

## Neural network model for the coordinated formation of orientation preference and orientation selectivity maps

M. Stetter,\* A. Müller, and E. W. Lang

*Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, POB 10 10 42, 93040 Regensburg, Germany*

(Received 17 February 1994)

Earlier models for the self-organization of orientation preference and orientation selectivity maps are explicitly designed to reproduce the functional structures observed in cortical tissue. They mostly use formal though biologically motivated implementations and artificial assumptions to achieve this result. In particular, orientation selective cells are usually encoded by doubling the orientation preference angle, which introduces an *ad hoc*  $180^\circ$  symmetry to the models. This symmetry is then reflected by the emerging  $\pm 180^\circ$  vortices, which parallel physiological findings. In this work a linear feed-forward neural network model is presented that is not designed to reproduce orientation maps but instead is designed to parallel the anatomical architecture of the early visual pathway. The network is trained using a general Hebb-type unsupervised learning rule and uncorrelated white noise as input. Arguments will be given that on average even strong intracortical interactions have only a weak influence on the learning dynamics of the afferent weights. An approximate description of the learning dynamics of these weights is then developed which strongly reduces computational expense without predetermining the receptive field properties, as earlier approaches do. For parameter regimes, where the most stable receptive fields form within the given model network, vortex structures containing singularities and fractures are observed. In addition, for strong lateral interactions, regions of reduced orientation selectivity appear, which coincide with these singularities. Thus, the present model suggests an implicit and biologically plausible coupling mechanism for the coordinated development of orientation preference and orientation selectivity maps.

PACS number(s): 87.10.+e

### I. INTRODUCTION

The representation of visual data in mammal area 17 is to a large extent performed by feature detecting neurons, the activities of which encode the presence of oriented contrast lines within their receptive fields [1–3]. The receptive fields of simple orientation selective neurons are subdivided into elongated, roughly stripe shaped regions, where either on or off response of the cell to small light stimuli is observed. These regions will henceforth be referred to as “lobes” [see Fig. 2(a) for a bilobed receptive field, which consists of one on and one off lobe]. Simple cell receptive fields in monkeys and cats exhibit at most two or three lobes [4,5] and can be found even in visually inexperienced animals [6]. For each cell, the stimulus orientation, which leads to maximum neural response, defines the orientation preference angle or *orientation preference*  $\varphi$ . Since  $\varphi$  and  $\varphi + \pi$  represent identical stimulus orientations, the orientation preference is restricted to the interval  $[0, \pi[$  and may be imagined as a bar or a nonpointed arrow. The amount, by which each cell prefers the optimally oriented stimulus with respect to other stimuli, is called its *orientation selectivity*.

Single-cell recordings along vertical penetrations through area 17 revealed only weak or no change of the

preferred orientation, but changes in complexity of the receptive field properties were observed. Horizontal penetrations of the primary visual cortex showed that the preferred orientations of cortical neurons vary continuously as one proceeds through the cortical tissue [2]. This continuous variation is frequently interrupted by orientation jumps or reversals of the orientation change. Hence in the primary visual cortex the neurons are arranged in orientation columns, where adjacent columns respond to similar orientations. This ordered arrangement of orientation preferences can also be found prior to visual experience [6].

Further studies of cortical topological orientation maps, which investigate their two-dimensional structure, were carried out using the 2-deoxyglucose method [7]. These investigations showed complex and patchy but in general stripelike iso-orientation domains. Unfortunately, this method does not allow us to separate regions of poor orientation selectivity from those with high orientation selectivity parallel to the previously applied stimulus. This separation can be done by the more recently developed *in vivo* differential imaging technique using voltage-sensitive dyes [8–10]. With this method it could be shown that in adult monkeys the preferred stimulus orientations are arranged in  $\pm \frac{1}{2}$  vortices, where the orientation preference changes by  $\pm 180^\circ$  for each counterclockwise surrounding of the vortex center. While these centers contain singularities of the orientation drift rates, the borders between adjacent vortices are often accompanied by one-dimensional discontinuities of orientation

\*FAX: +49 941-943-2479.

preference. Both types of singularities coincide with regions of reduced orientation selectivity as well as with cytochrome-oxidase rich zones [10–13]. Similar pinwheel-like orientation preference maps, though without variations in orientation selectivity as reported for the monkey, were also found in area 18 of the cat [14].

Several models have been suggested for the structure [15–20] as well as for the self-organization of orientation preference and orientation selectivity maps in mammal area 17 [21–26]. Some of these approaches suggest orientation preference distributions to consist of a system of  $\pm 1$  vortices [15,16], which could be shown to agree with the 2-deoxyglucose experiments [7] as well as quantitative evaluations of orientation drift rates [17]. Wolf *et al.* [19] showed that some structural analogy can be found between orientation preference maps and electric force vectors between conveniently positioned electric charges. Other authors use formal orientation preference vectors with doubled polar angles [18,20] in order to account for the  $180^\circ$  symmetry of orientation preference. Thereby, they introduce the  $180^\circ$  symmetry of the experimentally observed  $\pm \frac{1}{2}$  vortices into their systems.

The developmental models can be divided into several categories. Von der Malsburg and Cowan [22] showed that structured orientation preference maps can form in a system with predefined wave patterns of cortical activity and genetically predetermined subsets of orientation selective neurons. However, since the structures of the cortical activity patterns, which are applied to the network as input, determine the appearance of the emerging orientation map, the value of this model for the prediction of orientation preference structures is restricted.

Swindale [23,24] presented a general model for the formation of ordered structures within a two-dimensional array of two-dimensional vectors. Neither these vectors nor the dynamic equation of his system can be directly related to orientation selectivity and its time evolution in biological tissue. Therefore, as stressed by the author himself, this model makes no effort to suggest possible neural implementations of the introduced algorithm. Obermayer, Blasdel, and Schulten [25] present a sophisticated analysis of a Kohonen network, where the formation of columnar structures is achieved by training a cortical representation of a five-dimensional feature space describing orientation selectivity, ocular dominance, and retinotopic position. In this model, neither the learning rule nor the weight vector are designed to model biological synaptic plasticity. Instead, the Kohonen learning rule is applied to adjust formal weight vectors according to (in general correlated) input. Both groups adjust the properties of their models in order to obtain columnar structures that show optimal agreement to experimental findings.

Linsker [26] presented a linear feed-forward neural network with biologically motivated architecture, where the input neurons do not represent formal features but model biological neurons of the visual pathway. The author showed that Hebb training of initially random synaptic weights with uncorrelated white noise can lead to orientation maps with  $\pm \frac{1}{2}$  vortices as elementary structures. For the simulation of these maps, however, he applied an

approximate description that considered *ad hoc* three-lobed cortical receptive fields (with  $180^\circ$  symmetry) only and restricted the intracortical connections to weak lateral interaction.

All of these approaches make no effort to predict the vorticity of singularities and, therefore, the structure of orientation maps; rather, they artificially introduce the experimentally observed  $180^\circ$  symmetry either by doubling the orientation preference angle or through predefined receptive fields with even spatial symmetry. This *ad hoc* symmetry then determines the vorticities appearing in the orientation maps. Further, most of the approaches use formal quantities to describe properties of receptive fields and therefore fail to model the formation process of the cortical receptive field profiles themselves. In summary, the fact that collective adjustment of orientation preferences can only occur via the coupling of receptive field profiles (which in general show  $360^\circ$  symmetry), but not via the coupling of  $180^\circ$ -symmetric orientation preference angles themselves, is not taken into account so far.

The main purpose of the present work is not to model the formation of orientation maps using a system that is explicitly designed to perform this task. Instead, a neural network model is presented that explicitly incorporates into its architecture neuroanatomical data about the connectivity found in the early visual pathway. The maturation behavior of this model network under a general Hebb-type unsupervised learning rule [27] is then analyzed. The network consists of an array of linear perceptrons as introduced by Stetter, Lang, and Müller [27], which are shifted against each other and which interact via time independent lateral synaptic connections between their output neurons. Because prenatal developmental processes are modeled, the training is driven by uncorrelated white noise as input patterns. During the learning process, the profiles of the cortical receptive fields are changed by plasticity of the synaptic weights from the input layer to the output neurons. Therefore, the cortical receptive field profiles as well as the distributions of orientation preference and orientation selectivity, which are derived from their structure as observables, are allowed to undergo free development. This development is determined only by the receptive field properties of the input neurons (the model retinal ganglion cells) and the lateral interaction function.

For this network architecture, arguments will be given that on time average the lateral interaction, though dominant in synaptic strength [27,28], has only a weak influence on the learning dynamics of the afferent weights and can be treated as a perturbation. A method for the approximate description of the learning dynamics is then developed, which is similar to first-order perturbation theory. It allows for a strong reduction of computational expense without predetermining the receptive field structure, as Linsker's [26] approach does. Numerical simulations will give evidence that this approximation can even be used in case of a dominant influence of lateral cortical interactions on the learning dynamics. Finally, from zeroth-order approximations of degenerate receptive field states, an energy function is deduced. It provides a simple

gradient descent description for the learning dynamics, and its minima of which correspond to stable orientation preference maps.

The emerging orientation maps are characterized for purely excitatory, sombrero, purely inhibitory, and inverse sombrero-type lateral interaction functions. It turns out that vortex structures with singularities in the vortex centers and fractures can form in our anatomically motivated network architecture. In addition, for medium strong interaction strengths, patches of reduced orientation selectivity are found. Due to a coupling mechanism, which is inherent to the model, these patches coincide with the discontinuities of the corresponding orientation preference map. However, this is only the case for bilobed receptive fields, which were found previously to be the most stable type of cortical profiles [27]. Due to the 360° symmetry of these profiles, only  $\pm 1$  vortices are observed in the present model. This result is not in accord with more recent experimental findings [8–10] and demonstrates that  $\pm \frac{1}{2}$  vortices are not an inherent characteristic of Hebb-trained linear feed-forward model networks with an anatomically motivated architecture. In summary, there is yet no model network known in the literature that succeeds in predicting  $\pm \frac{1}{2}$  vortices within orientation preference maps as a result of cortical receptive field properties, which themselves develop during training under a given learning algorithm.

The paper is organized as follows. In Sec. II the network architecture, the learning rule, and the training conditions are specified. Further, an approximate description of the learning dynamics and the energy function for weak lateral interaction are deduced. Section III presents the training results obtained with the present model. Finally, Secs. IV and V contain a discussion of the results and a summary.

## II. NETWORK STRUCTURE AND ANALYTIC METHODS

### A. Network architecture, dynamics, and learning rule

In this work, a linear neural network with feed-forward connections is considered. Input as well as output neurons are arranged in two-dimensional layers. The input layer can be taken to model a part of the mammal retina or lateral geniculate nucleus (LGN), while the output layer corresponds to a part of layer IVc of the mammal primary visual cortex (area 17). The input neurons have mature sombrero-type receptive fields represented by a difference of Gaussians with radii  $R_{gc}$  and  $R_{gs}$ , respectively. These receptive fields act as filter functions on the input patterns and are henceforth referred to as input filters. Since random spatial variabilities of the properties of the input neurons are not considered explicitly in this paper, it is convenient to describe the input layer using the continuum limit, where the function  $v(\mathbf{r}, t)$  denotes the neural activity at position  $\mathbf{r}$  at time  $t$ . The activity function  $v(\mathbf{r}, t)$  for the input neurons is obtained by convolving the input pattern presented at time  $t - \tau$  with the input filter function. Each output neuron  $m$  receives feed-forward connections  $w_m(\mathbf{r} - \mathbf{r}_m)$  from position  $\mathbf{r}$  of

the input layer, where  $w_m(\mathbf{r})$  is nonvanishing only for retinotopic distances  $r$  less than the projection radius  $\rho$  (Fig. 1). This leads to a network that consists of a large number of linear perceptrons as described by Stetter, Lang, and Müller [27], each shifted against its neighbors by multiples of the grid constant of the output layer. In this configuration, the receptive fields of the output neurons strongly overlap (Fig. 1).

