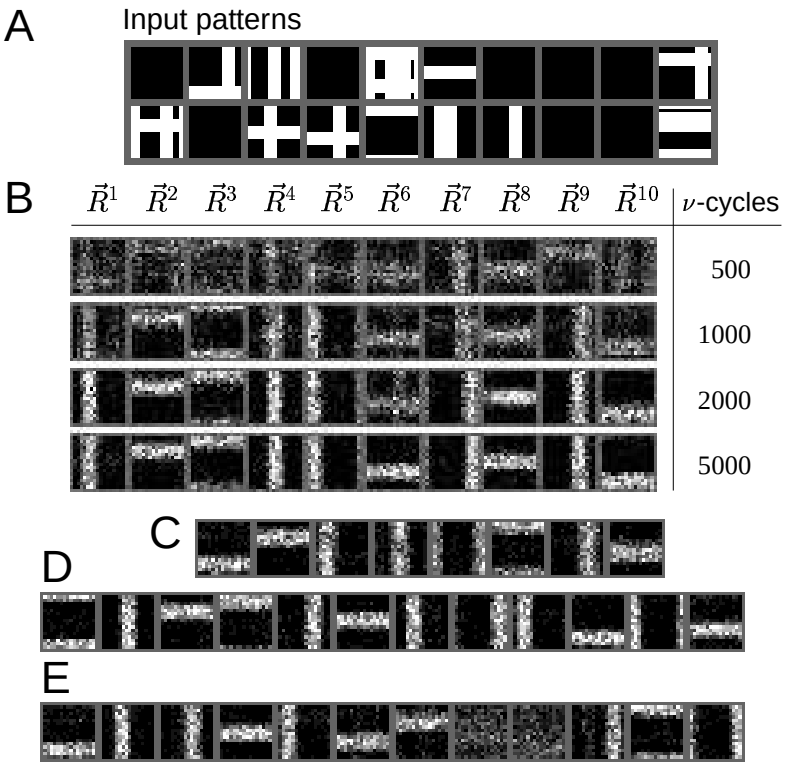

Figure 7: **A** 20 randomly chosen input patterns of a bars test with 8 bars and continuously varying bar positions. **B** RF modifications of a macrocolumn with $k = 10$ minicolumns if input of the form as displayed in **A** is presented. **C** RFs of a macrocolumn with $k = 8$ minicolumns after 5000 ν-cycles. **D** RFs of a macrocolumn with $k = 12$ minicolumns after 5000 ν-cycles. **E** RFs of a macrocolumn with $k = 12$ minicolumns and non-hierarchical learning after 5000 ν-cycles.

5000 ν-cycles. In all simulations with this input and $k = 10$ the x-axis positions of the vertical RFs and the y-axis positions of the horizontal RFs were evenly spaced after 5000 ν-cycles (see Fig. 8 for an example). Note that the system is stable against various perturbations of the bars test's parameters such as noise, bar width variations or variations of p_o. A macrocolumn with $k = 10$ minicolumns with an array of $N = 16 \times 16$ input neurons and standard set of parameters usually converges for a continuous bars test with $b = 32$ bars and $p_o = \frac{1}{b}$ to a final RF specialization within the first 2000 ν-cycles. If one RF specializes to a superposition of horizontal and vertical bars like $\vec{R}^6$ in Fig. 7B, the system needs up to about 3000 ν-cycles. Note that it is not possible to determine a time after which the system has found a representation of all bars. For other numbers of minicolumns, see Fig. 7C,D for $k = 8$ and $k = 12$, the outcome of the experiment is the same, i.e., half of the RFs evenly cover the space of vertical bars and the other half the space of horizonal bars. Occasionally, however, the numbers of vertical and horizontal RFs differ by a small number.

