Hamiltonian formalism for self-organization of formal neurons

1996 J. Phys. A: Math. Gen. 29 7389

(http://iopscience.iop.org/0305-4470/29/23/009)

J. Phys. A: Math. Gen. **29** (1996) 7389–7399. Printed in the UK

Hamiltonian formalism for self-organization of formal neurons

Sakari Inawashiro†, Yoshihide Tamori‡, Shogo Miyake§¶ and Jousuke Kuroiwa‖

† Department of Computer Science, Nihon University, Koriyama 963, Japan
‡ (Intelligence and its synthesis, Precursory Research for Embryonic Science and Technology (PRESTO) System, Research Development Corporation of Japan (JRDC)), Information Representation, Riken, Wako, Saitama 351-01, Japan
§ Department of Applied Physics, Tohoku University, Sendai 980-77, Japan
‖ Department of Communication Engineering, Tohoku University, Sendai 980-77, Japan

Received 24 June 1996, in final form 3 September 1996

Abstract. The self-organization of formal neurons due to external inputs is investigated in a two-layered model of a neural network. The synaptic connections are modified by a Hebbian rule. A steady state of the connections is attained after repeated inputs of a set of the patterns, and self-organization is accomplished. Equations describing the response of the output neurons in the steady state are transformed into the form of mean field equations for an Ising spin system. The mean field equations contain each *self-field* at a lattice site which is proportional to the spin average at the same site. A response property of the neural network is determined by a spin structure at a fixed low temperature. We show that a Hamiltonian of the Ising spin system and *self-consistency* conditions give the mean field equations. Based on the Hamiltonian, we propose a *self-consistent* Monte Carlo simulation as a practical method of finding a spin structure, i.e. a response property of the neural network. The self-fields are self-consistently determined in the Monte Carlo procedure. The result of the self-consistent Monte Carlo simulation qualitatively agrees with a numerical solution of the mean field equations in a simple case of self-organization.

1. Introduction

Orientation selectivity was first discovered by Hubel and Wiesel in microelectrode experiments on the striate cortex of cats and monkeys [1]. The cells in the orientation columns express a preference for certain orientations of bars and edges in their visual fields. Recent experiments using the optical technique revealed the detailed distribution of the orientation preference in visual cortices of monkeys [2] and cats [3].

The self-organization of the orientation preference map has been theoretically investigated by many authors [4–10]. While some of characteristic features of orientation preference map have been presented by recent works [6–8], nonlinear effects of outputs of formal neurons have not been fully considered in these works and influence of external inputs on the formation of the orientational map has not been explicitly treated. The study of self-organization taking into account the nonlinearity of neurons and influence of external inputs was started by von der Malsburg [4], and then developed by Takeuchi and Amari [5] using a continuous nerve field.

¶ E-mail address: miyake@camp.apph.tohoku.ac.jp

0305-4470/96/237389+11$19.50

In this paper we study self-organization of orientation selectivity due to external-inputs in a formal neuron model with a sigmoidal output function. A steady state of the model is attained by Hebbian learning [11] under the external inputs, and the outputs of the formal neurons are described by a set of nonlinear equations. The purposes of our study are, (1) to show that the nonlinear equations of the steady state are transformed into mean field equations which arise in the statistical mechanics of magnetism, (2) to show that a Hamiltonian of an Ising spin system and self-consistency conditions give the mean field equations in mean field approximation, and (3) to introduce a kind of Monte Carlo simulation based on the Hamiltonian in order to find the self-organization of the formal neurons.

In section 2 a two-layered model of a neural network with an inhibitory neuron pool and its related basic equations are described. In section 3 a steady state of the neural network is described by a set of nonlinear equations. In section 4 the nonlinear equations derived in section 3 are converted to a set of mean field equations which arise in the statistical mechanics of magnetism. In section 5 a Hamiltonian is defined for an Ising spin system which is characterized by self-fields. In section 6 a method of *self-consistent* Monte Carlo simulation is proposed. In a simple case of orientation selectivity it is shown that the result of self-consistent Monte Carlo simulation agrees qualitatively with a numerical solution of the mean field equations. Section 7 is devoted to conclusions.

