

## Critical phenomena in self-organizing feature maps: Ginzburg-Landau approach

R. Der and M. Herrmann\*

*Universität Leipzig, Institut für Informatik, D-04109 Leipzig, Augustusplatz 10/11, Germany*

(Received 6 July 1993; revised manuscript received 11 January 1994)

Self-organizing feature maps (SOFM's) as generated by Kohonen's algorithm are prominent examples of the cross fertilization between theoretical physics and neurobiology. SOFM's serve as high-fidelity models for the internal representation of the external world in the cortex. This is exploited for applications in the fields of data analysis, robotics, and for the data-driven coarse graining of state spaces of nonlinear dynamical systems. From the point of view of physics Kohonen's algorithm may be viewed as a stochastic dynamical equation of motion for a many particle system of high complexity which may be analyzed by methods of nonequilibrium statistical mechanics. We present analytical and numerical studies of symmetry-breaking phenomena in Kohonen's SOFM that occur due to a topological mismatch between the input space and the neuron setup. We give a microscopic derivation for the time dependent Ginzburg-Landau equations describing the behavior of the order parameter close to the critical point where a topology preserving second-order phase transition takes place. By extensive computer simulations we do not only support our theoretical findings, but also discover a first order transition leading to a topology violating metastable state. Consequently, close to the critical point we observe a phase-coexistence regime.

PACS number(s): 87.10.+e, 02.50.Ga, 05.70.Fh, 74.20.De

### I. INTRODUCTION

Recent years have witnessed an increasing cross fertilization between theoretical physics and biology, in particular concerning the science of the brain. The most prominent example of this development has been the analogy—established by Hopfield [1]—between a simplified model of a neural network and spin glasses. By the application of the highly developed methods of spin glass theory, the understanding of the behavior of neural nets was increased considerably. The enormous relevance of neural nets for both the understanding of the brain function and as a computing paradigm triggered one of the most spectacular booms in the sciences in the past decade.

The present paper is concerned with another type of neural network models, the so-called self-organizing ones which are of much interest for both biology and applications. The original model given by Kohonen was based on biological considerations in the attempt of understanding the self-organized generation of an internal representation of the environment in the brain; cf. [2]. In a later version this model has been reformulated into an algorithm which produces in a self-organized way a topographic mapping of the external world as seen by the sensors to the internal world of neural activities. The possible relevance of this algorithm for understanding somatotopic and retinotopic mappings in biology has been clarified in extensive computer simulations; cf. Ref. [3].

Besides its biological implications of providing a simple model of the self-organized generation of an internal representation of the environment in living systems, the Kohonen algorithm [2] is also of more general interest for the

following reasons. In practical applications the short-cut form of Kohonen's algorithm is used as a nonlinear generalization of principal component analysis which plays an important role in many problems of data analysis, above all for data in high dimensional data spaces. From the point of view of systems theory Kohonen's algorithm models a composite system with competition and cooperation between its constituents which may serve as a generic though degenerate example of the emergence of collective ordering phenomena in more complicated systems like economic or ecologic ones.

Finally, from the point of view of physics the synaptic vectors can be viewed as coordinates of a set of hypothetical particles with a complicated dynamics created by the Kohonen algorithm. The complexity of this system is reflected by phenomenalike spontaneous symmetry breaking (analogous to nonequilibrium phase transitions) and the occurrence of metastable states with partial ordering or the emergence of criticality under certain conditions [4].

Concerning practical applications these effects may well interfere with the intentions of using Kohonen's algorithm as a reliable tool for topology preserving feature mapping. On the other hand, if better understood these effects might be exploited for more sophisticated applications. In this sense, there is a great lack of theoretical study into the details of the map evolution. There is a general proof [7] based on a generalized central limit theorem of the convergence of the algorithm to a stable state under an implicitly specified cooling regime for the parameters. However, nothing is said about the topological order of this state or on the distribution of the synaptic vectors as a function of the input distribution. For the one dimensional case, i.e., the mapping of a one dimensional input space onto a one dimensional set of neurons, there is a rigorous proof [6] for the convergence of the map to the ordered state.

---

\*Electronic address: der@informatik.uni-leipzig.d400.de

Pioneering theoretical work has been performed by Ritter, Martinetz, and Schulten [5] in the framework of a linear stability analysis. They predicted a phase transition to occur due to dimensional conflicts between input space and net topology and calculated the critical value for the corresponding control parameter which measures the strength of the dimensional conflict.

In the present paper we use methods of nonlinear nonequilibrium statistical mechanics for the evaluation of the critical behavior of the order parameters in the self-organized feature map. These methods have been used successfully in the description of pattern formation by spontaneous symmetry breaking in reaction diffusion systems, cf. [8], and are based on deriving from first principles a time dependent Ginzburg-Landau equation describing the behavior of the order parameter in the vicinity of the critical point where the phase transition takes place.

The Ginzburg-Landau approach is quite general but for the explicit calculations we study here the simple case of mapping a rectangular data distribution in a two dimensional input space onto a one dimensional neuron chain. The pertinent order parameter in this case is just the amplitude of the folding the image of the chain develops if the algorithm tries to map the two dimensional onto the one dimensional space, which corresponds to the attempt to cover the rectangle by the chain of neurons.

## II. LANGEVIN PICTURE OF KOHONEN'S ALGORITHM

Let us consider a set of  $N$  neurons situated at sites  $\vec{r} \in \mathcal{N} \subseteq \mathcal{Z}^d$ .  $\mathcal{Z}^d$  is the  $d$  dimensional lattice with unity lattice constant, and  $\mathcal{N}$  is a subset of  $\mathcal{Z}^d$  containing  $N$  points. Each neuron is connected with the input units by synaptic connectivities  $\vec{w}_{\vec{r}} \in \mathcal{R}^n$ . The inputs to the network are given by random stimuli  $\vec{v} \in \mathcal{R}^n$  distributed according to the probability density  $P(\vec{v})$ .