The neurons of the output layer interact through intracortical synapses. This connectivity represents a lateral interaction between the output units, and the interaction strength between neurons  $m$  and  $n$  is denoted as  $I'_{mn} \equiv I'(\mathbf{r}_{mn})$  with the intracortical connection function  $I'(\mathbf{r})$ . The network is trained using a white noise function  $\xi(\mathbf{r}, t)$ , which models prenatal spontaneous photoreceptor activity patterns. The input function is assumed to obey

$$\langle \xi(\mathbf{r}, t) \xi(\mathbf{r}', t + t_0) \rangle_t \equiv g(t_0) \delta(\mathbf{r} - \mathbf{r}'), \quad (1)$$

where  $\langle \rangle_t$  denotes the time average over intervals short compared to the time constant of the learning dynamics, and  $g(t)$  is the time correlation function of the spatially uncorrelated photoreceptor activities. In contrast to the feed-forward connections, the lateral interaction is not trained but is introduced as a system parameter and kept fixed during the maturation process. In this paper, the influence of a finite lateral interaction on the profiles of the emerging cortical receptive fields is analyzed. The orientation preference angle as well as a measure of the orientation selectivity are determined for each output neuron by evaluating the quadrupole moment of the Fourier-transformed cortical profile. This procedure, which corresponds to the experimental determination of

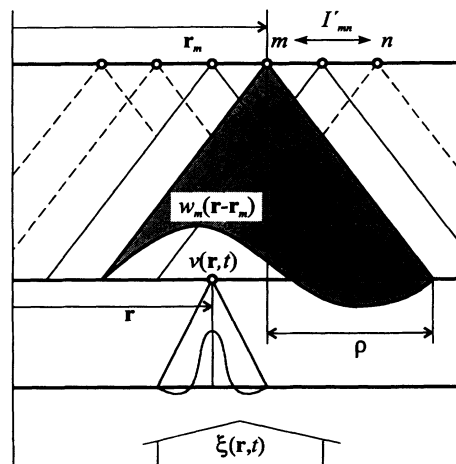


FIG. 1. Schematic plot of the neural network architecture used in this work. Uncorrelated white noise patterns  $\xi(\mathbf{r}, t)$  are convolved with the sombrero-type input filters and lead to activity distributions  $v(\mathbf{r}, t)$  of the input layer. The activities of the output neurons  $m$  are obtained by summation over these input activities weighted by the synaptic fields  $w_m(\mathbf{r} - \mathbf{r}_m)$ . The synaptic fields of different output neurons are shifted against each other to preserve strict retinotopic order. The output neurons  $m$  and  $n$  are connected via time independent lateral connections  $I'_{mn}$ .

orientation selectivity from the response to moving sine wave gratings, is described in more detail elsewhere [27]. The orientation preferences and the orientation selectivities of all output neurons then represent the orientation preference and orientation selectivity map for the considered network.

The activity  $s_m$  of output neuron  $m$  at time  $t$  is given by

$$s_m(t+\tau) = \int w_m(\mathbf{r}-\mathbf{r}_m)v(\mathbf{r},t)d\mathbf{r} + \sum_n I'(\mathbf{r}_{mn})s_n(t). \quad (2)$$

The constant  $\tau$  approximately corresponds to the time between a visual stimulus and the maximum response of the stimulated cortical neuron and is considered as an elementary time step of the network dynamics. It can be estimated to about 50 ms from evaluations of reverse correlations in cats [5]. The iterative equation (2) can be transformed into the expansion

$$\begin{aligned} s_m(t+\tau) = & \int w_m(\mathbf{r}-\mathbf{r}_m)v(\mathbf{r},t)d\mathbf{r} \\ & + \sum_n I'(\mathbf{r}_{mn}) \int w_n(\mathbf{r}-\mathbf{r}_n)v(\mathbf{r},t-\tau)d\mathbf{r} \\ & + \dots, \end{aligned} \quad (3)$$

where the signals of the last term, which propagate from the input neurons over one cortical interneuron to the cortical target neuron, are delayed by  $\tau$  with respect to the direct afferent signals.

The training is guided by a general Hebb-type learning rule, which is defined as

$$\begin{aligned} \delta w_m(\mathbf{r}-\mathbf{r}_m, t+\tau) = & s_m(t+\tau)v(\mathbf{r},t) \\ & - F(v, w_m)w_m(\mathbf{r}-\mathbf{r}_m). \end{aligned} \quad (4)$$

Insertion of (3) into (4) and time averaging as defined above leads to

$$\begin{aligned} \Delta w_m(\mathbf{r}) = & \int G(\mathbf{r}-\mathbf{r}')w_m(\mathbf{r}')d\mathbf{r}' - f(\|w_m\|)w_m(\mathbf{r}) \\ & + \sum_n I(\mathbf{r}_{mn}) \int G(\mathbf{r}-\mathbf{r}')w_n(\mathbf{r}'-\mathbf{r}_{mn})d\mathbf{r}', \end{aligned} \quad (5)$$

where  $\mathbf{r}_{mn} = \mathbf{r}_n - \mathbf{r}_m$ ,  $G(\mathbf{r}) \equiv \langle v(0,t)v(\mathbf{r},t) \rangle_t$  is the spatial correlation function of the input cell activities and  $f(\|w_m\|) = \langle F(v, w_m) \rangle_t$  is a general decay function, which is assumed to depend on the norm  $\|w_m\|$  of the synaptic weight function only. The last term in (5) describes the contribution of the intercortical connections to the learning process. It depends on a formal lateral interaction function

$$I(\mathbf{r}) := I'(\mathbf{r})g(\tau). \quad (6)$$

Since  $\tau$  represents the time constant of the neural dynamics, it should be similar to the correlation time of the spontaneous photoreceptor activities. Therefore, the time correlation function  $g(t)$  will vanish for time intervals much longer than  $\tau$ , i.e.,  $g(\tau) \ll g(0) = 1$  and  $g(t) \cong 0$  for  $t \geq 2\tau$ . With these assumptions, all higher terms of the lateral interaction in (3), which correspond to delays  $t \geq 2\tau$ , may be neglected in Eq. (5). In addition, the

lateral interaction function  $I(\mathbf{r})$  is small compared to the biologically motivated intracortical connection function  $I'(\mathbf{r})$  and can therefore be treated as a small perturbation of the time averaged learning dynamics. Its strength is measured by  $I^2 = \int I^2(\mathbf{r})d\mathbf{r}$ . The spatial correlation function of the input neurons  $G$  can be calculated as for the deterministic model in Stetter, Lang, and Müller [27]. The result for  $G$  is a combination of three Gaussians with radii  $\sqrt{2}R_{gc}$ ,  $\sqrt{2}R_{gs}$ , and  $R_{gcs} = (R_{gc}^2 + R_{gs}^2)^{1/2}$ .

To investigate the maturation behavior of the network, it is convenient to express the learning dynamics in terms of the eigenstates of the input correlation function  $G$  [26]. For this step, translational invariance of the correlation function is assumed. To perform the transformation to the eigenrepresentation of  $G$ , the synaptic fields are expanded with respect to the eigenstates  $c_\beta$  of the correlation function

$$\begin{aligned} w_m(\mathbf{r}) = & \sum_\beta a_{m,\beta} c_\beta(\mathbf{r}), \\ \text{with } \int d\mathbf{r}' G(\mathbf{r}-\mathbf{r}')c_\beta(\mathbf{r}') = & \lambda_\beta c_\beta(\mathbf{r}). \end{aligned} \quad (7)$$

Each quantum number  $\beta$  incorporates the radial and angular node indices  $(n, l)$  of the eigenlevel [29] and a number  $\kappa$ , which discriminates between degenerate eigenstates. Note that in general the learning rule does not lead to normalized synaptic fields, i.e.,  $\sum_\beta a_{m,\beta}^2 \neq 1$ . Replacing the weight functions  $w_m(\mathbf{r})$  in (5) by (7), multiplying by  $c_\alpha(\mathbf{r})$ , and integrating over  $\mathbf{r}$  yields

$$\begin{aligned} d(a_{m,\alpha})/dt = & [\lambda_\alpha - f(\|\mathbf{a}_m\|)]a_{m,\alpha} \\ & + \sum_n \sum_\beta Q_{\alpha\beta}(\mathbf{r}_{mn})a_{n,\beta}, \end{aligned} \quad (8)$$

$$Q_{\alpha\beta}(\mathbf{r}) = I(\mathbf{r})C_{\alpha\beta}(\mathbf{r}), \quad (9)$$

$$C_{\alpha\beta}(\mathbf{r}_{mn}) = \int d\mathbf{r} d\mathbf{r}' c_\alpha(\mathbf{r})G(\mathbf{r}-\mathbf{r}')c_\beta(\mathbf{r}'-\mathbf{r}_{mn}), \quad (10)$$

where  $\mathbf{a}_m$  is the vector with components  $a_{m,\alpha}$ .  $C_{\alpha\beta}(\mathbf{r}_{mn})$  is proportional to the correlation of the activities shown by the cortical (output) neurons  $m$  and  $n$  if their afferent synaptic fields are described by the eigenstates  $c_\alpha$  and  $c_\beta$ , respectively. It will be referred to as a *partial cortical correlation function*. In (8), the learning dynamics is determined by the time dependent behavior of the expansion coefficients  $a_{m,\alpha}(t)$ . The first term on the right hand side of (8) dominates at the beginning of the training process and leads to a growth of the norm values  $\|\mathbf{a}_m\|$ . At an intermediate stage, the first term decreases more and more and the learning process is dominated by the second term describing the influence of the lateral interaction. Then the development of cortical receptive field profiles is determined by the lateral interaction function  $I(\mathbf{r})$  and the matrix of the partial cortical correlations  $C_{\alpha\beta}(\mathbf{r})$ , which together form the lateral coupling matrix  $Q_{\alpha\beta}(\mathbf{r})$ . The coefficient  $a_{m,\alpha}$  tends to grow due to the influence of eigenstate  $\beta$  mixed to the receptive field of neuron  $n$  if the matrix element  $Q_{\alpha\beta}(\mathbf{r}_{mn})$  is positive, i.e., if the lateral interaction  $I(\mathbf{r}_{mn})$  and the partial cortical correlation  $C_{\alpha\beta}(\mathbf{r}_{mn})$  have the same sign. This may be explained by

the fact that for positive  $C_{\alpha\beta}$ , neurons  $m$  and  $n$  tend to fire more synchronously due to the presence of states  $\alpha$  and  $\beta$  in their receptive fields, respectively. For a positive lateral interaction, large activities of both neurons will therefore mutually strengthen each other, leading to a stabilization of the synaptic structures  $c_\alpha$  and  $c_\beta$  contributing to these activities.

