Self-organization of orientation maps in a formal neuron model using a cluster learing rule

doi: 10.1016/S0893-6080(99)00081-7

Self-organization of orientation maps in a formal neuron model using a cluster learning rule

Jousuke Kuroiwa[†] , Sakari Inawashiro[††] ,
Shogo Miyake[‡] and Hirotomo Aso[§]

[†] Division of Mathematical and Information Science, Hiroshima University, Higashi-Hiroshima, 739-8521, JAPAN

[††] Department of Computer Science, Nihon University, Koriyama, 963-8642, JAPAN

[‡] Department of Applied Physics, Tohoku University, Sendai, 980-8579, JAPAN

[§] Department of Electrical Communications, Tohoku University, Sendai, 980-8579, JAPAN

Abstract

Self-organization of orientation maps due to external stimuli in the primary visual area of the cerebral cortex is studied in a two-layered neural network which consists of formal neuron models with a sigmoidal output function. A cluster learning rule is proposed as an extended Hebbian learning rule, where a modification of synaptic connections is influenced by an activation of neighboring output neurons. By making use of self-consistent Monte Carlo method, we evaluate output responses of neurons against explicit inputs after the learning. An orientation map calculated from the output responses reproduces characteristic features of biological ones. Moreover quantitative analysis of our results are consistent with those of experimental results. It is shown that the cluster learning rule plays an important role in forming smooth changes of preferred orientations.

Keywords: Self-organization, orientation map, external stimuli, neuron model, cluster learning rule

1 INTRODUCTION

Hubel and Wiesel (1962, 1968, 1974) found that neurons in the primary visual area preferably respond to visual stimuli of oriented slits or bars and the preferred orientations change continuously as a microelctrode moves through the cortex parallel to its surface. They proposed an arrangement of preferred orientations which is characterized by linear changes of preferred orientations along one direction.

Recently optical recording techniques have been drastically developed, and have been used in experiments, which revealed various characteristic features of orientation preference maps. (Blasdel and Salama, 1986; Blasdel, 1992a, 1992b; Bonhoeffer and Grinvald, 1991, 1993). In these maps, preferred orientations change smoothly in all directions except for regions of pinwheel singularities and fractures. The pinwheel singularity is a point-like region around which orientation preferences change by 180 $^\circ$ along a closed path. There appear two kinds of singularty, +1/2 and -1/2 singularity: preferred orientations increase with counterclockwise (clockwise) motion around the $+1/2$ $(-1/2)$ pinwheel singularity. About the same number of $\pm 1/2$ pinwheel singularities are present in the orientation maps.

Von der Malsburg (1973) theoretically studied the self-organization of orientation selectivity for explicit input patterns using a model of a formal neuron of a linear output function with a threshold. However pinwheel singularities were not clearly formed. Of course the existence of pinwheel singularities in the map was not known at that time.

Linsker(1986), Miyashita and Tanaka (1992), and Miller (1994) investigated the formation of an orientation preference map with random

spontaneous activity provided as input. However, explicit external stimuli were not used, and the nonlinearity of the output function was not considered. Obermayer et al. (1992) produced an orientation preference map by the application of SOFM (self-organizing feature map) technique (Kohonen, 1987), where an input stimulus was represented by a set of five parameters instead of a definite pattern.

Although orientation preference is weakly organized before birth (Nicholas et al., 1992), an influence of external inputs after birth can not be ignored in sharpening and maintaining orientation preference. For instance, deprivation of vision in one eye in young kittens causes the map for this eye in area 18 to vanish (Kim and Bonfoffer, 1994; Gödecke and Bonfoffer, 1996). It is important to investigate whether a plausible orientation map containing definite pinwheel similarities is formed due to external input stimuli in a formal neuron model with a nonlinear output function.

The purposes of the present paper are to clarify the problem whether formal neurons and external stimuli can produce a plausible map of orientation preference. In particular, we study the following two problems: (i) Can a naive model of the Malsburg type produce the orientation map ? (ii) If not, what kind of improvements on the naive model are required in order to produce the map ? Here we adopt a formal neuron model that is based on a discrete version of the nerve field model of Takeuchi and Amari (1979).

In Section 2, a two-layered neural network model, a cluster learning rule and basic equations are described. In Section 3, a Hamiltonian formalism (Inawashiro et al., 1996) is briefly presented. In Section 4, results obtained by a self-consistent Monte Carlo method is presented, and the the-

oretical and experimental orientation maps are compared qualitatively and quantitatively. In Section 5, the roles of the cluster learning rule and of the inhibitory neuron pool are discussed. Section 6 is devoted to conclusions.

2 NEURAL NETWORK MODEL WITH INHIBITORY NEURON POOL

2.1 Model

We consider a two-layered neural network model with an inhibitory neuron pool (Amari and Takeuchi, 1978; Takeuchi and Amari, 1979) as shown in Figure 1. An input layer consists of L presynaptic neurons and an output layer of N postsynaptic neurons. We use a formal neuron model of McCulloch and Pitts (1943) with a sigmoidal output function.

(figure 1)

Input neurons simply send input signals to output neurons. An output neuron i receives three kind of inputs; (i) an explicit external input X_k from the input neuron k through an excitatory synaptic connection s_{ik}, (ii) an inhibitory input X_0 from the inhibitory neuron pool through an inhibitory synaptic connection s_i, and (iii) a feedback contribution from a neighboring neuron j through a lateral connection w_{ij}. For simplicity it is assumed that the inhibitory neuron pool provides a constant inhibitory input, $X_0 = 1$. The indexes i and k denote a two-dimensional vectors indicating the location of each neuron within the output and input layer, respectively.

An inner state of a neuron i is described by an averaged membrane potential u_i, referred to as a membrane potential for simplicity hereafter.

