Neural Computation, 16(3), 501 – 533, 2004

Rapid Processing and Unsupervised Learning in a Model of the Cortical Macrocolumn

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

*also at LCBV, University of Southern California, Los Angeles, USA

Abstract

We study a model of the cortical macrocolumn consisting of a collection of inhibitorily coupled minicolumns. The proposed system overcomes several severe deficits of systems based on single neurons as cerebral functional units, notably limited robustness to damage and unrealistically large computation time. Motivated by neuroanatomical and neurophysiological findings the utilized dynamics is based on a simple model of a spiking neuron with refractory period, fixed random excitatory interconnection within minicolumns, and instantaneous inhibition within one macrocolumn. A stability analysis of the system's dynamical equations shows that minicolumns can act as monolithic functional units for purposes of critical, fast decisions and learning. Oscillating inhibition (in the gamma frequency range) leads to a phase-coupled population rate code and high sensitivity to small imbalances in minicolumn inputs. Minicolumns are shown to be able to organize their collective inputs without supervision by Hebbian plasticity into selective receptive field shapes, thereby becoming classifiers for input patterns. Using the bars test, we critically compare our system's performance with that of others and demonstrate its ability for distributed neural coding.
Keywords: Cerebral Cortex, Columnar Organization, Neural Dynamics, Stability Analysis, Neural Coding, Hebbian Plasticity, Unsupervised Learning

1 Introduction

Simulations of artificial neural networks (ANNs) are a standard way to study neural information processing. Although a large amount of data about biological neural networks is available there remain uncertainties regarding the way in which neurons process incoming action potentials, the way the neurons are interconnected, and the way in which interconnections change dynamically over time. These uncertainties have generated a broad variety of different models of neural networks. They are based on different assumptions for connectivity (e.g., feed-forward or symmetrically interconnected), neuron models (e.g., McCulloch-Pitts, integrate-and-fire, Hodgkin-Huxley), and different modification rules for synaptic weight changes (e.g., Hebbian learning, back-propagation). ANNs like the Hopfield network (Hopfield, 1982; Hopfield and Tank, 1986) or perceptrons afford deep functional insight on the basis of mathematical analysis that (1) allowed the networks to be successful in various technical applications and (2) influenced our views on learning and information processing in biological neural networks significantly. By now, it has become obvious, however, that the classical ANNs fall short in modeling the generalization abilities or computation times of biological networks. Many important reactions in the brain take place in times so short that individual neurons had time to transmit only very few or just a single action potential (see Nowak and Bullier (1997) and Thorpe et al. (1996) for reaction times in the visual system). If graded signals are to be processed, models based on a single neuron rate code fail to model the measured reaction times. Further, most ANNs do not reflect biologically plausible connectivity because they were motivated by the view that biological information processing is continuously distributed over the cortical surface or that information is processed strictly feed-forward through layers of equal neurons. However, in the last decades a large amount of anatomical and physiological data was accumulated suggesting that the cortex is hierarchically organized in cellular columns as principal building blocks (see Mountcastle (1997) and Buxhoeveden and Casanova (2002) for overviews and Jones (2000) for a critical discussion). Columnar organization is advantageous (1) with respect to implementation of a neural *population* rate code able to overcome the computational speed limitations of *single neuron* rate codes and (2) with respect to connectivity and robustness. With evolutionary growth of the brain, individual building blocks had to connect to more and more other elements. Groups of neurons can support many more connections than individual neurons and a network based on neural columns as principal units can be expected to be much more robust against the loss of connections or drop-out of neurons. In the cerebral cortex of mammals neural columns can be identified on different scales. The *minicolumn* is believed to be the smallest neural entity consisting of several tens up to a few hundred neurons, which are stacked orthogonally to the cortical surface (Peters and Yilmaze, 1993). The minicolumns themselves combine to what is called a *macrocolumn* or *segregate* (Favorov and Diamond, 1990) (see Mountcastle (1997) for an overview). Like minicolumns, 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). In the primary somatosensory cortex of the cat macrocolumns were found to contain approximately 80 minicolumns. Although mini- and macrocolumns are best studied in primary sensory areas they are found in higher cortical areas as well (Peters et al., 1997; Constantinidis et al., 2001) and are believed to represent the basic building blocks of all areas of cortices of higher vertebrates (Mountcastle, 1997; Buxhoeveden and Casanova, 2002). The main part of a minicolumn is a collection of excitatory cells grouped around bundles of dendrites (Peters and Yilmaze, 1993) and axons (Peters and Sethares, 1996). Together with physiological findings (Thomson and Deuchars, 1994) this suggests that the excitatory cells of a minicolumn are mutually strongly interconnected (Mountcastle, 1997; Buxhoeveden and Casanova, 2002). For inhibitory feedback double-bouquet cells and basket (clutch) cells play a central role (DeFelipe et al., 1989, 1990; Peters and Sethares, 1997; Budd and Kisvarday, 2001). Dendritic branch and axonal field analysis suggests that the inhibitory cells are stimulated by the activities within the excitatory cells of their minicolumn and project back to a number of minicolumns in their neighborhood.

In this paper we study a model of the macrocolumn which is motivated by the above findings. We model the macrocolumn as a collection of inhibitorily coupled minicolumns which consist themselves of a collection of randomly interconnected excitatory neurons. The excitatory neurons are modeled explicitly. The neuron model is a very abstract one but it takes into account the neurons' spiking character. The assumptions made allow for a detailed mathematical analysis which captures the basic properties of the neuron dynamics. It turns out that the spiking character of the neurons in combination with the columnar interconnection structure and fast inhibitory feedback leads to a dynamics with ideal properties for computing input mediated by afferent fibers to the macrocolumn. We will show that in the absence of input the macrocolumn dynamics possesses stationary points, i.e. states of ongoing neural activity. The number of these depends exponentially on the number of minicolumns. The stability of the stationary points is controlled by a single parameter of inhibition and changes at a single critical value of this parameters. The system is operated by letting the inhibition parameter oscillate about its critical value. An isolated macrocolumn is hereby successively forced to spontaneously break the symmetry between alternate stationary points. If the dynamics is weakly coupled to input by afferent fibers which are subject to Hebbian plasticity, a self-organization of minicolumnar RFs is induced. The self-organization makes the macrocolumn to a decision unit with respect to the input. The result of a decision is identified with the active state of the macrocolumn at the maximum of one inhibitory oscillation (or ν-*cycle*). Possible ν-cycle periods may be very short, suggesting cortical oscillations in the γ-frequency range as their biological correlates and allowing very rapid functional decisions.

The macrocolumn model will be shown to be able to classify input patterns and to extract basic features from the input. The latter capability is demonstrated using the bars test (Földiák, 1990) and it can be shown that the presented model is competitive to all other systems able to pass the test. During the bars test a system has

to learn the basic components of input patterns made up of superpositions of these components. In the superpositions the components do not add up linearly which presents an additional difficulty. Systems passing the test must be able to represent an input pattern by exploiting constituent combinatorics. The passing of the bars test can be seen as a prerequisite for systems which are intended to handle large and varying natural input because natural input is expected to be most efficiently represented by combining elementary features.

The presented network owes its computational abilities to the properties of the internal macrocolumnar neuron dynamics, which itself is emergent from the columnar interconnection structure, the spiking nature of neurons, and background oscillation. The network reflects, on the one hand, major properties of biological neural information processing and is, on the other hand, competitive to a class of systems which focus on functional performance in tasks of basic feature extraction. In combining these two aspects the system distinguishes itself from all other column-based systems, e.g., (Fukai, 1994; Favorov and Kelly, 1996; Somers et al., 1995; Fransen and Lansner, 1998; Lao et al., 2001) and systems whose ability of extracting basic input constituents was demonstrated using the bars benchmark test, e.g., (Földiák, 1990; Saund, 1995; Dayan and Zemel, 1995; Marshall, 1995; Hinton et al., 1995; Harpur and Prager, 1996; Frey et al., 1997; Hinton and Ghahramani, 1997; Fyfe, 1997; Charles and Fyfe, 1998; Hochreiter and Schmidhuber, 1999; O'Reilly, 2001; Spratling and Johnson, 2002).

In Sec. 2 we define the macrocolumn model and analyze its dynamical properties. In Sec. 3 we introduce Hebbian plasticity of afferents to the macrocolumn and study the resulting input driven self-organization of the minicolumns' RFs. In Sec. 4 the computational abilities of the system are systematically studied for the problem of pattern classification and for the bars problem. In Sec. 5 we finally discuss the system's general properties in comparison with systems which were applied to the bars problem and discuss our system's relation to neuroscience.

2 Neural Dynamics of the Macrocolumn

We first define and analyze the dynamics of a model of a single minicolumn and than proceed by studying the dynamical properties of the macrocolumn as a set of inhibitorily coupled minicolumns.

2.1 Model of the Minicolumn

A minicolumn we take to consist of excitatory neurons which are randomly interconnected as motivated by the above mentioned findings (see Mountcastle (1997) and Buxhoeveden and Casanova (2002) for review). The excitatory neurons are modeled as McCulloch-Pitts neurons with a refractory period of one time step. The dynamics of a minicolumn consisting of m neurons is described by the following set of

difference equations ($i = 1, \ldots, m$):

$$n_i(t+1) = \mathcal{H}(\sum_{j=1}^{m} T_{ij}\, n_j(t) \;-\; \Theta) \cdot \underbrace{\mathcal{H}(1 - n_i(t))}_{refraction}, \;\; \mathcal{H}(x) := \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} \;. \quad (1)$$

For the interconnection T_{ij} we assume that each neuron receives s synapses from other neurons of the minicolumn. We further assume that the dendrites and axons interconnect randomly such that a natural choice for the probability to receive a synapse from a given neuron is $\frac{1}{m}$ (compare Anninos et al. (1970)). The synaptic weights we take to be equal to the constant c. Note that for any choice of c the threshold Θ can be chosen such that the resulting dynamics is the same. Without loss of generality we therefore choose $c = \frac{1}{s}$ such that $\frac{1}{m} \sum_{i,j=1}^{m} T_{ij} = 1$. To further analyze the dynamics we describe equations (1) in terms of the activation probability of a neuron, $p(t)$, at time t. The probability depends first on the number of received inputs and second on the probability that the neuron was active in the preceding time step. Due to the interconnection (T_{ij}) the probability $f_{bn}(x)$ of a neuron to receive x non-zero inputs from its pre-synaptic neurons is given by the binomial distribution

$$f_{bn}(x) \;\; = \;\; \left(\begin{array}{c} s \\ x \end{array} \right) p^x (1-p)^{s-x}. \quad (2)$$

For $s \gg 1$ the distribution can be approximated by a Gaussian probability density (theorem of Moivre-Laplace) of the form

$$f_g(x) \;\; = \;\; \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}(\frac{x-a}{\sigma})^2}, \; a = sp, \; \sigma = \sqrt{sp(1-p)}\,. \quad (3)$$

The probability $p_A(t+1)$ that a neuron receives enough input to exceed threshold at $(t+1)$ is thus given by the integration of all $f_g(x)$ at t with $x > \frac{\Theta}{c} = s\Theta$. The probability $p(t+1)$ that the neuron is activated at time $(t+1)$ further depends on the probability $p_B(t+1)$ that it is not refractory at $(t+1)$. The probability $p_B(t+1)$ is directly given by the complement of the probability that the neuron was active the time step before, $p_B(t+1) = (1 - p(t))$.

