Cortical Functionality Emergence: General Theory & Quantitative Results

Hans-Otto Carmesin, Institute for Theoretical Physics, University Bremen, 28334 Bremen, Germany, Fax: 0421 218 4869, email: Carmesin@theo.physik.uni-bremen-de, WWW: http://schoner.physik.uni-bremen.de/~carmesin/

In: Frank Schweitzer (Ed.): Self-Organization of Complex Structures: From Individual to Collective Dynamics, vol. I, chapt. 18, 215-233, London: Gordon and Breach, 1996.

Abstract: The human genotype represents at most ten billion binary informations, whereas the human brain contains more than a million times a billion synapses. So a differentiated brain structure is essentially due to self-organization. Such self-organization is relevant for areas ranging from medicine to the design of intelligent complex systems. Many brain structures emerge as collective phenomenon of a microscopic neurosynaptic dynamics: a stochastic dynamics mimics the neuronal action potentials, while the synaptic dynamics is modeled by a local coupling dynamics of type Hebb-rule, that is, a synaptic efficiency increases after coincident spiking of pre- and postsynaptic neuron. The microscopic dynamics is transformed to a collective dynamics reminiscent of hydrodynamics. The theory models empirical findings quantitatively: Topology preserving neuronal maps were assumed by Descartes in 1664; their self-organization was suggested by Weiss in 1928; their empirical observation was reported by Marshall in 1941; it is shown that they are neurosynaptically stable due to ubiquitous infinitesimal short range electrical or chemical leakage. In the visual cortex, neuronal stimulus orientation preference emerges; empirically measured orientation patterns are determined by the Poisson equation of electrostatics; this Poisson equation orientation pattern emergence is derived here. Complex cognitive abilities emerge when the basic local synaptic changes are regulated by valuation, emergent valuation, attention, attention focus or combination of subnetworks. Altogether a general theory is presented for the emergence of functionality from synaptic growth in neurobiological systems. The theory provides a transformation to a collective dynamics and is used for quantitative modeling of empirical data.

1 Introduction

The human brain consists of roughly 10^{12} neurons. A typical neuron transmits along its axon electric membrane potentials, most effectively by solitary wave like action potentials. The neurons are connected at so-called synapses, each transmitting an arriving electric potential to the neighboring neuron, with roughly 10000 synapses per neuron. So the brain establishes a neuronal network, the performance and functionality of which depends crucially on the synaptic connections. Hence the essential question is: Where do the synaptic connections come from?

While the biophysics of action potentials is understood quantitatively since Hodgkin and Huxley [13], that of synaptic change is hardly understood quantitatively. However an essential qualitative understanding is established by the so-called Hebb-rule [12, 16]: Synaptic change is provided by *local metabolic events* and thus depends on pre- and postsynaptic activity; in particular a synaptic efficiency increases after coincident pre- and postsynaptic activity. This simple rule is extended here in the same spirit by the influence of membrane potentials of neighboring neurons and analogously in [8] by slightly time delayed membrane potentials. This extended rule is sufficient for the analysis of the emergence and stabilization of many global synaptic structures, in quantitative agreement with experiment.

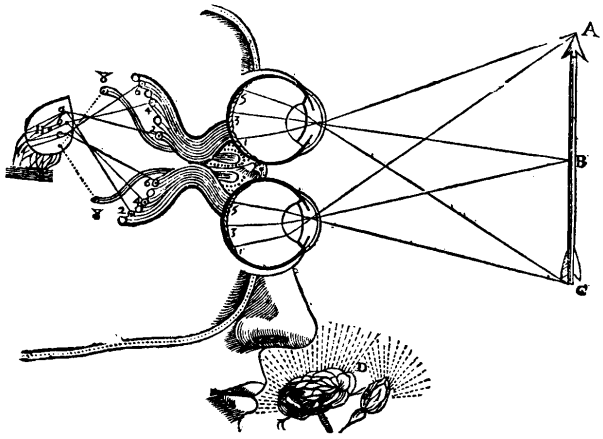

Figure 1: **Signal processing according to Descartes.** Scheme as indicated by Descartes in 1664. An arrow is projected onto the retina according to geometrical optics. Then it is projected further via axonic transmission to a cortical map. Thereby neighborhood relations are preserved.

2 Microscopic dynamics

For the present system of neurons and synapses, the microscopic dynamics is defined as the dynamics of neuronal and synaptic states.

2.1 Neuronal dynamics

The axonal transmission of membrane potentials over distances exceeding one centimeter is practically exclusively due to action potentials. Due to the dynamical nature of the latter, at any instant of time a neuron does either initiate an action potential or not. Thus it is adequate to model a neuron as a two state system

$$n_i(t) = \begin{cases} 1 & \text{action potential initiated;} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For simplicity, we model discrete time steps $t = 0, 1, 2, ...$ here; more detailed modeling of the time structure of membrane potentials including action potentials is possible, but not very relevant for the present investigation [8].

Next we characterize the stimulation $h_i(t)$ arriving at a neuron n_i at a time step t. A presynaptic neuron n_j, coupled to the neuron n_i via a synaptic efficiency alias coupling W_{ij} contributes an electrical potential $W_{ij}n_j(t-1)$ in the membrane of n_i; these potentials add up; so the considered stimulation is

$$h_i(t) = \sum_j W_{ij} n_j(t-1). \tag{2}$$

At this point we want to know, whether or not the postsynaptic neuron n_i exhibits an action potential alias spike as a consequence of a stimulation $h_i(t)$. If we neglect thermodynamic fluctuations, then the postsynaptic neuron n_i spikes when its membrane potential exceeds a threshold λ_i. The fraction of the membrane potential due to thermodynamic fluctuations is roughly equal to the product of Boltzmann constant and physical temperature T_{phys} divided by the electrostatic energy of the stimulated membrane potential h_i at the threshold [8]; at 310 Kelvin this fraction is typically one percent, so thermodynamic fluctuations are relevant for spiking. Formally, we consider fluctuations with a formal temperature T and a Boltzmann-type probability function for the neuronal state of the postsynaptic neuron [8]:

$$P[n_i(t)] = \frac{\exp\left(n_i(t)\frac{h_i(t)-\lambda_i}{T}\right)}{1+\exp\left(\frac{h_i(t)-\lambda_i}{T}\right)}. \tag{3}$$

In the following λ_i is not essential and set to zero.