We assume that the time dependence of the membrane potential is given by

$$\tau' \frac{du_i}{dt} = -u_i + \sum_j w_{ij} z_j + \sum_k s_{ik} X_k - s_i X_0, \tag{1}$$

where τ' denotes a time constant of the membrane potential of order of milliseconds. Here z_i represents an output of a neuron i defined by

$$z_i = f(u_i - u_{\rm th}), \tag{2}$$

where $u_{\rm th}$ denotes a threshold for the excitation of a neuron. The output function $f(x)$ is defined by a sigmoidal function,

$$f(x) = \frac{1}{1 + \exp(-2\beta x)}, \tag{3}$$

where $(\beta/2)$ represents a gradient at $x = 0$. We assume that a time constant of the membrane potential, τ', is much smaller than the time duration of presentation of an input pattern. In this situation, the steady value of the membrane potential is immediately achieved, and is given by

$$u_i = \sum_j w_{ij} z_j + \sum_k s_{ik} X_k - s_i X_0. \tag{4}$$

It is assumed that the synaptic connections s_{ik} and s_i are modified through the learning based on a Hebbian learning rule and a cluster learning rule described in Subsection 2.2. The self-organization of an orientation map is accomplished when the modification reaches saturation. The lateral connection w_{ij} is a function of the distance between output neurons i and j. We use a "Mexican-hat" interaction defined by

$$w_{ij} = (E + I) \exp(-\frac{|i-j|^2}{2r_E^2}) - I \exp(-\frac{|i-j|^2}{2r_I^2}), \tag{5}$$

which is excitatory for a pair of nearby neurons and inhibitory for a pair of distant ones. Here E denotes an excitatory strength, I an inhibitory

strength, r_E an excitatory range, and r_I an inhibitory range. The lateral connection is assumed to remain fixed during modification of the synaptic connections. A self-feedback connection w_{ii} is assumed to be zero.

2.2 Cluster Learning Rule

In a Hebbian learning rule, a synaptic connection s_{ik} between an input neuron k and an output neuron i is strengthened by simultaneous activation of the neurons as shown in Figure 2(a). The synaptic modification is not directly influenced by an activation of neighboring output neurons.

(figure 2)

Now, we propose a "cluster learning rule", in which the synaptic modification of s_{ik} is strengthened by a simultaneous activation of an input neuron k and an output neuron i together with its neighboring neurons as shown in Figure 2(b). The cluster learning rules for the synaptic connections s_{ik} and s_i are given by

$$\tau \frac{ds_{ik}}{dt} = -s_{ik} + c_1 z_i X_k + c'_1 \sum_j e_{ij} z_j X_k, \tag{6}$$

$$\tau \frac{ds_i}{dt} = -s_i + c_2 z_i X_0 + c'_2 \sum_j e_{ij} z_j X_0, \tag{7}$$

where e_{ij} represents a contribution factor from an output neuron j to an output neuron i, τ a learning time constant, and c_1, c'_1, c_2 and c'_2 are learning constants which control learning efficiencies of the synaptic connections. We assume e_{ij} to be positive and e_{ii} to be zero. In the right hand side of equations (6) and (7) the first term denotes the decay effect which ensures the saturation of the synaptic connections, the second term a increment due to the Hebbian learning rule, and the third term a con-

tribution due to the cluster learning rule. The role of the cluster learning rule is explained in Section 5.

2.3 Steady State of the Learning

We assume that an input pattern X^μ labeled by μ ($\mu = 1, 2, ..., P$) is presented to the output layer through the input layer. For simplicity we take bar-shaped patterns as input patterns X^μ although we think that more natural images are appropriate. In this study we are interested in mechanisms of the network itself such as the inhibitory neuron pool and the cluster learning rule. If a plausible map of orientation preference is produced in our model, the mechanism of our network will work for more appropriate external inputs. The pattern $\mathbf{X}^\mu$ is chosen at random from an input ensemble $\{\mathbf{X}\}$ which contains all the patterns $\mathbf{X}^\mu$ with the uniform probability, $1/P$. The learning constant τ is assumed to be much larger than the time required for presenting a full set of the patterns. In this situation, the synaptic connections are little modified during the presentation of an input pattern, and the learning equations (6) and (7) are approximated by an ensemble average over the input ensemble (Geman, 1979). Moreover if input patterns are presented continuously for a much longer duration than the learning time constant, a steady state of the learning will be attained and steady values of the synaptic connections are given by (Inawashiro et al., 1996)

$$S_{ik} = \frac{c_1}{P} \sum_\mu Z_i^\mu X_k^\mu + \frac{c'_1}{P} \sum_\mu \sum_j e_{ij} Z_j^\mu X_k^\mu, \tag{8}$$

$$S_i = \frac{c_2}{P} \sum_\mu Z_i^\mu X_0 + \frac{c'_2}{P} \sum_\mu \sum_j e_{ij} Z_j^\mu X_0, \tag{9}$$

where the capital letters denote the quantities in the steady state, and Z_i^μ represents a steady output of a neuron i against the μth input pattern.

After the system reached the steady state, the learning is stopped and response of the system against input patterns is studied. Using equations (2), (4), (8) and (9), a steady output of a neuron i against the μth input pattern is given by

$$\begin{aligned} Z_i^\mu &= f\left(\sum_j w_{ij} Z_j^\mu + \frac{1}{P}\sum_\nu (c_1 v_{\mu\nu} - c_2 X_0 X_0) Z_i^\nu \right. \\ &\quad \left. + \frac{1}{P}\sum_\nu \sum_j (c'_1 v_{\mu\nu} - c'_2 X_0 X_0)\, e_{ij} Z_j^\nu - u_{\rm th}\right) \end{aligned} \tag{10}$$

where $v_{\mu\nu}$ represents the spatial correlation between two input patterns $\mathbf{X}^\mu$ and $\mathbf{X}^\nu$ defined by

$$v_{\mu\nu} = \sum_k X_k^\mu X_k^\nu. \tag{11}$$

In general a self-correlation $v_{\mu\mu}$ takes a much larger positive value than the others. The learning constants c_1, c'_1, c_2 and c'_2 are chosen so that $(c_1 v_{\mu\mu} - c_2 X_0 X_0)$ and $(c'_1 v_{\mu\mu} - c'_2 X_0 X_0)$ are positive. If a neuron i would be excited by the μth input pattern, the large value of $(c_1 v_{\mu\mu} - c_2 X_0 X_0)$ plays an important role in maintaining the excitation. In the right hand side of equation (10), the third term comes from the cluster learning. We are interested in the response property of the neurons which is represented by a solution $\{Z_i^\mu\}$ of (10).