To further analyze the dynamics so called coherence effects are to be considered, i.e. effects which are caused by repeating (cycling) neural activity states. Such effects are a direct consequence of the random but fixed interconnection matrix (T_{ij}) and thus interdependent neural activation and refraction probabilities. As we will discuss at the end of the section, the coherence effects can be suppressed, e.g., by neural threshold noise. If the effects are sufficiently suppressed, we can assume the probabilities $p_A(t+1)$ and $p_B(t+1)$ to be approximately independent, $p(t+1) = p_A(t+1)\, p_B(t+1)$. The assumption permits a compact description of dynamics (1) in terms of the activation probability $p(t)$ (see Appendix for details):

$$p(t+1) = \Phi_s\left(\frac{p(t) \;-\; \Theta}{\sqrt{p(t)\,(1-p(t))}}\right) \;(1 - p(t))\,, \quad (4)$$

where $\Phi_s(x) \;\; = \;\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{s}\,x} e^{-\frac{1}{2}y^2}\, dy$ is the Gaussian error integral parameterized by s. The inhibitory feedback to the minicolumnar activity is modeled indirectly

as a rise of the threshold Θ. It is taken to be present already in the next time step and to be equally sensed by all neurons, which can be motivated by the axonal distribution of inhibitory double-bouquet neurons (DeFelipe et al., 1990; Peters and Sethares, 1997). The inhibitory neurons receive input from the excitatory neurons of the minicolumn. The inhibitory feedback we choose to depend linearly on the over-all activity, $\frac{1}{m}\sum_{i=1}^{m} n_i(t)$, of the minicolumn,

$$\Theta \;=\; \nu\,\frac{1}{m}\sum_{i=1}^{m} n_i(t) + \Theta_o \;=\; \nu\,p(t) + \Theta_o\,, \tag{5}$$

where ν is the proportionality factor of inhibition and Θ_o the constant threshold of the neurons. The choice represents a natural approximation of the feedback and allows for a further analytical treatment. Inserting (5) into (4) we get:

$$p(t+1) = \Phi_s\Big(\frac{(1-\nu)\,p(t)\;-\;\Theta_o}{\sqrt{p(t)\,(1-p(t))}}\Big)\;\;(1-p(t))\,. \tag{6}$$

The difference equation (6) can be shown to possess a point of non-zero stable stationary activity for a wide range of parameters s, Θ_o, and ν, given by:

$$\mathcal{P}_\nu^{s,\Theta_o} := \max\big\{\,p\,\big|\,p\;=\;\Phi_s\big(\tfrac{(1-\nu)\,p-\Theta_o}{\sqrt{p\,(1-p)}}\big)\;(1-p)\big\}\,. \tag{7}$$

$\mathcal{P}_\nu^{s,\Theta_o}$ (or $\mathcal{P}_\nu$ for short) can be numerically determined and its value is in good agreement with activity probabilities obtained by directly simulating (1) with inhibition (5). The dynamics (1) with (5) has to be simulated with additional neural threshold noise in order to suppress the coherence effects mentioned above. Fluctuations of the inhibitory feedback caused by a finite number of simulated neurons, m, also contribute to the noise but are on their own for a wide range of parameters not sufficient for an appropriate suppression of the effects. Note that the coherence effects are most efficiently suppressed if the interconnection matrix (T_{ij}) is re-randomized at each time step. In this case the assumption $p(t+1) = p_A(t+1)\,p_B(t+1)$ and also the computation of $p_A(t+1)$ via the binomial distribution can be adopted from (Anninos et al., 1970). There the matter is thoroughly discussed for a dynamics with another type of inhibition, but the essential arguments carry over to our dynamics. Neural threshold noise as an alternative to re-randomization was first described in (Lücke et al., 2002). In specific simulations with fixed interconnections and threshold noise we have validated that their behavior closely matches that of simulations with successively re-randomized interconnections, which shows that the analytical results are also applicable for the biologically realistic case of fixed interconnections, as used in this work.

2.2 Model of the Macrocolumn

As motivated by the distribution of synapses of inhibitory neurons (DeFelipe et al., 1989, 1990; Peters and Sethares, 1997; Budd and Kisvarday, 2001) the macrocolumn is modeled as a collection of inhibitorily coupled minicolumns. With the same

assumption as above the dynamics is given by the following set of $k\,m$ difference equations,

$$n_i^\alpha(t+1) = \mathcal{H}(\sum_{j=1}^{m} T_{ij}^\alpha n_j^\alpha(t) \; - \; \mathcal{I}(t) \; - \; \Theta_o) \cdot \mathcal{H}(1 - n_i^\alpha(t))\,, \tag{8}$$

where $\alpha = 1, \ldots, k$ counts the minicolumns, $i = 1, \ldots, m$ the neurons of the minicolumn, and where $\mathcal{I}(t)$ denotes a time dependent inhibition. Note that equation system (8) assumes that there are no direct connections between excitatory neurons of the different minicolumns. For a fixed α the interconnection (T_{ij}^α) is of the same type as in the previous section. By the statistical considerations of Sec. 2.1 we can replace the $k\,m$ difference equations by a set of k difference equations in terms of activation probabilities p_α of neurons of different minicolumns:

$$p_\alpha(t+1) \; = \; \Phi_s(\frac{p_\alpha(t) \; - \; \mathcal{I}(t) \; - \; \Theta_o}{\sqrt{p_\alpha(t)(1-p_\alpha(t))}}) \;\; (1 - p_\alpha(t))\,. \tag{9}$$

The inhibitory feedback $\mathcal{I}(t)$ is modeled as the maximum of the over-all activities in the minicolumns, $p_\beta = \frac{1}{m}\sum_{i=1}^{m} n_i^\beta(t)$,

$$\mathcal{I}(t) \;\; = \;\; \nu \max_{\beta=1,\ldots,k} \{p_\beta(t)\}\,, \tag{10}$$

where the maximum operation is assumed to be implemented by the system of inhibitory neurons of the macrocolumn. The maximum operation can be biologically implemented in various ways (Yu et al., 2002). Some possibilities are based on shunting inhibition (compare also Reichardt et al. (1983)) whereas others use subtractive inhibition. On the functional side, inhibition proportional to the maximal minicolumnar activity (10) results in a qualitatively different and favorable behavior compared to a dynamics with inhibition proportional to the average activity as was studied in (Lücke et al., 2002). The dynamical difference and its functional implications will be further discussed later in this section.

2.3 Stability Analysis

The dynamical properties of the macrocolumn model can now be studied with a stability analysis of a system of k coupled non-linear difference equations ($\alpha = 1, \ldots, k$):

$$p_\alpha(t+1) \; = \; \Phi_s(\frac{p_\alpha(t) - \; \nu \; \max_\beta\{p_\beta(t)\} \; - \; \Theta_o}{\sqrt{p_\alpha(t)(1-p_\alpha(t))}}) \;\; (1-p_\alpha(t)) \;\; =: \;\; G_\alpha(\vec{p}(t)) \tag{11}$$

First note that the system possesses the following set of potentially stable stationary points,

$$\mathcal{Q} = \{\vec{q}\,|\,\forall i = 1, \ldots, k\,(q_i = 0 \vee q_i = \mathcal{P}_\nu)\,\}\,, \tag{12}$$

e.g., for $k = 3$, $(0,0,0)$, $(\mathcal{P}_\nu, 0, 0)$, $(0, \mathcal{P}_\nu, 0)$, $\ldots$, $(\mathcal{P}_\nu, \mathcal{P}_\nu, 0)$, $\ldots$, $(\mathcal{P}_\nu, \mathcal{P}_\nu, \mathcal{P}_\nu)$, where $\mathcal{P}_\nu$ is given in (7). The magnitude of $\mathcal{Q}$ is $|\mathcal{Q}| = 2^k$. The vector with smallest norm, e.g., $(0,0,0)$, will be called $\vec{q}_{min}$, and the vector with largest norm, e.g.,

$(\mathcal{P}_\nu, \mathcal{P}_\nu, \mathcal{P}_\nu)$, will be called $\vec{q}_{max}$. The set of $\mathcal{Q}$ without the trivial stationary point $\vec{q}_{min}$ will be denoted by $\mathcal{Q}^+$, $\mathcal{Q}^+ = \mathcal{Q} - \{\vec{q}_{min}\}$. To analyze the stability of the stationary points in $\mathcal{Q}$ we first approximate $\mathcal{I}(\vec{p}) = \nu \max_\beta\{p_\beta\}$ with a differentiable function:

$$\tilde{\mathcal{I}}_\rho(\vec{p}) \;:=\; \nu \left(\sum_{\beta=1}^{k} (p_\beta)^\rho \right)^{\frac{1}{\rho}} \quad \Rightarrow \quad \lim_{\rho\to\infty} \tilde{\mathcal{I}}_\rho(\vec{p}) \;=\; \mathcal{I}(\vec{p}). \tag{13}$$

A stationary point $\vec{p}$ is stable if and only if the magnitudes of all eigenvalues of the Jacobian of $\vec{G}(\vec{p})$ (see (11)) are smaller than one. Thanks to symmetries of equation system (11) and to the substitution by $\tilde{\mathcal{I}}_\rho(\vec{p})$ the eigenvalues can be computed in general and are, in the limit $\rho \to \infty$, given by[1]:

$$\begin{aligned}
\lambda_0 \;&=\; 0\,,\;\; \lambda_1 \;=\; \frac{1}{2\sqrt{\mathcal{P}_\nu\,(1-\mathcal{P}_\nu)}}(1-\nu+\frac{1-2\mathcal{P}_\nu}{\mathcal{P}_\nu}\,\Theta_o)\,\Phi_s'(h(\nu)) \;-\; \Phi_s(h(\nu)) \\
\lambda_2 \;&=\; \frac{1}{2\sqrt{\mathcal{P}_\nu\,(1-\mathcal{P}_\nu)}}(1+(1-2\mathcal{P}_\nu)\,\nu+\frac{1-2\mathcal{P}_\nu}{\mathcal{P}_\nu}\,\Theta_o)\,\Phi_s'(h(\nu)) \;-\; \Phi_s(h(\nu)) \\
&\text{where}\;\; h(\nu) \;=\; \frac{(1-\nu)\,\mathcal{P}_\nu \;-\; \Theta_o}{\sqrt{\mathcal{P}_\nu\,(1-\mathcal{P}_\nu)}}\,.
\end{aligned}$$

If for a given vector $\vec{q} \in \mathcal{Q}$, $l(\vec{q})$ is the number of non-zero entries, then λ_0 is of multiplicity $(k-l(\vec{q}))$, λ_1 is of multiplicity 1, and λ_2 is of multiplicity $(l(\vec{q})-1)$. For fixed parameters s and Θ_o the magnitudes of all eigenvalues are smaller than one for ν smaller than a critical value ν_c. For $\nu > \nu_c$, λ_2 gets larger than one, which implies that all $\vec{q} \in \mathcal{Q}$ with $l(\vec{q}) \geq 2$ become unstable. Hence, a set of $(2^k - k - 1)$ stationary points of $\mathcal{Q}^+$ lose their stability at the same value ν_c. In Fig. 1 and 2 the properties of the dynamics are visualized using bifurcation diagrams. The critical value ν_c can be computed numerically and is, for $s = 15$ and $\Theta_o = \frac{1}{s}$, given by $\nu_c \approx 0.69$. For $k = 2$ the set $\mathcal{Q}^+$ consists of three non-trivial stationary points, which are all stable for $\nu < \nu_c$. Apart from the points in $\mathcal{Q}^+$ there exist two unstable stationary points, which are numerically computed and are given by the dotted lines in Fig. 1. For small values of ν the unstable points lie in the vicinity of the anti-symmetric stable stationary points $(\mathcal{P}_\nu, 0)$ and $(0, \mathcal{P}_\nu)$, which indicates that they attract only a small volume of neighboring phase space. For increasing ν the unstable points approach the symmetric stable stationary point $\vec{q}_{max} = (\mathcal{P}_\nu, \mathcal{P}_\nu)$, which indicates that the phase space volume of points attracted by $\vec{q}_{max}$ gets gradually smaller. In the point of structural instability, $\nu = \nu_c$, $\vec{q}_{max}$ finally loses its stability when it meets the unstable stationary points in a subcritical pitchfork bifurcation.