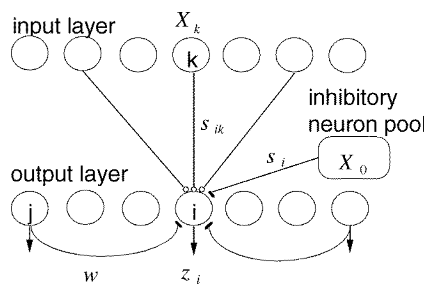

Figure 1. Two-layered model of neural network.

2. Model and basic equations

We consider a two-layered model of a neural network [5] which consists of an input layer and an output layer of formal neurons as shown in figure 1. A neuron in the output layer receives an input pattern $\boldsymbol{X} = \{X_k\}$ from the input layer, and input signals from the output layer and an inhibitory neuron pool. An output signal from an output neuron is denoted by z_i. Here the input and the output layer are two-dimensional lattices of neurons, and k and i denote two-dimensional vectors indicating the locations of neurons within each layer. An excitatory connection s_{ik} between the output neuron i and the input neuron k is modified by unsupervised learning due to the input patterns. The connection w_{ij} between two neurons i and j within the output layer is a function of the distance between i and j. It is of a lateral-inhibition type, that is, it is excitatory for a pair of neurons in the neighbourhood and inhibitory for a pair of neurons far from each other. This type of interaction is sometimes called a 'Mexican-hat' interaction [12]. We use the Mexican-hat interaction of the form,

$$w_{ij} = (E + I) \exp\left(-\frac{|i - j|^2}{r_E^2}\right) - I \exp\left(-\frac{|i - j|^2}{r_I^2}\right) \tag{1}$$

where E denotes the excitatory strength, I the inhibitory strength, r_E the range of the excitatory interaction, and r_I the range of the inhibitory interaction. We also introduce an inhibitory connection s_i between the output neuron i and an inhibitory neuron pool [5] which provides a constant input signal $X_0 = 1$. An output from the output neuron i is expressed by

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

where u_i denotes the average membrane potential of the output neuron i, and u_{th} a threshold for activity. Here the output function f is a monotonically increasing nonlinear function with a saturation property, which is approximated in this paper by a sigmoid function

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

where β is a characteristic constant that determines the gradient of the output function, $\beta/2$, at $x = 0$.

The average membrane potential, referred to as the membrane potential hereafter for simplicity, is assumed to satisfy the time-dependent equation [5]

$$\tau' \frac{\mathrm{d}u_i}{\mathrm{d}t} = -u_i + \sum_k s_{ik} X_k + \sum_j w_{ij} z_j - s_i X_0 \tag{4}$$

where τ' denotes a time constant of the membrane potential of the order of milliseconds, and the resting state is taken as $u_j = 0$. In the right-hand side of (4) the first term represents the decay effect which ensures the saturation of the membrane potential, the second term the external input to the output neuron i through the synaptic connections, the third term the contribution from the output neuron j through the 'Mexican-hat' interaction, and the fourth term the inhibitory contribution from the inhibitory neuron pool.

It is assumed that the synaptic connections s_{ik} and s_i are modified according to a so-called 'Hebbian rule', that is, it is increased by the simultaneous activities of both the presynaptic and postsynaptic neurons. Then the learning process is described by the time development of the synaptic connections,

$$\tau \frac{\mathrm{d}s_{ik}}{\mathrm{d}t} = -s_{ik} + c z_i X_k \tag{5a}$$

$$\tau \frac{\mathrm{d}s_i}{\mathrm{d}t} = -s_i + c' z_i X_0 \tag{5b}$$

where τ denotes a time constant of the learning much greater than τ' of the membrane potential, and c and c' are constants which control the efficiency of learning. The time constants in (5*a*) and (5*b*) are chosen to be the same for the sake of simplicity.

When an input pattern is presented to an output neuron i, the membrane potential u_i immediately reaches to a steady value. Hereafter we simply replace the potential u_i by the steady value

$$u_i = \sum_k s_{ik} X_k + \sum_j w_{ij} z_j - s_i X_0 \tag{6}$$

for an input pattern $\boldsymbol{X}$.

