Stimulus-dependent correlations in stochastic networks

H. J. Kappen
RWCP Novel Function SNN Laboratory, Department of Biophysics, University of Nijmegen, Geert Grooteplein 21,
NL 6525 EZ Nijmegen, The Netherlands

(Received 27 September 1996; revised manuscript received 8 January 1997)

It has been observed that cortical neurons display synchronous firing for some stimuli and not for others. The resulting synchronous cell assemblies are thought to form the basis of object perception. In this paper this ‘‘dynamic linking’’ phenomenon is demonstrated in networks of binary neurons with stochastic dynamics. Analytical treatment within the mean field theory and linear response theory is possible and is compared with simulations. We establish that correlations are a sensitive function of the spatial coherence in the stimulus. We discuss the possibility to use these correlations as a mechanism for scene segmentation.
[S1063-651X(97)07705-2]

PACS number(s): 87.10.+e, 02.70.−c, 05.50.+q, 87.22.As

I. INTRODUCTION

It is well established that the behavior of sensory neurons in the visual cortex can be described by a receptive field: A neuron is sensitive to certain specific stimuli and not to others [1]. It is often assumed that the role of individual cells is to represent *local* visual features, such as edges, corners, velocities, colors, etc. These representations may coexist on several length scales. The representation of local receptive fields or features is encoded in the feed-forward synaptic connections of individual neurons. This representation is thought to be an efficient information-theoretic description of the local structure of images [2].

Objects are generally believed to be represented by a collection of local features. The neurons that represent the local features of the object become active and constitute a so-called cell assembly [3]. The cell assembly is a neural representation of the object.

Since a visual image generally contains many objects simultaneously, many cell assemblies can be active at the same time. Therefore some labeling mechanism must exist to distinguish whether active neurons belong to the same cell assembly or to different cell assemblies. There exist various proposals to facilitate such a mechanism. One proposal is based on the synchronization of the firing patterns between neurons [4–6]. It is assumed that the resulting synchronous subpopulations of neurons form the basis of segmentation and object perception [7,8].

There is some experimental evidence that neurons in the visual cortex display synchronous firing for some stimuli and not for others [9–12]. In particular, some studies show that synchrony depends on the amount of conflict in the stimulus presented [13,14]. Thus if features are part of the same object, the corresponding neurons synchronize. If the same features are not part of the same object, no such synchronization occurs. The observed synchrony has in fact two components: one is the presence or absence of a central peak in the crosscorrellograms [11,14]. An additional aspect is the presence or absence of an oscillatory component in the autocorrellograms and crosscorrellograms [9,10]. Both phenomena could play a functional role as a mechanism for feature linking.

So far, most models have been based on oscillations and have addressed two key questions. One question is how to implement dynamic feature linking, i.e., how synchrony between neurons can arise for some stimuli and not for others. In [15] a network of bursting neurons is considered. In this model, stimulus-dependent assembly formation is based on fast synaptic modulations. References [16–18] introduce a network of pairs of nonlinear oscillators which models an orientation column. The network involves specific delayed synchronizing and desynchronizing connections that can be learned. Reference [19] discusses a network of integrate-and-fire neurons organized in orientation columns. Both these models display stimulus-dependent assembly formation in the sense that oscillations synchronize for spatially coherent stimuli and can be made to desynchronize for incoherent stimuli, without changing the synaptic strengths. Similar findings are reported in [20]. In [21] an overview is given of various network models that can give rise to oscillatory behavior.

In [22] a nonoscillatory model is introduced and correlations between rate coded neurons are studied. It is shown that correlations are strongest for neurons firing neither too fast nor too slow. As a result, correlation based couplings depend on the mean firing activities of the two neurons involved, and thus provide in principle a mechanism for feature binding. This property will also emerge in the present paper, but in the context of binary neurons instead of rate coding. The issue of how the stimulus affects the correlations is not explored in [22].

The second question is how synchrony can play a functional role for scene segmentation when various objects are present. An attractive model for representing various objects in a visual scene in a translationally invariant manner was proposed by [23]. The translational invariance is achieved by learning strong lateral connections encoding rigid relations between object features all over the retinal image. As a result, several *orbit assemblies* are activated for each object, which are detected by individual neurons in a separated layer. An additional set of lateral couplings between these neurons is defined. The result is, more or less, that excitatory connections develop between neurons that both participate in the same object and inhibitory connections between neurons that participate exclusively in different objects. By assuming an oscillatory neuron model, segmentation of the image in a number of objects is achieved in the temporal domain. This model was given a solid computational basis and was analyzed theoretically in [24,25].

In this paper we propose correlations that arise in networks of stochastic binary neurons as a mechanism to account for both feature linking and segmentation. Stochastic networks provide an attractive model for several reasons. Assuming detailed balance, the stochastic dynamics of these networks leads asymptotically to the Boltzmann-Gibbs distribution. Therefore the effect of stimulus-dependent correlations can be analyzed in equilibrium in the mean field framework and the linear response theory. Such analysis is more complicated or not possible for oscillatory models. This approach was first done in [26], where (time-delayed) correlations were studied in networks composed of several subpopulations of stochastic binary neurons. The issue of how the correlations depend on the stimulus was not addressed there.

Another advantage of the equilibrium formulation is that it offers an immediate solution to learning based on correlated activity using the Boltzmann machine learning paradigm [27] which has a clear information-theoretic basis. Learning in more complex networks involving various types of inhibition, causing competition in subnetworks, can be achieved using the approach outlined in [28].

A third advantage of the proposed approach is that higher order statistics may also play an important functional role in artificial networks. The experimentally observed stimulus-dependent (two point) correlations are only the simplest example. The proposed Boltzmann machine neural network is the simplest artificial system to study these phenomena.

Last, but not least, models based on oscillations tend to oscillate all the time. Setting up the dynamics such that oscillations arise under some conditions and not under others is in general difficult. Therefore it is difficult to obtain feature linking in these models. This problem was partly overcome in [18]. On the other hand, to obtain stimulus-dependent correlations in stochastic models is quite straightforward, as we will see.

The proposed mean field treatment is different to what is usually done in attractor neural networks [29,30]. Those analyses are typically applied to networks for which in the large N limit the mean field predictions become exact (for example, fully connected networks). Therefore no nontrivial correlations exist in these networks: $\langle s_1 s_2 \dots s_k \rangle = m_1 m_2 \dots m_k$, with m_i the mean field activity. To obtain nontrivial correlations, one must therefore necessarily look at models where the mean field prediction is only approximately correct. This is generally the case in models where the number of connections per neuron does not grow proportional to the system size as well as in models with multimodal equilibrium distributions [26]. As an example we consider here the simplest case of a two-dimensional Ising model.

