Biol. Cybern. 57, 197–206 (1987)

Biological Cybernetics

A Cognitive and Associative Memory

S. Shinomoto

Department of Physics, Kyoto University, Kyoto, 606 Japan

Abstract. By introducing a physiological constraint in the auto-correlation matrix memory, the system is found to acquire an ability in cognition i.e. the ability to identify an input pattern by its proximity to any one of the stored memories. The physiological constraint here is that the attribute of a given synapse (i.e. excitatory or inhibitory) is uniquely determined by the neuron it belongs. Thus the synaptic coupling is generally not symmetric. Analytical and numerical analyses revealed that the present model retrieves a memory if an input pattern is close to the pattern of the stored memories; if not, it gives a clear response by going into a special mode where almost all neurons are in the same state in each time step. This uniform mode may be stationary or periodic, depending on whether or not the number of the excitatory neurons exceeds the number of inhibitory neurons.

1 Introduction

Many intriguing mathematical models for the associative memory of the human brain have been suggested in the last three decades. Though the relationship between such models and the real nervous systems is yet to be clarified, one of the promising and interesting models is the correlation matrix memory, which has been studied intensively by many researchers (Kohonen 1972, 1984; Kohonen et al. 1976; Nakano 1972; Anderson 1972; Cooper 1973; Amari 1977; Little and Shaw 1978; Hopfield 1982, 1984).

The auto-correlation matrix memory is characterized by a rapid retrieval of its own memory when it is closely related to an applied input pattern. The process of the retrieval was discussed by Hopfield (1982) who adopted an asynchronous processing algorithm applied to the system composed of the McCulloch-Pitts or binary elements. The dynamics then bears a strong resemblance to the Monte-Carlo relaxation processes of the spin glasses at zero temperature, where the evolution proceeds until the system finds one of the local minima of the Lyapunov or the energy function. Hopfield also estimated the limit of the ratio between the number of memories M to the total number of elements N, as $\alpha = M/N \sim 0.15$, below which the system can retrieve stored memories almost correctly (see also Amit et al. 1985).

In spite of a number of its interesting properties, what is called the Hopfield model still has plenty of room for improvement. Various attempts appeared which led to some improvements of Hopfield's theory (Hopfield et al. 1983; Fukushima 1984, 1986; Dotsenko 1985; Parga and Virasoro 1986; Tsuda et al. 1987). Among others, we take up here the problems of

(*a*) how to include physiological constraint, and

(*b*) how to provide the system with the cognitive ability to distinguish whether or not an input signal is close to any of the stored memories.

There have been several attempts about these. It was pointed out that (*a'*) the retrieval is possible even in the presence of a physiological constraint such that each synaptic coupling does not change its sign in the course of learning (Toulouse et al. 1986; Personnaz et al. 1986; Buhman and Schulten 1986), and (*b'*) the cognitive ability such as (*b*) is possible by adding an asymmetric random matrix to the Hopfield coupling (Parisi 1986). These attempts, however, appear to be still insufficient for the full resolution of the problems (*a*) and (*b*).

The physiological constraint in the mammalian central nervous system is supposed to be stronger than that of the previous models proposed in relation to (*a'*). It is found from physiological observations that the attribute of a given synapse (i.e., excitatory or inhibitory) is mainly determined uniquely by the kind of neuron it belongs (Eccles 1977; Kuffler et al. 1984). The so called Dale's principle states that the same transmitter is liberated from all synapses belonging to the same neuron. Though the exceptions to Dale's principle have been found, it may be said that in most cases

the attribute of the synaptic coupling, apart from the kinds of chemical transmitters, is uniquely determined by the neuron which sends signals.

A model under the above physiological constraint is proposed in the present paper. The statistical dynamics of some macroscopic quantities in the synchronous processing algorithm is investigated analytically by a generalized version of Amari and Kinzel's method which was originally developed for the synchronous processing of the auto-correlation matrix memory (Amari 1977; Kinzel 1985). As a result, it is found that our model system may have an additional global basin or a pair of the global basins apart from the basins of stored memories. An input pattern which has almost no correlation with any of the stored memories (i.e. the pattern which is out of basins of memories) eventually approaches a special mode which is irrelevant to stored memories. This irrelevant mode is characterized by the simultaneous firing or resting over almost all neurons in each time step. We shall call this a uniform mode. The firing or resting state persists in the excitatory-dominant case, which means the number of excitatory neurons exceeds the number of inhibitory neurons. Contrary to this, in the inhibitory-dominant case, the system becomes periodic where the firing and resting states appear alternatively. Some underlying mechanisms of microscopic origin is investigated by numerical simulations. It is found that our model actually works even better than expected from analytical arguments. Thus the two problems (*a*) and (*b*) are resolved simultaneously in our model.

The present paper is organized as follows. In Sect. 2, the Hopfield model is introduced, and its statistical dynamics of the process of retrieval is described. Our model is introduced in Sect. 3. We shall discuss the statistical dynamics of the macroscopic quantities, leaving its derivational details to the Appendix. Bifurcation scenario leading to the coexistence of the basins of retrieval and uniform mode is derived. In Sect. 4, some results of our numerical experiments are shown and compared to the same calculation using the Hopfield model. Thus, some advantages of our model compared to previous models become clear. Discussions on our model in relation to real nervous systems are presented in the final section.

2 Process of Retrieval of the Auto-Correlation Matrix Memory

2.1 The McCulloch Pitts Model of a Neuron

Each element is assumed to have two states whose symmetric representation is

$$s_j=\begin{cases}+1: j\text{-th neuron is firing,}\\ -1: \text{non-firing or resting,}\end{cases}\quad (j=1,\ldots,N).$$

The firing neuron sends its signal to the others via its synaptic coupling, and the post-synaptic potential of the i-th neuron aroused by the j-th neuron is given by $K_{ij}\times(s_j+1)$, where $2K_{ij}$ or K_{ij} is the synaptic strength. Each neuron fires if and only if the sum of such post-synaptic potentials or a membrane potential, $U_i=\sum_j K_{ij}(s_j+1)$, exceeds its own threshold value H_i.

Thus, each element readjusts its state according to the rule

$$s_i\to\mathrm{sgn}(v_i), \tag{1}$$

if the reduced input signal, $v_i=\sum_j K_{ij}(s_j+1)-H_i$, is finite. For the sake of simplicity, we shall rewrite the input signal as

$$v_i=\sum_j K_{ij}s_j-L_i, \tag{2}$$

where $L_i=H_i-\sum_j K_{ij}$ is the reduced threshold value.