The vectors of synaptic weights  $\vec{w}_{\vec{r}}$  may be viewed as a kind of pointers into the input space, i.e., the components of  $\vec{w}_{\vec{r}}$  may be viewed as the coordinates of the image of the neuron situated at lattice site  $\vec{r}$ . Kohonen's learning rule is given by

$$\Delta \vec{w}_{\vec{r}}(t) = -\epsilon h_{\vec{r}, \vec{r}^*} (\vec{w}_{\vec{r}} - \vec{v}), \quad (1)$$

where  $\vec{r}^*$  denotes the winner neuron defined as the one with the best match (least distance) between input and synaptic vector  $\vec{w}_{\vec{r}}$

$$\|\vec{w}_{\vec{r}} - \vec{v}\| \geq \|\vec{w}_{\vec{r}^*} - \vec{v}\| \quad \forall \vec{r} \neq \vec{r}^*, \quad (2)$$

where  $\|\dots\|$  denotes the Euclidean norm in  $\mathcal{R}^n$ , and

$$h_{\vec{r}, \vec{r}^*} = \exp\left(-\frac{(\vec{r} - \vec{r}^*)^2}{2\sigma^2}\right) \quad (3)$$

is a Gaussian neighborhood function of width  $\sigma$  measuring the degree by which the neurons in the vicinity of the winner may participate in the learning step.

The winner rule provides a mapping between the input and neuron space by the agreement that all stimulus vectors which make the neuron at site  $\vec{r}$  the winner are mapped onto this neuron, its pointer being the code book vector for the set of these stimuli. This also subdivides the input space into a disjoint set of regions called the domains of the neurons each domain being mapped onto one lattice site (Voronoi tessellation). On the other hand, we may also map the lattice into the input space by just connecting the images of the neurons in the input space in the same way as the neurons in the lattice.

Kohonen's learning rule is understood best by noting that the increment of  $\vec{w}$  is proportional to the difference vector between the stimulus vector  $\vec{v}$  and the pointer (or synaptic) vector  $\vec{w}_{\vec{r}}$ , hence the images of the neurons are moved in each learning step into the direction of the stimulus, with step length dictated by the prefactor  $\epsilon h_{\vec{r}, \vec{r}^*}$ . In physical terms the (images of the) neurons may be interpreted as particles moving in input space under the influence of attractive forces exerted by the stochastically applied input stimuli  $\vec{v}$ . The neighborhood function acts as a kind of interaction between the particles. In fact, the stimulus attracts not only the particle closest to it but also its neighbors, the neighborhood, however, being defined by the topology in the neuron space (the  $d$  dimensional lattice in our case). Hence neighbored particles (images of the neurons in input space) move coherently only if they belong to neurons which are neighbors with respect to the topology in the neuron space. This way the algorithm produces a dynamic grouping of particles according to the topology of the neuron space and hence tries to reconcile the topology of the input with that of the neuron space. By a convenient shrinking of the learning rate  $\epsilon$  and the width  $\sigma$  of the neighborhood starting from a sufficiently large value, the topology of the neuron space eventually is reflected in the stationary positions of the particles in the input space.

The data driven algorithm Eq. (1) defines a Markovian stochastic process. Equation (1) may also be viewed as a stochastic difference equation with multiplicative noise which is verified by noting that the neighborhood function depends in a highly nonlinear fashion on the stimulus vector  $\vec{v}$  since the site of the winner neuron is  $\vec{r}^* = \vec{r}^*(\vec{v})$ . Our approach is based on rewriting (1) into the form of a generalized Langevin equation which essentially is a stochastic difference equation with additive noise. This is possible close to the stationary state of the system as shown below. The Langevin equation is obtained by separating the right hand side of Eq. (1) into a systematic and a random part. Introducing  $\mathbf{w}$  as the combined vector ( $\vec{w}_{\vec{r}} | \vec{r} \in \mathcal{N}$ ) and the average over the input distribution

$$\langle \dots \rangle = \int d^n v P(\vec{v}) \dots, \quad (4)$$

we may write

$$\Delta \vec{w}_{\vec{r}} = \epsilon \vec{\Phi}_{\vec{r}}(\mathbf{w}) + f_{\vec{r}}(\vec{v}), \quad (5)$$

where the systematic part

$$\vec{\Phi}_{\vec{r}}(\mathbf{w}) = -\langle h_{\vec{r},\vec{r}^*}(\vec{w}_{\vec{r}} - \vec{v}) \rangle \quad (6)$$

essentially is the average increment of the synaptic vectors with respect to the stimulus distribution  $P(\vec{v})$ . Hence the noise term is

$$f_{\vec{r}}(\vec{v}) = \epsilon h_{\vec{r},\vec{r}^*}(\vec{v} - \vec{w}_{\vec{r}}) - \epsilon \vec{\Phi}_{\vec{r}}(\mathbf{w}). \quad (7)$$

We have suppressed the remnant dependence of  $f$  on  $\mathbf{w}$  for the sake of clarity. In fact, we will see below that close to the stationary state the dependence on  $\mathbf{w}$  can be neglected in the noise so that we may consider Eq. (5) as a stochastic difference equation with additive noise,  $\Phi$  denoting the deterministic part.

