Artificial Neural Networks
T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (Editors)
© Elsevier Science Publishers B.V. (North-Holland), 1991

A neural network model for the formation and for the spatial structure of retinotopic maps, orientation- and ocular dominance columns

K. Obermayer[1], G. G. Blasdel[2] and K. Schulten[1]

[1]Beckman-Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, Email: oby@lisboa.ks.uiuc.edu
[2]Harvard Medical School, Harvard University, Boston, MA 02115, USA

Abstract:

We demonstrate that important features in the spatial structure of retinotopic maps, orientation- and ocular dominance columns in the primary visual cortex, can be explained as the result of a mapping from a five-dimensional feature space onto a two-dimensional cortical surface under the constraint that (*i*) the mapped features vary smoothly along the cortical surface, and (*ii*) the mapping is established by an activity-based self-organizing process. We generate our model maps by using the self-organizing feature map algorithm [1,2], which is known to implement the above mentioned principles in a biologically plausible way. We characterize the spatial structure of the model maps by their Fourier transform and correlation functions, and we study the interaction between both (model-) column systems, and between them and the retinotopic map. Numerical simulations are supplemented by a mathematical analysis. Results are compared with experimental data obtained from area 17 of the macaque.

1. Introduction

It has been proposed that the spatial structure of cortical maps results from the requirement of establishing an appropriate representation of relevant stimulus features along the two-dimensional cortical surface [3,4]. In the case of feature maps found in the primary visual area(s), these spatial representations seem to be topographic to some extent, such that neighboring patches of cortex represent similar stimulus features (or feature combinations) [5-10]. Since the number of relevant features is higher than two, the reverse does not hold, and similar features are not necessarily always mapped to adjacent regions in cortex. The spatial structure of these maps must be understood as a compromise, preserving some neighborhood relations at the expense of others.

There are, however, many ways that this can be achieved. In a previous study [11], we investigated the joint formation of a retinotopic projection and an orientation column system under two additional constraints: (*i*) the response properties of the cells should vary over the cortical surface as smoothly as possible ("principle of continuous mapping"), and (*ii*) the spatial structure of maps must be consistent with a – to some degree random – activity dependent, self-organizing process.

In this contribution we restrict ourselves to a "low-dimensional" version of this model. However, we extend these investigations along three lines: (*i*) we include ocular dominance as an additional feature and present results on the correlation in the spatial structure of both systems, (*ii*) we characterize the model-maps by their Fourier spectra and correlation functions and compare the results with an experimentally observed map from area 17 of the macaque, and (*iii*) we provide a mathematical analysis of the model with respect to the origin of the observed feature hierarchy.

The capability of this model to explain a variety of experimental findings with a small set of simple rules indicates that, although there are several different mechanisms at work in biological systems, their effects might be functionally similar, indicating functional redundancy.

Description of the model

In our model, the cortical surface is divided into $N \times N$ small patches designated by a position vector $\vec{r}$, which are considered as "units" of a two-dimensional lattice (network layer) with periodic boundary conditions (to avoid edge effects).

The functional properties of all neurons that are located in a common patch $\vec{r}$ are characterized by a feature vector $\vec{w}_{\vec{r}}$ associated with each unit $\vec{r}$. The components $(\vec{w}_{\vec{r}})_k$ are interpreted as receptive field properties of these neurons. The feature vectors $\vec{w}_{\vec{r}}$ as a function of unit locations $\vec{r}$, describe the spatial distribution of feature selectivity of cells over the cortical layer, i.e. the cortical map.