3 STATISTICAL MECHANICAL DESCRIPTION

3.1 Mean Field Equations

We introduce a set of new variables $m_{i\mu}$ by

$$Z_i^\mu = \frac{1}{2}(1 + m_{i\mu}). \tag{12}$$

Using a hyperbolic tangent function, the steady outputs (10) are rewritten in terms of $m_{i\mu}$ as

$$\begin{aligned} m_{i\mu} \;=\; \tanh\beta \Bigg(& \sum_{j\neq i} J_{ij}^{(xy)} m_{j\mu} + \sum_{\nu\neq\mu} J_{\mu\nu}^{(z)} m_{i\nu} \\ & + \sum_{j\neq i}\sum_{\nu} J_{ij\mu\nu}^{(xyz)} m_{j\nu} + h_{i\mu} \\ & + h_{i\mu}^{\text{self}} + h_{i\mu}^{\text{self}} m_{i\mu} \Bigg). \end{aligned} \tag{13}$$

Here $J_{ij}^{(xy)}$, $J_{\mu\nu}^{(z)}$ and $J_{ij\mu\nu}^{(xyz)}$ are exchange interactions defined by

$$J_{ij}^{(xy)} = \frac{1}{2} w_{ij}, \tag{14}$$

$$J_{\mu\nu}^{(z)} = \frac{c_1}{2P}(v_{\mu\nu} - \frac{c_2}{c_1} X_0 X_0), \tag{15}$$

$$J_{ij\mu\nu}^{(xyz)} = \frac{c'_1}{2P}(v_{\mu\nu} - \frac{c'_2}{c'_1} X_0 X_0) e_{ij}, \tag{16}$$

and $h_{i\mu}$ and $h_{i\mu}^{\text{self}}$ are local effective fields defined by

$$h_{i\mu} = \sum_{j\neq i} J_{ij}^{(xy)} + \sum_{\nu\neq\mu} J_{\mu\nu}^{(z)} + \sum_{j\neq i}\sum_{\nu} J_{ij\mu\nu}^{(xyz)} - u_{\text{th}} \tag{17}$$

$$h_{i\mu}^{\text{self}} = J_{ii}^{(xy)} + J_{\mu\mu}^{(z)} + J_{ii\mu\mu}^{(xyz)}. \tag{18}$$

The equations (13) represent mean field equations for an Ising spin system at a fixed temperature $T = 1/\beta$ in statistical physics of magnetism

(Ashcroft and Mermin, 1987), and $m_{i\mu}$ is a mean field average of an Ising spin located at a lattice site $i\mu$ in a three-dimensional space where i denotes xy coordinates and μ z coordinate. Output responses of neurons given by a solution of the steady outputs (10) are evaluated from a solution of the mean field equations (13) through the equations (12).

It is noted that $J_{ij}^{(xy)}$ represents an exchange interaction between two spins located at $i\mu$ and $j\mu$ on the same xy plane, $J_{\mu\nu}^{(z)}$ an exchange interaction between two spins located at $i\mu$ and $i\nu$ along the same z axis, and $J_{ij\mu\nu}^{(xyz)}$ an exchange interaction between two spins located at $i\mu$ and $j\nu$ $(i \neq j)$ in the xyz space. In the right hand side of equation (13) the last term $h_{i\mu}^{\rm self} m_{i\mu}$ is a kind of a mean field at $i\mu$ which is proportional to the mean field average of the Ising spin $i\mu$ itself, and is called a "self-field". The self-field does not appear in an ordinary form of mean field equations in physics. We believe that the self-field represents one of biological features of self-organization in a formal neuron model with a sigmoidal output function. It is to be noted that the self-field acts upwards because the self-correlation $v_{\mu\mu}$ takes a large positive value.

3.2 Hamiltonian Formalism

We consider an Ising Hamiltonian in the three-dimensional lattice space (Inawashiro et al., 1996),

$$\begin{aligned} H &= -\frac{1}{2}\sum_{i}\sum_{\mu\neq\nu}\sum_{\nu} J_{\mu\nu}^{(z)}\sigma_{i\mu}\sigma_{i\nu} \\ &\quad -\frac{1}{2}\sum_{i\neq j}\sum_{j}\sum_{\mu} J_{ij}^{(xy)}\sigma_{i\mu}\sigma_{j\mu} \\ &\quad -\frac{1}{2}\sum_{i\neq j}\sum_{j}\sum_{\mu}\sum_{\nu} J_{ij\mu\nu}^{(xyz)}\sigma_{i\mu}\sigma_{j\nu} \end{aligned}$$

$$-\sum_i \sum_\mu (h_{i\mu} + h_{i\mu}^{\rm self} + h_{i\mu}^{<\sigma>})\sigma_{i\mu} \tag{19}$$

where $\sigma_{i\mu}$ denotes an Ising spin operator which takes -1 or $+1$, and $h_{i\mu}^{<\sigma>}$ denotes a certain local effective field at $i\mu$. The local effective field $h_{i\mu}^{<\sigma>}$ is subject to an additional constraint,

$$h_{i\mu}^{<\sigma>} = h_{i\mu}^{\rm self} < \sigma_{i\mu} >, \tag{20}$$

where $< \sigma_{i\mu} >$ represents a thermal average of a local Ising spin $i\mu$ defined by

$$< \sigma_{i\mu} >= \frac{{\rm Tr}\{\sigma_{i\mu} \exp(-\beta H)\}}{{\rm Tr}\{\exp(-\beta H)\}}. \tag{21}$$

Here Tr denotes the trace operation over all Ising spin variables $\sigma_{i\mu}$. Under the constraint (20), the local effective field $h_{i\mu}^{<\sigma>}$ at $i\mu$ is proportional to the thermal average of the very Ising spin $i\mu$, and is also called a self-field in the Hamiltonian (19). The constraint (20) are referred to as "self-consistency condition".

The Ising spin system described by the Hamiltonian (19) is characterized by (i) the strong self-fields $h_{i\mu}^{<\sigma>}$, (ii) long-range antiferromagnetic interactions along z axis due to the inhibitory neuron pool, (iii) long-range antiferromagnetic interactions within xy plane due to the lateral connections, and (iv) short-range ferromagnetic and long-range antiferromagnetic interactions in the xyz space due to the cluster learning rule. It is to be noted that the interaction $J_{ij\mu\nu}^{(xyz)}$ effectively acts between two spins at $i\mu$ and $j\nu$ with neighboring two-dimensional coordinates i and j while the interaction $J_{\mu\nu}^{(z)}$ acts between two spins at $i\mu$ and $i\nu$ along the same z axis. The strong self-fields support a small cluster of up spins, and these clusters can be located in a variety of different distributions. Different distributions

of self-fields leads to different distributions of spin averages. The self-fields and these long-range antiferromagnetic interactions cause a variety of orientation maps calculated from spin averages.

The mean field equations (13) are derived from the Hamiltonian (19) under the self-consistency conditions (20). The mean field average of a spin is approximately given by a thermal average of a spin,

$$m_{i\mu} \simeq < \sigma_{i\mu} > . \tag{22}$$

We evaluate the thermal averages of spins in the next section.

4 SIMULATION OF ORIENTATION MAPS

4.1 Method

There are two methods to calculate the output responses of neurons: (i) a numerical calculation of the steady outputs (10) by an iterative method, and (ii) an indirect method which uses the Hamiltonian (19) to evaluate spin averages. The iterative procedure for (10) often fails to converge starting from random distributions of spin averages, at least in interesting ranges of parameters (Kuroiwa et al.). The difficulty of the iterative method partly comes from the existence of the strong self-fields and of three long-range antiferromagnetic interactions.

