Pattern recognition in a neural network with chaos

Z. Tan and M. K. Ali*
Department of Physics, University of Lethbridge, Lethbridge, Canada T1K 3M4
(Received 17 February 1998; revised manuscript received 6 May 1998)

Chaos is introduced into the Gardner model [J. Phys. A **21**, 257 (1988); **22**, 1969 (1989)] by reducing the number of connections among neurons. It is shown that patterns can be recognized in this chaotic model by means of chaos control focusing on the history of evolution of the states. Fixed points are not required for pattern recognition in this scheme. [S1063-651X(98)11109-1]

PACS number(s): 87.10.+e, 05.45.+b, 75.10.Nr

Chaos in neural networks has attracted much interest in recent years (see, for example, [1–5]). There are speculations that chaos plays important roles in neural networks. However, a definitive study on the role of chaos in neural networks is still missing. A fully developed chaotic network does not have any disjoint open set in the phase space, meaning that every point can be reached from every other point by running the dynamics. How can such a network perform information processing such as pattern recognition? This is an important and nontrivial question to answer. In this work we address this issue of pattern recognition using a model chaotic network obtained from the well known Gardner model [6].

Tsuda [3] has presented a model for a dynamic link of memory in nonequilibrium neural network. Adachi and Aihara, Nagashima *et al.*, and Nara *et al.* [4] have shown that in the presence of chaos the dynamics wanders among the learned patterns. It is not clear how the correct memory is retrieved from the wandering dynamics. Kushibe *et al.* used the parameter control to carry out recognition by reducing a chaotic model to the Hopfield model [5]. In this paper we demonstrate numerically that the history of trajectories contains information about patterns and we propose an approach to extract this hidden information. Our ansatz is based on synchronization of chaotic systems using the feedback method. In what follows, we first introduce chaos in the Gardner model and then discuss the process of pattern recognition. In the Gardner model [6], N neurons are *fully* connected to one another through the synapses J_{ij} $(i,j=1,\dots,N)$. The state $S_i(t)$ of the ith neuron at time t is the spin variable $S_i(t)=\pm 1$. In order to store p patterns, $\xi_i^\mu=\pm 1,\mu=1,\dots,p$, the synapses are trained so that the condition

$$\frac{1}{\sqrt{N}}\xi_i^\mu\sum_{j\neq i} J_{ij}\xi_j^\mu > K \geqslant 0, \quad \mu=1,\dots,p, \tag{1}$$

is satisfied. This model can work as an associate memory: It is capable of remembering a maximum of $2N$ patterns and retrieving these patterns, based on partial information about them [6–10]. The patterns coded in synapses are the fixed points under the dynamics of the system. The storage capacity $\alpha=\lim_{N\to\infty}p/N$ is well known [6] to depend on K as

$$\alpha=\left[\int_{-K}^{+\infty}\frac{(t+K)^2}{\sqrt{2\pi}}e^{-t^2/2}dt\right]^{-1},$$

when the net has a maximal (saturation) storage, as to be discussed in this paper. For example, $K=2.00$, 1.00, and 0.68 correspond to $\alpha=0.20$, 0.52, and 0.76, respectively. The limit $\alpha=2$ is reached when $K=0$. The larger the parameter K, the smaller the storage capacity and the larger the basins of fixed points. Details of the Gardner model are available in the literature. In the following, we will discuss the relationship among pattern recognition, storage capacity, and chaotic behavior in the model. Without a loss of generality, we treat the net with maximal storage. Our results show that the ability of recognition depends on K (or storage capacity α), the initial condition, and the number of synapses removed randomly. The net with $N=80$ is used to generate the bulk of our results since this choice gives reliable thermodynamic approximations [10]. We will also discuss below the size effects of the net.

Now we introduce chaos into this model by cutting randomly a number of synapses. Chaos appears if a neuron is connected to fewer than the maximum of $N-1$ neurons in a network of N neurons. The number of missing synapses, chosen randomly, is represented as a fraction c of the total $N-1$ synapses for any neuron. This way of introducing chaos could be significant since a network with partially connected neurons is similar to the human brain in which the neurons are not fully connected [11] and the loss of synaptic connections in the human brain may also occur because of brain damage [2]. In the Gardner model, the cutting off of synapses (CUTS) has been discussed earlier as a diluted approach [12], where the effect of the dilution on the retrieval of patterns corresponding to the fixed points was considered. However, the existing work concentrates exclusively on the concept of pattern recognition being persistent under dynamics without any analysis of its relation to chaos.