This leads to a more detailed physical interpretation of the Kohonen algorithm valid in the vicinity of the topologically ordered state. Following Eq. (5) we may consider in this case the Kohonen algorithm as the dynamical equation of a system of  $N$  massless particles with space vectors  $\vec{w}_{\vec{r}}$ ,  $\vec{r} \in \mathcal{N}$  (the vector  $\vec{r}$  may be considered just an index numbering the particles) moving in a medium with normalized friction under forces given by  $\Phi$  and an additional stochastic force represented by  $f$ . The new feature here consists in the fact that the forces are explicitly specified by the  $\Phi$ , the explicit expressions for the forces being derived below. The analogy to physics can be made even more explicit by noting that for  $\epsilon$  sufficiently small the difference equation can be interpreted as a differential equation which identifies it with a physical Langevin equation. The transition to a stochastic differential equation can also be made rigorously for finite

$\epsilon$  if we present the stimuli not at discrete times (constant waiting time), but with the waiting time between two stimuli to be Poissonian distributed; cf. the analogous considerations for the derivation of a Fokker-Planck equation in Ref. [9].

In order to obtain explicit expressions for the deterministic part of the Langevin equation we follow Ref. [5], Eqs. (216)–(218) by introducing the probability that for the actual  $\mathbf{w}$  the neuron at site  $\vec{r}$  will become the winner

$$\hat{P}_{\vec{r}}(\mathbf{w}) = \int_{\mathcal{D}_{\vec{r}}(\mathbf{w})} d^n v P(\vec{v}), \quad (8)$$

with  $\mathcal{D}_{\vec{r}}$  denoting the domain of neuron  $\vec{r}$ , i.e., that region in input space which makes neuron  $\vec{r}$  the winner. The average value of  $\vec{v}$  in this domain is

$$\vec{v}_{\vec{r}} = \frac{1}{\hat{P}_{\vec{r}}(\mathbf{w})} \int_{\mathcal{D}_{\vec{r}}(\mathbf{w})} d^n v P(\vec{v}) \vec{v} \quad (9)$$

in terms of which

$$\vec{\Phi}_{\vec{r}}(\mathbf{w}) = -\sum_{\vec{r}^*} h_{\vec{r},\vec{r}^*}(\vec{w}_{\vec{r}} - \vec{v}_{\vec{r}^*}) \hat{P}_{\vec{r}^*}(\mathbf{w}), \quad (10)$$

where  $\vec{r}^*$  now is just a summation index.

$\Phi$  depends on  $\mathbf{w}$  in a complicated way. Manageable expressions are obtained close to the stationary state of the stochastic process  $\mathbf{w}(t)$  by Taylor expanding  $\vec{\Phi}$  in terms of the deviations  $\vec{u}_{\vec{r}} = \vec{w}_{\vec{r}} - \vec{w}_{\vec{r}}^0$ , where  $\vec{w}_{\vec{r}}^0$  denotes the average over the stationary state distribution. Then, we obtain from Eq. (5) omitting terms of order higher than two,

$$\Delta u_{\vec{r},m} = \epsilon \sum_{\vec{r}',m'} B_{mm'}(\vec{r},\vec{r}') u_{\vec{r}',m'} + \epsilon \sum_{\vec{r}',\vec{r}'',m',m''} C_{mm'm''}(\vec{r},\vec{r}',\vec{r}'') u_{\vec{r}',m'} u_{\vec{r}'',m''} + f_{\vec{r},m}(\vec{v}). \quad (11)$$

While the coefficients  $B_{m,m'}$  of the linear term in the case of an one dimensional chain receiving two dimensional inputs ( $n = 1$ ,  $d = 2$ ) have been obtained before, cf. [5], the nonlinear terms are treated in the Appendix. The exclusion of higher order terms seems justified by the fact that we are interested in a stability analysis with small values of the order parameters and hence of the  $\vec{u}_{\vec{r}}$ . Moreover, by tentatively carrying out the expansion up to third order we could show that these terms are negligible as compared to the third order terms produced from Eq. (11) in the Ginzburg-Landau equation; see below.

The evaluation of the correlation function of the noise corresponds to that of the diffusion matrix  $\mathbf{D}_{\vec{r},m,\vec{r}',m'}(\mathbf{w})$  [5], Eq. (219), taken at the stationary point  $\mathbf{w} = \mathbf{w}^0$ , where  $\vec{w}_{\vec{r}}^0 = \vec{v}_{\vec{r}}$ ,

$$\langle f_{\vec{r},m} f_{\vec{r}',m'} \rangle = \frac{\epsilon^2}{2sN} \sum_{\vec{r}^*} h_{\vec{r},\vec{r}^*} h_{\vec{r}',\vec{r}^*} (w_{\vec{r},m}^0 - w_{\vec{r}^*,m}^0)(w_{\vec{r}',m'}^0 - w_{\vec{r}^*,m'}^0) + \int_{\mathcal{D}_s(\mathbf{w})} (v_m v_{m'} - w_{\vec{r}^*,m}^0 w_{\vec{r}^*,m'}^0) P(\vec{v}) d\vec{v}. \quad (12)$$

Equation (12) is the lowest order term in the expansion in powers of  $u$  in the correlation function of the noise. Since this term is nonzero the noise is essentially additive for sufficiently small  $u$ .

In the symmetry breaking phenomenon to be discussed below the stationary state is not unique.  $\vec{w}_{\vec{r}}^0$  then corresponds to the unstable asymptotic state average the symmetry of which is broken. By introducing the Fourier transforms (mode amplitudes)

$$\vec{u}_{\vec{k}} = \frac{1}{\sqrt{Nd}} \sum_{\vec{r}} \exp(i\vec{k}\vec{r}) \vec{u}_{\vec{r}}, \quad (13)$$

where  $N$  is the number of neurons and the Taylor expansion of  $\Phi$  we rewrite Eq. (5) as

$$\Delta u_{\vec{k},m} = \epsilon \sum_{m'} B_{mm'}(\vec{k}) u_{\vec{k},m'} + \epsilon \sum_{\vec{q},m',m''} C_{mm'm''}(\vec{k},\vec{q}) u_{\vec{k}-\vec{q},m'} u_{\vec{q},m''} + f_{\vec{k},m}(\vec{v}). \quad (14)$$

Explicit expressions for the coefficients  $C_{ijk}(\vec{k}, \vec{q})$  are given in the Appendix for a specific example. The nonlinear term couples the amplitudes of all modes so that Eq. (14) is still very complicated. The nonlinear coupling of the amplitudes of all modes appeared as the essential mechanism of the self-organized structure formation.

