Journal of Statistical Physics, Vol. 58, Nos. 5/6, 1990

A Study of Network Dynamics

Steve Renals[1] **and Richard Rohwer**[1]

Received June 21, 1989; revision received September 12, 1989

The dynamics of a discrete-time neural network model are investigated. First, a numerical survey of network power spectra is reported for networks of varying size with random weight matrices and initial states. The steepness of the logistic function and a symmetry measure of the weight matrix are taken as control parameters. Summary statistics are presented to give gross measures of the model's temporal activity in parameter space. Second, a detailed study of the dynamics of a particular network is described. Complex dynamical behavior is observed, including Hopf bifurcations, the Ruelle–Takens–Newhouse route to chaos (showing mode-locking at rational winding numbers and the destruction of an invariant torus), and the period-doubling route to chaos.

KEY WORDS: Neural networks; dynamics; chaos; routes to chaos; nonlinear systems.

1. INTRODUCTION

Research into neural network models has generally been directed toward models that perform tasks such as associative memory and pattern classification through the use of fixed point attractors.[1,2] Although many powerful algorithms for customizing the fixed points of stable networks have been developed and the current effort in understanding and exploiting computation with fixed points is exciting, it is apparent that computation with non-fixed-point attractors has great potential with regard to temporal applications such as associative memory, speech recognition, and motor control. Complex dynamical behavior in the brain has recently been observed and discussed[3–5] and it has been suggested that chaotic dynamics[6] may provide a positive "don't know" state in associative memory.[5,7,8]

Recently, several researchers have reported chaotic dynamics in neural

[1] Department of Physics and Centre for Speech Technology Research, University of Edinburgh, Edinburgh EH1 1HN, United Kingdom.

0022-4715/90/0300-0825$06.00/0

822/58/5-6-3

network models from theoretical considerations,[9–12] from computer simulation,[13] and from electronic implementation.[14] Sompolinsky *et al.*[11] studied a continuous-time system with random weight matrices and demonstrated that as $N \to \infty$ the dynamics will be chaotic, provided there is no zero fixed point. Kürten and Clark[13] studied a continuous-time system numerically, with random weight matrices and limited fan-in to each node. They observed chaotic behavior in certain systems with greater than 25 nodes. Marcus and Westervelt[14] studied analog electronic networks with time delay: they observed period doubling to chaos in an asymmetric three-node network, using the time delay as a control parameter. Early work in neural network modeling was concerned with oscillation and system timing: e.g., Sloane[15] investigated a class of neural network models with binary neurons that executed limit cycles with lengths increasing exponentially with the size of the network.

Here a discrete-time neural network model is investigated, consisting of N fully connected nodes, with the dynamical law

$$\begin{aligned} y_i(t+1) &= f\left(\sum_{j=1}^{N} r w_{ij}\, y_j(t)\right), \qquad i = 1, 2, \ldots, N \\ f(x) &= \frac{1}{1+e^{-x}} \end{aligned} \tag{1}$$

where $y_i(t)$ is the output of node i at time t and w_{ij} is an $N \times N$ weight matrix; f is the node transfer function, with nonlinearity r. An initial random weight matrix w'_{ij} was randomly generated, so as to provide an RMS activation of 1.0 for each node's input by initializing each element to lie between $\mp 3/\sqrt{N}$. This matrix was then decomposed into symmetric S_{ij} and antisymmetric A_{ij} components

$$S_{ij} = w'_{ij} + w'_{ji} \tag{2a}$$

$$A_{ij} = w'_{ij} - w'_{ji} \tag{2b}$$

and the weight matrix used in (1) was generated using a symmetry parameter ε:

$$\begin{aligned} w_{ij} &= \sigma S_{ij} + \alpha A_{ij} \\ \alpha + \sigma &= 1, \qquad \varepsilon = \alpha/\sigma \end{aligned} \tag{3}$$

Two control parameters were used: r, which describes the nonlinearity

(slope of the transfer function f), and a weight matrix symmetry parameter a, which may be obtained from ε as

$$a = \frac{1 - \varepsilon^2}{1 + \varepsilon^2} \tag{4}$$

Hence $a = -1.0$ corresponds to an antisymmetric weight matrix, $a = 1.0$ corresponds to a symmetric weight matrix, and $a = 0.0$ corresponds to a random (uncorrelated) weight matrix.

Two investigations have been performed. First an attempt was made to start cataloging the richness of dynamics offered by this neural network model by varying the parameters a and r with networks with 2–25 nodes. This survey was performed by computing power spectra for networks with randomly selected weight matrices and initial states. From the power spectra that were collected a set of summary statistics was computed to give gross measures of the temporal activity of this class of neural network model. The second investigation was a detailed study and characterization of the dynamics of a particular network.

2. A SURVEY OF NETWORK POWER SPECTRA

In an attempt to catalog systematically the dynamics accessible to this neural network model, power spectra of nodes in representative networks (2–25 nodes) were numerically computed. For each network of size N nodes, 10 randomly chosen initial weight matrices were chosen. The parameters a and r were varied, and each resultant system was run with 10 random initial states.

2.1. Computer Simulations

For a given network, the power spectrum of each node i was numerically computed (after 75 iterations to run out transients[2]) using a 1024-point fast Fourier transform (FFT) ($T = 1024$), giving a maximum detectable period of 512,

$$P_i(k) = \left| \frac{1}{T} \sum_{t=0}^{T-1} y_i(t) \exp\left(\frac{-2\pi i k t}{T} \right) \right|^2, \qquad k = 0, 1, \ldots, \frac{T}{2} \tag{5}$$

[2] Transients lasting for considerably more than 75 iterations may be observed in certain networks (see Section 3).

The peaks of the FFT were interpreted as activity at specific periods $T(K)$ or frequencies $\omega(k)$,

$$\omega(k) = \frac{2\pi k}{T}, \qquad k = 0, 1, ..., \frac{T}{2} \tag{6a}$$

$$T(k) = \frac{2\pi}{\omega(k)} = \frac{T}{k}, \qquad k = 0, 1, ..., \frac{T}{2} \tag{6b}$$

The raw FFT data were then summarized into two numbers per node:

1. μ_i^P, mean (over frequency) power

$$\mu_i^P = \frac{2}{T} \sum_{k=1}^{T/2} P_i(k) \tag{7}$$

2. S_i, "entropy" measure of power (over frequency)

$$S_i = -\sum_{k=1}^{T/2} P_i'(k) \log P_i'(k) \tag{8}$$

where $P'(k)$ is the normalized power spectrum.

These statistics were then averaged over all the nodes in the network to give two mean values $\bar{\mu}^P$ and $\bar{S}$, summarizing the dynamics of a particular network. The mean power measure gives an indication of the amplitudes of oscillations present in the network. This will clearly be large valued in situations in which the network is oscillating between saturated parts of the nodal transfer function, i.e., at large values of r. The entropy measure gives an indication of how broad the power spectrum is: a spectrum with a single peak will have $\bar{S} = 0$, whereas larger values of $\bar{S}$ indicate that the spectrum has peaks at many different frequencies, indicating a complex periodic oscillation (with many harmonics) or a chaotic oscillation (which is characterized by a broadband power spectrum).