Suppose that a pattern $\boldsymbol{X}^\mu$ labelled by μ, ($\mu = 1, 2, \ldots, P$), is chosen at random from an input ensemble, $\{\boldsymbol{X}\}$, which contains the pattern $\boldsymbol{X}^\mu$ with probability p_μ, and is presented to the output layer as an input pattern. The learning time constant is assumed to be much greater than the time duration in which each kind of labelled pattern is presented to the output layer. During the presentation of the input patterns the synaptic connections are little modified, and the learning equations are approximated by [5]

$$\tau \frac{\mathrm{d}s_{ik}}{\mathrm{d}t} = -s_{ik} + c\langle z_i X_k \rangle \tag{7a}$$

$$\tau \frac{\mathrm{d}s_i}{\mathrm{d}t} = -s_i + c'\langle z_i X_0 \rangle \tag{7b}$$

where the brackets in the right-hand side of (7*a*) and (7*b*) denote the ensemble average

$$\langle z_i X_k \rangle = \sum_{\mu=1}^{P} p_\mu z_i^\mu X_k^\mu \tag{8a}$$

$$\langle z_i X_0 \rangle = \sum_{\mu=1}^{P} p_\mu z_i^\mu . \tag{8b}$$

Here z_i^μ is the response of the output neuron to the input pattern $\boldsymbol{X}^\mu$. The approximation used in the derivation of (7*a*) and (7*b*) is well known in physics as 'adiabatic approximation'.

3. Steady state

If we continue to apply the input ensemble to the output layer for a time duration much greater than the learning time constant τ, steady values of synaptic connections will be attained,

$$S_{ik} = c \sum_{\mu=1}^{P} p_\mu Z_i^\mu X_k^\mu \tag{9}$$

$$S_i = c' X_0 \sum_{\mu=1}^{P} p_\mu Z_i^\mu \tag{10}$$

where the capital letters mean the quantities in the steady state. After the steady state has been attained, the learning process in the neural network stops. The response of the neural network is completely described by the set of steady outputs $\{Z_i^\mu\}$. Using (2), (6), (9) and (10) the steady output Z_i^μ is expressed as

$$Z_i^\mu = f\left[\sum_{\nu=1}^{P} (cv_{\mu\nu} - c'v_0) p_\nu Z_i^\nu + \sum_j w_{ij} Z_j^\mu - u_{\mathrm{th}} \right] \tag{11}$$

where $v_{\mu\nu}$ denotes the spatial correlation of the input patterns $\boldsymbol{X}^\mu$ and $\boldsymbol{X}^\nu$ and is defined by

$$v_{\mu\nu} = \sum_k X_k^\mu X_k^\nu \tag{12}$$

and v_0 denotes the self-correlation of the input from the inhibitory neuron pool and is defined by

$$v_0 = X_0 X_0 \ . \tag{13}$$

The spatial self-correlation $v_{\mu\mu}$ is positive and large compared to $v_{\mu\nu}$ ($\nu \neq \mu$) as shown later in section 6. We choose the learning constants c and c' so that $(cv_{\mu\mu} - c'v_0)$ is positive.

From (11) it is easily seen that the term proportional to $c'v_0$ would play the role of inhibition against the input pattern μ, if the neuron made a response to the input pattern ν. Therefore, the inhibitory neuron pool prevents a neuron from making redundant responses to input patterns which are weakly correlated with the optimum pattern.

4. Mean field equations and an Ising spin system

In this section we assume that the probabilities in the input ensemble are uniform, i.e.

$$p_\mu = \frac{1}{P} \qquad \text{for all } \mu. \tag{14}$$

Using the hyperbolic tangent function, the output function (3) is expressed by

$$f(x) = \tfrac{1}{2}(1 + \tanh \beta x). \tag{15}$$

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

$$Z_i^\mu = \tfrac{1}{2}(1 + m_{i\mu}). \tag{16}$$

Then equation (11) is rewritten in terms of $m_{i\mu}$ as,

$$m_{i\mu} = \tanh \beta \left[\frac{1}{2} \sum_j w_{ij}(1 + m_{j\mu}) + \frac{1}{2P} \sum_\nu^P (cv_{\mu\nu} - c'v_0)(1 + m_{i\nu}) - u_{\text{th}} \right]. \tag{17}$$

Equation (17) has the form of a mean field equation for an Ising spin system in statistical physics of magnetism [13]. Here the variable $m_{i\mu}$ is interpreted as the mean field average of an Ising spin located at a three-dimensional lattice site, $i\mu$, where i stands for a two-dimensional lattice site in the xy plane and μ a one-dimensional lattice site on the z axis. It is to be noted that the z coordinate of each lattice site represents the index number of the input pattern.

