PHYSICAL REVIEW A VOLUME 38, NUMBER 2 JULY 15, 1988

Temporal sequences and chaos in neural nets

U. Riedel, R. Kühn, and J. L. van Hemmen
Sonderforschungsbereich 123 an der Universität Heidelberg, D-6900 Heidelberg, West Germany
(Received 13 November 1987)

Exact deterministic equations for the evolution of temporal sequences with a delay stabilization and heat-bath dynamics are derived. The temperature may be arbitrary and both parallel and sequential updating are allowed. The equations display limit-cycle oscillations with a broad diversity of waveforms. For exponential delay it is shown how an increase of the transition amplitude leads to local chaos.

Temporal associations, such as reciting a poem, counting, and the control of rhythmic movements, are integral parts of the repertoire of our nervous system. Though an extensive formalism[1-6] has been developed to explain the associative recollection of a memory, as yet no analytic theory is available to describe the neurodynamics involved in temporal associations. In this paper, such a theory is developed, and it is shown under what conditions regular sequences are present, and how a transition into chaos can occur.

By now the collective behavior of neural nets with symmetric couplings is rather well understood.[1-6] The basic idea is to introduce an energy function or Hamiltonian

$$H_N = -\tfrac{1}{2}\sum_{i,j} J_{ij} S_i S_j , \tag{1}$$

with suitable symmetric couplings $J_{ij}=J_{ji}$, to store the data in the J_{ij}, model the neurons by Ising spins $S_i=\pm 1$, $1\le i\le N$, and let the system perform a *downhill* motion in the (free-) energy landscape associated with H_N. However, this type of dynamics does not allow a temporal sequence: a downhill motion ends up in a (free-) energy minimum and the system stays there.

Transitions between patterns require a certain amount of asymmetry. To this end, the synaptic efficacies are split up into two parts, a symmetric part $T_{ij}^{(1)}=\sum_{\mu=1}^{q}\xi_{i\mu}\xi_{j\mu}$, which is chosen so as to stabilize a set of q binary random patterns $\{\xi_{i\mu};1\le i\le N\}$ with $1\le\mu\le q$, and an asymmetric part, $T_{ij}^{(2)}=\sum_{\mu=1}^{p}\xi_{i\mu+1}\xi_{j\mu}$, $p\le q$, which is expected (see below) to induce transitions between the patterns. We have a *cycle* if $p+1\equiv 1 (\mathrm{mod} p)$, which we assume from now on.

The above setup can be simplified. First, p is taken to be finite. (Nobody has ever recited a poem of a length comparable to the total amount of the data stored.) The case of extensively many patterns (p finite, $q=\alpha N$ with $\alpha>0$) will be treated at the end of this paper. For finite q we can also assume $p=q$, since, as we will see shortly, the patterns outside the cycle do not play any active role.

If we take $J_{ij}=J_{ij}^{(1)}+\epsilon J_{ij}^{(2)}$ with $J_{ij}^{(\alpha)}=N^{-1}T_{ij}^{(\alpha)}$, $\alpha=1,2$, then[1] nothing will happen for small ϵ, whereas for larger values of ϵ, the sequence nearly instantaneously gets mixed completely. The reason is that $J_{ij}^{(2)}$ wants to induce yet another transition, $\mu\to\mu+1$, *as soon as* μ appears. A solution to this problem was found by Kleinfeld and by Sompolinsky and Kanter.[7-9] They introduced a *delay* which stabilizes the system in each state μ before it makes the transitions to $\mu+1$. If

$$\bar{S}_j(t)=\int_0^\infty ds\, w(s) S_j(t-s) \tag{2}$$

is the delay term which samples the past up to time t with positive weight w, then the postsynaptic potential (local field) experienced by neuron i is assumed to be given by

$$h_i(t)=\sum_j J_{ij}^{(1)} S_j(t)+\epsilon\sum_j J_{ij}^{(2)}\bar{S}_j(t) . \tag{3}$$

The weight or memory kernel w is normalized to one. Typical weight functions, which all have a natural decay time τ, are $w(t)=\delta(t-\tau)$ for δ-function delay, $w(t)=\tau^{-1}$ for $0\le t\le\tau$ and vanishing elsewhere for step-function delay, and $w(t)=\tau^{-1}\exp(-t/\tau)$ for exponential delay. From a physiological point of view,[10] exponential and δ-function delay are the most natural ones.