The main result of this paper is to show how a network of binary neurons can display stimulus-dependent feature linking: correlations between neurons are a sensitive function of the *spatial coherence* of the stimulus, without altering the synaptic connections between the neurons. We restrict our analysis to objects that can be defined simply in terms of the *amount* of local supportive evidence in a compact region of the stimulus space. Examples of such objects are lines, bars, or patches of constant texture: they involve only neurons that are sensitive to the same, or similar, feature values. A spatially incoherent object has by definition a large variability in features. A spatially coherent object has a clear dominance of one feature value. We will show how this behavior of feature linking can be computed analytically. In addition, we will briefly sketch how this mechanism can also account for segmentation of objects in a scene.

In Sec. II we introduce the basic model of stochastic neuron dynamics and its relation to spiking neurons. In Sec. III we introduce an abstract model for the visual cortex consisting of a two-dimensional grid of hypercolumns. Assuming nearest neighbor interaction between neurons that code for identical feature values and absence of interactions between different feature values, the model factorizes as a product of Ising models. In Sec. IV A we consider the case of a stimulus that consists of a number of spatially coherent patches of constant stimulus value. The model reduces to a simple two-dimensional Ising model with constant external field. We review how the mean firing rate and the correlations can be computed as a function of the stimulus intensity and the lateral coupling, using mean field theory and linear response theory. We discuss how these results apply to feature linking when the image consists of several objects. In Sec. IV B we obtain our main result on dynamic feature linking showing how the spatial coherence of an object, i.e., the amount of local evidence in support of a spatially constant feature value, affects the correlations between neurons. We perform a perturbation expansion around the coherent solution of Sec. IV A. Our analytical and simulation results show the dependence of the mean firing rate and the correlations on the spatial coherence in the stimulus. In the discussion, we will briefly address the issue of segmentation and outline how correlations can segment images consisting of several previously learned objects. We plan to make full treatment of this topic the subject of a future paper.

II. STOCHASTIC NEURON DYNAMICS

In this section we introduce our basic model. We use binary neurons, which can be in two states $s_i = \pm 1$. In order to arrive at an equilibrium description, we use so-called sequential dynamics (sequential dynamics is not strictly necessary for an equilibrium formulation, see, for instance, [31,32]). Neurons are randomly selected one at a time at discrete time steps. The probability of firing for neuron i, given the current state of the network $\vec{s}$, is

$$T(s_i' = 1 | \vec{s}) = \tfrac{1}{2}[1 + \tanh(\beta l_i)], \qquad (2.1)$$

where $l_i = \sum_{j=1}^{n} w_{ij} s_j + h_i$ (h_i denotes a threshold or external field contribution for neuron i). After long times, the probability to observe the network in a state $\vec{s}$ becomes independent of time. When the weights of the network are chosen symmetrically, this time-independent equilibrium distribution is the Boltzmann distribution and is given by

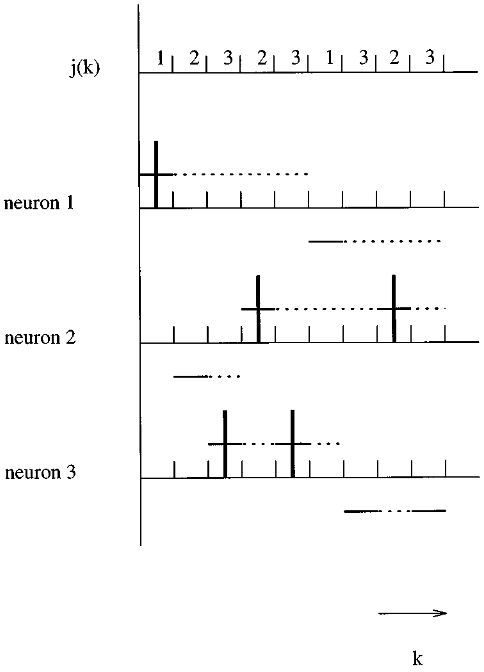

FIG. 1. Spike interpretation for network of stochastic binary neurons for the simple case of $n=3$. Time for update of the neuron states is discretized as $t=k\tau_0$, $k=1,\ldots$. Top line: For each k one neuron $j(k)$ is chosen at random. Bottom three lines: Neuron $j(k)$ is updated using Glauber dynamics (solid horizontal lines). The state s_i of each neuron remains unchanged when other neurons are updated (dashed lines). Spikes are emitted when the neuron update *and* the new state is $s_i=1$ (vertical solid lines).

$$p(\vec{s})=\frac{1}{Z}\exp\{-\beta E\}, \tag{2.2}$$

with

$$E=-\frac{1}{2}\sum_{i,j} w_{ij}s_is_j-\sum_i h_is_i$$

and

$$Z=\sum_{\vec{s}} \exp\{-\beta E\}.$$

Note that the form of Eqs. (2.1) and (2.2) allows us to assume $\beta=1$ without loss of generality.

Spike interpretation

In order to study synchronous firing we need a spike interpretation of the binary neurons. Updating occurs one neuron at a time at discrete time steps $k\tau_0$, $k=1,\ldots$ as shown in Fig. 1. Let the neuron that is updated at iteration k be denoted by $j(k)$. Let $y_i(k)=1,0$ denote whether or not neuron i spikes at iteration k. Thus $y_i(k)=1\Leftrightarrow[s_i(k)=1\wedge j(k)=i]$.

For large networks, each neuron is updated approximately every $n\tau_0$ seconds, with n the number of neurons in the network. If we choose $n\tau_0=\tau$, with τ fixed of the order of the refractory period of the neuron, every neuron is updated approximately every refractory period. For large n, the average number of spikes emitted between t and $t+\tau$ is given by $\Sigma_{k=1}^{n}\langle y_i(k)\rangle=(1/n)\Sigma_{k=1}^{n}\frac{1}{2}[s_i(k)+1]\approx\frac{1}{2}[s_i(t)+1]$. In the last step, we have made the assumption that the probability of firing is approximately constant on the fast time scale τ. The average $\langle\ \rangle$ is over possible random choices of $j(k)$ only and not over ensembles of networks as is done in Eq. (2.2). Thus we can interpret $s_i(t)=\pm1$ as ''one or no spike emitted in the interval $[t,t+\tau]$,'' respectively. By construction, no more than one spike can be emitted in this time interval when τ is chosen as the refractory period.

Therefore in terms of spikes the dynamical rule Eq. (2.1) becomes that the neuron integrates all incoming signals with zero time delay over a time τ and each incoming spike gives a contribution w_{ij} to the postsynaptic potential. This spike interpretation is consistent in the sense that first translating a spin state $\vec{s}(t)$ to a spike state and then performing spike dynamics yields the same result as first performing spin dynamics, Eq. (2.1), and then translating a spin state in a spike state.