One hundred simulations were performed for each value of the (r, a) parameter pair, obtained by taking 10 random weight matrices and using 10 random initial states for each weight matrix. The means and standard deviations of $\bar{\mu}^P$ and $\bar{S}$ were computed over these 100 simulations, giving four numbers to characterize the dynamics of an N-node network at a given (r, a) value. These simulations were performed using networks of various sizes ($2 \leqslant N \leqslant 25$) and the results are shown graphically for a network of size $N = 8$ (Fig. 1). Figure 2 shows just the entropy measures for a network of size $N = 15$.

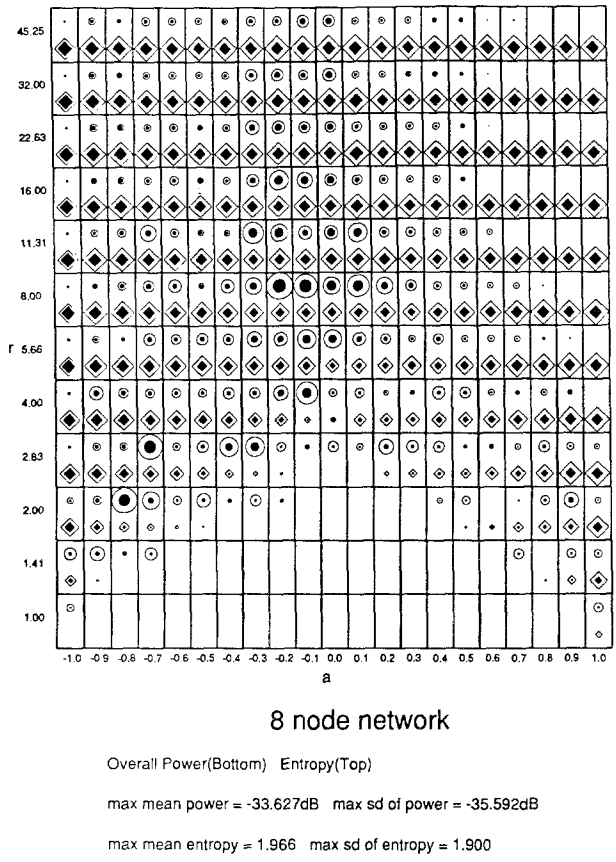


Fig. 1. Summary statistics from power spectra taken for networks with $N = 8$. Each box represents 100 simulations at a given (r, a) value: 10 initial random weight matrices each run with 10 random initial states. The upper circles in each box represent the entropy $\bar{S}$ and the lower diamonds represent the overall power $\bar{\mu}^P$. The filled-in shapes represent means, the surrounding outlines standard deviations (statistics over simulations).

2.2. Discussion

The (r, a) phase space of neural network temporal activity may be qualitatively divided into regions. At low values of r, the transfer function is virtually linear and the weight matrix will be of low magnitude, hence the eigenvalues will usually fall inside the unit circle and the network will display fixed-point behavior. When r is large, the transfer function is close to a step function; when the network is symmetric $(a = 1.0)$ this is the Hopfield case with parallel update, and it may be demonstrated

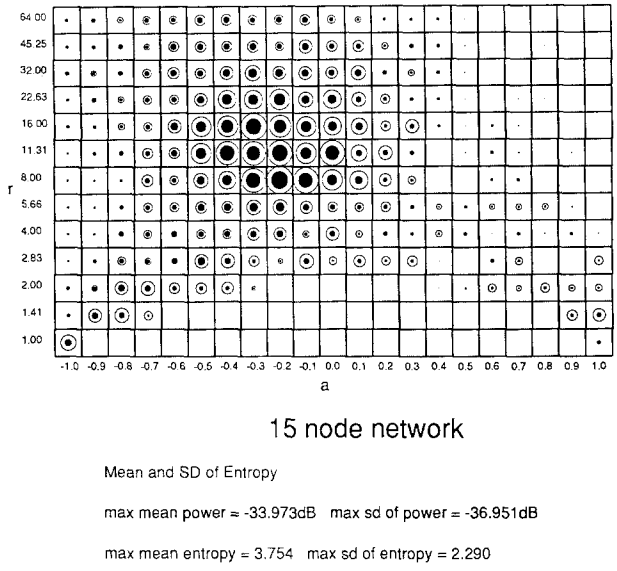


Fig. 2. Mean and standard deviations (over simulations) of average entropy $\bar{S}$ for networks with $N = 15$.

analytically that the system will display fixed-point or limit-cycle period-2 dynamics[16]; when the network is antisymmetric ($a = -1.0$), then it may be demonstrated that the system will exhibit limit-cycle period-4 dynamics.[17]

Away from these extremes, more complex dynamics with longer temporal correlations may be observed. In Fig. 2 these areas are indicated by the large values of the mean and standard deviation of the entropy. For networks of all sizes a region of higher activity may be observed, at approximately $4 \leqslant r \leqslant 32$ (i.e., at values of r in which the transfer function is neither close to linear nor close to a step function) and at approximately $-0.8 \leqslant a \leqslant 0.0$, i.e., random networks or networks with a greater antisymmetric component. When long-term temporal activity first occurs in the network (with respect to varying r), large entropy values are observed with corresponding extremely low overall power values. This may be interpreted as a bifurcation away from a fixed point to a very small-amplitude (quasi-) periodic oscillation. This oscillation increases in power as either a becomes more negative or r increases. Additionally, an "island of activity" may be observed for positive values of a and low values of r. These oscillations are generally at low power, so a possible explanation is that they are an artifact caused by numerical noise arising from the accuracy of the floating point

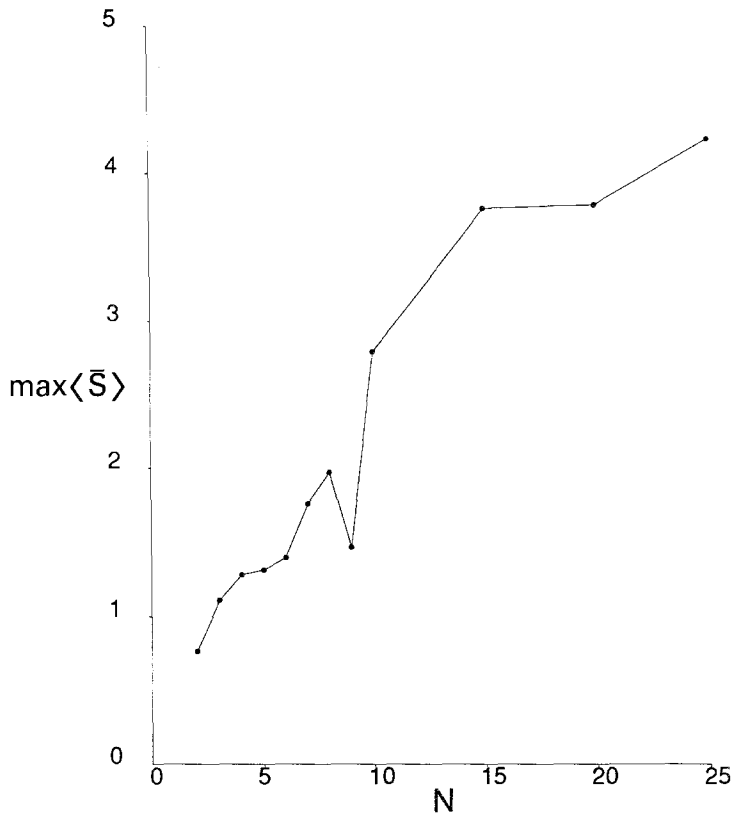


Fig. 3. Plot of maximum (over a and r) mean (over simulations) entropy ($\max\langle\bar{S}\rangle$) with respect to N. The maximum possible value for $\max\langle\bar{S}\rangle$ is $\log(t/2) = \log(512) = 6.23$.

processor.[3] However, some of the activity occurs at appreciable power levels, so it appears likely that there is some genuine long-time correlated activity occurring in these islands of activity.