Using the overlaps $m_\mu(t)=N^{-1}\sum_i\xi_{i\mu}S_i(t)$, we can write

$$h_i(t)=\sum_\mu \xi_{i\mu} m_\mu(t)+\epsilon\sum_\mu \xi_{i\mu+1}\bar{m}_\mu(t) . \tag{4}$$

If a system operating at zero temperature ($T=0$) has made a transition $\nu-1\to\nu$ at time $t=0$, then we expect it to make another transition $\nu\to\nu+1$ only after a time of the order τ and *provided* $\epsilon>1$. Indeed, under these conditions, stable cycles were shown to exist,[7,8] mainly through numerical simulations on small samples.

The equations of motion in the N-dimensional phase space ($N\to\infty$) generated by either a sequential or a parallel dynamics of the Glauber type at zero or finite temperature. In this paper, we show how to reduce these equations, which are not amenable to an analytic treatment, to a finite system of nonlinear (ordinary differential) equations with delay. This reduction is *exact*. It allows a detailed investigation of rather surprising results on the behavior of the solutions as a function of time and the role of temperature.

We start by taking $J_{ij}^{(1)}=N^{-1}Q^{(1)}(\xi_i;\xi_j)=J_{ji}^{(1)}$ and $J_{ij}^{(2)}=N^{-1}Q^{(2)}(\xi_i;\xi_j)\neq J_{ji}^{(2)}$ where $\xi_i=\{\xi_{i\mu};1\le\mu\le q\}$ is the local information available to neuron i. The case studied until now has $J_{ij}^{(\alpha)}=\phi(T_{ij}^{(\alpha)})$, $\alpha=1,2$, with $\phi(x)=x$ and is therefore called linear. However, *double clipping* with $\phi(x)=\mathrm{sgn}(x)$ will be shown to work equally well. It is relevant to hardware realizations.

Given q binary random patterns, there are 2^q different

positions available to the random vectors ξ_i, viz. the 2^q corners $\mathbf{x}$ of the q-dimensional hypercube $C^q=[-1,1]^q$. Each corner $\mathbf{x}$ has a probability $p(\mathbf{x})$. Introducing[5] the sublattices $I(\mathbf{x})=\{i;\xi_i=\mathbf{x}\}$ and sublattice magnetizations,

$$m(\mathbf{x};t)=|I(\mathbf{x})|^{-1}\sum_{i\in I(\mathbf{x})}S_i(t)\,, \tag{5}$$

one easily verifies that the local field at i is

$$h_i(t)=\sum_{\mathbf{y}\in C^q}[Q^{(1)}(\xi_i;\mathbf{y})m(\mathbf{y};t) +\epsilon Q^{(2)}(\xi_i;\mathbf{y})\bar{m}(\mathbf{y};t)]p_N(\mathbf{y})\,,$$

where $p_N(\mathbf{y})=N^{-1}|I(\mathbf{y})|$ converges to $p(\mathbf{y})$ with probability 1 as $N\to\infty$. For unbiased patterns, $p(\mathbf{y})=2^{-q}$. Finally, we note that if i belongs to $I(\mathbf{x})$, then $h_i(t)$ only depends on $\mathbf{x}$ since $\xi_i=\mathbf{x}$ for all i in $I(\mathbf{x})$. In the bulk limit $h_i(t)$ therefore converges to

$$h(\mathbf{x};t)=\sum_{\mathbf{y}\in C^q}[Q^{(1)}(\mathbf{x};\mathbf{y})m(\mathbf{y};t)+\epsilon Q^{(2)}(\mathbf{x};\mathbf{y})\bar{m}(\mathbf{y};t)]p(\mathbf{y})\,, \tag{6}$$

whatever i in $I(\mathbf{x})$.

The effect of temperature, i.e., a heat bath, is simulated by requiring that the probability of flipping a spin ($\Delta S_i\neq 0$) be given by $[1+\exp(-\beta h_i(t)\Delta S_i)]^{-1}$. Updating may be either sequential or parallel.

Suppose a parallel updating is performed after each elementary time step Δt. We then find in the limit $N\to\infty$, that for all i in $I(\mathbf{x})$, the thermal average of $S_i(t+\Delta t)$ is $\tanh[\beta h(\mathbf{x};t)]$. The size $|I(\mathbf{x})|$ grows with N so that by (5), (6), and the law of large numbers, the dynamics converges to

$$m(\mathbf{x};t+\Delta t)=\tanh[\beta h(\mathbf{x};t)]\,. \tag{7}$$

This is a recursion relation, which has to be iterated.