To visualize the crucial and analytically derived property that $(2^k - k - 1)$ stationary points of $\mathcal{Q}^+$ lose their stability for the same value of ν, the bifurcation diagram of a network for $k = 3$ is given in Fig. 2. In the diagram all stationary points of $\mathcal{Q}^+$

[1]Due to the symmetries in the equations we get, for $\vec{q} \in \mathcal{Q}$, a symmetric Jacobian $\vec{G}'(\vec{q})$ which eigenvalues are simple functions of the Jacobian matrix entries. The explicit equations for the eigenvalues can then be obtained by long but straight forward calculations.

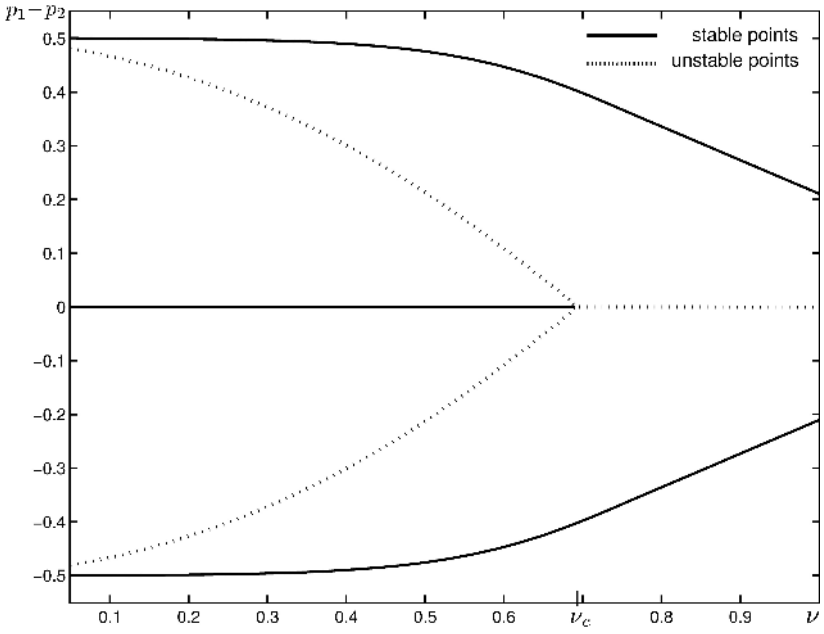

Figure 1: Bifurcation diagram of equation system (11) for parameters $s = 15$, $\Theta_o = \frac{1}{s}$ and $k = 2$. The points of $\mathcal{Q}^+ = \{(\mathcal{P}_\nu, 0),\ (0, \mathcal{P}_\nu),\ (\mathcal{P}_\nu, \mathcal{P}_\nu)\}$ are plotted together with the two unstable points of the dynamics. To obtain a 2-dimensional bifurcation diagram the stationary points for given ν are projected onto the 1-dimensional space with normal vector $\frac{1}{\sqrt{2}}(1, 1)$. The only stationary point not plotted is $\vec{q}_{min}$ because it projects onto the same point as $\vec{q}_{max}$. The solid lines mark stable stationary points, the dotted lines mark unstable points.

are plotted and we find for $\nu < \nu_c$ the set's $(2^k - 1)$ stable stationary points. Apart from the points in $\mathcal{Q}^+$ we get a number of unstable stationary points, which all lie, for small ν, at the same distance from $\vec{q}_{max}$ and in the vicinity of the other points of $\mathcal{Q}^+$. As ν increases, the unstable points are getting closer to the stable points with $l(\vec{q}) \geq 2$ non-zero entries and for $\nu = \nu_c$ these stable points of $\mathcal{Q}^+$ lose their stability when they are met by the unstable points in subcritical bifurcations.

Note that an inhibition proportional to the mean minicolumnar activity instead of the maximum as in (10) results in a dynamics whose stationary points lose their stability for different values of ν (compare also Lücke et al. (2002)). This dynamic property is reflected by eigenvalues of the dynamics's Jacobian which depend on $l(\vec{q})$ for $\tilde{\mathcal{I}}_\rho(\vec{p})$ with $0 < \rho < \infty$.

2.4 Input

For the neuron dynamics (8) we have gained, by our stability analysis, far reaching insight into the dynamical properties of the macrocolumnar model. The knowledge can now be exploited to investigate the dynamical behavior of the system if it is subject to perturbations in the form of externally induced input. As is customary

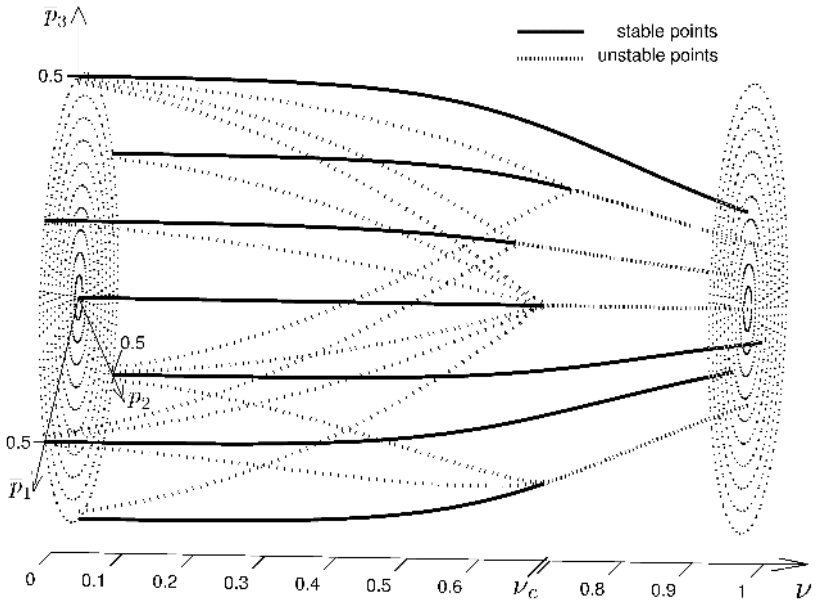

Figure 2: Bifurcation diagram of equation system (11) for parameters $s = 15$, $\Theta_o = \frac{1}{s}$, and $k = 3$. For each ν the stationary points of the 3-dimensional phase space are projected to the plane given by the normal vector $\frac{1}{\sqrt{3}}(1, 1, 1)$. The vectors $\overline{p}_1$, $\overline{p}_2$, and $\overline{p}_3$ are projections of the trihedron of the phase space onto this space. The only stationary point of the system which is not plotted is $\vec{q}_{min}$ because it projects onto the same point as $\vec{q}_{max}$. The unstable stationary points are plotted as dotted lines, the stable stationary points, which are always elements of $\mathcal{Q}^+$, are plotted as solid lines. For $\nu < \nu_c$ all elements of $\mathcal{Q}^+$ are stable but for $\nu > \nu_c$ only $k = 3$ stable points remain. All other stable points lose their stability in subcritical bifurcations for the same value of ν.

in biology, we will denote the positive contribution of a presynaptic neuron to the input of the postsynaptic neuron *excitatory postsynaptic potential* (EPSP) and we will say that a neuron emits a *spike* at time t if it is active at time t. For the dynamics as investigated in the previous section there are essentially three different modes of operation possible:

- For $\nu > \nu_c$ the macrocolumn can serve as a memory unit being able to stabilize k different macroscopic states, i.e. stable stationary points of equation system (11). The switching between the states is possible by sending a sufficiently large quantity of EPSPs to the respective minicolumn.

- For $\nu < \nu_c$ the macrocolumn is able to stabilize $2^k - 1$ different macroscopic states. The transition between the states would be possible by inducing a sufficiently large quantity of EPSPs to an appropriate subset of minicolumns. If ν is chosen to be only slightly smaller than ν_c and if one starts with the stable stationary point $\vec{q}_{max}$, already small differences in the input to the minicolumns are sufficient for the macrocolumn to change to a corresponding macroscopic

state.

- If, for an initial state $\vec{q}_{max}$, ν is continuously increased from $\nu < \nu_c$ up to a value larger than ν_c, the system will change to another stable stationary point for some value of $\nu < \nu_c$ depending on the input. For larger values of ν the system can again change between stable points until ν is finally larger than ν_c and the dynamics is forced to one of the remaining stable stationary points where just one minicolumn is active. Consider, for instance, external EPSP input to a macrocolumn with $k = 3$ minicolumns. If the numbers of EPSPs induced per time step, M^{α}_{EPSP}, are different for the three minicolumns, e.g.,
$$M^1_{EPSP} \,:\, M^2_{EPSP} \,:\, M^3_{EPSP} \,=\, 0 \,:\, 1 \,:\, (1+\epsilon)\,, \;\; \text{for } \; 0 < \epsilon \ll 1,$$
the dynamics will stabilize the initial state $\vec{q}_{max}$ for small ν. If ν gets larger, the system will change to the stable point $(0, \mathcal{P}_\nu, \mathcal{P}_\nu)$ for some $\nu_1 < \nu_c$ because this point is less deflected by the input than $\vec{q}_{max}$. The deflection of $(0, \mathcal{P}_\nu, \mathcal{P}_\nu)$ caused by the input is sufficiently large, however, if ν is further increased. The system will therefore finally stabilize the point $(0, 0, \mathcal{P}_\nu)$.

The third possibility is the one with the most useful features. For given inputs the dynamics first successively switches off the minicolumns with smallest inputs. These macro-state transitions occur the earlier the larger the differences between the inputs are and can therefore encode neural population rate differences. If a new stable stationary point is reached, the process of switching off a minicolumn continues, each time without the perturbing influence of the input of the already quiescent columns.

For a dynamics whose stationary points lose their stability for different values of ν (see Lücke et al. (2002)) the number of active minicolumns is determined by ν and not by the input. Note in this context that (10) is not the only type of inhibition that results in a single critical value of ν. Other inhibition functions, e.g. the average of active columns,

$$\mathcal{I}_{ac}(t) \;\; := \;\; \nu \, \frac{1}{|A|} \sum_{\beta \in A} p_\beta(t)\,, \;\; A \,=\, \{\beta \,|\, p_\beta > \tilde{p}\}\,, \tag{14}$$

with $0 < \tilde{p} \ll 1$, can also be shown to possess this property. In general, the contribution of quiescent minicolumns to the inhibition has to be negligible as a prerequisite for a dynamics with qualitative behavior comparable to the one with inhibition (10). The simplicity of (10) and its good functional performance were the reasons to choose an inhibition proportional to the maximum in this work.