In the sequel we will consider the following receptive field properties: position of the receptive field centers in visual space $(x_{\vec{r}},\ y_{\vec{r}})$, preferred orientation $(\phi_{\vec{r}})$, and two quantities which qualitatively can be interpreted as orientation specificity $(q_{\vec{r}})$ (see e.g. [12]) and ocular dominance $(z_{\vec{r}})$ (see e.g. [13]). If $q_{\vec{r}}$ is zero, then the units are unspecific for orientation, and the larger $q_{\vec{r}}$ becomes, the sharper the units are tuned to their preferred orientation. "Binocular" units are characterized by $z_{\vec{r}} = 0$, "monocular" units by $z_{\vec{r}} \to \pm\infty$. These properties are encoded by the 5-dimensional feature vectors $\vec{w}_{\vec{r}}$:

$$\vec{w}_{\vec{r}} = (x_{\vec{r}},\ y_{\vec{r}},\ q_{\vec{r}}\cos(2\phi_{\vec{r}}),\ q_{\vec{r}}\sin(2\phi_{\vec{r}}),\ z_{\vec{r}}) \tag{1}$$

where the orientation coordinates are given in their cartesian forms (see [12]).

The input patterns presented to the network layer are described by a feature vector also, which is of the same dimensionality as $\vec{w}_{\vec{r}}$. Its components,

$$\vec{v} = (x,\ y,\ q\cos(2\phi),\ q\sin(2\phi),\ z) \tag{2}$$

correspond to the stimulus properties "position in the visual field" $(x,\ y)$, "orientation" ϕ, and to two quantities q and z qualitatively describing pattern ellipticity, and the distribution of activity between both eyes, respectively. Circular stimuli correspond to $q = 0$, elongated patterns to $q > 0$. A "binocular" stimulus corresponds to $z = 0$, while purely "monocular" stimuli correspond to $z \to \pm\infty$.

To generate a representation of features across the network layer, we use the self-organizing feature map algorithm [1,2]. This algorithm is an iterative procedure. At each step a feature vector $\vec{v}$ is chosen at random according to a probability distribution $P(\vec{v})$. Then the unit $\vec{s}$, whose feature vector $\vec{w}_{\vec{s}}$ is closest to the input pattern $\vec{v}$, is selected and the feature vectors in the network layer are changed according to

$$\vec{w}_{\vec{r}}(t+1) = \vec{w}_{\vec{r}}(t) + \varepsilon(t)h(\vec{r},\vec{s},t)(\vec{v} - \vec{w}_{\vec{r}}(t)) \tag{3}$$

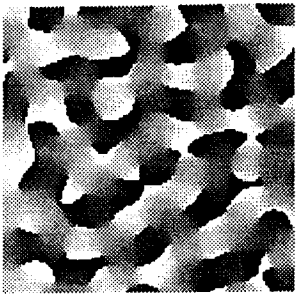
where $h(\vec{r},\vec{s},t)$, the neighborhood function, is given by:

$$h(\vec{r},\vec{s},t) = \exp\left(-(r_1 - s_1)^2)/\sigma_{h1}^2(t) - (r_2 - s_2)^2/\sigma_{h2}^2(t)\right). \tag{4}$$

Since little is known about the statistical properties of the afferent patterns that drive the map formation process, we decided to use an "unbiased" probability distribution $P(\vec{v})$ and drew the patterns with homogeneous probability from the manifold

$$V = \{\vec{v} \mid x, y \in [0,d],\ \phi \in [0,\pi],\ q < q_{pat},\ |z| < z_{pat}\}, \tag{5}$$

i.e. all stimuli characterized by q and $|z|$ smaller than a given value were chosen equally often.



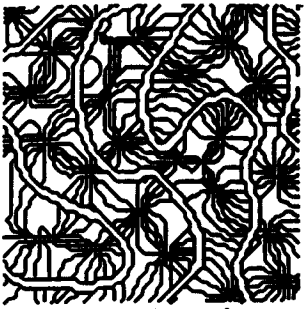


Fig. 1: (a) Left: Spatial distribution of orientation preference (isotropic neighborhood function). (b) Right: Contour plot of orientation preference (in 16^o steps). The borders of the ocular dominance stripes are indicated by white lines.

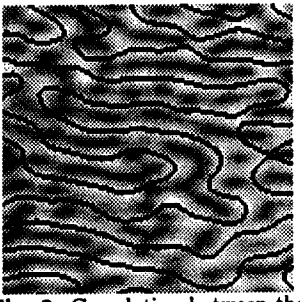
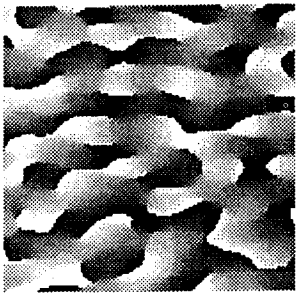


Fig. 3: Correlation between the orientation specificity and ocular dominance values (anisotropic neighborhood function).