2.2 Modes of Processing Algorithms

There are two representative modes of processing algorithm for the evolution rule (1). In the synchronous processing algorithm, the rule (1) is applied simultaneously to all the neurons or

$$s_i(t+1)=\mathrm{sgn}(v_i(t)),$$
$$v_i(t)=\sum_j K_{ij}s_j(t)-L_i.$$

The set of the above equations is a nonlinear transformation of the vector, $\mathbf{s}(t)=(s_1(t),\ldots,s_N(t))$, to another vector $\mathbf{s}(t+1)$, i.e.,

$$\mathbf{s}(t+1)=\mathbf{f}(\mathbf{s}(t)), \tag{3}$$

where, $\mathbf{f}(\mathbf{s})=\mathrm{sgn}(\mathbf{v})$, and $\mathbf{v}=\overleftrightarrow{K}\mathbf{s}-\mathbf{L}$. Let us remark here some symmetry included in this rule. The total input signal $\mathbf{v}$ changes its sign if $\mathbf{s}$ and $\mathbf{L}$ are inverted in sign simultaneously. Thus if $\{\mathbf{s}(t)\}_t$ is the solution of the evolution rule (3) of a system with parameters $\{\overleftrightarrow{K},\mathbf{L}\}$, then $\{-\mathbf{s}(t)\}_t$ is the orbit of the dual system characterized by $\{\overleftrightarrow{K},-\mathbf{L}\}$. In the case $\mathbf{L}=\mathbf{0}$, the system is self-dual. Then, for any orbit $\{\mathbf{s}(t)\}_t$, the system at the same time possesses an orbit $\{-\mathbf{s}(t)\}_t$. These two orbits may be separated from each other, or otherwise fused into one limit-cycle of even period.

Secondly, in the asynchronous processing algorithm, the rule (1) is applied randomly to every neuron at a given time with the mean attempt rate $W(<1)$. No symmetry as mentioned above exists, but in statistical sense such symmetry may be restored.

2.3 The Auto-Correlation Matrix Memory or the Hopfield Model

The auto-correlation matrix is an ad-hoc choice of the synaptic coupling as

$$T_{ij}=\begin{cases}(1/N)\sum_m s_i^{(m)}s_j^{(m)}, & \text{if } i\neq j,\\ 0, & \text{otherwise},\end{cases} \tag{4}$$

where, $\mathbf{s}^{(m)}=(s_1^{(m)},\ldots,s_N^{(m)})$ is the m-th firing pattern to be memorized in the network ($m=1,\ldots,M$). The factor $(1/N)$ in (4) is inserted simply to normalize the input signal and is irrelevant to the dynamics in the systems without noises. We have introduced the notation T_{ij} in place of K_{ij} in order to avoid confusion because the same notation K_{ij} appears in our model to be presented in Sect. 3.

Assuming the mutual independence of the memories, we have a pseudo-orthogonality;

$$\sum_j s_j^{(m)}s_j^{(\mu)}/N=\delta_{m\mu}\pm(1-\delta_{m\mu})O(1/\sqrt{N}).$$

If one applies one of the memories $\mathbf{s}=\mathbf{s}^{(\mu)}$ as an initial input pattern, the reduced input signal to the i-th neuron becomes

$$v_i=s_i^{(\mu)}-L_i\pm O(\sqrt{M/N}). \tag{5}$$

Thus in the case $\alpha=M/N\ll 1$, it is possible by choosing $\mathbf{L}=\mathbf{0}$ to let each of the memory pattern be a fixed point. Hopfield pointed out that a Lyapunov function which decreases monotonically until the system finds its stable fixed point exists, provided the coupling is symmetric, i.e. $T_{ij}=T_{ji}$, and the processing algorithm is asynchronous. It was also found that each memory pattern becomes a global attractor if $\alpha\lesssim 0.15$ and thus a small difference of an initial pattern $\mathbf{s}$ from $\mathbf{s}^{(\mu)}$ is washed out in the course of the processing (Hopfield 1982). This process is called the retrieval of the memory or the association.

2.4 Process of the Retrieval

Process of the retrieval in the autocorrelation matrix memory was analysed by Amari (1977) and Kinzel (1985) though in the synchronous processing algorithm. Before making use of this idea for our case (which will be done in the next section) we review it with some modifications.

Let us define a direction cosine c_μ between a pattern $\mathbf{s}$ and a memory pattern $\mathbf{s}^{(\mu)}$, by

$$c_\mu=\mathbf{s}\cdot\mathbf{s}^{(\mu)}/N=(1/N)\sum_j s_j s_j^{(\mu)}. \tag{6}$$

Some other macroscopic quantities are related to this quantity, e.g., direction angle $\Phi_\mu=\arccos(c_\mu)$, and the Hamming distance $D_\mu=N(1-c_\mu)/2$.

Let us next define a referenced input signal $\mathbf{u}=\{u_i\}_i$, for a specific test pattern $\mathbf{x}$, as $u_i=x_i v_i$, where $x_i=+1$ or -1 and v_i is the input signal to the i-th neuron (2). Thus the test pattern $\mathbf{x}$ is identical to the resulting pattern $\tilde{\mathbf{s}}$ if all u_i's are positive. For general $\mathbf{x}$, the direction cosine between $\mathbf{x}$ and $\tilde{\mathbf{s}}$ is given by

$$\mathbf{x}\cdot\tilde{\mathbf{s}}/N=(1/N)\sum_i \operatorname{sgn}(x_i v_i)=\int_{-\infty}^{\infty} du\varrho(u)\operatorname{sgn}(u), \tag{7}$$

where

$$\varrho(u)=(1/N)\sum_i \delta(u-u_i) \tag{8}$$

is a normalized distribution function of u_i's. The above direction cosine is identical to the resulting direction cosine $\tilde{c}_\mu$ if we put $\mathbf{x}=\mathbf{s}^{(\mu)}$.