The general portrait of network dynamics with respect to (r, a) appears to be qualitatively independent of N, the number of nodes in the network ($N > 3$). However, the dynamics becomes more complex as N increases: this can be seen by observing the increase of the maximum (over a and r) mean (over simulations) entropy, $\max\langle\bar{S}\rangle$, with increasing N (Fig. 3). As would be expected, very small networks ($N = 2, 3$) do not display complex dynamical behavior in any region of parameter space.

[3] The simulations in this section were performed using 32-bit floating-point arithmetic. In the next section the simulations were performed using 64-bit floating-point arithmetic.

3. DYNAMICS OF A SPECIFIC NETWORK

The survey of network power spectra gave some general indications of the kind of dynamics accessible to the neural network model, with respect to the number of nodes in the network, the symmetry of the weight matrix, and the nonlinearity of the transfer function. However, this gives no specific information about the dynamics exhibited by individual N-node networks with respect to r and a. To illustrate the dynamics that these systems exhibit, we numerically simulated an eight-node network with a particular random weight matrix and initial state. First the effect of varying r was investigated while a was held constant, then r was held constant and a varied.

3.1. Random Weight Matrix ($a = 0$)

When $a = 0$ the weight matrix had equal symmetric and antisymmetric parts, so there was no correlation between the elements. The network under investigation ($N = 8$) was first investigated by varying r at this value of a.

The dynamics may be viewed by constructing bifurcation diagrams for

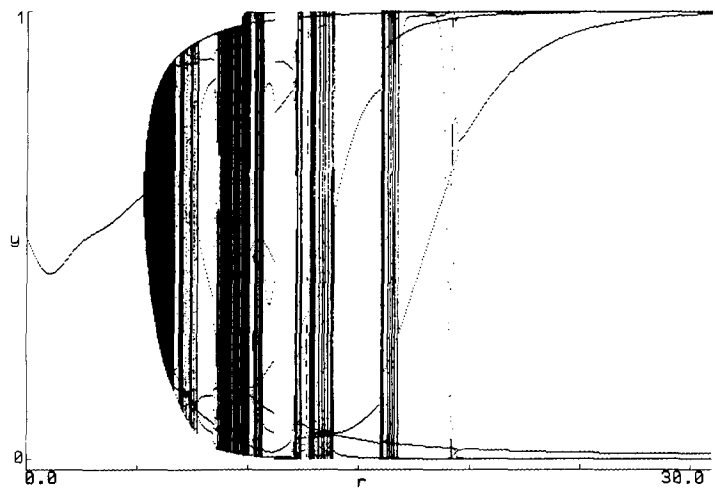


Fig. 4. Bifurcation diagram for one node of an eight-node network ($a = 0.0$) with respect to varying r ($0.0 \leqslant r < 31.0$). One column of the diagram represents 10,000 iterations of the network (after 10,000 iterations to run out transients) at a particular value of r. The y axis represents the output of a particular node of the network.

each node showing the output of that node versus r, as r increases in small steps. Bifurcation diagrams were constructed by running the system for 20,000 iterations and recording the last 10,000 states of the network (which were presumed to be on the attractor) and plotting these on a bifurcation diagram.

Figure 4 shows a bifurcation diagram for one node of the network for $0.0 \leqslant r < 31.0$, with r increasing in steps of 0.05. With small values of r we have a close to linear regime with low weight values and hence the dynamics evolves to a fixed point. At $r = 5.282$ a Hopf bifurcation occurs with the system moving from fixed-point to oscillatory behavior. As expected, the regime of r values just prior to this bifurcation is marked by very long-lived transients (up to 10,000 iterations). The network continues to display complex behavior until $r > 20$, when the attractor becomes a limit cycle of period 4, with activity on all nodes moving to the saturated regions of the transfer function (which becomes close to a step function).

Figure 5 shows the region $5.0 < r < 8.1$ in greater detail; these dynamics may also be examined via the power spectrum of one node of the network (Fig. 6). The system bifurcates from a fixed point to one frequency oscillatory behavior at $r = 5.2822$, indicated by a single peak in the power spectrum; a second bifurcation occurs at $r = 5.2840$ when a second incom-

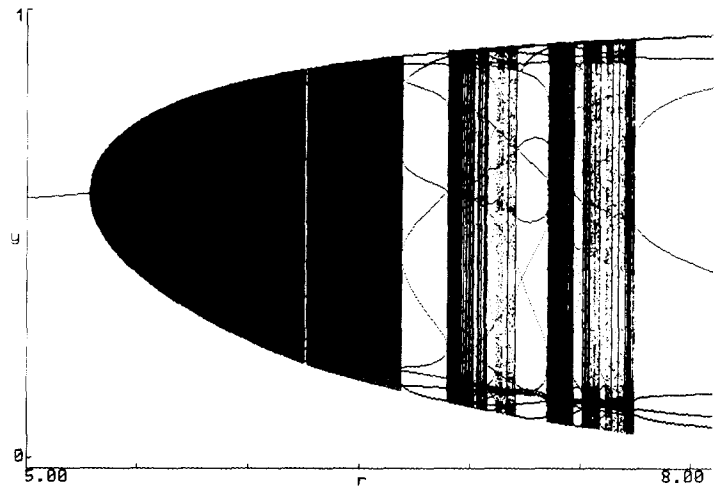


Fig. 5. Bifurcation diagram for one node of an eight-node network ($a = 0.0$) with respect to varying r ($5.0 \leqslant r < 8.1$). The regions of mode locking at winding numbers of 3/11, 5/18, and 2/7 may be observed.

mensurate frequency emerges (Fig. 6). However, if a section of the attractor is plotted, a one-dimensional circle map results (Fig. 7a) with a correspondingly linear "θ-map" (Fig. 7b).

The orbit in the plane is restricted to a one-dimensional circle, which may be parametrized in terms of the angles θ_t of the points on the circle (mod 1 in a circle of 2π) with respect to the centroid of the points, plotting θ_{t+1} vs. θ_t (Fig. 7b). This is approximately a straight line, indicating that the system is rotating in phase space with a constant angular velocity.