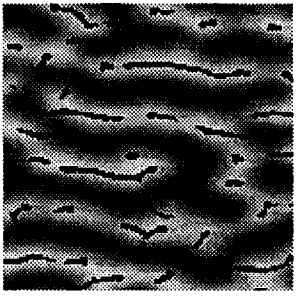


Fig. 2: (a) Left: Spatial distribution of orientation preference (anisotropic neighborhood function). (b) Right: Correlation between ocular dominance values (anisotropic neighborhood function) and "vortices" and "fractures" of the orientation map.

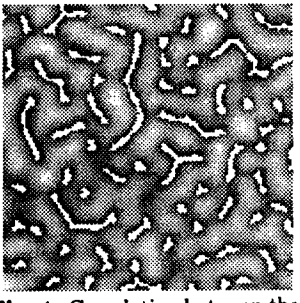


Fig. 4: Correlation between the "fractures", the "vortices" and the orientation specificity values.

The model provides an intermediate-level desciption of map formation, based on a small set of simple rules. Although these rules can be derived from developmental principles underlying activity driven processes [14,15], no detailed assumptions about a "neural implementation" have to be made. The results[1] are, therefore, valid for a whole class of mechanisms, which is an advantage in a situation where there are not enough data available to determine the large number of parameters a detailed model necessarily must have.

Spatial structure of model maps

Figure 1a shows the spatial distribution of "orientation preference" $\phi_{\vec{r}}$ (white $\rightarrow$ black: $0^o \rightarrow 180^o$) across the network layer for an artificial map generated with an isotropic neighborhood function. Units with similar orientation preference form domains of continuously changing orientation, in which iso-orientation regions are organized as elongated patches ("slabs"). This is even more visible in Fig. 1b which shows a contour plot of the orientation values for an enlarged section of Fig. 1a. In addition Fig. 1b shows the borders (white lines along which $z_{\vec{r}} = 0$) that separate

[1] as long as they are robust against parameter variations.

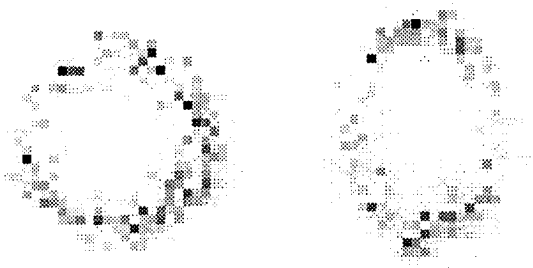


Fig. 5: Two-dimensional Fourier transform of the orientation coordinates for maps generated with an isotropic (a) (left) and anisotropic (b) (right) neighborhood function.

regions of opposite ocular dominance. The slabs start and end at vortices around which orientation preference changes by 180° in a clockwise- or counterclockwise fashion, and which seem to be preferably located in the center of the ocular dominance stripes. The slabs originating at these vortices often cross the border of the ocular dominance stripes at a steep angle, connecting vortices located in adjacent bands. Neighboring domains have similar slab-orientations, but on a larger length scale the directions of the domains are distributed isotropically.

Figure 2 shows the spatial distribution of orientation preference $\phi_{\vec{r}}$ (Fig. 2a) and ocular dominance values $z_{\vec{r}}$ ($z_{\vec{r}} < 0$: dark, $z_{\vec{r}} > 0$: bright) (Fig. 2b) across the network layer for a map generated with an anisotropic neighborhood function. The black lines in Fig. 2b indicate regions of high magnitude of the gradient $|\nabla_{\vec{r}}\phi_{\vec{r}}|$ (larger than 50° per patch diameter) of orientation preference ("fractures" and "vortices"). Again vortices and fractures exhibit a tendency to lie in the center of ocular dominance stripes. For maps generated using an anisotropic neighborhood function there exists a preferred direction along which iso-orientation slabs as well as ocular dominance stripes are aligned[2]. Anisotropic arrangement of slabs or bands has been observed, e.g. in the macaque (ocular dominance [16]) or cat (both systems [9]), the degree of anisotropy, however, might be different for the ocular dominance and orientation systems [6]. The model as formulated by eqs. (3) - (4) predicts an approximately equal degree of anisotropy for both systems. In order to account for observed differences the model needs to be extended by introducing an additional parameter.