### III. DERIVATION OF THE TIME-DEPENDENT GINZBURG-LANDAU EQUATIONS

For a discussion of Eq. (14) we study the case of the mapping of a two dimensional rectangle onto a one dimensional chain of neurons, i.e., the distribution of the input stimuli is

$$P(\vec{v}) = \begin{cases} \frac{1}{2sN} & \text{if } \vec{v} \in [-s, s] \times [0, N] \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

and  $\vec{k}, \vec{r}$  now are scalars  $k, r$  and the matrix  $\mathbf{B}$  is two dimensional; cf. Ref. [5]. If the height  $s$  of the rectangle is sufficiently small ( $s \ll \sigma$ ), i.e., if the input space is pseudo one dimensional, the chain is mapped onto the median which cuts the rectangle symmetrically into two parts and the virtual neuron sites are  $\vec{w}_r^0 = \begin{pmatrix} r \\ 0 \end{pmatrix}$ . For  $s > s_c$ , however, this situation is not stable any longer, i.e., the symmetric mapping is broken spontaneously and the image of the chain begins to fold into the input space.

In the limit of a long-ranged neighborhood function ( $1 \ll \sigma \ll N$ ) the matrix  $\mathbf{B}$  is diagonal [5],

$$B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad (16)$$

where the eigenvalue corresponding to the transversal mode is

$$\lambda_2(k) = \sqrt{2\pi\sigma} \left( \frac{2}{3} s^2 (1 - \cos k) \exp(-k^2 \sigma^2 / 2) - 1 \right) |N \quad (17)$$

and correspondingly for the longitudinal mode

$$\lambda_1 = \sqrt{2\pi\sigma} \left[ \left( \frac{1}{2} (1 + \cos k) - \sigma^2 k \sin k \right) \times \exp(-k^2 \sigma^2 / 2) - 1 \right] / N. \quad (18)$$

$\lambda_2(k)$  is seen to change sign at  $s = s_c \cong 2.02\sigma$  at a critical value of  $k = k_0 \cong \sqrt{2}/\sigma$ . Hence the mode with wave vectors of norm  $k_0$  would grow unrestrictedly for  $s > s_c$  if the nonlinear term would be missing in Eq. (14). With the nonlinear term we get a competition between the modes. Let us try an ansatz as a solution for Eq. (14) by assuming that there is just one mode winning the competition, i.e.,

$$|u_{k_0,2}| \gg |u_{k,i}|, \quad (19)$$

valid for  $i = 1, 2$  if  $k \neq k_0$  and for  $i = 1$  if  $k = k_0$ , and  $k_0$  being the wave number of the winner mode. The am-

plitude of the winning mode will be the order parameter  $\psi(t)$  in the phase transition from the straight line to the folded image of the neuron chain in the input space, i.e., the absolute value of the order parameter

$$\psi(t) = u_{k_0,2} \quad (20)$$

directly measures the amplitude of the sinusoidal folding. Moreover, we assume that the contribution to the coupling term involving the  $k = 0$  mode can be neglected. This is justified since  $u_{k,1}$  represents for  $k = 0$  just the longitudinal motion of the "center of gravity" of the neurons which may be assumed to be in rest.

The ansatz (19) allows us to derive from Eq. (14) self-consistently a closed equation of motion for  $\psi(t)$ . By virtue of (19) the only terms that can contribute in leading order to the nonlinear terms in (14) are those with  $q = 2k_0$  or  $q = 2\pi - k_0$ , i.e.,

$$C_{2m'm''}(k_0, 2k_0) u_{-k_0, m'} u_{2k_0, m''}$$

and

$$C_{2m'm''}(k_0, 2\pi - k_0) u_{2k_0 - 2\pi, m'} u_{2\pi - k_0, m''}.$$

Using the properties of  $C$  (see Appendix), (19), and the relation  $u_{2\pi - k, m} = u_{-k, m} = u_{k, m}^*$  the only contribution surviving in lowest order is

$$[C_{221}(k_0, 2k_0) + C_{212}(k_0, -k_0)] u_{2k_0, 1} u_{k_0, 2}^*. \quad (21)$$

Hence, using the diagonality of  $B$  from (14) we obtain in leading order

$$\Delta\psi = \epsilon \lambda_2(k_0) \psi + \epsilon [C_{221}(k_0, 2k_0) + C_{212}(k_0, -k_0)] u_{2k_0, 1} u_{k_0, 2}^* + f_{k_0, 2}(\vec{v}). \quad (22)$$

The accompanying equation for  $\vec{u}_{2k_0}$  obtained analogously from (14) reads

$$\Delta u_{2k_0, 1} = \epsilon \lambda_1(2k_0) u_{2k_0, 1} + \epsilon C_{122}(2k_0, k_0) \psi \psi + f_{2k_0, 1}(\vec{v}). \quad (23)$$