We first show that chaos is indeed introduced in this way. The usual way to identify chaos in a system is to calculate the Lyapunov exponents or autocorrelation function [13,14]. We have used both of these techniques to examine the presence of chaos. Here we present only the results of the calculation of autocorrelation function, which is defined as

*Electronic address: ali@hg.uleth.ca

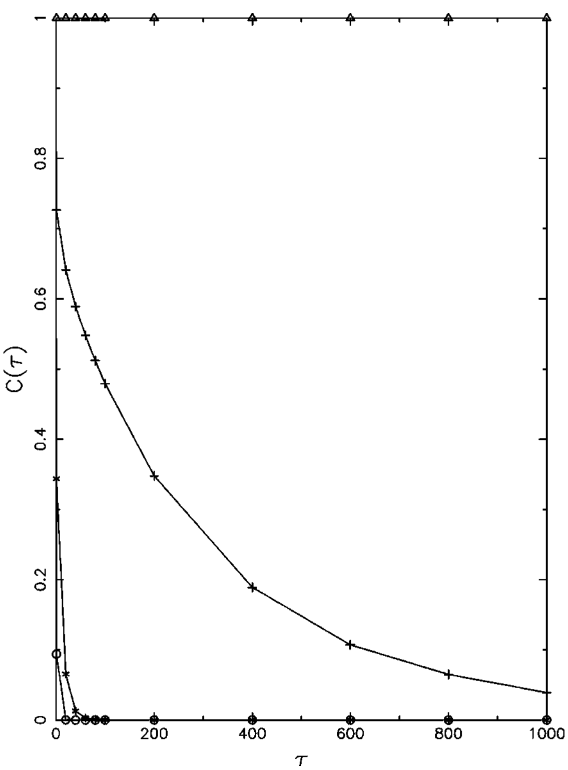

FIG. 1. Autocorrelation function of the Gardner model of $N=80$ at $K=2$ for $c=0$ (triangles), 0.5 (crosses), 0.6 (asterisks), and 0.8 (circles). Points are linked for a better perception.

$$C(\tau)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T} S_i(t)S_i(t+\tau), \qquad (2)$$

where $S_i(t)$ is the state of neuron i at time t. $C(\tau)$ is independent of the choice of i. The criterion of the autocorrelation function states that if $C(\tau)$ is a decreasing function of τ, the system is chaotic. The speed at which $C(\tau)$ decreases is related to the rate at which the systems turns chaotic. The dependence of $C(\tau)$ on τ is presented in Fig. 1 for $N=80$. The figure shows that when $c=0$, $C(\tau)$ remains a constant while for $c>0$, $C(\tau)$ decreases from a finite value to zero. For larger values of c, $C(\tau)$ goes to zero faster, implying a rapid growth of chaotic behavior. It should be pointed out that $C(\tau)$ is also dependent on K. For smaller K (not shown), even a small c will produce chaos.

When chaos is introduced, the network dynamics will not drive an initial state to a fixed point (pattern) as in the nonchaotic model without CUTS. The condition (1) is violated and patterns ξ^μ are no longer the fixed points. In fact, a chaotic system has a large number of unstable period orbits [15,16]. Although the dynamics does not drive an initial state to a fixed point, patterns may still be recognized. The synapses are trained before the CUTS and therefore we believe that some information about the patterns may still be present in the remaining J_{ij}. Before a fully chaotic behavior sets in, it may be possible to use the network for pattern identification. We are pointing here to short-term memories. The problem now is to extract this information, if there is any. To this end, we turn to synchronization of identical chaotic systems. The approach to synchronization of chaotic trajectories is not unique [15,16]. Here we utilize the feedback method proposed in [15] and select the trajectory stimulated initially by