The dynamics of this system may be characterized by the winding number

$$W = \lim_{t \to \infty} \left(\frac{\theta_t - \theta_0}{t} \right) \tag{9}$$

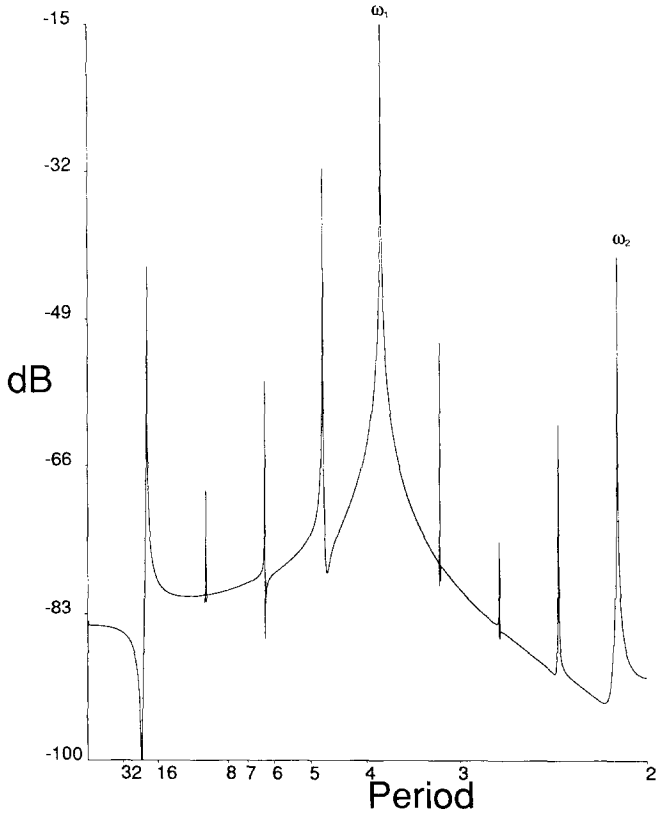


Fig. 6. Power spectrum of one node of an eight-node network ($a = 0.0$) at $r = 6.0$, showing two-frequency quasiperiodic motion, with several harmonics ($t_1 = 3.796$, $t_2 = 2.110$).

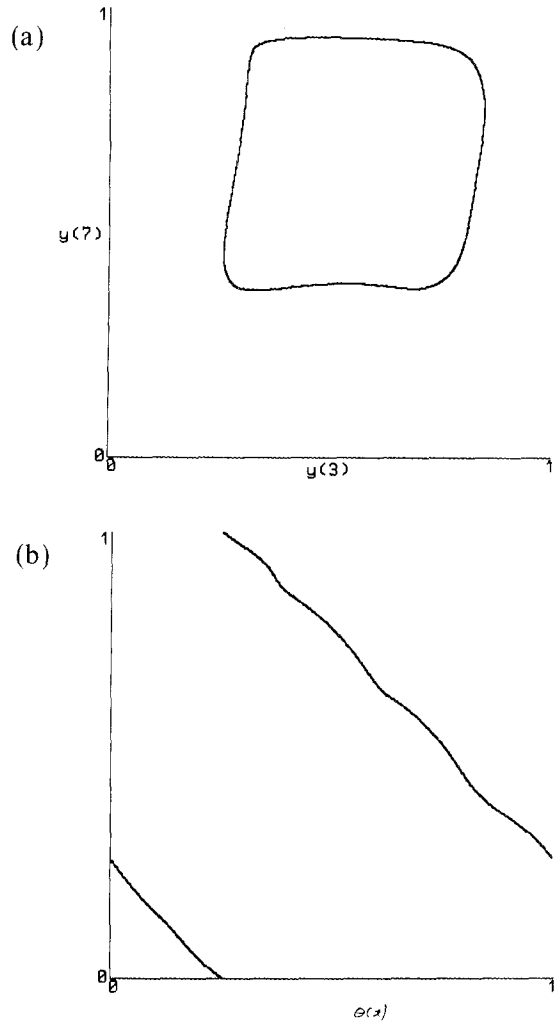


Fig. 7. (a) Attractor section y_3 vs. y_7 and (b) corresponding θ-map, when $r = 6.00$, $a = 0.0$. Although there are two incommensurate frequencies, the θ-map still shows one-dimensional (linear) characteristics.

The winding number measures the average shift of the angle θ per iteration; θ is not taken mod 1 in this expression. The value of the winding number at a given r value may also be obtained from the power spectrum: it corresponds to the dominant peak. The value of the winding number may be plotted against r (Fig. 8). The plot shows the characteristic "devil's staircase," a self-similar structure. Each step of the staircase corresponds to

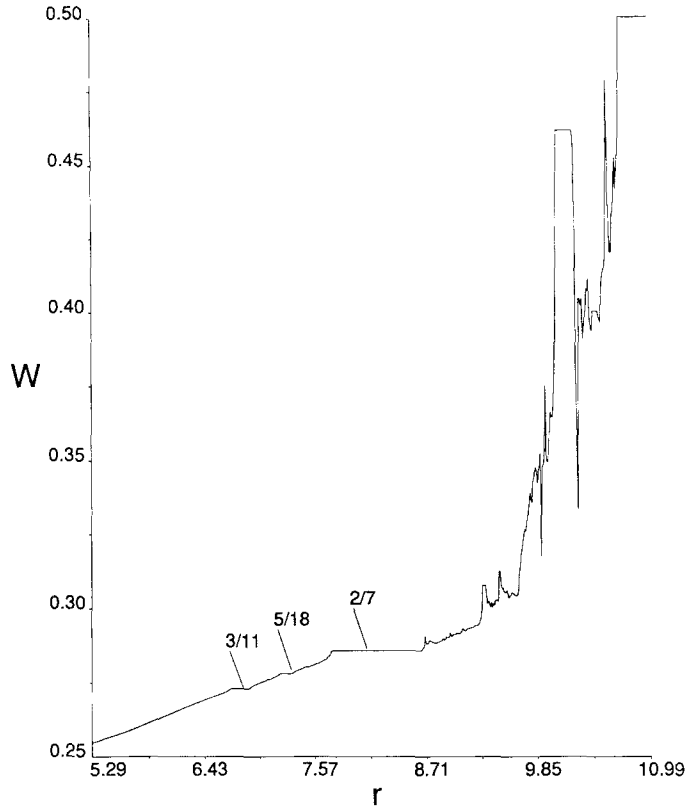


Fig. 8. Plot of winding number W against r at $a = 0.0$. The mode-locking phenomenon is apparent in the self-similar devil's staircase structure—note that 5/18 is the Farey daughter of 3/11 and 2/7. The breakup of the devil's staircase indicates a transition to chaos.

a rational winding number at which the system is mode locked; these rational numbers correspond to those generated by the Farey tree of rational numbers. These dynamics are similar to those of the circle map[18,19]:

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi\theta_n) \bmod 1 \tag{10}$$

When $r = 6.9$ another bifurcation occurs and a third frequency may be observed in the power spectrum (Fig. 9) along with numerous harmonics and linear combinations; some broadband noise is also visible in the power spectrum, indicating a transition to chaos. The attractor section (Fig. 10a) shows the circle starting to backfold and the corresponding θ-map