For sufficiently small  $\epsilon$  we may work in a continuous time approximation. For this purpose we introduce a stochastic force for real-valued  $t$  as

$$f_{k,n}(t) = \sum_{t'=0}^{\infty} f_{k,n}(\vec{v}_{t'}) \delta(t - t'), \quad (24)$$

where  $\vec{v}_{t'}$  is the input stimulus  $\vec{v}$  applied at time  $t'$ , where  $t' = 0, 1, 2, \dots$ . Now, for  $\epsilon$  sufficiently small, Eq. (23) may be written as

$$\dot{u}_{2k_0, 1} = \epsilon \lambda_1(2k_0) u_{2k_0, 1} + \epsilon C_{122}(2k_0, k_0) \psi \psi + f_{2k_0, 1}(t), \quad (25)$$

which in the stationary state is solved as

$$u_{2k_0, 1}(t) = \int_{-\infty}^t dt' U_{11}(t - t') [C_{122}(k_0) \psi(t') \psi(t') + f_{2k_0, 1}(t')], \quad (26)$$

where the matrix  $\mathbf{U}(t-t') = e^{-\mathbf{B}(2k_0)(t-t')}$ , i.e.,  $U_{nn} = e^{-(t-t')\lambda_n(2k_0)}$  was introduced. The time dependence of  $u(t')$  is purely noisy. Let us assume that the time scale of the order parameter fluctuations is well separated from the time scales of the stable modes specifying the decay of  $\mathbf{U}(t)$  which is justified for  $s$  sufficiently close to  $s_c$  (see below). Omitting the noisy contributions to the  $\psi$  term in Eq. (26) we obtain from (23)

$$u_{2k_0,1}(t) = \lambda_1^{-1}(2k_0)C_{122}(2k_0, k_0)\psi(t)\psi(t) + \int_{-\infty}^t dt' U_{11}(t-t')f_{2k_0,1}(t') \quad (27)$$

$$= \lambda_1^{-1}(2k_0)C_{122}(2k_0, k_0)\psi(t)\psi(t) + \phi_{k_0}(t), \quad (28)$$

where the noise  $\phi$  is now colored with a correlation time given by the decay time of the modes with wave vector  $2k_0$ , i.e.,  $\lambda_1^{-1}(2k_0)$  and  $\lambda_2^{-1}(2k_0)$ . However, staying sufficiently close to the critical point the characteristic times relevant to the order parameter can be made arbitrarily large, so that we may still consider the noise as a white one in the derivations of the order parameter equation given below.

Using the above results in Eq. (22) we obtain

$$\Delta\psi = \lambda_2(k_0)\psi + g(k_0)|\psi|^2\psi + F_{k_0}(t), \quad (29)$$

where

$$g(k_0) = [C_{221}(k_0, 2k_0) + C_{212}(k_0, -k_0)] \times \lambda_1^{-1}(2k_0)C_{122}(2k_0, k_0) \quad (30)$$

and

$$F_{k_0}(t) = f_{k_0}(\bar{v}(t)) + \phi_{k_0}(t). \quad (31)$$

In order to reproduce the standard form of the time-dependent Ginzburg-Landau equation we introduce a free energy as

$$\mathcal{H}(\psi) = \mathcal{H}_0 + \frac{1}{2}\lambda_2(k_0)|\psi|^2 - \frac{1}{4}g(k_0)|\psi|^4 \quad (32)$$

and write Eq. (29) as

$$\dot{\psi}(t) = -\frac{\partial}{\partial\psi^*}\mathcal{H}(\psi) + F(t), \quad (33)$$

which in fact may be viewed as the standard form of the time-dependent Ginzburg-Landau equation describing the time evolution of the amplitude of the winning mode which serves as the order parameter in the symmetry breaking phenomenon.

This equation is valid for values of  $s \approx s_c$ , i.e., just about the critical value of  $s$  where the dimensional conflict begins to make itself felt by a spontaneous symmetry breaking. For these values of  $s$  approximate the eigenvalue  $\lambda_2$  of the transversal mode (cf. [5]) as

$$\lambda_2 = (s - s_c)Q(k_0), \quad (34)$$

where

$$Q(k_0) = -\frac{\sqrt{8\pi\epsilon\sigma}}{3N}s_c(1 - \cos k_0)\exp(-k_0^2\sigma^2/2). \quad (35)$$

The evaluation of  $g$  is rather involved since it requires the expansion of  $\Phi$  in second order. We obtain for the

amplitude of the mode by means of Eqs. (29), (34), and (35) the following expression:

$$|\langle\psi\rangle| = 0 \quad \text{for } s \leq s_c, \quad (36)$$

$$|\langle\psi\rangle| = \alpha\sqrt{s - s_c} \quad \text{for } s > s_c, \quad (37)$$

where

$$\alpha = \sqrt{\frac{Q(k_0)}{g(k_0)}}. \quad (38)$$

The above theoretical expressions are typical for a second order phase transition from a state with order parameter  $\psi = 0$  into another state with order parameter  $\psi \neq 0$ .

#### IV. SIMULATIONS

In the following a series of computer experiments is described which have been performed in order to test the results of the Ginzburg-Landau approach. The simulations support clearly the validity of our theory; cf. Fig. 1. Moreover, our simulations reveal the existence of a further metastable state corresponding to a larger amplitude  $\psi$ . As discussed in the Appendix, this state is connected with the loss of the topographic properties of the map due to the intersection of non-nearest-neighbor bordering lines inside the input data distribution.

Including these effects into the free energy  $\mathcal{H}$  of Eq. (33) implies that the exact  $\mathcal{H}$  has not only a minimum