Equation (17) is rewritten in a simple form

$$m_{i\mu} = \tanh \beta H_{i\mu} \tag{18}$$

where $H_{i\mu}$ is a local effective field defined by

$$H_{i\mu} = \sum_{j \neq i} J_{ij}^{(xy)} m_{j\mu} + \sum_{\nu \neq \mu} J_{\mu\nu}^{(z)} m_{i\nu} + h_{i\mu} + h_{i\mu}^{\text{self}} + h_{i\mu}^{\text{self}} m_{i\mu}. \tag{19}$$

The set of equations (18) and (19) constitute a set of the mean field equations at a fixed temperature $T = 1/\beta$ in the theory of magnetism. The constants in (19) are given by

$$J_{ij}^{(xy)} = \tfrac{1}{2} w_{ij} \tag{20}$$

$$J_{\mu\nu}^{(z)} = \frac{1}{2P}(cv_{\mu\nu} - c'v_{\mu\nu}^0) \tag{21}$$

$$h_{i\mu} = \sum_{j \neq i} J_{ij}^{(xy)} + \sum_{\nu \neq \mu} J_{\mu\nu}^{(z)} - u_{\text{th}} \tag{22}$$

and

$$h_{i\mu}^{\text{self}} = J_{ii}^{(xy)} + J_{\mu\mu}^{(z)} \tag{23}$$

where $J_{ij}^{(xy)}$ is interpreted as an exchange interaction between the two Ising spins at $i\mu$ and $j\mu$ on the plane of $z = \mu$, and $J_{\mu\nu}^{(z)}$ as that between the two Ising spins at $i\mu$ and $i\nu$ on an axis parallel to the z-axis. It is noted that the symmetry in (21) with respect to μ and ν is ensured by the assumption of the uniform probability in the input ensemble. In (19), $h_{i\mu}$ and $h_{i\mu}^{\text{self}}$ denote local effective fields acting upon the Ising spin at $i\mu$. The last term in the right-hand side of (19), $h_{i\mu}^{\text{self}} m_{i\mu}$, is a kind of mean field at the lattice site $i\mu$ which is proportional to the spin average at the same site $i\mu$, and is referred to as the *self-field* in this paper. The similar term appears in a dynamical rule for a Hopfield model and is called a self-coupling term [14]. Note that the coefficient of the self-field, $h_{i\mu}^{\text{self}}$, is positive definite, because of the large value of the self-correlation $v_{\mu\mu}$. The notion of the self-field represents one of the biological features in our formulation of self-organization. The self-fields will play an important role in the formation of orientation selectivity.

5. Hamiltonian

We consider a Hamiltonian in the three-dimensional $i\mu$ lattice space,

$$H = -\tfrac{1}{2}\sum_{i}\sum_{\mu\neq\nu}\sum_{\nu} J^{(z)}_{\mu\nu}\sigma_{i\mu}\sigma_{i\nu} - \tfrac{1}{2}\sum_{i\neq j}\sum_{j}\sum_{\mu} J^{(xy)}_{ij}\sigma_{i\mu}\sigma_{j\mu} - \sum_{i}\sum_{\mu}(h_{i\mu} + h^{\mathrm{self}}_{i\mu} + h^{\langle\sigma\rangle}_{i\mu})\sigma_{i\mu} \tag{24}$$

where $\sigma_{i\mu}$ denotes a spin operator which takes the value 1 or -1, and $h^{\langle\sigma\rangle}_{i\mu}$ denotes a local effective field acting upon the Ising spin at $i\mu$ and undetermined constant. The thermal average of each Ising spin is given by

$$\langle\sigma_{i\mu}\rangle = \frac{\mathrm{Tr}\{\sigma_{i\mu}\exp(-\beta H)\}}{\mathrm{Tr}\{\exp(-\beta H)\}} \tag{25}$$

where Tr means taking the trace over all the Ising spin variables $\sigma_{i\mu}$.

