Theory of ocular dominance pattern formation

O. Scherf, K. Pawelzik, F. Wolf, and T. Geisel
Max-Planck-Institut für Strömungsforschung, Göttingen, Germany
and SFB 185 ''Nichtlineare Dynamik,'' Universität Frankfurt am Main, Frankfurt, Germany
(Received 22 July 1998)

We investigate a general and analytically tractable model for the activity-dependent formation of neuronal connectivity patterns. Previous models are contained as limiting cases. As an important example we analyze the formation of ocular dominance patterns in the visual cortex. A linear stability analysis reveals that the model undergoes a Turing-type instability as a function of interaction range and receptive field size. The phase transitions is of second order. After the linear instability the patterns may reorganize which we analyze in terms of a potential for the dynamics. Our analysis demonstrates that the experimentally observed dependency of ocular dominance patterns on interocular correlations of visual experience during development can emerge according to two generic scenarios: either the system is driven through the phase transition during development thereby selecting and stabilizing the first unstable mode or a primary pattern reorganizes towards larger wavelength according their lower energy. Experimentally observing the time course of ocular dominance pattern formation will decide which scenario is realized in the brain. [S1063-651X(99)12705-3]

PACS number(s): 87.10.+e, 47.54.+r, 07.05.Mh

I. INTRODUCTION

Patterns of neuronal connections emerge and change depending on neuronal activity. It is widely assumed that this enables the brain to adapt its processing machinery to the structure of environmental stimuli, i.e., sensory experience, and to build precise and useful circuitry from initially crude and imprecise patterns of connections. The development of neuronal connections in the visual cortex is an important experimental model system for studying the mechanisms and principles involved in the refinement of neuronal circuitry [1–12]. This system has also been a focus of attempts to formalize the basic processes of the activity-dependent refinement of neuronal circuits which promise also a deeper understanding of the properties implied by such mechanisms on a network level (for reviews see Refs. [13,14]). In this study we investigate one central phenomenon in the development of visual cortical circuits from a pattern formation perspective, namely the emergence of the pattern of cortical domains specialized for processing information from the left or the right eye.

Our study refers to experimental observations in the visual cortices of cats and monkeys. In layer IV of the primary visual cortex, inputs from the left and right eye are segregated into spatially distinct domains called ocular dominance columns (ODCs) [15,16]. Neurons in individual domains preferentially respond to stimulation of either the left or the right eye [17,18]. In the primary visual cortex of cats ODCs form a roughly repetitive pattern [16,19–21]. During development the pattern arises between the third and the seventh postnatal week by gradually segregating the initially overlapping afferents of the two eyes [19,22]. Many lines of evidence indicate that this process is driven by activity-dependent competition for cortical territory between the geniculocortical afferents subserving the two eyes [6,11,23]. At the level of individual neurons and synapses this competition presumably results from an activity-dependent refinement of synaptic connections whereby ''improper'' connections are removed and ''appropriate'' connections are elaborated [24,25,10]. Recently it was shown that the spacing of ODCs in squinting cats was significantly larger than in normally raised animals [26] (Fig. 1). This dependence of ODC spacing on visual experience has also been suggested from model simulations [27] and similar observations have meanwhile been reported from cats that were raised with alternating monocular occlusion [28]. Because a global change in columnar spacing cannot be easily produced by shifting ocular dominance borders in a preexisting grid these experimental observations rather indicate that initially emerging patterns of ODCs form spontaneously and are not determined by yet unobserved prepatterns.

The development of differently spaced ODCs depending on visual experience is also intriguing from a theoretical point of view. One very elaborate and biological plausible framework for studying the emergence of columnar patterns is the class of correlation-based semilinear models developed by Miller and co-workers [29,30]. Within this model class it has been shown that the spacing of ODCs is insensitive to the structure of afferent activity and primarily determined by the extent of connections that link neurons in the cortical layer to their neighbors. At first sight this appears very plausible. Following Hebbian rules of synaptic plasticity [31], neurons that are simultaneously active will develop similar afferent connectivity patterns. Connections that link neighboring neurons will induce a tendency of local groups of neurons to be simultaneously active and therefore to develop connections to the same eye. Within correlation-based models the structure of afferent activity patterns only influences whether ODCs develop at all. Another prominent class of models are the so-called self-organizing maps (SOMs) [32–37,27]. These models, although biologically not very detailed, constitute a straightforward idealization of the important aspects of activity-dependent modifications: (1) Afferent activity patterns induce activity patterns within the cortex. (2) The selectivities of activated neurons are modified as a function of presynaptic and postsynaptic activities. Interest-

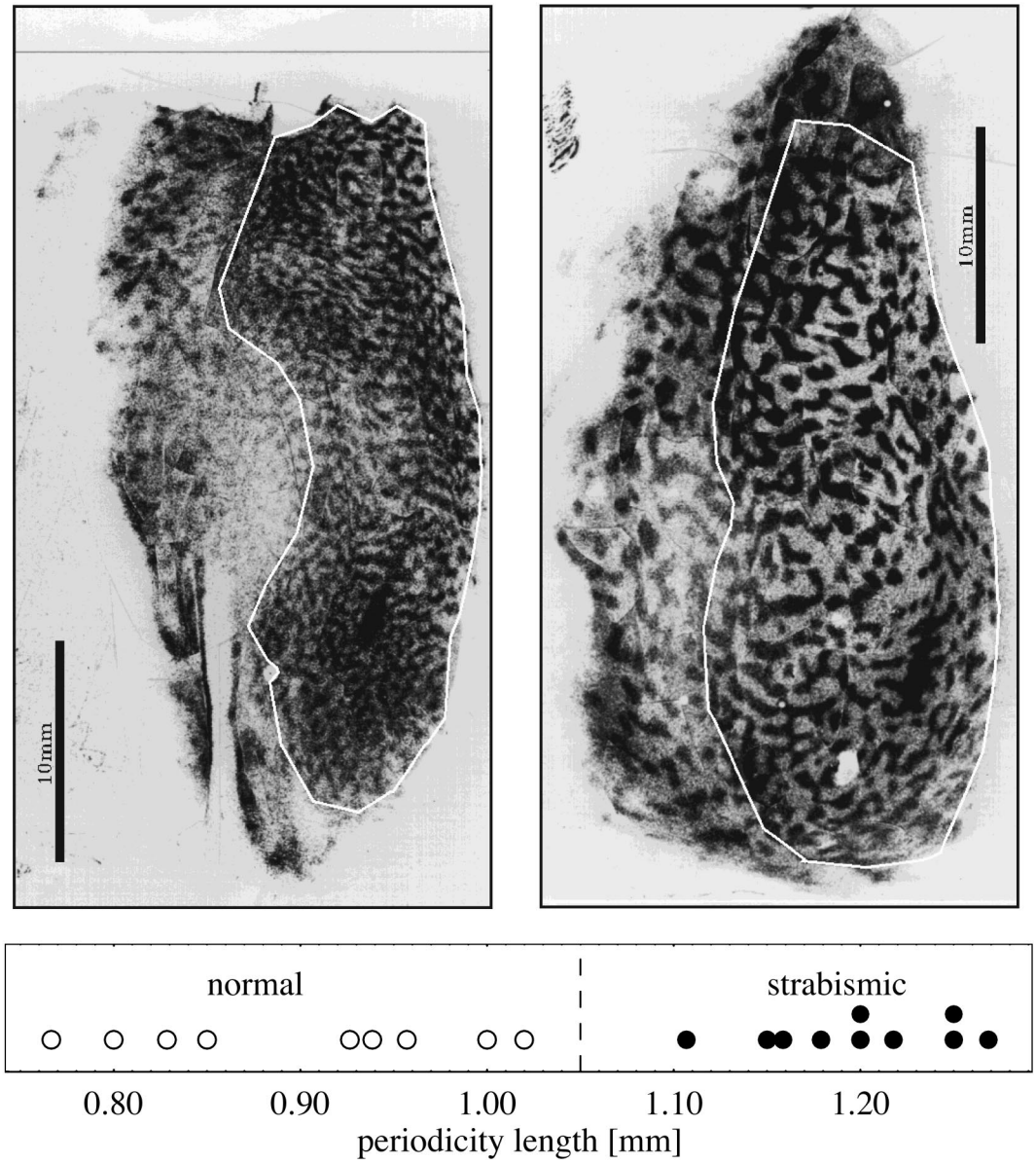

FIG. 1. Upper pictures: autoradiographs from the visual cortex of cats (with friendly permission by Siegrid Löwel). White lines enclose area 17. The dark regions indicate neurons that are more strongly activated by the right eye while the brighter regions indicate neurons driven by the left eye. The left picture shows an ocular dominance pattern of a normal cat and the right picture of a strabismic cat. The display below shows the typical length scales of the ocular dominance patterns from 19 autoradiographs (same data as in Ref. [26]).

ingly simulations of a model within this class suggested that the spacing of columns in squinting animals should be larger as compared to normal [27].

A central mathematical difference between the two modeling approaches is the treatment of cortical activity patterns. The derivation of correlation based models involves describing the activity in the cortical layer as a linear response to the afferent activity patterns [29,30]. In the family of SOM models activity in the cortical layer is described by a stereotyped local activity blob [32,33] and is therefore a highly nonlinear function of the afferent activity patterns. The rational behind this assumption is that in homogeneous layers of nonlinear model neurons with a lateral coupling of sombrero type, activity naturally appears in the form of spatially localized domains of activated neurons [32,33,38,39]. Recently, evidence has begun to emerge indicating that this picture may capture important aspects of cortical responses. Quantitative aspects of the selectivity of visual cortical neurons to the orientation of visual edge stimuli can be naturally explained when it is assumed that cortical responses are strongly shaped by intracortical interactions [40–42]. Ben-Yishai *et al.* have demonstrated that under such assumptions cortical responses appear in the form of rather stereotyped and localized activity patterns [42]. Also during development neurons in the cortex are active in local groups coupled electrically through so-called gap junctions [43]. SOM based models are useful starting point for studying the consequences of such strongly nonlinear activity responses for the activity-dependent development and plasticity of connectivity patterns. Analyzing this model class, we will show that the nonlinear nature of

cortical responses in fact appears as a central prerequisite to describe the dependence of ODC spacing on visual experience.

For an understanding of the map formation processes it appears to be useful to generalize some of the mostly used models rather than to add new models to the family of algorithms. Some work has been done in this respect [44–48], mostly based on the mathematical properties of the models.

In this contribution we start from general phenomenological principles of map formation and derive the ''convolution model,'' which includes the widely used Kohonen model [33] and the elastic net [49] as limiting cases. We then investigate the dynamics of pattern formation of the convolution model with particular emphasis on the correlation dependency of the characteristic length scale of the ocular dominance pattern. The analysis includes the linear and nonlinear stages of pattern formation and the bifurcation behavior. We show that there are two generic scenarios reproducing the effect which should be experimentally distinguishable. The contribution is organized as follows. In Sec. II we introduce the convolution model. In Sec. III we apply the convolution model to a derivation of a simple model of ocular dominance pattern formation. In Sec. IV we show the pattern formation behavior of the model with numerical simulations and distinguish two essentially different scenarios. The next three sections consist of analytical investigations of the pattern formation process: in Sec. V the linear stage, in Sec. VI the early nonlinear stage and in Sec. VII the strongly nonlinear stage. In Sec. VIII we show how the convolution model is related to other feature map models and we discuss our approach in Sec. IX.

II. OUTLINE OF THE MODEL

Self-organized neural maps can be seen as mesoscopic models based on three biological principles of activity-dependent change of connectivity in the cortex. The first principle is the so-called Hebb-rule [31]: it is supposed that the connectivity strength between two neurons is enforced if the activity of both neurons are positive correlated. The second principle is the competition between the neurons in the neural sheet for activity. An input stimulus leads to a localized excitation in the neural area next to the neurons that are best optimized to the stimulus [50,51]. As a consequence neurons in this excited domain specialize to the given stimulus after the Hebb rule, because input activity (stimulus) and output activity (excitation in the neural area) are correlated. The third principle is cooperation. Even a very localized stimulus leads to an extended activity patch in the neural area. Neighboring neurons have correlated activity. As a result the adaptation process is a so-called neighborhood-preserving map: neighboring neurons are specialized to similar stimuli.