### B. Formal solution for weak lateral interactions

In this section it will be shown that for weak lateral interactions most of the expansion coefficients  $a_{m,\alpha}$  can be neglected. For a vanishing lateral interaction, the stable fix points of (5) and (8) are eigenstates  $\mathbf{w}_m$  to the largest eigenvalue  $\lambda_0$  of the input cell correlation function (principal components):

$$\mathbf{w}_m = \sum_{\kappa} a_{m,0\kappa} \mathbf{c}_{0\kappa}, \quad \mathbf{G} \mathbf{c}_{0\kappa} = \lambda_0 \mathbf{c}_{0\kappa}, \quad f(\|\mathbf{a}_m\|) = \lambda_0. \quad (11)$$

In this notation, the index  $\alpha \equiv 0$  denotes the quantum numbers  $(n, l)$  characterizing the state with the largest eigenvalue, and a separate index  $\kappa$  discriminates the degenerate states of this eigenlevel. A solution of  $f(\|\mathbf{a}_m\|) = \lambda_0$ , which determines the norm of the resulting synaptic field, is assumed to exist. Further, a shorthand operator notation is used, and functions and integral operators are printed in bold face. Only nondegenerate  $(n, 0)$  and twofold degenerate  $(n, 1)$  principal component eigenstates exist [27], which can be written as  $c_0(\mathbf{r}) = g_0(r)$  for  $(n, 0)$  domains and  $c_{0,1}(\mathbf{r}) = g_n(r) \cos(\varphi)$ ,  $c_{0,2}(\mathbf{r}) = g_n(r) \sin(\varphi)$  for  $(n, 1)$  domains [29].

Now a finite lateral interaction is considered, which is assumed to be weak compared to the average strength of afferents projecting onto cortical neurons, i.e.,  $I_r := \|I\| / \langle \|\mathbf{a}_m\| \rangle_m \ll 1$  (see Sec. II A). To obtain first-order solutions with respect to the relative strength  $I_r$  of the lateral interaction, normalized states  $\psi_m(\mathbf{r}) := w_m(\mathbf{r}) / \|\mathbf{a}_m\|$  and the perturbation function  $V_m(\mathbf{r}) := U_m(\mathbf{r}) / I_r$ , with

$$U_m(\mathbf{r}) = \sum_n I(\mathbf{r}_{mn}) / \|\mathbf{a}_m\| \int d\mathbf{r}' G(\mathbf{r} - \mathbf{r}') w_n(\mathbf{r}' - \mathbf{r}_{mn}),$$

are introduced. Then, the fix points of (8) satisfy

$$\mathbf{G} \Psi_m + I_r \mathbf{V}_m = \mu_m \Psi_m, \quad f(\|\mathbf{a}_m\|) = \mu_m. \quad (12)$$

In analogy to perturbation theory, one may write the first-order solutions and eigenvalues as

$$\Psi_{m,\alpha} := \sum_{\kappa} a_{m,\alpha\kappa} \mathbf{c}_{\alpha\kappa} + I_r \Psi_1, \quad \mu_m := \lambda_0 + I_r \mu_1. \quad (13)$$

These expressions are inserted into (12) and ordered by powers of  $I_r$ . Then, for nonzero interaction strength, the principal components transform into the following first-order fix points:

$$\Psi_m = \sum_{\kappa} a_{m,0\kappa} \mathbf{c}_{0\kappa} + \sum_{\beta \neq 0, \kappa} a_{m,\beta\kappa} \mathbf{c}_{\beta\kappa}, \quad (14)$$

$$a_{m,0\kappa} = (\mathbf{c}_{0\kappa} \cdot \mathbf{U}_m) / N_m,$$

$$a_{m,\beta\kappa} = (\mathbf{c}_{\beta\kappa} \cdot \mathbf{U}_m) / (\lambda_0 - \lambda_\beta), \quad (15)$$

$$\beta \neq "0",$$

$$f(\|\mathbf{a}_m\|) = \lambda_0 + N_m, \quad (16)$$

where  $(\cdot)$  denotes the scalar product defined by the integral over the two inserted functions, and  $N_m^2 = \sum_{\kappa} (\mathbf{c}_{0\kappa} \cdot \mathbf{U}_m)^2$  assures normalization of the vector  $\mathbf{a}_{m,0}$ . The first-order solutions obtained using this method are no longer orthogonal. Yet if the separation of the largest eigenvalue from smaller ones is large compared to the perturbation, i.e.,  $\lambda_0 - \lambda_\beta \gg \|\mathbf{U}_m\| \forall m, \beta \neq "0"$ , the solutions for  $\alpha \equiv "0"$  given in (14)–(16) remain the only stable ones. Note that this condition is best fulfilled for the all excitatory (0,0), the bilobed (0,1), and the circularly symmetric (1,0) receptive field profiles [27].

Equation (15) shows that the expansion coefficients  $a_{m,\beta\kappa}$  for nonprincipal component states  $\beta$  decrease in size with increasing difference of the corresponding eigenvalues  $\lambda_\beta$  from the principal eigenvalue  $\lambda_0$ . Therefore, only the coefficients of the few eigenstates with the largest eigenvalues are important for the description of the resulting first-order synaptic fields. This fact strongly reduces time consumption of corresponding computer simulations.

In nondegenerate (0,0) and (1,0) domains, the receptive fields are only changed by small admixtures of nonprincipal component states to the rotationally symmetric zeroth-order principal component. Therefore, only small orientation selectivity and hardly developed orientation preference maps can emerge for weak lateral interaction. In particular, the assumed three-lobed receptive fields with  $180^\circ$  symmetry, which were used by Linsker [26] to reproduce the formation of  $\pm \frac{1}{2}$  vortices in orientation preference maps, can be found in the present model only for sufficiently strong lateral interaction. Hence, in the weak coupling approximation, the most interesting case is the formation of orientation preference maps in the twofold degenerate (0,1) parameter domain. Here the first-order solution points into the same direction as the projection of the perturbation function onto the principal component subspace. For the special case of an excitatory lateral interaction, this means that the receptive field profile of neuron  $m$  forms in a way that maximizes the correlation of its activity with the average activities of the surrounding neurons (as pointed out by Linsker [26] for a similar model).

Finally, it must be mentioned that the perturbation  $\mathbf{U}_m$  acting on the receptive field of neuron  $m$  depends on the synaptic fields of all other neurons  $n$ . Therefore, (14)–(16) can only give a formal solution of (8), while explicit training results must be obtained numerically.

### C. Energy function for the (0,1) domain

This section develops an energy function for the learning dynamics in considering the special case of a very weak lateral interaction in the (0,1) domain. This approach neglects variations of orientation selectivity, but allows a semianalytical and very efficient description of stable orientation preference maps, which represent local

energy minima. For the description of the learning dynamics, only the two orthogonal functions spanning the principal component subspace, namely  $c_x(\mathbf{r})=g_0(r)\cos(\varphi)$  and  $c_y(\mathbf{r})=g_0(r)\sin(\varphi)$ , are considered, i.e., all coefficients except  $a_{m,01}:=A_m\hat{a}_{m,x}$  and  $a_{m,02}:=A_m\hat{a}_{m,y}$  are neglected in Eq. (14). This corresponds to first-order perturbation theory for degenerate states. In the present notation, the norm  $A_m:=\|\mathbf{a}_m\|$  is written separately. Because  $\hat{a}_{m,x}^2+\hat{a}_{m,y}^2=1$  holds, these coefficients can be written as  $\hat{a}_{m,x}:=\cos(\varphi_m)$  and  $\hat{a}_{m,y}:=\sin(\varphi_m)$ , respectively, and the first-order solution (14) becomes  $w_m(\mathbf{r})=g_0(r)\cos(\varphi-\varphi_m)$ . Thus, the angle  $\varphi_m$  is always orthogonal to the preferred orientation, while  $\hat{\mathbf{a}}_m:=(\hat{a}_{m,x},\hat{a}_{m,y})$  is the normalized orientation preference vector for the receptive field of neuron  $m$  (Fig. 2). Note that the receptive field profile is described by a pointed vector, which shows  $360^\circ$  symmetry, while the corresponding orientation preference has  $180^\circ$  symmetry. The total of all orientation preference angles of the cortical neurons fully determines the structure of an emerging orientation preference map. Proceeding from these assumptions, an energy function determining the time evolution of the orientation preference angles can be derived. With the help of the orientation preference vectors, (8) may be rewritten as

$$(dA_m/dt)\hat{\mathbf{a}}_m + A_m d\hat{\mathbf{a}}_m/dt = [\lambda_0 - f(A_m)] A_n \hat{\mathbf{a}}_m + \sum_n \mathbf{Q}(\mathbf{r}_{mn}) A_n \hat{\mathbf{a}}_n, \quad (17)$$

where  $\mathbf{Q}$  is the matrix containing the intracortical coupling coefficients (9). Because the orientation preference vector  $\hat{\mathbf{a}}_m$  is normalized to unity,  $d\hat{\mathbf{a}}_m/dt$  is orthogonal to  $\hat{\mathbf{a}}_m$ , i.e.,  $d\hat{\mathbf{a}}_m/dt = \mathbf{t}_m d\varphi_m/dt$  with  $\mathbf{t}_m := (-\sin(\varphi_m), \cos(\varphi_m))$ . Therefore, the radial and angular parts of (17) obtained by forming the scalar product with  $\hat{\mathbf{a}}_m$  and  $\mathbf{t}_m$ , respectively, are

$$dA_m/dt = [\lambda_0 - f(A_m)] A_m + \sum_n \hat{\mathbf{a}}_m \mathbf{Q}(\mathbf{r}_{mn}) \hat{\mathbf{a}}_n A_n, \quad (18)$$

$$d\varphi_m/dt = \sum_n (A_n/A_m) \mathbf{t}_m \mathbf{Q}(\mathbf{r}_{mn}) \hat{\mathbf{a}}_n. \quad (19)$$

From the approximate solution of (18),

$$A_m = A_0 \left\{ 1 + \sum_n \hat{\mathbf{a}}_m \mathbf{Q}(\mathbf{r}_{mn}) \hat{\mathbf{a}}_n A_n / [A_0^2 f'(A_0)] \right\},$$

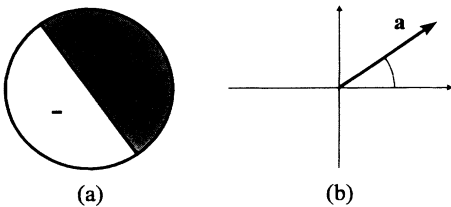


FIG. 2. (a) Simplified profile of a bilobed orientation selective (0,1) receptive field. “+” and “-” denote the sign of the synaptic field  $w$ . It is obtained as a linear combination of the two (0,1) basis states. (b) The orientation preference vector  $\mathbf{a}$  of the receptive field in (a). The vector is orthogonal to the preferred orientation of the cell.

with the unperturbed amplitude  $A_0$ , it follows that the amplitudes differ from the common zeroth-order amplitude only by a term of order  $O(I)$  (since the coupling matrix is of order  $I$ ). Therefore, the ratio of the amplitudes in (19) is close to unity in the case of a weak lateral interaction. Hence the radial and angular equations become approximately decoupled at an intermediate stage of the training process. If one neglects the small influence of the amplitudes in (19), i.e.,  $A_n/A_m \cong 1$ , the dynamics of the orientation preference angles is fully determined by the structure of the lateral coupling matrix  $\mathbf{Q}$ . The energy function, for which (19) performs gradient descent and the minima of which define stable orientation preference maps, then becomes

$$E(\varphi_1, \dots, \varphi_N) = -\frac{1}{2} \sum_{\substack{m,n \\ m \neq n}} \hat{\mathbf{a}}_m \mathbf{Q}(\mathbf{r}_{mn}) \hat{\mathbf{a}}_n. \quad (20)$$

It is worth mentioning that this energy function is similar to that obtained for the two-dimensional  $xy$  model, where interacting spins located on a two-dimensional grid rotate constrained to the plane of this lattice [29]. Since the unit matrix of the  $xy$  model is replaced by the coupling matrix  $\mathbf{Q}$ , the energy function (20) can be considered to describe a generalized two-dimensional  $xy$  model [30].