In Fig. 7E the RFs of non-hierarchical self-organization are shown for $\chi = 0.55\frac{1}{k}$

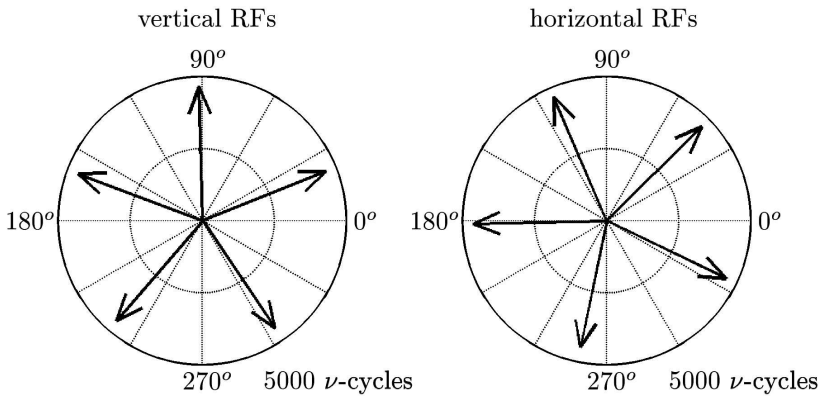

Figure 8: Visualization of RF specialization for a macrocolumn with $k = 10$ minicolumns after 5000 ν-cycles if input of the type as in Fig. 7A is presented. Each arrow corresponds to a RF. An arrow in the left-hand-side (lhs) diagram corresponds to a vertical RF and an arrows in the right-hand-side (rhs) diagram corresponds to a horizontal RF. An angle of an arrow in the lhs diagram corresponds to the x-axis position of a vertical RF, $\phi_{\text{left}} = \frac{x}{16}360^o$, and the angle of an arrow in the rhs diagram corresponds to the y-axis position of a horizontal RF, $\phi_{\text{right}} = \frac{y}{16}360^o$. An arrow's length is zero if no bar is preferred and one (the circles radii) if exactly one bar is preferred.

after 5000 ν-cycles. As can be observed, the system does not use all its RFs to represent the space of input patterns, which leads to a less accurate representation than in the hierarchical case.

4.2.3 The Classical Bars Test

The system can, of course, also be applied to the classical bars test and we will shorty put forward some results in order to be able to compare hierarchical RF self-organization with non-hierarchical and with results of other suggested systems. The input data of the bars test consist of patterns of superpositions of disjunct horizontal and disjunct vertical bars. In an input pattern (16×16 pixel in this case) each bar occurs with the same probability p_o (here $p_o = \frac{1}{b}$) and all bars are of equal size (four pixel width in our case).

On first sight, hierarchical self-organization does not promise to be superior to non-hierarchical self-organization in this experiment because the input does not contain a hierarchy of similarity classes. However, in all experimental settings the hierarchical approach shows a shorter learning time and smaller variations between simulations. For $b = 8$ bars and $k = 10$ minicolumns there is a 50% probability that the system has found all bars after about 650 ν-cycles. After about 500 ν-cycles there is a 20% probability and after 800 we have a probability of about 80% that all bars are represented (we used 100 simulations for the measurements). A non-hierarchical system with constant $\chi = 0.55\frac{1}{k}$ needs longer for learning (about 1000