III. ARCHITECTURE

Experimental findings indicate that neurons in the visual cortex that encode similar features have a larger probability of being connected than neurons that encode dissimilar features. In addition, these connections are short range and the probability to find a connection decays with distance. (See [33] for orientation selectivity, [34] for color selectivity.) Neurons that encode for different features are presumed to be less connected. Here we will take a simplified approach and assume (1) that features can take a discrete number of values $\alpha=1,\ldots,m$, (2) that neurons encoding for different feature values are not connected, and (3) neurons encoding for the same feature value at neighboring retinal positions are connected with excitatory symmetric connections w. Thus the model becomes a product of independent Ising models, one for each feature value α.

The equilibrium distribution of the feature detecting neurons s in feature layer α, given a stimulus x, is given by

$$p_\alpha(s|x)=\frac{1}{Z_\alpha(x)}\exp\left(\frac{1}{2}\sum_{i,j} w_{ij}s_is_j+\sum_i h_{i,\alpha}(x)s_i\right). \tag{3.1}$$

$s_i=\pm1$, $i=1,\ldots,n$ denote the firing of the neuron with feature preference α at grid location i. w_{ij} is the connectivity matrix, which is w between nearest neighbors in the grid and zero otherwise.

x denotes the external stimulus, i.e., it consists of a two-dimensional array of pixel values. $h_{i,\alpha}(x)$ describes the stimulus dependence of the neuron with feature preference α at grid location i on the stimulus x. It is well known that nearby neurons in the cortex have overlapping receptive fields. As a result, the sensory activity reaching nearby neurons can generally not be varied independently. However, here we choose to ignore this fact and assume that the stimulus at each grid location can be varied independently, $x=x_1,\ldots,x_n$, and $h_{i,\alpha}(x)=h_\alpha(x_i)$.

Although sensory neurons have a preferred stimulus, this preference is usually not very specific (coarse coding). That

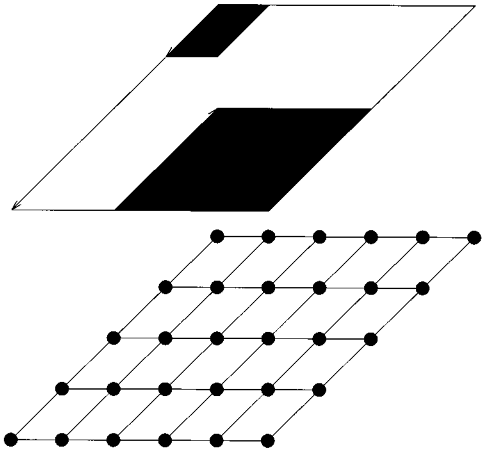

FIG. 2. In the simple Ising model, connections are only between nearest neighbors with identical feature value, which implies that objects are ''patches'' of constant feature value. Stimulus values in the stimulus layer only affect neurons at the same location in the feature layer(s). In regions where the stimulus value $x_i=\alpha$ (dark areas) the local field contribution to neuron s_i in layer α is h_+. In the remaining regions $x_i\neq\alpha$ (light areas) and the local field contribution to neuron s_i in layer α is h_-.

is, neurons in layer α can have graded responses depending on the amount of overlap with the stimulus. In our model we will ignore coarse coding. We assume that the stimulus x_i is either compatible with feature α, and $h_\alpha(x_i)=h_+$ or x_i is incompatible with feature α, and $h_\alpha(x_i)=h_-$. In the rest of the paper, we will analyze only layer α and drop the index α. For this layer, only the presence or absence of feature value α at location i is relevant. Therefore we will redefine $x_i=\pm 1$ to indicate the presence or absence of feature α at location i, i.e., $h_\alpha(x_i)=\frac{1}{2}(1+x_i)h_+ + \frac{1}{12}(1-x_i)h_-$. h_- can be interpreted as the neural threshold and h_+ as the sum of the external stimulus and the neuron threshold.

IV. STIMULUS-DEPENDENT CORRELATIONS

Consider a visual stimulus that may contain various objects. It is a basic assumption of the present study that objects are detected through the cooperative effect of the external input and the lateral excitation or inhibition. Thus objects are ''encoded'' in the lateral connectivity structure of the network in the sense that if the stimulus is ''sufficiently similar'' to the lateral structure the neurons involved in the structure will fire synchronously.

In the simple Ising model as introduced in the preceding section, connections are only between nearest neighbors with identical feature value, which implies that objects are ''patches'' of constant feature value, as shown in Fig. 2. A *coherent* object is therefore a patch of constant features. Incoherence arises when a subset of the stimulus elicits other feature responses. The coherence is a spatial property of the stimulus and measures the amount of local evidence in favor of the hypothesis ''patch of feature value α is here.'' A family of stimuli is considered, such that $p(x_i\pm 1)=p_\pm$. Thus $p_+=\frac{1}{2}$ corresponds to a fully incoherent stimulus and $p_+=1$ corresponds to a fully coherent stimulus.

In this section we will study how the synchrony depends on the parameters in the network, w, h_+, and h_-, and on the coherence of the stimulus. We first consider in Sec. IV A a fully coherent stimulus and analyze the correlations as a function of the lateral coupling and the stimulus strength. From this analysis we will find under which conditions a visual stimulus composed of constant patches will display correlated firing within each patch and uncorrelated firing between patches.

Subsequently, in Sec. IV B we will analyze how the correlations within one patch depend on the coherence in the stimulus. We will see that correlations gradually disappear when the incoherence increases.

A. Correlated firing in assemblies

We can perform a mean field computation of the mean firing rate in each of the patches. In addition, we can compute the correlations as well, making use of the linear response theorem.

The energy of the system is given, in accordance with Eq. (3.1), by

$$-E=\sum_i s_i h_i(\vec{x})+\frac{1}{2}\sum_{i,j} w_{ij}s_i s_j .$$

Consider the mean field (MF) energy

$$-E_{\mathrm{MF}}=\sum_i s_i\{h_i(\vec{x})+H_i\}, \tag{4.1}$$

where we have introduced n mean fields H_i that approximate the lateral interactions. Define the mean field partition function

$$Z_{\mathrm{MF}}=\sum_{\vec{s}} \exp(-E_{\mathrm{MF}})=\Pi_i 2\cosh(h_i+H_i).$$

The partition function can be computed in the mean field approximation [35]:

$$\begin{aligned} Z&=\sum_{\vec{s}} \exp(-E)=\sum_{\vec{s}} \exp(-E_{\mathrm{MF}}+E_{\mathrm{MF}}-E)\\ &=Z_{\mathrm{MF}}\langle\exp(E_{\mathrm{MF}}-E)\rangle_{\mathrm{MF}}\\ &\approx Z_{\mathrm{MF}}\exp(\langle E_{\mathrm{MF}}-E\rangle)=Z'. \end{aligned}\tag{4.2}$$