In general, we have a m-dimensional input space Ω and a n-dimensional neural area $\mathcal{N}$. Each neuron at the position $\mathbf{r} \in \mathcal{N}$ responds to an input activity $\mathbf{v} \in \Omega$ according to its connectivity strength $\mathbf{w_r}$ which is called synaptic weight vector. In the context of the model we identify the weight vector with the so-called receptive field (RF). The receptive field of a neuron describes the set of stimuli which induce activity in the neuron. We are interested in the stimulus driven evolution of receptive fields which define the neural map. A large number of stimuli and a (convenient) slow adaptation process leads to a weight dynamics, which depends on the time average of the distribution of the stimuli. Hence, the weight dynamics can be given by a differential equation, which describes a deterministic dynamical system. Because of the high density and irregular positions of the neurons in the cortex, we consider a continuous formulation that simplifies analytical treatment. (The discrete version can be seen as a neighborhood-preserving vector quantizer.)

A formalization of the first adaptation or *learning* principle should not only include a specialization dynamics on the stimuli but also has to account for the shortage of resources: the weights cannot be strengthened infinitely. On the other hand, an optimal specialization of the weights to a stimulus should not lead to a further adaptation of the weights. These demands are fulfilled by a simple equation

$$\tau \frac{\partial}{\partial t} \mathbf{w_r} = \langle (\mathbf{v} - \mathbf{w_r}) e_{\mathbf{r}}(\mathbf{v}) \rangle, \tag{2.1}$$

where the parameter τ is the time constant of the evolution (set to 1 in the following) and $\langle \cdots \rangle$ denotes the average over the input activity distribution $P(\mathbf{v})$, which we assume to exist. The term $-\langle \mathbf{w_r} e_{\mathbf{r}}(\mathbf{v}) \rangle$ describes a activity-independent decay which guarantees weight normalization [34]. The factor $e_{\mathbf{r}}$ denotes the excitation in the neuronal array and guarantees that only the neurons which are excited by the stimulus can change their weights. The evaluation of $e_{\mathbf{r}}$ as a result of a given input activity $\mathbf{v}$ depends not only on the weights $\mathbf{w_r}$ but also has to account for the lateral connectivity structure in the neural area.

The short range interactions in the cortex are dominated by excitatory connections and the inhibitory interneurons dominate at a longer range. In combination this leads to a sombrero shaped connectivity. As a result an intracortical feedback dynamics leads to an activity in the neural area which is localized around those neurons which are relatively well specialized for the given input [33] On the other hand the excitation $e_{\mathbf{r}}$ is influenced by a kind of lateral cooperation which is induced by local connections between nearby neurons in the input space (smooth input) [52] or spread of activation evoked by the lateral excitatory connectivity in the neural area [41,42]. As we will show later, the strength of the competition of cortical neurons for activity and the spatial scale of the lateral cooperation are the dominating parameters of the pattern formation process [53].

It is convenient to use an explicit ansatz for the steady state of the excitation [33]. We introduce the above considerations into the model by a convolution of a competition term with a cooperative lateral neighborhood function. In other words, we model the neuronal activation e as a result from a global competition for activity and a local spread of activation from neighboring neurons.

First, let us consider the competition term. To measure the degree of optimization of the receptive field $\mathbf{w_r}$ to a given

stimulus $\mathbf{v}$, we propose a distance measure $d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})$, which considers the implicit influence of the neighboring neurons (this distance measure is usually not a metric, e.g., the triangle inequality generally does not hold):

$$d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w_r}) := \int_{\mathcal{N}} h_n(d_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))(\mathbf{v}-\mathbf{w}'_{\mathbf{r}})^2 d\mathbf{r}'. \tag{2.2}$$

Here h_n is the neighborhood function and $d_{\mathcal{N}}(\mathbf{r},\mathbf{r}')$ the distance measure in the neuronal area, which usually is the Euclidian distance. It can be seen as a collective optimization measure, which considers the RF of the whole group of neurons that would respond to the stimulus. To obtain the Euclidian distance we choose $h_n=\delta(\mathbf{r}-\mathbf{r}')$.

To enforce the activity response of the neurons whose distance measures are small we use a monotonic decreasing ''response-function'' of $d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})$: the Gaussian

$$h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})) \propto \exp(-d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})/2\sigma^2_{\mathcal{R}}). \tag{2.3}$$

To introduce competition we normalize the response function by the response of the whole neural area.

$$g_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})) := \frac{h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r}))}{\int_{\mathcal{N}} h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r}))d\mathbf{r}}. \tag{2.4}$$

Note that the parameter $\sigma_{\mathcal{R}}$ controls the competition strength: setting $\sigma_{\mathcal{R}}$ to infinity results in a constant relative response function $g_{\mathcal{R}}$ for all neurons which means no competition. If $\sigma_{\mathcal{R}}=0$, we obtain with probability 1 $g_{\mathcal{R}}=\delta(\mathbf{r}-\mathbf{r}_*)$, with $\mathbf{r}_*$ the neuron whose RF $\mathbf{w_r}$ is next to the stimulus $\mathbf{v}$. Here we have ''hard competition,'' where the activity is localized only around the ''winning'' neuron $\mathbf{r}_*$ which is called the ''winner takes all''(WTA) case. The competition term $g_{\mathcal{R}}$ can be interpreted as relative receptive field effect, since $h_{\mathcal{R}}$ can be seen as an ansatz for a receptive field of a width $\sigma_{\mathcal{R}}$. Another interpretation is Bayesian: $g_{\mathcal{R}}$ denotes the probability that the respective neuron is the best optimized to the given stimulus [47,54].

In the neural area the activation spreads into the neighborhood of the neuron $\mathbf{r}$ due to local neuronal interactions, described by a kernel $h_{\mathcal{N}}(d_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))$, where $d_{\mathcal{N}}(\mathbf{r},\mathbf{r}')$ is the distance measure in the neural area and $h_{\mathcal{N}}$ is the neighborhood function. As an important example consider the Euclidian distance measure

$$d^2_{\mathcal{N}}(\mathbf{r},\mathbf{r}') = (\mathbf{r}-\mathbf{r}')^2 \tag{2.5}$$

and a Gaussian neighborhood function

$$h_{\mathcal{N}}(d_{\mathcal{N}}(\mathbf{r},\mathbf{r}')) \propto \exp(-d^2_{\mathcal{N}}(\mathbf{r},\mathbf{r}')/2\sigma^2_{\mathcal{N}}). \tag{2.6}$$

Combining this, we get an ansatz for the excitation in the neural area

$$e_{\mathbf{r}}(\mathbf{v}) := \int_{\mathcal{N}} g_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r}'))h_{\mathcal{N}}(d_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))d\mathbf{r}'. \tag{2.7}$$

With this ansatz our ''convolution model'' [Eqs. (2.1),(2.7)] is strongly related to the well-known Kohonen model [33] and elastic net [49]. We will discuss this topic later in detail. One special property of the convolution model is the existence of two different explicit length scales: the width of the receptive field $\sigma_{\mathcal{R}}$ in the input space and the width of the neighborhood in the neural area $\sigma_{\mathcal{N}}$. We want to stress here that the basic principles of pattern formation in our model are independent of a special choice of $d_{\mathcal{R}}$, $h_{\mathcal{R}}$, and $h_{\mathcal{N}}$.

To examine the pattern formation behavior of the model it is very helpful to have a potential or Lyapunov function $E(\mathbf{w})$. That is, a function which decreases monotonically in time if $\mathbf{w_r}(t)$ is a solution of Eq. (2.1). A sufficient condition is $\partial_t \mathbf{w_r} = -\partial E/\partial \mathbf{w_r}$. The advantage of potential systems are twofold [55]. First, only fixed-point attractors can occur. Second, if there are many fixed points we can compare them by their value of E. This gives us a criterion which fixed point the system may select dynamically. Our model is in general nonpotential. It only has a potential if the distance measure $d_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})$ includes the same neighborhood function as the cooperation term in the convolution, i.e., $h_n=h_{\mathcal{N}}$ (see Ref. [56]).

With $h_{\mathcal{R}}$ and $h_{\mathcal{N}}$ as Gaussians, we get the potential

$$E(\mathbf{w}) := \left\langle -\sigma^2_{\mathcal{R}} \ln\left[\int_{\mathcal{N}} \exp\left(-\frac{d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w_r})}{2\sigma^2_{\mathcal{R}}} \right) d\mathbf{r} \right] \right\rangle \tag{2.8}$$

with the distance measure after Eq. (2.2).

III. MODELING OCULAR DOMINANCE PATTERN FORMATION

To understand the formal principles and mechanisms which underly the development and activity-dependent stripe width of the ocular dominance pattern in the primary visual cortex, we consider two input surfaces L and R, representing the retinae or lateral geniculate nuclei (LGN) and one neural sheet $\mathcal{N}$ representing the primary visual area in the cortex [27] (see Fig. 2). Every neuron $\mathbf{r}\in\mathcal{N}$ gets input activity $\mathbf{A}=(\mathbf{A}^L(x,y),\mathbf{A}^R(x,y))$, $(x,y)\in L,R$ from the left and right eye, respectively. The neurons react to this input according to their distribution of connectivity weights $\mathbf{W_r}=(\mathbf{W}^L_{\mathbf{r}}(x,y),\mathbf{W}^R_{\mathbf{r}}(x,y))$, which determines their receptive field. We consider ocular dominance formation as a special case of [Eqs. (2.1),(2.7)]:

$$\frac{\partial}{\partial t}\mathbf{W_r} = \langle (\mathbf{A}-\mathbf{W_r})e_{\mathbf{r}}(\mathbf{A})\rangle. \tag{3.1}$$

If the receptive fields are described by (or we are interested in) only a few features, it is useful to consider the problem in terms of a feature map. In the case where the features are *linear* functionals of the stimuli (e.g., gravity centers of the stimuli distribution) the general form of Eq. (2.4) holds for the features as well. (For nonlinear features such as the orientation selectivity the relation to the corresponding full activity distribution is not obvious.) For the ocular dominance map, we represent the distributed activity $\mathbf{A}$ by $\mathbf{v}=(v_x,v_y,v_z)$ with $v_x=\int_{\mathcal{A}}(\mathbf{A}^L+\mathbf{A}^R)xdxdy$, $v_y=\int_{\mathcal{A}}(\mathbf{A}^L+\mathbf{A}^R)ydxdy$, and $v_z=\int_{\mathcal{A}}(\mathbf{A}^L-\mathbf{A}^R)dxdy$ while the synaptic

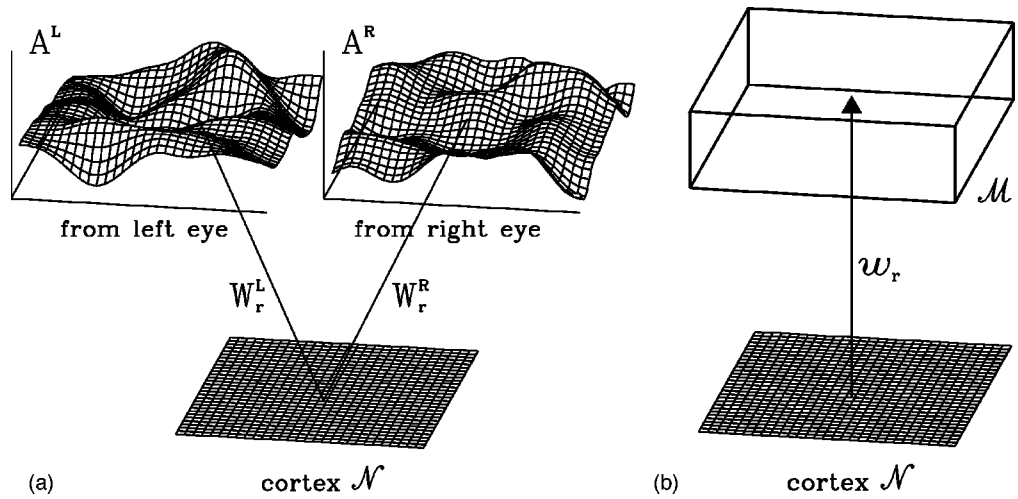

FIG. 2. Left: high-dimensional model. Distributed stimuli A to the two retinas provide input to neuron $\mathbf{r}$ via weight vectors W. Right: inputs and weights are represented by their corresponding centers of gravity and the averaged interocular differences as points in a reduced representation (''feature space'') where the large extensions denote the location of the stimulus while the height represents ocularity.