The problem of finding the evolution process of c_μ for specific $\mathbf{s}$ and $\overleftrightarrow{T}$ is equivalent to the problem of finding original evolution for $\mathbf{s}$ (3). One will find, however, some simplification for the statistical evolution of c_μ if one introduces an ensemble of the set of memory patterns $\{\mathbf{s}^{(m)}\}_m$. Assume the statistical independence between the elements of the vector, i.e.

$$\langle s_j^{(m)}s_{j'}^{(m')}\rangle_{c_\mu}=\delta_{jj'}\delta_{mm'},$$

where the square bracket represents an average operation over the ensemble, with the restriction that $\mathbf{s}\cdot\mathbf{s}^{(\mu)}/N$ is kept constant at the value c_μ, or more strongly,

$$\langle s_j s_{j'}^{(m)}\rangle_{c_\mu}=c_\mu\delta_{jj'}\delta_{m\mu}.$$

If $N\gg 1$, we may safely approximate $\varrho(u)$ by the normal distribution;

$$\varrho(u)=(\sqrt{2\pi}\,\Delta u_\mu)^{-1}\exp[-(u-\bar{u}_\mu)^2/2(\Delta u_\mu)^2], \tag{9}$$

where the mean value and the variance of u_i are readily evaluated as

$$\bar{u}_\mu=\langle u_i\rangle_{c_\mu}=c_\mu,$$
$$(\Delta u_\mu)^2=\langle(u_i-\bar{u}_\mu)^2\rangle_{c_\mu}=M/N=\alpha,$$

the terms of $O(1/N)$ being neglected in the above.

Thus we finally get the evolution equation for the direction cosine $c_\mu(t)$ in the form of a map

$$c_\mu(t+1)=\sqrt{\frac{2}{\pi}}\int_0^{v_\mu} dx\, e^{-x^2/2}, \tag{10}$$

where $v_\mu=\bar{u}_\mu/\Delta u_\mu=c_\mu(t)/\sqrt{\alpha}$. The right-hand side is a monotonically increasing function of c_μ and thus the map shows a pitchfork bifurcation only once as α is decreased. A trivial fixed point $c_\mu=0$ becomes unstable if $\alpha=M/N<\alpha_c=2/\pi$, and then, a pair of stable fixed points $c_\mu=c_\mu^*(>0)$ and $-c_\mu^*(<0)$ appear. The appearance of the pair implies the appearance of the global multibasin structure for all μ's. Especially, in the

case $\alpha \ll 2/\pi$, c_μ^* is near unity and the Hamming distance is small. The critical parameter value $\alpha_c = 2/\pi \sim 0.64$ is large compared to the criterion of Hopfield, $\alpha \sim 0.15$, for which the fairly good proximity of the retrieved pattern to the memory is required.

3 Cognitive Memory Composed of Excitatory and Inhibitory Neurons

3.1 Model

We assume that the sign of the synaptic coupling is uniquely determined by the alternative attributes of the neurons which send signals, as discussed in Sect. 1. In other words, the sign of the synaptic couplings $\{K_{ij}\}_i$ is the same over all i's for a given j. We shall note $\zeta_j = +1$, if $K_{ij} \geqq 0$ and then the j-th neuron is called excitatory; if $\zeta_j = -1$, (i.e. if $K_{ij} \leqq 0$) the j-th neuron is called inhibitory. This property should be contrasted to the previous heterogeneous neuron, whose synaptic couplings may be positive or negative for a given j (see Fig. 1).

We take this fact into account and assume

$$K_{ij} = 2T_{ij}\theta(\zeta_j T_{ij}), \tag{11}$$

where T_{ij} is the synaptic coupling of the Hopfield model (4) and $\theta(x)$ is the Heaviside step function ($\theta(x)=1$, if $x \geqq 0$ and $=0$, otherwise). Here, the coupling T_{ij} which is not obedient to the attribute of the j-th neuron has been made to vanish (see Fig. 2). The numerical factor 2, appearing above is simply to normalize the input signal and is again irrelevant to the present processing algorithm without noises. The reduced threshhold L_i is assumed to be zero. We consider the case that the excitatory and inhibitory neurons are distributed independently. An additional parameter characterizing this system is the number of inhibitory neurons, N_I, or that of excitatory neurons, N_E. For the sake of symmetric representation, we introduce a parameter q by

$$q = (N_E - N_I)/N = \sum_j \zeta_j / N, \qquad -1 \leqq q \leqq 1. \tag{12}$$

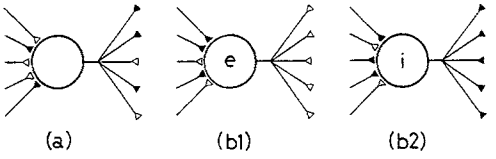

Fig. 1a and b. Diagrammatic representation of neurons. Central circle and line represent the cell body and axon while initial segment of the axon is sometimes omitted. Excitatory and inhibitory synapses are represented by △ and ▲, respectively. Signal produced in the cell body is transmitted to the other cells via axon. **a** A neuron with heterogeneous attributes. **b1** and **b2** Excitatory and inhibitory neurons we shall adopt in the present model

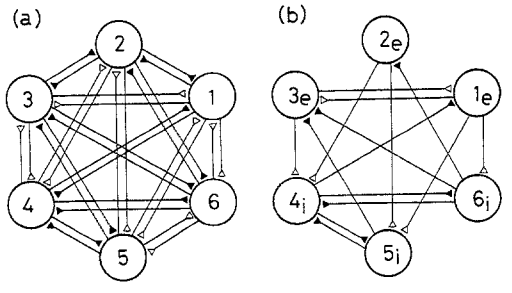

Fig. 2a and b. Examples of the diagrams of the neural networks of $N=6$. **a** An example of the Hopfield model $\{T_{ij}\}$, **b** coupling $\{K_{ij}\}$ formed from the $\{T_{ij}\}$ by our rule ($\zeta_i = 1$ for $i = 1, 2, 3$, and -1 for $i = 4, 5, 6$)

3.2 Preliminary Consideration

Though the coupling in our model is asymmetric the retrieval is still possible. The i-th input signal obtained by the application of a memory pattern $\mathbf{s} = \mathbf{s}^{(\mu)}$ is

$$v_i^\mu = (2/N) \sum_m s_i^{(m)} \sum_{j(\neq i)} s_j^{(m)} s_j^{(\mu)} \theta\left(\zeta_j \sum_l s_i^{(l)} s_j^{(l)}\right).$$

In the absence of correlation between ζ_j and T_{ij}, the argument of the Heaviside step function may be positive or negative with equal probability. Thus the above quantity is estimated as

$$v_i^\mu \sim s_i^{(\mu)} \pm O(\sqrt{2M/N}). \tag{13}$$

Note that the magnitude of the fluctuation of v_i^μ differs from that of the Hopfield model by the factor $\sqrt{2}$. This is because the effective number of summands is reduced to the half of N due to the presence of the Heaviside step function. Still the retrieval of the memory seems possible also in our model.