The mean field approximation is in the last step and is related to the convexity of the exponential function $\langle \exp f\rangle \geqslant \exp\langle f\rangle$. $\langle\ \rangle_{\mathrm{MF}}$ denotes expectation with respect to the MF distribution:

$$\langle f\rangle_{\mathrm{MF}}=\frac{1}{Z_{\mathrm{MF}}}\sum_{\vec{s}} f(\vec{s})\exp(-E_{\mathrm{MF}}). \tag{4.3}$$

From Eq. (4.3) we obtain $\langle s_i\rangle_{\mathrm{MF}}=\tanh(h_i+H_i)=m_i$ and $\langle s_i s_j\rangle_{\mathrm{MF}}=m_i m_j$, where we have introduced the mean field magnetization m_i. Thus we obtain the mean field free energy

$$-F=\ln Z'=\sum_i \ln[2\cosh(h_i+H_i)]-\sum_i H_i m_i + \frac{1}{2}\sum_{i,j} w_{ij}m_i m_j . \tag{4.4}$$

The mean fields H_i are given by minimizing the free energy:

$$\frac{\partial F}{\partial H_i}=(1-m_i^2)\left(H_i-\sum_j w_{ij}m_j\right) \tag{4.5}$$

or

$$m_i=\tanh(h_i+H_i)=\tanh\left(\sum_j w_{ij}m_j+h_i\right). \tag{4.6}$$

We can go beyond the mean field prediction $\langle s_i s_j\rangle_{\mathrm{MF}}=m_i m_j$ in the following way. First observe that true correlation is

$$\langle s_i s_j\rangle=\frac{1}{Z}\frac{d^2 Z}{dh_i dh_j}\approx\frac{1}{Z'}\frac{d^2 Z'}{dh_i dh_j}.$$

When we now make use of Eq. (4.4), we must be aware that the mean fields H_i depend on the external fields h_i through Eq. (4.6). Therefore, using the approximate free energy of Eq. (4.4),

$$\frac{d}{dh_i}\ln Z'=\left(\frac{\partial}{\partial h_i}+\sum_j \frac{\partial H_j}{\partial h_i}\frac{\partial}{\partial H_j}\right)\ln Z'=m_i .$$

In the last step we have used Eq. (4.6), by which all contributions proportional to $\partial H_j/\partial h_i$ vanish. Thus

$$\langle s_i s_j\rangle\approx\frac{1}{Z'}\frac{d}{dh_j}(Z' m_i)=m_i m_j+\frac{dm_i}{dh_j}. \tag{4.7}$$

Equation (4.7) is known as the linear response theorem and describes how spins correlate around the mean field solution $\langle s_i s_j\rangle_{\mathrm{MF}}=m_i m_j$.

By differentiating Eq. (4.6) we derive that

$$\sum_j \left(\frac{\delta_{ij}}{1-m_i^2}-w_{ij}\right)dm_j=dh_i .$$

Thus

$$\langle s_i s_j\rangle-\langle s_i\rangle\langle s_j\rangle=\frac{dm_i}{dh_j}=A_{ij}, \tag{4.8}$$

with $A_{ij}^{-1}=\delta_{ij}/(1-m_i^2)-w_{ij}$.

The matrix A^{-1} is well known and controls the linear stability of mean field solutions as a function of the coupling. Negative eigenvalues of A^{-1} indicate bifurcation to broken solutions with $\vec{m}\neq 0$. In [36–39], such a bifurcation analysis is performed for a large class of neural networks. In the present work we restrict our attention to stable solutions and use A to investigate the dependence of the correlations as defined in Eq. (4.8) on the stimulus coherence.

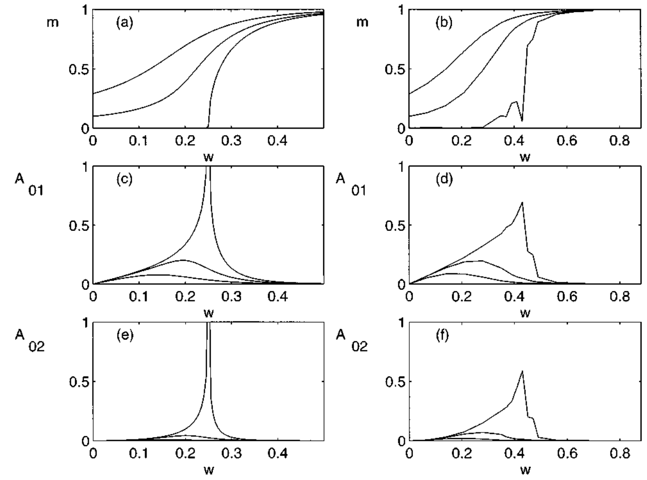

FIG. 3. Average neuron activity and correlations for coherent stimulus ($x_i=1$ for all i) as a function of lateral coupling for various values of stimulus strength $h_+=0$ (solid), $h_+=0.1$ (dashed), and $h_+=0.3$ (dotted). (a) and (b) Average neuron activity m versus coupling w. (c) and (d) Nearest neighbor correlations A_{01} versus coupling w. (e) and (f) Next-nearest neighbor correlations A_{02} versus coupling w. (a), (c), and (e) are results of the mean field computation. (b), (d), and (f) are simulations. The simulations are obtained with a grid of 10×10 neurons with periodic boundary conditions. Results are computed by temporal averaging over 5000 updates per neuron. Errors in all quantities due to spatial averaging are less than 0.05.

When $m_i=m$ independent of i, $A=A^0$ can be computed using the Fourier transform. For the cubic two-dimensional Ising lattice we find

$$A_{kl}^0=\frac{1}{(2\pi)^2}\int d\vec{p}\; G\left(\vec{p},\frac{1}{1-m^2}\right)\exp[i(\vec{k}-\vec{l})\cdot\vec{p}], \tag{4.9}$$

with $G(\vec{p},y)=[y-2w(\cos p_1+\cos p_2)]^{-1}$ and $\int d\vec{p}=\int_{-\pi}^{\pi}dp_1\int_{-\pi}^{\pi}dp_2$. $\vec{k},\vec{l}$ denote the two-dimensional coordinate vectors for the location of neuron k,l in the grid, respectively. The result Eq. (4.9) is a straightforward generalization of results by [40], obtained for $h=m=0$. Equation (4.9) can be numerically integrated, using standard methods.