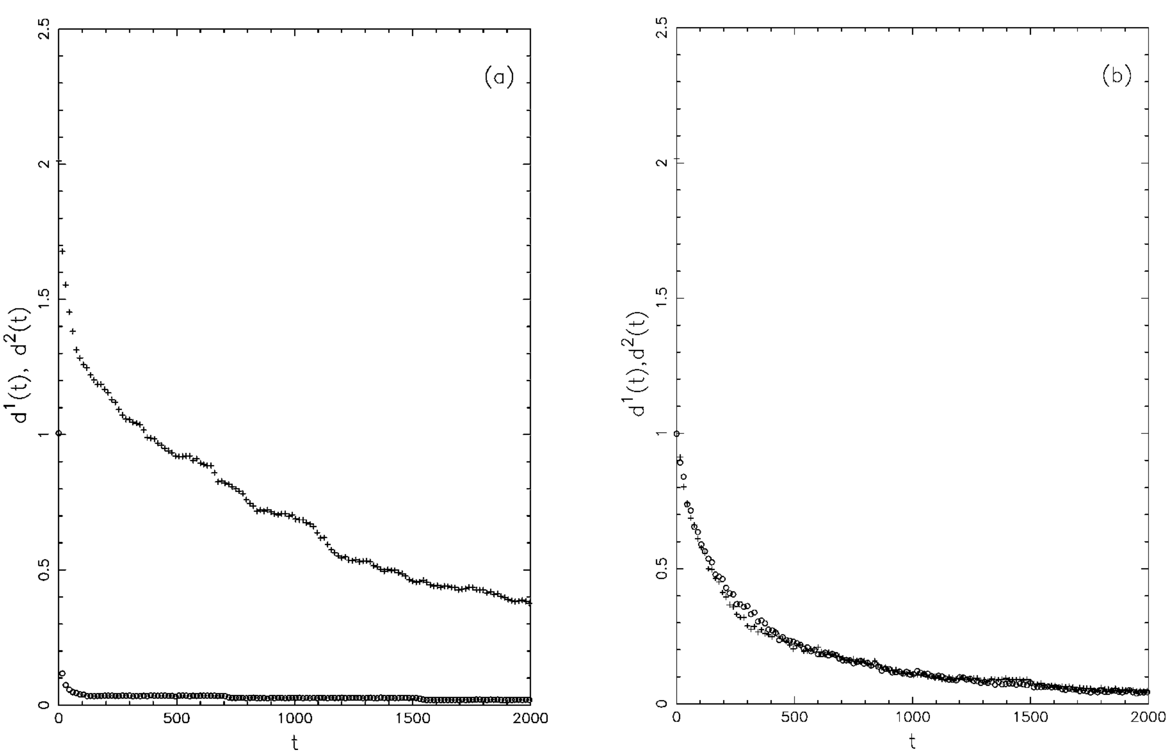

FIG. 2. Dynamic distances $d^1(t)$ (circles) and $d^2(t)$ (crosses) vs time t in the fully connected Gardner model ($c=0$) for $d^1(0)=1$: (a) $K=2$ and $k=0.08$ and (b) $K=0.68$ and $k=0.3$. The pattern can [cannot] be recognized in (a) [(b)].

