Topology-preservation emergence by the Hebb rule with infinitesimal short-range signals

Hans-Otto Carmesin*
Institut für Theoretische Physik, Universität Bremen, 28334 Bremen, Germany

(Received 1 March 1995; revised manuscript received 29 August 1995)

Topology preservation is a ubiquitous phenomenon in the mammalian nervous system. What are the necessary and sufficient conditions for the self-organized formation of topology preservation due to a Hebb mechanism? A relatively realistic Hebb rule and neurons with stochastic fluctuations are modeled. Moreover, the reasonable growth law is used for coupling growth that the biomass increase is proportional to the present biomass under the constraint that the biomass is limited at a neuron. It is proven for such general Hebb-type networks that infinitesimal lateral signal transfer to neighboring neurons is necessary and sufficient for the emergence of topology preservation. As a consequence, observed topology preservation in nervous systems may emerge with or without purpose as a byproduct of infinitesimal lateral signal transfer to neighboring neurons due to ubiquitous chemical and electrical leakage.

PACS number(s): 87.90.+y

I. INTRODUCTION

Since the 1930's it has been empirically established that many nervous systems exhibit several so-called *cortical maps* [1]. For instance, the sensor stimuli of the skin are transferred via axonic connections to the gyrus of the parietal cortex. These connections exhibit two nontrivial properties: they establish a mapping from locations at the skin to locations at the cortex and neighboring skin locations are preferentially mapped to neighboring cortex locations; the latter phenomenon is called topology preservation. The basic role of such topology preservation is illustrated by the fact that such topology preservation is inherent already to Descartes brain models [2,3].

The *formation* of these maps takes place in several steps during ontogeny. First the cells of the embryo migrate to their destination areas, then the axons grow according to markers, chemical markers, for instance, and finally the axonic connections change according to the neuronal activity in a self-organized manner. Such self-organization has been discussed since the 1920's [4,5]. Nowadays such self-organization is usually studied with two-layer neural networks with some explicit or implicit *lateral interaction mechanism*. However, there remained still some important open questions. For instance, it was not clear what range and what intensity such lateral interaction should have in order to stabilize topological order [5–8]. In the present paper, a neural network model is developed and solved exactly in the framework of a general neurostatistical field theory [9–11], similar to hydrodynamics, with the Navier-Stokes equations as prototye, for instance.

As a result it turns out that infinitesimal short-range signal transfer is sufficient. This implies that such topology preservation is a typical phenomenon in self-organizing neuronal systems with only slight short-range lateral signal transfer. This implies in turn that the mere observation of such topological order without further evidence should not yet be interpreted as purposeful, because such order may easily emerge as a byproduct of a neuronal self-organization process.

For comparison, two clocks with a pendulum at the same wall are infinitesimally coupled, as a result, they synchronize after a while, as discovered by Huygens in 1665 [12].

II. NETWORK MODEL

A neural network usually has two dynamical rules: the neuronal dynamics models the activity of neurons and the coupling dynamics models the change of synaptic connections [13]. In addition, a network architecture and a stimulation should be specified for an adapting neural network [9–11].

A. Network architecture

The network consists of S sensor neurons n_j, I cortical neurons $\tilde{n}_i$ and of all possible afferent couplings W_{ij} (that is from sensor neurons to cortical neurons; it turns out later that many of these couplings take the value 0). For the sake of a simple formalism, the case $Sn_0=I$ is considered; the general case of arbitrary I and S is similar.

The sensor neurons and the cortical neurons exhibit an arbitrary topology, including arbitrary dimension, as follows. The sensor neurons exhibit arbitrary neighborship relations. The neighbors of a sensor neuron n_j are formalized by the *sensor neighbor set* $\nu(j)$. Analogously, the neighbors of a cortical neuron $\tilde{n}_i$ are formalized by the *cortical neighbor set* $\tilde{\nu}(i)$. For the particular case of a one-dimensional neural network, the architecture is illustrated in Fig. 1.

B. Network dynamics

The neurons take values 0 or 1 at discrete time steps $t=1,2,3,\ldots,$ that is $n_j(t)=0$ or 1 and $\tilde{n}_i(t)=\tilde{n}_i=0$ or 1. Here and in the following, the time index is omitted if it is t. The sensor neurons are stimulated by mutually un-

*FAX: 421 218 4869. Electronic address: Carmesin@theo.physik.uni-bremen.de

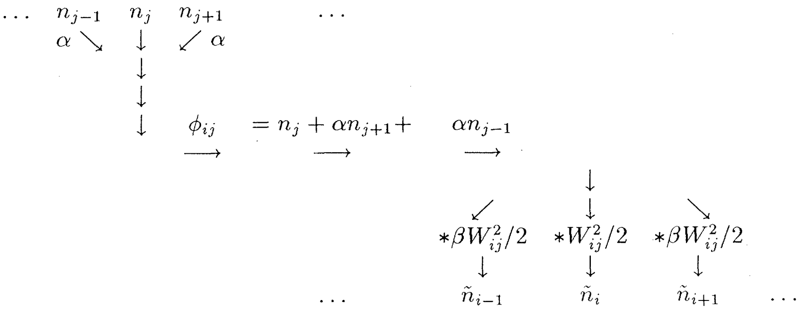
... n_{j-1} n_j n_{j+1} ...
α
α
ϕ_{ij}
$= n_j + \alpha n_{j+1} +$
αn_{j-1}
$*\beta W_{ij}^2/2$
$*W_{ij}^2/2$
$*\beta W_{ij}^2/2$
... $\tilde{n}_{i-1}$ $\tilde{n}_i$ $\tilde{n}_{i+1}$ ...

FIG. 1. Network architecture. Illustration for the special case of the one-dimensional model. n_j, input neuron; $\tilde{n}_i$, cortical map neuron; ϕ_{ij}, axonic membrane potential; α, presynaptic lateral contribution parameter; β, postsynaptic lateral contribution parameter; W_{ij}, coupling.

correlated equally distributed random stimuli, these are expressed by $n_j(t)=0$ or 1 with equal probability.