#### D. Evaluation of the coupling matrix

The partial cortical correlations  $C_{\alpha\beta}(\mathbf{r})$ , and therefore the coupling matrix elements  $\mathbf{Q}_{\alpha\beta}(\mathbf{r})$ , can be evaluated analytically. For convenience, gabor functions will be used to represent the eigenstates

$$c_x(\mathbf{r}) = \exp(-r^2/2P^2) \sin(kx), \\ c_y(\mathbf{r}) = \exp(-r^2/2P^2) \sin(ky).$$

Fitting  $P$  and  $k$  to numerically obtained (0,1) receptive field profiles corresponding to a projection radius  $\rho$  gives  $P \approx 0.5\rho$ ,  $k \approx 2/\rho$ . The evaluation of (9) using these functions yields

$$\mathbf{Q}_{\alpha\beta}(\mathbf{r}) = I(\mathbf{r}) [S_{\alpha\beta}(\mathbf{r}, R_{gc}) - 2S_{\alpha\beta}(\mathbf{r}, R_{gs}) + S_{\alpha\beta}(\mathbf{r}, R_s)], \quad (21)$$

$$S_{xx}(\mathbf{r}, R) = N \exp(-r^2/2R_g^2) [\cos(k_0x) - b^2], \quad (22)$$

$$S_{yy}(\mathbf{r}, R) = N \exp(-r^2/2R_g^2) [\cos(k_0y) - b^2], \quad (23)$$

$$S_{xy}(\mathbf{r}, R) = S_{yx}(\mathbf{r}, R) \\ = -2bN \exp(-r^2/2R_g^2) \\ \times [\sin(k_0x/2)\sin(k_0y/2)], \quad (24)$$

where  $R_g^2 = 2P^2 + R^2$ ,  $k_0 = (2P^2/R_g^2)k$ ,  $b = \exp[-k^2P^4/2(P^2 + R^2)]$ , and  $N$  is a norm factor depending on  $k$ ,  $P$ , and  $R$ .

From these results, the interaction energy  $E(\varphi_1, \varphi_2)$  between two cortical neurons can be evaluated. Consider two isolated neurons 1 and 2, where 1 is located at the origin with its orientation preference vector pointing to the positive  $y$  axis. The energy function (20) then reduces to  $E(\varphi_1 = \pi/2, \varphi_2(\mathbf{r}_{12}))$ , which only depends on the vec-

tor connecting the two output neurons. Minimizing this energy with respect to the orientation preference angle  $\varphi_2$  yields the equilibrium orientation preference of receptive field 2 as a function of the distance vector to the central neuron 1. The equilibrium angle  $\varphi_2 \equiv \varphi(\mathbf{r}_{12}) \equiv \varphi(\mathbf{r})$  is then defined by

$$\tan[\varphi(\mathbf{r})] = Q_{yy}(\mathbf{r})/Q_{xy}(\mathbf{r}), \quad (25)$$

with the additional constraint

$$Q_{xy}(\mathbf{r})\cos[\varphi(\mathbf{r})] + Q_{yy}(\mathbf{r})\sin[\varphi(\mathbf{r})] > 0. \quad (26)$$

The orientation preference vector fields obtained by replacing the matrix elements  $Q_{\alpha\beta}(\mathbf{r})$  with (21) are shown in Fig. 3(a) for all excitatory lateral interactions and in Fig. 3(b) for sombrero-type lateral interactions. The structures obtained simply express the fact that the lateral coupling maximizes the cortical correlation (which is roughly the overlap between the two receptive fields) for positive lateral interaction [Fig. 3(c)] and minimizes it otherwise. Further, these vector fields resemble the magnetic field of a classical magnetic dipole or the exchange field of the  $xy$  model. While these latter interactions lead to ferromagnetic equilibrium states, the structure of the field seen in Fig. 3 can induce more complex orientation preference patterns.

### III. RESULTS

The development of orientation preference maps was investigated for a neural network model as defined in Sec. II. The diameter of the output layer was chosen between 20 and 80 units and periodic boundary conditions were applied. The simulations were carried out using a discrete layer of input neurons with the same grid constant as the output layer. The network was trained using Yuille, Kammen, and Cohen's [31] learning rule, which is a special case of (5) with  $f(\|w\|) = \|w\|^2$ . For nearly all simulations, the projection radius was set to  $\rho = 6$ , which corresponds to about 100 afferent synapses per output neuron. To obtain zeroth-order solutions in the  $(0,1)$  domain,  $R_{gc} = \rho/3$  and  $R_{gs} = 2R_{gc}$  was used in most stimulations [27]. The lateral interaction function was represented either by Gaussian excitatory, sombrero, Gaussian inhibitory, or inverse sombrero-type profiles.

The center radius of these functions is denoted by  $R_c$  and the surrounding radius for sombrero-type functions is  $R_s = 2R_c$ . The strength of the lateral interaction was adjusted using the parameter  $I_0 = I(0)$ . For each simulation, the norm  $I$  (the total strength) of the lateral interaction function was scaled to the average norm of the resulting afferent synaptic fields. This relative interaction strength  $I_r := \|I\| / \langle \|w_m\| \rangle_m$  provides a measure of the strength of intracortical connections with respect to the afferents enervating any given cortical neuron. The interaction function was cut at  $2R$ , where  $R$  is the maximum Gaussian radius. Test simulations for larger cutoff radii did not lead to different results, justifying  $R_{\max} = 2R$ .

#### A. Behavior of the expansion coefficients

As a first step, the relative magnitudes of the expansion coefficients  $a_{m,\alpha}$  were determined. Simulations using the first 20 eigenstates were carried out and the average coefficients  $a_\alpha = \langle a_{m,\alpha} \rangle_m$  were calculated. Figure 4 shows the dependence of these coefficients on the distance of the corresponding eigenlevel from the principal component level. For the simulation, a sombrero-type lateral interaction with  $R_c = 1.5$  and an interaction strength corresponding to  $I_r = 0.84$  was used. First-order perturbation theory predicts, according to (15), a reciprocal behavior, but the coefficients show a stronger than reciprocal decrease that is more nearly exponential. This behavior can be understood from the structures of the external perturbation  $U$  and the eigenstates  $c_\alpha$ . For all cases, where continuous orientation maps evolve, the perturbation  $U$  is a smooth function with only a few zero crossings. On the other hand, within the parameter regimes considered here, the eigenstates  $c_\alpha$  show an increasing number of radial and angular nodes with increasing difference of their eigenvalues  $\lambda_\alpha$  from the principal eigenvalue  $\lambda_0$ . Therefore, their overlap with the perturbation function, which appears in the numerator of (15), will strongly decrease with the difference  $\lambda_0 - \lambda_\alpha$ . This tendency leads to a stronger than reciprocal decrease of the average coefficients in (15) and additionally justifies the neglect of all but the few highest eigenlevels for an approximate description of the network behavior.

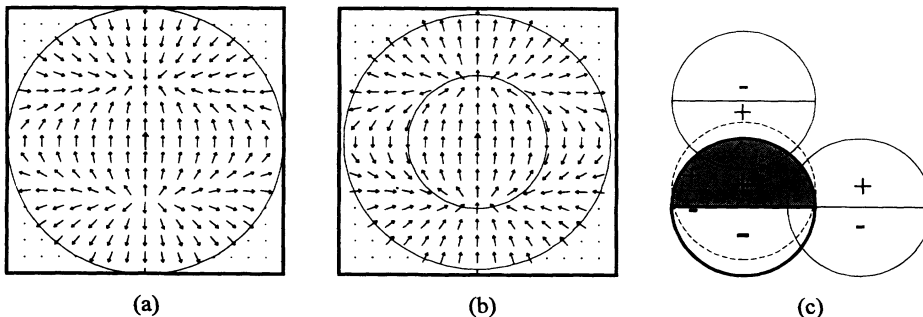


FIG. 3. Orientation of minimal energy as a function of the position for an output neuron, which only interacts with a central neuron with vertical orientation preference vector (a) for purely excitatory Gaussian lateral interaction and (b) for sombrero-type lateral interaction. (c) shows some receptive field profiles drawn for case (a). It can be seen that the receptive fields are oriented to obtain maximum overlap to the central receptive field.



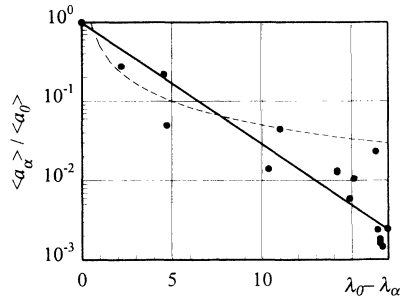


FIG. 4. The averaged expansion coefficients  $a_\alpha$  to the 20 highest eigenstates of the input correlation function for sombrero lateral interaction with relative strength  $I_r=0.84$ . The dashed line represents reciprocal behavior, the solid line an exponential dependence. The coefficients show exponential rather than reciprocal decrease with increasing deviation of the corresponding eigenvalue  $\lambda_\alpha$  from the principal (largest) eigenvalue  $\lambda_0$ . Due to this strong decrease, only the largest coefficients need to be considered to obtain a good approximate description of the network behavior.

### B. Structural organization of orientation preference maps

In this section and in Sec. III C, the orientation preference distributions that emerge in the model network are analyzed by systematically changing the parameters of the lateral interaction function and characterizing the emerging orientation preference map.

First, the type and range of the lateral interaction function were varied for the (0,1) domain and for a weak interaction strength  $I_0=0.01$ . In the simulations, the first eight eigenstates were used. Typical results are presented in Fig. 5, where the bars encoding the orientation preference of the nodes of the output layer are plotted. The length of the bars is related to the strength of the orientation preference, while their orientation is orthogonal to the preferred orientation of the corresponding output neuron [27]. In the special case of (0,1) receptive fields, they are parallel to the orientation preference vector as defined in Fig. 2. Since they characterize receptive fields with arbitrary spatial symmetry, the orientation preference bars do not express any directionality.