In the Hamiltonian formalism, a Monte Carlo simulation is often used to evaluate thermal averages of Ising spins in complex systems even though it is computationally expensive. However we cannot apply an ordinary Monte Carlo procedure because we do not know distributions of self-fields $h_{i\mu}^{\rm self} < \sigma_{i\mu} >$ in the Hamiltonian (19) before starting calculations. We

use a "self-consistent Monte Carlo method" (referred as SCMC method, Inawashiro et al., 1996; Kuroiwa et al.). We start from an initial condition in which certain values are assigned to self-fields. We perform a Monte Carlo procedure over many Monte Carlo (MC) periods. In a MC period self-fields are given by $h_{i\mu}^{\rm self}$ multiplied by MC averages of spins which are calculated over the immediately preceding MC period, and the procedure is continued until the self-fields converge with a certain accuracy. After the convergence has been accomplished, the self-consistency conditions (20) are satisfied approximately, and the output responses of neurons are approximated by

$$Z_i^\mu = \frac{1}{2}(1+ < \sigma_{i\mu} >_{\rm MC}), \tag{23}$$

where $< \sigma_{i\mu} >_{\rm MC}$ denotes a MC average of an Ising spin $i\mu$. It is to be noted that SCMC method is able to give output responses independent of parameters and of initial distribution of spin averages (Kuroiwa et al.). The stability of SCMC method in giving spin averages help us to find appropriate values of parameters by trial and error.

4.2 Parameters

We use fifteen bar-shaped patterns as explicit external inputs for simplicity as shown in Figure 3(a). The spatial correlation (11) between a pair of input patterns is easily calculated as shown in Figure 3(b). The lateral connection w_{ij} is chosen as shown in Figure 3(c), and its parameters are given in Table 1. The contribution factor e_{ij} in the cluster learning rule is chosen equal to the positive part of the lateral connection w_{ij}. The threshold $u_{\rm th}$ is chosen so that a neuron is not excited effectively ($Z_i^\mu < 0.5$ for all i and μ) in case

of no inputs.

(figure 3)

(table 1)

We calculate MC averages over every period of 4000 MC steps. Here one MC step represents a MC procedure where spin-flips are tried a number of times which is equal to the total number of spins. Note that first 4000 MC steps are discarded to exclude transient configurations of Ising spins. Initial self-fields $h_{i\mu}^{<\sigma>}$ are chosen to be $+h_{i\mu}^{\rm self}$ or $-h_{i\mu}^{\rm self}$ at random, and an initial spin configuration to be $+1$ and -1 at random. The other parameters are given in Table 1. A periodic boundary condition for the output layer is assumed.

4.3 Self-Consistency

In order to verify the self-consistency conditions (20) in SCMC simulation we calculate a root mean square deviation Δ defined by

$$\Delta = \sqrt{\frac{1}{PN}\sum_{\mu,i}\left(<\sigma_{i\mu}(l)>_{\rm MC} - <\sigma_{i\mu}(l-1)>_{\rm MC}\right)^2} \qquad (24)$$

where $<\sigma_{i\mu}(l)>_{\rm MC}$ denotes MC averages of Ising spins over "the lth MC period". The smallness of Δ ensures that the self-consistency conditions (20) are approximately satisfied within a certain accuracy.

(figure 4)

The behavior of Δ for every MC period is shown in Figure 4. The deviations Δ decreases monotonically as MC period increases. Between 80001 and 84000 MC steps, the root mean square deviation Δ becomes 0.019, and the self-consistency conditions (20) are satisfied approximately.

4.4 Orientation Maps

Monte Carlo averages of local Ising spins are calculated over a MC period between 80001 and 84000 MC steps where the system is approximately in a thermal equilibrium. An output response of a neuron against the μth input pattern is easily calculated from the MC average of a spin within the $z = \mu$ plane through the equation (23). All the responses of neurons are depicted on the two-dimensional triangular lattice of the output layer by line segments where the magnitudes of output responses are represented by the lengths of the line segments and the orientations of input patterns by those of the line segments (Figure 5). An output neuron makes effective responses against more than two input patterns, which are indicated by more than two line segments at a lattice site in Figure 5. It is to be noted that we used a random distribution of self-fields and a random spin configuration as the initial condition of the simulation in order to avoid any artificial arrangements to produce these structures.

(figure 5)

In the orientation map (Figure 5), we observe linear zones, $+1/2$ and $-1/2$ pinwheel singularities, and fractures as characteristic features of the local structures. These local features occur in a dispersed manner in the orientation map with no long-range orders and the map shows a global disorder as observed in the experiments (Blasdel and Salama, 1986; Blasdel, 1992a, 1992b; Bonhoeffer and Grinvald, 1991, 1993). The number of $+1/2$ pinwheel singularities is almost equal to that of $-1/2$ ones; there are twenty-one $+1/2$ pinwheel singularities and twenty $-1/2$ ones in the maps. Pinwheel singularities are smoothly linked, and this smoothness

leads to global smooth changes of preferred orientations except for regions of the pinwheel singularities and fractures. The orientation map obtained by using the cluster learning rule is qualitatively similar to biological observations.

In this paper the orientation selectivity is defined as the strength of the maximal output of a neuron. One of characteristic features of the orientation map of our model is that 80% neurons take an orientation selectivity between $[0.9, 1.0]$. In other words, the orientation selectivity in our model is tuned much shaper than that obtained by use of spontaneous activities (Miller, 1994). Learning due to explicit external inputs is important in producing a sharp tuning.

We present typical structures of $\pm 1/2$ pinwheel singularities on the triangular lattice. These structures are derived by manual assignment of orientations on the triangular lattice sites without computer simulation. We observe that a lot of $\pm 1/2$ pinwheel singularities in Figure 5 are similar in their structure to those in Figure 6.

(figure 6)

4.5 Quantitative Comparisons

In Figure 5, we can see two types of nearest neighbors of pinwheel singularities; each of 34 (82.9%) singularities has the nearest-neighbor singularity of the opposite sign and each of 7 singularities has that of the same sign. Recently Obermayer and Blasdel (1996) analyzed data obtained from squirrel monkeys and macaque monkeys of different ages, and reported that the approximately 80% of singularities have their own nearest neighbors of the opposite sign.