On the other hand, in the presence of imbalance between the populations of excitatory and inhibitory neurons, which means $q \neq 0$, our model with synchronous processing algorithm may have other fixed points or a limit cycle orbit irrelevant to the stored memories. For instance, by the application of $\mathbf{1} = (1, \ldots, 1)$ as an input pattern, the input signal becomes

$$v_i^+ = \sum_j 2T_{ij}\theta(\zeta_j T_{ij}) \sim \sum_j T_{ij}[(1+q)\theta(T_{ij}) + (1-q)\theta(-T_{ij})].$$

If T_{ij}'s are equally distributed around zero with variance $(\Delta T)^2 = (\sqrt{M/N})^2$, then v_i^+ is approximately given by

$$v_i^+ \sim q\sqrt{2M} \pm O(\sqrt{2M/N}). \tag{14}$$

Thus if $0 < q \lesssim 1$ (excitatory-dominant) there is a pair of fixed points, $\mathbf{s}^* \sim \mathbf{1}$ and $-\mathbf{1}$. On the contrary, if $0 > q \gtrsim -1$ (inhibitory-dominant), the system has a periodic solution with period 2 such that $\mathbf{1}$ and $-\mathbf{1}$ appear alternatively.

3.3 Statistical Dynamics of the Direction Cosines

From the above preliminary consideration, the existence of two kinds of global attractors are expected: one corresponding to the retrieval of the memories and the other to the uniform mode.

One may find statistical dynamics of the set of direction cosines c_μ and c_+ defined by

$$c_\mu = \mathbf{s}\cdot\mathbf{s}^{(\mu)}/N, \tag{15}$$

$$c_+ = \mathbf{s}\cdot\mathbf{1}/N, \tag{16}$$

under the condition that the direction cosine $c_{\mu+}$ between $\mathbf{s}^{(\mu)}$ and $\mathbf{1}$ is fixed at some value (see Fig. 3). We leave the exact derivation of the map of

$$\mathbf{c}(t) = (c_\mu(t),\, c_+(t))^T \tag{17}$$

to the Appendix, and discuss the resulting dynamics, especially in the case $c_{\mu+}=0$.

The referenced input signals u_i^μ and u_i^+ whose test patterns are $\mathbf{s}^{(\mu)}$ and $\mathbf{1}$ repectively are characterized by their means,

$$\bar{u}_\mu = c_\mu,$$

$$\bar{u}_+ = q\sqrt{2M}\,c_+,$$

and variances,

$$\langle (u_i^\mu - \bar{u}_\mu)^2\rangle = q^2 2M c_+^2 + 2M/N,$$

$$\langle (u_i^+ - \bar{u}_+)^2\rangle = c_\mu^2 + 2M/N,$$

$$\langle (u_i^\mu - \bar{u}_\mu)(u_i^+ - \bar{u}_+)\rangle = q\sqrt{2M}\,c_\mu c_+,$$

where c_μ and c_+ stand for the set of the direction cosines of the input pattern $\mathbf{s}(t)$, i.e., $c_\mu(t)$ and $c_+(t)$. Then the two-dimensional map of $\mathbf{c}(t)$ is obtained as

$$c_\mu(t+1) = \int\int du_\mu du_+ \varrho(u_\mu, u_+)\,\mathrm{sgn}(u_\mu), \tag{18}$$

$$c_+(t+1) = \int\int du_\mu du_+ \varrho(u_\mu, u_+)\,\mathrm{sgn}(u_+), \tag{19}$$

where $\varrho(u_\mu, u_+)$ is the normal distribution with the means and variances given above. Let this nonlinear map be expressed by

$$\mathbf{c}(t+1) = \mathbf{g}(\mathbf{c}(t)). \tag{20}$$

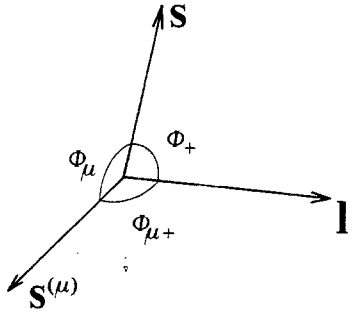

Fig. 3. Schematic representation of the vectors in N-dimensional vector space

In the equations for the means and variances, c_+ is always accompanied by q which may be positive or negative, and the map may exhibit a subharmonic bifurcation leading to the periodic motion of period two. The sign of q becomes irrelevant if we consider the iterated map,

$$\mathbf{c}(t+2) = \mathbf{G}(\mathbf{c}(t)) = \mathbf{g}(\mathbf{g}(\mathbf{c}(t))). \tag{21}$$

Let the fixed point of this map $\mathbf{G}$ be denoted by $\mathbf{c}^*$, i.e.

$$\mathbf{c}^* = \mathbf{G}(\mathbf{c}^*). \tag{22}$$

The linearized map around this fixed point is

$$\delta\mathbf{c}(t+2) = \overleftrightarrow{\partial\mathbf{G}/\partial\mathbf{c}}\,\Big|_{\mathbf{c}^*}\,\delta\mathbf{c}(t)$$

$$= \overleftrightarrow{\partial\mathbf{g}/\partial\mathbf{c}}\,\Big|_{\mathbf{g}(\mathbf{c}^*)}\,\overleftrightarrow{\partial\mathbf{g}/\partial\mathbf{c}}\,\Big|_{\mathbf{c}^*}\,\delta\mathbf{c}(t). \tag{23}$$

The map has a trivial fixed point at the origin, $\mathbf{c}^*=\mathbf{0}$. Here, the matrix $[\overleftrightarrow{\partial\mathbf{G}/\partial\mathbf{c}}]$ is diagonal and has the eigenvalues, $1/\pi\alpha$ and $2q^2M/\pi\alpha$. Thus the origin is a stable node if $1/\pi<\alpha$ and $M<\pi\alpha/2q^2$. When one of the above inequalities is reversed as the change of the parameters (N, M, and q), a pair of nodes appear on the instability axis through the pitchfork bifurcation, and the origin becomes a saddle point. Next, the origin becomes an unstable node if

$$1/\pi > \alpha \quad \text{and} \quad M > \pi\alpha/2q^2, \tag{24}$$

and a pair of saddle points appear on the second instability axis. The scenario is depicted schematically in Fig. 4a–c. The symmetry of these bifurcations comes from the self-duality explained in Sect. 2.2 and the condition $c_{\mu+}=0$.