Figure 3 displays the spatial distribution of orientation specificity ($q_{\vec{r}} = 0$: black, $q_{\vec{r}}$ =max.: white) for the map shown in Fig. 2a. The borders of the ocular dominance stripes ($z_{\vec{r}} = 0$) are marked by black lines. The network layer shows a tendency to segregate into "binocular" regions with higher orientation specificity and "monocular" regions with lower orientation specificity. Evidence for the occurence of such segregation in the macaque was recently found [17,18]. A comparison between Figs. 2b and 3 already indicates a correlation between "fractures" and "vortices" with regions containing units unspecific for orientation. This is more explicitly demonstrated in Fig. 4 where regions of high magnitude of the orientation gradient $|\nabla_{\vec{r}}\phi_{\vec{r}}|$ (white lines), are superimposed onto a map of orientation specifity $q_{\vec{r}}$ (Fig. 4 correponds to the map shown in Fig. 1).

Figure 5 shows the two-dimensional complex Fourier spectrum

$$\hat{w}_{\vec{k}} = \sum_{\vec{r}} e^{i\vec{k}\vec{r}} q_{\vec{r}}(\cos(2\phi_{\vec{r}}) + i\sin(2\phi_{\vec{r}})) \tag{6}$$

of the orientation coordinates $w_{\vec{r}3}$ and $w_{\vec{r}4}$ for maps generated with an isotropic (Fig. 5a) and an anisotropic (Fig. 5b) neighborhood function. The origin of the $\vec{k}$-plane is marked by a dot. Each pixel corresponds to a single mode $\vec{k}$ and its brightness indicates the mean square amplitude $|\hat{w}_{\vec{k}}|^2$ of the mode $\vec{k}$. For an isotropic neighborhood function the orientation map is characterized by wave vectors from a ring shaped region in the Fourier domain (Fig. 5a), which becomes eccentric with increasing σ_{h1}/σ_{h2} (Fig. 5b) until the ring dissolves into two separate groups of modes. Phases are random. Further analysis shows that the feature coordinates $w_{\vec{r}3}$, $w_{\vec{r}4}$ and $w_{\vec{r}5}$ as a function of unit coordinates $\vec{r}$ are all characterized by either a ring shaped or by a two group-Fourier spectrum.

[2]However, if the iso-orientation slabs cross ocular dominance borders they still do it at steep angles

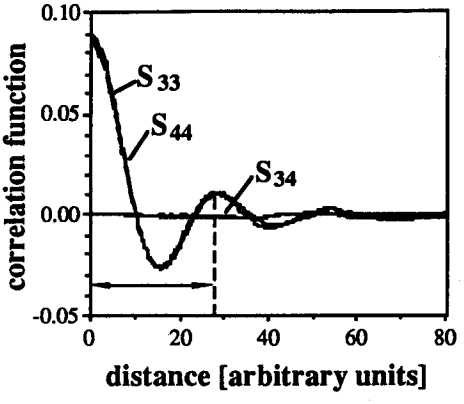


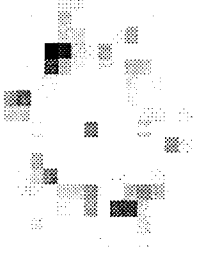
Fig. 6: Correlation functions of orientation values for a map generated with an isotropic neighborhood function.

This indicates that a random superposition of long-crested waves (as proposed e.g. by [19]) captures some essential aspects of the spatial structure of model maps (as well as of "real" maps, see Fig. 7 below).