weights $\mathbf{W}$ are reduced to the receptive field centers $\mathbf{w}=(w_x,w_y,w_z)$ with $w_x=\int_{\mathcal{A}}(\mathbf{W}^L+\mathbf{W}^R)x\,dx\,dy$, $w_y=\int_{\mathcal{A}}(\mathbf{W}^L+\mathbf{W}^R)y\,dx\,dy$, and $w_z=\int_{\mathcal{A}}(\mathbf{W}^L-\mathbf{W}^R)dx\,dy$. In this representation the average weight dynamics Eq. (2.1) reads

$$\frac{\partial}{\partial t}\mathbf{w}_\mathbf{r}=\int_{\mathcal{M}}p(\mathbf{v})(\mathbf{v}-\mathbf{w}_\mathbf{r})\,\tilde{e}_\mathbf{r}(\mathbf{v})\,d\mathbf{v}, \tag{3.2}$$

where p is the projection of the high-dimensional stimulus distribution P into the reduced representation (''feature space'') $\mathcal{M}$ (Fig. 2). The activation $\tilde{e}_\mathbf{r}$ in general might depend on more than the centers of gravity and the interocular differences of the stimuli. Here, however, we use the feature vectors instead of the activity distribution and RF as a first approximation in the framework of the convolution model.

Note that the dimension reduction not necessarily involves an approximation but instead shifts the attention from the high-dimensional problem to the dynamics of centers of gravity and interocular differences, respectively. A structure which appears in this representation therefore also should emerge in the full problem. What then remains is to derive the stimulus distribution in the feature space from the distribution of all inputs P.

Intereye correlations are reflected by the variance of p along the ocularity coordinate of the feature space. To see this we consider correlated inputs constructed as in Ref. [27]: Locally random stimuli are smoothed with a Gaussian kernel of width σ_s providing surfaces $(\tilde{\mathbf{A}}^L,\tilde{\mathbf{A}}^R)$ which then are mutually correlated with strength κ, i.e., $\mathbf{A}^L=\kappa\tilde{\mathbf{A}}^R+(1-\kappa)\tilde{\mathbf{A}}^L$, $\mathbf{A}^R=\kappa\tilde{\mathbf{A}}^L+(1-\kappa)\tilde{\mathbf{A}}^R$ (Fig. 2). A simple calculation involving the central limit theorem shows that in this case p is a normal distribution along v_z with a variance $\langle v_z^2\rangle=(1-2\kappa)^2\sigma_s^2\pi/6$. We also derived the stimulus distribution in the feature space from other assumptions (such as, e.g., rare localized stimuli) with the general result that the variance is always $\langle v_z^2\rangle\propto(1-2\kappa)^2$.

IV. TWO SCENARIOS OF OCULAR DOMINANCE PATTERN FORMATION

Pattern formation in our model occurs as a result of a dimensional conflict. The two-dimensional neural surface is forced to represent an essentially three-dimensional input space: the two dimensions are attributed to space and the third to ocularity. The neural map ignores the third dimension if the variance of the related feature is small compared to the effective receptive field size which depends on $\sigma_\mathcal{N}$ and $\sigma_\mathcal{R}$. Beyond a critical value of $\sigma_\mathcal{N}$ and $\sigma_\mathcal{R}$ the map folds in and patterns such as stripes and patches emerge.

We investigated the formation of ''ocular dominance patterns'' in our model in many simulations. We started from a binocular [i.e. $w_z(\mathbf{r})=0$] and roughly retinotopic map which is a fixed point of Eq. (3.2) (see next section). It turns out that it is important to distinguish between two different scenarios of pattern formation (compare with Ref. [57]).

Figure 3 shows the result of a typical simulation. In the

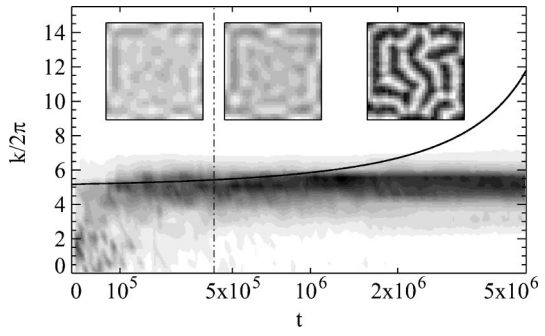

FIG. 3. Mode selection in a simulation of 32×32 neurons (open boundary conditions, $\sigma_\mathcal{R}=0$ method from Ref. [33] with learning rate 0.1). The inputs are samples from a normal distribution with variance $\langle v_z^2\rangle=(1-2\kappa)^2\sigma_s^2\pi/6$ at $\kappa=0.3$. $\sigma_\mathcal{N}$ was decreased linearly from 1.35 to 0.5 (neural units) during $t=5\times10^6$ iterations. The grey levels show the evolution of the normalized power spectra of the OD patterns (inset). When $\sigma_\mathcal{N}$ crosses its critical value (dash-dotted line) the pattern forms and retains its wavelength although the maximum of the spectrum of the linearized dynamics moves further to higher values (thick line). Note the logarithmic time scale.

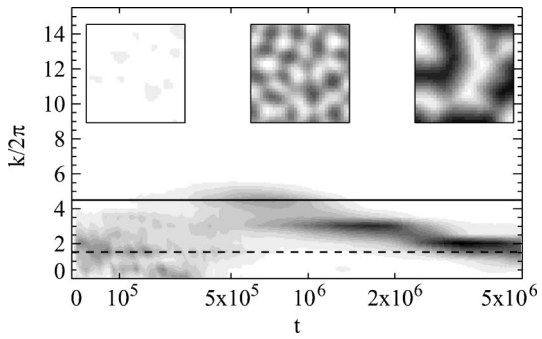

FIG. 4. Evolution of ocular dominance of the same model as above. $\sigma_{\mathcal{N}}$ was constant at 1.5 (neural units) during $t=5\times10^6$ iterations and $\kappa=0.0$. The eigenvalue of $\lambda_{max}=3.6$ boosts the noise to a great extent. First the modes with the greatest eigenvalue (line) were amplified. Then the pattern reorganizes and the frequency shifts towards the edge of the positive spectrum (broken line).

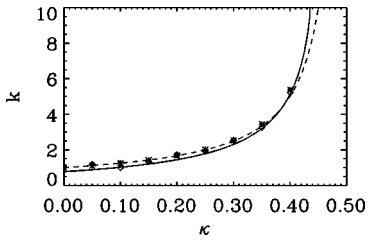

FIG. 5. The wave number k of the pattern normalized by the wave number of the pattern at $\kappa=0$ as a function of the correlation parameter κ. The diamonds denote the wave number at the first scenario where the width parameters shrinks during development. They are determined by the wave number of the first positive eigenvalue (broken line). The stars describe the wave number of the stationary solution of the pattern after the reorganization in the second scenario. They are just above the left edge of the positive spectrum (line).

simulations we use a stochastic, Markovian adaptation method as in Ref. [33] to get a robust result. As described in Ref. [37], the fluctuations are strongest near the largest eigenvalue of the linearized dynamics (see next section). When the parameters $\sigma_{\mathcal{R}}$, $\sigma_{\mathcal{N}}$, and σ_n are below the critical value, the noise pattern becomes amplified: it grows exponential. The critical values depend on the correlation parameter κ and the wavelength of the amplified pattern depends on the critical values of the width parameters. Further shrinking during development leads to *no change* of the wavelength, the pattern becomes sharper and more stripe-shaped only. The stripe width resulting from this scenario are larger for smaller correlations (see Fig. 5). We shall discuss this in the next section.

In the other scenario the width parameters are fixed at subcritical values. Thus the pattern grows right from the beginning. In the first part of the evolution the wavelength of the pattern is at the largest eigenvalue of the linearized dynamics as in the first scenario. But then reorganization occurs due to the nonlinearity of the dynamics. The stripe width of the pattern shifts on a logarithmic time scale to longer wavelengths (see Fig. 4). The stationary length scale of the pattern depends on both the width parameters and the correlation parameter. Here, the stripe width are larger for smaller correlation too (see Fig. 5). We shall discuss this in Sec. VII.

Both scenarios qualitatively predict the same dependency between the correlation and the length scale of the stationary map. However, the evolution of the pattern formation process differs in time scale and amplitude. In the next sections, we investigate the formal principles of the pattern formation mechanisms in both scenarios. We will focus specifically on the determination of the length scale of the emerging pattern.

V. PATTERN FORMATION IN FEATURE MAP MODELS

Generally, pattern formation occurs if the homogeneous binocular map is unstable against small fluctuations. In the following sections we will analyze ocular-dominance pattern formation in the most simple case of a two-dimensional input space and a one-dimensional chain of neurons. In a first approximation this can be seen as a vertical cut through the feature space along the ocular dominance direction, parallel to one side, say, the y direction. The validity of the results in the case of higher-dimensional feature maps or the fully connected problem with two input sheets projecting onto one neural area will be discussed in the last section.

In the two-dimensional feature space with one space coordinate x and the ocular dominance coordinate y, we perform a linear stability analysis. This allows us to determine which modes are stable and which become amplified.

The evolution of the weights $\mathbf{w}=(w_x,w_y)$ in our reduced problem is given as

$$\frac{\partial}{\partial t}\mathbf{w}_r=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P(\mathbf{v})\int_{-\infty}^{\infty}\frac{h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{r'}))}{\int_{-\infty}^{\infty}h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{r'}))dr'}h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))dr'(\mathbf{v}-\mathbf{w}_r)dv_x dv_y\,.$$

We choose for the calculations $h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))$ after Eq. (2.6) and $h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_r))$ after Eq. (2.3) with

$$h_n(d_{\mathcal{N}}(r,r')):=\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{(r-r')^2}{2\sigma_n^2}\right).$$

The case $\sigma_{\mathcal{N}}=\sigma_n$ represents the potential version and the case $\sigma_n=0$ stands for the continuous formulation of the standard nonpotential Kohonen model [33].

The homogeneous solution of the map is $\overline{\mathbf{w}}_r=(r,0)$. This is a stationary solution of the map for all parameters $\sigma_{\mathcal{R}}$, σ_n, and $\sigma_{\mathcal{N}}$ if $P(\mathbf{v})$ is constant in the x direction and the mean vanishes in the y direction because the integration

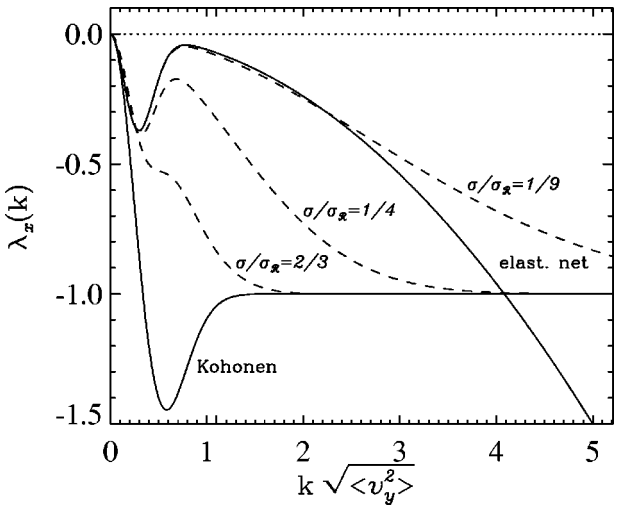

FIG. 6. The spectra in the stable x direction for several parameters ($\sigma_{\mathcal{N}}=\sigma_n$, $\sigma=2\sigma_{\mathcal{N}}$).