1. *Signal transfer*

A presynaptic neuron n_j transfers a signal to a postsynaptic neuron $\tilde{n}_i$ via a coupling W_{ij}; in addition, this coupling transfers signals also to and from neighboring neurons as indicated in Fig. 1. This is formalized as follows. The axonic membrane potential ϕ_{ij} that gives rise to the signal transfer from n_j to $\tilde{n}_i$ is established additively by the values n_j of the presynaptic neuron and the *presynaptic lateral contribution parameter* α times n_m of the neighboring neurons

$$\phi_{ij} = n_j + \alpha \sum_{n_m \in \nu(j)} n_m \ . \tag{1}$$

This membrane potential ϕ_{ij} contributes a *stimulating local field* $h_{ij,k}$ to the kth cortical neuron $\tilde{n}_k$ as follows. The membrane potential ϕ_{ij} gives rise to a stimulating local field $h_{ij,i}(t+1)$ at the next time step $t=1$ at the postsynaptic neurons $\tilde{n}_k=\tilde{n}_i$ proportional to the positive term $W_{ij}^2/2$ and it gives rise to a stimulating local field $h_{ij,k}(t+1)$ at the neighbors $\tilde{n}_k$ of the postsynaptic neuron $\tilde{n}_i$ proportional to the positive term $W_{ij}^2/2$ times the *postsynaptic lateral contribution parameter* β (see Fig. 1). That is

$$\begin{aligned} h_{ij,i}(t+1) &= \phi_{ij}\frac{W_{ij}^2}{2} \ , \\ h_{ij,k}(t+1) &= \phi_{ij}\frac{W_{ij}^2}{2}\beta \quad \text{for } \tilde{n}_k \in \tilde{\nu}(i) \ , \\ h_{ij,k}(t+1) &= 0 \quad \text{for } k \neq i, \text{ and not } \tilde{n}_k \in \tilde{\nu}(i) \ . \end{aligned} \tag{2}$$

One might be surprised that a square of the couplings occurs here. This is convenient. Moreover it is not essential, because the theory can as well be developed with the transformation $K_{ij}=W_{ij}^2$, it is clear from theoretical reasons and it has been shown explicitly in similar theories that this transformation has no effect; see Eq. (7) below or [8–10].

As a consequence, such contributions from all sensor neurons n_j add up at a cortical neuron $\tilde{n}_i$ as follows

$$\tilde{h}_i(t+1) = \sum_j^N \left[\tilde{h}_{ij,i}(t+1) + \sum_{\tilde{n}_k \in \tilde{\nu}(i)} \tilde{h}_{kj,i}(t+1) \right] \ . \tag{3}$$

By inserting Eq. (2) one gets

$$\tilde{h}_i(t+1) = \sum_j^N \left[\phi_{ij}\frac{W_{ij}^2}{2} + \beta \sum_{\tilde{n}_k \in \tilde{\nu}(i)} \phi_{kj}\frac{W_{kj}^2}{2} \right] \ . \tag{4}$$

2. *Stochastic neuronal dynamics*

The cortical neurons $\tilde{n}_i$ prefer to fire according to the stimulating local field $\tilde{h}_i$, however there is the possibility that the cortical neurons fire differently due to random fluctuations. This is formalized by the Boltzmann probability with a *fluctuation parameter* T as follows:

$$P(\tilde{n}_i) = \frac{\exp[\tilde{h}_i \tilde{n}_i / T]}{1+\exp[\tilde{h}_i/T]} \ . \tag{5}$$

3. *Coupling dynamics*

A coupling weight W_{ij} is increased, if the presynaptic membrane potential ϕ_{ij} and the postsynaptic firing stimulated by that membrane potential are in accordance. This is modeled as follows:

$$\Delta W_{ij}^{\text{Hebb}} = a W_{ij} \left[n_j + \alpha \sum_{n_m \in \nu(j)} n_m \right] \times \left[\tilde{n}_i(t+1) + \beta \sum_{\tilde{n}_k \in \tilde{\nu}(i)} \tilde{n}_k(t+1) \right] \ . \tag{6}$$

Hereby, the coupling change $\Delta W_{ij}^{\text{Hebb}}$ is proportional to a *learning parameter a*, to the present coupling W_{ij} (in typical biological growth processes the biological matter increase is proportional to the present biological matter), to the presynaptic activity $n_j + \alpha \sum_{n_m \in \nu(j)} n_m$ and to the postsynaptic activity $\tilde{n}_i(t+1) + \beta \sum_{\tilde{n}_k \in \tilde{\nu}(i)} \tilde{n}_k(t+1)$.

This ansatz is immediately motivated by the Hebb rule, due to the product of presynaptic and postsynaptic activities. This Hebb rule makes sense physiologically, because the presynaptic and postsynaptic activities give rise to local metabolic changes that induce the coupling change. Moreover, the functional form of this coupling change is quite general, because any function f of a neuronal variable n_j may be presented in terms of a power series $f(n_j) = f_0 + f_1 n_j + f_2 n_j^2 + \cdots = f_0 + f_1 n_j$, due to

$n_j^2 = n_j = 0$ or 1; here the constant f_0 does not depend on n_j and is irrelevant, while f_1 corresponds to α and β.

For the sake of transparency of the square in the formal field $\tilde{h}_i$, it is shown that the coupling growth law takes the same form for the transformed couplings $K_{ij} = W_{ij}^2$: Using the partial derivative $\partial K_{ij} / \partial W_{ij} = 2W_{ij}$, one gets $\Delta K_{ij}^{\text{Hebb}} = 2W_{ij} \Delta W_{ij}^{\text{Hebb}}$, thus

$$\Delta K_{ij}^{\text{Hebb}} = 2aK_{ij} \left[n_j + \alpha \sum_{n_m \in \nu(j)} n_m \right] \times \left[\tilde{n}_i(t+1) + \beta \sum_{\tilde{n}_k \in \tilde{\nu}(i)} \tilde{n}_k(t+1) \right] . \tag{7}$$

Moreover, the above coupling dynamics is modeled with the effective constraint that the total coupling weight at a presynaptic neuron is constant. This is formalized in terms of a Euclidean norm and a radius r in coupling space

$$\sum_i^I W_{ij}^2 = n_0 r^2 . \tag{8}$$

Analogously, the above coupling dynamics is modeled with the effective constraint that the total coupling weight at a postsynaptic neuron is constant.

$$\sum_j^S W_{ij}^2 = r^2 . \tag{9}$$

Both constraints are in agreement with the empirical observation that the connectivity is quite fixed at a neuron [14].