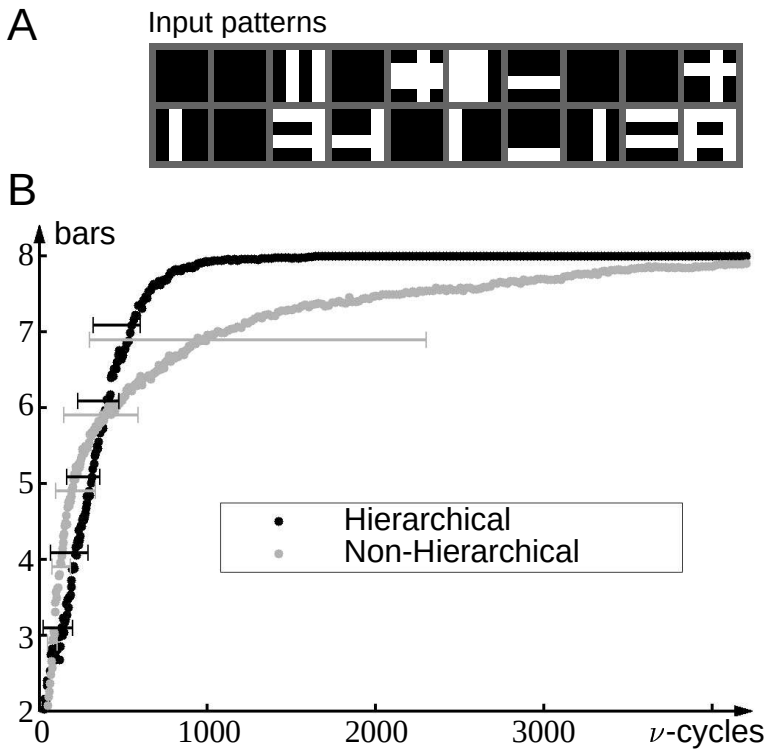

Figure 9: **A** 20 randomly chosen input patterns of a bars test with 8 bars. **B** Hierarchical and non-hierarchical learning of bars. The plots show self-organization in the course of a bars test with $k = 10$ minicolumns and $b = 8$ bars. A bar is taken to be represented if a minicolumn remains active in at least nine of ten ν-cycles if the bar is shown. The number of represented bars is measured at different time steps (every 5 ν-cycles before 500 ν-cycles, every 10 between 500 and 1000, and every 20 after 1000 ν-cycles) and the average number of represented bars is computed using 200 simulations. The black graph shows hierarchical learning and the grey graph shows non-hierarchical learning with $\chi = 0.55\frac{1}{k}$. Left limits of the error bars show the learning time after which there is a 20% probability that 3, 4, 5, 6, and 7 bars are represented, respectively. Right limits show the time after which there is an 80% probability. Black error bars belong to hierarchical learning and grey bars to non-hierarchical.

ν-cycles for a 50% probability that all bars are represented) and requires about 2400 ν-cycles to find all bars with a probability of 80%.

In Fig. 9B RF specialization for $b = 8$ bars and $k = 10$ minicolumns for hierarchical and non-hierarchical ($\chi = 0.55\frac{1}{k}$) learning is plotted. As can be seen, the learning time average of both systems is comparable until about six bars are represented. The non-hierarchical system subsequently needs on average a very long time to represent the remaining bars (see Spratling and Johnson, 2002, for a similar effect). For other numbers of bars or minicolumns (including perturbations such as noise of bar width variations) the qualitative differences between hierarchical and non-hierarchical learning are the same and in all test settings hierarchical learning has shown to be superior to non-hierarchical learning.

The classical bars test has become a benchmark test for feature extracting systems since it was introduced in (F¨oldiák, 1990). In the classical bars test minicolumnar RF self-organization was shown to be highly competitive to other suggested systems (e.g. F¨oldiák, 1990; Dayan and Zemel, 1995; Hinton et al., 1995; Hinton and Ghahramani, 1997; Charles and Fyfe, 1998; Hochreiter and Schmidhuber, 1999; O'Reilly, 2001; Spratling and Johnson, 2002) and, furthermore, proved to be especially robust with respect to various perturbation. Although hierarchical self-organization is more suitable for data with hierarchical similarity structure (see, e.g., Fig. 4B), it was shown in the experiment of Fig. 9 that it also improves performance in the classical bars test. Learning time is reduced significantly, e.g., for $k = 10$ minicolumns hierarchical learning needed about 650 ν-cycles[12] to find all $b = 8$ bars while the system described in (L¨ucke and von der Malsburg, 2004) needed about 1150 ν-cycles[13]. For $k = 18$ hierarchical learning needed about 4800 ν-cycles[14] to find $b = 16$ bars whereas the non-hierarchical system in (L¨ucke and von der Malsburg, 2004) needed about 9100 ν-cycles[15]. These differences are comparably significant for other numbers of minicolumns and bars and under different perturbations.