The macro state transitions which depend on input differences but which are induced by an increased parameter ν are all performed near to symmetry breaking points. The transitions are theoretically infinitely sensitive to input differences such that a macrocolumn can serve as an ideal decision unit (see also Lücke et al. (2002)). For appropriate parameters s and Θ_o the stabilization of stationary activity is performed in a few iteration steps such that the time to increase ν from its minimal to its maximal value lies within a few tens of time steps, which makes decisions very fast in addition. If the inhibition parameter oscillates, the macrocolumn can repeatedly select the strongest input or inputs. In the next chapter this mode of operation is

exploited and further discussed for a macrocolumn with explicitly modeled afferent fibers.

3 Afferent Fibers and Hebbian Plasticity

We consider the situation that the excitatory neurons of the macrocolumn, n_i^α, receive input from an input layer of N external neurons n_j^I which are of the same type as the excitatory neurons of the minicolumns. In the following we think of the neurons of the input layer as extra-cortical neurons in order to analyze the dynamical properties of the macrocolumn in a more convenient way. However, the input neurons can also be considered to be excitatory neurons of other macrocolumns, which would account for lateral excitation within the cortex. An afference from an input neuron, n_j^I, to a neuron of a minicolumn, n_i^α, will be denoted by R_{ij}^α (see Fig. 4A for a sketch of the system). Analogously to the internal connectivity we demand that one neuron n_i^α receives a fixed number of r synapses from neurons of the input layer and that the synaptic weight of a synapse is given by $c = \frac{1}{s}$. The receptive field vector of a minicolumn, $\vec{R}^\alpha \in \{0, c, 2c, \ldots\}^N$, is defined as the sum of the RF vectors $\vec{R}_i^\alpha = (R_{i1}^\alpha, \ldots, R_{iN}^\alpha)$ of all neurons[2] of the minicolumn α, $\vec{R}^\alpha = \sum_{i=1}^m \vec{R}_i^\alpha$.

Instead of reanalyzing the dynamics statistically it is (for r significantly smaller than s) sufficient to treat the external input to the macrocolumn as perturbation of the internal dynamics. The macrocolumn will be operated by repeatedly increasing the inhibition factor ν from a minimal value ν_{min} to a maximal value ν_{max}. The system is hereby forced to select the column(s) with strongest input at the end of each period or *ν-cycle* (as we will call it from now on). In the beginning of a ν-cycle the system has to be in the state $\vec{q}_{max}$, which can be achieved under the influence of noise by setting ν to a sufficiently small value before starting to increase ν at ν_{min} (see Fig. 3B). If for a macrocolumn of, e.g., $k = 3$ minicolumns, the RFs, $\vec{R}^\alpha$, are already given, the system is able to distinguish even strongly overlapping input patterns. The system first switches off the minicolumns with RFs very different to the presented stimulus and then decides between the remaining minicolumns with RFs similar to the stimulus (see Fig. 3). In this way the system can also gracefully handle simultaneously presented patterns. If a superposition of two patterns corresponding to the RFs of two minicolumns is presented, the dynamics switches off all irrelevant minicolumns except the two corresponding ones, whose activities are symmetrized. It then depends on the choice of ν_{max} whether this is the final state or whether the symmetry is broken to favor one of the patterns.

We now proceed by introducing Hebbian plasticity of the afferent fibers to match neurophysiological experiments which show input-dependent changes of neuron RFs. As the RFs of neurons, $\vec{R}_i^\alpha$, change, the RFs of the minicolumns, $\vec{R}^\alpha$, consequently change in time as well. As update rule for the synaptic change, $\Delta R_{ij}^\alpha(t) = R_{ij}^\alpha(t) - R_{ij}^\alpha(t-1)$, we use elementary Hebbian plasticity, i.e. an afferent connection R_{ij}^α is increased if the pre-synaptic neuron was active at the time-step directly preceding the firing of the post-synaptic neuron. The state of maximal

[2] In the neurons' RF vectors only afferent connections are considered.

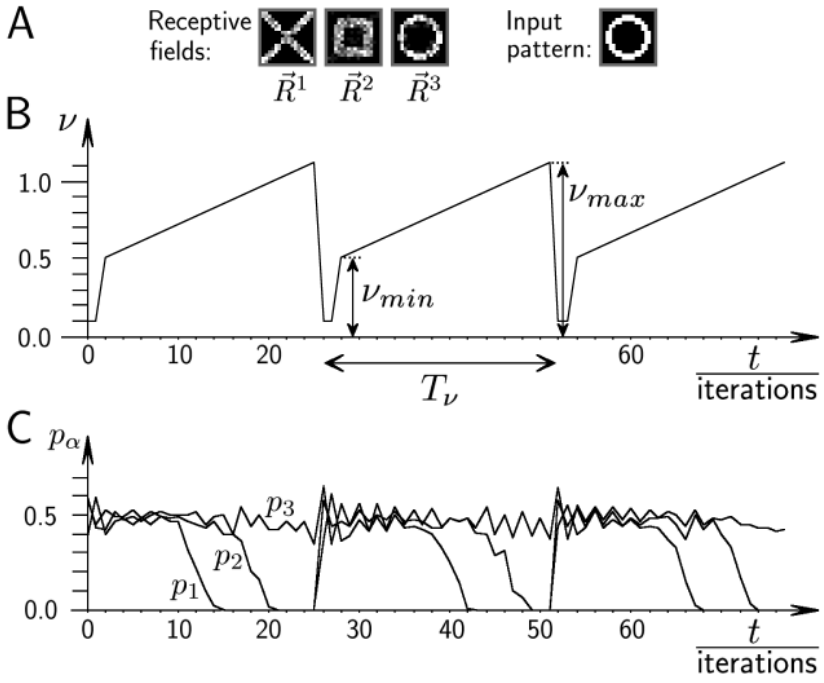

Figure 3: Operation and dynamic behavior of a system with parameters, $m = 100$, $s = 15$, $r = 7$, $\Theta_o = \frac{1}{s}$, with $k = 3$ minicolumns, and an input layer of $N = 16 \times 16$ neurons. In **A** the RFs, $\vec{R}^\alpha$, of the minicolumns $\alpha = 1, 2, 3$ are given as two-dimensional plots of the 16^2 vector entries. The entries are visualized as gray levels (black $= 0$). To make the RFs more conceivable we have chosen them to be of the form of simple two dimensional patterns. The input pattern is chosen to correspond to the RF of minicolumn $\alpha = 3$. During the operation of the system all neurons of the input layer which correspond to white pixels spike with probability $\frac{1}{3}$, neurons which correspond to black pixels are not spiking. In **B** the periodical change of the parameter of inhibition, ν, is visualized. After a short period with $\nu = 0.1$ which serves to reset the dynamics to $\vec{q}_{max}$, ν is linearly increased from ν_{min} to $\nu_{max} = 1.12$. Three ν-cycles with period length $T_\nu = 25$ are displayed. The dynamic behavior of the system is visualized in **C** where the activities $p_\alpha(t)$ for the minicolumns $\alpha = 1, 2, 3$ are plotted against time. In the beginning of a ν-cycle the dynamics tends to symmetrize the activities as predicted by the theoretical results and the bifurcation diagrams. The symmetry is first broken when minicolumn $\alpha = 1$ is switched off because its RF receives the smallest number of EPSPs from the presented input. Afterwards the stationary point $(0, \mathcal{P}_\nu, \mathcal{P}_\nu)$ is stabilized, i.e. the remaining two minicolumn activities are symmetrized, until minicolumn $\alpha = 2$ becomes quiescent because it receives less input than minicolumn $\alpha = 3$. The qualitative behavior for each ν-cycle is the same but quantitative differences exist due to threshold noise of the neurons and due to finitely many neurons per minicolumn.

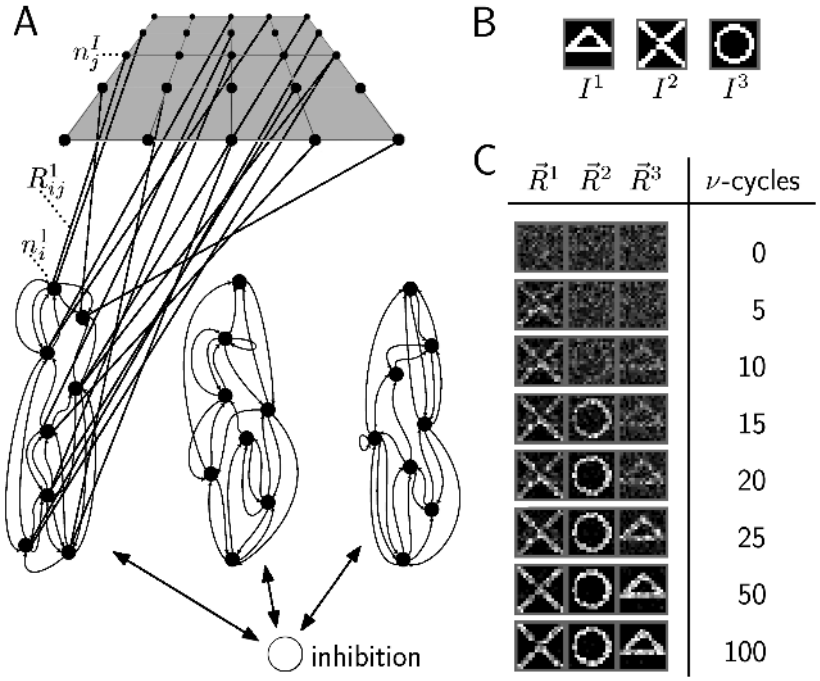

Figure 4: **A** Sketch of a macrocolumn with $k = 3$ minicolumns connected to an input layer of $N = 25$ neurons. The $m = 8$ neurons per minicolumn are randomly interconnected, each minicolumnar neuron receives $s = 3$ synapses from within its minicolumn. The inhibition is symbolically sketched as one inhibitory neuron receiving input from all minicolumns and projecting back to all of them. Each minicolumnar neuron receives $r = 2$ synapses from neurons of the input layer. 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, there are no connections of neurons within the input layer. **B** Set of three different input patterns of 16×16 pixels. **C** Modifications of RFs of a macrocolumn with $k = 3$ minicolumns and parameters $m = 100$, $s = 15$, $r = 7$, $\Theta_o = \frac{1}{s}$, $\mathcal{E} = 0.03$, $\xi = 55$, and $N = 256$. For 0 ν-cycles the random initialization of the RFs is displayed. After five ν-cycles (and five presentations of patterns randomly chosen from the set of input patterns) $\vec{R}^1$ is slowly specializing to pattern 2 and after 10 and 15 ν-cycles $\vec{R}^2$ and $\vec{R}^3$ specialize to the patterns 3 and 1, respectively. After 15 ν-cycles the RF specialization further increases until the maximal specialization is reached after about 100 ν-cycles. From 100 ν-cycles on the degree of specialization remains unchanged.

macrocolumnar activity $\vec{q}_{max}$ is the same for all inputs. Only after the selection process, i.e. at lower levels of activity due to minicolumn inactivation, the activity state is able to distinguish between inputs. Therefore, synaptic plasticity has to be predominant at low levels of macrocolumnar activity, $B(t) = \sum_{\alpha=1}^{k} \sum_{i=1}^{m} n_i^\alpha(t)$, in order to generate discriminating RFs. A simple and, as it turned out, functionally advantageous way to do this is enabling synaptic modification only if $B(t)$ is smaller than a threshold ξ. As activity oscillations are ubiquitous in the cortex, it seem plausible that synaptic plasticity is phase coupled (see Wespatat et al. (2003) for recent evidence of phase coupled synaptic modification). For the dynamics of synaptic change we further demand as boundary condition that the number of synapses received by a minicolumnar neuron is limited to r in order to avoid unlimited synaptic growths. We get as dynamic equations for the synaptic weights ($\alpha = 1, \ldots, k$; $i = 1, \ldots, m$; $j = 1, \ldots, N$):

$$\Delta R_{ij}^\alpha(t) \;=\; A_t^{\mathcal{E}}\, n_i^\alpha(t)\, n_j^I(t-1) \quad \text{iff } \; B(t) < \xi\,, \tag{15}$$

$$\forall i, \alpha : \frac{1}{c} \sum_{j=1}^{N} R_{ij}^\alpha(t) \;=\; r\,. \tag{16}$$

As our synaptic weights are discrete values, $A_t^{\mathcal{E}}$ is not a real valued growth factor but a probability[3] that the synaptic weight is increased by $c = \frac{1}{s}$. If R_{ij}^α is increased for given (α, i), the neuron n_i^α removes randomly one afferent from the input layer in order to fulfill the boundary condition.