The above effective constraints are achieved roughly by the following additional coupling changes.

$$\Delta W_{ij}^{\text{additional sensor}} = -\frac{\partial V_j^{\text{sensor}}}{\partial W_{ij}} \quad \text{with } V_j^{\text{sensor}} c \left[\sum_i^I W_{ij}^2 - n_0 r^2 \right]^2 \quad \text{and constraint parameter } c . \tag{10}$$

$$\Delta W_{ij}^{\text{additional cortical}} = -\frac{\partial V_i^{\text{cortical}}}{\partial W_{ij}} \quad \text{with } V_i^{\text{cortical}} = c \left[\sum_j^S W_{ij}^2 - r^2 \right]^2 . \tag{11}$$

Altogether, the total coupling change is the sum

$$\Delta W_{ij} = \Delta W_{ij}^{\text{Hebb}} + \Delta W_{ij}^{\text{additional sensor}} + \Delta W_{ij}^{\text{additional cortical}} . \tag{12}$$

Each such additional coupling change depends on those couplings that are connected with the corresponding neuron. Thus such an additional dynamics models a mechanism by which the presynaptic neuron may decrease or increase the coupling weight incrementally at its end of the coupling and analogously the postsynaptic neuron may decrease or increase the coupling weight incrementally at its end of the coupling. Such a mechanism may be regarded as local at a neuron.

In addition to the identification of the empirical plausibility of the effective constraints, a further discussion of the locality of these constraints makes sense: In the framework of a computer simulation, it is clear that the performance of such a constraint requires the performance of S or I couplings at a neuron, while a completely global coupling norm would require the performance of all possible SI neurons at once.

4. *Constraint approximation*

For appropriate values of the constraint parameter c, the additional coupling changes give rise to coupling states that obey the effective constraints in Eqs. (8) and (9) roughly. The focus of the present study is topological order, so it is completely adequate to approximate the effect of the additional coupling changes by a rescaling of the coupling state $W'_{ij} = W_{ij} + \Delta W_{ij}^{\text{Hebb}}$ so that the effective constraints are roughly obeyed. Within the constraint approximation, one gets

$$W'_{ij} = [W_{ij} + \Delta W_{ij}^{\text{Hebb}}]_{\text{rescaled}} .$$

For short, the bracket is not explicated in the following, i.e.,

$$W'_{ij} = W_{ij} + \Delta W_{ij}^{\text{Hebb}} . \tag{13}$$

C. Fixed simulation potential

The network model is introduced above. An immediate property of the Hebb-type coupling dynamics [see Eq. (6)] is that it may be derived from a potential for each stimulation:

1. *Fixed stimulation potential theorem*

For a fixed stimulation by a sensor configuration $\mu = \{n_j\}$, the resulting Hebb-type coupling change $\Delta W_{ij}^{\text{Hebb}}$ is the derivative of a potential function H^μ as follows.

$$\Delta W_{ij}^{\text{Hebb}} = -a \frac{\partial H^\mu}{\partial W_{ij}} \tag{14}$$

with the formal energy function

$$H^\mu = -\sum_{i=1}^I \tilde{h}_i(t+1) \tilde{n}_i(t+1) \quad \text{for a fixed stimulation } \mu . \tag{15}$$

2. *Principle underlying the proof*

The proof relies on the symmetry of the neighborship-relation.

3. *Proof*

By inserting Eq. (4) into Eq. (15) one gets

$$\frac{\partial H^{\mu}}{\partial W_{ij}}=\frac{\partial}{\partial W_{ij}}\left\{-\sum_{km}\tilde{n}_k(t+1)\times\left[\phi_{km}\frac{W_{km}^2}{2}+\beta\sum_{\tilde{n}_q\in\tilde{\nu}(k)}\phi_{qm}\frac{W_{qm}^2}{2}\right]\right\}. \quad (16)$$

The above partial derivative yields nonzero terms only, if the indices of W_{ij} are equal to those of W_{km} or W_{qm}. Thus nonzero terms occur, when m takes the value j and when i is equal to k or $i=q=$neighbor index of k. Thus one obtains conversely $k=i$ or $k=$neighbor index of i, due to the symmetry of the neighborship relation. So one gets

$$\frac{\partial H^{\mu}}{\partial W_{ij}}=-W_{ij}\phi_{ij}\left[\tilde{n}_i(t+1)+\beta\sum_{\tilde{n}_q\in\tilde{\nu}(i)}\tilde{n}_q(t+1)\right]. \quad (17)$$

By comparing this expression with Eqs. (1) and (6) one obtains

$$\frac{\partial H^{\mu}}{\partial W_{ij}}=-\Delta W_{ij}^{\mathrm{Hebb}}/a\ . \quad (18)$$

Q.E.D.

III. FIELD-THEORETIC SOLUTION OF THE NETWORK

We give here an overview for the solution method. In order to solve the above-introduced network model, one should specify the coupling matrices that emerge as a result of the combined neuronal and coupling dynamics. This is achieved here as follows (for a very detailed description see [9–11]): First the combined dynamics is identified as taking place in the combined set of states $(\mathbf{n},\mathbf{W})$ of neurons and couplings. This set may be regarded as being embedded in a *vector space* with continuous values for neurons and couplings and with the neuronal space and coupling space as subspaces. In this state set, the combined dynamics establishes a *Marcov process,* by construction. In the rest of the paper, this Markov process is assumed to be ergodic; this is reasonable for many cortical maps with rapid coupling changes [15]; see also the discussion below.

As a consequence, the averaged changes may be described by a vector field. The fast neuronal variables may be solved first in a so-called adiabatic limit. The remaining coupling dynamics is characterized by an ordinary differential equation that may be derived from a scalar potential. The stationary states of the coupling dynamics are the local potential minima. These stationary states represent the possible emerging networks. As a consequence, the possible emerging networks may be investigated by analyzing the potential minima. As a result of that analysis one obtains precise conditions for the emergence of topologic order.

A. Field theory

1. Discussion of ergodicity

If each state of a Markov process may be taken with finite probability from any of its states, then the process is ergodic. So the emergence of a cortical map with the possibility of relatively global coupling changes may be adequately modeled with an ergodic Markov process. Such cortical map formations are quite typical in the nervous system. For instance, the couplings may change significantly when the stimulation from a part of the sensory input stops due to a lesion [15]. More specifically, cortical maps exhibit especially high plasticity during early ontogeny, in special critical periods, after lesions or during regeneration. Altogether it appears adequate to model cortical map formation in terms of an ergodic Markov process; for a particular comparison with experiments, one may adjust the learning rate a and the fluctuation rate T to the data.