5 Discussion

Activity dependent Hebbian plasticity (6)-(10) coupled to dynamics of spiking neurons based on columnar organization and background oscillation was shown to result in hierarchical self-organization of minicolumnar RFs. Self-organization appropriately adapted the minicolumn sensitivities to the patterns of different databases. Hereby, the database can contain nested similarity classes as in experiments 4.1.1 and 4.1.2, continuously overlapping patterns as in Exp. 4.2.1, superpositions of patterns (Exp. 4.2.3), or superpositions of continuously overlapping patterns (see Exp. 4.2.2). The experiments further show that hierarchical learning is functionally advantageous to non-hierarchical learning as presented in (L¨ucke and von der

[12] time after which the bars are found in 50% of 100 considered simulations
[13] less than about 500 ν-cycles in 20% and less than 2400 in 80% of the cases
[14] less than about 3100 in 20% of the cases and less 8400 in 80% of the cases
[15] less than 3700 in 20% of the cases and less 18500 in 80% of the cases

Malsburg, 2004). Hierarchical self-organization develops stronger discriminating RFs, uses all RFs to represent the input, and is faster than non-hierarchical self-organization. The superiority of the hierarchical approach is especially significant for larger numbers (about $k > 10$) of minicolumns, which further supports the necessity of an hierarchical approach if we consider an estimation of about 80 minicolumns per macrocolumn, e.g., in primary cortical areas (Favorov and Diamond, 1990; Mountcastle, 1997).

The ability of the hierarchical system to adapt its neuron sensitivities such that the input pattern space is uniformly covered is similar to the ability of self-organizing maps (SOMs) (Kohonen, 1990; Kohonen, 1995) to appropriately cover a space of input data. However, SOMs require a predefined distance measure for the input space as well as a predefined dimensionality of the SOM output space. The here presented system forms an appropriate topological representation of the input space solely on the basis of input pattern overlap. It can, therefore, also process input data whose topology is not equivalent to the one of an n-dimensional Euclidean space. Thus, the system is rather comparable to a Growing Neural Gas[16] (GNG) (Fritzke, 1995) with respect to its topological flexibility. Note, however, that also for the GNG a predefined distance measure between points of the input space is used whereas our system uses overlap of extended patterns.

A second and more significant difference between the system here presented and SOMs or GNGs is the ability of our system to learn from pattern superpositions and to extract the basic constituents of the patterns as was demonstrated in the bars test. Furthermore, the neural architecture of our column based system with spiking neurons is quite different from the one of SOMs or GNGs and we refer to (L¨ucke and von der Malsburg, 2004) for a discussion of functional and neuroscientific aspects of the macrocolumnar neural dynamics. A system which was applied to feature extraction and to input like the one in Exp. 4.2.1 was described in (Spratling, 1999). In contrast to our system, the computational units of that network had to be explicitly one-dimensionally interconnected in order to solve the task of Exp. 4.2.1.

There are various unsupervised systems able to find clusters of mutually similar or neighboring input. If the input data is continuously distributed over a manifold with or without well-defined dimensionality, SOMs and GNGs are, respectively, a popular choice to appropriately represent the data. Again other systems have successfully been applied to the bars test, e.g. (F¨oldiák, 1990; Dayan and Zemel, 1995; Hinton et al., 1995; Hinton and Ghahramani, 1997; Charles and Fyfe, 1998; Hochreiter and Schmidhuber, 1999; O'Reilly, 2001; Spratling and Johnson, 2002), and thus demonstrated their ability for distributed neural coding. The hierarchical type of self-organization presented in this paper results, for Exp. 4.1.2, in a natural nested series of RF subdivisions and in a high discrimination ability due to competition between already similar RFs. For continuously overlapping input patterns it results in RFs which evenly cover the corresponding input space. And, though it is not obvious on first sight, it also decreases learning time in the classical bars test. Hierarchical self-organization distinguishes the system from the majority of other especially neu-

[16] a further development of SOMs

ronal systems which can be applied to the one or other input type. Only for clustering in the field of data mining hierarchical approaches are frequently used.