In order to determine $h^{\langle\sigma\rangle}_{i\mu}$, we add constraints

$$h^{\langle\sigma\rangle}_{i\mu} = h^{\mathrm{self}}_{i\mu}\langle\sigma_{i\mu}\rangle. \tag{26}$$

The equations (25) and (26) constitute a set of simultaneous equations which determine $h^{\langle\sigma\rangle}_{i\mu}$.

In physics a set of mean field equations is derived from the Hamiltonian of interacting Ising spins, assuming that in thermal equilibrium the influence of other spins on a particular spin is approximately represented by a mean field which is a sum of terms proportional to the thermal averages of surrounding spins. The mean field derived from the Hamiltonian (24) is given by a local effective field,

$$H^{\langle\sigma\rangle}_{i\mu} = \sum_{j\neq i} J^{(xy)}_{ij} m_{i\mu} + \sum_{\nu\neq\mu} J^{(z)}_{\mu\nu} m_{i\nu} + h_{i\mu} + h^{\mathrm{self}}_{i\mu} + h^{\langle\sigma\rangle}_{i\mu} \tag{27}$$

where

$$m_{i\mu} = \tanh\beta H^{\langle\sigma\rangle}_{i\mu}. \tag{28}$$

The average of each Ising spin in mean field approximation is given by (28), and the constraints (26) are replaced by

$$h^{\langle\sigma\rangle}_{i\mu} = h^{\mathrm{self}}_{i\mu} m_{i\mu}. \tag{29}$$

Equations (27), (28) and (29) give the mean field equations (18) and (19).

The Hamiltonian (24) subject to the constraints (26) is referred to as a *pseudo*-Hamiltonian in this paper. The local fields defined by (26) are referred to as self-fields in the pseudo-Hamiltonian. The constraints (26) are called *self-consistency conditions* for the pseudo-Hamiltonian.

The physical system of the output neurons and the external environments which are represented by the correlations between input patterns are incorporated into the single pseudo-Hamiltonian in three-dimensional lattice space spanned by the two-dimensional output layer and the one-dimensional fictitious axis of input patterns. The self-consistency condition (26) determines the self-fields only at the fixed temperature T as constant external fields in the pseudo-Hamiltonian. Biologically, therefore, the pseudo-Hamiltonian is meaningful only at the fixed temperature for the self-consistency. It should be noted that a pseudo-Hamiltonian has its own distribution of upward and downward self-fields. In other words, different distributions of self-fields correspond with different pseudo-Hamiltonians.

6. Mean field equations and self-consistent Monte Carlo simulation

There are two alternative methods of obtaining a solution of the mean field equations. A direct method that requires a numerical calculation based on iteration of the mean field equations (17) themselves and an indirect method that uses the Hamiltonian to calculate the thermal averages via (25).

We usually use an iterative method to solve the mean field equations (17). A set of local magnetizations, $\{m_{i\mu}(0)\}$, is assumed first. This naturally determines a set of self-fields $\{h_{i\mu}^{\text{self}} m_{i\mu}(0)\}$, referred to as an initial distribution of self-fields. Substituting the initial local magnetizations into the right-hand side of (17), we get a new set of magnetization $\{m_{i\mu}(1)\}$ as the results of the first iteration. Then we continue by substituting the new magnetizations, $m_{i\mu}(1)$, into the right-hand side of (17), and so forth. If the iterative procedure converges to give a set of $\{m_{i\mu}\}$ within a certain accuracy, a solution of the mean field equations is obtained. It is to be noted that the iteration sometimes fails to converge, depending on the initial conditions above mentioned. We guess that this is due partly to an inadequate distribution of initial self-fields, partly to inhibitory (antiferromagnetic) effects in the lateral interactions, w_{ij}, partly to an inappropriate choice of parameter values, and/or partly to nonlinear effects. The important things are as follows.

(1) There are many solutions in the mean field equations (17). This is a situation quite different from mean field equations in magnetism where usually only one or a few solutions exist.

(2) The solutions depend strongly on the initial distribution of self-fields, $\{h_{i\mu}^{\text{self}} m_{i\mu}(0)\}$, because of the large coefficient, $h_{i\mu}^{\text{self}}$, originally coming from the large self-correlations of the input patterns.