If we use sequential updating, only one spin per elementary time step Δt is updated. To obtain an extensive change as N becomes large one has to rescale time so that $\Delta t\propto N^{-1}$; for details, see Ref. 11. Then one obtains a set of 2^q coupled nonlinear ordinary differential equations

$$\dot{m}(\mathbf{x};t)=-\Gamma\{m(\mathbf{x};t)-\tanh[\beta h(\mathbf{x};t)]\}\,, \tag{8}$$

where Γ is the mean attempt rate per (rescaled) unit time. Both (7) and (8) are exact, whatever the neural activity level and the synaptic kernels Q.

Because of the delay contained in $h(\mathbf{x};t)$, Eq. (8) represents a so-called *functional differential equation.* Explicit solutions are not known, but (8) can be solved numerically. The outcomes are presented most conveniently in terms of the overlaps $m_\mu(t)$. Plainly,

$$m_\mu(t)=\sum_{\mathbf{x}\in C^q}p(\mathbf{x})x_\mu m(\mathbf{x};t)\equiv\langle x_\mu m(\mathbf{x};t)\rangle\,. \tag{9}$$

In the case of the linear model, and only here, we can reduce (8) to a set of q equations for the overlaps themselves,

$$\dot{m}_\mu(t)=-\Gamma\left[m_\mu(t)-\left\langle x_\mu\tanh\left\{\beta\sum_\nu x_\nu[m_\nu(t)+\epsilon\bar{m}_{\nu-1}(t)]\right\}\right\rangle\right] \tag{10}$$

One easily verifies that for patterns ν not involved in the cycle we can put $m_\nu(t)=0$.

We now turn to the results. First, whatever the updating, double clipping works as well as the linear model. This is exemplified by Figs. 1(a) and 1(b). In our numerical work, the patterns were taken to be unbiased.

Second, for $\epsilon>1$, cycles may run at temperatures far above the critical temperature of the static model ($\epsilon=0$) [see Fig. 1(c)]. This implies that at *high* temperatures, the picture of $\epsilon J_{ij}^{(2)}$ pushing the system through a free-energy landscape created by $J_{ij}^{(1)}$ is incorrect.

Third, for $\epsilon<1$, thermal noise ($T>0$) allows the persistence of (pure) cycles, which, as we have seen, cannot exist at zero temperature. [See Fig. 2 for a typical phase diagram in the (β,ϵ) plane.] As the inverse temperature β increases, we first have a continuous transition where the cycle appears out of the completely symmetric solution $m_\mu(t)=m$, $1\le\mu\le q$; at high temperatures the latter is stabilized by $\epsilon>0$. As β increases further, there is a first-order, discontinuous, transition when the cycle gets stuck in one of the free-energy valleys which it visits during its journey in phase space.[12]

At low temperatures, the picture of $J_{ij}^{(1)}$ creating a (free-) energy landscape where $\epsilon J_{ij}^{(2)}$ simply induces transitions between the valleys functions reasonably well. For instance, in the original Hopfield model[1] m-symmetric states ($m\le q$) are stable at low temperatures if m is *odd.*[3] Taking now $q=4$, say, and starting with the 3-

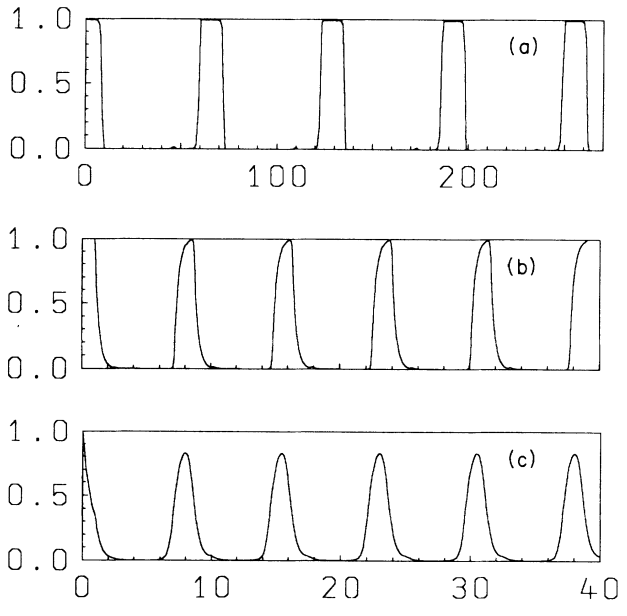

FIG. 1. (a) Overlap with the first pattern as a function of time for double clipping with step-function delay, parallel updating, and $\tau=15$ iterations. (b) Double clipping with δ-function delay and sequential updating. Here and elsewhere (except Fig. 4), $\tau=1$ and $\Gamma=3$. As in (a), $\beta=5$. (c) The *linear* model with the same delay and updating as in (b), but with $\beta=0.7<\beta_c^{\text{Hopfield}}$. In all three cases, $\epsilon=1.5$ and $q=5$.