The most interesting case appears after each saddle point on the axis of the second instability becomes a stable node by producing a pair of saddle points (see Fig. 4d). This final flow diagram implies the coexistence of the basins corresponding to the retrieval and the uniform mode. Thus, if we choose the parameters so as to make the basins of retrieval sufficiently small, our system has a cognitive ability to distinguish whether or not an input signal is close to any one of the memories. In the uniform mode, almost all the neuron

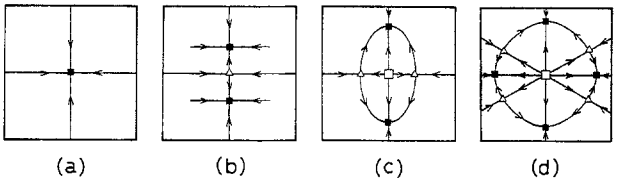

Fig. 4a–d. Bifurcation scenario leading to the coexistence mode (from **a** to **d**). In each flow diagram, ■, □, and △ represent stable node, unstable node, and saddle point, respectively. Shaded region in **d** is the basin of retrieval if the horizontal axis is c_μ

states are identical. They are fixed in time in the excitatory-dominant case ($q>0$) or shows a flip-flop motion with time in the inhibitory-dominant case ($q<0$).

Finally, let us derive the stability conditions for the retrieval state $(c_\mu^*, 0)$ and the uniform mode $(0, c_\pm^*)$. The off-diagonal elements of the matrix $[\overleftarrow{\partial \mathbf{G}/\partial \mathbf{c}}]$ vanish in each of the fixed points, and the eigenvalues are

$$1/\pi\alpha \cdot \exp(-c_\mu^{*2}/2\alpha), \quad \text{and} \quad 4q^2M/[\pi(2\alpha+c_\mu^{*2})], \tag{25}$$

for the fixed point $(c_\mu^*, 0)$ and

$$1/[\pi(\alpha+q^2Mc_\pm^{*2})], \quad \text{and} \quad 2q^2M/\pi\alpha \times \exp(-c_\pm^{*2}q^2M/\alpha), \tag{26}$$

for the fixed point $(0, c_\pm^*)$. In order to get the flow diagram as in Fig. 4d, all of these four quantities should be smaller than unity, while the origin should be an unstable node (24).

4 Numerical Experiments

Though we obtained the statistical dynamics of some macroscopic quantities, a more detailed mechanism of a microscopic origin still remains unclear. For instance, although the statistical dynamics of the Hopfield model implies the existence of the global basins of retrieval in $\alpha<2/\pi$, there are at the same time a number of local basins which we call spurious memories. In the computer simulations of the deterministic processing of our model, we also found spurious memories other than the uniform mode, especially in the parameter region where the coexistence of retrieval and uniform mode is possible (Fig. 4d). It was found, however, that our model system can remove many of the spurious memories by choosing the parameter q so that the retrieval state $(c_\mu^*, 0)$ is slightly unstable against the direction of c_+ (Fig. 4c). In the actual simulations, our system is able to retrieve its memories even in this parameter region. For instance, we shall investigate the cases such as $q=\pm 0.5$, $N=200$ and $M=10$. It is readily found that the second eigenvalue in (25) is larger than unity for this set of parameters.

We prepared the synaptic coupling according to (11) with the memory patterns chosen randomly. Input pattern is arranged in a random manner for each of the Hamming distances from a memory pattern $\mathbf{s}^{(\mu)}$. Here the direction cosines c_+, c_m $(m\neq\mu)$, and $c_{\mu+}$ are not taken into consideration for the choice of the patterns, and thus they are expected to be $\pm O(1/\sqrt{N})$, respectively. Note that the present case does not exactly correspond to the previous analysis, because $c_{\mu+}$ is not exactly zero. After τ time steps of the synchronous processing of the system, the direction cosines $c_\mu(\tau)$ and $c_+(\tau)$ and a reversion-activity are calculated. The reversion activity is defined as the ratio of the number of neurons which should flip in the next time, i.e., $N_R(\tau) = \sum_m \theta(-s_i(\tau)v_i(\tau))$, to the total number of neurons N.

Usually, the state $\mathbf{s}(t)$ of the system enters a final limit cycle or a fixed point after transient steps of $O(1)$ in the present model. In general, there may be transient

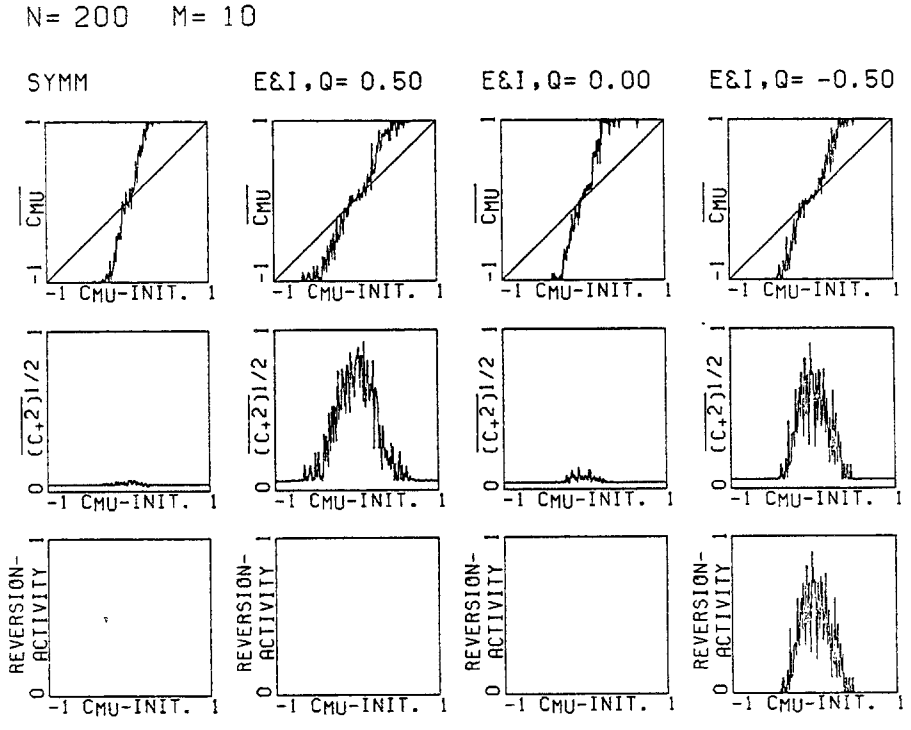

Fig. 5. Results of numerical simulations of the Hopfield model and the present model ($N=200$, $M=10$). Horizontal axis for each graph stands for the direction cosine $c_\mu(0)$ of the input pattern. See the text for further detail