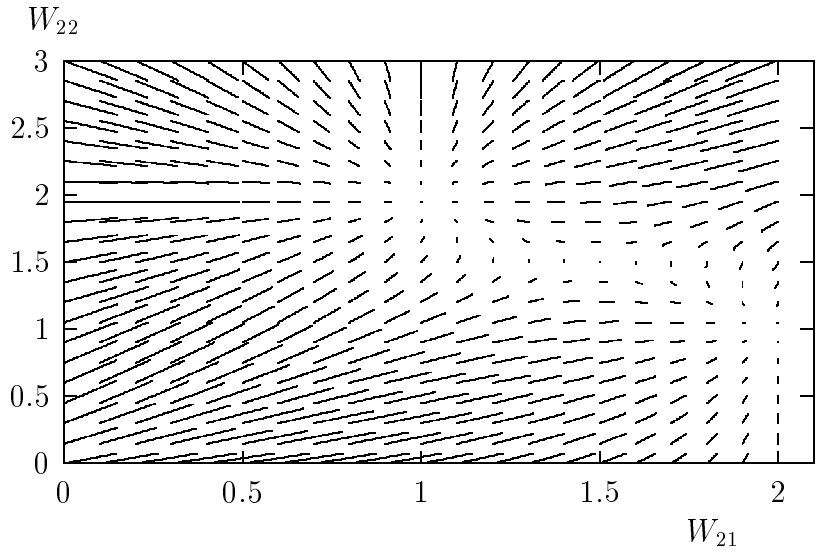

Figure 2: **Vector field.** Abscissa: Coupling weight W_{21}. Ordinate: Coupling weight W_{22}. For details see [2,8]. Vector field: Averaged changes $(\langle \Delta W_{21} \rangle, \langle \Delta W_{22} \rangle)$. Stationary states = emerging networks = zeros of the vector field. First network (I) at $(W_{21}, W_{22}) \approx (1, 2)$. Second network (II) at $(W_{21}, W_{22}) \approx (2, 1)$. Third network (III) at $(W_{21}, W_{22}) \approx (1.5, 1.5)$, not stable in contrast to (I) and (II).

2.2 Coupling dynamics

A presynaptic neuron n_j and a postsynaptic neuron n_i are connected at the coupling W_{ij}. Assuming that the coupling can increase by local metabolic events only, we express the coupling change $\Delta W_{ij}(t) := W_{ij}(t) - W_{ij}(t-1)$ as a function of the presynaptic state $n_j(t-1)$ at the previous time step and the postsynaptic state $n_i(t)$ at the current step, that is, $\Delta W_{ij}(t) = \Delta W_{ij}[n_i(t), n_j(t-1)]$. A power series expansion of this function yields four terms only, due to the fact that $n_i^2 = n_i$; thereby the only associative term is $n_i(t)n_j(t-1)$; because we are interested in the emergence of connections due to correlated activity, we keep this term only as an adequate approximation; thus we get the Hebb-type coupling change

$$\Delta W_{ij}^{Hebb}(t) = a n_i(t) n_j(t-1), \tag{4}$$

thereby a is the power series coefficient called *learning rate* in the following. Altogether Eqs. (1) to (4) establish the microscopic neurosynaptic dynamics.

3 Macroscopic dynamics

In many neurosynaptic systems, the number of neuronal events is very large ($\approx 10^{12}$ neurons, each spiking every few milliseconds). So large overall changes may take place and the resulting structure may exhibit a dynamical equilibrium state. Accordingly one may assume ergodicity and thus exchange time averages and ensemble averages, for details see [3, 2, 8]. Hence one may consider for each neurosynaptic state $(\vec{W}, \vec{n})$ the ensemble average $\langle (\Delta \vec{W}, \Delta \vec{n}) \rangle|_{(\vec{W},\vec{n})}$ of all changes of couplings and neurons; this establishes a vector field in the formal space of neurosynaptic states (for the case of two synaptic states only, such a vector field is presented in Fig. (2)).

This averaging procedure is analogous to weather forecasting using the wind dynamics, usually called hydrodynamics: Microscopically, weather forecasting might be based on calculating the trajectory according to Newton's axioms for each molecule of the wind; such a procedure is practically at best possible for few nanoseconds yielding quite a short forecasting period. Macroscopically one may instead first derive the averaged change of molecule positions (this is essentially the wind) and its dynamics from Newton's axioms, thus one gets the wind dynamics alias hydrodynamics as introduced first by Euler (see Fig. (3)).

Forecast of	Weather	Neuronal Adaptation
Micro-dynamics	Molecular motion (Newton 1686)	Single synapse change
Computable	Nanosecond	$\approx$ Hours
Averaging	$\downarrow$(Euler 1760)	$\downarrow$
Diff. eq. for	Wind	(Synaptic) change field
Computable	Days	Mathematical solutions
Novel phenomena	Wing	Topological charges

Figure 3: **Comparable Field Theories.** Both field theories are obtained by averaging the microscopic dynamics and yield predictability and novel phenomena.

Formally the averaged dynamics may be expressed in terms of the conditioned probability $P[\vec{n}(t+1)|\vec{n}(t),\vec{W}(t),T]$ of postsynaptic states at a time step $t+1$ for provided neurosynaptic states at the preceding time step t and fixed formal temperature T:

$$\langle(\Delta\vec{n},\Delta\vec{W})\rangle|_{(\vec{W},\vec{n})} = \sum_{\{n_i(t+1)\}}^{2^N}\prod_{i=1}^{N} P[n_i(t+1)|\vec{n},\vec{W},T]\ [\vec{n}(t+1)-\vec{n},\Delta\vec{W}(t)]. \tag{5}$$

Thereby N is the number of neurons. Here and in the following we often omit the time index t and express other time indices such as $t+1$; the above conditioned probabilities are determined by Eq. (3) and the above coupling changes $\Delta\vec{W}(t)$ are determined by Eq. (4). The above formula may be interpreted biophysically as follows: The events of subsequent pre- and postsynaptic neuronal spiking give rise to the coupling changes $\Delta\vec{W}(t)$ and are characterized globally and quantitatively by the conditioned probabilities $P[n_i(t+1)|\vec{n},\vec{W},T]$; both are combined and averaged by the sum $\sum_{\{n_i(t+1)\}}^{2^N}$ over all possible neuronal configurations. In a continuous time limit one may express the above difference equation by a differential equation [2, 8].

Adiabatic limit. In most neuronal systems, the synaptic changes per time are relatively small compared to the neuronal changes per time [16]. Accordingly one may analyze the dynamics in the so-called adiabatic limit [11, 10]. In this limit, a representative number of presynaptic neuronal states $\vec{n}(t)$ occurs, before the couplings change significantly; so one may perform the average over the presynaptic neuronal states in the above equation. For this purpose we introduce the probability function $P[\vec{n}(t)]$ for the neuronal states of the presynaptic neurons, so we get

$$\langle\Delta\vec{W}\rangle|_{\vec{W}} = \sum_{\vec{n}}^{2^N} P[\vec{n}] \sum_{\{n_i(t+1)\}}^{2^N}\prod_{i=1}^{N} P[n_i(t+1)|\vec{n},\vec{W},T]\Delta\vec{W}(t). \tag{6}$$

The above equation explicates the so-called [2, 8] synaptic change field or *change field* for short, for an illustration see Fig. (2).

4 Retinotopy

Phenomenon of topology preservation. Many brain areas are organized as so-called cortical maps, especially simple are cortical maps with each map neuron receiving input directly or indirectly from a sensor neuron of an area of sensor neurons; examples are the retinotopic maps (see Fig. (1)) with sensor neurons in the retina and somatosensory maps with sensor neurons at the skin [16]. In these maps neighboring sensor neurons project to neighboring map neurons (up to non-generic discontinuities), that is, the *topology is preserved*.