From a more formal point of view, one may derive the ergodicity from an assumption of limited coupling resolution [9,11].

2. Vector field

Because the combined dynamics is assumed ergodic, it makes sense to characterize the mean changes of combined states in terms of the ensemble average of changes of combined states. This average may be expressed in terms of the conditioned probability $P(\{\tilde{n}_i(t+1)\}|\{n_j(t)\},\{\tilde{n}_i(t)\},\tilde{W}(t),T)$ that a neuronal configuration $\{\tilde{n}_i(t+1)\}$ is taken at the time step $t+1$ under the condition that at the time step t the stimulation is $\{n_j(t)\}$ and the combined state is $\{\tilde{n}_i(t)\}$, $\mathbf{W}(t)$ and the fluctuation parameter is T. In particular, this average is the following sum over the possible 2^N configurations of the neuronal states at time t and over the possible 2^N configurations of the neuronal states at time $t+1$, multiplied with the probability $1/2^N$ due to the uniform distribution of the stimulating states:

$$\langle(\Delta\mathbf{n},\Delta\mathbf{W})\rangle=\sum_{\{n_j\}}^{2^N}\frac{1}{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}P(\{\tilde{n}_i(t+1)\}|\{n_j\},\{\tilde{n}_i\},\mathbf{W},T)(\mathbf{n}(t+1)-\mathbf{n},\Delta\mathbf{W})\ . \quad (19)$$

As a consequence, for each T the mean changes $\langle(\Delta\mathbf{n},\Delta\mathbf{W})\rangle$ establish a vector field in the combined space, because such mean changes are functions of the combined state due to the condition of the above-conditioned probability and due to the fact that after averaging such mean changes do not depend on the stimulation.

3. *Differential equation*

One may introduce the continuous time limit by using a variable time interval Δt between two successive time steps t and $t+\Delta t$. That is, the time steps, t, $t+\Delta t$, $t+2\Delta t$, $t+3\Delta t, \ldots$ are modeled. With it one may divide the above equation by that time interval and take the limit of zero time interval, that is,

$$\frac{\partial\langle(\mathbf{n},\mathbf{W})\rangle}{\partial t}=\lim_{\Delta t\to 0}\frac{\langle(\mathbf{n}(t+\Delta t)-\mathbf{n},\mathbf{W}(t+\Delta t)-\mathbf{W})\rangle}{\Delta t}$$

$$=\frac{1}{2^N}\lim_{\Delta t\to 0}\sum_{\{n_j(t)\}}^{2^N}\sum_{\{\tilde{n}_i(t+\Delta t)\}}^{2^N}P(\{\tilde{n}_i(t+\Delta t)\}|\{n_j\},\{\tilde{n}_i\},\mathbf{W},T)\frac{(\mathbf{n}(t+\Delta t)-\mathbf{n},\mathbf{W}(t+\Delta t)-\mathbf{W})}{\Delta t}\ . \qquad (20)$$

Thus the mean change of combined states obeys the above explicitly expressed ordinary differential equation in the combined space.

4. *Adiabatic limit*

Typically, the neurons change on the time scale of milliseconds, whereas the couplings change on the much longer time scale of minutes to years. Thus one may solve the motion of the fast neurons first by means of an adiabatic limit (that is the leading order of a systematic adiabatic approximation as described in [16]), then one may solve the changes of the slower couplings.

To this end one may proceed as follows. One may consider a fixed value of the slow couplings and perform the average over the neuronal states [see Eq. (19)] and one may use the fact that the state $\{\tilde{n}_i(t+1)\}$ is generated independently from the state $\{\tilde{n}_i\}$. As a result one obtains for the mean change of couplings

$$\langle\Delta\mathbf{W}\rangle=\frac{1}{2^N}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}P(\{\tilde{n}_i(t+1)\}|\{n_j\},\mathbf{W},T)\Delta\mathbf{W}\ . \qquad (21)$$

These mean coupling changes establish another vector field in coupling space. The corresponding differential equation may be obtained either by applying the usual adiabatic approximation in leading order to the former differential equation [see Eq. (20)], or by performing the continuum limit to the above Eq. (21). As a result one gets

$$\frac{\partial\langle\mathbf{W}\rangle}{\partial t}=\frac{1}{2^N}\lim_{\Delta t\to 0}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+\Delta t)\}}^{2^N}P(\{\tilde{n}_i(t+\Delta t)\}|\{n_j\},\mathbf{W},T)\frac{\mathbf{W}(t+\Delta t)-\mathbf{W}}{\Delta t}\ . \qquad (22)$$

For the sake of explicitness, one may express the mean coupling change [see Eq. (21)] in terms of the components

$$\langle\Delta W_{ij}\rangle=\frac{1}{2^N}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}P(\{\tilde{n}_i(t+1)\}|\{n_j\},\mathbf{W},T)\Delta W_{ij}\ . \qquad (23)$$

5. *Necessary and sufficient condition for the adequate applicability of the adiabatic limit*

For the purpose of a general applicability of the present approach in cortical maps, one would like to be sure that the adiabatic limit may be applied quite generally. A trivial condition for the applicability is that the time scales should differ significantly; this is typically the case in the nervous system. A nontrivial condition is that the neuronal dynamics "comes to a result" sufficiently fast so that the ensemble average $\sum_{\{\tilde{n}_i(t+\Delta t)\}}^{2^N}P(\{\tilde{n}_i(t+\Delta t)\}|\{n_j\},\mathbf{W},T)$ over neuronal states corresponds to a time average for those time scales at which the couplings do change only slightly. Here this is probably the case for all neuronal processes that are effectively finished within a few minutes. Nowadays, slower neuronal process can hardly be measured, for instance, event related potentials can be recorded at most for few seconds. These two conditions should be sufficient and necessary for the adequate applicability of the adiabatic limit. Thus the adiabatic limit is adequate for practically all nervous events that are measurable nowadays.

In the present case of the self-organization of topological order, the adiabatic limit can be applied especially easily, because the neuronal states $\{\tilde{n}_i(t+1)\}$ do not depend on the previous states $\{\tilde{n}_i(t)\}$, due to the feed forward network architecture. So the above nontrivial condition is obeyed in a trivial manner here.