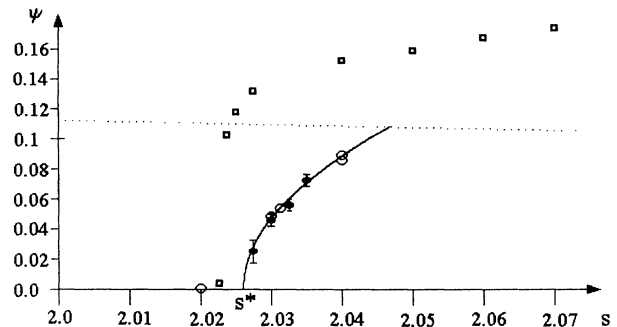


FIG. 1. The absolute value of the averaged order parameter  $\psi \equiv u_{k_0,2}$  corresponding to the amplitude of the winning mode vs the height  $s$  of the rectangular input space which measures the strength of the dimensional conflict and serves as the control parameter of the considered phase transition. The order parameter  $\psi$  measures the amplitude of the sinusoidal deviations of the synaptic vectors (images of the neurons in input space) from the line of symmetry which is the stable position of the synaptic vectors below the critical value of the order parameter  $s < s_c$ . The full line indicates theoretical results of the Ginzburg-Landau solution. Circles and boxes refer to computer experiments. Open symbols refer to a simulation with  $N = 222$ ,  $k^* = 10$ , filled ones to  $N = 89$ ,  $k^* = 4$ . As described in the text there are two types of states with broken symmetry. Circles refer to topology conserving states that are in accordance with the Ginzburg-Landau theory, whereas boxes denote states that violate topological order. These are reached by a first-order phase transition. The broken line restricts from above the range of validity of the Ginzburg-Landau solution.

at the Ginzburg-Landau solution [obtained from cutting off the power series expansion of the friction force  $\Phi$  in (10)] but also for still larger values of  $\psi$ . Hence there are two metastable solutions—a low- and a high-amplitude one. Correspondingly there are two kinds of phase transitions possible. On the one hand, there is a second order one to the low-amplitude Ginzburg-Landau state which is still topology preserving. On the other hand, there is a first order transition to the high-amplitude state with nonpreservation of topology. Both these effects are very clearly seen in our computer simulations. The first order transition will be discussed in more detail in a forthcoming paper.

In order to not become trapped immediately in the larger basin of attraction of the high-amplitude state, we allowed only very small learning parameters ( $\epsilon = 0.0001$ , constant). Periodic boundary conditions in the longitudinal direction have been used, and the number of neurons  $N$  was chosen such that in a good approximation an integer wave number  $\frac{k^*}{2\pi}N$  results. The neural chain was exposed to 20 000 000 stimuli. Most experiments displayed the following scenario: after an initial period the state remains close to the theoretical solution given by the Ginzburg-Landau theory. A few million steps later the system used to escape to another metastable state with higher-amplitude folding. Transitions from the high- to the low-amplitude state are expected to occur less frequently, and indeed have not been observed in the present experiments.

## V. CONCLUSION

In conclusion we remark that we have developed a theoretical description of the phase transition taking place in self-organizing topographic maps if the dimensional conflict between input space and net topology exceeds a certain critical value. By means of a single mode ansatz we could derive the time-dependent Ginzburg-Landau equations describing the behavior of the order parameter close to the critical point where the phase transition takes place.

In the present paper the derivation of the Ginzburg-Landau equation was explicitly performed for a simple example, namely the self-organizing map of a two dimensional homogeneous input space onto a one dimensional chain of neurons. This particular case was chosen to analyze explicitly the folding phenomenon, noticed for the first time in [2].

There are many problems for which the proposed approach may turn out to be of relevance. We mention the structure formation in the visual cortex [10] where there are several models which in computer simulations have proven to reproduce the phenomenon at least in principle. However, these simulations are very expedient and there is no general agreement as to the biological relevance of these models. Our theory may prove helpful here since it allows us to treat the ordering phenomena produced by different models in a common framework. Another perspective concerns the application of self-organizing feature maps in data analysis [11]. This includes on the one hand maps of high dimensional inputs onto a two

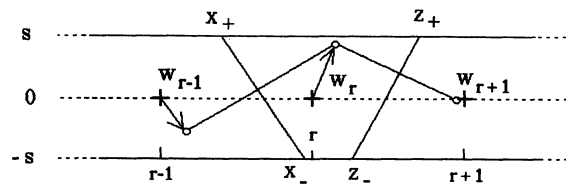


FIG. 2. The domains in a two dimensional rectangular input space of three neurons from a one dimensional chain. All neural domains border with nearest neighboring domains only or with the outer margin of the input set.

dimensional neural layer where the dimensional conflict may lead to more sophisticated patterns formed due to the symmetry breaking. In particular, investigation of behavior in the case of inhomogeneous (e.g., Gaussian) input distributions would be of interest for the practical applications of Kohonen's algorithm. A further challenge arises if the dimensional conflict is enhanced much beyond this point. Then, for sufficiently many neurons the neural vectors form a Peano-like curve in the input space. This effect can be explained as hierarchy of phase transitions similar to the case described here. The purpose of the present paper was, however, to present the analytical treatment of the most simple problem. Applications of the theory together with a more detailed analysis of the first-order phase transition will be discussed in a forthcoming paper.

## ACKNOWLEDGMENT

The reported results are based on work done in the project LADY sponsored by the German Federal Ministry of Research and Technology under Grant No. 01 IN 106B/3.

## APPENDIX: CALCULATION OF THE FRICTION COEFFICIENTS AND RANGE OF VALIDITY OF THE GINZBURG-LANDAU APPROACH

### 1. Calculation of the friction coefficients

In the following the coefficient matrices  $B$  and  $C$  in Eq. (14) will be determined in the case  $n = 1$ ,  $d = 2$ . For this purpose we Fourier transform Eq. (10). Consider

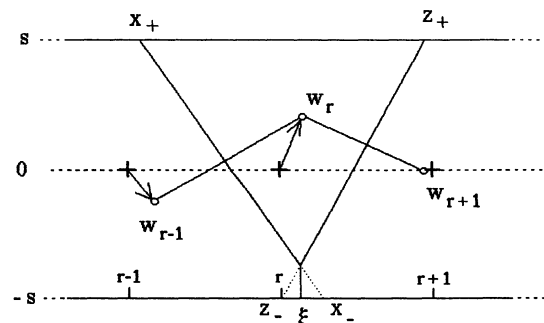


FIG. 3. Same as Fig. 2, but with non-nearest-neighbor neural domains having a common border region.