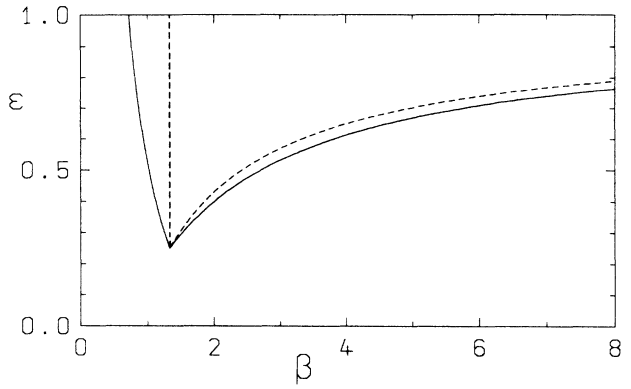

FIG. 2. Exact dynamical phase diagram for a cycle of length $q=2$ with sequential updating, δ-function delay (solid curve) or exponential delay (dashed curve). The lowest point $(\beta,\epsilon)=[1+\gamma^{-1},(1+\gamma)^{-1}]$ with $\gamma=\Gamma\tau$ is apparently universal. In the V-shaped region starting here, a pure cycle exists.

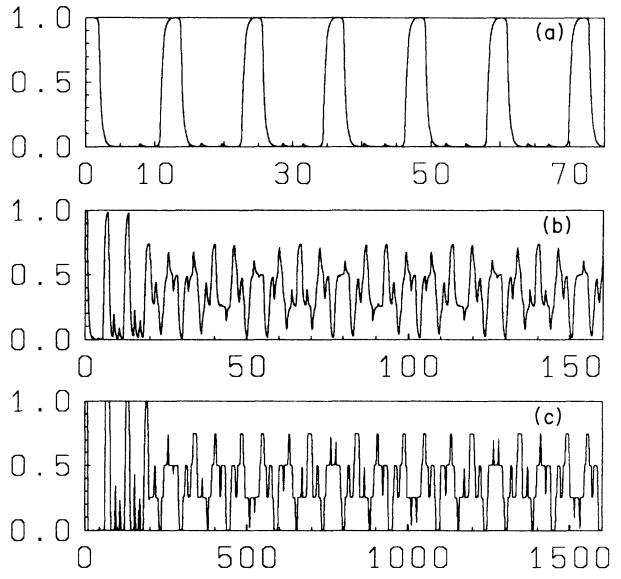

FIG. 3. Overlap with the first pattern as a function of time for exponential delay with $q=4$ and $\beta=10$. We have sequential dynamics in (a) and (b) and parallel dynamics in (c). The pure cycle in (a) has $\epsilon=1$. In (b) and (c), a globally attracting solution is shown for $\epsilon=1.80$. Note the transient pure cycle and the local chaos.

symmetric state (1,2,3) one finds a so-called 3-symmetric cycle $(1,2,3)\rightarrow(2,3,4)\rightarrow\cdots$ if ϵ in $\epsilon J_{ij}^{(2)}$ is large enough. A generous increase of temperature, however, will destroy the symmetric cycle and leave one with the pure $1\rightarrow 2\rightarrow 3\rightarrow 4\rightarrow 1$, in a way which is surprisingly similar to the statics.[3]

Finally, we turn to the rather fascinating behavior of the cycles as we increase ϵ. Whatever $\epsilon>1$, δ-function and step-function delay always allow pure cycles; cf. Fig. 1. Exponential delay, however, gives rise to a rather different behavior. In itself a very natural choice,[10] it samples the remote past much better than the previous two memory kernels.

Let us fix the temperature T and increase $\epsilon\geq 1$. At low T, we start with pure cycles but, as ϵ increases, the system apparently wants to perform a transition to a *fully symmetric* state. This reminds us of the malfunctioning of the model without delay and is consistent with the exponential distribution sampling the remote past. For odd q, the q-symmetric state is stable and invariant, for $\epsilon\geq\epsilon_c$ it is globally attracting, so wherever the system starts, it finally ends up here. The threshold ϵ_c depends on T. At $T=0$, we have $\epsilon_c\approx 1.8$.