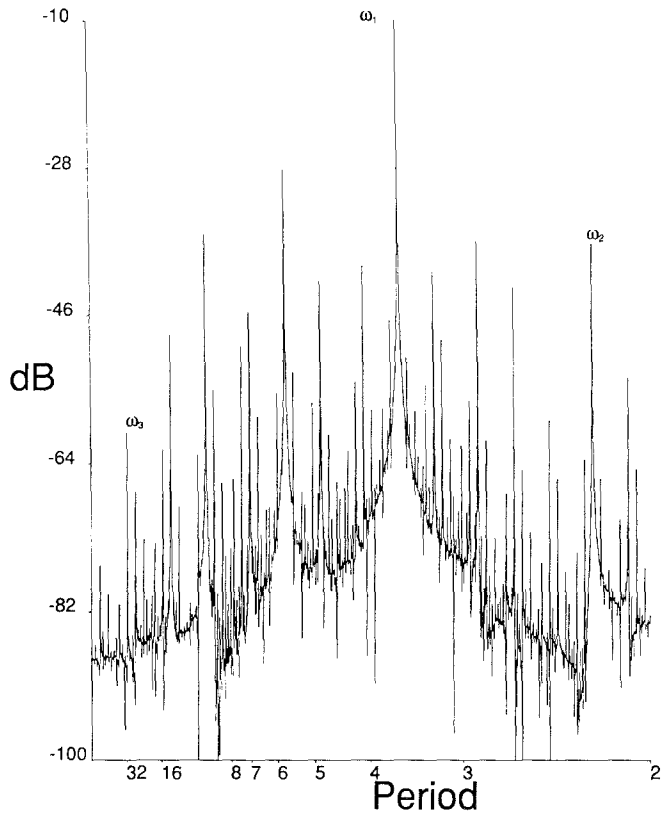
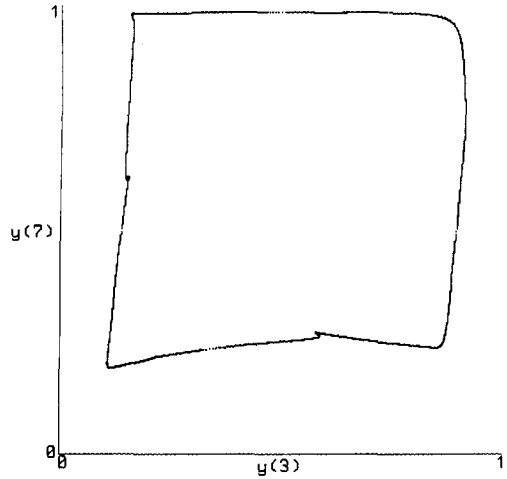


Fig. 9. Power spectrum of one node of the network at $r = 7.07$ and $a = 0.0$. A third incommensurate frequency has been introduced.

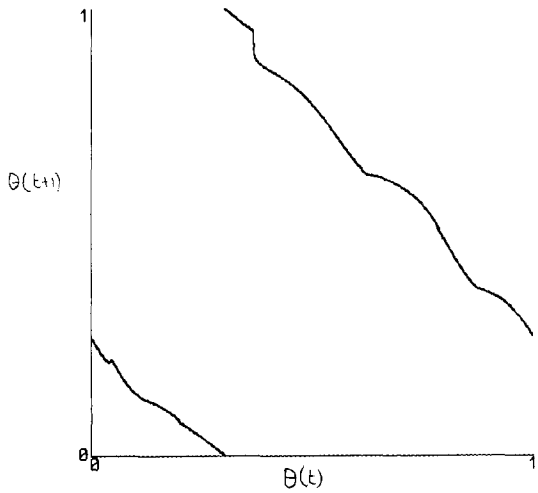
(Fig. 10b) shows a cubic point of inflection. This is characteristic of the Ruelle–Takens–Newhouse route to chaos.[20] As r increases, more backfolding occurs and the transition to chaos is complete, which may also be observed in the W vs. r plot (Fig. 8): the devil's staircase breaks up when the system becomes chaotic.

A section of the attractor may be plotted in two dimensions (Fig. 11): however, at least three dimensions are needed in order to illustrate the full structure and complexity of this attractor. At $r = 10.69$ the correlation dimension (D_2) was computed using the Grassberger–Procaccia[21] method with 32,000 points assumed to be on the attractor, having run the system for 10,000 iterations to discard transients. A value of $D_2 = 2.19 \pm 0.01$ was computed.

Figure 12 shows a bifurcation diagram in the range $11.0 \leqslant r \leqslant 14.1$. As



(a)



(b)

Fig. 10. (a) Attractor section y_3 vs. y_7 and (b) corresponding θ-map, when $r = 7.07$, $a = 0.0$. The introduction of a third frequency causes the circle to begin to break up (a) and a cubic point of inflection appears in the θ-map (b), marking the appearance of a transition to chaos.

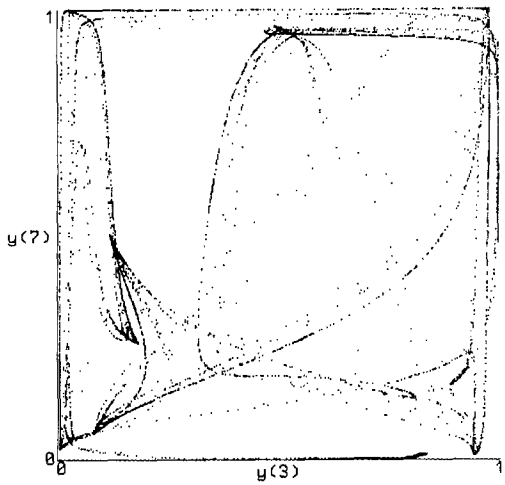


Fig. 11. Section of the chaotic attractor at $r = 10.69$, $a = 0.0$.

r is increased to around 12.0, the dynamics displays period doubling to chaos, which may be observed in the bifurcation diagram; the period doubling may also be tracked using power spectra, with the corresponding new low-frequency peak appearing at each period doubling. This is a classical route to chaos[22] characterized by Feigenbaum's constant δ,

$$r_n = r_\infty - \text{const} \cdot \delta^{-n}, \qquad n \gg 1 \tag{11}$$

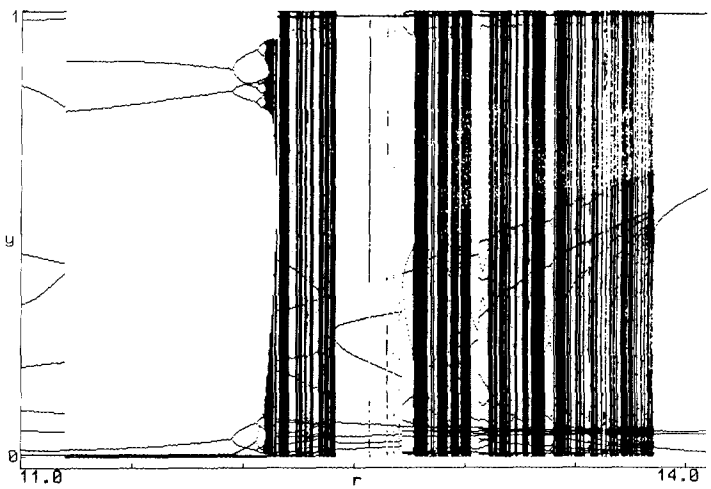


Fig. 12. Bifurcation diagram for one node of an eight-node network ($a = 0.0$) with respect to varying r ($11.0 \leqslant r < 14.1$). Period doubling to chaos may be observed around $r = 12$.