In Fig. 3 we show the mean firing rates and the correlations as a function of the lateral coupling strength w for various values of the stimulus h. The left-hand figures are the theoretical predictions from the mean field computation, Eq. (4.6), and from the linear response function, Eq. (4.9). The right-hand figures are the corresponding numerical simulations. It is well known that the critical coupling $w_c=0.44$ is incorrectly predicted by the mean field computation $w_{c,\mathrm{MF}}=0.25$. Nevertheless, the mean field computation qualitatively reproduces the main characteristics that are found in the simulations. Sizable correlations for nearest neighbors are found for small h and $w<w_c$. Long-range correlations (next-nearest neighbor and more) require $h\approx 0$ and $w\approx w_c$. We are mainly interested in the correlations at distance 1, because experimental findings indicate that significant correlations fall off within several mm [41]. Ana-

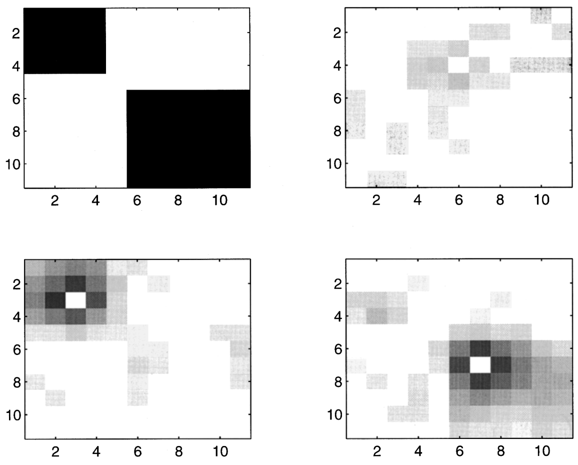

FIG. 4. Top left: Sensory input to layer α is present in the two black areas ($h=h_+=0$) and absent elsewhere ($h=h_-=-4$), $w=0.4$. Top right: Correlation A_{ij} with i the neuron located at lattice site (6,4). White (black) encodes $\langle s_i s_j\rangle-\langle s_i\rangle\langle s_j\rangle=0,1$, respectively. Bottom left: Correlation with point (3,3). Bottom right: Correlation with point (7,7).

tomical studies show that the probability of direct synaptic connections is high when neurons are separated by this order of distance.

We can apply the above analysis in each of the patches of constant stimulus. By choosing $w\approx w_c$, $h_+=0$, and $h_-<0$ we assure that (1) in regions of the network that receive coherent input α, correlations establish and neurons fire at approximately half their maximum firing rate and (2) in the remaining regions the (α sensitive) neurons are more or less quiescent. Simulations in a network consisting of an 11×11 grid of neurons with open boundary conditions are shown in Fig. 4.

As is clear from the figure, all cells belonging to a coherently stimulated part of the stimulus are highly correlated, whereas cells belonging to different regions (same or different α) are not correlated.

B. Coherence-dependent correlations

In this section we will study how correlations depend on the coherence in the stimulus. A family of stimuli is considered, such that $p(x_i \pm 1)=p_\pm$.

For a fixed stimulus, the network can be divided into two populations of neurons, those that are stimulated by feature α with local field h_+ ($x_i=1$) and the remaining neurons with local field h_- ($x_i=-1$). We introduce two mean fields $H_{+,-}$ which approximate the average contribution from the lateral interactions in the + and − populations, respectively. Thus the mean fields in Eq. (4.1) become $H_i=\frac{1}{2}(1+x_i)H_+ + \frac{1}{2}(1-x_i)H_-$. In terms of the average quantities $H_\pm$ and $h_\pm$ the free energy Eq. (4.4) becomes

$$\langle F\rangle_x = -p_+\ln[2\cosh(h_+ + H_+)] - p_-\ln[2\cosh(h_- + H_-)] - \frac{\nu w}{2}(p_-^2 m_-^2 + p_+^2 m_+^2 + 2p_+p_-m_+m_-) + p_+H_+m_+ + p_-H_-m_-, \tag{4.10}$$

where we have introduced the mean field magnetizations $m_\pm$ for neurons coupling to the stimulus $h_\pm$, respectively. $\langle\ \rangle_x$ denotes spatial averaging $\langle y\rangle_x=(1/n)\Sigma_i y_i = p_+y_+ + p_-y_-$ for some quantity y. ν denotes the number of neighbors of each neuron [$\nu=4$ for the two-dimensional (2D) Ising model].

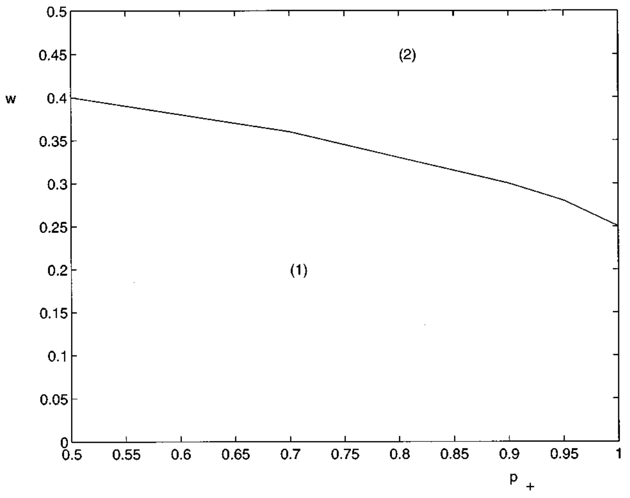

FIG. 5. Phase plot as a function of lateral coupling w and stimulus coherence p_+. $h_+=0$ and $h_-=-0.5$.

The mean fields $H_\pm$ are determined by extremizing the free energy, giving $H_+=H_-=H$, with

$$H=\nu w(p_+m_+ + p_-m_-), \quad m_\pm = \tanh(h_\pm + H). \tag{4.11}$$

Thus in this approximation the lateral contributions to the mean firing rates are identical ($H_+=H_-=H$) in the two populations. The coupled system of Eq. (4.11) can be solved using standard fixed point iteration. The phase plot is given for w and p_+ for the choice of stimulus strength $h_+=0$ and $h_-=-0.5$ in Fig. 5. First note that for fully coherent stimulus ($p_+=1$) the critical coupling is $w=0.25$, as mentioned before. For incoherent stimuli also a critical coupling exists which increases with increasing incoherence. In phases 1 and 2, the network response is ''data dominated'' and ''prior dominated,'' respectively. In phase 1 the neural activity is more determined by the contribution from the stimulus than by the contribution from the lateral coupling and in phase 2 vice versa. In phase 1, $H\approx -\nu w$, except on the line $p_+=1$ where $H=0$. In phase 2, $H\approx \pm\nu w$.

When the stimulus is incoherent, i.e., it takes different values at different sites in the network, the neural activity $m_i=m_\pm$ [Eq. (4.11)] is also site dependent. The site dependence breaks the translational invariance in the network and the Fourier transformation, used to arrive at Eq. (4.9), can no longer be applied. We can, however, perform a perturbation expansion in $\epsilon_i=1/(1-m_i^2)-1/(1-m^2)$ around the translationally invariant solution:

$$A=(A_0^{-1}+\epsilon)^{-1}=A^0[1-\epsilon A_0+(\epsilon A_0)^2+\cdots],$$

where A_0 is the matrix given by Eq. (4.9) and ϵ is a diagonal matrix. m is the value of the constant neural activity around which we perturb, whose numerical value will be fixed later.