We calculate an autocorrelation function defined by (Obermayer et al, 1992)

$$C(|i-j|) =< q(i)q(j)\cos 2\phi(i)\cos 2\phi(j) >, \tag{25}$$

where $< \cdot >$ means the average operation over the all lattice points of the output layer. We define a preferred orientation $\phi(i)$ as the orientation of an input pattern which gives the maximal output of a neuron i and the selectivity $q(i)$ as the strength of the maximal output. The autocorrelation function calculated from the orientation map (Figure 5) is shown in Figure 7(a). It takes a Mexican-hat shape with orientation preferences anticorrelated for the average distance between singularities that amounts to 5.8 in unit of triangular lattice constant. The distance is interpreted as the half length of a hypercolumn. The autocorrelation function oscillates and its amplitude decreases as the distance increases, indicating a global disorder in the orientation map.

An intersection angle $I(i)$ represents the magnitude of the correlation between orientation preference coordinates and cortical coordinates (Erwin et al., 1996), and is given by

$$I(i) = \min(|\phi(i) - g(i)|, 180^\circ - |\phi(i) - g(i)|), \tag{26}$$

where $g(i)$ is the angular component of the gradient $\nabla\phi(i)$. From Figure 7(b) we can see that there are no preferred angle of the intersection between orientation preference and its angle of gradient. These three results mentioned above are consistent with data of the biological experiments (Erwin et al., 1996). (figure 7)

5 ROLES OF THE CLUSTER LEARNING RULE AND THE INHIBITORY NEURON POOL

5.1 Orientation Maps Obtained by Using Hebbian Learning Rule only

What kinds of roles does the cluster learning rule play in the formation of the orientation map as shown in Figure 5 ? We switch off the cluster learning rule, and use only the Hebbian learning rule in a formation of an orientation map. In the simulation we set e_{ij} to be zero, and c_1 to be 4.5. The other parameters remain the same as used in the previous simulation. The orientation map is shown in Figure 8. We can see that (i) structures of singularities are deformed and the identification of pinwheel singularities is more difficult than in Figure 5, and (ii) the orientation map contains only a few links between singularities and a lot of fractures instead.

(figure 8)

These results show that the cluster learning rule plays an important role in organizing smooth structures of pinwheel singularities, in forming smooth links between them, and in preventing the formation of fractures in the orientation map.

5.2 Analysis of Effects of the Inhibitory Neuron Pool and of the Cluster Learning Rule

Some of spin averages in the planes of $z = \mu$ and ν in the Ising spin space are shown in Figure 9. A cluster of positive spin averages on the μ plane corresponds to a cluster of firing neuron against the μth input pattern

in the output layer. A part of the orientation map is also shown at the bottom in Figure 9, where two clusters of firing neurons with line segments of orientation preferences are located nearby with an abrupt changes of preferred orientations.

(figure 9)

Local effective fields in the Hamiltonian (19) are divided into two classes, constant fields $h_{i\mu}$ and m-dependent fields $h_{i\mu}^{\rm self}(1+m_{i\mu})$. The former fields $h_{i\mu}$ are negative and have a strong effect to align spins downwards. A spin average $m_{i\mu}$ is approximately equal to 1 or -1 at low temperature, and the latter field becomes approximately $2h_{i\mu}^{\rm self}$ for a positive $m_{i\mu}$, or approximately vanishes for a negative $m_{i\mu}$. Because of the large positive value of $v_{\mu\mu}$, $2h_{i\mu}^{\rm self}$ is larger than the absolute value of $h_{i\mu}$, causing to establish an up spin, i.e., an active neuron.

The excitatory part of the exchange interactions $J_{ij}^{(xy)}$ aligns the neighboring spins to form a cluster of up spins within the μ plane. The up-spin cluster within the μ plane represents a cluster of output neurons responding to the μth input pattern. The inhibitory part of the interaction prevents the clusters from growing too large and keeps these spin clusters apart from each other of the order of inhibitory range r_I within the μ plane.

For an appropriate positive value of c_2, the exchange interaction $J_{\mu\nu}^{(z)}$ along the z axis becomes negative for a pair of μ and ν distant apart. This tends to prevent a pair of spins distant apart along the same z axis from aligning upwards simultaneously. Thus the inhibitory neuron pool prevents an output neuron to respond simultaneously to different input patterns.

Suppose that a fracture is formed in the output layer, that is, two active clusters are formed close to each other with abrupt changes of preferred

orientations as shown at the bottom in Figure 9. This means that in the Ising spin space two clusters of positive spin averages are located nearby in xy coordinates and distant apart from each other in z coordinates. We emphasize that in case of no inhibitory interactions between these two clusters many fractures tends to be formed without any obstructions and smooth changes of preferred orientations are lost all over the output layer as shown in Figure 8. The inhibitory interactions are provided by the negative part of $J^{(xyz)}_{ij\mu\nu}$ as explained below.

The cluster learning rule generates the exchange interaction $J^{(xyz)}_{ij\mu\nu}$ as seen in (16). The interaction is positive for a pair of spins $i\mu$ and $j\nu$ located nearby in xyz coordinates, and negative for a pair of spins $i\mu$ and $j\nu$ located nearby in xy coordinates and distant apart from each other in z coordinates. Positive parts of the interactions tend to align a pair of spins in the same direction, and negative parts of the interactions tend to align them in the opposite directions to each other. This means that a pair of neurons with nearby xy coordinates tends to fire simultaneously against similar input patterns μ and ν, and to avoid simultaneous firing against input patterns of a large difference in their orientations. In the orientation map formations of fractures are suppressed, and overall smoothness in changes of preferred orientations is achieved by the cluster learning rule as shown in Figure 5.

6 CONCLUSIONS

In the present paper we investigate self-organization of orientation maps due to explicit external stimuli in a formal neuron model, and propose the cluster learning rule where a modification of synaptic connections is

strengthened by an activation of neighboring output neurons. The simulation using the cluster learning rule reproduces orientation maps qualitatively similar to biological ones, and the quantitative analysis of our results also agree with those of experimental results.

In order to show the effects of the cluster learning rule we carry out the simulation without the cluster learning rule using only the Hebbian learning rule. It turns out that the resultant map contains abundant fractures and deformed structures of singularities, and the smoothness of preferred orientations is lost. Therefore we concluded that the cluster learning rule plays an important role in producing smooth changes of preferred orientations in the map. The effects of the cluster learning rule are also explained by the Ising spin formalism. It is an interesting problem to investigate whether an effect similar to the cluster learning rule is observed in biological systems.

The inhibitory neuron pool plays also an important role in forming orientation preference, preventing a output neuron from making multiple responses to several input patterns. Since there exist a large number of inhibitory neurons in many parts of cerebral cortices, we are interested in whether inhibitory neurons in the cortex really play a similar role to the inhibitory neuron pool in this study.

Acknowledgements

The authors are grateful to Professor S. Amari for enlightening discussions.