We operate the system by periodically changing ν as in Fig. 3B. Throughout the duration of a ν-cycle we present a pattern randomly chosen from a set of input patterns. An input neuron which corresponds to a white pixel is spiking and a neuron which corresponds to a black pixel is not. The RFs of the neurons are randomly initialized and are modified according to (15) and (16). If the set of training patterns is structured, e.g., in the sense that it contains a small number of patterns as in Fig. 4B, we can observe a specialization of the RFs of the minicolumns to the different input patterns. In Fig. 4C the modification of the RFs, $\vec{R}^\alpha$, of a macrocolumn with $\alpha = 1, 2, 3$ minicolumns is displayed and it can be seen that the system organizes its RFs such that the macrocolumn becomes a decision unit for the input patterns.

In the beginning an input pattern effects all minicolumns equally such that the system selects a subset of minicolumns by symmetry breakings. As soon as, initiated by random selection, a RF specializes for one class of input patterns, the corresponding minicolumn is more likely to be activated by patterns of this class, which further increases the specialization of the RF. This is the positive feedback loop of the self-organizing process, which amplifies small fluctuation and finally leads to an ordered state of the RFs. Additionally to self-organizing aspects, we have a competition due to the minicolumn selection process and competition between afferent fibers induced by (16). In order to avoid mutual weakening of different patterns stored in the same minicolumnar RF, the system specializes its RFs to adequately different input patterns.

[3]To be more precise, for each (i, j, α) $A_t^{\mathcal{E}} \in \{0, c\}$ is an independent Bernoulli sequence with probability $P(0) = 1 - \mathcal{E}$ and $P(c) = \mathcal{E}$.

4 Experiments

We have seen that the system is able to specialize its RFs to different input patterns. So far we presented three different patterns to a network of three minicolumns (Fig. 4). We will now investigate two more general situations. In the first experimental setting we will present to the network different patterns which can be grouped into different classes. In the second setting the network's task is to extract basic constituents of a class of patterns generated by combining different bars, a task known as the *bars problem* (Földiák, 1990).

For both tasks one and the same network is used with the same set of parameters. All experiments use an input layer of 16×16 input neurons. If a binary (black and white) pattern of 16×16 pixels is presented, the input neuron, n_j^I, of a given pixel spikes with probability $\frac{1}{3}$ if the corresponding pixel is *white* and is not spiking if the corresponding pixel is *black*. Gray levels can be coded by intermediate firing rates but in the following, for simplicity, only binary input is considered. We use a network with $m = 100$ neurons per minicolumn. Each neuron receives $s = 15$ synapses from pre-synaptic neurons of the same minicolumn and $r = 7$ synapses from neurons of the input layer. The neurons' constant threshold is set to $\Theta_o = \frac{1}{s} \approx 0.067$, it is chosen such that a single EPSP is not sufficient to activate a neuron. The constant threshold is subject to Gaussian threshold noise with zero mean and a variance of $(\sigma_\Theta)^2 = 0.01$.

The oscillation of the inhibition is essentially governed by the parameters $\nu_{min} = 0.5$, $\nu_{max} = 1.12$, and the length of a ν-cycle is $T_\nu = 25$ time steps. To allow short ν-cycle periods we use a ν-oscillation as given in Fig. 3B where the first part (with $\nu = 0.1$ and additional noise) serves to reset the dynamics to the stationary point $\vec{q}_{max}$. Note, however, that self-organization of RFs is also possible with other, e.g., sinusoidal, types of ν-oscillations.

Hebbian plasticity (15)(16) is determined by the synaptic modification rate $\mathcal{E} = 0.03$ and the parameter $\xi = 55$ which determines the network activity for which synaptic modification is possible. The latter is chosen such that synaptic modification is only enabled close to the end of a ν-cycle (note that a 2-3 times larger value for ξ with simultaneously reduced $\mathcal{E}$ results in a system with comparable qualitative behavior).

All parameters are independent of the number of minicolumns k which we allow to change for different experiments. The parameters are partly chosen to reflect anatomical data as in the case of the number of neurons per minicolumns $m = 100$ and partly to optimize performance in the experiments. In the following we will refer to these parameters as *the standard set of parameters*.

4.1 Pattern Classification

We have seen that the system is able to specialize its RFs to be sensitive to a number of input patterns. More realistic input would not consist of a repeated presentation of exactly the same patterns as in Fig. 4B,C but rather of different patterns which can be grouped into different classes. In Fig. 5 a pattern classification experiment

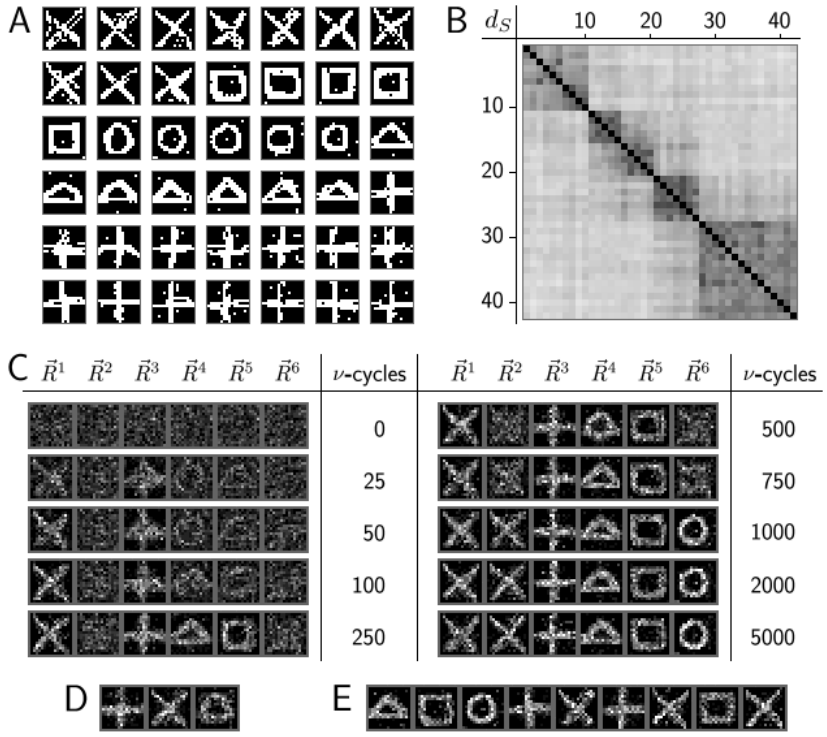

Figure 5: In **A** the set of input patterns is displayed. During each ν-cycle one randomly chosen pattern of this set is presented. **B** Distance matrix generated using the distance measure d_S. The line and column index enumerates the 42 input patterns in the same order as they appear in **A**. In **C** the modification of the RFs of a macrocolumn with $k = 6$ minicolumns and the standard set of parameters is displayed. After 250 and 1000 ν-cycles four and six different pattern classes are represented, respectively. The RFs' degree of specialization further increases thereafter and it can be seen that RFs $\vec{R}^1$ and $\vec{R}^2$ further subdivide the pattern class formerly represented by $\vec{R}^1$ only. **D** Final RF specialization (after 250 ν-cycles) if a macrocolumn with $k = 3$ minicolumns is used with the same input. **E** Final RF specialization (after 10000 ν-cycles) if a macrocolumn with $k = 9$ is used.

for such a kind of input is illustrated. For input patterns $V^a \in \{0,1\}^{256}$ as displayed in Fig. 5A we can define the distance measure,

$$d_S(V^a, V^b) \;:=\; \frac{|A \triangle B|}{|A \cup B|}\,, \tag{17}$$

where $A = \{i|V_i^a = 1\}$, $B = \{i|V_i^b = 1\}$, and where $(A \triangle B) = (A \cup B) - (A \cap B)$ is the symmetric difference of sets. Distance function (17) can directly be derived by analyzing RF mediated input to the minicolumns (further detail would go beyond the scope of this paper) and can be used to group the input patterns into different classes of mutually similar patterns as can be seen in Fig. 5B. In Fig. 5C a typical modification of RFs of a macrocolumn with six minicolumns is displayed and it can be observed that the system builds up representations of all classes identifiable in Fig. 5B. If fewer minicolumns than pattern classes are available, the system builds up larger classes of mutually similar patterns (see Fig. 5D). If more minicolumns than major classes are available, the system further subdivides the pattern classes (see Fig. 5E). The subdivision may in this respect slightly differ from simulation to simulation. E.g., for $k = 9$ the 'square class' is in many cases only represented by one and the 'plus class' by three instead of two minicolumns as in Fig. 5E. In the experiment it can further be observed that the final representation rather depends on the substructure of the pattern classes than on their size, e.g., the 'plus' pattern appears more frequently than the 'St. Andrew's cross' pattern but the 'St. Andrew's cross' tends to be represented by more RFs (see Fig. 5C,E). Furthermore, the classification is independent of the number of white pixels per pattern because the impact of the patterns on the minicolumns is normalized by boundary condition (16). The independence is only affected by patterns approximately filling out the whole input layer or by patterns having a number of white pixels close to zero.