The first order correction is given by

$$\begin{aligned}\delta A^{(1)}_{kl} &= -\sum_j A^0_{kj}\epsilon_j A^0_{jl}\\ &= -\frac{\langle\epsilon\rangle_x}{(2\pi)^2}\int d\vec{p}\ G\left(\vec{p},\frac{1}{1-m^2}\right)^2 \exp[i(\vec{k}-\vec{l})\cdot\vec{p}].\end{aligned} \tag{4.12}$$

The second order correction is given by

$$\begin{aligned}\delta A^{(2)}_{kl} &= \sum_{ij} A^0_{ki}\epsilon_i A^0_{ij}\epsilon_j A^0_{jl}\\ &= \frac{\langle\epsilon^2\rangle_x}{(2\pi)^4}\int d\vec{p}_1\int d\vec{p}_2 G\left(\vec{p}_1,\frac{1}{1-m^2}\right)^2 G\left(\vec{p}_2,\frac{1}{1-m^2}\right)\\ &\quad\times\exp[i(\vec{k}-\vec{l})\cdot\vec{p}_1] + \frac{\langle\epsilon\rangle_x^2}{(2\pi)^2}\int d\vec{p}\ G\left(\vec{p},\frac{1}{1-m^2}\right)^3\\ &\quad\times\exp[i(\vec{k}-\vec{l})\cdot\vec{p}].\end{aligned} \tag{4.13}$$

In arriving at Eqs. (4.12) and (4.13) we have used that $\Sigma_k y_k \exp(i\vec{k}\cdot\vec{p}) \approx (2\pi)^2\langle y\rangle_x \delta(\vec{p})$ for $y_k = \epsilon_k, \epsilon_k^2$, respectively.

In this perturbation expansion, we have the freedom to choose the homogeneous solution m around which we expand. We chose m such that $\langle\epsilon\rangle_x = 0$, which yields $1/(1-m^2) = \langle 1/(1-m^2)\rangle_x$ and which minimizes $\langle\epsilon^2\rangle_x = p_+ p_-[1/(1-m_+^2) - 1/(1-m_-^2)]^2$.

Finally, we obtain

$$\begin{aligned}A_{kl} &= \frac{1}{(2\pi)^2}\int d\vec{p}\ G\left(\vec{p},\left\langle\frac{1}{1-m^2}\right\rangle_x - \langle\epsilon^2\rangle_x C\right)\\ &\quad\times\exp[i(\vec{k}-\vec{l})\cdot\vec{p}] + O(\epsilon^3),\end{aligned} \tag{4.14}$$

with

$$C = \frac{1}{(2\pi)^2}\int d\vec{p}\ G\left(\vec{p},\left\langle\frac{1}{1-m^2}\right\rangle_x\right). \tag{4.15}$$

We are now able to compute the effect of stimulus coherence on the correlations between stimulated neurons. We chose the lateral coupling $w = 0.35$ in our simulations to be close to the critical coupling but not too close to avoid problems with mixing of phases. For each coherence, we compute the mean firing rates from Eq. (4.11). Subsequently, we compute the correlations from Eqs. (4.14) and (4.15). The results are given in Fig. 6.

The results from our analytical computation are in qualitative agreement with the simulations. In Figs. 6(a) and 6(b) we see a monotone increase of the correlations between pairs of stimulated neighboring neurons with the coherence in the stimulus. In addition, we see that also the average firing of these neurons is strongly dependent on the coherence. Thus for incoherent stimuli we observe low incoherent firing rates and for coherent stimuli we observe a correlated firing at $\frac{1}{2}$ their maximal firing rate $1/\tau$.

We observe that the relation between coherence and correlations is strongly influenced by the strength of the stimulus h_+. h_+ should be close to zero, which means that the external stimulus and the neuron threshold should have similar values. Deviations from this assumption are shown in Figs. 6(c) and 6(d) and Figs. 6(e) and 6(f), respectively. For $h_+ > 0$ a fully coherent stimulus leads to too high mean firing rates, which reduces the correlations [see Eq. (4.8)]. In this case intermediate coherence leads to maximal correlations. For $h_- < 0$ for no stimulus there are sufficiently high firing rates to produce strong correlations.

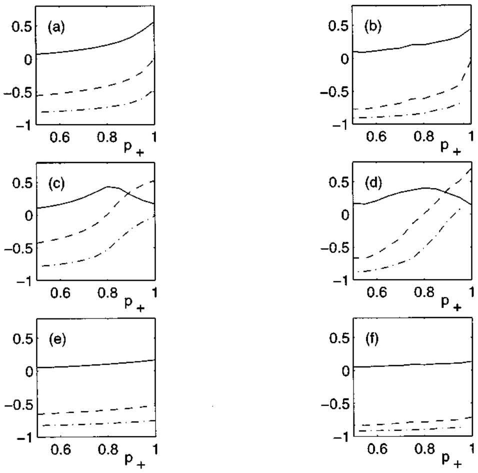

FIG. 6. Correlations A_{01} (solid line), m_+ (dashed line), and m_- (dash-dotted line) as a function of stimulus coherence p_+. Left-hand figures are analytical results with $w = 0.23$. Right-hand results are simulations with $w = 0.35$ in a 10×10 grid with periodic boundary conditions. Results are computed by temporal averaging over 5000 updates per neuron. Errors in all quantities due to spatial averaging are less than 0.05 (a) and (b) $h_+ = 0$ and $h_- = -0.5$. (c) and (d) $h_+ = 0.1$ and $h_- = -0.5$. (e) and (f) $h_+ = -0.1$ and $h_- = -0.5$.

In Fig. 7 we give an example of the spiking behavior of the network under various stimulus conditions.

V. DISCUSSION

A. Feature linking

We have proposed to use a network of binary spins to study the experimentally observed phenomenon of stimulus-dependent correlations in the visual cortex. As a crude approximation to model the cortex we have proposed a separate Ising model for each of a number of distinct feature values.

We have shown how the correlations depend on the strength of the stimulus, on the strength of the lateral connectivity, as well as on the coherence of the stimulus. These results were obtained using a mean field computation for the average firing rates in the stimulated and nonstimulated populations, and using a linear response calculation for the leading order correlations. These calculations were verified with numerical simulations.

We conclude that correlations between connected neurons can be present or absent depending on the coherence in the stimulus. This effect of dynamic linking is achieved without fast synaptic changes and is caused by the coherence in the stimulus only. In addition, we observe that also the mean firing rates are strongly affected by the coherence in the stimulus.