The system presented in this paper was, with one set of parameters, not only successfully applied to all different input types just discussed but, furthermore, is able to process input consisting of mixtures of the different types. In Exp. 4.2.2 the system's RFs formed two groups, one specialized to horizontal and one group specialized to vertical bars. Simultaneously, each group's RFs organized such that they evenly covered the space represented by their group (see Fig. 8). In being able to pass the continuous form of the bars test the system offers unequaled opportunities especially if compared to other artificial neural networks.

In this paper we considered self-organization of afferents to a single model macrocolumn. In the neocortex, however, a number on the order of, coarsely, 10^6 mutually interconnected macrocolumns is estimated (Mountcastle, 1997). In a large number of experiments, the interconnection structure between the macrocolumns or rather between the minicolumns were investigated, see e.g. (Rathjen et al., 2002) for a recent study in the visual cortex. The interconnections were found to change in time (Callaway and Katz, 1990; Chapman et al., 1996) and to reflect the diversity of the environment in which the subject grew up, e.g. (Schmidt et al., 1997). The statistical prevalence of collinear contours in real world images (see e.g. Kr¨uger, 1998; Kaschube et al., 2001), for instance, is reflected by the prevalence of interconnections between corresponding orientation sensitive columns in the visual cortex (see, e.g., Schmidt et al., 1997). In the primary visual cortex the interconnections can be shown to be unspecific at first, i.e., no specialization to specific columns, which is measured via the clustering of horizontal interconnections, occurs during the first postnatal days or weeks (Ruthazer and Stryker, 1996; Chapman et al., 1996). The interconnections only gradually specialize to form interconnections between orientation tuned columns. They first form coarsely clustered interconnections which gradually refine in time. The refinement can hereby be shown to be the result of synaptic plasticity rather than cell death and to be input dependent (Ruthazer and Stryker, 1996; Chapman et al., 1996). Although the interconnection structure and its development is best studied in primary sensory areas the same mechanisms are believed to underly horizontal interconnection organization in all cortical areas.

For the intercolumnar connections to be able to specialize to appropriate cells an input-driven organization principle can be assumed which has properties similar to the type of self-organization presented here. In networks of macrocolumns hierarchical self-organization of minicolumnar RFs has the advantage to gradually structure the network from coarse to fine. In non-hierarchical self-organization a newly (and formerly poorly) specialized RF would cause other inter-macrocolumnar connections to reorganize in order to consider the new RF. Reorganization is much less efficient than refinement induced by hierarchical learning, however. Furthermore, a coarse to fine process is not only observed neuroanatomically but is also consistent with our view of high level learning.

The RFs of biological minicolumns of a macrocolumn have to differentiate to appropriately represent input which can be expected to consist of superpositions of

spike patterns with continuous overlaps. E.g., orientation columns in the primary visual cortex are sensitive to basic constituents of natural images which are often mathematical abstracted using Gabor wavelet-filters (Jones and Palmer, 1987). Another example are columns in MT[17] which are sensitive to different directions of motion (Albright et al., 1984). The biological columns equidistantly cover the space of all possible orientations and all possible directions of motion, respectively. Natural input consist, e.g., in the case of orientation columns, of superpositions of stimuli for different orientations which almost certainly lie in between of the represented orientations. This is, however, the situation which was studied in the continuous bars test in which hierarchical self-organization has shown to result in appropriate RFs. A more detailed analysis, e.g., of natural image processing, would go beyond the scope of this paper especially if considering advanced preprocessing which is already done before the optical nerve fibers connect to cortical neurons.