We usually started from a random distribution of magnetizations, and sometimes obtained a rather weakly ordered distribution of up-spin clusters dispersed in the sea of down-spin magnetization, and sometimes non-convergent results. We wonder what sorts of solutions might exist for the mean field equations besides those obtained in the manner just described.

As an alternative, we use the Hamiltonian to search for other types of solutions. The partition function of Ising spins in three dimensions has not been calculated analytically. Therefore, analytical calculation of the right-hand side of (25) is almost impossible. The thermal averages $\{\langle\sigma_{i\mu}\rangle\}$ can be evaluated numerically by a Monte Carlo simulation [15–17]. In performing the Monte Carlo simulation, we start with an initial distribution of self-fields represented by local effective fields, $h_{i\mu}^{\text{self}}$, and with an initial configuration of Ising spins $\sigma_{i\mu} = 1$ or -1. Then we take account of the self-consistency conditions (26) as follows. The local effective fields, $h_{i\mu}^{\langle\sigma\rangle}$, in the Hamiltonian (24) are kept constant throughout a certain period of Monte Carlo steps (MCS). Spin averages are calculated over the period, and the averages multiplied by $h_{i\mu}^{\text{self}}$ are used as new local fields $h_{i\mu}^{\langle\sigma\rangle}$ in the following period of MCS. The procedure is continued until the local fields converge within a certain accuracy. Then the self-consistency conditions (26) are satisfied and the converged local fields give the self-fields in thermal equilibrium. Hereafter this simulation is referred to as 'a *self-consistent* Monte Carlo simulation'.

It is to be noted that, contrary to the usual situation in physics, the pseudo-Hamiltonian is the result of an approximation to the self-organization in our neural network model. Therefore, it is necessary to ensure that the result of the self-consistent Monte Carlo simulation agrees at least qualitatively with the numerical solution of (17). In the following we compare results obtained by each method for the simplest case of self-organization of the orientation selectivity. We adopt triangular lattices for the input and the output layer of