Local neurosynaptic self-organization of topology preservation. In order to model the retinotopy emergence we introduce a prototypical 1D model (for a general dimension and

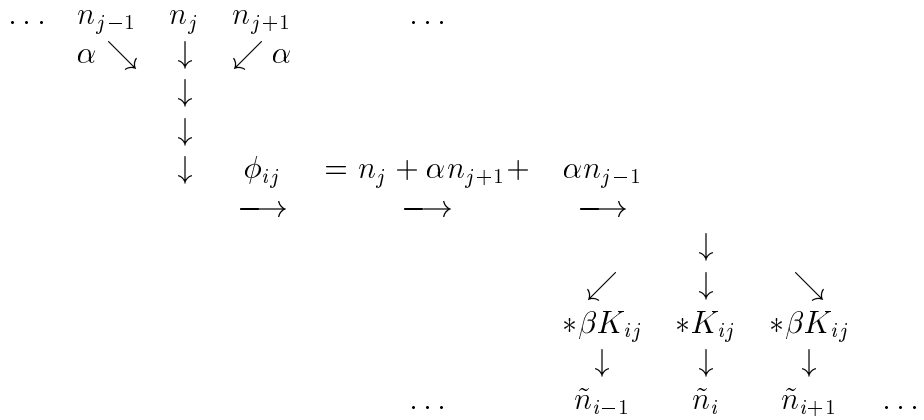

Figure 4: **Network architecture.** Illustration for the special case of the one dimensional model. n_j: Input neuron. $\tilde{n}_i$: Cortical map neuron. ϕ_{ij}: Axonic membrane potential. α: Presynaptic lateral contribution parameter. β: Postsynaptic lateral contribution parameter. K_{ij}: Coupling.

topology see [2, 8, 9, 7, 5]) with N sensor neurons n_j, N inner neurons $\tilde{n}_i$ and couplings K_{ij} from sensor neurons to inner neurons; we do not consider boundary effects. We generalize the coupling dynamics (see Eq. (4)) as follows. First we assume that neighboring presynaptic neurons contribute proportional to a presynaptic lateral contribution parameter α to the membrane potential ϕ_{ij} of the coupling K_{ij} (see Fig. (4)); analogously we assume that the membrane potential is transferred to the postsynaptic neuron $\tilde{n}_i$ proportional to the coupling K_{ij} and to the neighbors of $\tilde{n}_i$ with a postsynaptic lateral contribution parameter β as additional proportionality factor. Second we assume that the coupling increase is proportional to the current coupling (typical for biomass growth), to the learning parameter a and to the pre- and postsynaptic signals, so we get

$$\Delta K_{ij}^{Hebb} = 2aK_{ij}[\tilde{n}_i + \beta(\tilde{n}_{i+1} + \tilde{n}_{i-1})][n_j(t-1) + \alpha(n_{j-1}(t-1) + n_{j+1}(t-1))]. \tag{7}$$

Third we assume constant total coupling biomass at each neuron as follows

$$\sum_{i=1}^{N} 2K_{ij} = r^2 = \sum_{j=1}^{N} 2K_{ij}. \tag{8}$$

Altogether the topology preservation network architecture and neurosynaptic dynamics are defined. Next we transform the couplings according to $K_{ij} = W_{ij}^2/2$ and accordingly $\Delta K_{ij} = W_{ij}\Delta W_{ij}$. The resulting change field is a scalar potential of a so-called *change potential* V as follows.

Potential theorem. *For the above neurosynaptic dynamics (see Eqs. (1) to (8)) and in the adiabatic limit, the mean coupling change (see Eq. (6)) is the gradient of a scalar potential as follows:*

$$\langle \Delta W_{ij} \rangle = -\frac{\partial V}{\partial W_{ij}} \tag{9}$$

with the change potential

$$V = -\frac{aT}{2^N} \sum_{\mu}^{2^N} P[\mu] \ln Z^{\mu}, \text{ where the stimulation } \vec{n} \text{ is denoted by } \mu \tag{10}$$

and the formal partition functions, one for each μ,

$$Z^{\mu} = \sum_{\{\tilde{n}_i(t+1)\}}^{2^N} \exp[-H^{\mu}/T] \tag{11}$$

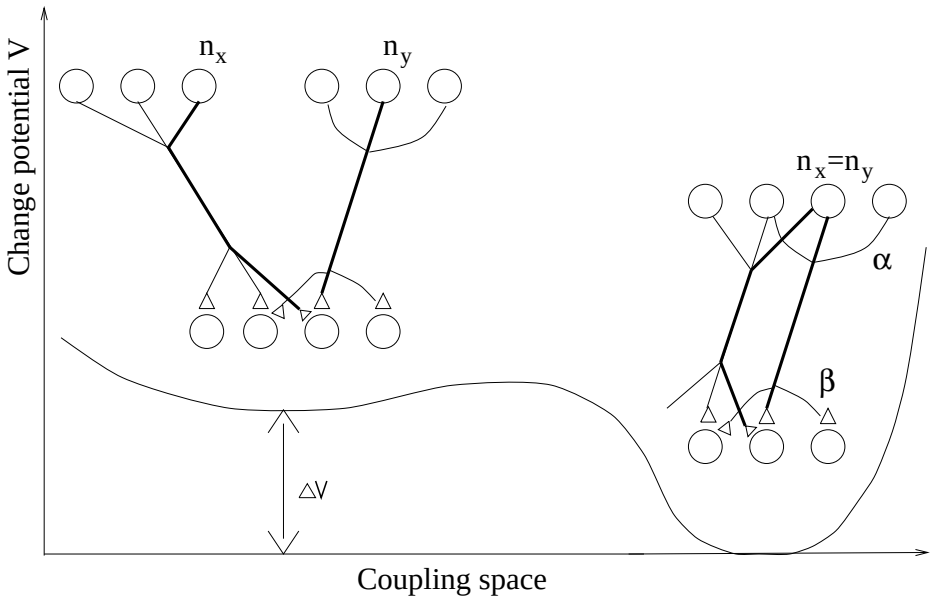

Figure 5: **Change potential difference.** Abscissa: Coupling space schematically. Ordinate: Change potential V. Left: Local change potential minimum without coincidence of signals along thick lines. Right: Change potential minimum due to coincidence as indicated by thick lines.

and with the formal energy functions

$$H^{\mu} = \sum_{i}^{N} h_i^{\mu}(t+1)\tilde{n}_i(t+1), \quad \text{where} \quad h_i^{\mu}(t+1) = \sum_{j}^{N} K_{ij} n_j|_{\mu}. \tag{12}$$

Accordingly, the stable emerging networks are the local minima of the scalar potential V.
For a proof see [2, 8, 9, 7, 5]. The structure of the solution is a sum of formal partition functions Z^{μ} as they are known from statistical physics. Nevertheless the system is by no means a system of equilibrium statistics, because it is open due to the stimulation of sensor neurons. In the rest of the paper, we consider random stimulation of sensor neurons with equal probability for spiking and for not spiking.

Bijective mapping theorem. *Each coupling state that is locally stable with respect to the neuronal fluctuations and with respect variations of the formal temperature T exhibits exactly one nonzero coupling from each sensor neuron and exactly one nonzero coupling to each inner neuron and thus establishes a one-to-one mapping from sensor to inner neurons.*
This result is proven in [2, 8, 9, 7, 5] and may be easily understood as follows: Large couplings grow most rapid (see Eq. (7)) and the total coupling weight at each neuron is constant (see Eq. (8)); thus only one coupling remains at each neuron; hence a one-to-one mapping is established globally.