References

Amari, S. & Takeuchi, A. (1978). Mathematical theory on formation of category detecting nerve cells. Biological Cybernetics, *29*, 127 - 136.

Ashcroft, T. W., & Mermin, N. D. (1987). *Solid State Physics* (p715). Saunders College Publishing.

Blakemore, C., & Van Sluyters, R. C. (1975). Innate and environmental factors in the development of kitten's visual cortex. *Journal of Physiology (London)*, *248*, 663-716.

Blasdel. G. G., & Salama., G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, *321*, 579-585.

Blasdel, G. G. (1992a). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *Journal of Neuroscience*, *12*, 3115-3138.

Blasdel, G. G. (1992b). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, *12*, 3139-3161.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*, 429-431.

Bonhoeffer, T., & Grinvald, A. (1993). The layout of iso-orientation domain in area 18 of cat visual cortex: optical imaging reveals a pinwheel-like organization. *Journal of Neuroscience*, *13*, 4157-4180.

Braitenberg, V., & Braitenberg, C. (1979). Geometry of orientation columns in the visual cortex. *Biological Cybernetics*, *33*, 179-186.

Erwin, E., Obermayer, K., & Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Computation*, *7*, 425-468.

Geman, E. (1979). Some averaging and stability results for random differential equations. *SIAM Journal of Applied Mathematics*, *36*, 86-105.

Gödecke, I., & Bonhoeffer, T. (1996). Development of identical orientation maps for two eyes without common visual experience. *Nature*, *379*, 251-254.

Göetz, K. G. (1988). Cortical templates for the self-organization of orientation-specific d- and l-hypercolumns in monkeys and cats. *Biological Cybernetics*, *58*, 213-223.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, *160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive field and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, *195*, 215-243.

Hubel, D. H., Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, *158*, 267-294.

Inawashiro, S., Tamori, Y., Miyake. S., & Kuroiwa, J. (1996). Hamiltonian formalism for self-organization of formal neurons. *Journal of Physics A: Mathematical and General*, *29*, 7389-7399.

Kim, D., & Bonhoeffer, T. (1994). Reverse occlusion leads to a precise restoration of orientation preference maps in visual cortex. *Nature*, *370*, 730-372.

Kohonen, T. (1987). *Self-Organization and Associative Memory.* Springer-Verlag, New York.

Kuroiwa, J., Inawashiro, S., Miyake. S., & Aso, H. Mean field theory and

self-consistent Monte Carlo method for self-organization of formal neuron model. In preparation.

Linsker, R. (1986). From basic network principles to neural architecture (series). *Proceedings of the National Academy of Sciences*, *83*, 7508-7512, 8390-8394, 8779-8783.

McCulloch, W. S., & Pitts, W. (1943). A logical calculation of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115-133.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *Journal of Neuroscience*, *14*, 409-441.

Miyashita, M., & Tanaka, S. (1992). A mathematical model for the self-organization of orientation columns in visual cortex. *Neuroreport*, *3*, 69-72.

Nicholls, J. G., Martin, A. R., & Wallace, B. G. (1992) *From Neuron to Brain* (p653). Sinauer Associate, Inc. Publishers.

Obermayer, K., Blasdel, G. G., & Schulten, K. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Physical Review A*, *45*, 7568-7589.

Obermayer, K., & Blasdel, G. G. (1996). Singularities in primate orientation maps. *Neural Computation*, *7*, 555-575.

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biological Cybernetics*, *35*, 63-72.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*, 85-100.

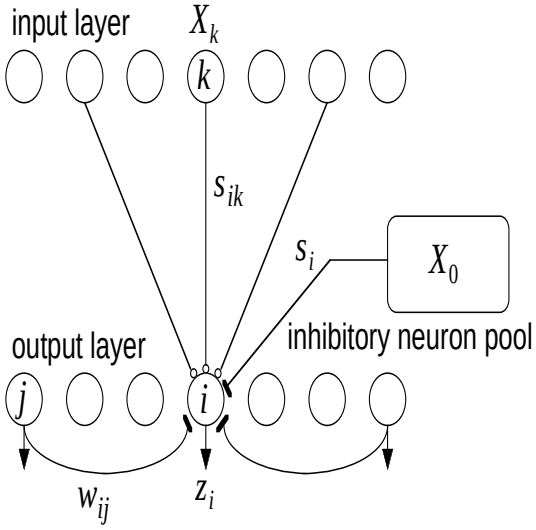

Figure 1: A two-layered neural network model. An input signal X_k from input neuron k is sent to an output neuron i through an excitatory synaptic connection s_{ik}. A constant inhibitory input X_0 $(= 1)$ is sent to the output neuron i through an inhibitory synaptic connection s_i. The output neuron i receives also a feedback contribution from a neighboring neuron j through a lateral connection w_{ij}. A sum of these three kinds of inputs leads to an output z_i of the neuron i through a sigmoidal output function.

Table 1: Parameters.

$L_x = 17$, $L_y = 17$, $N_x = 40$, $N_y = 40$,
$P = 15$, $T = 4.0$, $u_{\rm th} = 1.5$, $X_0 = 1.0$,
$c_1 = 1.05$, $(c_2/c_1) = 10.0$, $c'_1 = 1.05$, $(c'_2/c'_1) = 10.0$,
$E = 0.96$, $I = 2.04$, ${\sigma_E}^2 = 1.64$, ${\sigma_I}^2 = 2.5$,

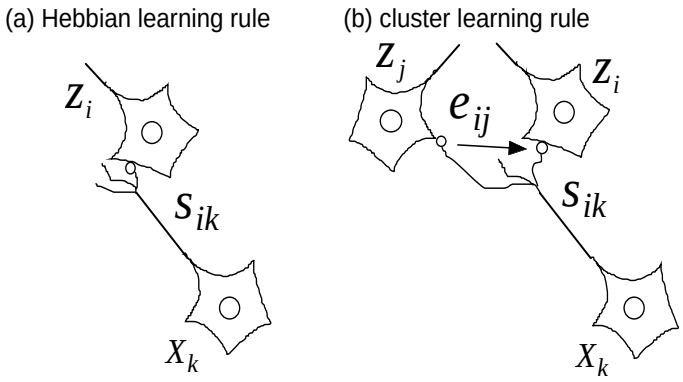

Figure 2: (a) Hebbian learning rule. (b) Cluster learning rule. In the Hebbian learning rule, a synaptic connection s_{ik} is strengthened by simultaneous activation of an input neuron k and an output neuron i. In the cluster learning rule, the synaptic connection is strengthened by a simultaneous activation of an input neuron and an output neuron i together with its neighboring neurons. A contribution factor from an output neuron j to an output neuron i in the learning is denoted by e_{ij} (> 0).