Within the (0,1) domain, only a few qualitatively different ordered structures were found. The most prominent ones are the radial and tangential vortices [Figs. 5(a) and 5(b)] and the parallel configuration. Further structures obtained are disordered vortices with nearly antiparallel vectors for neighboring cells [Fig. 5(c)] and the rarely appearing wave structures. Since only orientation preference is considered, parallel and antiparallel configurations lead to similar plots in Fig. 5. The parameter regimes leading to the corresponding structures are listed in Table I.

The orientation preference structures developing from inhibitory and inverse sombrero-type lateral interactions can be understood as follows. With very short interaction radii only nearest neighbors can interact. Then, within any mature orientation map the orientation preference vectors are antiparallel, which is optimal for inhibitory lateral interaction [as can be seen by inverting the arrows in Figs. 3(a) and 3(b)]. Note that in this configuration, the next nearest neighbors show parallel, nonoptimal arrangement with respect to each other. Slightly larger interaction radii lead to a weak interaction of these next nearest neighbors, which for the antiparallel structure contribute a small positive term to the total interaction energy. Minimal energy will therefore be obtained with a configuration, where the preferred orientations of the next nearest neighbors, and therefore of all neurons, are slightly tilted with respect to each other. This behavior can be seen in Fig. 5(c). For large interaction radii, in contrast, the positive interaction energy of the parallel next nearest neighbors is comparable to the negative next neighbor interaction term. Then, any approximately antiparallel configuration becomes unstable. Instead, all vectors avoid mutually parallel arrangements, which lead to the emergence of irregular structures with discontinuous behavior of the orientation preference angles. Since in biology the range of the lateral interaction most probably reaches beyond next neighbors, inhibitory and inverse sombrero lateral interactions are not suitable for the description of orientation map formation in the present model.

Purely excitatory interactions of long range as well as short range sombrero-type lateral interactions always lead to the formation of vortices. For both interactions,

TABLE I. The orientation preference maps obtained in different parameter regimes.

Interaction radius	Type of the lateral interaction function			
	Excitatory	Sombrero	Inhibitory	Inverse sombrero
Short range	Parallel	Tangential vortices irregularly arranged [Fig. 5(b)]	Irregular vortices [Fig. 5(c)]	Irregular vortices [Fig. 5(c)]
Long range	Radial vortices regularly arranged [Fig. 5(a)]	Wave structures Parallel radial vortices [Fig. 5(a)]	Irregular	Irregular



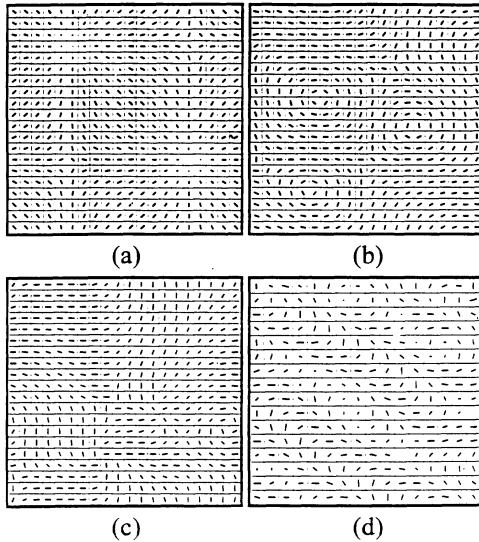


FIG. 5. (a)–(c) The emerging orientation preference maps for weak lateral interaction strengths  $I$ , between 0.015 and 0.025. (a) Radial vortices ( $R_c=2$ ), (b) tangential vortices ( $R_c=1.5$ ), (c) disordered vortices ( $R_c=1$ ). (d) Parameters as in (b) but with stronger and slightly elliptic lateral interaction functions. For (0,1) profiles, the bars are parallel to the orientation preference vector of the receptive field as defined in Fig. 2. The parameter regimes leading to each type of map are given in Table I.

$\pm 1$ -type vortices are found, where the orientation preference angle changes by  $\pm 360^\circ$  for every counterclockwise surrounding of a vortex center (for the notation, see also Baxter and Dow [17]).  $\pm \frac{1}{2}$ -type vortices cannot be obtained within the (0,1) domain, because this type of vortex requires the invariance of the receptive field profiles against  $\varphi \rightarrow \varphi + \pi$  symmetry operations. Only  $2n + 1$ -lobed (i.e., three-lobed) receptive fields obey this condition and lead to  $\pm \frac{1}{2}$  vortices (as shown by Linsker [26]). But in case of weak lateral interactions, three-lobed receptive fields do not form in the current network model under the learning algorithm considered in the simulations.

The organization of the orientation preference maps can be related to the structure of the interaction field shown in Figs. 3(a) and 3(b). Related with a short range excitatory lateral interaction (for example,  $I \neq 0$  only for nearest neighbors) are the minimum energy structures with a parallel arrangement of adjacent, hence all, vectors. If the interaction radius  $R_c$  exceeds  $\rho/3$ , i.e.,  $R_c > \rho/3$ , the nonzero area of the interaction field includes the two singularities of the field on the vertical axis that result from the zero crossings of the cortical correlation function  $C_{\alpha\beta}(r)$ . In this case, the number of vectors that are nearly optimally arranged with respect to each other is maximized if many vectors arrange in a radial vortex. The radial vortex structure of the field then translates into radial vortex patterns of all interacting output neurons of the network. Since the location of the singularities depends on the cortical correlation, and therefore on the projection radius  $\rho$ , the radius  $R_v$  of the

vortices is expected to scale with  $\rho$ . In fact,  $R_v$  exhibits a roughly linear dependence on the projection radius  $\rho$ , but only a weak dependence on the range  $R_c$  of the lateral interaction function.

The inherent underlying structure of the interaction field, in case of a sombrero-type lateral interaction, is a pair of distorted tangential vortices. They are arranged around the zero crossing of the lateral interaction function on the left and right hand side of the center of Fig. 3(b). Therefore, a tangential vortex leaves most of the concerned orientation preference vectors in a nearly minimum energy arrangement with respect to each other. In this case, the distance of the singularities of the field depends on the interaction radius  $R_c$ . Therefore, the vortex radii should scale with  $R_c$  for sombrero-type lateral interactions, which was in fact corroborated by the simulations.

These two types of lateral interactions lead to piecewise continuous changes of the orientation preference and also to zero- as well as one-dimensional discontinuities, and in this respect resemble the behavior of orientation maps found in mammal cortical tissue. Given these results, small uniform polarities in the Gaussian functions representing the input filters and polarities of the lateral interaction function were then introduced. A polar lateral interaction function leads to a deformation of the cortical interaction field. For not too small interaction strengths the vortices become prolonged [Fig. 5(d)]. Small polarities of the input filters induce anisotropic correlations of the input activities and directly force the unperturbed receptive fields to a unique fixed orientation preference angle. Under the influence of a weak lateral interaction, these polarities dominate and the system develops into a parallel configuration. For slightly stronger lateral interactions vortices form, but orientation preference angles near the angle supported by the input filter polarities become overrepresented. Note that for cats, locally uniform orientation biases of the retinal receptive fields were observed [32]. Hence the results given above suggest an overrepresentation of the preferred retinal orientation to occur in area 17.

### C. Evaluation of the energy function

The iterative simulations described above yield a qualitative characterization of the system behavior, i.e., they characterize the possible orientation preference structures that may be found. In this section, the energy per output neuron is calculated from (20) for isolated well-defined orientation preference structures, namely, the two types of vortices and the parallel structure observed in Sec. III B. All structures are defined on a two-dimensional grid with diameter  $2R_v$  [see Fig. 6(a) for a tangential vortex with  $R_v=5$ ]. This method represents an efficient way to examine the parameter domains, which lead to the emergence of vortex structures within the orientation preference maps. First, the influence of the absolute value of the projection radius  $\rho$  was investigated. The energy was calculated as a function of lateral interaction parameters for  $\rho=6, 8$ , and 10, while  $R_c/\rho$  and  $R_v/\rho$  were kept fixed. The relative behavior of the

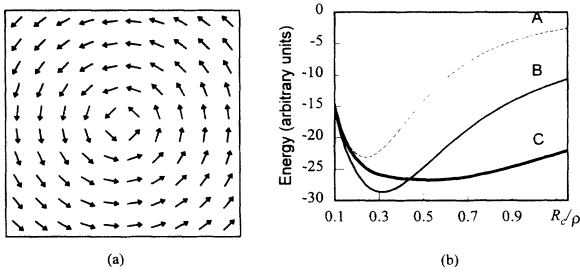


FIG. 6. (a) A predefined tangential vortex with radius  $R_v = 5$ . (b) The energy per neuron as defined in (20) for an isolated tangential vortex (A), a parallel structure (B), and an isolated radial vortex (C). The dependence on the range  $R_c$  of the purely excitatory lateral interaction function normalized to the projection radius  $\rho$  is shown. The curves for the parallel structure and the radial vortex intersect, which defines the boundary of two parameter domains.

energy per neuron was found to be nearly independent of the absolute size of the system. Therefore, the result of the training process is not much affected by the artificial grid constant of the output layer. Hence the important parameters are the relative values  $R_c/\rho$  and  $R_v/\rho$ , i.e., the radius parameters scale with  $\rho$ .

Figure 6(b) shows the dependence of the energy per neuron on the range of the excitatory lateral interaction for the three types of structures and a radius  $R_v = 5$ . It can be seen that the tangential vortex never has minimal energy for excitatory interactions. Since one can show that this structure is not a local energy minimum, it cannot be the result of iterative simulations. Further, there exists an intersection between the two lower curves. It separates two parameter domains where the parallel structure and the radial vortex, respectively, have minimal energy of all considered structures. Note that in general this is not necessarily the absolute minimum of the energy function. However, the results of the iterative simulations, which agree very well with the behavior of the energy curves in Fig. 6(b), suggest a high probability for the training process to be trapped in the corresponding minimum.

To obtain the optimal radius  $R_{v,opt}$  for the vortex structures, several calculations similar to those shown in Fig. 6(b) were carried out for different vortex radii. The parameter domains, where vortices form, were obtained

by evaluation of the intersections of the curves with lowest energies. In these domains, the optimum vortex radius  $R_{v,opt}$ , which minimizes the energy per neuron for a given lateral interaction, was determined. This radius is expected to provide an estimate of the vortex radii found in iterative simulations. The results of this procedure are plotted in Fig. 7(a) for excitatory lateral interaction and in Fig. 7(b) for sombrero-type lateral interaction. The estimated curve for the optimal vortex radius is inserted in all domains where vortices form. For excitatory lateral interactions, radial vortices form with values  $R_c/\rho > 0.5$ . Except near the domain boundary, the optimal vortex radius shows an approximately linear dependence on the projection radius  $\rho$ , as expected from the argumentation given in Sec. III B.

If, for sombrero-type structures, the lateral interactions have a short range, tangential vortices develop, whereas a large radius of the interaction leads to the formation of radial vortices. The latter result can be explained by the observation that according to (9) the inhibitory part of the interaction is cut off by the finite range of the cortical correlation function. Therefore, at the limit of large interaction radii, the sombrero system must behave like a Gaussian excitatory system. For fixed  $\rho$ , the radii of the tangential vortices are found to show a nearly linear dependence on  $R_c$ . This again agrees well with the suggestion made in Sec. III B.