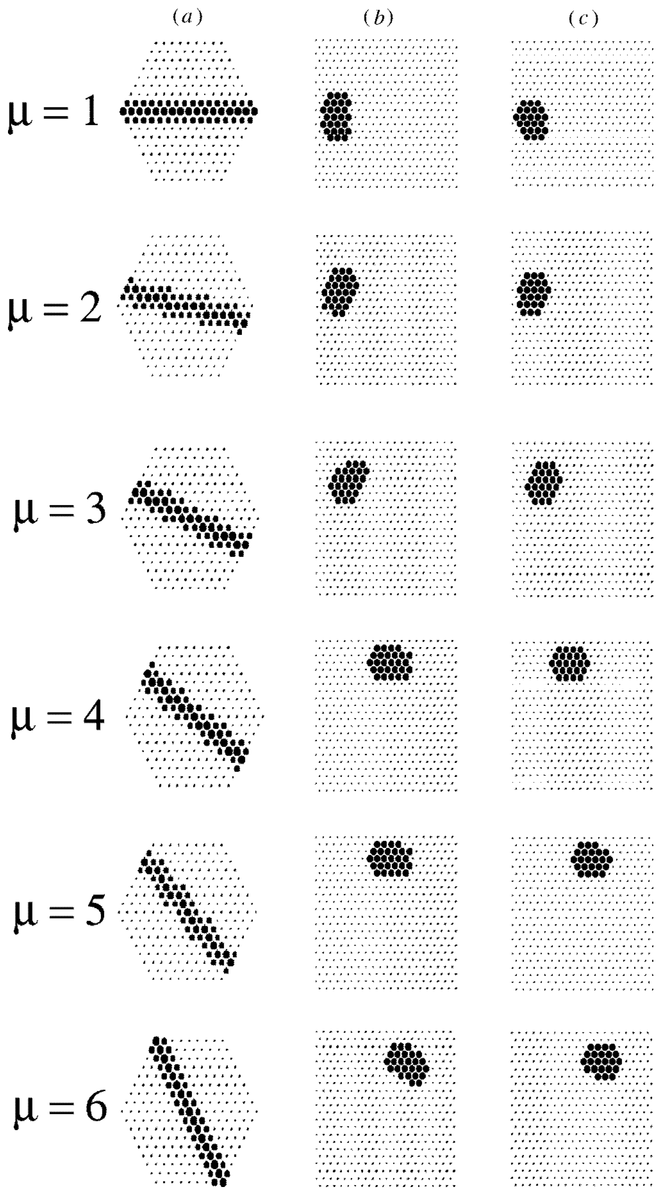

Figure 2. Input patterns and output patterns from $\mu = 1$–6. (a) Input patterns $\boldsymbol{X}^{\mu}$. (b) Output patterns Z^{μ} obtained from mean field equations with $T = 0.5$. (c) Output patterns Z^{μ} of self-consistent Monte Carlo simulations with $T = 0.3$. Parameters are chosen as follows. $E = 0.5$, $r_E = 2$, $I = 0.3$, $r_I = 15$, $c/P = 0.098$ and $h = 1.0$. The size of the input layer is given by $L_x \times L_y$ where $L_x = 17$ and $L_y = 17$, and that of the output layer by $N_x \times N_y$ where $N_x = 20$ and $N_y = 22$. Averages of spins are taken over 1000 MCS, and the last average is taken from the 5001st MCS to the 6000th MCS.

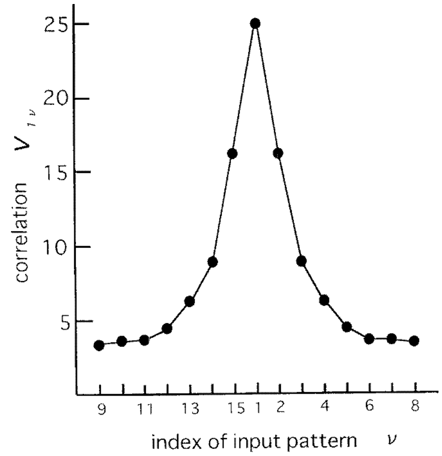

Figure 3. Correlation $v_{1\nu}$ between input patterns 1 and ν. The self-correlation v_{11} is very large compared to the other correlations and results in strong self-fields.

the neural network, and we use a simplified form of the 'Mexican-hat' interaction

$$w_{ij} = \begin{cases} E & \text{for } |i-j| \leqslant r_E \\ -I & \text{for } r_E < |i-j| \leqslant r_I \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

as an approximation for (1). Fifteen patterns of a bar shape with various orientations are chosen as input patterns, $\boldsymbol{X}^\mu$ ($\mu = 1, 2, \ldots, 15$). A few of them are shown in figure 2(a) for $\mu = 1$–6. Generally the spatial correlation $v_{\mu\nu}$ depends on the intersection angle $\theta_{\mu\nu}$ between the input patterns μ and ν, and is roughly given by

$$v_{\mu\nu} = \begin{cases} 2ab & \text{for } \mu = \nu \\ \dfrac{b^2}{\sin\theta_{\mu\nu}} & \text{for } \mu \neq \nu \end{cases} \tag{31}$$

where a and b denote the radius and the width of the input pattern $\boldsymbol{X}$ of a bar shape as shown in figure 2(a). As an example of numerical calculation based on (12), the correlation for $\mu = 1$ and ν is shown in figure 3 as a function of ν. We neglect the effect of the inhibitory neuron pool by setting the inhibitory learning constant $c' = 0$ for the sake of simplicity.

The mean field equations (17) are solved numerically and a few of the output patterns Z_i^μ converted from $m_{i\mu}$ are shown in figure 2(b) for $\mu = 1$–6, where the output neurons which fire for the input pattern μ are denoted by full circles, showing the response of the neural network after the learning has been accomplished.

On the other hand, Monte Carlo simulation for the Ising spin system in the three-dimensional $i\mu$ lattice space has been carried out starting from an initial distribution of self-fields similar to the above numerical solution. The self-fields and other quantities are averaged over 1000 MCS. Upon confirming that the averages of spins remain almost constant for two consecutive periods of 1000 MCS, the simulation is stopped. The output patterns are easily derived from the averages of local spins by the conversion (16) and a few of them are shown in figure 2(c) for $\mu = 1$–6. We can see that these output patterns acquired by the two methods for the same input patterns are qualitatively in agreement. This suggests that the Monte Carlo simulation for calculating the self-fields self-consistently gives a good approximation to the solution of the mean field equations.