Coincidence stabilization. So far we considered the emergence of a unique coupling at each neuron. Next we consider the relations among emerging couplings. For this purpose we consider an elementary key mechanism for the formation of related couplings, the stabilization of those coupling states that provide additional coincident signal transfers (see Fig. (5)). To begin with, one may express the change potential (see Fig. (10)) as a sum over *single neuron potentials* [8] as follows

$$V = \sum_{i}^{N} V_i \quad \text{with} \quad V_i = -\frac{aT}{2^N} \sum_{\mu} \ln\left(1 + \exp[\tilde{h}_i(t+1)/T]\right). \tag{13}$$

By explicating the local stimulation h_i (see Eq. (2)) and denoting by Q the number of presynaptic neurons that transfer signals to a postsynaptic neuron directly or indirectly (see Fig.

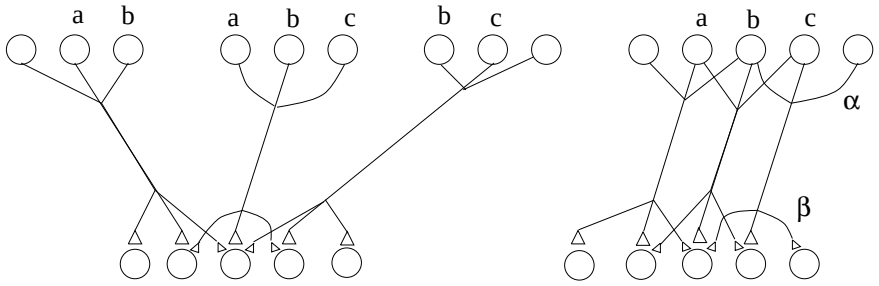

Figure 6: **Topology preservation in the 1D-network.** LEFT: Unrelated presynaptic neurons. RIGHT: Presynaptic neurons with four neuron identities as indicated minimize the change potential and provide topology preservation.

(5)) we get [8]

$$V_i = -\frac{aT}{2^Q} \sum_{n_q=0/1}^{2^Q} \ln\left(1 + \exp\left[\frac{\sum_q^Q \alpha_q n_q}{T}\right]\right) \; with \; \alpha_q \in \{0, 1, \alpha, \beta, \alpha\beta\}. \tag{14}$$

Potential difference lemma. *The single neuron change potential V_i in Eq. (14) with two unequal neurons $n_q = n_x$ and $n_q = n_y$ is decreased by the potential difference*

$$\Delta V_i = \frac{aT}{2^Q} \sum_{n_q=0/1; n_q \neq n_x; n_q \neq n_y}^{2^{Q-2}} \ln\left(1 + \frac{(e^{\alpha_x/T} - 1)(e^{\alpha_y/T} - 1)}{e^{-R/T} + e^{\alpha_x/T} + e^{\alpha_y/T} + e^{(\alpha_x+\alpha_y+R)/T}}\right), \tag{15}$$

if the two neurons n_x and n_y are made identical. Hereby R represents terms due to the other neurons; i.e. $R = \sum_{x \neq q \neq y} \alpha_q n_q$.
A proof is presented in [8]. For positive factors α_x and α_y, the potential difference ΔV_i is positive. Here these factors are positive (see Eq. (14)). *So the network with $n_x = n_y$ is more stable than that with $n_x \neq n_y$ (see Fig. (5)) due to a decrease of the change potential.* For the case of topology preservation (see Fig. (4)) the change potential is minimal, when the number of identical presynaptic neurons $n_x = n_y$ (see Fig. (5)) is maximal; in the 1D model the number Q of neurons that transmit signals to a postsynaptic neuron is nine (see Figs. (4) and (6)); by considering all possible arrangements of the nine couplings from these neurons one may easily derive that the number of resulting identical presynaptic neurons $n_x = n_y$ is obtained when the topology is preserved [2, 8, 9, 7, 5]; so we get:

1D topology preservation theorem. *For a globally stable coupling state that is also stable with respect to temperature variations holds: The neighborhood is necessarily preserved if and only if both lateral contribution parameters α and β are positive.*

Discussion. Earlier modeling of such self-organization of topology preservation [22, 17, 23] used some global mechanisms such as a so-called 'winner takes all mechanism' (i. e. only the map neuron with maximal stimulation spikes) or some lateral inhibition among distant map neurons; here the dynamics is completely local at a neuron; this fact and the modeled dynamics may be regarded as an adequate modeling of a third ontogenetic step of cortical maps; while in the first step neurons migrate to target positions and in the second step axons grow roughly to target areas, both according to chemical or other markers [16]. Moreover, these earlier models do not exhibit a potential function, while the present change potential is embedded in the theory of statistical physics and in a coherent theory of neuronal adaptation [2, 8].

The topology preservation stabilization is similar to Huygens pendulum clocks that are hanging at the same wall and thus coupled slightly and exhibit a common phase after a while: An infinitesimal coupling gives rise to global order. As a consequence, the ubiquitously observed