4.2 Feature Extraction

So far we have seen that if the training patterns contain v classes of patterns, the system is able to identify these classes if $k \geq v$. There are situations, however, where the training patterns cannot easily be grouped into pattern classes. This is the case, for instance, if we present from a number of v patterns not only the patterns themselves but also all possible superpositions. If $k < 2^v$, the system is not able to store all patterns in different RFs. We have already mentioned in Sec.3, however, that the internal dynamics of a macrocolumn is especially suitable to take into account pattern superpositions such that we can, nevertheless, expect the system to generate appropriate RFs. A method to evaluate the ability of a network to handle input which can only be represented by a combination of different patterns is the bars test. It was first introduced in (Földiák, 1990) and soon became a benchmark test for generalization and combinatorial abilities of learning systems. The training patterns of the bars test consist of horizontal and vertical bars. On a quadratic input layer, $\frac{b}{2}$ (non-overlapping) horizontal and $\frac{b}{2}$ (non-overlapping) vertical bars can be displayed (with b an even integer) each with probability $\frac{1}{b}$ and all of equal size (see Fig. 6A for some examples with $b = 8$). Note that overlapping horizontal and vertical bars do

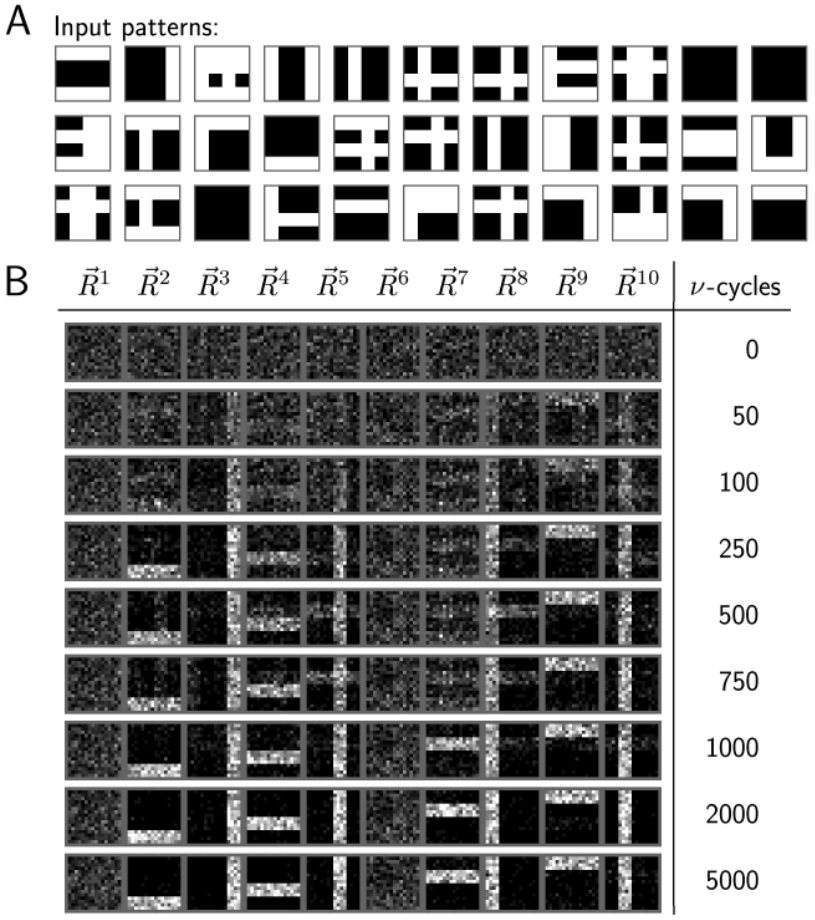

Figure 6: **A** A selection of 33 typical input patterns of the bars test of 8 different bars. **B** Typical example of the self-organization of the RFs of a macrocolumn with 10 minicolumns and the standard set of parameters. During each ν-cycle a randomly generated input pattern of the upper type is presented. After about 250 ν-cycles the network has already found representations of seven bars. After 1000 ν-cycles representations of all bars are found and are further stabilized.

not add up linearly because two overlapping white pixels do not add but result in a white pixel, as well. The bars test is passed if, after a training phase, the system has built up representations of all bars and is able to correctly classify input consisting of superpositions of the learned bars.

We can operate our system without modification, with the same set of parameters as for the experiment in Fig. 5, and it turns out that it passes the bars test without difficulties. The only prerequisite for a correct representation is that the number of minicolumns k is greater or equal to the number of different bars, $k \geq b$. If $k > b$ the RFs of some minicolumns remain uncommitted or specialize for a bar already represented by another minicolumn. For a bars test with $b = 8$ different bars Fig. 6B shows the modification of the RFs of a macrocolumn with $k = 10$ minicolumns. Starting from random initialization the RFs specialize to different single bars even though the input patterns consist mainly of bar superpositions. In Fig. 6B a representation of all bars is clearly visible after 1000 ν-cycles and the representation can be seen to further stabilize thereafter. During the learning phase a RF sometimes specializes to a combination two or more bars as can be seen by looking at RF $\vec{R}^8$ in Fig. 6 (after 500 ν-cycles). Such a RF is not stable, however, because the parts of the RF which correspond to different bars compete via equation (16). The RF therefore rapidly specializes for one bar if another RF becomes sensitive for the other. An example is given by RFs $\vec{R}^7$ and $\vec{R}^8$ in Fig. 6 from 500 to 1000 ν-cycles. In the experiment of Fig. 6 two RFs remain unspecialized. In other experiments or for a longer learning phase the two super-numerary RFs often specialize to an already represented bar and increase redundancy in this way.

The bars test was used in different versions with different numbers of bars and different systems. In (Hinton et al., 1995), for instance, 8 bars where used, (Hochreiter and Schmidhuber, 1999) and others used 10 bars, (Hinton and Ghahramani, 1997) 12 bars, and (Földiák, 1990) and others used 16 bars. To allow for comparison with these systems we measured the performance of our system for bars tests of 8, 10, 12, 14, and 16 bars. For all tests we used the same system and always with the standard set of parameters and an input layer of 16×16 neurons. The different bars tests required different bar widths in order to cover the input layer appropriately. For the bars test with $b = 8$ bars a bar width of four pixels was used (see Fig. 6), for $b = 10$ three pixels, and for $b = 12$, 14, 16 bars were of a width of two pixels. Consequently, the input layer is not uniformly covered for 10, 12, and 14 bars. In Fig. 7A,B,C the results of different test series are presented. In Fig. 7A the number of minicolumns is equal to the number of different bars, in Fig. 7B the number of minicolumns exceeds the number of bars by 2, and in Fig. 7C a surplus of 4 minicolumns is available. A measurement point in the diagrams corresponds to the number of ν-cycles after which there is a 50% probability that all bars are represented, e.g., in 200 runs with 8 bars and $k = 10$ minicolumns there were 100 runs in which a representation was found after 1050 ν-cycles (see first measurement point in Fig. 7B). A bar is taken to be represented by a minicolumn if the minicolumn remains active in 9 of 10 ν-cycles if the bar is presented. A macrocolumn is said to have found a representation of all bars if all bars are represented by at least one minicolumn and no minicolumn represents two different bars. In all runs the system

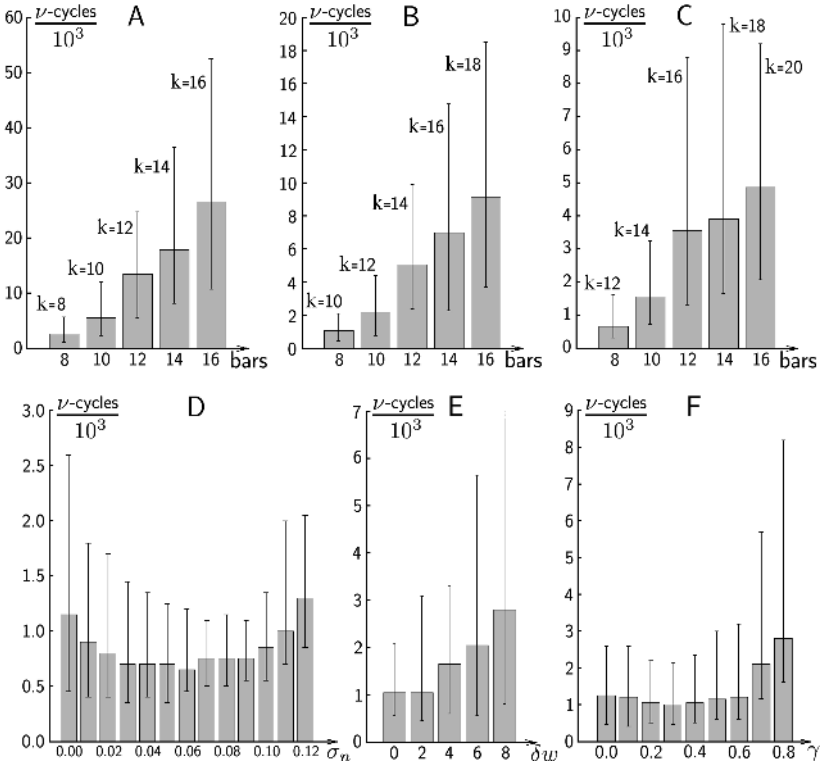

Figure 7: In **A**, **B**, and **C** results of bars tests with $b = 8, 10, 12, 14, 16$ bars and a macrocolumn with standard set of parameters are displayed. In **A** the number of minicolumns of the used macrocolumns is always equal to the number of different bars. In **B** the number of minicolumns exceeds the number of bars by two and in **C** there is a surplus of four minicolumns. In **D**, **E**, and **F** results of a bars tests with $b = 8$ bars and a macrocolumn with $k = 10$ minicolumns and standard parameters are given. In **D** the input patterns are perturbed with bit flip noise of 0 to 12%. In **E** the bar widths are varied and in **F** the generation of the input patterns is altered in the way that for each run four randomly chosen bars appear with probability $\frac{1}{8}(1 - \gamma)$ whereas the other four appear with probability $\frac{1}{8}(1 + \gamma)$. The measurement points of **A**, **B**, and **C** were obtained by taking 200 runs into account, the measurement points of **D**, **E**, and **F** with 100 runs, each. As result the numbers of ν-cycles is given after which a representation of all bars is found with a probability of 0.5. The lower and upper bounds of the error bars correspond to a probability of 0.2 and 0.8 respectively. For each run a newly generated macrocolumn with newly initialized RFs was used.

finally found a correct representation. Once a representation was found it remained stable in the sense that the minicolumns remained specialized for the same bars. For a given experimental setting there can be relatively large differences between individual runs, however. For 8 bars and $k = 10$, e.g., a correct representation of the bars was not found after about 2100 ν-cycles in 20% of the 200 runs (indicated by the upper bound of the error bar) whereas another 20% of the experiments found representations already after 400 ν-cycles (lower bound). The reason for this is that all bars but one find presentations very early whereas the remaining bar might consume a long time to be represented — an effect which is, for instance, also observable in (Spratling and Johnson, 2002) for the noisy bars test.

In Fig. 7B,C a large reduction of learning time can be observed if the number of minicolumns is larger than that of presented bars. A surplus of 2 minicolumns results in a reduction to less than half of the learning time for no surplus and a surplus of 4 minicolumns results in a learning time of coarsely a fourth.

For the results in Fig. 7A,B,C we used a newly generated bars image for every ν-cycle. The same experiments can be carried out, however, by choosing randomly from a fixed set of a number of u generated bars images. If u is several times larger than b the results are qualitatively and quantitatively comparable. For a bars test with 8 bars and $k = 10$ minicolumns, for instance, $u = 50$ input patterns are fully sufficient to build up a correct representation of single bars.

The results of Fig. 7A,B,C further show that learning time in terms of ν-cycles decreases if the number of bars does. This can be expected because the system has to learn a decreasing number of independent input constituents. On the other hand, there is an increasing overlap of bars, which makes it harder for the system to differentiate between two bars (compare Dayan and Zemel (1995) and Hochreiter and Schmidhuber (1999)). The positive effect of fewer constituents is predominant in our system. The negative effect of more overlap can be made responsible, however, for an increase of learning time if the bar widths are varied in an experiment discussed below.

Once a system has learned a correct representation of the bars it can be used to analyze bars images. To test the accuracy of the recognition we trained a macrocolumn with images generated according to the bars test until it found a representation. After some additional learning to further stabilize the representation it was tested with newly generated bars images of the same type as the training images. If an image is presented, the minicolumns corresponding to the bars appearing in the image remain active longer than minicolumns associated to bars not appearing in the test image. At the end of a ν-cycle a minicolumn is either active or not. A test image is considered to be correctly recognized if for all minicolumns which correspond to bars appearing in the image, the probability to remain active is above average and if the probabilities of minicolumns corresponding to all other bars lie below average. Six macrocolumns of 16 minicolumns trained with bars images of 16 bars were each tested 100000 times with images generated according to the bars test except that, for convenience, we required each image to contain at least one bar. The networks could classify the input correctly in all but two of the 600000 cases. In the first case one of seven bars was not recognized and in the second one of eight bars.