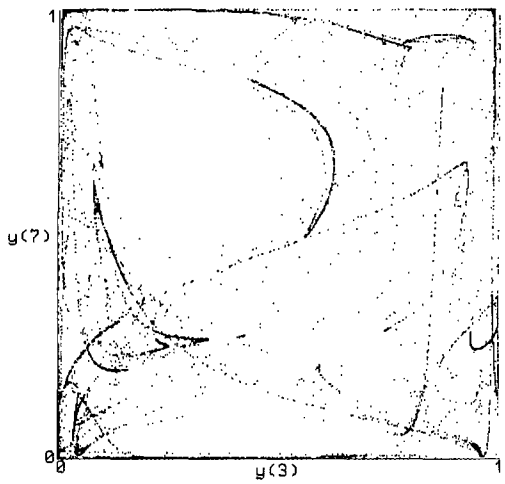
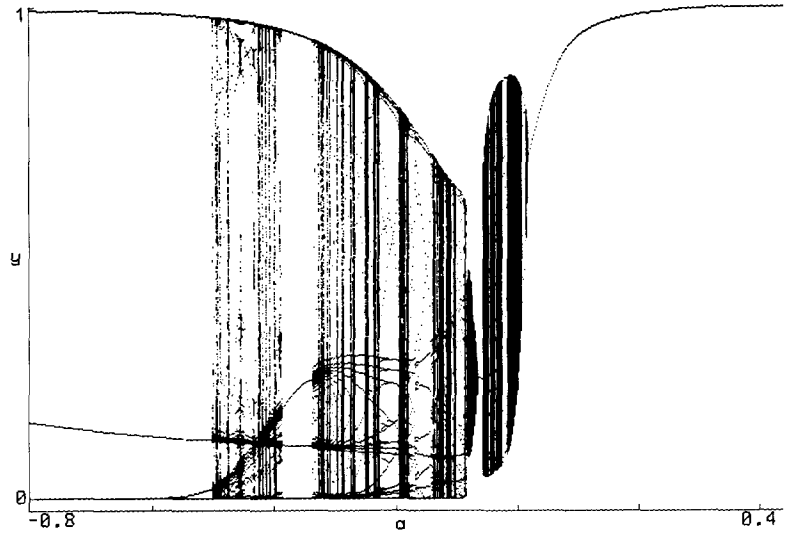


Fig. 13. Section (y_3 vs. y_7) of the chaotic attractor resulting from the period doubling to chaos at $r = 12.18$.



Node 3

Fig. 14. Bifurcation diagram for one node of an eight-node network ($r = 6.0$) with respect to varying a ($-0.8 \leqslant a < 0.41$).

where r_∞ is the accumulation point of the period doublings and r_n is the parameter value of the nth period doubling. Within the floating-point accuracy of the computer on which the system was simulated, we were able to determine r values for the first six period doublings, from which a value for δ was extracted, $\delta = 4.67 \pm 0.04$, which is in good agreement with the accepted value of $\delta = 4.6692...$. This route to chaos indicates that this system may be reduced to a one-dimensional system of quadratic form in this region of parameter space. A section of the resultant attractor is shown for $r = 12.18$ (Fig. 13). D_2 was again computed using the Grassberger–Procaccia[21] method. A value of $D_2 = 2.21 \pm 0.02$ was obtained at $r = 12.18$.

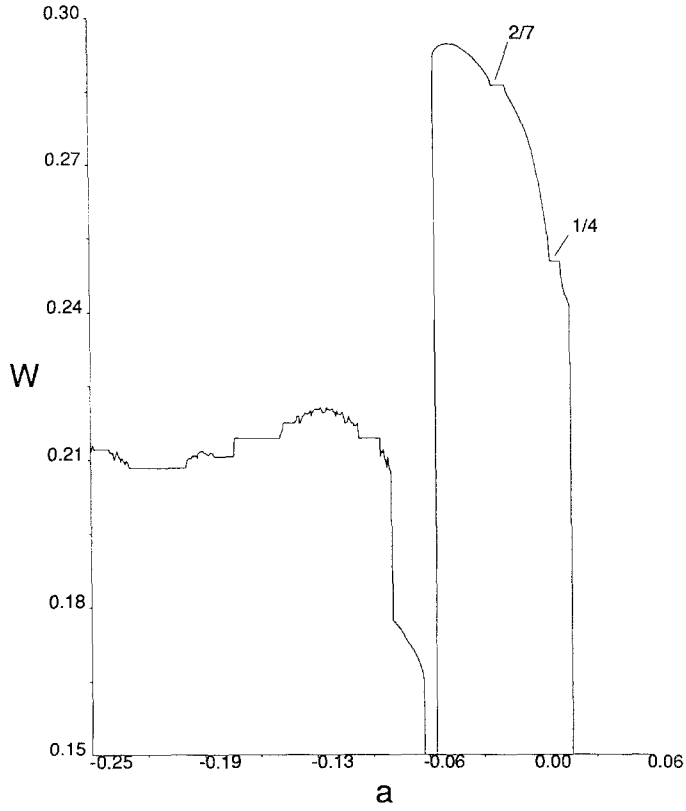


Fig. 15. Plot of winding number W against a at $r = 6.0$. Note that the devil's staircase is not monotonically increasing.

3.2. Symmetry Dependence

Simulations were performed on this eight-node system in which r was held constant and a was varied. A bifurcation diagram in the region $-0.8 \leqslant a < 0.44$ ($r = 6.0$) is shown in Fig. 14. A Hopf bifurcation occurs at $a = 0.0153$ and single-frequency quasiperiodic behavior ensues. The system displays mode locking at rational winding numbers and a plot of W vs. a may be made (Fig. 15). After mode locking at $W = 1/4$ a second bifurcation occurs at $a = 0.004$, introducing a second incommensurate frequency into the system. The system abruptly bifurcates back to fixed-point behavior at $a = -0.06$ (see the expanded bifurcation diagram, Fig. 16), bifurcating back to one-frequency behavior at $a = 0.675$; at $a = 0.685$ the system bifurcates to two-frequency behavior and the system again undergoes a transition to chaos via the Ruelle–Takens–Newhouse route, with an abrupt transition occurring at $a = -0.088$. A section of the resultant attractor (y_3 vs. y_1) is displayed (Fig. 17). Using the Grassberger–Procaccia[21] method, the correlation dimension (D_2) was computed for the attractor (32,000 points were computed after 10,000 iterations to run out transients) giving a value of $D_2 = 1.54 \pm 0.01$ at $a = -0.088$.

The dynamics of the system also appears to be more complicated than when varying r and holding a constant, with multiple coexisting attractors,

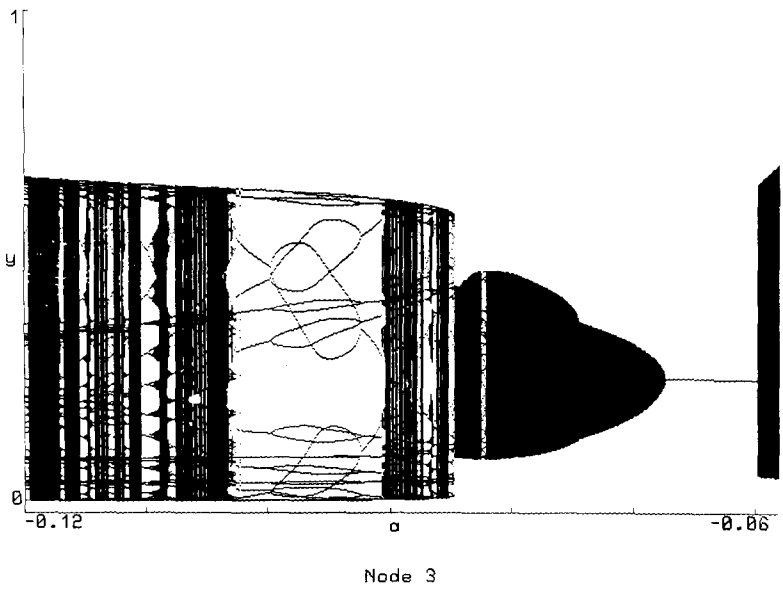


Fig. 16. Bifurcation diagram for one node of an eight-node network ($r = 6.0$) with respect to varying a ($-0.12 \leqslant a < -0.059$).