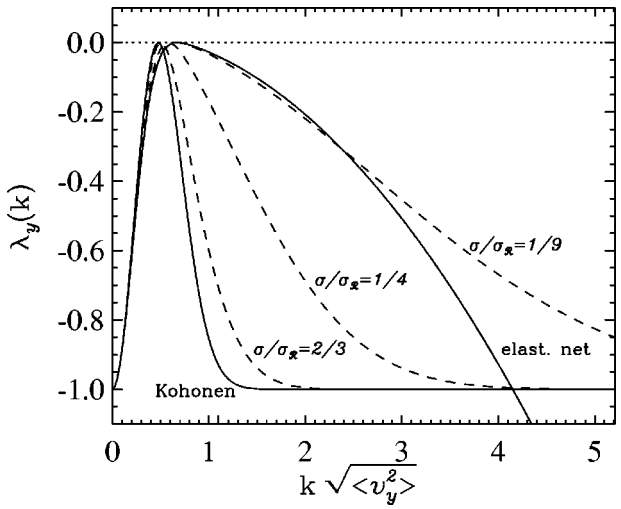

FIG. 7. Some spectra in the unstable y direction for different parameters sets which fulfill the marginal stability condition ($\lambda_{\max}=0$).

runs over an odd function (for the x direction around r and for the y direction around 0).

We examine the stability of this stationary solution by linearizing the evolution of a small perturbation $\boldsymbol{\delta}_r$. With the ansatz $\mathbf{w}_r=\overline{\mathbf{w}}_r+\boldsymbol{\delta}_r$ we get for the evolution in the neighborhood of the stationary solution

$$\dot{\boldsymbol{\delta}}_r \approx \mathcal{L}\boldsymbol{\delta}_r$$

with the linear operator $\mathcal{L}$. In Fourier space this translation invariant operator is diagonal. Therefore the computation of the spectrum is straight forward (see Appendix A). The linearized dynamics is given through a simple eigenvalue equation

$$\frac{\partial}{\partial t}\hat{\boldsymbol{\delta}}_k=\hat{\mathcal{L}}\hat{\boldsymbol{\delta}}_k=\lambda(k)\hat{\boldsymbol{\delta}}_k$$

with the spectra of eigenvalues

$$\lambda_x(k)=-1+\left(1+k^2\frac{\sigma_{\mathcal{N}}^2\sigma_n^2}{\sigma_{\mathcal{R}}^2}-k^2\frac{(\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2)(\sigma_{\mathcal{R}}^2+\sigma_n^2)}{\sigma_{\mathcal{R}}^2}\exp(-k^2\sigma_{\mathcal{R}}^2)\right)\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right), \tag{5.1}$$

$$\lambda_y(k)=-1+\langle v_y^2\rangle\frac{[1-\exp(-k^2\sigma_{\mathcal{R}}^2)]}{\sigma_{\mathcal{R}}^2}\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right). \tag{5.2}$$

If an eigenvalue $\lambda(k)$ is positive, then perturbations with wave number k grow exponentially, while perturbations with wave numbers k' with negative eigenvalues $\lambda(k')$ are damped away. Its easy to see that in the x direction (Fig. 6) all modes except the marginally stable translation are damped. This direction is stable.

In the y direction the behavior depends on the width of the input distribution $\langle v_y^2\rangle$: above a critical value the spectrum includes positive eigenvalues around the maximum (Fig. 7). The spectra have a maximum at

$$k_{\max}=\frac{1}{\sigma_{\mathcal{R}}}\sqrt{\ln\left(1+\frac{2\sigma_{\mathcal{R}}^2}{\sigma_{\mathcal{N}}^2+\sigma_n^2}\right)}. \tag{5.3}$$

A mode with this wave number is marginal stable at

$$\langle v_y^2\rangle=\left(\frac{\sigma_{\mathcal{N}}^2+\sigma_n^2}{2}+\sigma_{\mathcal{R}}^2\right)\exp\left[\frac{\sigma_{\mathcal{N}}^2+\sigma_n^2}{2\sigma_{\mathcal{R}}^2}\ln\left(1+\frac{2\sigma_{\mathcal{R}}^2}{\sigma_{\mathcal{N}}^2+\sigma_n^2}\right)\right]. \tag{5.4}$$

Combining these two formulas, we get the position of the first positive eigenvalue

$$k_* = \frac{1}{\sqrt{\langle v_y^2 \rangle}} \sqrt{\ln\left(1+\frac{2\sigma_{\mathcal{R}}^2}{\sigma_{\mathcal{N}}^2+\sigma_n^2}\right)\left[1+\frac{\sigma_{\mathcal{N}}^2+\sigma_n^2}{2\sigma_{\mathcal{R}}^2}\left(1+\frac{2\sigma_{\mathcal{R}}^2}{\sigma_{\mathcal{N}}^2+\sigma_n^2}\right)^{(\sigma_{\mathcal{N}}^2+\sigma_n^2)/2\sigma_{\mathcal{R}}^2}\right]}. \tag{5.5}$$

Applying this result to the model of ocular dominance pattern formation of Sec. III, we have to take into consideration that the parameters $\sigma_{\mathcal{N}}$, $\sigma_{\mathcal{R}}$, and σ_s are not known. Hence, we are only able to compare the resulting wave numbers of the pattern for the normal and the strabismic case. Assuming that the relation of the width parameters remain unchanged by introducing strabism, we normalize the wave number k_* by the wave number k_*^0 of the strabismic case ($\kappa=0$). We get for the κ dependency of the wave number of the resulting ocular dominance pattern

$$\frac{k_*}{k_*^0} = \frac{1}{1-2\kappa}. \tag{5.6}$$

A pattern with this wave number is selected in the scenario of shrinking receptive fields (compare Fig. 5).

VI. PHASE TRANSITIONS AT THE PATTERN FORMATION

Shrinking the width parameters below the critical values, lead to patterns with finite amplitude (''ocularity'') via a phase transition of second order. The stationary amplitude of the model near marginal stability can be derived by expanding the evolution equation [Eqs. (2.1),(2.7)]. For small amplitudes, only the mode of the critical wave number k_* in the unstable direction and the mode with the strongest coupling from the stable direction have to be considered. Using the symmetry of the equation we have the ansatz for the approximation of the solution

$$\mathbf{w}_r := [r + a_x \sin(2k_* r), a_y \cos(k_* r)].$$

The evolution equation is invariant against translations. This leads to the relative simple form of expansion around $(a_x, a_y) = (0,0)$

$$\dot{a}_x \approx \lambda_x a_x + c_x a_y^2,$$

$$\dot{a}_y \approx \lambda_y a_y + c_y a_x a_y + d_y a_y^3.$$

Using the spectrum λ_y as control parameter we have for the stationary amplitude of the mode

$$a_y = \sqrt{\frac{\lambda_x}{c_x c_y - \lambda_x d_y}} \sqrt{\lambda_y}.$$

The model shows a continuous phase transition at all parameters (see Appendix B and Figs. 8,9).

Near the WTA case ($\sigma_{\mathcal{R}}=0$) a second branch appears. Interestingly this behavior takes place even at positive $\sigma_{\mathcal{R}}$. This stationary amplitude could refer to higher-order terms of the expansion. The simulations show that the local minima at the continuous branch is small. Little noise pushes the amplitude to the branch with hysteresis (compare with the results in Ref. [58]).

VII. REORGANIZATION PHENOMENA OF PATTERNS

Undercritical width parameters usually lead to a whole range of unstable modes. For each of these modes a positive amplitude is more optimal than the homogeneous map. This corresponds in the potential case to a Lyapunov function which has a minimum at a positive amplitude at each of this modes. The wave number of deepest minimum at all this modes can be seen as the wave number of the stationary solution of the model.

The numerical simulations show that the map reorganizes after a first saturation of the amplitude at the wave number which corresponds to the largest eigenvalue of the linearized dynamics. This process is bounded by the positive part of the spectrum of eigenvalues (see Sec. IV). In the following, we derive an upper bound for the wave number of the stationary solution. We assume that the higher harmonics of the stationary map make only a small contribution to the energy and that the ''refinements'' to a simple harmonic map have more influence at lower wave numbers. This holds in all cases we know (see, e.g., Figs. 10,11).

In the case of undercritical width parameters, which are not too far from the critical values, it is possible to expand the energy respective the amplitude and the wave number and then evaluate the zeros of the equation system. This classical technique [55] is not very useful here because the resulting terms are very long and give only a little insight.

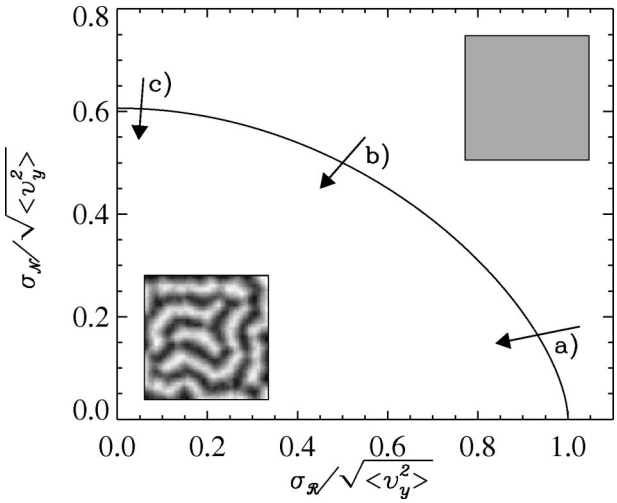

FIG. 8. The set of bifurcation points of the convolution model ($\sigma_{\mathcal{N}}=\sigma_n$). The line denotes the parameters, where the homogeneous solution is marginally stable [$\lambda_y(k_*)=0$]. Inside the line of marginal stability, the homogeneous solution is unstable and pattern formations occurs. The arrows denote different paths for the development of patterns (see Fig. 9).

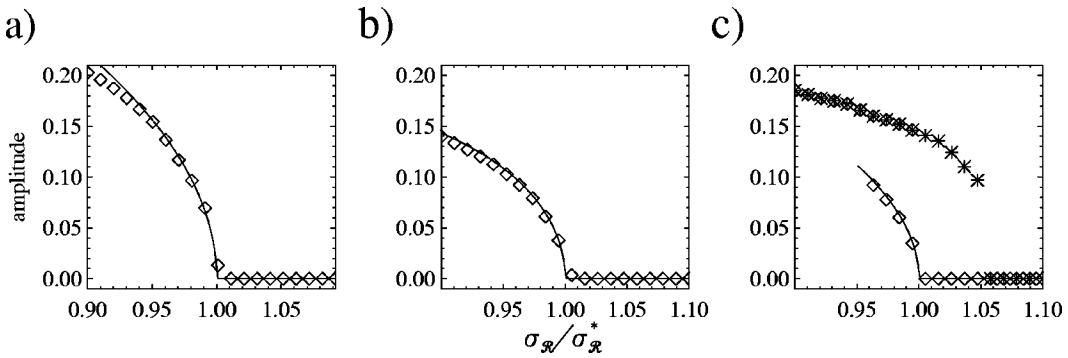

FIG. 9. Bifurcation diagrams of the convolution model at different parameters $\sigma_{\mathcal{R}}$ (compare with arrows at Fig. 8). All paths show a continuous phase transition behavior. The solid lines show the amplitude $a_y/\sqrt{\langle v_y^2\rangle}$ derived from the expansion of the evolution equations in Appendix B. Near the hard competition limit ($\sigma_{\mathcal{R}}=0$) a new branch with hysteresis appears. This behavior could result from higher order terms of the expansion.

Far away from the critical point the expansion breaks down. In this case it is possible to evaluate an estimation in the following way.