In the usual bars test an individual bar is always displayed identical, the bars are of the same size, and all bars occur with exactly the same probability. Systems solving the bars test can therefore be suspected to use these artificial assumptions. The system (Földiák, 1990), for instance, not only exploits the fact that the bars are occurring with the same probability but also needs to know the exact value of the bars' probability of occurrence. How much a system relies on the assumptions of the bars test can be tested by relaxing them and we present three test series showing the corresponding behavior of our system. For all three series we use a bars test with $b = 8$ bars and a macrocolumn with $k = 10$ minicolumns and standard set of parameters.

For the robustness against perturbed bar images we presented input images with bit flip noise during the learning phase (see Fig. 8). In Fig. 7D the learning time is plotted for different degrees of noise. As can be observed, low levels of noise even have positive effects. However, with an increasing noise level the final degree of specialization of the minicolumns' RFs is reduced. In Fig. 8, the final specialization degree corresponds to the displayed RFs after about 2000 or 5000 ν-cycles. If compared to the final degree of specialization in Fig. 6B, it can be seen that in the noisy case the RFs have more overlap. The overlap increases with increasing noise which leads to an increasing instability of a representation of all bars until the system cannot find a representation of the bars anymore. For the standard set of parameters and for a bars test with 8 bars no representations can be found for noise levels about above 12%. By decreasing the learning rate $\mathcal{E}$ the robustness against noise can be increased such that representations can be found for noise levels above 15%.

In the second test series the bar size is varied. For $b = 8$ the bars are usually $w = 4$ pixels wide. If $\vec{w} = (w_1, w_2, w_3, w_4)$ denotes the bar widths for the four vertical as well as for the four horizontal bars, we can define $\delta w = \sum_{i=1}^{b/2} |w_i - 4|$ as a measure for the bar width variation. In Fig. 7E the results for the test series $\vec{w} = (4, 4, 4, 4)$,$(3, 4, 4, 5)$, $(3, 3, 5, 5)$, $(2, 3, 5, 6)$, $(1, 3, 5, 7)$ are given. The learning time increases with increasing δw presumably because the maximal bar overlap increases, e.g., for $\vec{w} = (1, 3, 5, 7)$ the horizontal 7 pixel wide bar covers nearly half of the 1 pixel wide vertical bar.

The robustness of the system against relaxation of the assumption that all bars occur with equal probability is investigated in the third test series. We reduce the appearance probability of four randomly chosen bars to the value $p = p_o\,(1 - \gamma)$, and increase the appearance probability of the four other bars by the same value $p = p_o\,(1 + \gamma)$. Here γ is a parameter in the interval $[0, 1]$ and $p_o = \frac{1}{b} = \frac{1}{8}$ is the usual appearance probability. In Fig. 7F the results for $\gamma = 0.0, \ldots, 0.8$ are given, for $\gamma = 0.9$ the corresponding measurements are 3750, 8050, and 32850 ν-cycles for probabilities to get a correct representation of 0.2, 0.5, and 0.8, respectively. The measurements show that the system learns reliable a correct representation even if half of the bars appear nearly 20 times more frequently than the others and it only needs a longer learning phase if half of the bars occur more than four times more frequently.

The bar appearance probability can also be varied globally. If all bars appear with the same probability p_o and if p_o is increased to values larger than $\frac{1}{b}$, the prob-

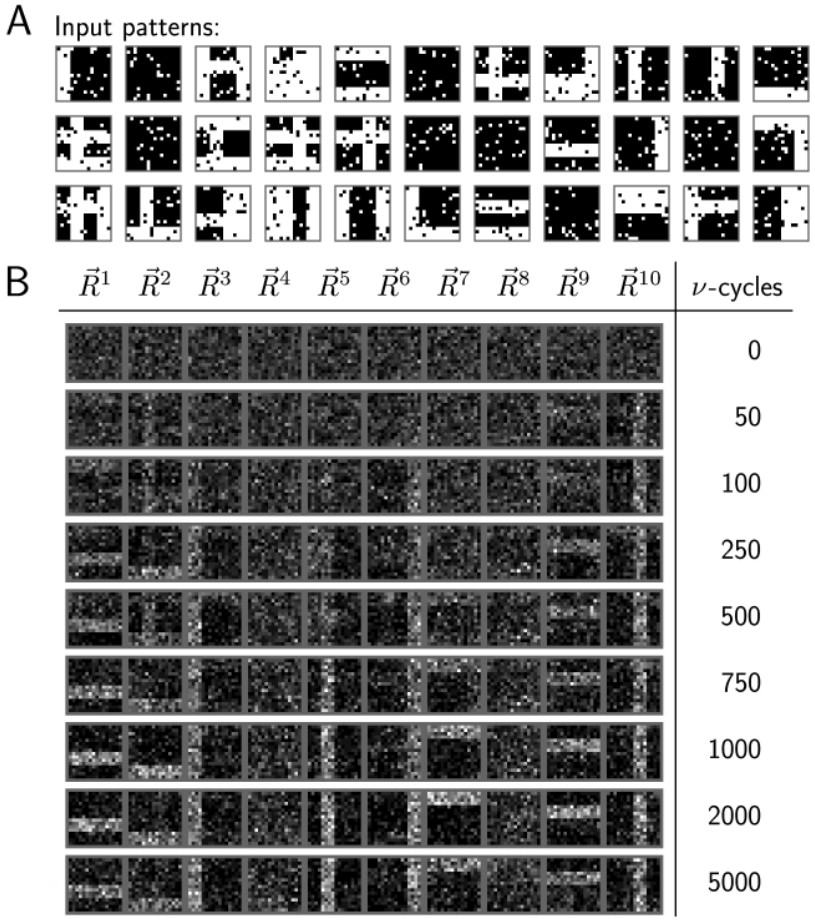

Figure 8: **A** A selection of 33 typical input patterns of a bars test with $b = 8$ different bars and 8% bit flip noise. **B** Typical RF specialization corresponding to this input. RFs of a macrocolumn with 10 minicolumns and the standard set of parameters are displayed. After about 500 ν-cycles representations of all bars are recognizable and after about 2000 ν-cycles the maximal degree of specialization is reached.

ability to find a few or a single bar in an input image gets gradually smaller. For $k = 10$, $b = 8$, and the standard set of parameters, learning time increases if p_o gets larger. For $p_o = 1.5\,\frac{1}{b}$ the system found a stable representation in all of 100 runs and needed less than 1600 ν-cycles to find representations in 50% of them. In addition to a longer learning phase, the RFs representing the different bars become less disjunct until the final representation gets unstable for values of p_o larger than about $1.8\,\frac{1}{b}$. Representations for input with higher values of p_o can be found, however, if the synaptic modification rate $\mathcal{E}$ is reduced, which in general stabilizes the representation. For $\mathcal{E} = 0.005$ instead of $\mathcal{E} = 0.03$ the system always finds stable representations for $p_o = 2.0\,\frac{1}{b}$ (after less than 18600 ν-cycles in 50% of 100 runs). However, even for very low values of $\mathcal{E}$ there is limit at p_o somewhat larger than $2.0\,\frac{1}{b}$ from which on no stable representation can be found anymore.

We have seen that one and the same network solves problems such as pattern classification and basic feature extraction. As demonstrated in the bars test, the network can build up a representation of the input which allows to classify patterns by using distributed neural coding. The network found correct representations of all bars in all 5700 simulations which were carried out to acquire the data given in Fig. 7. After the learning phase the classification for the usual bars test shows a reliability of virtually 100%. All experimental data given in Fig. 5 to Fig. 8 was obtained with the same parameters. Different sets of parameters lead to different results and for an individual task the parameters can be optimized to obtain shorter learning times or a higher robustness. We have chosen, however, to use for all experiments the standard set of parameters in order to demonstrate universality and robustness of the system's dynamics.

5 Discussion

From an elementary neuron model and a random but column-based interconnection we derived a neural dynamics with properties which make of the macrocolumn an ideal decision unit for input to its minicolumns. The dynamics is best exploited with an oscillating gain factor of the inhibition. If the afferent fibers to the macrocolumn are subject to elementary Hebbian plasticity which is also phase locked to the oscillation of inhibition, we get a system which self-organizes the RFs of its minicolumns. The system is able to classify input patterns into different groups or to extract basic constituents of the input patterns as was demonstrated using the bars test. The way the system represents the input depends only on the nature of the input, as for the pattern grouping task and for the bars test the same system with the same set of parameters was used.

5.1 Computational Aspects

There are various systems capable to learn without supervision. Important (not necessarily disjoint) classes are different types of ANN, probabilistic models, and, in a more general sense, independent component analysis (ICA) and principal compo-

nent analysis implementations. Among these systems, only few are able to build up efficient combinatorial representations of the input. The bars test is a means of testing this ability and it represents in this respect a hard problem because in general its components, the bars, do not add up linearly. Linear methods like ICA therefore fail to pass the test (see e.g. Hochreiter and Schmidhuber (1999)). The problem is (in its more or less difficult versions) solved by merely a small subset of systems (Földiák, 1990; Saund, 1995; Dayan and Zemel, 1995; Marshall, 1995; Hinton et al., 1995; Harpur and Prager, 1996; Frey et al., 1997; Hinton and Ghahramani, 1997; Fyfe, 1997; Charles and Fyfe, 1998; Hochreiter and Schmidhuber, 1999; O'Reilly, 2001; Spratling and Johnson, 2002). Some of them need additional knowledge about the input, e.g., (Földiák, 1990) and (Marshall, 1995) require that all bars occur with equal probability. Other systems, e.g. (Dayan and Zemel, 1995) and (Hinton and Ghahramani, 1997), use hierarchical approaches. If these systems are applied to the bars test, a pattern is first represented as containing horizontal or vertical patterns and then exact instances of the those patterns are represented in the next level. The system as presented in this paper is not hierarchical. However, the dynamics can be extended to allow for hierarchical learning in the sense that the input patterns are first subdivided into larger classes of patterns on the basis of the distance measure (17). Such an extended system increases the parameter ν_{max} during learning. A system based on this mechanism is currently studied in our lab. Note, however, that such a system is learning hierarchically but that it is not hierarchically representing a pattern as the systems (Dayan and Zemel, 1995) and (Hinton and Ghahramani, 1997) do.

To compare systems which solve the bars test, their behavior under the relaxation of the bars test assumptions is one important criterion, their reliabilities (some systems do not always find correct representations) and the time they need to find a representation are others. Comparison between the systems is difficult in many cases, however, because important data concerning, e.g., robustness or reliability, is often missing. Even if data is available, e.g., in terms of the number presentations of input images required to build up a correct representation, comparison remains difficult because systems specialized to the bars test assumptions[4] can be expected to be much faster than systems which can also be applied to more general input[5]. Our system was therefore tested against relaxations of the bars test assumptions and was shown to behave favorably (Fig. 7D,E,F). In terms of pattern presentations only the systems (Földiák, 1990) and (Spratling and Johnson, 2002)[6] are faster than the presented network. In (Földiák, 1990) the probability of bar occurrence has to be known ahead of time, however, and (Spratling and Johnson, 2002) don't report on the robustness of their system when bars are of different size or appear with different probability. A further possibility to compare systems is complexity of computation. A typical system with N input units and k internal computational units with all-to-all connectivity needs $O(Nk+k^2)$ elementary computations for one update in the learn-

[4] e.g., the system Földiák (1990) required learning time of 1200 presentations for 16 bars.