Figure 6 shows the two-point correlation functions $S_{ij}(\vec{s}) = < w_{(\vec{r}-\vec{s})i}\, w_{(\vec{r})j} >$, $(i,j \in \{3,4\})$ for the map shown in Fig.1a (isotropic neighborhood function). Since in this case the correlation functions S_{ij} seem to depend only on the absolute value $|\vec{s}|$, S_{ij} has been averaged over all directions of $\vec{s}$. The arrow indicates the wavelength λ associated with the wave number $|\vec{k}|$ of modes from the ring-shaped region in Fig. 5a. The autocorrelation functions have a Mexican-hat shape. At a separation of $\lambda/2$ there is a tendency for the preferred orientations $\phi_{\vec{r}}$ to be orthogonal. For an anisotropic neighborhood function the autocorrelation functions no longer have rotational symmetry. They essentially consist of a central "bar" of positive values accompanied by two "lobes". of negative values. The same type of spectra and correlation functions are also obtained for the ocular dominance coordinate $w_{\vec{r}5}$.

There is experimental evidence for the type of power spectra and correlation functions displayed in Figs. 5 and 6. Figure 7 shows power spectrum and correlation functions for an orientation map obtained from area 17 of the macaque [6], an animal with an almost "isotropic" orientation map. The spectrum (Fig. 7a) is an anulus around the origin of the $\vec{k}$-plane, phases seem to be random. The autocorrelation functions (Fig. 7b) have "Mexican hat" shape, with a minimum at half the wavelength of the modes located on the ring in Fig. 7a. There is also evidence from area 18 of the cat, whose orientation column system is anisotropic. Swindale et al. showed [8] that the angle-autocorrelation function[3] consists of a central bar of parallel preferred orientation accompanied by two lobes of orthogonal orientation values.

[3] In their paper [8] they defined the correlation functions in a slightly different way.



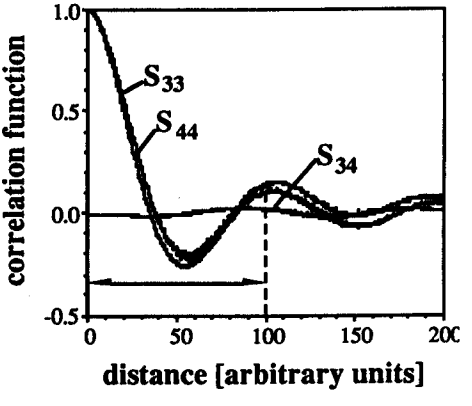


Fig. 7: 2d Fourier transform **(a)** (left) and correlation functions **(b)** (right) for an orientation map from area 17 of the macaque. Preferred orientation and orientation specificity were measured using voltage sensitive dyes [6].

Instability and phase transition

Although there is no general theory of the stationary states of the self-organizing feature map algorithm for a dimension reduction scenario, the observed power spectra and correlation functions can be qualitatively understood by a mathematical analysis for the case of small values of q_{pat} and z_{pat}.

If q_{pat} and z_{pat} are smaller than a certain threshold q_{thres}, then the stable stationary states of (3) - (4) correspond to a perfect topographic representation of visual space, with orientation and ocular dominance *not* represented in the map (i.e. $q_{\vec{r}} = z_{\vec{r}} = 0$). Let us consider an ensemble of networks, each characterized by a set $\{\vec{w}_{\vec{r}}\}$ of feature vectors, and denote the time-dependent distribution function of this ensemble by $S(\{\vec{w}\},t)$. Following a method derived in [20], we can describe the time-development of $S(\{\vec{w}\},t)$ near the stationary state by the Fokker-Planck equation

$$\frac{1}{\epsilon}\partial_t S(\{\vec{u}_{\vec{r}}\},t) = \sum_{\vec{r}m\vec{q}n} \frac{\partial}{\partial \vec{u}_{\vec{r}m}} B_{\vec{r}m\vec{q}n}\vec{u}_{\vec{q}n} S(\{\vec{u}_{\vec{r}}\},t) + \frac{\epsilon}{2}\sum_{\vec{r}m\vec{q}n} D_{\vec{r}m\vec{q}n}\frac{S(\{\vec{u}_{\vec{r}}\},t)}{\partial \vec{u}_{\vec{r}m}\partial \vec{u}_{\vec{q}n}} \qquad (7)$$