These results were compared to those obtained by the evaluation of iterative simulations with more than two eigenstates and  $I_0 = 0.01$ . The domain boundaries could be quantitatively reproduced for systems with sombrero-type couplings. In addition, explicit measurements of the vortex radii were performed for iterative simulations and showed qualitative agreement with the estimated curves. In contrast, the domain boundaries obtained for excitatory lateral interaction could not be reproduced quantitatively, although the qualitative behavior agrees well. This is due to the fact that even for rather small interaction strengths the radial vortices tend to contain near (1,0) (circularly symmetrical) profiles in their centers. Therefore, the neglect of the (1,0) state is a good approximation only for very small interaction strengths. Nevertheless, the evidence of vortex formation for large radii is not affected by this fact.

Finally, the calculation of the energy function was repeated for predefined structures with periodic boundary conditions, where +1 and -1 vortices are arranged to

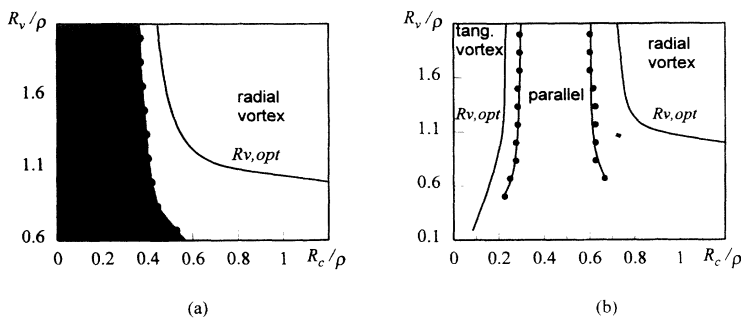


FIG. 7. The parameter domains, where the energy of isolated structures suggests the formation of radial vortices, tangential vortices, and parallel structures (a) for excitatory and (b) for sombrero-type lateral interaction. The inserted curves mark the estimated optimal vortex radius  $R_{v,opt}$  for the vortex domains. Except near the domain boundaries, the radial vortex radii approximately scale with  $\rho$ ; those of the tangential vortices scale with  $R_c/\rho$ .

yield a large continuous area of vortices. The domain boundaries and the qualitative behavior of the optimal vortex radius obtained for these systems agree well with those found for isolated vortices. Only the values for the optimum vortex radii show better quantitative agreement with those obtained for the iterative simulations. Therefore, the qualitative system behavior can be explained predominantly by the structure of isolated vortices, and the interaction between different vortices arranged within the output layer is evidently of minor importance only.

#### D. Stronger lateral interactions: Structured orientation selectivity maps

While the results given above concentrate on the orientation preference maps forming for weak lateral interaction, it will now be shown that with increasing strength of the lateral interaction, patches of reduced orientation selectivity emerge, which coincide with the singularities of the simultaneously developing orientation preference map. First, the strength of the lateral interaction function was varied for excitatory Gaussian and sombrero lateral interactions within the (0,1) domain, and the behavior of the expansion coefficients, averaged over all neurons of a single network, was analyzed. The dependence of the average expansion coefficients of the eight highest eigenlevels on the relative strength  $I_r$  of the lateral interaction is shown in Fig. 8(a) for excitatory interactions and in Fig. 8(b) for sombrero-type interactions. In both graphs, the coefficients show an approximately linear dependence on  $I_r$  for weak interaction strengths and become saturated for stronger  $I_r$ , where the condition  $\|U\| \ll \lambda_0 - \lambda_\beta \forall \beta$  is no longer fulfilled. For excitatory lateral interactions a qualitative change of the system behavior is found at  $I_r \cong 0.08$ . Below this value, all receptive fields approximately show (0,1) profiles, their correlation maximized by reorientation of the orientation preference vectors (structured orientation preference map). Above this value, phase shifts of the receptive fields maximize the correlation and result in parallel orientation preference vectors. This allows us to obtain  $I_{r,\max} \cong 0.08$  as a quantitative upper level for the relative interaction strength of weak lateral interactions. Phase shifted synaptic fields contain strong admixtures of the

(1,0) eigenstate and the two degenerate (0,2) eigenstates with different symmetry axes. Which of the two (0,2) states is more strongly represented in the synaptic fields depends on the direction of the orientation preference vectors, which is constant for the parallel structure. In the particular simulations in Fig. 8(a), the coefficient  $a_4$  dominates  $a_5 \cong 0$  above the boundary, because the symmetry axis of the (0,2) receptive field belonging to  $a_4$  is closely aligned to the orientation preference angle of the parallel orientation map resulting in these simulations.

From Fig. 8(a), one can derive that the radial vortex structures obtained for excitatory lateral interaction become unstable if the strength of the lateral interaction is increased above  $I_r = 0.08$ . In contrast, the formation of tangential vortices induced by sombrero-type lateral interactions is nearly independent of the strength of the intracortical connectivity. For both cases, however, Fig. 8 shows that even for strong lateral interactions, all but the coefficients of the few highest eigenstates remain small (see also Fig. 4, where in fact the relative interaction strength was set to  $I_r = 0.84$ , which is not small compared to unity). As mentioned above, this is due to the strongly decreasing overlap of the eigenstates with the external perturbation in (15). Thus the analysis demonstrates that the present approximation is not really confined to weak lateral interactions but instead remains valid also for large lateral interaction functions, and therefore for even dominant intracortical synaptic coupling strengths, as they are probably present in cortical tissue [ $I(r) \ll I'(r)$ , see Eq. (6)].

For increasing  $I_r$ , an increasing part of the receptive fields in the center of both types of vortices exhibit nearly circularly symmetrical receptive fields [Figs. 9(a) and 9(b)]. This is due to the fact that near the center the perturbation function  $U(r)$  shows approximately rotational symmetry and supports the coefficients of the circularly symmetric ( $n,0$ ) states. In other words, at vortex centers a nearly rotationally symmetric environment is provided to the developing synaptic fields, which strengthens contributions from eigenstates with circular symmetry and therefore reduces orientation selectivity of the corresponding output neuron. The nonorientation selective centers are much more pronounced for radial vortices, since this configuration enables a continuous transition

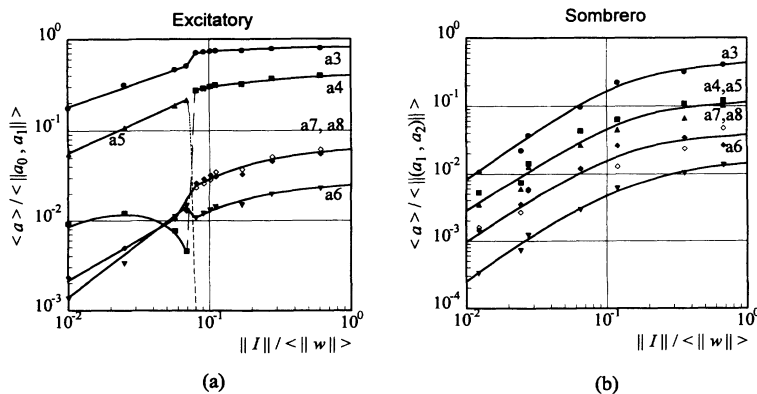


FIG. 8. Dependence of the eight largest averaged expansion coefficients on the strength of the lateral interaction  $I_r = \|I\| / \langle \|w_m\| \rangle_m$  (a) for purely excitatory interaction with  $R_c = 2$  and (b) for sombrero-type lateral interaction with  $R_c = 1.2$ . For both systems, the neglect of all but the few highest eigenstates for the system description is justified even for strong lateral interaction. The excitatory system shows a qualitative change of its behavior for  $I_r = 0.08$ , which yields a quantitative definition of weak lateral interaction in the sense of this paper. (●) =  $a_3$ , (■) =  $a_4$ , (▲) =  $a_5$ , (▼) =  $a_6$ , (◆) =  $a_7$ , (◇) =  $a_8$ .

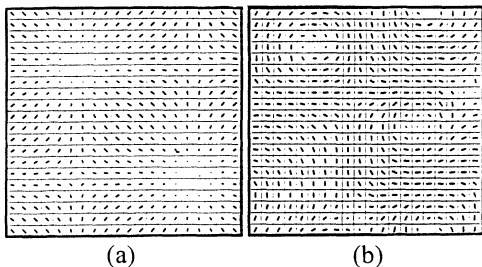


FIG. 9. Coordinated arrangement of orientation preference and orientation selectivity (a) for purely excitatory lateral interaction ( $R_c=2$ ) and (b) for sombrero-type interaction ( $R_c=1.5$ ) for systems with stronger lateral interactions ( $I_r=0.057$  and  $I_r=0.75$ , respectively). One observes reduced orientation selectivity in the vortex centers and partly at the boundaries between different vortices. The orientation of the bars encodes the preference angle, their length the orientation selectivity for each output neuron.

from bilobed (0,1) to circularly symmetric (1,0) receptive fields. Note that in contrast to the model of Wörgötter and Niebur [20], where reduction of orientation selectivity at discontinuities is achieved by explicitly convolving the orientation preference map with an appropriate kernel, the present model contains an inherent mechanism for the coordinated arrangement of orientation preference and orientation selectivity. It suggests a general principle for the coupling of discontinuities in orientation preference maps and dips in orientation selectivity maps, which is also observed in biology [10].

Finally, the (1,0) and (0,0) domains were investigated. As mentioned above, no strong orientation selectivity can result in these domains in the case of very weak lateral interactions. With increasingly stronger, purely excitatory lateral interactions, (0,0) receptive fields themselves maintain their profiles, but with sombrero-type lateral interactions a stripe pattern of alternating sign of the receptive field profiles is created. In the (1,0) domain, the receptive fields partially deform to three-lobed profiles [due to admixture of (0,2) states]. However, instead of expressing  $\pm \frac{1}{2}$  vortices as found by Linsker [26], the lateral interaction causes phase shifts of the receptive field profiles, which then orient into the parallel configuration.

#### IV. DISCUSSION

The main purpose of this paper is to investigate the training behavior of a simple neural network, the architecture of which is derived from experimental data for the anatomical structure of the prenatal early visual pathway in mammals. One of the most prominent features of this anatomically motivated network is a retinotopic projection of the input neurons to the output layer. Retinotopic axonal terminals lead to spatially restricted receptive fields of the output neurons, which show roughly circular symmetry and are shifted against each other in the same way as the output neurons they belong to. This approach is based on the chemoaffinity hypothesis [33], which suggests the development of retinotopy by mecha-

nisms of directed axonal growth rather than by synaptic plasticity. The network is trained using a general Hebb-type learning rule as introduced earlier [27], which at least for weak lateral interactions approximately extracts the principal component from the ensemble of input patterns. Since prenatal development is considered, these input patterns are spatially uncorrelated at the level of the photoreceptor layer and model spontaneous photoreceptor activity.

The present network represents an inductive model in that it first determines the architecture and the learning principle of the network according to biological data and then characterizes the structural features, in the present context the orientation maps, that are found to arise in the system during the training process. This approach is in contrast to models which are explicitly designed to show the structural features found in adult mammals [18–20] or to reproduce them by conveniently chosen developmental processes [21–25].

Von der Malsburg and Cowan [22] use an anatomically motivated network with predefined wave patterns of cortical activity and genetically predetermined subsets of orientation selective neurons. A training algorithm is used, which leads to iso-orientation domains with structures similar to the cortical activity patterns. These patterns, however, are introduced externally, so that this model cannot predict the structure of orientation maps from inherent developmental principles.