In the first step we evaluate the modified distance measure by using the harmonic ansatz $\mathbf{w}_r=(r,A\sin(kr))$:

$$d^2(\mathbf{v},\mathbf{w}_r)=\frac{1}{\sqrt{2\pi}\sigma_{\mathcal{N}}}\int_{-\infty}^{\infty}\exp\left(-\frac{(r-r')^2}{2\sigma_{\mathcal{N}}^2}\right)(\mathbf{v}-\mathbf{w}_{r'})^2dr'$$

$$=r^2-2rv_x+v_x^2+v_y^2-2v_yA\exp\left(-\frac{k^2\sigma_{\mathcal{N}}^2}{2}\right)\sin(kr)+\sigma_{\mathcal{N}}^2+\frac{A^2}{2}-\frac{A^2}{2}\exp(-2k^2\sigma_{\mathcal{N}}^2)\cos(2kr).$$

For a better approximation we translate the distance function to achieve nearly a saddle point approximation and replace the trigonometric function with the approximation up to second order

$$\sin(kr)\approx kr,\quad \cos(kr)\approx 1-\frac{k^2r^2}{2}.$$

We get as approximation for small $k,\sigma_{\mathcal{N}},\sigma_{\mathcal{R}}$, and large A (far away from linear instability) for the distance measure

$$d^2(\mathbf{v},\mathbf{w}_{r+v_x})\approx Pr^2+Qr+R$$

with

$$P:=1+A^2k^2\exp(-2k^2\sigma_{\mathcal{N}}^2),$$

$$Q:=2v_xA^2k^2\exp(-2k^2\sigma_{\mathcal{N}}^2)-2v_yAk\exp\left(-\frac{k^2\sigma_{\mathcal{N}}^2}{2}\right),$$

$$R:=\sigma_{\mathcal{N}}^2+\frac{A^2}{2}[1-\exp(-2k^2\sigma_{\mathcal{N}}^2)]+v_y^2+v_x^2A^2k^2$$

$$\times\exp(-2k^2\sigma_{\mathcal{N}}^2)-2v_yv_xAk\exp\left(-\frac{k^2\sigma_{\mathcal{N}}^2}{2}\right).$$

With this approximation we have for the energy

$$E\approx\left\langle-\sigma_{\mathcal{R}}^2\ln\left[\int_{-\infty}^{\infty}\exp\left(-\frac{Pr^2+Qr+R}{2\sigma_{\mathcal{R}}^2}\right)dr\right]\right\rangle$$

$$=\frac{1}{2}\left[\left\langle R-\frac{Q^2}{4P}\right\rangle+\sigma_{\mathcal{R}}^2\ln\left(\frac{P}{2\pi\sigma_{\mathcal{R}}^2}\right)\right]$$

$$=\frac{1}{2}\left[\sigma_{\mathcal{N}}^2+\frac{A^2}{2}[1-g^2(\sigma_{\mathcal{N}},k)]+\sigma_{\mathcal{R}}^2\ln\left(\frac{1+A^2k^2g^2(\sigma_{\mathcal{N}},k)}{2\pi\sigma_{\mathcal{R}}^2}\right)\right.$$

$$\left.+\frac{\langle v_y^2\rangle\{1+A^2k^2[g^2(\sigma_{\mathcal{N}},k)-g(\sigma_{\mathcal{N}},k)]\}+A^2g^2(\sigma_{\mathcal{N}},k)/3}{1+A^2k^2g^2(\sigma_{\mathcal{N}},k)}\right]$$

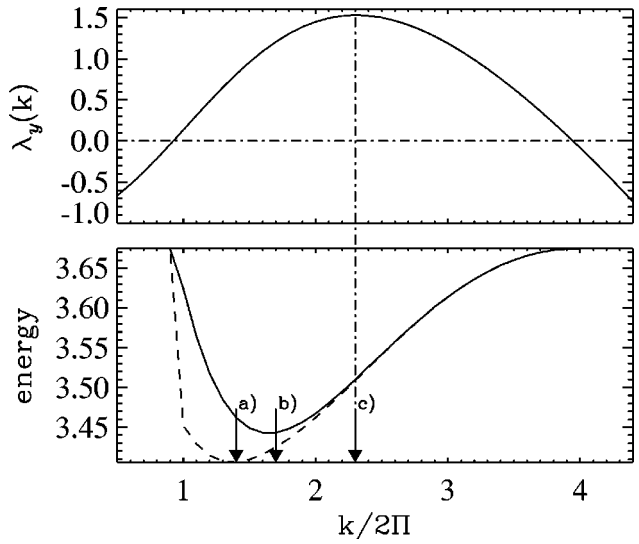

FIG. 10. Spectrum of eigenvalues (top) and energy of stationary solutions (broken line) compared to our harmonic ansatz (solid line) in the case of a Gaussian distribution $P(v_y)$. The optimal wave number of the stationary solution (a) is bounded by the minimum of our ansatz (b) and is below the position k_* of maximal eigenvalue $\lambda_y(k_*)$ (c).

with $g(\sigma_{\mathcal{N}},k) := \exp(-\sigma_{\mathcal{N}}^2 k^2)$. To determine, at which wave number k of the harmonic ansatz the energy has a minimum, we have to look for the zeros of the partial derivatives

$$\frac{\partial}{\partial A}E(A,k)=0, \quad \frac{\partial}{\partial k}E(A,k)=0.$$

Inserting the zero of the first equation into the second one, leads to a condition for the minima of the approximated energy

$$0=[\sigma_{\mathcal{R}}^2\exp(-2\sigma_{\mathcal{N}}^2k^2)]k+[1-\exp(-2\sigma_{\mathcal{N}}^2k^2)] - \left(\frac{2}{3}[\exp(-4\sigma_{\mathcal{N}}^2k^2)-\exp(-2\sigma_{\mathcal{N}}^2k^2)] + \sigma_{\mathcal{R}}^4k^2\exp(-4\sigma_{\mathcal{N}}^2k^2)+\frac{6}{3}\langle v_y^2\rangle \times[\exp(-\sigma_{\mathcal{N}}^2k^2)-\exp(-3\sigma_{\mathcal{N}}^2k^2)]\right)^{1/2}.$$

For small parameter $\sigma_{\mathcal{N}},\sigma_{\mathcal{R}}$ this equation holds near

$$k=\frac{1}{\sqrt{3(\langle v_y^2\rangle-\sigma_{\mathcal{R}}^2-\sigma_{\mathcal{N}}^2)}}$$

which shows a nearly $k\propto 1/\sqrt{\langle v_y^2\rangle}$ dependency also far away from linear instability. Hence, the dependency of the wave number normalized with the wave number of the strabismic case is similar to the dependency which holds for the wave number of the first positive eigenvalue in the shrinking receptive field case.

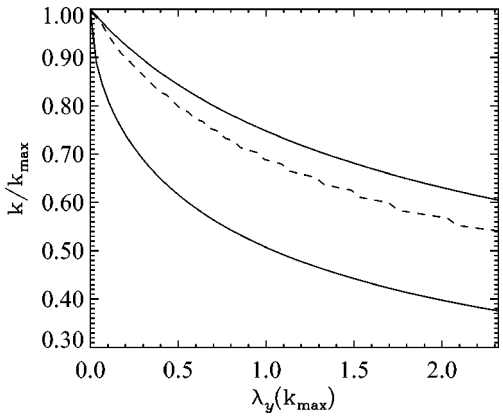

FIG. 11. The wave number of the stationary solution (broken line) and of the upper bound from the position of the energy minimum of the harmonic ansatz (upper solid line). The lower solid line marks the position of the left bound of the positive spectrum. Parameters are $\sigma_{\mathcal{R}}=\sigma_n=0$.

VIII. RELATION TO OTHER FEATURE MAP MODELS

Our convolution model is strongly related to the well known Kohonen algorithm [33,37] and to the elastic net algorithm [49]. The first one can be seen as a discrete, algorithmic approach of our model using $h_n=\delta$ and a vanishing input field width (''winner takes all'' limit). Correspondingly, the spectra of the linearized dynamics are the same as in Ref. [37], if we choose the Euclidian metric ($\sigma_n=0$) and then $\sigma_{\mathcal{R}}\to 0$:

$$\lambda_x(k)=-1+(1-k^2\sigma_{\mathcal{N}}^2)\exp\left(-\frac{k^2\sigma_{\mathcal{N}}^2}{2}\right),$$

$$\lambda_y(k)=-1+\langle v_y^2\rangle k^2\exp\left(-\frac{k^2\sigma_{\mathcal{N}}^2}{2}\right)$$

because we have

$$\lim_{\sigma_{\mathcal{R}}\to 0}\frac{1-\exp(-k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}=k^2.$$

Here, we can see, why Kohonen's algorithm is so successful (additional to the simple and algorithmically cheap form): because no eigenvalue is smaller than -1; numerically the algorithm is very robust.

If additionally to $\sigma_R=0$ we use $h_n=h_{\mathcal{N}}$, we get the potential version of the Kohonen's SOFM. In this case the model has an energy function which we obtain from the general energy function Eq. (2.8) with $\sigma_{\mathcal{R}}\to 0$

$$E := \left\langle \frac{1}{2}\int_{\mathcal{N}} h_{\mathcal{N}}(\mathbf{r}_*-\mathbf{r})(\mathbf{v}-\mathbf{w}_{\mathbf{r}})^2 d\mathbf{r}\right\rangle,$$

where the neuron $\mathbf{r}_*$ is the winner neuron in terms of the

distance measure

$$d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}_*}) \leqslant d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}}) \quad \forall \mathbf{r} \in \mathcal{N}.$$

The elastic net algorithm consists of two parts: a Hebbian soft competition term and smoothing elastic force term, which guarantees the neighborhood preservation. Unfortunately, the discretization induces an implicit length scale without obvious interpretation of the second term. The elastic net behaves similar to the convolution model with a small but finite neighborhood width $\sigma_{\mathcal{N}}$:

$$\frac{\partial}{\partial t}\mathbf{w}_{\mathbf{r}} = \left\langle \int_{\mathcal{N}} g_{\mathcal{R}}(d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}'}))h_{\mathcal{N}}(d^2_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))((\mathbf{v}-\mathbf{w}_{\mathbf{r}'})+(\mathbf{w}_{\mathbf{r}'}-\mathbf{w}_{\mathbf{r}}))d\mathbf{r}' \right\rangle$$

$$= \left\langle \int_{\mathcal{N}} g_{\mathcal{R}}(d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}'}))h_{\mathcal{N}}(d_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))(\mathbf{v}-\mathbf{w}_{\mathbf{r}'})d\mathbf{r}' \right\rangle + \left\langle \int_{\mathcal{N}} g_{\mathcal{R}}(d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}'}))h_{\mathcal{N}}(d^2_{\mathcal{N}}(\mathbf{r},\mathbf{r}'))(\mathbf{w}_{\mathbf{r}'}-\mathbf{w}_{\mathbf{r}})d\mathbf{r}' \right\rangle.$$

The first term reduces to the Hebb term of the elastic net. The second term behaves similar to the elastic force term, if the neighborhood $\sigma_{\mathcal{N}}$ is small and the width of the receptive fields $\sigma_{\mathcal{R}}$ is wide in relation to the width of the distribution of the stimuli

$$\frac{\partial}{\partial t}\mathbf{w}_{\mathbf{r}} \approx \langle g_{\mathcal{R}}(d^2_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{\mathbf{r}}))(\mathbf{v}-\mathbf{w}_{\mathbf{r}})\rangle + \beta\Delta\mathbf{w}_{\mathbf{r}}$$

with a small factor β. The right side is the continuous formulation of the elastic net, since the elastic force term can be seen as a discretization of the Laplace-operator: $\Sigma_{\mathbf{r}'}(\mathbf{w}_{\mathbf{r}'} - \mathbf{w}_{\mathbf{r}}) \approx \Delta\mathbf{w}_{\mathbf{r}}$ with $\mathbf{r}' \in$ neighborhood of $\mathbf{r}$.