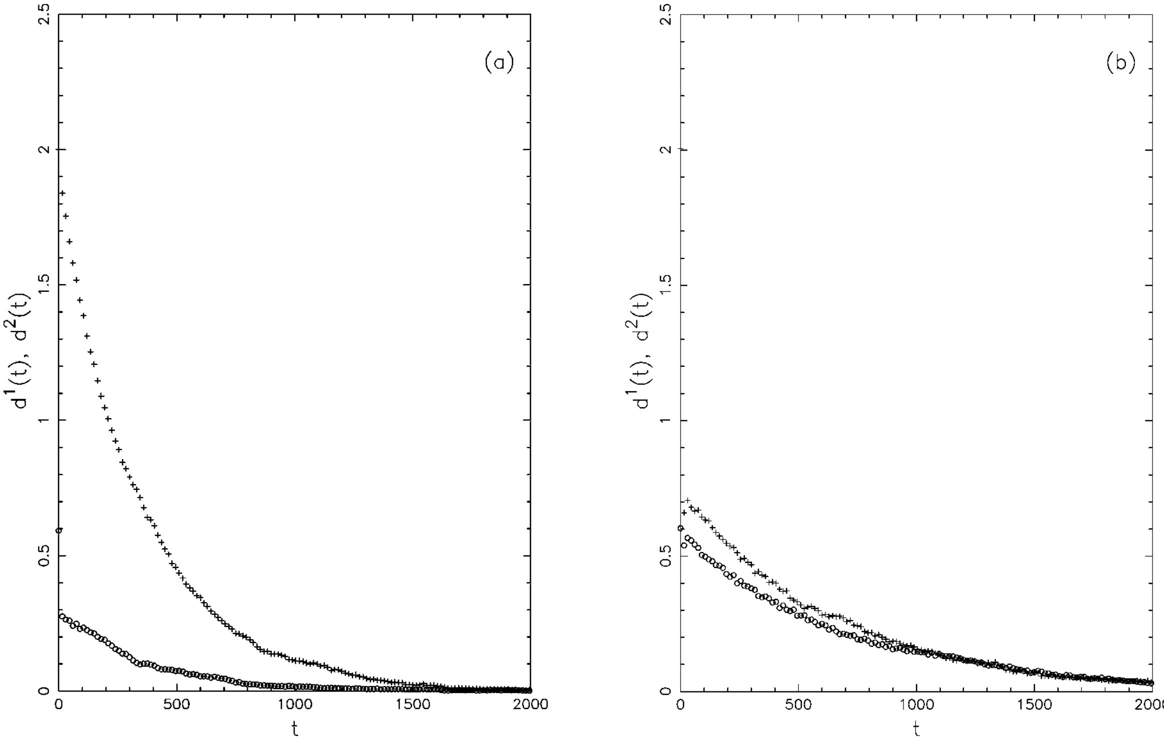

FIG. 3. Dynamic distances $d^1(t)$ (circles) and $d^2(t)$ (crosses) vs time t for $K=2$ and $d^1(0)=0.6$: (a) $c=0.5$ and $k=0.025$ and (b) $c=0.6$ and $k=0.32$. The pattern can [cannot] be recognized in (a) [(b)].

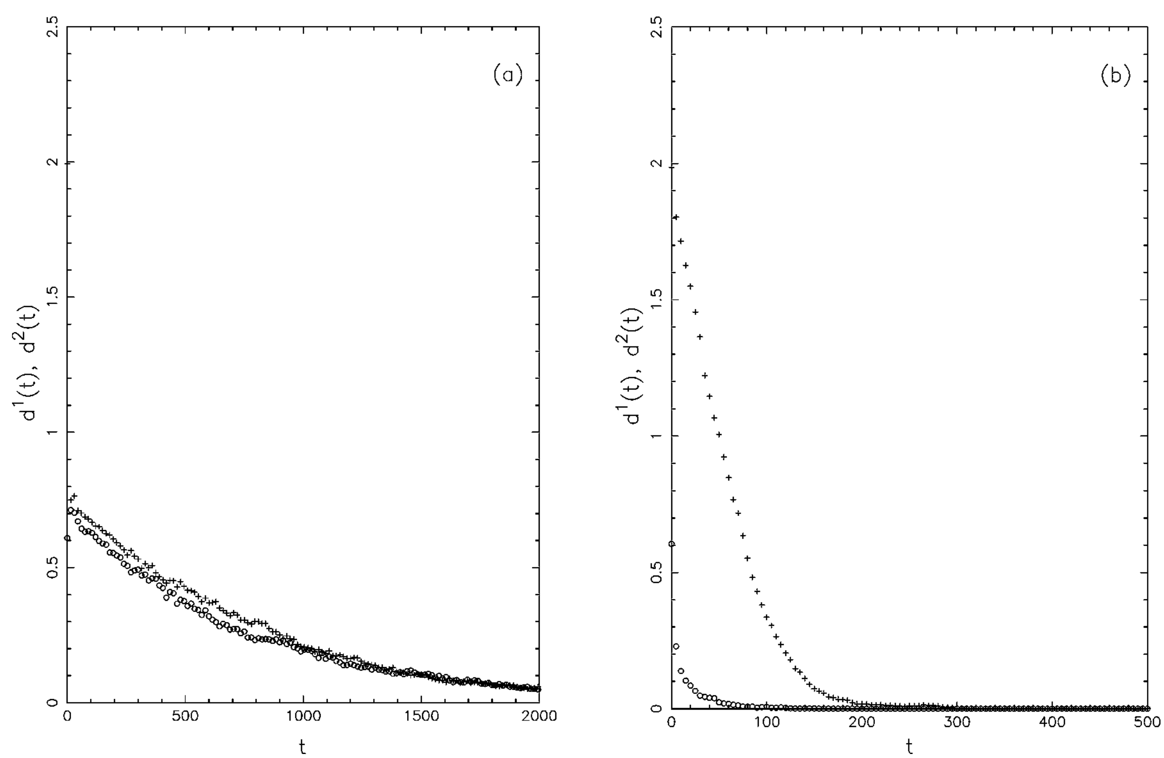

FIG. 4. Dynamic distances $d^1(t)$ (circles) and $d^2(t)$ (crosses) vs time t for (a) $K=1.5$, $c=0.5$, $k=0.4$, and $d^1(0)=0.6$ and (b) $K=2$, $c=0.49$, $k=0.045$, $d^1(0)=0.6$, and $N=200$. The pattern can [cannot] be recognized in (b) [(a)].