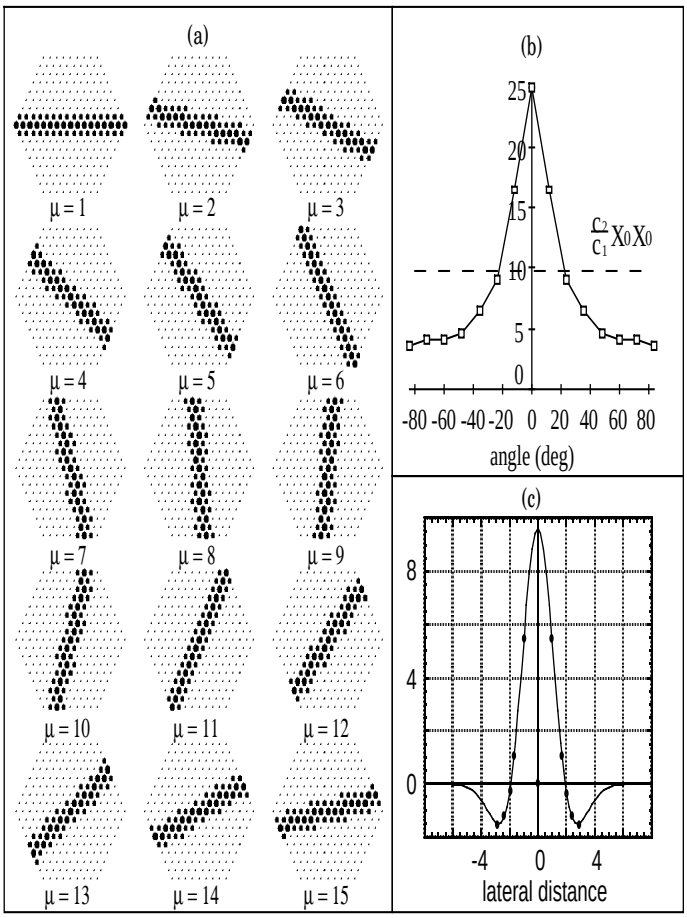

Figure 3: (a) Fifteen bar-shaped patterns. The input neurons are arranged on a triangular lattice. The magnitude of an input signal is chosen to be 0, 0.5 or 1, represented by a point, a small or large black dot, respectively. (b) Spatial correlation between the input pattern $\mathbf{X}^1$ and the others. The horizontal axis implies the intersecting angle between two input patterns and the vertical axis the spatial correlation. The broken line represents a value of $(c_2/c_1)X_0X_0$ $(= 10.0)$ used in the simulation. Note that $(c_1 v_{11} - c_2 X_0 X_0)$ takes a positive value. (c) A lateral connection. The horizontal axis implies a distance of a pair of neurons and the vertical axis the lateral connection. A black dot (•) corresponds to a distance of a pair of neurons on the triangular lattice. Note that we set a strength of self-feedback connection w_{ii} to be zero. The connection takes a positive value for 1st, 2nd and 3rd neighboring neurons.

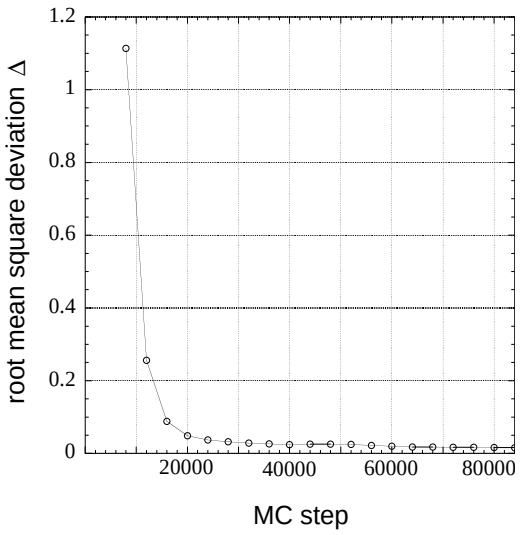

Figure 4: The root mean square deviation Δ of spin averages as a function of MC step. The horizontal axis represents MC step and the vertical axis Δ. Smallness of Δ ensures that the self-consistency conditions are satisfied approximately.

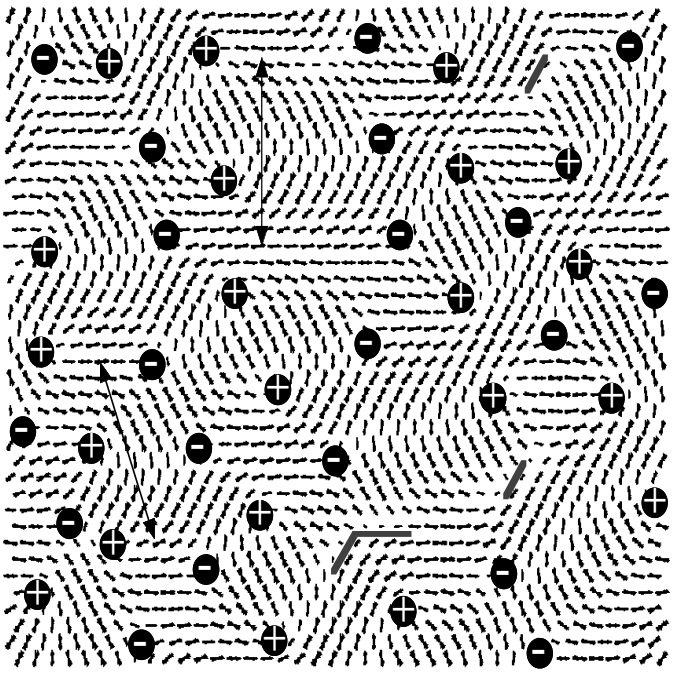

Figure 5: Orientation map obtained by SCMC method using cluster learning rule in the period between 80001 and 84000 MCS. The length of a line segment represents the magnitude of an output of a neuron and the orientation of the line segment the orientation of the input. An output neuron responds to more than two input patterns, which are indicated by more than two line segments at a lattice site. The symbol '+' represents +1/2 pinwheel singularity and '−' −1/2 one. A thin line with arrows shows a linear zone and a thick line a fracture.