A comparison of the dynamics linearized around the homogeneous solution makes the relation clearer. We obtain the spectrum of eigenvalues from the spectrum of the convolution model [Eqs. (5.1),(5.2)] by inserting the parameters $\sigma_n = \sigma_{\mathcal{N}} = 0$ and by adding the Fourier transformed linear Laplace-operator

$$\beta\Delta\boldsymbol{\delta}_{\mathbf{r}} \rightarrow -\beta k^2 \hat{\delta}_{\mathbf{r}}.$$

This yields

$$\lambda_x(k) = -k^2\sigma_{\mathcal{R}}\exp(-k^2\sigma^2_{\mathcal{R}}) - \beta k^2, \tag{8.1}$$

$$\lambda_y(k) = -1 + \langle y^2\rangle\frac{[1-\exp(-k^2\sigma^2_{\mathcal{R}})]}{\sigma^2_{\mathcal{R}}} - \beta k^2. \tag{8.2}$$

The spectrum in the y direction is maximal at

$$k_* = \frac{1}{\sigma_{\mathcal{R}}}\sqrt{\ln\left(\frac{\langle v_y^2\rangle}{\beta}\right)}$$

which depends on the variance $\langle v_y^2\rangle$ and becomes marginally stable at

$$\langle v_y^2\rangle = \sigma^2_{\mathcal{R}} + \beta\left[1 + \ln\left(\frac{\langle v_y^2\rangle}{\beta}\right)\right].$$

Comparing this to the spectrum of the convolution model in the unstable y direction we get an approximation of the spectrum of the elastic net around the maximum with a small but finite width $\sigma_{\mathcal{N}}$ of the neighborhood function. From the expansion around $\sigma_{\mathcal{N}} = 0$ we can estimate the elastic force factor β with an small but finite neighborhood in the neural area:

$$\beta = \frac{\sigma^2_{\mathcal{N}}}{2\sigma^2_{\mathcal{R}}}\langle v_y^2\rangle.$$

For higher frequencies, however, the spectrum of the elastic net differs from that of the convolution model. The Laplace operator damps the modes in this range much more than the neighborhood function. This damping behavior leads to numerical instabilities, which have to be considered in simulations. The energy function of the potential elastic net model can be obtained from the energy function of the potential convolution model by using the parameters above and adding an elastic energy term

$$\frac{\beta}{2}\int_{\mathcal{N}}(\nabla\mathbf{w}_{\mathbf{r}})^2 d\mathbf{r}.$$

The expansion of this term around the homogeneous map as in the previous section results in:

$$\frac{\beta}{2}\int_{-\infty}^{\infty}(\nabla\mathbf{w}_r)^2 dr = \frac{\beta}{2}\left(1 + \frac{A^2k^2}{2}\right).$$

An analysis of this energy function along the lines of the previous sections shows, that the elastic net has a phase transition of second order for all $\sigma_{\mathcal{R}}$ and we can expect a similar reorganization behavior as for the convolution model [59].

IX. DISCUSSION

In recent years a wide range of models with different mathematical properties and different degree of abstraction has been used to describe the development of ocular dominance patterns. On one side there are noncompetitive, mainly linear models [60–62] with local learning rules which are based only on the correlation between input cells and post synaptic cells. In these models the changes are governed in a linear way by the input patterns. On the other side there are the competitive, essentially nonlinear models [49,27,37] in which the learning rules are nonlinear and slightly nonlocal. In some comparisons between linear and nonlinear ap-

proaches it had been stressed that the competitive models seem to fit better to the physiological findings [57,63,14]. Hence, we restrict our investigation of principles of pattern formation to the latter class of models.

In this contribution we analyzed activity-dependent formation of neural connectivity patterns in the framework of the convolution model. The convolution model is derived from basic assumptions about activity-dependent neural map formation making it neurobiologically plausible. At the same time it generalizes a range of previous approaches. In particular, the widely used Kohonen model [33,37,64] is contained as a limiting case and the elastic net algorithm [49,45,65] can be approximated for the case of a weak elastic force. We applied our approach to the problem of ocular dominance pattern formation in visual cortex for which we performed simulations and a mathematical analysis of the basic mechanisms of pattern formation including the bifurcation behavior and reorganization phenomena. In particular we were interested in the origin of characteristic length scales of the resulting patterns and on how these scales depend on interocular correlations of the inputs to the two eyes during development.

We found that the experimental observation of increasing scales for lower correlation [26] is reproduced in the convolution model [37,66,27] following essentially one of two dynamic scenarios of pattern formation.

(1) In the first scenario the homogeneous solution is stable in the beginning, i.e., all neurons are binocular. In order to obtain patterns the intrinsic width parameters (effective receptive field size) must shrink during the formation process such that the map is driven through the instability. In this case the mode with the wave number of the eigenvalue first becoming positive grows exponential. The pattern saturates at an amplitude which can be determined by the higher order coefficients of the expansion of the energy function as shown in Sec. VI. Further shrinking the receptive fields leads to a deeper energy minimum and the pattern may become fixed.

(2) In the second scenario the parameters lead to pattern formation right from the beginning. In this case a pattern corresponding to the largest positive eigenvalues grows first until it reaches saturation. The pattern does not remain stable, it reorganizes on a long time scale which goes along with a lowering of the wave number of the pattern. The shift to lower wave numbers stops before it reaches the left boundary of the positive spectrum of eigenvalues.

Both scenarios correctly describe the effect of wider stripe width of ocular dominance patterns in strabismic cats (Fig. 5). The absolute wave length of the pattern is, however, not determined, because it depends on experimentally unknown physiologically values of length scales $\sigma_{\mathcal{R}}$, $\sigma_{\mathcal{N}}$, and σ_n. Based on the assumption, that at induced strabism the width parameters remain unchanged and correlation is zero, we obtain a relative wavelength as a function of the correlation parameter for the normal case. Figure 5 shows that the scenarios of OD pattern formation cannot be distinguished on the basis of the stationary solutions alone. If, however, the *dynamics* of the pattern formation is observed, both scenarios make very different predictions (compare Figs. 3 and 4). In case of shrinking receptive field sizes and/or shrinking neighborhood widths, at some point the pattern grows very fast with the first unstable wavelength which from then on does not change any further. In the other case of constant subcritical parameters, the pattern grows with a high spatial frequency right from the beginning and then reorganizes on a long time scale. This yields a map with a wider stripe width.

Which scenario is actually realized in the brain can be resolved by the experiment sketched below that is feasible with present day technology. Starting at about two to three weeks of age, ocular dominance columns can be visualized in the visual cortex of kittens by optical imaging of intrinsic signals [67,68]. It is known from monocular deprivation studies that the OD pattern is in a plastic state until the end of the third month of life, with the degree of plasticity beefing maximal at an age of six weeks. Practically it appears possible to visualize the pattern every two or three days. It should therefore be possible to obtain a sequence of patterns following its development from the third week to adulthood. Comparing such sequences in squinting animals and animals with normal visual experience should suffice to identify by which scenario a larger wavelength of ocular dominance columns is realized in the brain of squinting animals.

In our formal analysis we considered the evolution of feature maps in terms of a deterministic dynamical system. We neglected the influence of noise in the formulation of the model in order to guide the attention to the basic structure of the dynamics. However, knowing the properties of the deterministic system means knowing a lot about the related stochastic system. The linear approximation of the evolution equation corresponds to the drift term of the Fokker-Planck equation which determines the stability behavior. On the other hand, the energy function of the deterministic system determines the stationary invariant distribution of the maps $P(\mathbf{w}):P(\mathbf{w})\propto\exp[-2E(\mathbf{w})/\sigma^2]$, where σ is the strength of the additive noise.

The type of transition from the homogeneously binocular state to the ocular dominance pattern exhibits distinct signatures of the nonlinear contributions to the developmental dynamics. Our results demonstrate that the type of bifurcation (supercritical vs subcritical) depends on the qualitative nature of intracortical interactions. Assuming that the development of ocular dominance columns is not coupled to the retinotopic organization, the transition is supercritical (discontinuous) if the range $\sigma_{\mathcal{N}}$ over which activity spreads within the cortical layer is much larger than the receptive field size $\sigma_{\mathcal{R}}$ (see coefficient d_y in Appendix B). The bifurcation is continuous if the size of receptive fields is a considerable fraction of the intracortical range $\sigma_{\mathcal{N}}$. If the development of the columnar pattern is coupled to retinotopic distortions the transition becomes continuous for the complete range of parameters for which patterns can form. The latter result agrees with the findings of Herman and Der [58] who studied the phase transition, which occurs in the discrete version of the Kohonen model. Such detailed properties of the patterns forming transition might become experimentally accessible in the future. Our results suggest that in this case overall properties of the pattern forming transition could be used for testing the qualitative nature of neuronal interactions in the cortical area.

The formal analysis were made in the most simple case of a one-dimensional neural area and a two-dimensional input space. The extension to the higher-dimensional case is similar to the ''classical'' pattern formation systems in physics

such as the Swift-Hohenberg equations [55]. At the linear stage of the evolution of the map, the different dimensions are independent from each other. Hence, the phase transition occurs for each dimension individually. At the nonlinear stage of pattern formation the dimensions are no longer independent. Similar to systems described by amplitude equations of Ginzurg-Landau-type the map reorganizes to smoother patterns such as parallel stripes [69]. This reorganization process has only a little influence on the wave number of the resulting pattern.

Generally, it is very difficult to find a stationary solution for high-dimensional fully connected maps. This prohibits a similar analysis of the stability behavior as in the low-dimensional feature map models. Another way of investigating the pattern formation in a high-dimensional map was proposed by Bauer *et al.* [70]. They analyzed a heuristic cost function depending on an ansatz for the stationary map. They demonstrated the validity of their computations by comparing to numerical simulations. However, the starting point of their derivation of the distortion measure which was adopted from Ref. [71] is not valid even in the case of an ordered map in the low-dimensional feature map case (see Appendix C). A more general investigation will have to use a cost function, for which the validity was formally proofed.

After all, perhaps the most striking feature of the self-organized neural feature map model is the analogy with the classical pattern generating systems of physics which are described by equations of the Ginzburg-Landau type. We suspect, that more detailed investigations of instabilities in the nonlinear regime would reveal behaviors such as zig-zag instabilities, competition between different modes and especially will stress the relevance of the boundaries of the neural area for the hole layout of the pattern [59].

APPENDIX A: COMPUTATION OF THE SPECTRA OF THE CONVOLUTION MODEL

The evolution of the weights is in our reduced problem given as

$$\frac{\partial}{\partial t}\mathbf{w}_r = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\mathbf{v}) \int_{-\infty}^{\infty} \times \frac{h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{r'}))}{\int_{-\infty}^{\infty} h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{r'}))dr'} h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))dr' \times (\mathbf{v}-\mathbf{w}_r)dv_x dv_y .$$

We choose for the calculations

$$h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_r)) = \frac{1}{2\pi\sigma_{\mathcal{R}}^2}\exp\left(-\frac{d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_r)}{2\sigma_{\mathcal{R}}^2}\right)$$

with

$$d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_r) := \frac{1}{\sqrt{2\pi}\sigma_n}\int_{-\infty}^{\infty}\exp\left(-\frac{(r-r')^2}{2\sigma_n^2}\right)(\mathbf{v}-\mathbf{w}_{r'})^2 dr'$$

and

$$h_{\mathcal{N}}(d_{\mathcal{N}}(r,r')) = \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{N}}}\exp\left(-\frac{d_{\mathcal{N}}(r,r')}{2\sigma_{\mathcal{N}}^2}\right), \quad \text{with } d_{\mathcal{N}}(r,r') := (r-r')^2.$$

We take the ansatz

$$\mathbf{w}_r = \mathbf{w}_{0r} + \boldsymbol{\delta}_r = [r+\delta_x(r), \delta_y(r)].$$