where the origin of $S(.,t)$ was shifted to the stationary state $\{\vec{w}_{\vec{r}}\}$, using now the new argument variable $\vec{u}_{\vec{r}} = \vec{w}_{\vec{r}} - \vec{\bar{w}}_{\vec{r}}$. The matrices $\underline{B}$ and $\underline{D}$ depend only on the pattern distribution (5), i.e. on q_{pat} and z_{pat} and on the stationary state $\{\vec{\bar{w}}_{\vec{r}}\}$ (for details see [20]). The eigenvalues of $\underline{B}$ determine the stability of the stationary state, while $\underline{B}$ and $\underline{D}$ together govern size and time development of fluctuations $< u_{\vec{r}i}u_{\vec{s}j} >$.

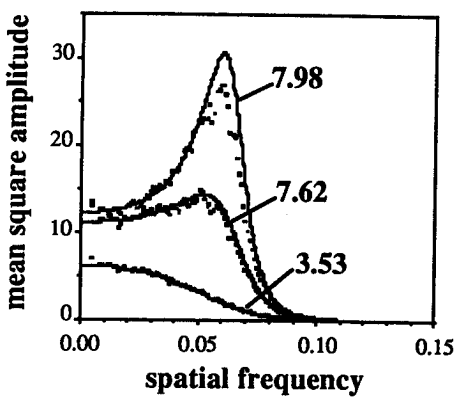


Fig. 8: Power spectrum of fluctuations of the orientation coordinates for a map generated with $q_{pat} < q_{thres}$ and an isotropic neighborhood function. Comparison of a Monte Carlo simulation (dots) with the analytically obtained solution of eq. (7) for three values of q_{pat}.

Let's define the Fourier modes $\vec{u}_{\vec{k}}$ of the equilibrium deviations $\vec{u}_{\vec{r}}$ by $\hat{\vec{u}}_{\vec{k}} = 1/N \sum_{\vec{r}} e^{i\vec{k}\vec{r}}\vec{u}_{\vec{r}}$. A mathematical analysis of (7) (details will be published elsewhere) shows that for q_{pat} or z_{pat} larger than

$$q_{thres} = \sqrt{e}\,\frac{d}{N}\min(\sigma_{h1},\sigma_{h2})$$
$$z_{thres} = \sqrt{\frac{3}{4}}e\,\frac{d}{N}\min(\sigma_{h1},\sigma_{h2}) \qquad (8)$$

the purely topographic stationary state becomes unstable and an orientation- and/or ocular dominance column system emerges. The threshold values given by eqs. (8) can

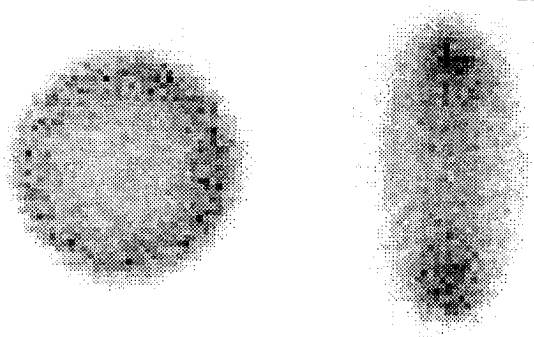


Fig. 9: 2d Fourier spectrum of the fluctuations in the orientation coordinates for q_{pat} slightly below q_{thres} for a map generated using an isotropic (a) (left) and an anisotropic (b) (right) neighborhood function.