B. Potential field

The mean coupling change in Eq. (23) is a vector field in coupling space. Next it is shown that this vector field turns out to be a gradient of a scalar potential, that is, a potential field.

1. *Potential theorem*

In the adiabatic limit, the mean coupling change [see Eq. (23)] is the gradient of a scalar potential as follows:

$$\langle\Delta W_{ij}\rangle=-\frac{\partial V}{\partial W_{ij}} \qquad (24)$$

with the scalar potential

$$V=-\frac{aT}{2^N}\sum_{\mu}^{2^N}\ln Z^{\mu} \text{ where } \textit{the stimulation } \{n_j\} \text{ is denoted by } \mu \tag{25}$$

and the formal partition functions

$$Z^{\mu}=\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\exp[-H^{\mu}/T] \; . \tag{26}$$

Accordingly, the stable emerging networks are the local minima of the scalar potential V.

2. *Principle underlying the proof*

Formally, one may interpret the determination of the potential as an integration, for instance, one may integrate Eq. (24), so one may get $V=-\int_0^{W_{ij}}\langle\Delta W_{ij}\rangle dW_{ij}$. In this sense the method and the obtained results are *quite general.* The fact that in the particular present case the resulting integral may in fact be expressed in terms of an explicit and *rather simple function* is due to the form of the probability [see Eqs. (1)–(5)] and of the whole network model. In particular, methods of statistical physics are applied here and generalized to the case of single objects like single neurons and couplings, whereas statistical physics deals with systems in the limit of an infinite number of objects.

3. *Proof*

To begin with, one may take Eq. (23) and express the conditioned probability in terms of the product of the probabilities in Eq. (5), because this product is the desired probability that a neuronal state $\{\tilde{n}_i(t+1)\}$ is taken. So one gets

$$\langle\Delta W_{ij}\rangle=\frac{1}{2^N}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\prod_{i=1}^{N}P(\tilde{n}_i(t+1)|\{n_j\},\mathbf{W},T)\Delta W_{ij} \; . \tag{27}$$

Next one may insert for the probabilities according to Eq. (5) and express these probabilities in the denominator with a sum of two exponentials. So one gets

$$\langle\Delta W_{ij}\rangle=\frac{1}{2^N}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\prod_{i=1}^{N}\frac{\exp[\tilde{h}_i(t+1)\tilde{n}_i(t+1)/T]}{\sum_{\tilde{n}_i(t+1)=0,1}^{2}\exp[\tilde{h}_i(t+1)\tilde{n}_i(t+1)/T]}\Delta W_{ij} \; . \tag{28}$$

This expression may be transformed into (for details see [9–11])

$$\langle\Delta W_{ij}\rangle=\frac{1}{2^N}\sum_{\{n_j\}}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\frac{\exp[-H^{\mu}/T]}{Z^{\mu}}\Delta W_{ij} \; . \tag{29}$$

For the purpose of a later comparison, one may perform the gradient of the scalar potential

$$-\frac{\partial V}{\partial W_{ij}}=-\frac{a}{2^N}\sum_{\mu}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\frac{\exp[-H^{\mu}/T]}{Z^{\mu}}i\frac{\partial H^{\mu}}{\partial W_{ij}} \; . \tag{30}$$

Next one may use the fixed stimulation potential theorem, so one gets

$$-\frac{\partial V}{\partial W_{ij}}=\frac{1}{2^N}\sum_{\mu}^{2^N}\sum_{\{\tilde{n}_i(t+1)\}}^{2^N}\frac{\exp[-H^{\mu}/T]}{Z^{\mu}}\Delta W_{ij} \; . \tag{31}$$

By comparison of this expression with Eq. (29) one obtains

$$\langle\Delta W_{ij}\rangle=-\frac{\partial V}{\partial W_{ij}} \; . \tag{32}$$

Q.E.D.

4. *Interpretation of the potential theorem*

The potential V makes possible an intuitive and simple understanding and analysis of the emerging networks in terms of local potential minima. Moreover one may derive for any desired stimulation (rather than equally distributed as above) the resulting emerging networks. Conversely, one may design for a desired network an appropriate stimulation that gives rise to it.

C. Emergence of an injective mapping

Due to the fact that the growth of the coupling biomass is fast at large couplings and limited at each neuron, one might expect that few couplings are nonzero at each neuron. In the following, two neurons n_j and $\tilde{n}_i$ are called *connected,* if they are coupled by a nonzero coupling $W_{ij}\neq 0$.

1. *Injective mapping theorem*

For networks that are locally stable with respect to stochastic fluctuations and with respect to variations of the formal temperature T holds: (1) Each sensor neuron is connected with n_0 cortical neurons. (2) Each cortical neuron is connected with exactly one sensor neuron. This establishes an injective mapping from the cortical neurons to the sensor neurons. (3) The injective mapping establishes the formation of cortical neuron sets of n_0 cortical neurons. This establishes a bijective mapping from the cortical neuron sets to the sensor neurons and vice versa. (4) Each emerging nonzero coupling has the weight r.

2. *Principle underlying the proof*

There are two main underlying reasons for the mapping theorem: First, biological matter does typically grow proportional to its present biomass, this fact is used for coupling growth here and it is inherent to the factor W_{ij} in the coupling dynamics [see Eq. (6)]. This gives rise to the fact that large couplings tend to grow faster than small couplings. Second, biological matter does typically grow within some limitations. Such limitation is used at a neuron in a quite local manner and it is expressed via the constraints [see Eqs. (8) and (9)], or alternatively via the additional dynamics [see Eqs. (10) and (11)], and these have been observed empirically for the case of synapses [14]. It is already clear intuitively that the combination of the first and second reasons gives rise to a tendency towards states with one coupling at a cortical neuron and n_0 couplings at a sensor neuron.