$$\begin{aligned} \mathcal{F}\left(\vec{\Phi}_r(\mathbf{w})\right) &= \mathcal{F}\left(-\sum_{r^*} h_{r,r^*}(\vec{w}_r - \vec{v}_{r^*})\hat{P}_{r^*}(\mathbf{w})\right) \\ &= \mathcal{F}\left[-\sum_{r^*} h_{r,r^*} \binom{r-r^*}{0} \hat{P}_{r^*}(\mathbf{w}) - \vec{u}_r \sum_r h_{r,r^*} \hat{P}_{r^*}(\mathbf{w}) + \sum_{r^*} h_{r,r^*} \vec{\partial}v_{r^*} \hat{P}_{r^*}(\mathbf{w})\right], \end{aligned} \quad (\text{A1})$$

where  $\vec{w}_r = \vec{w}_r^0 + \vec{u}_r = \binom{r}{0} + u_r$ , and  $\vec{v}_{r^*} = \vec{v}_{r^*}^0 + \vec{\partial}v_{r^*} = \binom{r^*}{0} + \vec{\partial}v_{r^*}$ ,

$$\begin{aligned} \mathcal{F}[\Phi_{r,n}(\mathbf{w})] &= -\frac{\partial h(k)}{\partial k} \mathcal{F}[\hat{P}_{r^*}(\mathbf{w})] \delta_{n,1} - \mathcal{F}(u_r) * \{h(k) \mathcal{F}[\hat{P}_{r^*}(\mathbf{w})]\} \\ &\quad + h(k) \{\mathcal{F}(\vec{\partial}v_{n,s}) * \mathcal{F}[\hat{P}_{r^*}(\mathbf{w})]\}, \end{aligned} \quad (\text{A2})$$

where  $f(i) * g(i) = \sum_j f(i-j)g(j) = \sum_j f(j)g(i-j)$  denotes the convolution, and assuming  $1 \ll \sigma \ll N$  the Fourier transform of  $h_{rs}$  is

$$h(k) = \mathcal{F}[h(r-r^*)] = 2\pi\sigma^2 \exp\left(-\frac{\sigma^2 k^2}{2}\right). \quad (\text{A3})$$

In order to find the Fourier transforms of the quantities  $\vec{\partial}$  and  $P_s(\mathbf{w})$  we rewrite these expressions. Figure 2 visualizes the situation.

Neglecting from now on terms higher than second order in the  $\vec{u}_r$  we find at a fixed position  $r$  for  $x_{\pm}$  and  $z_{\pm}$ :

$$x_{\pm}(r) = \pm s(u_{r-1,2} + u_{r,2})(1 + u_{r-1,1} - u_{r,1}) + \frac{1}{2}(u_{r,2}^2 - u_{r-1,2}^2 + 2r - 1 + u_{r-1,1} + u_{r,1}), \quad (\text{A4})$$

$$z_{\pm}(r) = \pm s(u_{r,2} + u_{r+1,2})(1 + u_{r,1} - u_{r+1,1}) + \frac{1}{2}(u_{r+1,2}^2 - u_{r,2}^2 + 2r + 1 + u_{r,1} + u_{r+1,1}), \quad (\text{A5})$$

where the weighted size of the domain of neuron  $r$  (8) is given by

$$\begin{aligned} \bar{P}(r) &= \int_{v \in \mathcal{D}(r)} P(\vec{v}) d\vec{v} \\ &= \frac{1}{2N} s[z_+(r) - x_+(r) + z_-(r) - x_-(r)] \\ &= \frac{1}{N} \left(1 - \frac{u_{r-1,1}}{2} + \frac{u_{r+1,1}}{2} + \frac{u_{r-1,2}^2}{2} - u_{r,2}^2 + \frac{u_{r+1,2}^2}{2}\right) \end{aligned} \quad (\text{A6})$$

and the mean value of the stimuli that make neuron  $r$  to be the winning unit (9) is given by

$$\vec{v}_r = \int_{v \in \mathcal{D}(r)} \vec{v} P(\vec{v}) d\vec{v}, \quad (\text{A7})$$

$$\begin{aligned} \bar{v}_{r,1} &= r + \frac{s}{3\bar{P}(r)} [z_+^2(r) + z_+(r)z_-(r) + z_-^2(r) - x_+^2(r) - x_+(r)x_-(r) - x_-^2(r)] \\ &= \frac{u_{r-1,1}}{4} + \frac{u_{r,1}}{2} + \frac{u_{r+1,1}}{4} - \frac{u_{r-1,1}^2}{8} - \frac{u_{r-1,2}^2}{4} + \frac{u_{r+1,1}^2}{8} + \frac{u_{r+1,1}^2}{4} - \frac{u_{r-1,1}u_{r,1}}{4} + \frac{u_{r,1}u_{r+1,1}}{4} \\ &\quad - \frac{s^2}{3} \left( \frac{u_{r-1,2}^2}{2} + \frac{u_{r+1,2}^2}{2} + u_{r-1,2}u_{r,2} - u_{r,2}u_{r+1,2} \right), \end{aligned} \quad (\text{A8})$$

$$\begin{aligned} \bar{v}_{r,2} &= \frac{s^2}{3\bar{P}(r)} \{[z_+(r) - x_+(r)] - [z_-(r) - x_-(r)]\} \\ &= \frac{s^2}{3} (-u_{r-1,2} + 2u_{r,2} - u_{r+1,2} - u_{r-1,1}u_{r-1,2} + u_{r,1}u_{r-1,2} + u_{r-1,1}u_{r,2} \\ &\quad - u_{r+1,1}u_{r,2} - u_{r,1}u_{r+1,2} + u_{r+1,1}u_{r+1,2}). \end{aligned} \quad (\text{A9})$$