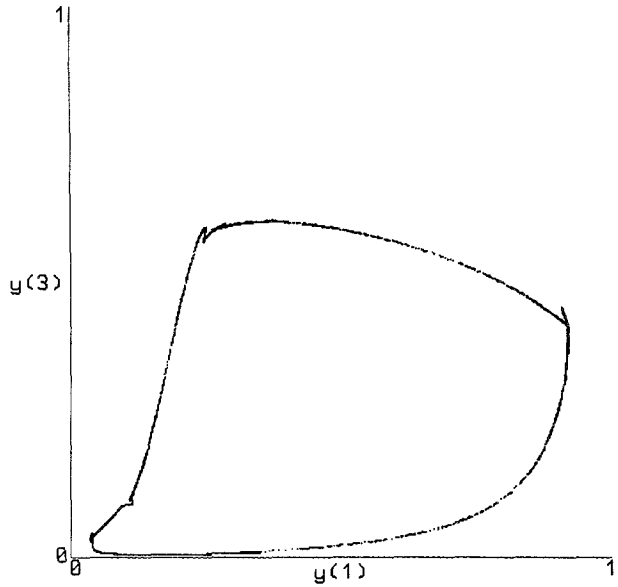


Fig. 17. Attractor section y_3 vs. y_1 for $r = 6.0$, $a = -0.088$. The resultant chaotic attractor has a very thin structure, corresponding to its low dimensionality ($D_2 = 1.54$).

period doubling cascades, and torus destructions (see Fig. 15). Further investigation is warranted in this area. (The dynamics in this region of parameter space is similar to that observed by Parlitz and Lauterborn[19] in a study of a driven van der Pol oscillator.)

The "circle map" behavior exhibited here is also shown by systems such as the one defined in (10). This system also has two control parameters (K, Ω), which, when plotted against each other, give a phase diagram showing periodic and nonperiodic behavior. The areas of periodic behavior appear on this phase diagram as *Arnol'd tongues*. We conjecture that our system is similar to (10) and that both K and Ω are nonlinear functions of (r, a).

4. DISCUSSION

A most important problem in engineering neural networks is that of customizing networks to produce attractors to perform a desired computation: this is the temporal training problem. Several algorithms have been proposed for this task,[23-29] but there is as yet no general algorithm to design a network to emulate a given dynamical process. Baird[30] proposes an algorithm that inverts the problem by designing a set of attractors and

their basins of attraction by specifying desired eigenvectors. Rosenblatt[31] suggested a perceptron with sequential memory: this extension to perceptron theory utilized a set of feedback units with system timing properties. Rosenblatt's specification for these units is that they should display non-repetitive deterministic dynamics, i.e., chaotic dynamics. A related problem is that of using a recurrent neural network to act as a temporal associative memory. Neural networks displaying chaotic dynamics may be important in this case[7,8] by allowing for a "knowing that you don't know situation" being signaled by being on a strange attractor: this would allow a broad search of state space, as well as serving as priming for a novel stimulus. Gardner *et al.*[32] have analytically calculated the storage capacity (where a stored pattern corresponds to a fixed point) of a fully-connected neural network model (utilizing a step transfer function) with respect to the symmetry (a) and a stability parameter (κ) corresponding to the size of the basins of attraction around fixed points. The theoretical curve that they derive gives the region of maximal storage to be when the weight matrix has a larger symmetric component ($a = 0.3$ when $\kappa = 0$, $a \to 1.0$ when $\kappa \to \infty$). To the extent that the maximal r values surveyed (Section 2) are good approximations to ∞, then we may locate the region of maximal storage in Figs. 1 and 2 as areas of high power and low spectrum entropy.

The phenomenon of intermittency[6] would also be useful for associative memory. In this route to chaos, a fixed point bifurcates (typically via a tangent bifurcation) and the resultant attractor has long, regular phases together with shorter bursts of irregular motion. This could clearly be useful in a time-dependent associative memory, as it would provide a mechanism for moving from a memory some time after recall without any external stimulus: a self-priming mechanism.

Chaotic attractors act as information generators in some directions and information compressors in others. By partitioning the phase space and labeling with symbols of an alphabet, it is feasible that a dynamical system featuring coexisiting chaotic attractors could be used as a method of generating strings of symbols. It may be hypothesized that certain attractors, partitioned in an appropriate way, may approximate to an nth-order Markov process. Nicolis *et al.*[33] show how the Rössler attractor may approximate to a fifth-order Markov chain. Such considerations may be important in applying dynamical systems methods to high-level cognitive and linguistic problems. Crutchfield and Young[34] present a technique that reconstructs a minimal set of equations of motion via a variable-order Markov model.

The observed regions of mode locking are extremely interesting. Systems displaying mode locking are generally multifrequency systems; e.g., they may have a natural frequency and a driving frequency. There is no

obvious mapping from the control parameters used here (r and a) to a pair of frequencies. It may be hypothesized that what is being observed is the interaction of two (or more) subnets that nonlinearly combine to produce the overall behavior. The frequencies of these subnets may be nonlinear functions of (r, a). This could be used for some kind of local system timing, should it be required in a particular network. The phenomenon of locking into rational winding numbers may also have implications for fault tolerance and learning: when a system moves into an appropriate region of parameter space, it may lock into a desired rational winding number.

In the region near the Hopf bifurcation from fixed-point behavior to (quasi-) periodic behavior very long-lived transients are observed. Kantz and Grassberger[35] suggest that chaotic transients are more robust against noise than the true attractors and they are in fact "more typical" of the dynamical system. This may be important for schemes in which learning dynamics is run concurrently with the network dynamics. A system could exist on a robust, long-lived transient according to the network dynamics, while the learning dynamics makes changes to the parameters of the dynamical system (and hence the less robust true attractor). Exploitation of these properties would allow the system to explore parameter space to find an attractor suitable for the computation at hand, while existing on a robust transient.

Studies of continuous-time counterparts to discrete mappings suggest that continuous systems tend to be more stable.[36] Typical of the continuous models is the one studied by Pineda[2]:

$$\frac{dy_i}{dt} = -y_i + f\left(\sum_j w_{ij} y_j + I_i\right) \tag{12}$$

where f is the transfer function and I_i represent external biases. A discretized version of this system is

$$y_i(t + \Delta t) = (1 - \Delta t)\, y_i(t) + \Delta t\, f\left(\sum_j w_{ij} y_j(t) + I_i\right) \tag{13}$$

When $\Delta t = 1$ and there is no external bias, then we have (1); Pineda[2] has suggested that discretizing with $\Delta t = 0.9$ imparts considerably stability into the system, compared with (1). Ottaway *et al.*[37] report that after numerical simulation of (12) with random initial weight matrices, all networks with 50 or less nodes possessed stable fixed points, and only 2 out of 500 100-node networks, with weights uniformly randomly initialized in the region ± 1, did not display fixed-point behavior. These results are similar to the discrete case, as these initialization conditions locate the networks in the low-r ($r < 3.3$) region of Fig. 2; the value of Δt in the discretization is

not reported. However, complex dynamical activity has been demonstrated in (12): Kürten and Clark[13] have reported chaotic behavior in such systems, with greater than 25 nodes, and a limited fan-in of eight weights inputing to each node; Pearlmutter[29] has trained small networks governed by dynamical law (12) to display complex limit cycles. Immediate future work will focus on comparing discrete dynamical systems with their continuous counterparts.