or chaotic phenomena where the transient length or the period of a final limit cycle orbit is of the non-polynomial order of N (Shinomoto 1986a). The latter problem is not the matter of present concern, and the processing of steps of $O(10)$ is sufficient to eliminate the transient. In the present calculation, $\tau = 20$.

One of the results of our simulation is shown in Fig. 5. The direction cosine $c_\mu(\tau)$ and the root mean square of the direction cosine $c_+(\tau)$, and the reversion activity are shown in each row, where the bar means the average over 10 samples. The first column is the result of the Hopfield model. The results of our model for several values of parameter q are drawn in the right side.

Previously mentioned cognition ability is observed in our model for $q = 0.5$ and -0.5. In fact, in both the excitatory-dominant and inhibitory-dominant cases, the mean square of c_+ shows a marked increase for

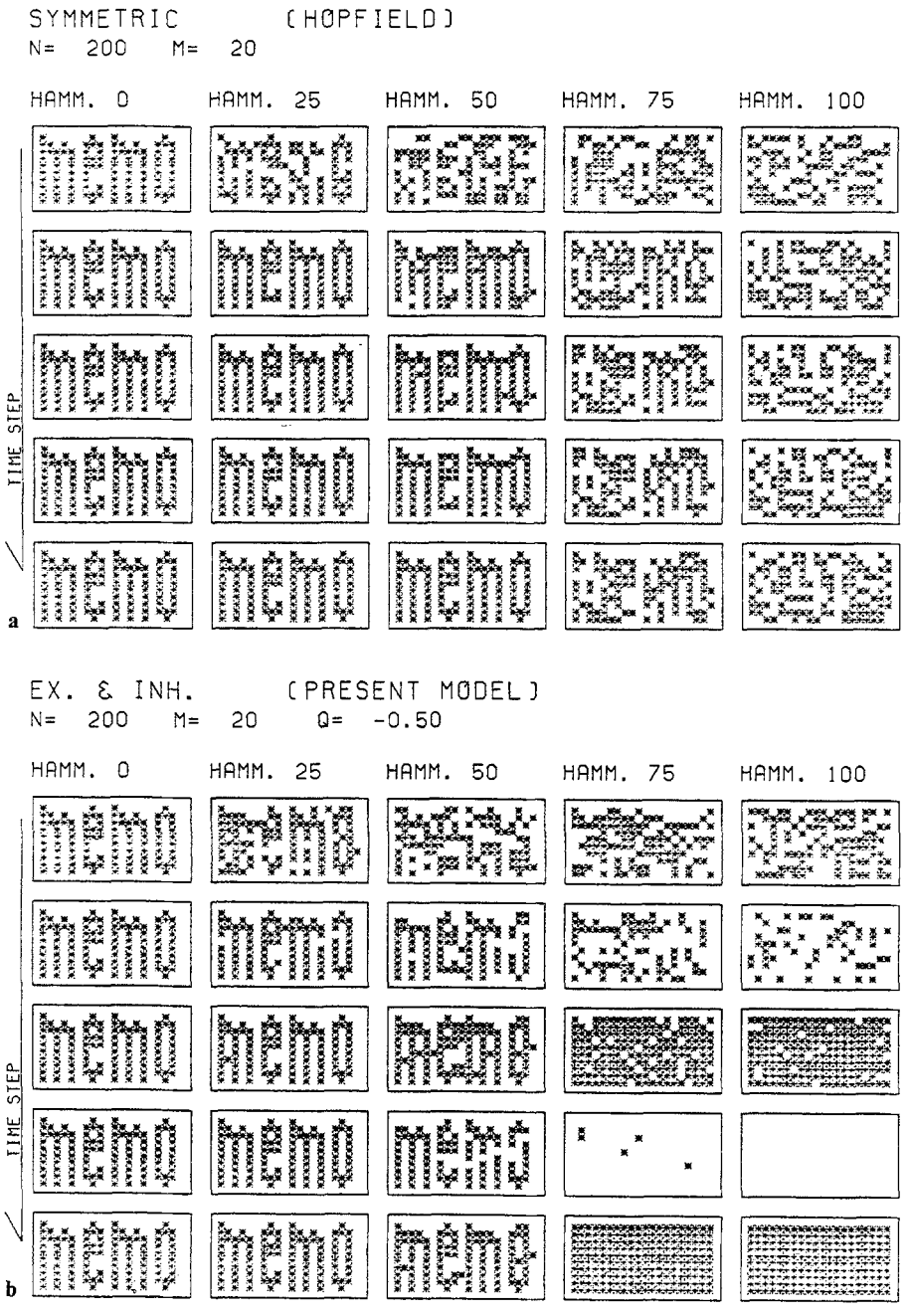

Fig. 6a and b. Retrieval processes of **a** the Hopfield model and **b** the present model ($N = 200$, $M = 20$). It is observed that the indistinctive patterns are finally trapped by a spurious memory in the former model, while they get into the uniform mode in our model. The number following to the character "HAMM." is the Hamming distance of the input pattern from the specific memory designed as "memo"

small $|c_\mu(0)|$. Thus the input pattern which is not associated with any of the memories is washed out to give a uniform pattern. In the inhibitory-dominant case $q=-0.5$, the final uniform mode is the periodic state oscillating between the uniformly firing state and the uniformly resting state. Thus the graph for the reversion-activity in this case is similar to the one of mean square of c_+.