We get

$$\frac{\partial}{\partial t}\mathbf{w}_r = \frac{\partial}{\partial t}\boldsymbol{\delta}_r = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\mathbf{v}) f_1(\mathbf{w}_{0r}+\boldsymbol{\delta}_r) f_2(\mathbf{w}_{0r}+\boldsymbol{\delta}_r) \times f_3(\mathbf{w}_{0r}+\boldsymbol{\delta}_r),$$

where we expand the four functions around the stationary solution $\mathbf{w}_{0r}$ up to terms of first order:

$$f_1(\mathbf{w}_{0r}+\boldsymbol{\delta}_r) := \int_{-\infty}^{\infty} h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{0r'}+\boldsymbol{\delta}_r)) h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))dr'$$

$$\approx \int_{-\infty}^{\infty} h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{0r'})) h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))dr'$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp[-(r'-r'')/2\sigma_n^2]}{\sqrt{2\pi}\sigma_n\sigma_{\mathcal{R}}^2}(\mathbf{v}-\mathbf{w}_{0r''})\boldsymbol{\delta}_{r''}dr'' h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{0r'})) h_{\mathcal{N}}(d_{\mathcal{N}}(r,r'))dr',$$

$$f_2(\mathbf{w}_{0r}+\boldsymbol{\delta}_r) := \left(\int_{-\infty}^{\infty} h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{0r'}+\boldsymbol{\delta}_r))dr'\right)^{-1}$$

$$\approx \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{R}}}\exp\left(\frac{v_y^2+\sigma_n^2}{2\sigma_{\mathcal{R}}^2}\right)\left(1+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[-(r'-r'')/2\sigma_n^2]}{2\pi\sigma_n\sigma_{\mathcal{R}}^2}(\mathbf{v}-\mathbf{w}_{0r''})\boldsymbol{\delta}_{r''}dr'' h_{\mathcal{R}}(d_{\mathcal{R}}(\mathbf{v},\mathbf{w}_{0r'}))dr'\right),$$

$$f_3(\mathbf{w}_{0r}+\boldsymbol{\delta}_r) := \begin{pmatrix} v_x - r - \delta_x(r) \\ v_y - \delta_y(r) \end{pmatrix}.$$

Neglecting terms with higher order than one and using the fact of vanishing mean in the y direction, we get

$$\frac{\partial}{\partial t}\boldsymbol{\delta}(r) = -\boldsymbol{\delta}(r) + \frac{1}{2\pi\sigma_{\mathcal{N}}\sigma_{\mathcal{R}}^3}\left(\boldsymbol{\psi}(r) - \frac{\sigma_{\mathcal{N}}}{\sqrt{\sigma_{\mathcal{N}}^2+\sigma_{\mathcal{R}}^2}}\boldsymbol{\varphi}(r)\right)$$

with

$$\psi_x(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(v_x - r)\frac{\exp[-(r'-r'')^2/2\sigma_n^2]}{\sqrt{2\pi}\sigma_n}(v_x - r')\,\delta_x(r'')\exp\left(-\frac{(v_x-r')^2}{2\sigma_{\mathcal{R}}^2}\right)\exp\left(-\frac{(r-r')^2}{2\sigma_{\mathcal{N}}^2}\right)dr''dr'dv_x,$$

$$\varphi_x(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(v_x - r)\exp\left(-\frac{(v_x-r)^2}{2(\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2)}\right)\frac{\exp[-(r'-r'')^2/2\sigma_n^2]}{\sqrt{2\pi}\sigma_n}(v_x - r')\,\delta_x(r'')\exp\left(-\frac{(v_x-r')^2}{2\sigma_{\mathcal{R}}^2}\right)dr''dr'dv_x$$

and in the y direction

$$\psi_y(r) = \langle v_y^2\rangle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp[-(r'-r'')^2/2\sigma_n^2]}{\sqrt{2\pi}\sigma_n}\delta_y(r'')\exp\left(-\frac{(v_x-r')^2}{2\sigma_{\mathcal{R}}^2}\right)\exp\left(-\frac{(r-r')^2}{2\sigma_{\mathcal{N}}^2}\right)dr''dr'dv_x,$$

$$\varphi_y(r) = \langle v_y^2\rangle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\frac{(v_x-r)^2}{2(\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2)}\right)\frac{\exp[-(r'-r'')^2/2\sigma_n^2]}{\sqrt{2\pi}\sigma_n}\delta_y(r'')\exp\left(-\frac{(v_x-r')^2}{2\sigma_{\mathcal{R}}^2}\right)dr''dr'dv_x.$$

If we consider the dynamics related to the wave numbers k of the perturbation, the convolutions in $\boldsymbol{\psi}(r)$ and $\boldsymbol{\varphi}(r)$ translates to simple products. We get

$$\frac{\partial}{\partial t}\hat{\boldsymbol{\delta}}(k) = -\hat{\boldsymbol{\delta}}(k) + \frac{1}{2\pi\sigma_{\mathcal{N}}\sigma_{\mathcal{R}}^3}\left(\hat{\boldsymbol{\psi}}(k) - \frac{\sigma_{\mathcal{N}}}{\sqrt{\sigma_{\mathcal{N}}^2+\sigma_{\mathcal{R}}^2}}\hat{\boldsymbol{\varphi}}(k)\right)$$

with

$$\hat{\psi}_x(k) = \hat{\delta}_x(k)2\pi\sigma_{\mathcal{R}}^3\sigma_{\mathcal{N}}\left(1 + k^2\frac{\sigma_{\mathcal{N}}^2\sigma_n^2}{\sigma_{\mathcal{R}}^2}\right)\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right),$$

$$\hat{\varphi}_x(k) = \hat{\delta}_x(k)2\pi\sigma_{\mathcal{R}}\sqrt{\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2}(\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2)(\sigma_{\mathcal{R}}^2+\sigma_n^2)k^2\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2+2\sigma_{\mathcal{R}}^2)}{2}\right),$$

and

$$\hat{\psi}_y(k) = \hat{\delta}_y(k)\langle v_y^2\rangle 2\pi\sigma_{\mathcal{R}}\sigma_{\mathcal{N}}\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right),$$

$$\hat{\varphi}_y(k) = \hat{\delta}_y(k)\langle v_y^2\rangle 2\pi\sigma_{\mathcal{R}}\sqrt{\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2}\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2+2\sigma_{\mathcal{R}}^2)}{2}\right).$$

The linearized evolution in the fourier space of a small perturbation is given by an eigenvalue equation

$$\frac{\partial}{\partial t}\hat{\boldsymbol{\delta}}(k) = \lambda(k)\,\hat{\boldsymbol{\delta}}(k)$$

with the spectra of eigenvalues

$$\lambda_x(k) = -1 + \left(1 + k^2\frac{\sigma_{\mathcal{N}}^2\sigma_n^2}{\sigma_{\mathcal{R}}^2} - k^2\frac{(\sigma_{\mathcal{R}}^2+\sigma_{\mathcal{N}}^2)(\sigma_{\mathcal{R}}^2+\sigma_n^2)}{\sigma_{\mathcal{R}}^2}\exp(-k^2\sigma_{\mathcal{R}}^2)\right)\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right),$$

$$\lambda_y(k) = -1 + \langle v_y^2\rangle\frac{[1-\exp(-k^2\sigma_{\mathcal{R}}^2)]}{\sigma_{\mathcal{R}}^2}\exp\left(-\frac{k^2(\sigma_{\mathcal{N}}^2+\sigma_n^2)}{2}\right).$$

APPENDIX B: THE EXPANSION OF THE EVOLUTION EQUATION

We expand the evolution equation [Eq. (3.2)], with the ansatz

$$\mathbf{w}_r := (r + a_x \sin(2kr), a_y \cos(kr))$$

around $(a_x, a_y) = (0,0)$. Using the symmetry of the equation and neglecting the influence of higher modes, we have

$$\dot{a}_x \approx \lambda_x a_x + c_x a_y^2, \tag{B1}$$

$$\dot{a}_y \approx \lambda_y a_y + c_y a_x a_y + d_y a_y^3. \tag{B2}$$

With similar computations as in Appendix A we have for the coefficients

$$c_x = k\left[\left((\sigma_{\mathcal{N}}^2+\sigma_{\mathcal{R}}^2)\frac{1-\exp(-4k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}-1\right)\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]-\langle y^2\rangle\right.$$
$$\left.\times\left((\sigma_{\mathcal{N}}^2+\sigma_{\mathcal{R}}^2)\frac{[1-\exp(-2k^2\sigma_{\mathcal{R}}^2)][1-\exp(-k^2\sigma_{\mathcal{R}}^2)]^2}{\sigma_{\mathcal{R}}^4}-\frac{1-\exp(-k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}\right)\exp[-k^2(\sigma_n^2+2\sigma_{\mathcal{N}}^2)]\right], \tag{B3}$$

$$c_y = 2k\left[\left((\sigma_n^2+\sigma_{\mathcal{R}}^2)\frac{1-\exp(-4k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}-1\right)\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]-\langle y^2\rangle\right.$$
$$\left.\times\left((\sigma_n^2+\sigma_{\mathcal{R}}^2)\frac{[1-\exp(-2k^2\sigma_{\mathcal{R}}^2)][1-\exp(-k^2\sigma_{\mathcal{R}}^2)]^2}{\sigma_{\mathcal{R}}^4}-\frac{1-\exp(k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}\right)\exp\left(-\frac{k^2}{2}(3\sigma_n^2+\sigma_{\mathcal{N}}^2)\right)\right], \tag{B4}$$

$$d_y = \frac{1}{8}\left\{\frac{1-\exp(-4k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2}\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]-\langle y^2\rangle\frac{[1-\exp(-2k^2\sigma_{\mathcal{R}}^2)][1-\exp(-k^2\sigma_{\mathcal{R}}^2)]^2}{\sigma_{\mathcal{R}}^4}\right.$$
$$\left.\times\left[\exp[-k^2(\sigma_n^2+2\sigma_{\mathcal{N}}^2)]+\exp\left(-\frac{k^2}{2}(5\sigma_n^2+\sigma_{\mathcal{N}}^2)\right)\right]-\langle y^4\rangle\frac{[1-\exp(-k^2\sigma_{\mathcal{R}}^2)]^4}{\sigma_{\mathcal{R}}^6}\exp\left(-\frac{k^2}{2}(3\sigma_n^2+\sigma_{\mathcal{N}}^2)\right)\right\}. \tag{B5}$$

Inserting the stationary solution of Eq. (B1) in the stationary solution of Eq. (B2), we have

$$a_y^2 = \frac{\lambda_x\lambda_y}{c_x c_y - \lambda_x d_y}. \tag{B6}$$

The amplitude equation Eq. (B6) shows the bifurcation behavior of the model. Normally the spectrum of eigenvalues λ_y is used as control parameter for the bifurcation. The factor $c_x c_y - \lambda_x d_y$ indicates the type of the of the bifurcation. Is this factor negative, then there is a foreward bifurcation (a continuous phase transition). In the other case a backward bifurcation with an discontinuous phase transition and some hysteresis occurs. The model shows a continuous phase transition at all parameters. In the winner takes all limit $(\sigma_{\mathcal{R}} = 0)$ we have

$$c_x = k\{(4\sigma_{\mathcal{N}}^2k^2-1)\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]+\langle y^2\rangle k^2$$
$$\times\exp[-k^2(\sigma_n^2+2\sigma_{\mathcal{N}}^2)]\}, \tag{B7}$$

$$c_y = 2k\left[(4\sigma_n^2k^2-1)\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]+\langle y^2\rangle k^2\right.$$
$$\left.\times\exp\left(-\frac{k^2}{2}(3\sigma_n^2+\sigma_{\mathcal{N}}^2)\right)\right], \tag{B8}$$

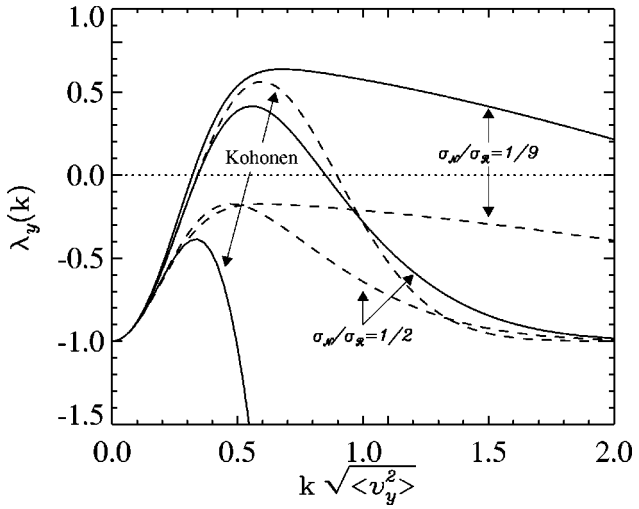

FIG. 12. Comparison of the spectra of the linearized evolution equations (broken lines) with the spectra of the corresponding quadratic part of the Lyapunov function (lines) related to a perturbation only in the y direction near marginal stability. In all cases is $\sigma_n = 0$. The Lyapunov function predicts a different stability behavior than the dynamical systems actually have.

$$d_y = \frac{k^2}{2}\exp[-2k^2(\sigma_n^2+\sigma_{\mathcal{N}}^2)]. \tag{B9}$$

In the SOFM limit $\sigma_n = 0$ we get the same results as in Ref. [58].