3. *Proof*

One may use multidimensional polar coordinates for the couplings as follows.

$$\begin{aligned} &W_{ij}=r\cos\vartheta_{ij} \ , \\ &W_{i+1j}=r\sin\vartheta_{ij}\cos\vartheta_{i+1,j} \ , \\ &W_{i-1j}=r\sin\vartheta_{ij}\sin\vartheta_{i+1,j}\cos\vartheta_{i-1,j} \ , \\ &\vdots \end{aligned} \tag{33}$$

(Nonsingular nature of the representation: Inherent to polar coordinates is the singularity at the pole, for each coupling state, this singularity may always be avoided by shifting the pole away from the coupling constant. Formally, the coupling W_{ij} might vary from positive to negative values in a nonanalytic manner; however, such variation is irrelevant, because $K_{ij}=W_{ij}^2$ is the actual coupling factor.) The networks that are locally stable with respect to stochastic fluctuations are specified by the local minima of the potential V. Consequently, the partial derivatives of the potential V with respect to the above angle variables vanish for these networks, that is,

$$\frac{\partial V}{\partial\vartheta_{mk}}=0 \ . \tag{34}$$

In order to determine the form of such a derivative, one may recall that any angular variable ϑ_{mj} in the potential V occurs in terms of a $\cos^2\vartheta_{mj}$ or in terms of a $\sin^2\vartheta_{mj}$, because the couplings enter the potential in terms of squares [see Eqs. (4), (33), (25), (26), and (15)]. Consequently, the derivative of the potential V with respect to such an angular variable ϑ_{mj} is proportional to $\sin\vartheta_{mj}\cos\vartheta_{mj}$, due to the chain rule. That is, each such derivative is of the form

$$\frac{\partial V}{\partial\vartheta_{mj}}=\sin\vartheta_{mj}\cos\vartheta_{mj}\text{rest}(T)=0 \ , \tag{35}$$

whereby rest(T) denotes the remaining factor, which is a function of the formal temperature. Consequently, a network that is locally stable with respect to stochastic fluctuations obeys either $\sin\vartheta_{mj}\cos\vartheta_{mj}=0$ or rest(T)=0. The networks that do not obey $\sin\vartheta_{mj}\cos\vartheta_{mj}=0$ do obey rest(T)=0; thus they vary with the formal temperature, hence they are not stable with respect to temperature variations. As a consequence, those networks that are locally stable with respect to stochastic fluctuations and with respect to variations of the formal temperature T do obey $\sin\vartheta_{mj}\cos\vartheta_{mj}=0$. This implies $\vartheta_{mj}=0$ or $\vartheta_{mj}=\pi/2$. Thus one gets [see Eq. (33)] W_{ik} is either 0 or r.

Moreover, one may recall the constraints [see Eqs. (8) and (9)] $n_0r^2=\sum_m W_{mj}^2$ and $r^2=\sum_j W_{mj}^2$. The first ($n_0r^2=\sum_m W_{mj}^2$) implies that at each sensor neuron n_j, there are n_0 nonzero couplings $W_{mj}=r$. Analogously, the second ($r^2=\sum_j W_{mj}^2$) implies that each cortical neuron $\tilde{n}_m$, there is exactly one nonzero coupling $W_{mj}=r$. That is, each sensor neuron is connected with n_0 cortical neurons and each cortical neuron is connected with one sensor neuron. This implies immediately the four items of the theorem. Q.E.D.

4. *Illustrative discussion of the mapping theorem*

Here the bijective mapping emerges between sets of cortical neurons and sensor neurons. This result is quite general. For instance, if there are metric relations among sensor neurons, these are related to cortical neurons and may be described in terms of densities.

D. Emergence of clusters and topology preservation

In this section it is shown how clustering and topology preservation emerge in the case of arbitrary topologies. In such a case, there may occur constraints that cannot be satisfied in a theoretically optimal manner, due to missing neighbors, for instance. In order to formally treat clustering and topology preservation emergence also for such constraints, the notions *cluster relation, topology-preservation relation,* and *relative stabilization* are introduced first as follows:

At a locally stable coupling state (see injective mapping theorem), two neighboring cortical neurons that are connected with the same sensor neuron exhibit the cluster relation, and two cortical neurons in the same neighbor set that are connected with the same sensor neuron exhibit the cluster relation, while two neighboring cortical neurons that are connected with two neighboring sensor neurons exhibit the topology-preservation relation. The potential V is a sum of local potentials V_i [for details see (1) of the following proof]. If a cluster relation or a topology-preservation relation gives rise to a decrease of a local potential V_i, then such a relation is called *relatively stable.*

1. Clustering and neighborship preservation theorem

For a coupling state that is locally stable with respect to stochastic fluctuations and with respect to variations of the formal temperature T holds: (1) A nonzero postsynaptic lateral contribution parameter β is a necessary and sufficient condition for the relative stabilization of a cluster relation. (2) Nonzero presynaptic and postsynaptic lateral contribution parameters α and β are necessary and sufficient for the relative stabilization of a topology-preservation relation.

2. Principle underlying the proof

The clustering and topology preservation emerge here due to four essential principles: First, the Hebb mechanism increases couplings that transfer signals among successively (that is almost coincidently) active neurons. Second, a slight lateral signal transfer among neighboring neurons, for instance, due to chemical or electrical leakage, can hardly be avoided in a realistic system. Such lateral signal transfers give rise to additional coincidences, if the topology is preserved. Third, presently large coupling biomass grows relatively fast (this is also essential for the mapping theorem above). Fourth, the total coupling biomass at a neuron is limited (this is also essential for the mapping theorem above). As a result, those couplings grow best and thus remain that provide most coincidences and these are the couplings that provide clusters and topology preservation.

3. Proof

(i) Separation of the potential V into local potentials V_i: First one may recall [see Eqs. (26) and (15)] that the partition function is

$$Z^{\mu}=\sum_{\{n_i\}}^{2^N}\prod_{i}^{N}\exp[\tilde{h}_i(t+1)\tilde{n}_i(t+1)/T] \ . \tag{36}$$

Next one may exchange the product and the sum according to the distributive law; so one gets

$$\begin{aligned} Z^{\mu}&=\prod_{i=1}^{N}\sum_{n_i=0,1}^{2}\exp[\tilde{h}_i(t+1)\tilde{n}_i(t+1)/T] \\ &=\prod_{i=1}^{N}(1+\exp[\tilde{h}_i(t+1)/T]) \ . \end{aligned} \tag{37}$$