It is to be noted that in the self-consistent Monte Carlo simulation local effective fields calculated in an averaging period of MCS change gradually from those calculated in the previous averaging period of MCS, while in an iterative solution of the mean field equations, the self-fields sometimes change drastically from those in the previous step. This gives a kind of stability to the self-consistent Monte-Carlo simulation. We consider the self-consistent Monte Carlo simulation to be a promising method of obtaining various results of self-organization by starting from various random and other appropriate initial conditions. A more detailed discussion concerning the comparison of solutions of the mean field equations and results of the self-consistent Monte Carlo calculation will be reported in another paper.

7. Conclusions

The detailed analysis of the steady-state equations describing the self-organization in the two-layered model of formal neurons leads to the following conclusions.

(1) The response of the output neurons for each input pattern in the steady state is described by a set of the mean field equations which arise in the statistical mechanics of magnetism.

(2) The mean field equations are derived in mean field approximation from the Hamiltonian of the Ising spin system and the self-consistency conditions.

(3) The response of the output neurons is represented by a thermal state at a fixed temperature of the Ising spin system which is described by the Hamiltonian in three-dimensional lattice space.

(4) The self-consistent Monte Carlo simulation is a new method of simulation and gives a satisfactory result as an approximation to the mean field equations.

In the Hamiltonian formalism, the distribution of upward self-field within the plain of $z = \mu$ represents the distribution of the active neurons responding to the input pattern μ. The formalism provides us with an overall perspective point of view by combining the neural network and the external environments represented by the correlations between input patterns. We expect that the self-consistent Monte Carlo simulation is a promising method of finding self-organization by formal neurons, because it provides an effective way of determining a spin structure at a fixed temperature, i.e. a response property of the neural network.

Acknowledgments

The authors thank Professor S Amari for illuminating discussions. One of authors (SI) thanks Professor R B Griffiths for his help in correcting the manuscript. He is also grateful to the College of Engineering, Nihon University for providing him with a Grant-in-Aid for academic research.

References

[1] Hubel D H and Wiesel T N 1962 Receptive fields, binocular interaction, and functional architecture in cat's visual cortex *J. Physiol.* **160** 106–54

[2] Blasdel G G and Salama G 1986 Voltage-sensitive dyes reveal a modular organization in monkey striate cortex *Nature* **321** 579–85

[3] Bonhoeffer T and Grinvald A 1991 Iso-orientation domains in cat visual cortex are arrange in pinwheel-like patterns *Nature* **353** 429–31

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

[5] Takeuchi A and Amari S 1979 Formation of topographic maps and columnar microstructures in nerve fields *Biological Cybernetics* **35** 63–72

[6] Linsker R 1986 From basic network principles to neural architecture (series) *Proc. Natl Acad. Sci.* **83** 7508–12, 8390–4, 8779–83

[7] Miyashita M and Tanaka S 1992 A mathematical model for the self-organization of orientation columns in visual cortex *Neuroreport* **3** 69–72

[8] 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 *J. Neurosci.* **14** 409–41

[9] Tamori Y 1991 Self-organization of neural network with lateral inhibitions *Dissertation* Department of Applied Physics, Tohoku University, Sendai, Japan (in Japanese)

[10] Inawashiro S, Musha Y, Miyake S and Matsubara F 1993 The Self-organization of the orientation selectivity through learning by inhibitory interneurons *Proc. Int. Joint Conf. on Neural Networks (Nagoya, Japan)* pp 2524–7

[11] Hebb D O 1949 *The Organization of Behaviour* (New York: Wiley)

[12] Kohonen T 1989 *Self-Organization and Associative Memory* 3rd edn (Berlin: Springer) p 123

[13] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders) pp 715–8

[14] Personnaz L, Guyon I and Dreyfus G 1985 Information storage and retrieval in spin-glass like neural networks *J. Physique Lett.* **46** L359–65

[15] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 Equations of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92

[16] Binder K and Heerman D W 1988 *Monte Carlo Simulation in Statistical Mechanics* (Berlin: Springer)

[17] Suzuki M, Shirakura T and Inawashiro S 1991 AC susceptibilities and relaxation time distributions for the two-dimensional and three-dimensional $\pm J$ Ising spin glasses *J. Phys.: Condens. Matter* **3** 3139–54