Finally, it must be emphasized that the design of general dynamical training algorithms is not a trivial task. However, the techniques used to study random networks in this work may be valuable tools with which to analyze temporal neural network training algorithms.

ACKNOWLEDGMENTS

We have benefitted from discussions with David Broomhead, David Wallace, David Willshaw, Sara Solla, and K. E. Kürten. One of the authors (S.R.) was supported by a U.K. Science and Engineering Research Council studentship. This work was supported by the Information Engineering Directorate/Science and Engineering Research Council as part of the IED/SERC Large Scale Integrated Speech Technology Demonstrator Project (SERC grants D/29604, D/29611, F/10309, F/10316) in collaboration with Marconi Speech and Information Systems and Loughborough University of Technology.

REFERENCES

1. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
2. F. J. Pineda, Dynamics and architecture in neural computation, *J. Complexity* **4**:216 (1988).
3. M. R. Guevara, L. Glass, M. C. Mackey, and A. Shrier, Chaos in neurobiology, *IEEE Trans. Syst. Man Cyber.* **SMC-13**:790 (1983).
4. A. Babloyantz and A. Destexhe, Low-dimensional chaos in an instance of epilepsy, *Proc. Natl. Acad. Sci. USA* **83**:3513 (1986).
5. C. A. Skarda and W. J. Freeman, How brains make chaos to make sense of the world, *Behav. Brain Sci.* **10**:161 (1987).
6. H. G. Schuster, *Deterministic Chaos*, 2nd ed. (VCH Verlagsgesellschaft, Weinheim, Federal Republic of Germany, 1988).
7. G. Parisi, Asymmetric networks and the process of learning, *J. Phys. A: Math. Gen.* **19**:L675 (1986).
8. M. Lewenstein and A. Nowak, Fully connected neural networks with self-control of noise levels, *Phys. Rev. Lett.* **62**:225 (1989).
9. M. Y. Choi and B. A. Huberman, Dynamic behaviour of nonlinear networks, *Phys. Rev. A* **28**:1194 (1983).

10. K. L. Babcock and R. M. Westervelt, Dynamics of simple electronic neural networks, *Physica* **28D**:305 (1987).
11. H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, *Phys. Rev. Lett.* **61**:259 (1988).
12. U. Riedel, R. Kühn, and J. L. van Hemmen, Temporal sequences and chaos in neural nets, *Phys. Rev. A* **38**:1105 (1988).
13. K. E. Kürten and J. W. Clark, Chaos in neural systems, *Phys. Lett.* **114A**:413 (1986).
14. C. M. Marcus and R. M. Westervelt, Dynamics of analog neural networks with time delay, in *Advances in Neural Information Processing Systems*, Vol. 1, D. S. Touretzky, ed. (Morgan Kaufmann, San Mateo, California, 1989), p. 568.
15. N. J. A. Sloane, Lengths of cycle times in random neural networks, Cognitive Systems Research Program Report No. 10, Cornell University, Ithaca, New York (1967).
16. A. Frumkin and E. Moses, The physicality of the Little model, *Phys. Rev. A* **36**:714 (1986).
17. H. Gutfreund, J. D. Reger, and A. P. Young, The nature of attractors in an asymmetric spin glass with deterministic dynamics, *J. Phys. A: Math. Gen.* **21**:2775 (1988).
18. M. H. Jensen, P. Bak, and T. Bohr, Transition to chaos by interaction of resonances in dissipative systems: I. Circle maps, *Phys. Rev. A* **30**:1960 (1984).
19. U. Parlitz and W. Lauterborn, Period doubling cascades and devil's staircases of the driven van der Pol oscillator, *Phys. Rev. A* **36**:1428 (1987).
20. S. Newhouse, D. Ruelle, and F. Takens, Occurrence of strange Axiom-A attractors near quasiperiodic flow on T^m, $m \leqslant 3$, *Commun. Math. Phys.* **64**:35 (1978).
21. P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Physica* **9D**:189 (1983).
22. M. J. Feigenbaum, The universal properties of nonlinear transformations, *J. Stat. Phys.* **19**:25 (1978).
23. P. Kanerva, Self-propagating search: A unified theory of memory, Ph.D. thesis, CSLI, Stanford University, Stanford, California (1984).
24. S. Dehaene, J.-P. Changeux, and J.-P. Nadal, Neural networks that learn temporal sequences by selection, *Proc. Natl. Acad. Sci. USA* **84**:2727 (1987).
25. I. Guyon, L. Personnaz, J.-P. Nadal, and G. Dreyfus, Storage and retrieval of complex sequences in neural networks, *Phys. Rev. A* **38**:6365 (1988).
26. R. J. Williams and D. Zipser, A learning algorithm for continually running fully connected neural networks, ICS Report 8805, University of California at San Diego, La Jolla, California (1988).
27. R. Rohwer, The "Moving Targets" training method, in *Proceedings of the DANIP Workshop*, J. Kindermann and A. Linden, eds. (GMD, Sankt Augustin, Federal Republic of Germany, 1989).
28. A. C. C. Coolen, J. J. Denier van der Gon, and T. W. Ruijgrok, An exact dynamical solution for generalized Hopfield networks, in *Neural Networks from Models to Applications*, L. Personnaz and G. Dreyfus, eds. (IDSET, Paris, 1989), p. 269.
29. B. A. Pearlmutter, Learning state-space trajectories in recurrent neural networks, *Neural Computation* **1**:263 (1989).
30. B. Baird, Bifurcation theory methods for programming static or periodic attractors and their bifurcations in dynamic neural networks, in *Proceedings IEEE 2nd International Conference on Neural Networks* (San Diego, California, 1988), I-9.
31. F. Rosenblatt, A model for experimental storage in neural networks, in *Computer and Information Sciences*, J. T. Tou and R. H. Wilcox, eds. (Spartan Books, Washington, D.C., 1964).
32. E. Gardner, H. Gutfreund, and I. Yekutieli, The phase space of interactions in neural networks with definite symmetry, *J. Phys. A: Math. Gen.* **22**:1995 (1989).

33. G. Nicolis, C. Nicolis, and J. S. Nicolis, Chaotic dynamics, Markov partitions, and Zipf's law, *J. Stat. Phys.* **54**:915 (1989).
34. J. P. Crutchfield and K. Young, Inferring statistical complexity, *Phys. Rev. Lett.* **63**:105 (1989).
35. H. Kantz and P. Grassberger, Repellers, semi-attractors and long-lived chaotic transients, *Physica* **17D**:75 (1985).
36. M. Y. Choi and B. A. Huberman, Digital dynamics and the simulation of magnetic systems, *Phys. Rev. B* **28**:2547 (1983).
37. M. B. Ottaway, P. Y. Simard, and D. H. Ballard, Fixed point analysis for recurrent neural networks, in *Advances in Neural Information Processing Systems*, Vol. 1, D. S. Touretzky, ed. (Morgan Kaufmann, San Mateo, California, 1989), p. 149.