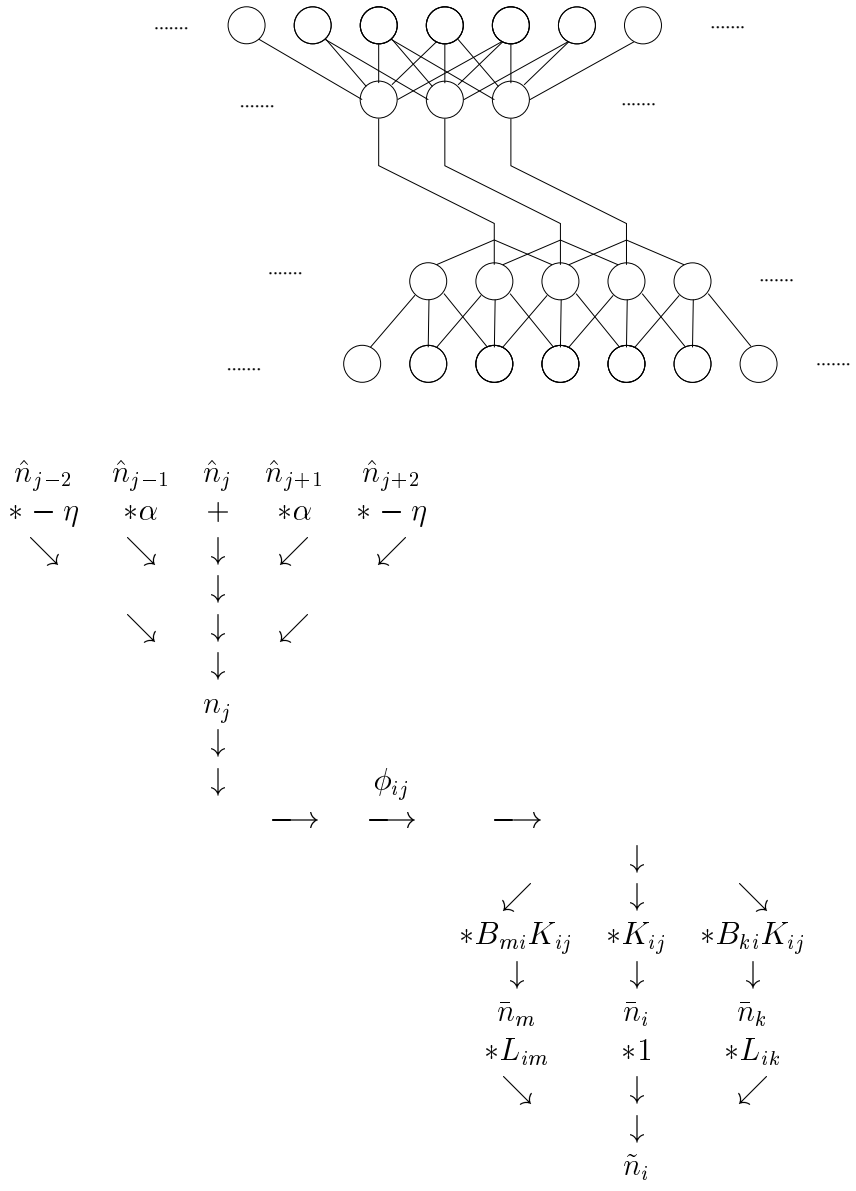

Figure 7: **Network architecture.** Monocular model. TOP: Full network architecture for the 1D case. BOTTOM: All signal pathways arriving at $\tilde{n}_i$ are explicated: n_j: Input neuron. $\tilde{n}_i$: Cortical map neuron. ϕ_{ij}: Axonic membrane potential; explicitly $\phi_{ij} = \hat{n}_j + \alpha(\hat{n}_{j-1} + \hat{n}_{j+1}) - \eta(\hat{n}_{j-2} + \hat{n}_{j+2})$. α: Presynaptic lateral contribution parameter for ON-center stimulation. η: Presynaptic lateral contribution parameter for OFF-surround stimulation. K_{ij}: Coupling. B_{mi} and L_{im}: Further couplings. For 1D-architecture: $m = i - 1$ and $k = i + 1$. β is neglected since it is regarded small compared to B_{ki}.

topology preserving brain maps (see Fig. (1)) may be due to ubiquitously present electrical or chemical leakages to neighbors, here modeled with α and β.

5 Orientation preference

Phenomenon of orientation preference. In mammals, the visual signals are transmitted from the retina via the lateral geniculate nucleus (LGN) to the primary visual cortex, in particular to the areas 17 and 18 (Brodmann notation) [16]. In these areas there are neurons that respond preferentially to a stimulation by edges with a particular orientation, as was discovered by Hubel and Wiesel [14, 15].

Model for local random stimuli based neurosynaptic self-organization of orientation preference. In order to model orientation preference emergence, we generalize the network architecture of Fig. (4) to that specified in Fig. (7), moreover we model a square lattice of neurons in each layer. In particular we introduce a retinal input layer with neurons $\hat{n}_j$

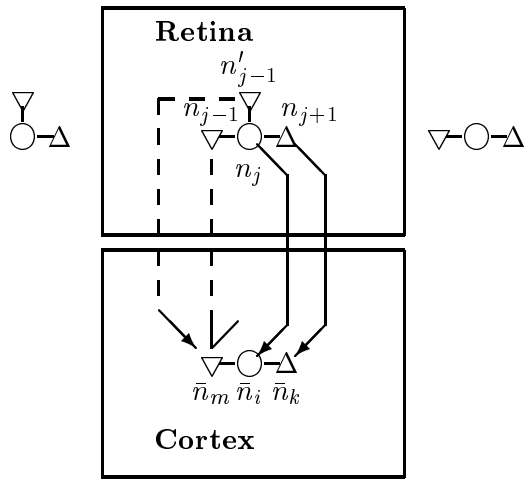

Figure 8: **Orthogonal versus parallel presynaptic arrangement.** Top: Retina. Bottom: Cortex. Lower circle: Considered cortical neuron. Upper circle: Corresponding presynaptic retinal neuron. Lower Δ, ∇: Cortical neighbor neurons. Upper Δ, ∇: Corresponding presynaptic neurons. Upper ∇: Occurs either in the same direction as Δ or in an orthogonal direction.

stimulated randomly. The retinal output neurons take values according to

$$n_j = \hat{n}_j + \alpha \sum_{k \in \nu(\hat{n}_j)} \hat{n}_k - \eta \sum_{k \in \nu'(\hat{n}_j)} \hat{n}_k, \tag{16}$$

where $\nu(\hat{n}_j)$ is the set of neighbors of $\hat{n}_j$ while $\nu'(\hat{n}_j)$ is the set of next nearest neighbors of $\hat{n}_j$. The cortical neurons $\bar{n}_i$ and $\tilde{n}_i$ are stimulated according to the couplings K_{ij}, B_{mi} and L_{im} indicated in Fig. (7) and spike according to the Boltzmann type probability specified in Eq. (3). The couplings change according to a general local dynamics [8] as follows

$$\Delta K_{ij} = 2aK_{ij}n_j(t-1)\Big[\bar{n}_i + \sum_{m \in \nu(\bar{n}_i)} B_{mi}\bar{n}_m\Big]. \tag{17}$$

$$\Delta B_{mi} = 2aB_{mi}K_{ij}n_j(t-1)\bar{n}_m; \tag{18}$$

the couplings L_{im} are set equal to B_{mi} (so the coincident stimulus transmission is realized) and there are two couplings B_{mi} at a coupling K_{ij} (see Fig. (7)) with

$$\sum_i^N 2B_{mi} = B^2 \quad \text{and} \quad \sum_m^N 2B_{mi} = B^2. \tag{19}$$

Results. As a first result, one may derive a change potential for this system [8]; here we derive the main result by investigating essential network architectures and by applying the above potential difference lemma (see p. 4): The two neurons related by B_{mi} to a cortical input neuron are shown in Fig. (8); due to the bijective mapping theorem there occur three corresponding presynaptic neurons being neighbors according to the topology preservation theorem and establishing either an orthogonal arrangement or a parallel arrangement. In the case of the parallel arrangement, the postsynaptic neuron spikes preferentially when a bar directed along the presynaptic arrangement is presented as stimulus, whereas no orientation preference occurs in the orthogonal arrangement [2, 8].