TABLE I. Critical values of c vs K for $d^1(0)=0.6$ and $N=80$. The numbers in parentheses are the values of α.

K	c
3.00 (0.1)	0.76
2.50 (0.14)	0.66
2.00 (0.20)	0.50
1.50 (0.31)	0.32

a pattern (e.g., ξ^1) as the desired orbit and synchronize other chaotic trajectories to this orbit. We try to get useful information during the synchronization.

We denote the desired trajectory as $\zeta^1(t)$ with $\zeta^1(0)=\xi^1$ and a trajectory to be synchronized as $s(t)$, which is near ξ^1 at $t=0$. The relation between $\zeta^1(t)$ and $s(t)$ is measured by a dynamic distance

$$d^1(t)=\left\langle \frac{1}{N}\sum_{i=1}^{N}[s_i(t)-\zeta_i^1(t)]^2\right\rangle_\xi, \tag{3}$$

where $\langle\rangle_\xi$ means the average over random patterns ξ in the Gardner model [6–9]. $d^1(0)$ gives the initial distance. Although $\zeta^1(t)$ is not ξ^1, it is generated by ξ^1 and in a deterministic chaotic system it is unique. Therefore, we "transfer" ξ^1 to $\zeta^1(t)$. If $d^1(t)$ goes to zero (or a small value), ξ^1 is retrieved from $s(0)$ containing only partial information of ξ^1. In this chaotic model, the dynamics will drive $s(t)$ away from $\zeta^1(t)$. $d^1(t>0)$ is larger than $d^1(0)$. However, if at every dynamic step we "help" $s(t)$ to go a little bit closer to $\zeta^1(t)$, we may get useful results. This is just the idea of chaos control by feedback. We now turn to chaos control and try to perform recognition in the chaotic system. The feedback is imposed by modifying the trajectory $s(t)$ at time t with

$$s_i(t)\rightarrow s_i^f(t)=s_i(t)+k[\zeta_i^1(t)-s_i(t)],\quad i=1,\ldots,N, \tag{4}$$

where $0<k<1$ is a constant called hereafter the feedback strength. For $k=0$, there is no feedback; for $k=1$, the trajectory is modified immediately to be the desired one. The state is $s^f(t)$ after the feedback process is utilized for the next dynamics. Because the feedback is applied directly and at every time step to the output of the system, $s^f(t)$ will be driven to $\zeta^1(t)$ after a sufficiently long time depending on k. After the synchronization, the feedback term goes to zero and the distance in Eq. (3) vanishes. However, we cannot believe that $d^1(t)=0$ in this case means that $s(t)$ is recognized because $s(t)$ may be synchronized in the same way with all other trajectories $\zeta^2,\ldots,\zeta^p$ coming initially from other patterns, $\xi^2,\ldots,\xi^p$, respectively. Therefore, we need another criterion to distinguish between the synchronized orbits.

The criterion that we have used in this work is based on the "time of synchronization." The phrase time of synchronization (TOS) is used here in a limited sense. The TOS is the time required for the distance between the drive and response systems to become less than a preset value d_0. According to this criterion, the TOS tells us which (target) pattern is retrieved. More precisely, we denote $d^\mu(t)$ as the distance by replacing index 1 in Eq. (3) with μ. The shortest time

$$t^r=\min\{t^1,\ldots,t^p\}, \tag{5}$$

where t^μ satisfy

$$d^\mu(t^\mu)<d_0,\quad \mu=1,\ldots,p, \tag{6}$$

determines that pattern r is the target pattern. Here d_0 is a small number. If the shortest time t^r does not exist (the t^μ, $\mu=1,\ldots,p$, are close to one another before d_0 is reached), then the recognition fails. In practice we do not have to compare all distances because after the average in Eq. (3) all patterns are equivalent. Below we set $s(0)$ to be related to ξ^1 with the initial distance $d^1(0)$ and compare the distance $d^1(t)$ between $s(t)$ and $\zeta^1(t)$ and the distance $d^2(t)$ between $s(t)$ and an arbitrary trajectory; see $\zeta^2(t)$ starting from ξ^2. Since the patterns are randomly orthogonal, $d^2(0)$ is 2. In the following, we choose $d_0=0.02$.