Coherence in the stimulus was controlled by varying the percentage of ‘‘on’’ stimuli, independently for each stimulus

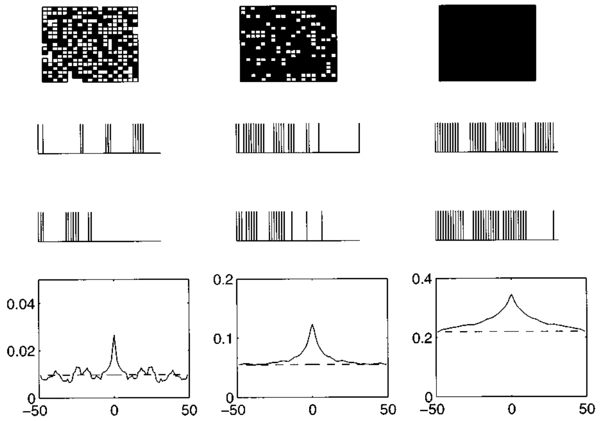

FIG. 7. Example of the spiking behavior of the network under various stimulus conditions. Top row shows three stimulus conditions with increasing coherence of feature α. Second and third rows show a short segment of the spike trains of two neighboring neurons that both receive stimulus α. The total length of the train is 50τ seconds. Bottom row shows time-delayed crosscorrellograms $\langle s_i(0)s_j(t)\rangle$ (solid line) and square mean firing rates $\langle s_i\rangle^2$ (dashed line) as a function of time difference t.

location. This gives a one parameter family of stimuli where coherence is in fact the ''luminance'' (fraction of pixels ''on''). Clearly, other families of stimuli can be chosen. For instance, in [42] the stimulus itself is modeled as an Ising model. The stimulus is now defined by two parameters, which are the lateral coupling and the external field. One can then consider the one-dimensional family of stimuli defined by varying the lateral coupling and with external field zero. Due to the lateral coupling, these stimuli have the property that for the same luminance, the coherence in the stimulus is larger than for those considered in this paper. Fully coherent stimuli and fully incoherent stimuli are the same in both approaches. One can analyze the phase diagram in the mean field approach, as was done by [42], and one can probably compute the correlations using the linear response computation, in a similar way as was done in this paper. It should be expected that the results from such an analysis will be qualitatively the same as those obtained in this paper, with the difference that one will observe increased correlations at the same luminance level, compared to the results presented in this study.

Clearly, we are not proposing the Ising model as a serious computational model for the cortex. An important restriction of the present work is that feature sensitivity of neurons has been discretized and neurons have been assumed to be only sensitive to one feature value. In addition, we assumed that only neurons that are sensitive to identical features are coupled horizontally. One should formulate models with more complex horizontal interactions, for instance, fully connected excitatory interaction within hypercolumns or inhibition within hypercolumns which leads to competition between feature detectors (Potts model). In the present model, receptive fields are nonoverlapping (spatially) and are strongly specialized. One should investigate the effects of redundancy such as spatial overlap and coarse coding on the correlations.

The analytical results obtained pertain to the equilibrium situation. To relate the correlations to functional behavior, it is important to establish at what time scales the correlations establish after onset of the stimulus. For unfrustrated systems of the type that we have studied so far, this may be analyzed within the linear response approach.

In the present work we have established the stimulus dependence of correlated firing for fixed lateral (and feedforward) connections. In a more realistic network, the lateral connectivity would arise from learning. The connections that will establish will be between those neurons that are correlated in the stimulus environment. It is interesting to note that the most straightforward learning paradigm for stochastic networks, i.e., the Boltzmann machine learning rule, is indeed based on correlated activity $\langle s_i s_j\rangle$.

B. Scene segmentation

In this paper we have shown how correlations can establish in stochastic networks, and how these correlations depend on the coherence in the stimulus ensemble. We have demonstrated how this coherence dependence can be analyzed theoretically using mean field and linear response theory.

However, the simple Ising model is quite far removed from how it is generally assumed that patterns are stored in the cortex. In addition, it is not clear how this mechanism can be used for scene segmentation. Therefore in this section we will give a heuristic argument for how the main ideas of this paper can be accommodated in a more realistic setting. We plan to provide a more thorough treatment in the future.

Consider a network of n neurons $s_i = \pm 1$, each encoding a different feature [25] (or orbit assembly [23]). Suppose that the objects are nonoverlapping, i.e., features appear uniquely in one object and not in others. Suppose the objects are represented neurally by p patterns $\xi_i^\mu = \pm 1, \mu = 1, \dots, p$. $\xi_i^\mu = \pm 1$ denotes the presence or absence of feature i in object μ. Suppose that as a result of training, positive connections w_+ develop between neurons encoding features of the same object and negative connections w_- develop between neurons encoding features of different objects. Examples of such learning rules are given in [23,25].

The energy of the system in the absence of external stimulus is given by

$$-E = \sum_i \sum_{j>i} w_{ij} s_i s_j + \theta \sum_i s_i .$$

By choosing $\theta = -w_- n(2/p - 1)$ one can easily show that the patterns ξ_i^μ are global minima of E. Thus the equilibrium distribution $p(s) = (1/Z)\exp[-\beta E(s)]$ has p peaks around the global minima. Additionally, local minima of E may give rise to small subpeaks, which we will ignore here. As a very crude approximation, therefore, we have

$$\langle s_i\rangle = \sum_s s_i p(s) \approx \frac{1}{p} \sum_\mu \xi_i^\mu = \frac{2-p}{p}$$

and

$$\langle s_i s_j\rangle - \langle s_i\rangle\langle s_j\rangle \approx \begin{cases} \dfrac{4}{p}\left(1-\dfrac{1}{p}\right) & \text{when } i,j \text{ belong to the same pattern} \\ -\dfrac{4}{p^2} & \text{when } i,j \text{ belong to different patterns.} \end{cases}$$

Thus in the absence of a stimulus all neurons fire with the same rate, but this firing is correlated depending on whether the neurons encode features belonging to the same or different objects.

Consider now that an external visual scene is presented consisting of a subset S of q objects out of the p objects ξ^μ. Now, an additional term should be added to E of the form $-\Sigma_i h_i^e s_i$, with $h_i^e = h\Sigma_{\mu\in S}\xi_i^\mu$ the external field contribution due to the subset of patterns that are present in the scene. h is a free parameter, related to the strength of the feed-forward connections between the retinal image and the present layer. The effect is that the global minimum of E will by attained by $\xi^\mu, \mu \in S$, whereas the remaining objects will become local minima, with energy $2hn/p$ higher than the minimal energy. By the same argument as above we have

$$\langle s_i\rangle \approx \begin{cases} \dfrac{2-q}{q} & \text{when } i \text{ belongs to } \mu\in S \\ -1 & \text{when } i \text{ belongs to } \mu\notin S \end{cases}$$

and

$$\langle s_i s_j\rangle - \langle s_i\rangle\langle s_j\rangle \approx \begin{cases} \dfrac{4}{q}\left(1-\dfrac{1}{q}\right) & \text{when } i \text{ and } j \text{ belong to the same } \mu\in S \\ -\dfrac{4}{q^2} & \text{when } i \text{ and } j \text{ belong to different } \mu,\nu\in S \\ \approx 0 & \text{when } i \text{ or } j \text{ belongs to } \mu\notin S. \end{cases}$$

Thus all neurons that encode features that are present in the scene fire with the same rate and all other neurons are quiescent. The firing between active neurons is correlated depending on whether the neurons encode features belonging to the same or different objects.