Finally, some of the simulation results on the Hopfield model and ours are compared in Fig. 6. We prepared a specific memory pattern "memo", while other memory patterns are chosen randomly. We have shown how the input pattern changes when its initial state is chosen randomly with the Hamming distance to the pattern fixed at each value. Some differences between the Hopfield model and ours arise when the initial input pattern largely deviates from the memory. In the Hopfield model, the pattern is then trapped by one of the spurious memories, while in our model the system assumes a uniform pattern. The latter is considered as the system's clear statement that the input pattern is not identifiable.

5 Discussion

We thus succeeded in obtaining a cognitive and associative memory by introducing the physiological constraint into the previous auto-correlation matrix memory. What we call the cognitive ability is the ability to identify an input pattern by its proximity to any one of the stored memories. Our system shows a clear response if the input is indistinctive to any of the memories. This point seems to be a great advantage compared with the Hopfield model in which the indistinctive pattern is eventually trapped by a spurious memory or otherwise any one of the memories is forced to be taken out.

Here, we note the relationship between the mathematically simple McCulloch-Pitts elements and real neurons. A real neuron takes roughly three states, i.e. firing, refractory, and resting states. During the retractory period of $O(1$ ms) after firing, a neuron is unable to fire or depolarize even if other firing conditions are satisfied. A signal produced by firing is transferred to the other neurons with a synaptic delay of $O(1$ ms). There are several mathematical models which takes the above fact into account (see for instance, Caianiello 1961). Then, the theory becomes highly complex compared with the present treatment. It would be possible, however, to avoid the difficulty by introducing the asynchronous processing algorithm with noises (Perreto and Niez 1986; Shinomoto 1986b).

First, we shall discuss the asynchronous processing of our model. The statistical dynamics in Sect. 3 can be extended to this case as far as the macroscopic quantities such as direction cosines are concerned. We generalize the evolution equation for the direction cosines (21), as

$$\mathbf{c}(t+1)=W\mathbf{g}(\mathbf{c}(t))+(1-W)\mathbf{c}(t).$$

Each fixed point $\mathbf{c}^*$ of the original equation (20) thus remains unchanged even if $0<W<1$. The main change caused by the asynchrony is the stability of the fixed point. The symmetry of the iterated map $\mathbf{G}$ on q is lost and especially, the appearance of the periodic state in $q<0$ occurs at even smaller q. Note, however, that the decrease of W from 1 does not change the period (two) of the periodic state. This is because there are only the alternative attributes of the neurons, excitatory and inhibitory, and this should be contrasted with the case where longer period appears (Wilson and Cowan 1972; Amari 1971, 1982).

Secondly, we shall note the reason for the necessity of introducing noises. Although a probabilistic nature is introduced in the asynchronous processing, the algorithm is still too restrictive. In fact, once a fixed point $\mathbf{s}^*=\mathbf{f}(\mathbf{s}^*)$ is attained, the above algorithm no longer changes the state. It would be more realistic to assume that the rule (1) is obscured in some sense if the absolute value of the input signal v_i is too small. The obscurity here arises mainly from the microscopic dynamics of the neuron, and it would be natural to substitute some stochasticity for the microscopic dynamics. A possible form of the probability function of readjustment is suggested by Little (1974). With the inclusion of weak noises, the system can escape from the local basin of a supurious memory mentioned in Sect. 4, and finally enters a global basin. The escape from the global basin will take an enormously long time. Thus the flow diagram such as Fig. 4 will remain useful as far as the noises are sufficiently weak.

Though the relationship of our model with real nervous systems is yet to be clarified, the existence of the uniform mode as obtained in the present study is quite suggestive. It is known that the intensity of the α-rhythm of the human brain is increased when our eyes are closed. This, for instance, implies some similarity of the α-rhythm to our uniform mode of inhibitory-dominant cases. Although such a problem is still beyond the scope of our study, there may be some physiological relevance about the appearance of the uniform mode in response to indistinctive inputs.

Acknowledgements. The author would like to express his gratitude to Y. Kuramoto for his continual support and advices. He is also thankful to S. Amari, I. Tsuda, K. Satoh, T. Todani, S. Nara, G. Parisi, and I. Morgenstern for providing him with information on the recent progress on related fields, and M. Sakurai, and H. Kasai for the knowledge on the physiology.

Appendix

In this Appendix, we shall derive exactly the average and the variance of the referenced input signals, u_i^{μ} and u_i^{+}.

First, we calculate the average of u_i^{μ} and u_i^{+} as

$$\bar{u}_{\mu}=\langle\langle s_i^{(\mu)} v_i\rangle_q\rangle_c,$$

$$\bar{u}_{+}=\langle\langle v_i\rangle_q\rangle_c,$$

where $\langle\ \rangle_q$ and $\langle\ \rangle_c$ represent average operations over the distribution of $\{\zeta_i\}_i$ and $\{s^{(m)}\}_m$ under the fixed values of q and $\mathbf{c}=(c_{\mu}, c_{+}, c_{\mu+})$, respectively.

By using the integral formula of the Heaviside step function,

$$\theta(x)=\lim_{\varepsilon\to+0}\frac{1}{2\pi i}\int_{-\infty}^{\infty}d\omega\frac{1}{\omega-i\varepsilon}e^{i\omega x},$$

one may reduce $\bar{u}_{\mu}$ and $\bar{u}_{+}$ as

$$\bar{u}_{\gamma}=\sum_{j(\neq i)}\frac{1}{2\pi i}\int_{-\infty}^{\infty}d\omega\left[q\frac{1}{\omega}+i\pi\delta(\omega)\right]\frac{1}{i}\frac{d}{d\omega}h_{\gamma}(\omega),$$