The latter is achieved by Swindale [23,24], who represents orientation selectivity as a two-dimensional vector. The components of this vector develop under interactions that tend to adjust the preferred orientation of each considered neuron along the average orientation of the surrounding vectors. Thus he implements an iterative smoothing procedure of the orientation preference map that is similar to the approach given by Baxter and Dow [17]. As mentioned by the author himself, this model makes no effort to relate these mechanisms to biologically plausible self-organization processes of cortical receptive fields.

Obermayer, Blasdel, and Schulten [25] present a Kohonen network, where the five-dimensional input vector encodes retinotopic position, orientation preference, orientation selectivity, and ocular dominance. These features are projected onto a two-dimensional output layer via a Kohonen algorithm applied to the weight vectors belonging to each output neuron. The authors give a sophisticated analysis of the structures formed by the model and show that orientation preference, orientation selectivity, and ocular dominance maps form in a coordinated manner if the average orientation selectivity or the average ocular dominance within the ensemble of input patterns (the order parameters of the model) exceed a threshold value. In this case, their results can be closely related to columnar structures found in biology. Since before birth neither strong orientation selectivity nor ocular dominance of the random stimuli is given, only fluctuations of the maps around a flat equilibrium distribution with binocular nonorientation selective cells are predicted by the authors for the prenatal stage of development. Thus the model is not suited to model prenatal formation

of orientation selectivity as well as its ordered arrangement (as observed for the monkey by Wiesel and Hubel [6]), which is the focus in the present work. Further, the Kohonen algorithm adjusts formal synaptic weights, which represent properties of cortical profiles and are not directly related with the development of biological synaptic coupling strengths.

One common feature which is characteristic to the latter two models is the explicit introduction of a  $180^\circ$  symmetry to the receptive field properties of their model neurons. This is done by doubling the polar angle of the preferred orientation, which was introduced by Swindale [23] and repeated in more recent approaches [24,25]. In these models, the developmental process describes the time dependence of each orientation preference as a function of all other preferences. Because these quantities remain identical for changes about  $\pm 180^\circ$ , the developing orientation preference maps also show this symmetry. In particular,  $\pm \frac{1}{2}$  vortices are commonly found in the orientation preference maps reported by these authors. These models, though correctly reproducing orientation preference maps found in mammal area 17, neglect the fact that in biology the development of orientation preference and selectivity occurs via the coordinated self-organization of receptive field profiles. These profiles, however, do not generally show  $180^\circ$  symmetry, though their orientation preference does. Therefore, it seems difficult to relate these models to a developmental process based on plasticity of cortical receptive field profiles as it is probably implemented in biological systems. This is confirmed by the simulation results observed in the present network, which focus on the development of cortical receptive field profiles and the concomitant formation of orientation preference and orientation selectivity maps. The most stable orientation preference maps consist of  $\pm 1$  vortices, which reflect the  $360^\circ$  symmetry of the most stable bilobed cortical receptive fields.

A linear feed-forward neural network with an anatomically motivated architecture, where the input neurons do not represent formal features but model biological neurons of the visual pathway, was given by Linsker [26] (see also [34]). The author showed that Hebb training of initially random synaptic weights with uncorrelated white noise can lead to orientation maps with  $\pm \frac{1}{2}$  vortices as elementary structures (note that Linsker's model does not perform principal component analysis). This result is achieved using a simplified description of the learning dynamics, where the structure of the receptive fields is fixed to three-lobed profiles (which also show  $180^\circ$  symmetry), the orientation preference angles of which change according to the influence of neighboring neurons, thus forming orientation preference maps. This model, in contrast to the network of the present paper, assumes the appropriate symmetry of the individual receptive fields, which translates into the emerging orientation map. Further, since all profiles are identical, Linsker's model does not allow variations of orientation selectivity (which requires deformation of the profiles) and therefore cannot model the development of orientation selectivity maps. Finally, the approach is restricted to weak lateral interactions, which is in contrast to biological evidence (in mon-

keys, each neuron receives about 2300 synapses [28], where at most 300 belong to afferent fibers [27]) and to purely excitatory intracortical connections.

The present model takes into account the structural development of cortical receptive fields, the profiles of which develop during the maturation process. Since this procedure is computationally expensive, the receptive fields were expanded with respect to the eigenstates of the correlation function of the neural activities within the input layer, and the training algorithm was transformed to a set of equations determining the time evolution of the expansion coefficients. In this representation, it turns out that only a few expansion coefficients can take large values in a mature system, since the symmetry of the remaining eigenstates differs strongly from the environmental symmetry determined by the lateral interaction and the partial cortical correlation function. Thus even for strong lateral interactions, only the few highest eigenstates need to be taken into account for the training process, which strongly reduces simulation times. Therefore, the present model neural network provides an efficient way to simulate the coordinated development of cortical receptive fields, which in turn determines the structure of orientation preference as well as orientation selectivity maps.

For the implemented Hebb-type learning rule, it was found that the time averaged influence of intracortical connections to the learning dynamics remains small compared to the dominant contribution from the afferent synaptic weights. This results from a weak correlation of intracortical and afferent signals arriving at a given cortical neuron in contrast to a strong correlation between afferent signals to this neuron, provided the input patterns do not vary too slowly. A weak interaction approximation could therefore be applied to the learning algorithm, which renders the observed quenching of all but the highest eigenstates plausible even in the case of biologically relevant strong lateral couplings.

Using this approximation, it was found that three-lobed receptive fields and concomitant orientation maps with  $180^\circ$  symmetry do not form in our model in the case of weak lateral interactions. But the application of strong interactions, where partly three-lobed receptive fields evolve, leads to the development of all parallel orientation maps instead of vortex structures. Thus Linsker's results (as well as those of other models that presuppose  $180^\circ$  symmetry of their elements) cannot be considered as inherent features of an anatomically motivated network with freely developing cortical receptive field profiles. Instead, structured orientation preference maps were observed at the (0,1) domain, where the stable bilobed receptive fields form. Due to the symmetry of these profiles, the emerging orientation maps contain  $\pm 1$  vortices. For the twofold degenerate (0,1) eigenstate, the weak lateral interaction approximation can be simplified to obtain an energy function, the minima of which correspond to stable orientation preference maps. While dropping variations of orientation selectivity, this approach focuses on a semianalytical characterization of the orientation preference maps that emerge for different lateral interaction functions.

Proceeding from these results, stronger lateral interactions were applied. In this case, reduced orientation selectivity was found to coincide with regions where the continuity of the orientation map is distorted. The coupling of both maps occurs via a change in environmental symmetry for neurons located near discontinuities. This leads to an increased admixture of nonprincipal component states to deformed resulting receptive fields and therefore to reduced orientation selectivity. Thus the coordinated formation of orientation preference maps with piecewise linear orientation drift interrupted by discontinuities, and a reduction of orientation selectivity at these discontinuities, is an inherent feature of our network (see, in contrast, [20]).

In the following, some quantitative features of the orientation maps emerging in the model are compared to those observed in mammal area 17. As shown in Blasdel and Salama [8] and Blasdel [9,10], orientation preference in monkey area 17 is arranged in systems of  $\pm\frac{1}{2}$  vortices with average center distances of about 500–600  $\mu\text{m}$ , which agrees with previously measured orientation drift rates [35] (see Sherk and Stryker [36] for kittens). The discontinuities coincide with regions of reduced orientation selectivity. From axonal degeneration in monkey area 17 after small intracortical lesions, lateral interactions have been estimated by Fiskens, Garey, and Powell [37] to show roughly sombrero shape with radii corresponding to  $R_c=0.5$  mm and  $R_s=1$  mm. However, the total range of the intracortical connections varies strongly between 0.2 mm in layer IV up to 4 mm in layer IIIc. Other estimates using punctual injections of horseradish peroxidase [38] also lead to a range of lateral connections up to 1–2 mm. Finally, physiological investigations on complex cells in cats' visual cortex showed that intracortical inhibition plays a considerable role for neural processing [39].

If one identifies the output layer of the model network with the cortical layer IV, where the lateral interaction shows very short range ( $R_c/\rho=0.6$ ,  $\rho=0.15$  mm and  $R_c/\rho=0.2$ ,  $\rho=0.5$  mm for parvo and magnocellular input, respectively [27]), the model which uses sombrero-type lateral interactions suggests the formation of tangential vortices with  $R_v/\rho\approx 0.8$  for the magnocellular system [see Fig. 7(b)]. The resulting vortex radius of  $R_v=0.3$  mm agrees even quantitatively with the neurobiological findings.

Further, comparison of the present simulation results to quantitative evaluations of orientation drift rates [17] also suggests that regions with  $\pm 1$  vortices might exist in area 17. The strongest reduction of orientation selectivity in the present model occurs in  $+1$  centers of the vortices (though also present at other discontinuities). The concentration to  $+1$  centers is most evident for the radial vortices obtained with purely excitatory lateral interactions but can also be found for tangential vortices (see Fig. 9). This result strongly supports the *E1* model of Baxter and Dow [17] ( $\pm 1$  vortices, where also the  $+1$  but not the  $-1$  centers correspond to patches of nonorientation selective cells), which in turn was found by these authors to yield the closest match to orientation drift rates obtained from tangential electrode penetrations in area

17.

However, some restrictions of the present model must be mentioned also. Its main disadvantage is its failure to reproduce the experimentally observed  $\pm\frac{1}{2}$  vorticity of the orientation preference maps. This shows that the  $180^\circ$  symmetry presupposed by other authors is difficult to implement in an anatomically motivated network architecture. In analogy to Linsker's model [26], the formation of  $\pm\frac{1}{2}$  vortices could easily be achieved by the present neural network if the cortical receptive fields were forced into profiles with even spatial symmetry (i.e., three-lobed profiles). Therefore, the fact that  $\pm\frac{1}{2}$ -vortex structures do not evolve in the present model can be reduced to the observation that cortical profiles with even spatial symmetry do not represent a stable solution in this system. Since the present approach uses a general Hebb-type learning rule and investigates several types of lateral interaction functions, this lacking stability of profiles with  $180^\circ$  symmetry was shown to appear in a whole class of anatomically motivated neural networks. The modification of this presently investigated network in a way that stable orientation selective cortical receptive fields with even spatial symmetry evolve (which in turn leads to the formation of  $\pm\frac{1}{2}$ -vortex structures in an anatomically motivated network architecture) is thus a challenge for future investigations.

Another restriction is that the quantitative agreement of drift rates to biological data can only be achieved for the magnocellular but not for the parvocellular system. One possible explanation of this phenomenon might be an earlier development of receptive field profiles for cortical neurons which receive magnocellular input. These neurons would first show correlated activities with respect to each other and be responsible for the formation of the orientation map. The development of the parvocellular part of layer IV might then be dominated by the existing magnocellular structure. A model that explicitly includes both magnocellular and parvocellular afferents might provide new insight into the coordinated development of magnocellular vortex-shaped orientation maps and the parvocellular nonorientation selective color system located in the vortex centers.