[5] e.g., Hochreiter and Schmidhuber (1999) needed 5000 passes through a training set of 500 patterns for 10 bars.

[6] The system Spratling and Johnson (2002) needs 210 cycles to get a correct representation in the majority of runs for 16 bars and a specially chosen set of parameters.

ing phase, (Spratling and Johnson, 2002) report $O(Nk^2)$ computations, whereas our system needs $O(Nk + k)$ because it is not using internal all-to-all connectivity[7].

5.2 Neuroscientific Aspects

As discussed in the introduction we designed our model of the cortical macrocolumn in accordance with relevant neuroanatomical and neurophysiological facts. We show that on discrimination and learning tasks the resulting system can overcome two serious problems raised by the concept of single neurons as the brains' decision units, reaction time and limited fault tolerance. The essential components of our model are column-based interconnections, discrete neural spike signals, oscillatory activity, and Hebbian plasticity. These neural characteristics, usually seen as independent of each other, are shown here to form a natural alliance, with important functional consequences. The model requires little genetic information, being based on sparse, asymmetric and random interconnections within the minicolumn. Our model makes several simplifying assumptions, using an abstract neuron model, discrete time and direct inhibition. Experimental predictions of the model should therefore be treated with caution. A fundamental property of our system is the ability to sustain neural activity without input. The property is based on a random interconnection matrix within a minicolumn. A relatively high number of EPSPs per time step results in a relatively high number of EPSPs in the next. The amount of EPSPs is controlled by inhibitory feedback and refractoriness of the neurons. As studies of continuous time systems suggest (e.g. Wilson and Cowan (1973)), this mechanism can be implemented in a continuous time version of the presented minicolumn model as well, such that, with a continuous inhibition between the minicolumns similar to (10), the qualitative dynamical behavior of the discrete model can be expected to carry over to a continuous one, which is based, e.g., on an integrate-and-fire or Hodgkin-Huxley neuron model. It can even be expected that convergence to stable stationary points of the dynamics is faster than in the discrete time case, which would allow for a shorter ν-cycle period and consequently an even faster reaction time. For this reason, and because of the possibility of a better comparison with neurophysiology, continuous time systems are the subject of further studies.

Our system realizes neural populations with well-defined *global* behavior, while realistically using *local* update rules for individual neurons and synapses. The resulting population code is based on a collective firing rate, evaluated by the macrocolumnar dynamics as average over each minicolumn's population at a particular phase relative to oscillating inhibition. We tentatively identify our inhibitory cycle with cortical oscillations in the gamma frequency range, ca. 30–60 Hz. Recent neurophysiological experiments (Perez-Orive et al. (2002), see Singer (2003) for review) support this view of a phase-coupled population rate code. For evidence for phase dependence of Hebbian modification see (Wespatat et al., 2003), where membrane potential oscillations of 20–40 Hz were artificially induced in pyramidal cells.

[7] If N grows, the number of available afferents per input unit can be kept constant by proportionally increasing r and reducing the synaptic weights of the afferents accordingly.

A central issue for understanding the brain is the neural code. The currently dominant view is the single neuron hypothesis (Barlow, 1972), according to which essential decisions of the brain can be linked directly to firing decisions of individual neurons. A fundamental difficulty for this view are reaction times of the brain. These can be so short that single neurons can fire only once. This makes it impossible to express graded signals (see, however, the time-of-arrival hypothesis of Thorpe (1988), which also advocates a firing phase). On the other hand, a population code can be the basis for very fast information processing. In our model with standard set of parameters, individual neurons typically fire only 2-10 times before the macrocolumn makes a decision, and yet the decision is based in a precise graded fashion on the input (if T_ν is reduced to $T_\nu = 10$ the system shows qualitatively the same behavior but neurons spike only 1-4 times before the first macro-state transition).

The other fundamental weakness of the single neuron hypothesis is lack of robustness against damage and accidents of wiring. The usual proposal to repair this weakness is a population code, and our model may be seen as an essential step at establishing one. The minicolumn has a collective receptive field. This makes it fault tolerant with respect to accidents in the afferent connections; the same can be said about intra-cortical connections. Moreover, self-tuning of the activity dynamics of minicolumns (e.g., of the parameters ν_{min} and ν_{max}) can make them robust to lesion or imperfections in ontogenesis.

In summary, our model, motivated by macrocolumn connectivity, has neurodynamical properties that solve important conceptual problems of neurophysiology. The spiking character of neurons, column based interconnection structure, oscillatory inhibition, and Hebbian plasticity are shown to combine together to form an advanced information processing system which allows to solve problems such as pattern classification and, most specifically, the bars benchmark problem, where it is highly competitive with other recent systems.

Appendix

Derivation of the neuron dynamics in terms of neuron activation probability

For the dynamics (1) with the above described interconnection T_{ij} the probability, $p(t+1)$, that a neuron is activated at time $(t+1)$ can be approximated by the product of the probability, $p_A(t+1)$, that it receives enough input to exceed threshold and the probability, $p_B(t+1)$, that the neuron is not refractory. Using equation (3) we get in the limit $s \to \infty$:

$$\begin{aligned} p(t+1) &= p_A(t+1)\, p_B(t+1) \\ &= \int_{s\Theta}^{\infty} f_g(x)\, dx\, (1-p(t)) \\ &= \int_{s\Theta}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}(\frac{x-a}{\sigma})^2} dx\, (1-p(t))\,,\ a{=}sp(t),\ \sigma{=}\sqrt{sp(t)(1-p(t))} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-s\Theta}{\sigma}} e^{-\frac{1}{2}x^2} dx\, (1-p(t)) \end{aligned}$$

$$= \Phi_s(\frac{p(t) - \Theta}{\sqrt{p(t)\,(1-p(t))}})\,(1-p(t))\,,\ \Phi_s(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\sqrt{s}\,x} e^{-\frac{1}{2}y^2}\,dy\,.$$

The approximation has proven to be applicable even for a relatively low neuron number m and for relatively small values of s.

Acknowledgments. We would like to thank Rolf P. Würtz for many discussions and suggestions and Jan D. Bouecke for helping us in implementing parts of the system. Further, partial funding by the EU in the RTN MUHCI (HPRN-CT-2000-00111), the German BMBF in the project LOKI (01 IN 504 E9), and the Körber Prize awarded to C. von der Malsburg in 2000 is gratefully acknowledged.

References

Anninos, P. A., Beek, B., Csermely, T. J., Harth, E. M., and Pertile, G. (1970). Dynamics of neural structures. *J. Theo. Biol.*, 26:121 – 148.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1:371–394.

Budd, J. M. L. and Kisvarday, Z. F. (2001). Local lateral connectivity of inhibitory clutch cells in layer 4 of cat visual cortex. *Exp. Brain Res.*, 140(2):245 – 250.

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

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

Constantinidis, C., Franowicz, M. N., and Goldman-Rakic, P. S. (2001). Coding specificity in cortical microcircuits: A multiple-electrode analysis of primate prefrontal cortex. *J. of Neuroscience*, 21:3646 – 3655.

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

DeFelipe, J., Hendry, S. H. C., Hashikawa, T., Molinari, M., and Jones, E. G. (1990). A microcolumnar structure of monkey cerebral cortex revealed by immunocytochemical studies of double bouquet cell axons. *Neuroscience*, 37:655 – 673.

DeFelipe, J., Hendry, S. H. C., and Jones, E. G. (1989). Synapses of double bouquet cells in monkey cerebral cortex. *Brain Res.*, 503:49 – 54.

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. *J. Neuroscience*, 20:1 – 5.

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

Favorov, O. V. and Kelly, D. G. (1996). Stimulus-response diversity in local neuronal populations of the cerebral cortex. *Neuroreport*, 7(14):2293 – 2301.

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.

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

Fransen, E. and Lansner, A. (1998). A model of cortical associative memory based on a horizontal network of connected columns. *Netw.-Comp.Neur.Sys.*, 9(2):235 – 264.

Frey, B. J., Dayan, P., and Hinton, G. E. (1997). A simple algorithm that discovers efficient perceptual codes. In Jenkin, M. and Harris, L. R., editors, *Computational and psychophysical mechanisms of visual coding*. Cambridge University Press.

Fukai, T. (1994). A model of cortical memory processing based on columnar organization. *Biological Cybernetics*, 70:427 – 434.

Fyfe, C. (1997). A neural network for PCA and beyond. *Neural Processing Letters*, 6:33 – 41.

Harpur, G. F. and Prager, R. W. (1996). Development of low entropy coding in a recurrent network. *Network-Computation in Neural Systems*, 7:277 – 284.

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.

Hopfield, J. J. (1982). Neural networks and physical systems with collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554 – 2558.

Hopfield, J. J. and Tank, D. W. (1986). Computing with neural circuits: a model. *Science*, 233:625 – 633.

Jones, E. G. (2000). Microcolumns in the cerebral cortex. *PNAS*, 97:5019 – 5021.

Lao, R., Favorov, O. V., and Lu, J. P. (2001). Nonlinear dynamical properties of a somatosensory cortical model. *Information Science*, 132:53 – 66.

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. *J. of Neuroscience*, 20:5300 – 5311.

Lücke, J., von der Malsburg, C., and Würtz, R. P. (2002). Macrocolumns as decision units. In *Artificial Neural Networks – ICANN 2002*, LNCS 2415, pages 57 – 62. Springer.

Marshall, J. A. (1995). Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity, and scale. *Neural Networks*, 8:335 – 362.

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

Nowak, L. G., and Bullier, J. (1997). The timing of information transfer in the visual system. *Cerebral Cortex*, 12:205-241

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

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., and Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297(5580):359 – 365.

Peters, A., Cifuentes, J. M., and Sethares, C. (1997). The organization of pyramidal cells in area 18 of the rhesus monkey. *Cerebral Cortex*, 7:405 – 421.

Peters, A. and Sethares, C. (1996). Myelinated axons and the pyramidal cell modules in monkey primary visual cortex. *J. of Comp. Neurology*, 365:232–255.

Peters, A. and Sethares, C. (1997). The organization of double bouquet cells in monkey striate cortex. *J. of Neurocytology*, 26:779 – 797.

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

Reichardt, W., Poggio, T., and Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.*, 46 (Suppl.):1 – 30.

Saund, E. (1995). A multiple cause mixture model for unsupervised learning. *Neural Computation*, 7:51 – 71.

Singer, W. (2003). Synchronization, binding and expectancy. In Arbib, M. A., editor, *The handbook of brain theory and neural networks*, pages 1136–1143. MIT Press.

Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. of Neuroscience*, 15:5448 – 5465.

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

Thomson, A. M. and Deuchars, J. (1994). Temporal and spatial properties of local circuits in neocortex. *Trends in Neuroscience*, 17:119 – 126.

Thorpe, S. (1988). Identification of rapidly presented images by the human visual system. *Perception*, 17(A77):415.

Thorpe, S., Fize, F., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520-522.

Wespatat, V., Tennigkeit, F., and Singer, W. (2003). Phase sensitivity of Hebbian modifications in oscillating cells of rat visual cortex. subm.

Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55 – 80.

Yu, A. J., Giese, M. A., and Poggio, T. A. (2002). Biophysiologically plausible implementations of the maximum operation. *Neural Computation*, 14:2857 – 2881.