Using the relations

$$u_{r,m} = \frac{1}{\sqrt{N}} \sum_k \exp(ikr) \hat{u}_{k,m}, \quad (\text{A10})$$

$$u_{r\pm 1,m} = \frac{1}{\sqrt{N}} \sum_k \exp(\pm ik) \exp(ikr) \hat{u}_{k,m}, \quad (\text{A11})$$

we collect the coefficients of terms of  $\mathcal{F}[\Phi_{r,n}(\mathbf{w})](k)$  that include  $u_{k-q,m} u_{q,m'}$  into  $C_{nm m'}(k, q)$ . The resulting expression is given in the case  $n = 1, d = 2, \vec{v} \in [0, N] \times [-s, s], P(\vec{v}) = \frac{1}{2sN}$  subject to the symmetry relations  $C_{mm'm''}(k, q) = C_{mm''m'}(k, k - q)$  as

$$\begin{aligned} C_{112}(k, q) &= C_{121}(k, q) = C_{211}(k, q) = C_{222}(k, q) = 0, \\ C_{111}(k, q) &= i \left( \frac{1}{2} [\hat{h}(k - q) \sin(k - q) + \hat{h}(q) \sin(q)] - \frac{1}{4} \hat{h}(k) [\sin(k) + \sin(k - q) + \sin(q)] \right), \\ C_{122}(k, q) &= i \left( \frac{\partial}{\partial k} \hat{h}(k) [\cos(k) - 1] + \frac{1}{2} \hat{h}(k) \sin(k) + \frac{s^2}{3} \hat{h}(k) [\sin(k) + \sin(k - q) + \sin(q)] \right), \end{aligned} \quad (\text{A12})$$

$$\begin{aligned} C_{221}(k, q) &= C_{212}(k, q) \\ &= i \left( \frac{1}{4} [\hat{h}(k - q) \sin(k - q) + \hat{h}(q) \sin(q)] - \frac{s^2}{3} \hat{h}(k) [\sin(k) - \sin(k - q) - \sin(q)] \right). \end{aligned}$$

## 2. Range of validity of the Ginzburg-Landau approach

In the above calculations of each neural domain we assumed that the border lines between the neuron domains do not cross inside the rectangle representing the input data distribution. For a larger amplitude  $\psi$  of the folding this will be violated. For a rough estimation of this effect we consider a continuous arrangement of neural domains in the input space  $(0, N) \times (-s, s)$  with transversal components

$$w_2(x) = \psi \cos(kx). \quad (\text{A13})$$

The now infinitesimally wide neural domains are bounded by straight lines orthogonal to the graph of the cosine function. In the case  $\psi = 0$  all borders lines are parallel. An effect is to be expected if the points of intersection that are present for  $\psi \neq 0$  fall inside the input space, as in the example given in Fig. 3. This will lead to the breakdown of the series expansion of the friction force in the generalized Langevin equation, so that for these values of  $\psi$  the Ginzburg-Landau approach is not valid. Whether or not this happens is determined by the

maximal curvature  $\frac{1}{\rho}$  of the chain,

$$\frac{1}{\rho} = \left\| \left. \frac{d^2 w_2(x)}{dx^2} \right|_{x=0} \right\| = \|\psi k^2 \cos(kx)|_{x=0}\|. \quad (\text{A14})$$

The geometrical assumption made above is correct only if  $\rho > s + \psi$ , i.e., using  $k^* = \frac{\sqrt{2}}{\sigma}$ , for

$$\psi < \psi_c = \frac{1}{2} (\sqrt{s^2 + 2\sigma^2} - s). \quad (\text{A15})$$

For relevant parameters  $\sigma = 5$  and  $s \approx 10$  the result obtained within the Ginzburg-Landau approach for a one dimensional chain in a rectangle of height  $s$  is valid as long as  $\psi < \psi_c \approx 0.1s$ , indicated by the broken line in Fig. 1. This boundary for the validity of the Ginzburg-Landau approach is also well reflected in our computer simulations. Moreover, our simulations clearly indicate the existence of a further metastable state for larger values of  $\psi$ . This state is characterized by the intersection of domain-bordering lines of non-nearest-neighbor neurons *inside* the data distribution (cf. Fig. 3). Consequently, input stimuli which are very close to each other may have non-neighboring winning neurons, i.e., the topology preservation of the mapping is lost.

- [1] J. J. Hopfield, *Proc. Natl. Acad. Sc. USA* **79**, 2554 (1982).
- [2] *Self-Organization and Associative Memory*, 3rd ed., edited by T. Kohonen, Springer Series in Information Sciences Vol. 8 (Springer, Berlin, 1989).
- [3] K. Obermayer, H. Ritter, and K. Schulten, *Parallel Computer* **14**, 381 (1990).
- [4] I. Csabai, T. Geszti, and G. Vattay, *Phys. Rev. A* **46**,

- R6181 (1992).
- [5] H. Ritter, T. Martinetz, and K. Schulten, *Neural Computation and Self-Organizing Maps* (Addison-Wesley, Reading, MA, 1992).
- [6] E. Erwin, K. Obermayer, and K. Schulten, *Biol. Cybern.* **67**, 35 (1992); **67**, 47 (1992).
- [7] H. Yin and N. M. Allinson, in *New Trends in Neural Computation*, edited by J. Mira, J. Cabestany, and A.



- Prieto (Springer, Berlin, 1993), pp. 291–296.
- [8] *Synergetics. An Introduction*, 3rd ed., edited by H. Haken, Springer Series in Synergetics Vol. 1 (Springer, Berlin, 1983).
- [9] T. Heskes and B. Kappen, *Phys. Rev. A* **44**, 2718 (1991).
- [10] K. Obermayer, G. Blasdel, and K. Schulten, *Phys. Rev. A* **45**, 7568 (1992).
- [11] R. Der, H. Englisch, M. Funke, and M. Herrmann, *Neural Networks World* **6**, 699 (1993).