APPENDIX C: COUNTER EXAMPLE TO THE SUPPOSED LYAPUNOV FUNCTION OF THE KOHONEN MODEL

We investigate the validity of an ansatz for a Lyapunov function of the convolution model in the case of nonpotentiality $h_{\mathcal{N}} \neq h_n$:

$$V := \frac{1}{2}\int_{\mathcal{M}} p(\mathbf{v}) \int_{\mathcal{N}} (\mathbf{v}-\mathbf{w}_r)^2 e_r(\mathbf{v}) dr d\mathbf{v}.$$

A simple method to test that V is a Lyapunov function, is to compare the local stability conditions of the dynamical system and the Lyapunov function

$$\mathcal{L} \overset{!}{<} 0 \Leftrightarrow \mathcal{Q} > 0$$

with

$$\frac{\partial}{\partial t}(\mathbf{w}_0+\boldsymbol{\delta}) \approx \mathcal{L}(\mathbf{w}_0)\boldsymbol{\delta},$$

$$V(\mathbf{w}_0+\boldsymbol{\delta}) \approx V(\mathbf{w}_0) + \frac{1}{2}\boldsymbol{\delta}\mathcal{Q}(\mathbf{w}_0)\boldsymbol{\delta}.$$

The operator $\mathcal{L}$ describes the linearized evolution of the dynamical system and $\mathcal{Q}$ is the quadratic term of expansion of the lyapunov function V. The implications made above have to hold for any perturbation. Hence, we can restrict us to the case of a perturbation only in the y direction. Then the problem reduces to a simple comparison between the spectrum λ_y of the operator $\mathcal{L}$ in the y direction [Eq. (5.2)] and the negative spectrum of eigenvalues of $\mathcal{Q}_{yy} = \partial^2_{\delta_y} V(\mathbf{w}_0)$. We consider the stability behavior around the stationary solution $\mathbf{w}_0 = (r,0)$, which is a ordered, smooth map. This allows for an examination of the results in Ref. [71].

In Fourier space the operator $\mathcal{Q}_{yy}$ is diagonal, as a result of the translation invariance. We have

$$\hat{\mathcal{Q}}_{yy}(k) = \frac{1}{2}\left[1 - \langle v_y^2\rangle \frac{1-(2+\sigma_{\mathcal{R}}^2 k^2)\exp(-k^2\sigma_{\mathcal{R}}^2)}{\sigma_{\mathcal{R}}^2} \times \exp\left(-\frac{(\sigma_{\mathcal{N}}^2+\sigma_n^2)k^2}{2}\right)\right].$$

Figure 12 shows, that V is in general *not* a lyapunov function of the neural map model of Eqs. (2.1),(2.7). The spectrum of the linearized system predicts another stability behavior than the quadratic term of the lyapunov function.

[1] W. Singer, in *The Neuronal and Molecular Basis of Learning*, edited by J. Changeux and M. Konishi (Wiley, Chichester, 1987), pp. 301–336.
[2] W. Singer, Science **270**, 758 (1995).
[3] C. J. Shatz, Neuron **5**, 745 (1990).
[4] C. J. Shatz, Curr. Opin. Neurobiol. **2**, 78 (1992).
[5] M. P. Stryker, *Neuroscience Research Program Bulletin* (MIT Press, Cambridge, MA, 1979), Vol. 15, No. 3, Chap. VII, pp. 454–462.
[6] M. P. Stryker, in *Development of the Visual System*, Vol. 3 of *Proceedings of the Retina Research Foundation Symposia*, edited by D. M. Lam and C. J. Shatz (MIT Press, Cambridge, MA, 1991), Chap. 16, pp. 267–288.
[7] J. H. Kaas, in *Neurobiology of Neocortex*, edited by P. Rakic and W. Singer (Wiley, New York, 1988), pp. 101–113.
[8] M. Stryker *et al.*, in *Neurobiology of Neocortex*, edited by P. Rakic and W. Singer (Wiley, New York, 1988), pp. 115–136.
[9] L. C. Katz, Curr. Opin. Neurobiol. **3**, 93 (1993).
[10] L. C. Katz and C. J. Shatz, Science **274**, 1133 (1996).
[11] M. Constantine-Paton, H. T. Cline, and E. Debski, Annu. Rev. Neurosci. **13**, 129 (1990).
[12] N. W. Daw, *Visual Development* (Plenum, New York, 1995).
[13] K. D. Miller, in *Models of Neural Networks III*, edited by E. Domany, J. van Hemmen, and K. Schulten (Springer-Verlag, New York, 1995).
[14] N. V. Swindale, Network **7**, 161 (1996).
[15] S. LeVay, D. H. Hubel, and T. N. Wiesel, J. Comp. Neurol. **159**, 559 (1975).
[16] C. J. Shatz, S. Lindström, and T. Wiesel, Brain Res. **131**, 103 (1977).
[17] D. H. Hubel and T. N. Wiesel, J. Neurophysiol. **28**, 1041 (1965).
[18] D. H. Hubel and T. N. Wiesel, Proc. R. Soc. London, Ser. B **198**, 1 (1977).
[19] C. J. Shatz and M. P. Stryker, J. Physiol. (London) **281**, 267 (1978).
[20] S. Löwel and W. Singer, Exp. Brain Res. **68**, 661 (1987).
[21] P. Anderson, J. Olavaria, and R. V. Sluyters, J. Neurosci. **8**, 2183 (1988).
[22] S. LeVay, M. P. Stryker, and C. Shatz, J. Comp. Neurol. **179**, 223 (1979).
[23] C. S. Goodman and C. J. Shatz, Neuron **10**, 77 (1993).
[24] G. S. Stent, Proc. Natl. Acad. Sci. USA **70**, 997 (1973).
[25] J.-P. Changeux and A. Danchin, Nature (London) **264**, 705 (1976).
[26] S. Löwel, J. Neurosci. **14**, 7451 (1994).
[27] G. J. Goodhill, Biol. Cybern. **69**, 109 (1993).
[28] S. Tieman and N. Tumosa, Visual Neurosci. **14**, 929 (1997).
[29] K. D. Miller, J. Keller, and M. Stryker, Science **245**, 605 (1989).
[30] K. D. Miller and M. P. Stryker, in *Connectionist Modeling and Brain Function: The Developing Interface*, edited by S. Han-

son and C. Olson (MIT Press/Bradford, Cambridge, Mass, 1990), pp. 255–350.
[31] D. O. Hebb, *The Organisation of Behavior* (Wiley, New York, 1949).
[32] T. Kohonen, Biol. Cybern. **43**, 59 (1982).
[33] T. Kohonen, *Self-Organisation and Associative Memory*, No. 8 in *Springer Series in Information Sciences* (Springer-Verlag, Heidelberg, 1984).
[34] H. Ritter and K. Schulten, Biol. Cybern. **54**, 99 (1986).
[35] H. Ritter and K. Schulten, Biol. Cybern. **60**, 59 (1988).
[36] K. Obermayer, H. Ritter, and K. Schulten, Proc. Natl. Acad. Sci. USA **87**, 8345 (1990).
[37] K. Obermayer, G. G. Blasdel, and K. Schulten, Phys. Rev. A **45**, 7568 (1992).
[38] A. Takeuchi and S. Amari, Biol. Cybern. **35**, 63 (1979).
[39] H. R. Wilson and J. D. Cowan, Kybernetik **13**, 55 (1973).
[40] R. Douglas *et al.*, Science **269**, 981 (1995).
[41] D. C. Somers, S. B. Nelson, and M. Sur, J. Neurosci. **15**, 5448 (1995).
[42] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proc. Natl. Acad. Sci. USA **92**, 3844 (1995).
[43] R. Yuste, D. A. Nelson, W. W. Rubin, and L. C. Latz, Neuron **14**, 7 (1995).
[44] D. Burr, in *Proceedings of the IEEE International Conference on Neural Networks* (IEEE Press, New York, 1988), pp. 69–76.
[45] P. Dayan, Neural Comput. **5**, 392 (1993).
[46] F. Mulier and V. Cherkassky, Neural Comput. **7**, 1165 (1995).
[47] P. D. Simic, Network **1**, 89 (1990).
[48] A. Yuille and D. Kammen, in *Models for the Development of the Visual Cortex* (Addison Wesley, Reading, MA, 1988), pp. 393–410.
[49] R. Durbin and G. Mitchinson, Nature (London) **343**, 644 (1990).
[50] A. Das and C. D. Gilbert, Nature (London) **375**, 780 (1995).
[51] A. Grinvald, E. E. Lieke, R. D. Frostig, and R. Hildesheim, J. Neurosci. **5**, 2545 (1994).
[52] B. Krekelberg and J. G. Taylor, in *ICANN'95 International Conference on Artificial Neural Networks*, ICANN, edited by F. Fogelman-Soulie and P. Gallinari (EC2 & Cie, Paris, 1995), pp. 525–530.
[53] K. D. Miller, Neuron **17**, 371 (1996).
[54] A. L. Yuille, Neural Comput. **2**, 1 (1990).
[55] M. C. Cross and P. C. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).
[56] T. M. Heskes and B. Kappen, in *Proceedings of the IEEE International Conference on Neural Networks 1993* (IEEE Press, New York, 1993), p. 1219.
[57] G. Goodhill, Ph.D. thesis, University of Sussex at Brighton, 1992.
[58] R. Der and M. Herrmann, Phys. Rev. E **49**, 5840 (1994).
[59] O. Scherf, K. Pawelzik, and T. Geisel, in *Computational Neuroscience: Trends in Research 1997*, edited by J. M. Bower (Plenum, New York, 1997), pp. 485–489.
[60] K. D. Miller, in *Correlation-Based Models of Neural Development*, edited by M. A. Gluck and D. E. Rumelhart (Lawrence Erlbaum Associates, Hillsdale, NY, 1990), Chap. 7, pp. 267–353.
[61] R. Linsker, Proc. Natl. Acad. Sci. USA **83**, 8779 (1986).
[62] S. Tanaka, Biol. Cybern. **65**, 91 (1991).
[63] E. Erwin, K. Obermayer, and K. Schulten, Neural Comput. **7**, 425 (1995).
[64] H. Ritter, T. Martinetz, and K. Schulten, in *Neuronale Netze: Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke, Reihe Künstliche Intelligenz*, edited by W. Wahlster (Addison-Wesley, Bonn, 1991).
[65] F. Hoffsümmer *et al.*, in *Computational Neuroscience*, edited by J. M. Bower (Academic, San Diego, 1996), pp. 197–202.
[66] G. J. Goodhill and D. J. Willshaw, Neural Comput. **6**, 615 (1994).
[67] G. G. Blasdel and G. Salama, Nature (London) **321**, 579 (1986).
[68] M. C. Crair, D. C. Gillespie, and M. P. Stryker, Science **279**, 566 (1998).
[69] P. Manneville, in *Dissipative Structures and Weak Turbulence*, edited by H. Araki, A. Libchaber, and G. Parisi (Academic, London, 1990).
[70] H. U. Bauer, M. Riesenhuber, and T. Geisel, Phys. Rev. E **54**, 2807 (1996).
[71] E. Erwin, K. Obermayer, and K. Schulten, Biol. Cybern. **67**, 47 (1992).