Next we investigate the stability of the parallel and orthogonal arrangements. Stable arrangements exhibit minimal change potential; for simplicity we consider relatively large OFF-surround parameter η here. Then such minima occur at a minimum of identical presynaptic next nearest neighbor neurons, because such identical neurons give rise to a negative potential difference ΔV_i equivalent to an effective repulsive interaction among the presynaptic neurons transmitting signals to the same postsynaptic neuron directly or indirectly; this is due to the

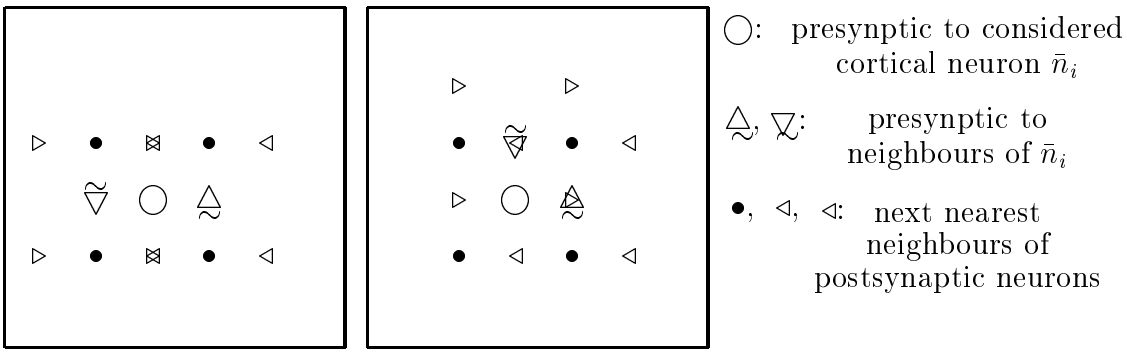

Figure 9: **Orthogonal and parallel presynaptic neighbor arrangements.** Retinal arrangement of presynaptic neurons and their *next nearest* neighbors. Presynaptic neurons as in the above figure. Next nearest neighbors analogous to neighbors in the above figure. Essence: Effective *repulsive interaction* in the orthogonal arrangement due to coincidence of presynaptic with next nearest neighbor neurons; but effective *attractive interaction* in the parallel arrangement due to coincidence of next nearest neighbor neurons.

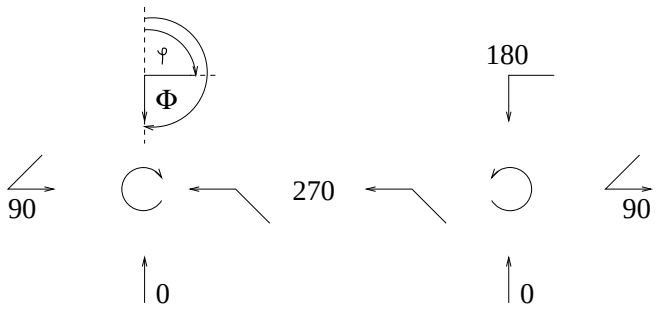

Figure 10: **Pinwheel-order illustrated with vectors and orientations.** Orientations, indicated in the form of lines without arrows, explicated for the eight neurons in the immediate vicinity of the topological singularities. Round arrows: Topological singularities of previous figure. Vectors: Indicators for the polar angles Φ that obey the Poisson equation. Additional lines without arrows: Observed orientation preferences; these are constructed from the polar angle $\varphi = \Phi/2$; whereby the angles Φ obey the Poisson equation.

negative sign of η in Fig. (7), see potential difference lemma. Such a minimum of identical neurons transmitting signals to the same postsynaptic neuron is achieved by the parallel arrangement, see Fig. (8). Thus the parallel arrangement is stabilized.

Discussion. The orientation preference of a cortical input neuron $\bar{n}_i$ is induced by the lengthy arrangement of retinal neurons transmitting signals to $\bar{n}_i$ directly or indirectly. Such a lengthy alias parallel arrangement is stabilized by an effective repulsive force calculated with the potential difference lemma and based on the OFF-surround activity of retinal neurons [2, 8]. The cortical output layer is not yet needed here but will become essential in the next part about orientation patterns. Some other models of orientation preference emergence did use lengthy stimuli [19]; such stimulation is unreasonable because orientation preference emerges before birth. Other models [20] do neither identify effective repulsive forces nor provide a soluble model.

6 Orientation patterns

Discovery. Bonhoeffer and Grinvald observed area 18 in the cat's brain with an electronic

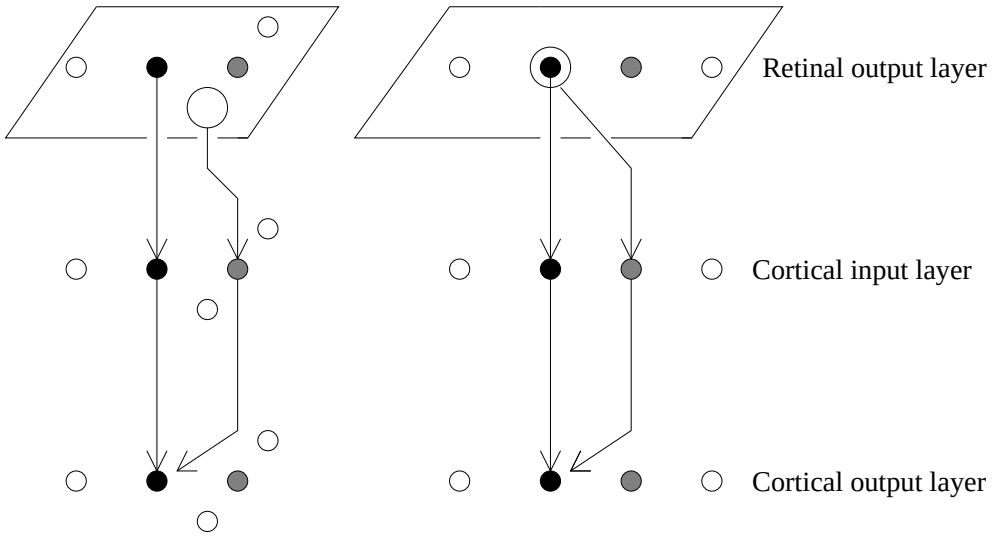

Figure 11: **T-shape versus straight-line.** Three layers occurring in two connectivity arrangements: T-shape versus orthogonal. Full circles: Two considered cortical neurons with their presynaptic neurons. Big circle: Emphasized retinal neuron. Arrows: Emphasized signal transfers in the two arrangements from the big circle via the same neuron in the cortical input layer to the same neuron in the cortical output layer. Essence: Random stimulation gives rise to additional signal coincidences in the cortical output layer in the straight-line configuration only. This induces an effective force towards the straight-line configuration, according to the potential difference lemma.

camera, using light with wavelength 600 nanometer [1]. At this wavelength, one observes oxygenated, but no unoxygenated hemoglobin; therefrom one obtains a picture of the neuronal activity. For eight orientations Bonhoeffer and Grinvald stimulated with bars of that orientation, took the corresponding activity picture, determined for each area 18 location the orientation of maximal neuronal activity and produced a map of these activities (see Fig. (10)). This map exhibits regions of smoothly varying orientation preferences and singular points with all orientation preferences in their neighborhood. At these singular points, the orientations vary by the full amount of π in a regular clockwise or counter-clockwise manner (see Fig. (10)).

Neurosynaptic stabilization of pinwheels. We analyze further the network model introduced above. As derived so far, each cortical output neuron is related to a cortical input neuron related in turn to a parallel arrangement of retinal output neurons as indicated in Fig. (11). In this figure, two neighboring cortical output neurons are investigated. Their corresponding parallel arrangements in the retina exhibit either a T-shape or a straight-line arrangement.