Here one may apply the logarithm

$$\ln Z^{\mu}=\sum_{i=1}^{N}\ln(1+\exp[\tilde{h}_i(t+1)/T]) \ . \tag{38}$$

So the potential V [see Eq. (25)] may be separated into local potentials V_i, one for each cortical neuron $\tilde{n}_i$ as follows:

$$V=\sum_{i}^{I}V_i$$

$$\text{with } V_i=-\frac{aT}{2^N}\sum_{\mu}\ln\{1+\exp[\tilde{h}_i(t+1)/T]\} \ . \tag{39}$$

(ii) Form of the local fields h_i at a local minimum of the potential V: Next one may explicate the formal local field $\tilde{h}_i$ by using Eq. (4):

$$\tilde{h}_i(t+1)=\sum_{j}^{N}\left[\phi_{ij}\frac{W_{ij}^2}{2}+\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}\phi_{qj}\frac{W_{qj}^2}{2}\right] \ . \tag{40}$$

As a consequence of the injective mapping theorem, in the above sum $\sum_j$ there are exactly the following nonzero terms. The corresponding nonzero couplings may be expressed by using a sensor neuron index $u(i)$ conjugated to the index i of the connected cortical neuron $\tilde{n}_i$, a sensor neuron index $u(q)$ conjugated to the index q of the connected cortical neuron $\tilde{n}_q$. So one gets

$$\begin{aligned} &W_{i,u(i)}=r \quad \text{to } \tilde{n}_i \text{ from } n_{u(i)} \ , \\ &W_{q,u(q)}=r \quad \text{to } \tilde{n}_q \text{ from } n_{u(q)} \text{ with } \tilde{n}_q\in\tilde{v}(i) \ . \end{aligned} \tag{41}$$

Next one may insert these explicit expressions into the above Eq. (40). So the sum over j yields nonzero terms for $j=u(i)$ and $j=u(q)$. So one gets

$$\tilde{h}_i(t+1)=\frac{r^2}{2}[\phi_{i,u(i)}+\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}\phi_{q,u(q)}] \ . \tag{42}$$

Here one may explicate the membrane potentials ϕ_{ij} in terms of the sensor neuron states n_j [see Eq. (1)]. So one obtains

$$\begin{aligned} \tilde{h}_i(t+1)=\frac{r^2}{2}\Bigg[&n_{u(i)}+\alpha\sum_{n_m\in v[u(i)]}n_m+\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}n_{u(q)} \\ &+\alpha\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}\sum_{n_m\in v[u(q)]}n_m\Bigg] \ . \end{aligned} \tag{43}$$

Moreover, one may add an upper index μ for the stimulation, in order to indicate the inherent dependence of the above terms on the particular stimulation:

$$\begin{aligned} \tilde{h}_i^{\mu}(t+1)=\frac{r^2}{2}\Bigg[&n_{u(i)}^{\mu}+\alpha\sum_{n_m\in v[u(i)]}n_m^{\mu}+\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}n_{u(q)}^{\mu} \\ &+\alpha\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}\sum_{n_m\in v[u(q)]}n_m^{\mu}\Bigg] \ . \end{aligned} \tag{44}$$

(iii) Form of a local potential V_i at a local minimum of the potential V: Due to Eqs. (39) and (44), each local potential V_i is an average over the states of sensor neurons n_q; explicitly, V_i takes the form

$$V_i=-\frac{aT}{2^Q}\sum_{n_q=0,1}^{2^Q}\ln\left[1+\exp\left[\frac{\sum_q^Q\alpha_q n_q}{T}\right]\right] \ . \tag{45}$$

(iv) Reduction of a local potential V_i due to the identity of two sensor neurons n_q: In this part, the basic reason for the stabilization of clusters and topology preservation is derived, namely, the potential reduction as a consequence of equal presynaptic neurons n_q in Eq. (45); this may be interpreted as an effective force that tends to make presynaptic neurons equal.

One may begin with the above form for V_i, consider two neurons $n_q=n_x$ and $n_q=n_y$ explicitly and denote the terms due to the other $n_q$by R. So one gets

$$V_i=-\frac{aT}{2^Q}\sum_{n_q=0,1}^{2^Q}\ln\left[1+\exp\left[\frac{\alpha_x n_x+\alpha_y n_y+R}{T}\right]\right] . \tag{46}$$

Next one may consider two local minima of the potential V so that the neurons n_q of the potential V_i are the same for both minima, except that in the second minimum one has $n_x=n_y$. In such a case the local potential V_i of the second local minimum is reduced by the term $\Delta V_i=V_{\text{with } n_x \text{ not identical to } n_y}-V_{n_x \text{ identical to } n_y}$. This difference is determined as follows:

$$\Delta V_i=\frac{aT}{2^Q}\sum_{n_q=0,1;n_q\neq n_x;n_q\neq n_y}^{2^{Q-2}}\ln\left[1+\frac{(e^{\alpha_x/T}-1)(e^{\alpha_y/T}-1)}{e^{-R/T}+e^{\alpha_x/T}+e^{\alpha_y/T}+e^{(\alpha_x+\alpha_y+R)/T}}\right] . \tag{47}$$

This ΔV_i is positive, because the factors α_x and α_y are both positive for positive lateral coupling parameters α and β.

(v) Proof of two parts of the theorem: In general, local stability is achieved, if and only if subindices of neuron-variables that occur in Eq. (44) are equal [see part (iv)].

First, if α is zero and β is nonzero, then Eq. (44) takes the form

$$\tilde{h}_i^\mu(t+1)=\frac{r^2}{2}\left[n^\mu_{u(i)}+\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}n^\mu_{u(q)}\right] . \tag{48}$$

Here, equal subindices can be achieved only if $u(i)=u(q)$, that is if the cortical neuron $\tilde{n}_i$ and its neighbor $\tilde{n}_q$ are connected to the same sensor neuron, or if $u(q')=u(q)$, that is if the cortical neuron $\tilde{n}_q$ and another cortical neuron $\tilde{n}_{q'}$ in the same neighbor set are connected to the same sensor neuron. Thus

$\beta\neq 0$ is a sufficient condition for relative

stabilization of a cluster relation (49)

and

$\beta\neq 0$ is not a sufficient condition for relative

stabilization of a topology-preservation

relation . (50)

Second, if α is nonzero and β is zero, then Eq. (44) takes the form

$$\tilde{h}_i^\mu(t+1)=\frac{r^2}{2}\left[n^\mu_{u(i)}+\alpha\sum_{n_m\in v[u(i)]}n^\mu_m\right] . \tag{51}$$