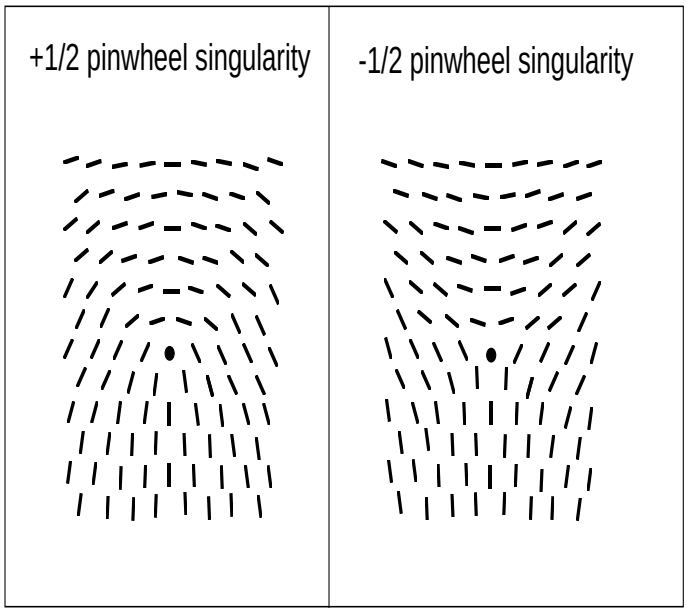

Figure 6: $\pm 1/2$ pinwheel singularities on the triangular lattice. Orientations are manually assigned on the lattice sites without computer simulation.

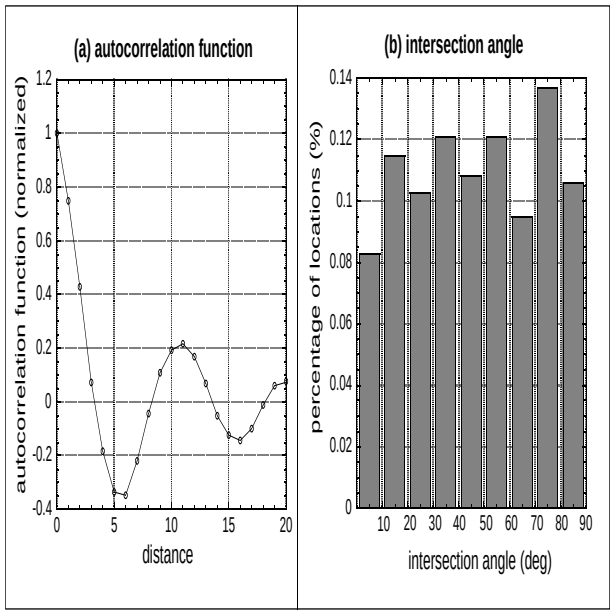

Figure 7: (a) Autocorrelation function of distance between a pair of two neurons i and j in unit of the triangular lattice constant. (b) The histogram shows the percentage of locations with an intersection angle $[0, 90°]$ between a preferred orientation $0 \leq \phi \leq 180°$ and its gradient $0 \leq g \leq 180°$.

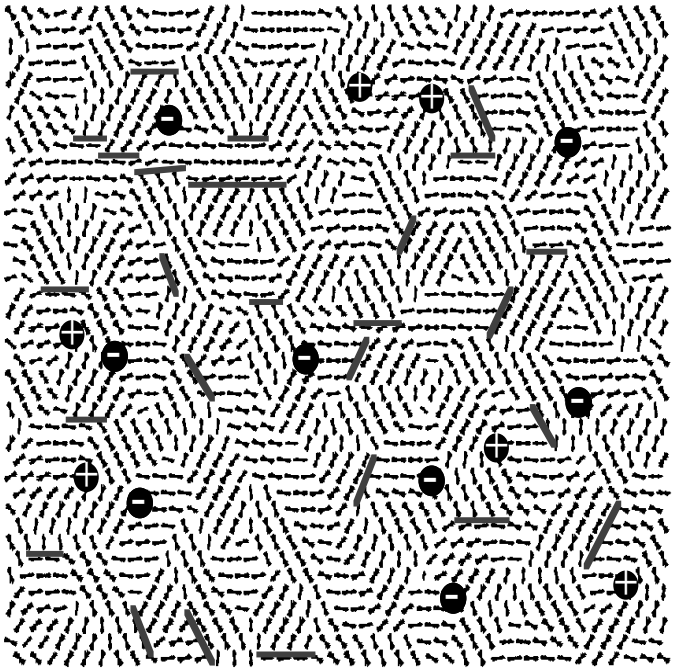

Figure 8: Orientation map calculated in the period between 80001 and 84000 MCS by use of the Hebbian learning rule only. In the simulation we set $e_{ij} = 0$ and $c_1 = 4.5$, and the other parameters are the same as in table 1.

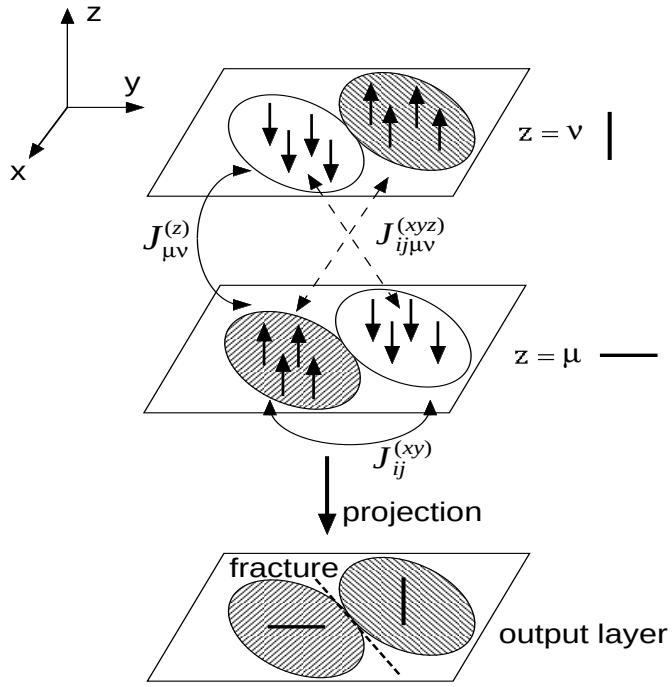

Figure 9: The schematic representation of the Ising spin system and the output layer. The planes of $z = \mu$ and ν belong to a three-dimensional Ising spin space. The plane at the bottom represents the output layer of the neuron system. Spin averages in the three-dimensional Ising spin space are projected onto the two-dimensional output layer. Clusters of positive spin averages are represented by shaded regions. In the output layer, a fracture is formed between two cluster with a large difference of preferred orientations. Exchange interactions $J^{(xy)}_{ij}$ act between two spins $i\mu$ and $j\mu$ on the same xy plane and $J^{(z)}_{\mu\nu}$ act between two spins $i\mu$ and $i\nu$ along the same z axis, while $J^{(xyz)}_{ij\mu\nu}$ act between two spins $i\mu$ and $j\nu$ with different i and j.