Next we study the stability of these arrangements for the case in which the cortical coupling parameter B (see Eq. (19)) is large compared to the retinal parameters α and η. In this case one may neglect the influence of the retinal neighbors (see Fig. (7)); so Fig. (11) contains all neurons contributing significantly to the stabilization of T-shape or straight-line arrangements. The stable arrangement exhibits the lower change potential. Thus it exhibits coincident signal processing among the relevant neurons (see Fig. (11)). Hence the straight-line arrangement is stable, because it provides coincident signal processing (see Fig. (11)). For a quantitative analysis one may introduce a two neuron change potential $V_{ij} := V_i + V_j$, a change potential V_{ij}^T of the T-shape arrangement and a change potential $V_{ij}^|$ of the straight-line arrangement and the difference $\Delta V_{ij} := V_{ij}^| - V_{ij}^T$, an *effective orientation potential*. The latter may be calculated according to the potential difference lemma and is $\Delta V_{ij} < 0$.

Generalization of the lattice model. So far we considered a square lattice of neurons in

each layer. Next we allow any neuron positions in each layer and note that the whole analysis may be performed analogously and yields similar results due to the continuity inherent to the present theory. Thus one gets in the generalized theory similar results for possible T-shape arrangements and straight-line arrangements and one gets interpolating results for orientation angle differences $\phi_i - \phi_j$ in between. This interpolation is performed adequately with a multipole expansion yielding the quadrupolar term as leading order, due to the π-periodicity of orientation preferences. So we get the orientation potential

$$V_{ij}(\varphi_i - \varphi_j) = \Delta V_{ij} \cos^2(\varphi_i - \varphi_j). \tag{20}$$

Angle transformation. For the purpose of a later analysis, it is convenient to transform the orientation angles φ_i into their twofold values $\Phi_i = 2\varphi_i$. These twofold angles Φ_i are the polar angles of corresponding arrows as indicated in Fig. (10). The transformed orientation interaction may be determined by transforming the cosine, so one gets

$$V_{ij}(\Phi_i - \Phi_j) = \frac{1}{2}\Delta V_{ij} \cos(\Phi_i - \Phi_j) + \underbrace{\frac{1}{2}\Delta V_{ij}}_{irrelevant\ constant}. \tag{21}$$

The constant in the above equation is irrelevant (because it does not give rise to a nonzero derivative) and neglected in the following. The cosine is the scalar product of planar vectors $\vec{v}_i$ and $\vec{v}_j$ indicated by the arrows in Fig. (10).

Total orientation interaction. Next we sum up all orientation interactions of the neighboring orientations. So we get the formal orientation energy as follows.

$$H = \frac{1}{2} \sum_{\langle i,j \rangle} V_{ij}. \tag{22}$$

Here the brackets below the sum indicate the summation over next nearest neighbors and the factor 1/2 compensates the double summation of pairs. This formal energy is that of the so-called x-y-model [18]. By definition, an x-y-model is a model of interacting dipoles.

Resulting orientation patterns. It has been shown for the above x-y-model formal energy [18, 8] that the vector angles Φ_i obey the Poisson equation

$$2\pi\rho(\vec{r}) = \frac{\partial^2 \Phi}{\partial \vec{r}^2} \tag{23}$$

(ρ = charge density) of electrostatics with the clockwise topological charges (see Fig. (10)) corresponding to positive electrical charges and with counter-clockwise topological charges corresponding to negative electrical charges, or vice versa. This establishes a quantitative equivalence between planar electrostatic systems and orientation patterns [6, 8]; this is in quantitative agreement with experimental findings.

In particular, the resulting patterns depend on a formal orientation temperature, that is induced by the neuronal orientations via the Hebb-rule, for a quantitative analysis thereof see [8]. At zero fluctuation rate $T = 0$ all orientations are parallel (see Fig. (12)); below a critical fluctuation rate T_c neutral pairs of topological charges emerge (see Fig. (12)); while the topological charges are located randomly above T_c. So far only randomly placed topological charges have been observed, according to the present theory we predict a phase transition at sufficiently low fluctuation rate (see Fig. (12)).

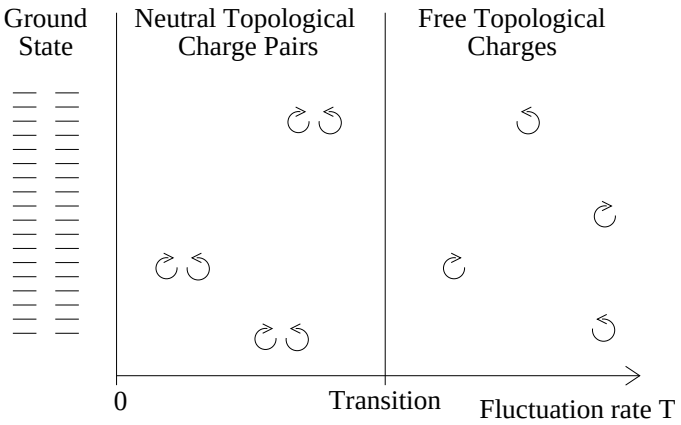

Figure 12: **Phase diagram.** Abscissa: Fluctuation rate T. Partial circles with arrows: Indicators for topological charges. High T: Randomly placed topological charges. Intermediate T: Neutral pairs of charges. Zero T: Parallel orientations.

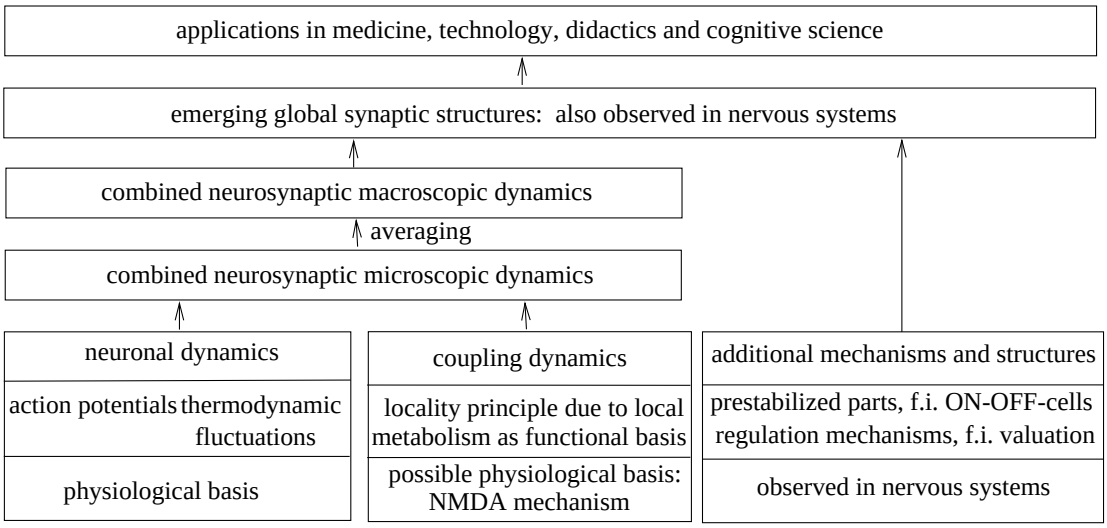

Figure 13: **The physical bases of the present neuronal adaptation theory.** Left lower box: Neuronal dynamics. Middle lower box: Coupling dynamics. Right lower box: Additional mechanisms relevant for emerging structures. Two boxes of intermediate length: neurosynaptic dynamics. Upper box: Resulting emergent synaptic structures. Arrows: Indication of consequences.