Our system models self-organization of RFs of a single cortical macrocolumn (Favorov and Diamond, 1990; Mountcastle, 1997) and is therefore not directly comparable to large scale models (Baxter and Dow, 1989; Obermayer et al., 1990; Tanaka, 1990; Niebur and W¨org¨otter, 1993; Miller, 1994; Bednar et al., 2002; Prod¨ohl et al., 2003, and others) which intend to study map creation mainly in the visual cortex (see Erwin et al., 1995; Swindale, 1996, for overviews). We rather suggest hierarchical self-organization as an underlying mechanism which can lead to structured afferent connections as well as to structured inner-cortical connections. In this context, the presented synaptic plasticity has proven to adapt minicolumnar RFs to various kinds of inputs in a for pattern representation especially suitable and compared to other systems extraordinarily flexible manner. Hierarchical self-organization matches properties required from the plasticity of biological RFs and is in particular modeling their hierarchical formation.

References

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178 – 1183.

Albright, T. D., Desimone, R., and Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology*, 51(1):16 – 31.

Baxter, W. T. and Dow, B. M. (1989). Horizontal organization of orientation-sensitive cells in primate visual cortex. *Biological Cybernetics*, 61:171 – 182.

Bednar, J. A., Kelkar, A., and Miikkulainen, R. (2002). Modeling large cortical networks with growing self-organizing maps. *Neurocomputing*, 44:315 – 321.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn and evolution of the brain. *Brain, Behavior and Evolution*, 60:125–151.

[17] Note, that more examples can be found in somatosensory areas, auditory areas etc.

Callaway, M. C. and Katz, L. C. (1990). Emergence and refinement of clustered horizontal connections in cat striate cortex. *Journal of Neuroscience*, 10:1134–1153.

Chapman, B., Stryker, M. P., and Bonhoeffer, T. (1996). Development of orientation preference maps in ferret primary visual cortex. *Journal of Neuroscience*, 16:6443–6453.

Charles, D. and Fyfe, C. (1998). Modelling multiple-cause structure using rectification constraints. *Network: Computation in Neural Systems*, 9:167 – 182.

Dayan, P. and Zemel, R. S. (1995). Competition and multiple cause models. *Neural Computation*, 7:565 – 579.

Elston, G. N. and Rosa, M. P. G. (2000). Pyramidal cells, patches, and cortical columns: a comparative study of infragranular neurons in TEO, TE, and the superior temporal polysensory areas of the macaque monkey. *Journal of Neuroscience*, 20:1 – 5.

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual-cortex: a critical comparison. *Neural Computation*, 7:425–468.

Favorov, O. V. and Diamond, M. (1990). Demonstration of discrete place-defined columns, segregates, in cat SI. *Journal of Comparative Neurology*, 298:97 – 112.

Favorov, O. V. and Whitsel, B. L. (1988). Spatial organization of the peripheral input to area 1 cell columns I. The detection of 'segregates'. *Brain Res.*, 472:25 – 42.

Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27:45 – 56.

F¨oldiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, 64:165 – 170.

Fritzke, B. (1995). A growing neural gas network learns topologies. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems*, volume 7. MIT Press, Cambridge MA.

H¨aussler, A. F. and von der Malsburg, C. (1983). Development of retinotopic projections: An analytical treatment. *Journal of Theoretical Neurobiology*, 2:47–73.

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. (1995). The 'wake-sleep' algorithm for unsupervised neural networks. *Science*, 268:1158 – 1161.

Hinton, G. E. and Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Phil. Trans. Royal Soc. London, Sereies B, Biological Sciences*, 352:1177 – 1190.

Hochreiter, S. and Schmidhuber, J. (1999). Feature extraction through LOCOCODE. *Neural Computation*, 11:679 – 714.

Jones, E. G. (2000). Microcolumns in the cerebral cortex. *Proceedings of the National Academy of Sciences of the USA*, 97:5019 – 5021.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233 – 1258.