be interpreted as essentially the distance in feature space which corresponds to the range of the neighborhood function in the network layer. The Fourier modes, which become unstable first, are characterized by $\tilde{\vec{u}}_{\vec{k}} \perp (x,y)$-plane and

$$|\vec{k}| = 2/\sigma_h \;\; if \;\; \sigma_{h1} = \sigma_{h2}, \quad \left.\begin{array}{rcl} k_x &=& \pm 2/\sigma_{h1} \\ k_y &=& 0 \end{array}\right\} \;\; if \;\; \sigma_{h1} < \sigma_{h2} \tag{9}$$

i.e. either a ring of modes (for $\sigma_{h1} = \sigma_{h2}$) or two modes with opposite sign of $\vec{k}$ (for $\sigma_{h1} \neq \sigma_{h2}$) become unstable. These modes do not couple, hence the phases of $\hat{w}_{\vec{k}}$ defined by eq. (6) are independent and random. The instability leads to large fluctuations in the amplitude of these modes and of nearby modes close to the threshold (8). Figure 8 shows the mean square amplitude (6) of fluctuations parallel to the orientation feature dimensions for an isotropic neighborhood function. The results from Monte Carlo simulations (dots) are confirmed by the analytical solution of eq. (7) (solid lines), and both are shown for three values of q_{pat} ($q_{thres} = 8.23$). The spectrum depends on the wavenumber only and evolves to an annulus around the origin of the $\vec{k}$-plane if q_{pat} approaches the threshold. This is demonstrated in Fig. 9a, which shows the mean square amplitude (0: white, max.: black) as a function of $\vec{k}$. Each pixel corresponds to one mode in the $\vec{k}$-plane.

Figure 9b displays the power spectrum of fluctuations for an anisotropic neighborhood function near threshold. Amplitudes are maximal in two groups of modes located around the two modes given by eq. (9). The fact, that either a ring or two groups of modes become unstable at the threshold, is reflected in the maps above threshold and gives rise to the observed Fourier spectrum shown in Figs. 5 and 6.

Acknowledgement: The authors would like to thank R. Brady and R. Kufrin for their help and support in all technical matters concerning the use of the Connection Machine on which calculations were carried out, and the Boehringer-Ingelheim Fonds for financial support by a scholarship to K. Obermayer. This research has been supported by the National Science Foundation (grant number 9017051). Computer time on the Connection Machine CM-2 has been made available by the National Center for Supercomputer Applications at Urbana-Champaign and the Pittsburgh Supercomputing Center, both supported by the National Science Foundation.

References

[1] Kohonen T. (1982a), Biol. Cybern. **43**, 59-69
[2] Kohonen T. (1982b), Biol. Cybern.**44**, 135-140
[3] Nelson M.E. and Bower J.M. (1990), TINS **13**, 401-406
[4] Durbin R. and Mitchison M. (1990), Nature **343**, 644-647
[5] Hubel D.H. and Wiesel T.N. (1974), J. Comp. Neurol. **158**, 267-294
[6] Blasdel G.G. and Salama G. (1986), Nature **321**, 579-585
[7] Grinvald A. et al. (1986), Nature **324**, 361-364
[8] Swindale N.V. et al. (1987), J. Neurosci. **7**, 1414-1427
[9] Löwel S. et al. (1987), J. Comp. Neurol. **255**, 401-415
[10] Ts'o D.Y. et al. (1990), Science **249**, 417-420
[11] Obermayer K. et al. (1990), Proc. Natl. Acad. Sci. USA **87**, 8345-8349
[12] Swindale N.V. (1982), Proc. R. Soc. Lond., **B215**, 211-230
[13] Goodhill G.J. and Willshaw D.J. (1990), Network **1**, 41-59
[14] von der Malsburg C. (1973), Kybernetik **14**, 85-100
[15] Kohonen T. (1983), Self-Organization and Associative Memory, Springer-Verlag, New York
[16] Hubel D.H. and Wiesel T.N. (1977), Proc. R. Soc. B **198**, 1-59
[17] Livingstone M.S. and Hubel D.H. (1984), J. Neurosci. **4**, 309-356
[18] Blasdel G.G. (1991), J. Neurosci., submitted
[19] Rojer A.S. and Schwarz E.L. (1990), Biol. Cybern. **62**, 381-391
[20] Ritter H. and Schulten K. (1989), Biol. Cybern. **60**, 59-71