The approach has been first checked for the fully connected Gardner model ($c=0$). Figures 2(a) and 2(b) give the distances $d^1(t)$ (circles) and $d^2(t)$ (crosses) versus t with initial distance $d^1(0)=1$ for $K=2$ and 0.68, respectively. In Fig. 2(a) $d^1(t)$ is always smaller than $d^2(t)$ and this leads to the result that t^1 is smaller than t^2, while in Fig. 2(b) $d^1(t)$ and $d^2(t)$ have the same value after certain steps. Therefore, according to our discussion, the network is able to retrieve pattern ξ^1 in Fig. 2(a) but not in Fig. 2(b). This is in agreement with the well-known results obtained by means of fixed points [7–9].

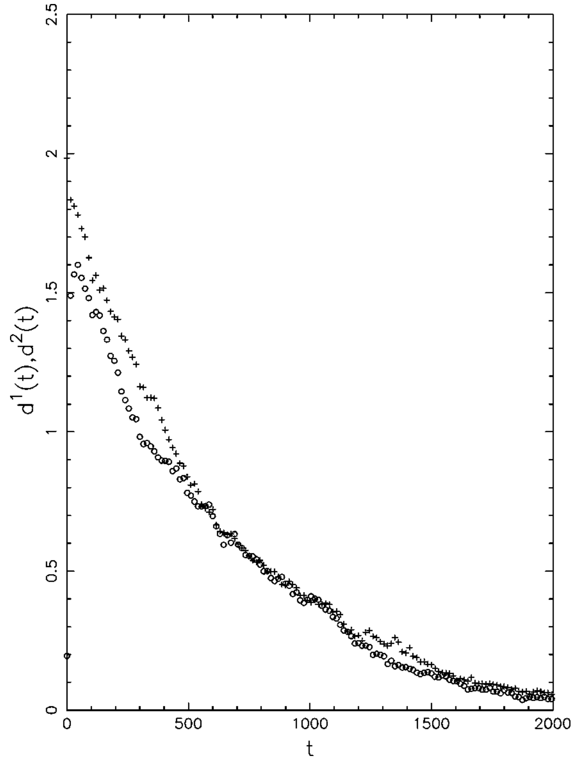

FIG. 5. Dynamic distances $d^1(t)$ (circles) and $d^2(t)$ (crosses) started from two random states for $K=2$, $d^1(0)=0.2$, $c=0.5$, and $k=0.023$. The distances are indistinguishable after certain steps.

Now we turn to the chaotic model. Figure 3 gives the dynamic distance $d^1(t)$ (circles) and $d^2(t)$ (crosses) versus t for $K=2$, $d^1(0)=0.6$, and different fractions of CUTS and feedback strengths. In Fig. 3(a) with $c=0.5$, $d^1(t)$ goes to zero much faster than $d^2(t)$ and this implies that t^1 is smaller than t^2 when the condition (6) is fulfilled. When $c=0.6$ in Fig. 3(b), we cannot find any difference between $d^1(t)$ and $d^2(t)$ after certain steps. This will not lead to different t^1 and t^2 when the condition (6) is satisfied. Therefore, the pattern is recognized in Fig. 3(a) but not in Fig. 3(b).

It is natural that when c is small, the structure of the fully connected model will not be affected dramatically. Specially, when the fixed points have larger basins (e.g., $K=2$), i.e., smaller storage α, the strong convergent tendency persists. When c is large, things change noticeably. When the basin gets smaller (storage gets larger), even small CUTS may produce chaos.