Kaschube, M., Wolf, F., Geisel, T., and L¨owel, S. (2001). The prevalence of colinear contours in the real world. *Neurocomputing*, 38:1335–1339.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78:1464 – 1480.

Kohonen, T. (1995). *Self-Organizing Maps*. Springer-Verlag.

K¨orner, E., Gewaltig, M. O., K¨orner, U., Richter, A., and Rodemann, T. (1999). A model of computation in neocortical architecture. *Neural Networks*, 12:989 – 1005.

Kr¨uger, N. (1998). Collinearity and parallism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8:117 – 129.

Lubke, J., Egger, V., Sakmann, B., and Feldmeyer, D. (2000). Columnar organization of dendrites and axons of single and synaptical coupled excitatory spiny neurons in layer 4 of the rat barrel cortex. *Journal of Neuroscience*, 20:5300 – 5311.

L¨ucke, J. and von der Malsburg, C. (2004). Rapid processing and unsupervised learning in a model of the cortical macrocolumn. *Neural Computation*, 16:501 – 533.

L¨ucke, J., von der Malsburg, C., and W¨urtz, R. P. (2002). Macrocolumns as decision units. In *ICANN 2002*, LNCS 2415, pages 57 – 62. Springer.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *Journal of Neuroscience*, 14:409 – 441.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120:701 – 722.

Niebur, E. and W¨org¨otter, F. (1993). Orientation columns from first principles. *Computation and Neural Systems*, 92:409 – 414.

Obermayer, K., Ritter, H., and Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences of the USA*, 87:8345 – 8349.

O'Reilly, R. C. (2001). Generaliztion in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, 13:1199 – 1241.

Peters, A. and Yilmaz, E. (1993). Neuronal organization in area 17 of cat visual cortex. *Cerebral Cortex*, 3:49 – 68.

Prod¨ohl, C., W¨urtz, R. P., and von der Malsburg, C. (2003). Learning the gestalt rule of collinearity from object motion. *Neural Computation*, 15(8):1865 – 1896.

Rathjen, S., Schmidt, K. E., and L¨owel, S. (2002). Two-dimensional analysis of the spacing of ocular dominance columns in normally raised and strabismic kittens. *Experimental Brain Research*, 145:158 – 165.

Ruthazer, E. S. and Stryker, M. P. (1996). The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *Journal of Neuroscience*, 16:7253–7269.

Schmidt, K. E., Kim, D. S., Singer, W., Bonhoeffer, T., and L¨owel, S. (1997). Functional specificity of long-range intrinsic and interhemispheric connections in the visual cortex of strabismic cats. *Journal of Neuroscience*, 17:5480–5492.

Spratling, M. W. (1999). Pre-synaptic lateral inhibition provides a better architecture for self-organising neural networks. *Network: Computation in Neural Systems*, 10:285–301.

Spratling, M. W. and Johnson, M. H. (2002). Preintegration lateral inhibition enhances unsupervised learning. *Neural Computation*, 14:2157 – 2179.

Swindale, N. V. (1996). The development of topography in the visual cortex: a review of models. *Network: Computation in Neural Systems*, 7:161–247.

Tanaka, S. (1990). Theory of self-organization of cortical maps: Mathematical framework. *Neural Network*, 3:625 – 640.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892 – 896.

von der Malsburg, C. (1985). Nervous structures with dynamical links. *Berichte der Bunsengesellschaft für Physikalische Chemie*, 89:703 – 710.

von der Malsburg, C. (2003). Self-organization and the brain. In Arbib, M., editor, *The handbook of brain theory and neural networks*. MIT Press.

Watt, A. J., von Rossum, M. C. W., MacLeod, K. M., Nelson, S. B., and Turrigiano, G. G. (2000). Activity co-regulates quantal AMPA and NMDA current at neocortical synapses. *Neuron,* 23:659 – 670.