The subindices occurring here cannot be made equal, so neither clustering nor topology preservation emerges in this case. Thus

$\beta\neq 0$ is a necessary condition for relative

stabilization of a topology-preservation

relation (52)

and

$\beta\neq 0$ is necessary for the relative

stabilization of a cluster relation . (53)

Third, if α and β are nonzero, then the clustering occurs as in the first case above, and in addition the term proportional to $\alpha\beta$, namely,

$$\alpha\beta\sum_{\tilde{n}_q\in\tilde{v}(i)}\sum_{n_m\in v[u(q)]}n^\mu_m , \tag{54}$$

gives rise to additional equal subindices as follows: Here the subindex m corresponds to a sensor neuron n_m that is a neighbor of another sensor neuron $n_{u(q)}$ that is connected with a cortical neuron $\tilde{n}_q$ that is a neighbor of another cortical neuron $\tilde{n}_i$. Such sensor neurons n_m can be equal to the sensor neuron $n_{u(i)}$ connected to the cortical neuron $\tilde{n}_i$ or to a sensor neuron $n_{u(q)}$ connected to the cortical neuron $\tilde{n}_q$ neighboring $\tilde{n}_i$, due to topology preservation. Thus

$\frac{r^2}{2}\alpha\beta\neq 0$ is a sufficient and necessary

condition for relative stabilization

of a topology-preservation relation

Q.E.D. (55)

4. *On relative stability*

Except for possible constraints due to the arbitrary topology of a network, the minimization of the potential V and of the potentials V_i are identical. So the relative stability is the same as the usual stability, except for such constraints. Thus, in systems without such constraints, the relative stability is the same as the usual stability.

For instance, the one-dimensional model (see Fig. 1) exhibits no such constraints, so each globally stable coupling state exhibits perfect topology preservation; this special case is treated explicitly in [8–10]. The general case of relative stability corresponds to the experimental findings of almost perfect topology preservation in cortical maps [15].

5. *Interpretation of the topological order theorem*

The theorem proves two essential facts. First, in a Hebb-type system, already infinitesimal lateral contributions to neighbor neurons give rise to the emergence of topology preservation. This can hardly be understood in the framework of Kohonen-type networks, because the latter contain the so-called "winner takes all" mechanism from the very beginning; as a consequence, the necessary quantitative considerations are practically completely "overshadowed" by the assumed "winner takes all" mechanism. Moreover, such infinitesimal lateral contributions to neighbor neurons occur practically in any system, due to chemical or electrical leakages. As a consequence, one should expect topological order to occur in Hebb-type systems, irrespective of a possible purpose. Second, topological order can be avoided in Hebb-type systems, if at least one of the lateral contributions (presynaptic or postsynaptic) is eliminated or compensated.

IV. CONCLUSION

Topology preservation is a ubiquitous phenomenon in the mammalian nervous system [1,15]. Today it is quite clear that self-organization processes of plastic synapses are essential for the formation of such order and that plastic synapses are adequately describable by a Hebb mechanism [5]. So the following question arises. What are the necessary and sufficient conditions for the self-organized formation of topology preservation due to a Hebb mechanism?

Previous modeling applied either Kohonen networks, however, these make such crude assumptions (for instance, a winner takes all mechanism) that the present question cannot be investigated, or they used a slightly unrealistic Hebb rule [5,8]. In contrast, in the present paper a quite realistic Hebb rule and neurons with stochastic fluctuations are modeled. Moreover, the reasonable growth law is used for coupling growth that the coupling biomass increase is proportional to the present coupling biomass under the constraint that the coupling biomass at a neuron is limited. Infinitesimal lateral signal transfer to neighbor neurons, as it occurs in any nervous system due to chemical and electrical leakage, gives rise to topology preservation.

In this manner it is proven for the present quite general Hebb-type networks that such infinitesimal lateral signal transfer to neighbor neurons is necessary and sufficient for the emergence of topology preservation. As a consequence, observed topology preservation in nervous systems may emerge with or without purpose as a byproduct of infinitesimal lateral signal transfer to neighbor neurons due to chemical and electrical leakage.

The modeling of topological order in terms of a simple Hebb rule makes possible the future investigation and possible understanding of combined learning and self-organization mechanisms in the brain. Or alternatively, one may implement technical neural networks that simultaneously self-organize topologically and learn via reinforcement with the same basic Hebb mechanism.

ACKNOWLEDGMENTS

I am grateful for stimulating discussions with Hans Flohr, Gerhard Roth, Olaf Scherf, and Helmut Schwegler.

[1] W. H. Marshall, C. N. Woolsey, and P. Bard, J. Neurophysiol. **4**, 1 (1941).

[2] R. Descartes, *L'Homme* (Chez Theodore Girard, Paris, 1664).

[3] P. Corsi, *The Enchanted Loom* (Oxford, Oxford, 1991).

[4] P. Weiss, Naturwissenschaften **16**, 626–636 (1928).

[5] D. J. Willshaw and C. von der Malsburg, Proc. R. Soc. London B **194**, 431 (1976).

[6] T. Kohonen, *Self-Organization and Associative Memory* (Springer, Berlin, 1989).

[7] H. Ritter, T. Martinetz, and K. Schulten, *Neuronale Netze* (Addison-Wesley, Bonn, 1991).

[8] H.-O. Carmesin, in *International Conference on Applied Synergetic and Synergetic Engineering Proceedings*, edited by F. G. Bbel and T. Wagner (Fraunhofer-Gesellschaft, Erlangen, 1994), pp. 53–60.

[9] H.-O. Carmesin, *Theorie neuronaler Adaption* (Köster, Berlin, 1994).

[10] H.-O. Carmesin, *Neuronal Adaptation Theory* (Peter-Lang, Frankfurt, 1996).

[11] H.-O. Carmesin, Phys. Essays **8** (1), 38 (1995).

[12] S. H. Strogatz and I. Stewart, Spektrum **2/94**, 74–81 (1994).

[13] F. J. Pineda, Phys. Rev. Lett. **59**, 2229 (1987).

[14] H. Reichert, *Neurobiologie* (Thieme, Stuttgart, 1990).

[15] E. R. Kandel, J. H. Schwarz, T. M. Jessell, *Principles of Neural Science* (Elsevier, New York, 1991).

[16] H. Haken, *Advanced Synergetics* (Springer, Berlin, 1983).