Figure 4(a) shows the effect of the basin (storage capacity). Compared with Fig. 3(a) for $K=2(\alpha=0.2)$, Fig. 4(a) for $K=1.5$ ($\alpha=0.31$) does not give distinguishable t^1 and t^2 when the condition (6) is valid. The recognition thus fails in Fig. 4(a). In order to check if there is any size effect, we treated a net with $N=200$ and $K=2$. The results are given in Fig. 4(b). Comparing Fig. 3(a) ($N=80$ and $K=2$) and Fig. 4(b) ($N=200$ and $K=2$), we believe that, for pattern recognition, our results represent large N (thermodynamic) behaviors. We have also checked the nets with $N=40$, 80, and 160 and have found that the results for different sizes are qualitatively the same, except for some quantitative differences. Our results of the effects of different sizes are in agreement with the discussions given in [10].

In short, the ability of recognition depends on the parameters K (which determines the basins of the fully connected nonchaotic model), c (which defines the fraction of CUTS), and $d^1(0)$ (which is the initial distance). Table I gives the critical values of c as a function of K for $d^1(0)=0.6$ and $N=80$. The numbers in parentheses are the corresponding values of α. If the cut ratios c are below the values listed in the table, the recognition is successful; otherwise patterns cannot be recognized. It may be pointed out that the qualitative property is not sensitive to the feedback strength k. k determines the time required for satisfying the condition (6) and does not affect the order in magnitudes among different t^μ.

Using chaos control and the criterion of time of synchronization, we have shown that the information hidden in the history of evolution can be exploited. When chaos exists, the process of pattern recognition can be performed in this way. We emphasize that this conclusion is related not only to the initial condition, but also to the structure of the network that has embedded the patterns into synapses J_{ij}. This can be verified by starting ζ^1 and ζ^2 by two random states instead of ξ^1 and ξ^2. Figure 5 gives the distance $d^1(t)$ between $\zeta^1(t)$ and $s(t)$ and the distance $d^2(t)$ between $\zeta^2(t)$ and $s(t)$. Even for very small initial distance $[d^1(t)=0.2]$ and strong convergent tendency ($K=2$), $d^1(t)$ and $d^2(t)$ quickly become similar and we cannot treat them the way we did in Fig. 3(a). This is not surprising because the initial random states are not stored in the J_{ij}. The network ''does not know'' the random states. The approach of pattern recognition in chaotic networks proposed in this paper can be easily applied to other chaotic systems. It may be interesting to study the effects of gradual and sudden cutoffs (or regeneration) of synapses on such issues as metastable states [17].

The work has been supported through a grant to M. K. A. from the Defence Research Establishment Suffield under the Contract No. W7702-6-R607/001/EDM.

[1] C. A. Skarda and W. J. Freeman, Behav. Brain Sci. **10**, 161 (1987).
[2] J. Glanz, Science **277**, 1758 (1997).
[3] I. Tsuda, Neural Networks **5**, 313 (1992).
[4] M. Adachi and K. Aihara, Neural Networks **10**, 83 (1997); T. Nagashima, J. Miyazaki, Y. Shiroki, and I. Tokuda, Int. J. Chaos Theory Appl. **2**, 1 (1997); S. Nara, P. Davis, M. Kawachi, and H. Totsuji, Int. J. Bifurcation Chaos Appl. Sci. Eng. **5**, 1205 (1995).
[5] M. Kushibe, Y. Liu, and J. Ohtsubo, Phys. Rev. E **53**, 4502 (1995).
[6] E. Gardner, J. Phys. A **21**, 257 (1988).
[7] E. Gardner, J. Phys. A **22**, 1969 (1989).
[8] T. B. Kepler and L. F. Abbott, J. Phys. (Paris) **49**, 1657 (1988).
[9] Z. Tan and L. Schülke, Int. J. Mod. Phys. B **26**, 3549 (1996).
[10] W. Krauth and M. Mezard, J. Phys. A **20**, L745 (1987).
[11] J. W. Clark, Phys. Rep. **158**, 91 (1988).
[12] M. Bouten, A. Engel, A. Komoda, and R. Serneels, J. Phys. A **23**, 4643 (1990).
[13] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer-Verlag, Berlin, 1983).
[14] P. W. Milonni, M.-L. Shih, and J. R. Ackerhalt, *Chaos in Laser-Matter Interaction* (World Scientific, Singapore, 1987).
[15] K. Pyragas, Phys. Lett. A **170**, 421 (1992).
[16] M. K. Ali and J.-Q. Fang, Phys. Rev. E **55**, 5285 (1997).
[17] Z. Tan and M. K. Ali, Phys. Rev. E **57** R3739 (1998).