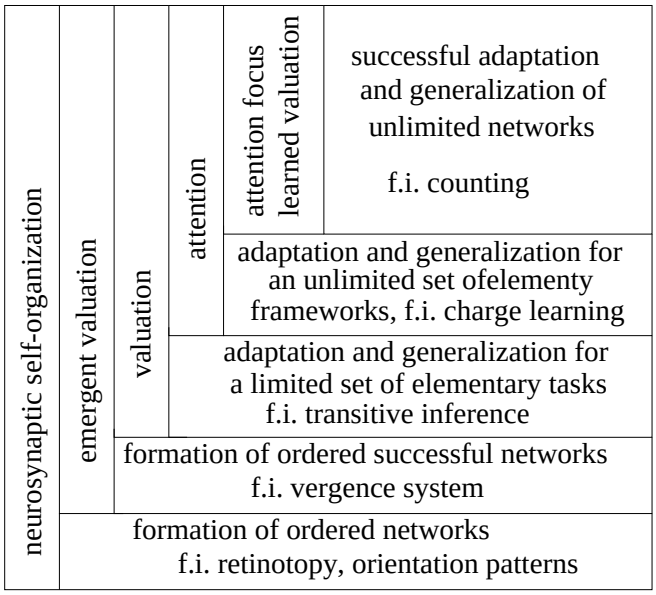

Figure 14: **Overview for regulation mechanisms.** Vertical boxes: Adaptation processes. Horizontal boxes: Resulting adaptations and generalizations with examples.

Discussion. The present derivation of the orientation patterns is in quantitative agreement with experiment and may be regarded as a success of this neurosynaptic theory, based on the biologically plausible locality assumption for synaptic change and on the biophysically established action potentials always influenced by thermodynamic membrane potential fluctuations [8]. Other models of orientation patterns are not based on such simple and biologically plausible model assumptions [24, 21] and do not provide such quantitative agreement with experiment, analytical treatment and proof of equivalence to planar electrostatic systems as a result. A particularly beautiful aspect of the present theory is that the Poisson equation may be derived by purely topological considerations here, see f. i. [8], while in electrostatics the physical nature of electrical charges must be assumed.

7 Outlook and discussion

Based on a stochastic neuronal dynamics modeling spikes and on a Hebb-type coupling increase after coincident pre- and postsynaptic spiking extended by a slight influence of neighboring neurons (see Fig. (13)), we showed the ubiquitous emergence of topology preservation, the emergence of orientation preference and the emergence of orientation patterns with topological charges equivalent to electrical charges. More generally, one may extend the local Hebb-type dynamics further by investigating the influence of small slightly time delayed signals [8], of valuation mechanisms, attention, attention focus and combining of sub-networks; as a result one obtains emerging networks with increasing cognitive skills such as successful sensor-motor performance, operant conditioning, generalization, transitive inference, learning of abstract objects such as electrical charges, learning of counting without limitation as a possible solution of Wittgenstein's paradox (see Fig. (14)). Altogether I developed and investigated the present neurosynaptic theory with the mentioned extensions in my habilitation thesis [2] and in more detail and depth in my new book [8]; the investigation of such neurosynaptic systems may be regarded as a research program, actually performable analytically and numerically with many proven non-trivial quantitative agreements with experimental observations [4, 2, 8], with basic results in neurophysics, biology and cognitive sciences and with applications ranging from medicine to intelligent systems technology. **Acknowledgements.** I am grateful for fruitful discussions with Stefan Arndt, Hans Flohr, Olaf Scherf, Helmut Schwegler and Fred Wolf.

References

[1] T. Bonhoeffer and A. Grinvald. Iso-orientation domains in cat visual cortex are arranged in pinwheel- like patterns. *Nature*, 353:429–431, 1991.

[2] H.-O. Carmesin. *Theorie neuronaler Adaption.* Köster, Berlin, 1994. ISBN 3-89574-020-9; 2. Ed. 1996.

[3] H.-O. Carmesin. Neurophysics of adaption. *Physics Essays*, 8(1):38–51, 1995.

[4] H.-O. Carmesin. Statistical neurodynamics: A model for universal properties of EEG-data and perception. *Acta Physica Slovaca*, 44:311–330, 1994.

[5] H.-O. Carmesin. Topological order in networks selforganized through local dynamics. In F. G. Böbel and T. Wagner, editors, *ICASSE Proceedings*, pages 53–60, Erlangen, 1994. Fraunhofer - Gesellschaft.

[6] H.-O. Carmesin. Self-organization of pinwheels in the visual cortex by stochastic Hebb dynamics: Equivalence to the Kosterlitz-Thouless model. *Acta Physica Slovaca*, 45(2):93–102, 1995.

[7] H.-O. Carmesin. Topology preservation by Hebb-rule with infinitesimal short range signals. Technical Report 4/95, ZKW Universität Bremen, Bremen, 1995.

[8] H.-O. Carmesin. *Neuronal Adaptation Theory.* Peter Lang, Frankfurt am Main, 1996. ISBN 3-631-30039-5.

[9] H.-O. Carmesin. Topology preservation emergence by Hebb rule with infinitesimal short range signals. *Phys. Rev. E*, 53(1):993–1003, 1996.

[10] W. Ebeling and R. Feistel. *Physik der Selbstorganisation und Evolution.* Akademie Verlag, Berlin, 1986.

[11] H. Haken. *Advanced Synergetics.* Springer, Berlin, 1983.

[12] D. O. Hebb. *The Organization of Behaviour.* Wiley, New York, 1949.

[13] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (London)*, 117:500–544, 1952.

[14] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.*, 148:574–591, 1959.

[15] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154, 1962.

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

[17] T. Kohonen. *Self-Organization and Associative Memory.* Springer, Berlin - Heidelberg, 1989.

[18] J.M. Kosterlitz and D.J. Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys, C*, 6:1181–1203, 1973.

[19] C. Malsburg, von der. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100, 1973.

[20] K. D. Miller. A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on-center-off-surround inputs. *J. Neuroscience*, 14, 1994.

[21] M. Stetter, A. Müller, and E. W. Lang. Neural network model for the coordinated formation of orientation preference and orientation selectivity maps. *Phys. Rev. E*, 50:4167–4181, 1994.

[22] W. Weiss. Eine neue Theorie der Nervenfunktion. *Die Naturwissenschaften*, 16:626–636, 1928.

[23] D. J. Willshaw and C. von der Malsburg. How patterned neural connections can be set up by self- organization. *Proc. Royal Soc. London*, B194:431–445, 1976.

[24] F. Wolf, K. Pawelzik, T. Geisel, D.-S. Kim, and T. Bonhoeffer. "Ein Feldenergiemodell zur Beschreibung von Orientierungspräferenzkarten im Visuellen" Kortex. In Deutsche Physikalische Gesellschaft, editor, *Verhandlungen DPG (VI)*, volume 29, pages 955–956. DPG, DPG, 1994.