A comment is in order here on the validity of the approximation to replace the sum over all states by just the maxima of the probability distribution. When $\beta\rightarrow\infty$ this approximation is exact. However, in this limit, the transition times between the q different phases also become infinite, which implies that any biologically reasonable dynamics will get stuck in one of the phases. In other words, ergodicity is broken and ensemble average and time average can no longer be identified. Thus β should be chosen small enough such that the transition times between the optima are reasonably small. For lower β, the bold approximation above gets worse and worse, because also suboptimal states will contribute significantly to the sum over states. However, as was shown in [43] for continuous variables, a Gaussian approximation can summarize effectively the contribution of all states in the q optimal bases of attraction. It should be expected that these contributions do not qualitatively change the conclusions drawn above.

The difference between the mechanism for feature binding based on oscillations and the above mechanism is quite striking. The oscillatory solution to segmentation is to represent the different objects one after another in time like a periodic movie [23,25]. The solution based on correlated activity is, on the other hand, not periodic but stationary. There exists a time-independent equilibrium probability distribution and the network is given a stochastic dynamics such that over long times all states are visited with this probability. As we saw, this leads to time-independent correlations between neurons depending to which object they belong.

ACKNOWLEDGMENTS

I would like to thank Tom Heskes for useful discussions on this paper. I would like to thank Marcel Nijman for helping with some of the numerical simulations.

[1] D.H. Hubel and T.N. Wiesel, J. Neurophysiol. **26**, 994 (1963).

[2] J.J. Atick, Network **3**, 213 (1992).

[3] D. Hebb, *The Organization of Behaviour* (Wiley, New York, 1949).

[4] P.M. Milner, Psychol. Rev. **81**, 521 (1974).

[5] Chr. von der Malsburg, in *Models of Neural Networks II*, edited by E. Domany, J.L. van Hemmen, and K. Schulten, Physics of Neural Networks (Springer-Verlag, Berlin, 1994).

[6] H.J. Reitboeck, in *Synergetics of the Brain*, edited by E. Basar (Springer, Berlin, 1983), pp. 174–181.

[7] B. Julesz, *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, 1971).

[8] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1982).

[9] C.M. Gray, P. König, A.K. Engel, and W. Singer, Nature (London) **338**, 334 (1989).

[10] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H.J. Reitboeck, Biol. Cybern. **60**, 121 (1988).

[11] J.I. Nelson, P.A. Salin, M.H.J. Munck, M. Arzi, and J. Bullier, Visual Neurosci. **9**, 21 (1992).

[12] R.Chr. deCharms and M.M. Merzenich, Nature (London) **381**, 610 (1996).

[13] A.K. Engel, P. König, and W. Singer, Proc. Natl. Acad. Sci. USA **88**, 9136 (1991).

[14] A.K. Kreiter and W. Singer, J. Neurosci. **16**, 2381 (1996).

[15] Chr. von der Malsburg and W. Schneider, Biol. Cybern. **54**, 29 (1986).

[16] P. König and T.B. Schillen, Neural Comput. **3**, 155 (1991).

[17] T.B. Schillen and P. König, Neural Comput. **3**, 167 (1991).

[18] P. König, B. Janosch, and T.B. Schillen, Neural Comput. **4**, 666 (1992).

[19] T. Chawanya, T. Aoyagi, I. Nishikawa, K. Okuda, and Y. Kuramoto, Biol. Cybern. **68**, 483 (1993).

[20] M. Arndt, P. Dicke, M. Erb, R. Eckhorn, and H.J. Reitboeck, in *Neural Network Dynamics*, edited by J.G. Taylor (Springer-Verlag, Berlin, 1992), pp. 140–155.

[21] W. Gerstner, Phys. Rev. E **51**, 738 (1995).

[22] T.J. Sejnowski, Biol. Cybern. **22**, 203 (1976).

[23] H. Neven and A. Aertsen, Biol. Cybern. **67**, 309 (1992).

[24] W. Gerstner, R. Ritz, and J.L. van Hemmen, Biol. Cybern. **68**, 363 (1993).

[25] R. Ritz, W. Gerstner, U. Fuentes, and J.L. van Hemmen, Biol. Cybern. **71**, 349 (1994).

[26] I. Ginzburg and H. Sompolinsky, Phys. Rev. E **50**, 3171 (1994).

[27] D. Ackley, G. Hinton, and T. Sejnowski, Cognitive Sci. **9**, 147 (1985).

[28] H.J. Kappen, Neural Networks **8**, 537 (1995).

[29] J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[30] D.J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).

[31] P. Peretto, Biol. Cybern. **50**, 51 (1984).

[32] J.L. van Hemmen and R. Kühn, in *Models of Neural Networks* (Ref. [5]), pp. 1–106.

[33] C.D. Gilbert and T.N. Wiesel, J. Neurosci. **9**, 2432 (1989).

[34] M.S. Livingstone and D.H. Hubel, J. Neurosci. **4**, 309 (1984).

[35] C. Itzykson and J-M. Drouffe, *Statistical Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 1989).

[36] J.L. van Hemmen, D. Grensing, A. Huber, and R. Kuhn, Z. Phys. B **65**, 53 (1986).

[37] J.L. van Hemmen and R. Kuhn, Phys. Rev. Lett. **57**, 913 (1986).

[38] J.L. van Hemmen, D. Grensing, A. Huber, and R. Kuhn, J. Stat. Phys. **50**, 231 (1988).

[39] J.L. van Hemmen, D. Grensing, A. Huber, and R. Kuhn, J. Stat. Phys. **50**, 259 (1988).

[40] G. Parisi, *Statistical Field Theory*, Frontiers in Physics, Vol. 66 (Addison-Wesley, Reading, MA, 1988).

[41] A.K. Engel, P. König, C.M. Gray, and W. Singer, Eur. J. Neurosci. **2**, 588 (1990).

[42] J.M. Pryce and A.D. Bruce, J. Phys. A **28**, 511 (1995).

[43] T. Heskes, E. Slijpen, and B. Kappen, Phys. Rev. A **46**, 5221 (1992).