Further, the parameter domain for the formation of vortices is narrow, and the small energy difference between the tangential vortex structure and the totally parallel structure seems to indicate a lacking robustness of the results to perturbations such as randomized lateral interaction functions. Finally, in the present work, orientation selectivity was only defined via the structure of the afferent connectivity, which is only justified for weak lateral interaction. However, as shown by Wörgötter and Koch [40], the main influence of a strong lateral connectivity is probably the sharpening of the orientation tuning of cortical neurons, provided that orientation maps of the afferents do already exist. Therefore, it seems justified to consider the time development of afferent structures only.

## V. SUMMARY

In this work, a neural network model for the self-organization of simple orientation selective cortical receptive fields and their ordered arrangement into orienta-

tion preference and orientation selectivity maps was presented. In contrast to earlier models, it was designed to parallel the anatomy of the early visual pathway rather than to reproduce the experimentally observed orientation maps. It thus represents an inductive, anatomically motivated approach, where the formation of orientation maps occurs via the self-organization of cortical receptive field profiles under the influence of different intracortical connections. For this model, orientation preference maps consisting of vortex structures with singularities at their centers and fractures at their borders were found to emerge. Due to a coupling mechanism inherent to the model, patches of reduced orientation selectivity evolve simultaneously and coincide with the discontinuities of

the corresponding orientation preference map, which parallels recent biological findings. However, since three-lobed receptive fields with even spatial symmetry ( $180^\circ$  symmetry) do not represent stable solutions within the considered class of networks, no  $\pm\frac{1}{2}$  vortices as observed in area 17 were found to form in our model, only  $\pm 1$  vortices with  $360^\circ$  symmetry. Thus the present results offer a challenge for further investigations, which should focus on designing a modified, anatomically motivated network architecture yielding stable cortical receptive fields with even spatial symmetry. It should lead to a model for the formation of  $\pm\frac{1}{2}$ -vortex structures and superimposed orientation selectivity maps via the coordinated development of cortical receptive field profiles.

- 
- [1] D. H. Hubel and T. N. Wiesel, *J. Physiol. (London)* **148**, 574 (1959).
- [2] D. H. Hubel and T. N. Wiesel, *J. Physiol. (London)* **195**, 215 (1968).
- [3] D. H. Hubel and T. N. Wiesel, *Proc. R. Soc. London, Ser. B* **198**, 1 (1977).
- [4] P. H. Schiller, B. L. Finlay, and S. F. Volman, *J. Neurophysiol.* **39**, 1288 (1976).
- [5] J. J. Jones and L. A. Palmer, *J. Neurophysiol.* **58**, 1187 (1987).
- [6] T. N. Wiesel and D. H. Hubel, *J. Comp. Neurol.* **158**, 307 (1974).
- [7] D. H. Hubel, T. N. Wiesel, and M. P. Stryker, *J. Comp. Neurol.* **177**, 361 (1978).
- [8] G. G. Blasdel and G. Salama, *Nature* **321**, 579 (1986).
- [9] G. G. Blasdel, *J. Neurosci.* **12**, 3115 (1992).
- [10] G. G. Blasdel, *J. Neurosci.* **12**, 3139 (1992).
- [11] J. C. Horton and D. H. Hubel, *Nature* **292**, 762 (1981).
- [12] D. H. Hubel and M. S. Livingstone, *Soc. Neurosci.* **7**, 357 (1981).
- [13] M. S. Livingstone and D. H. Hubel, *J. Neurosci.* **4**, 309 (1984).
- [14] T. Bonhoeffer and A. Grinvald, *Nature* **353**, 429 (1991).
- [15] V. Braitenberg and C. Braitenberg, *Biol. Cybern.* **33**, 179 (1979).
- [16] B. M. Dow and R. Bauer, *Biol. Cybern.* **49**, 189 (1984).
- [17] W. T. Baxter and B. M. Dow, *Biol. Cybern.* **61**, 171 (1989).
- [18] N. V. Swindale, in *Models of the Visual Cortex*, edited by D. Rose and V. G. Dobson (Wiley, Chichester, 1985).
- [19] F. Wolf, K. Pawelzik, T. Geisel, D. S. Kim, and T. Bonhoeffer, in *Proceedings of the International Conference on Artificial Neural Networks, Amsterdam, The Netherlands, 1993*, edited by S. Gielen and B. Kappen (Springer, London, 1993).
- [20] F. Wörgötter and E. Niebur, *Biol. Cybern.* **70**, 1 (1993).
- [21] K. G. Götz, *Biol. Cybern.* **58**, 213 (1988).
- [22] C. von der Malsburg and J. D. Cowan, *Biol. Cybern.* **45**, 49 (1982).
- [23] N. V. Swindale, *Proc. R. Soc. London Ser. B* **215**, 211 (1982).
- [24] N. V. Swindale, *Biol. Cybern.* **66**, 217 (1992).
- [25] K. Obermayer, G. G. Blasdel, and K. Schulten, *Phys. Rev. A* **45**, 7568 (1992).
- [26] R. Linsker, *Proc. Natl. Acad. Sci. USA* **83**, 8779 (1986).
- [27] M. Stetter, E. W. Lang, and A. Müller, *Biol. Cybern.* **68**, 465 (1993).
- [28] J. O'Kusky and M. Colonnier, *J. Comp. Neurol.* **210**, 278 (1982).
- [29] R. Linsker, in *Proceedings of the International Joint Conference on Neural Networks, (IJCNN), Washington, DC, 1990*, edited by M. Caudill (Erlbaum, Hillsdale, NJ, 1990).
- [30] J. M. Kosterlitz and D. J. Thouless, *J. Phys. C* **6**, 1181 (1973).
- [31] A. L. Yuille, D. M. Kammen, and D. S. Cohen, *Biol. Cybern.* **61**, 183 (1989).
- [32] W. R. Levick and L. N. Thibos, *Nature* **286**, 389 (1980).
- [33] R. W. Sperry, *Proc. Natl. Acad. Sci.* **50**, 703 (1963).
- [34] D. J. C. MacKay and K. D. Miller, *Network* **1**, 257 (1990).
- [35] D. H. Hubel and T. N. Wiesel, *J. Comp. Neurol.* **158**, 267 (1974).
- [36] H. Sherk and P. Stryker, *J. Neurophysiol.* **39**, 63 (1976).
- [37] R. A. Fiskens, L. J. Garey, and T. P. S. Powell, *Brain Res.* **53**, 208 (1973).
- [38] K. S. Rockland and J. S. Lund, *J. Comp. Neurol.* **216**, 303 (1983).
- [39] C. Blakemore and E. A. Tobin, *Exp. Brain Res.* **15**, 439 (1972).
- [40] F. Wörgötter and C. Koch, *J. Neurosci.* **11**, 1959 (1991).



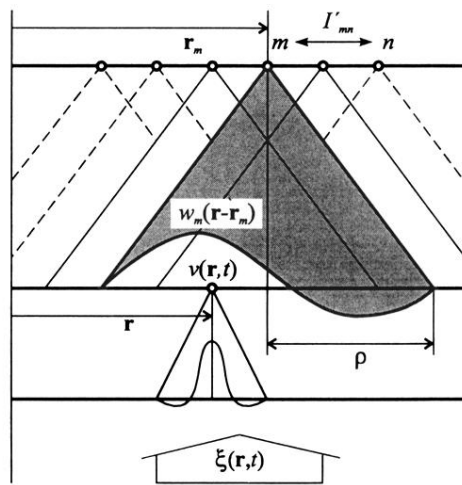
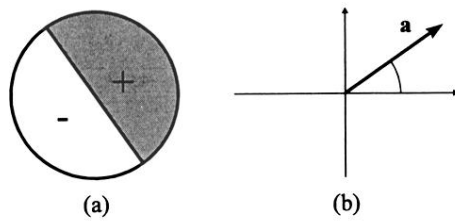


FIG. 1. Schematic plot of the neural network architecture used in this work. Uncorrelated white noise patterns  $\xi(\mathbf{r}, t)$  are convolved with the sombrero-type input filters and lead to activity distributions  $v(\mathbf{r}, t)$  of the input layer. The activities of the output neurons  $m$  are obtained by summation over these input activities weighted by the synaptic fields  $w_m(\mathbf{r}-\mathbf{r}_m)$ . The synaptic fields of different output neurons are shifted against each other to preserve strict retinotopic order. The output neurons  $m$  and  $n$  are connected via time independent lateral connections  $I'_{mn}$ .



**FIG. 2.** (a) Simplified profile of a bilobed orientation selective (0,1) receptive field. “+” and “-” denote the sign of the synaptic field  $w$ . It is obtained as a linear combination of the two (0,1) basis states. (b) The orientation preference vector  $\mathbf{a}$  of the receptive field in (a). The vector is orthogonal to the preferred orientation of the cell.

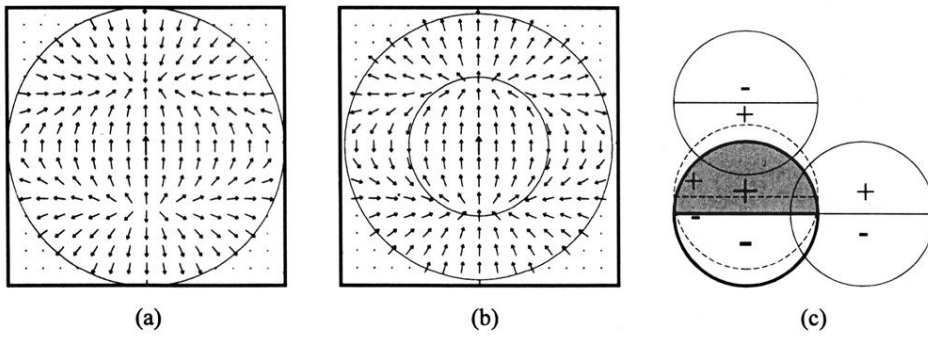


FIG. 3. Orientation of minimal energy as a function of the position for an output neuron, which only interacts with a central neuron with vertical orientation preference vector (a) for purely excitatory Gaussian lateral interaction and (b) for sombrero-type lateral interaction. (c) shows some receptive field profiles drawn for case (a). It can be seen that the receptive fields are oriented to obtain maximum overlap to the central receptive field.

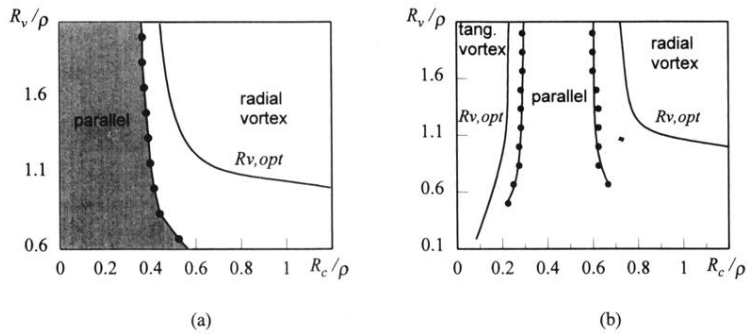


FIG. 7. The parameter domains, where the energy of isolated structures suggests the formation of radial vortices, tangential vortices, and parallel structures (a) for excitatory and (b) for sombrero-type lateral interaction. The inserted curves mark the estimated optimal vortex radius  $R_{v,opt}$  for the vortex domains. Except near the domain boundaries, the radial vortex radii approximately scale with  $\rho$ ; those of the tangential vortices scale with  $R_c/\rho$ .