where γ stands for μ or $+$ and

$$h_{\mu}=\langle s_i^{(\mu)} s_j\exp(i\omega 2T_{ij})\rangle_c,$$

$$h_{+}=\langle s_j\exp(i\omega 2T_{ij})\rangle_c.$$

h_{μ} and h_{+} are easily calculated by making use of the relations

$$\exp(ivs_i^{(m)}s_j^{(m)})=\cos v+is_i^{(m)}s_j^{(m)}\sin v,$$

and

$$\langle s_i^{(m)}s_i\rangle_c=\delta_{ij}\delta_{m\mu}c_{\mu},$$

$$\langle s_j\rangle_c=c_{+},$$

$$\langle s_j^{(m)}\rangle_c=\delta_{m\mu}c_{\mu+}.$$

Thus we get

$$h_{\gamma}=[a_{\gamma}\cos v+ib_{\gamma}\sin v]\cos^{M-1}v,$$

where $v=2\omega/N$ and

$$a_{\mu}=c_{+}c_{\mu+},\quad\text{and}\quad b_{\mu}=c_{\mu},$$

$$a_{+}=c_{+},\quad\text{and}\quad b_{+}=c_{\mu}c_{\mu+}.$$

In the final integration, an integral

$$A=\frac{1}{\pi}\int_{-\infty}^{\infty}dv\frac{\sin v}{v}\cos^{M-1}v=\prod_{l=1}^{m}\left(1-\frac{1}{2l}\right)$$

appears, where $m=\mathrm{Int}((M-1)/2)$. It is easy to find A in the limit $M\gg 1$, as

$$A\sim m^{-1/2}\sim\sqrt{2/M}.$$

Finally, we obtain

$$\bar{u}_{\mu}=q\sqrt{2M}\,c_{+}c_{\mu+}+c_{\mu},$$

$$\bar{u}_{+}=q\sqrt{2M}\,c_{+}+c_{\mu}c_{\mu+},$$

where we have neglected the terms of $O(1/N)$ or $O(1/M)$.

Secondly, we calculate the second order moments of u_i^{μ} and u_i^{+} to obtain the second order cumulants. The method of calculation is similar to the above. After a somewhat lengthy calculation we get

$$\langle\langle u_i^{\mu 2}\rangle_q\rangle_c=\langle\langle u_i^{+2}\rangle_q\rangle_c=\langle\langle u_i^{2}\rangle_q\rangle_c$$
$$=q^2 2Mc_{+}^2+2q\sqrt{2M}\,c_{+}c_{\mu}c_{\mu+}+c_{\mu}^2+2M/N,$$

and

$$\langle\langle u_i^{\mu}u_i^{+}\rangle_q\rangle_c=\langle\langle s_i^{(\mu)}v_i^2\rangle_q\rangle_c$$
$$=q^2 2Mc_{+}^2c_{\mu+}+2q\sqrt{2M}\,c_{+}c_{\mu}$$
$$+c_{\mu}^2c_{\mu+}+2M/Nc_{\mu+},$$

where we have neglected the terms of $O(1/N)$, $O(1/M)$ or $O(\sqrt{M/N})$.

References

Amari S (1971) Characteristics of randomly connected threshold-element networks and network systems. Proc IEEE 59:35–47

Amari S (1977) Neural theory of association and concept formation. Biol Cybern 26:175–185

Amari S (1982) Competitive and cooperative aspects in dynamics of neural excitation and self-organization. In: Amari S, Arbib MA (eds) Competition and cooperation in neural nets. Springer, Berlin Heidelberg New York, pp 1–28

Amit DJ, Gutfreund H, Sompolinsky H (1985) Storing infinite number of patterns in a spin-glass model of neural networks. Phys Rev Lett 55:1530–1533

Anderson JA (1972) A simple neural networks generating iterative memory. Math Biosci 14:197–220

Buhmann J, Schulten K (1986) Assiciative recognition and storage in a model network of physiological neurons. Biol Cybern 54:319–335

Caianiello ER (1961) Outline of a theory of thought-process and thinking machines. J Theor Biol 2:204–235

Cooper LN (1973) A possible organization of animal memory and learning. In: Lundqvist B, Lundqvist S (eds) Proceeding of the Nobel symposium on collective of physical systems. Academic Press, New York, pp 252–264

Dotsenko VS (1985) Ordered spin glass: a hierarchical memory machine. J Phys C18:L1017–L1022

Eccles JC (1977) The understanding of the brain, 2nd edn. McGraw-Hill, New York

Fukushima K (1984) A hierarchical neural network model for associative memory. Biol Cybern 50:105–113

Fukushima K (1986) A neural network model for selective attention in visual pattern recognition. Biol Cybern 55:5–15

Hopfield JJ (1982) Neural networks and physical systems with emergent collective abilities. Proc Natl Acad Sci USA 79:2554–2558

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci USA 81:3088–3092

Hopfield JJ, Feinstein DI, Palmer RG (1983) Unlearning has a stabilizing effect in collective memories. Nature 304:158–159

Kinzel W (1985) Learning and pattern recognition in spin glass models. Z Phys B – Condensed Matter 60:205–213

Kohonen T (1972) Correlation matrix memories. IEEE Trans. Computers C-21:353–359

Kohonen T (1984) Self-organization and associative memory. Springer, Berlin Heidelberg New York

Kohonen T, Reuhkala E, Mäkisara K, Vainio L (1976) Associative recall of images. Biol Cybern 22:159–168

Kuffler SW, Nicholls JG, Martin AR (1984) From neuron to brain. Sinauer, Massachusetts
Little WA (1974) The existence of persistent states in the brain. Math Biosci 19:101–120
Little WA, Shaw GL (1978) Analytic study of the memory strage capacity of a neural network. Math Biosci 39:281–290
Nakano K (1972) Associatron: a model of associative memory. IEEE Trans Syst Man Cybern SMC-2:380–388
Parga N, Virasoro MA (1986) The ultrametric organization of memories in a neural network. J Phys (Paris) 47:1857–1864
Parisi G (1986) Asymmetric neural networks and the process of learning. J Phys A19:L675–L680
Peretto P, Niez JJ (1986) Stochastic dynamics of neural networks. IEEE Trans SMC-16:73–83
Personnaz L, Guyon I, Dreyfus G, Toulouse G (1986) A biologically constrained learning mechanism in networks of formal neurons. J Stat Phys 43:411–422
Shinomoto S (1986a) Statistical properties of neural networks. Prog Theor Phys 75:1313–1318
Shinomoto S (1986b) Talk at the 41st annual conference of the Physical Society of Japan
Toulouse G, Dehaene S, Changeux JP (1986) Spin glass model of learning by selection. Proc Natl Acad Sci USA, 83:1695–1698
Tsuda I, Koerner E, Shimizu H (1987) Memory dynamics in asynchronous neural networks. Prog Theor Phys (to be published)
Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12:1–24

Received: April 2, 1987

Dr. S. Shinomoto
Department of Physics
Kyoto University
Kyoto
606 Japan