For even q, a new type of solution appears, since at low temperatures the q-symmetric state is not stable. In the case of $q=4$ and $T=0.1$, a "3-symmetric" cycle is globally attracting for $\epsilon>1.75$ and, with it, apparently chaotic behavior appears. For instance, at $\epsilon=1.80$, the orbit looks very chaotic [see Fig. 3(b)]. Indeed, locally it is, though a closer examination reveals that there exists something like a global period, much longer than the one of the pure cycle at $\epsilon=1$. As shown in Fig. 3(c), parallel dynamics does not allow any periodicity. In both cases, however, a sensitive dependence upon the initial conditions[13] could not be found yet. The feature that is new is the local complexity, which increases with β and ϵ.

We now turn to the question how to describe a regular and finite cycle of length p in the presence of $q=\alpha N$ patterns. The intuitive idea is that the system spends most of its time in a specific valley (ergodic component), where it can be described by equilibrium thermodynamics. We make, therefore, a slight detour and turn to the computation of the free energy[3,6] for the model without any transition term ($\epsilon=0$). As in the ordinary Hopfield case,[3] we single out the finitely many patterns which are involved in the cycle (to be labeled by μ) and average over the rest. Whatever the nonlinearity in Q, the expression for the free energy[6] then consists of three parts. The first refers to the μ patterns we concentrate on, the second only depends on the order parameter $\mathcal{q}$ which represents the noise generated by the other, extensively many patterns, and the last describes the interaction between the two groups of patterns in terms of $\mathcal{q}$ and the overlaps m_μ.

Fixing $\mathcal{q}$, one straightforwardly verifies that the first and third term of the free energy can be derived from the effective Hamiltonian

$$H_{\text{eff}}=-\tfrac{1}{2}N\sum_\mu\Lambda_\mu\left[N^{-1}\sum_{i=1}^{N}\xi_{i\mu}S_i\right]^2-\sqrt{\mathcal{q}r}\sum_{i=1}^{N}h_iS_i\,, \quad (11)$$

where the h_i are independent Gaussians with mean zero and variance one while $r=r(\mathcal{q})$ is a given function[6] of $\mathcal{q}$ and Λ_μ is the embedding strength[6] of pattern μ. So the extensively many patterns outside the cycle produce a Gaussian random field of variance $\mathcal{q}r$. We then find, for instance, instead of (10),

$$\dot{m}_\mu=-\Gamma\left\{m_\mu-\left\langle\left\langle x_\mu\tanh\left[\beta\left(\sum_\nu x_\nu(m_\nu+\epsilon\bar{m}_{\nu-1})+\sqrt{\mathcal{q}r}z\right)\right]\right\rangle\right\rangle\right\}. \quad (12)$$

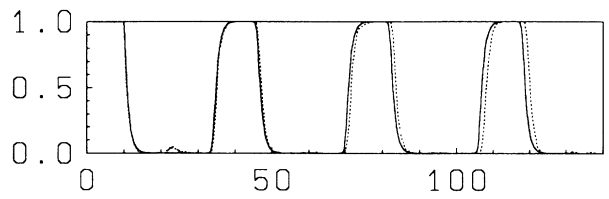

FIG. 4. Comparison between the analytic solution to Eq. (12) (solid curve) and numerical simulation on a finite sample of size $N=1000$ with Glauber dynamics (dashed curve). Here we have $\beta=10$, $\epsilon=1$, and δ-function delay with $\tau=10$ and $\Gamma=1$. A cycle of length three is running in the presence of 100 other patterns ($\alpha=0.10$). Only the overlap with the first of the cycle is shown.

The double angular brackets denote an average not only over the $x_\mu=\pm 1$ but also over the Gaussian z. A similar equation holds for the spin-glass order parameter q. As compared to (10), the role of the extensively many patterns is to smoothen the solution through the Gaussian z. In fact, their role is quite similar to the one of *temperature*.

The above procedure has no claim of rigor but as yet no exact solution to the problem of treating extensively many patterns has been found. Figure 4 shows, however, that our ansatz contains the essential physics. The agreement between the solution to (12) and the numerical simulation on a finite sample ($N=1000$) is excellent. The pulse forms agree exactly and the periods differ by only 2% (mainly due to finite-size effects).

In summary, we have derived exact equations for the dynamical evolution of temporal sequences, at arbitrary temperature and with either parallel or sequential updating. The method applies not only to pure cycles but also to cycles with repetitions (and multispin interactions), such as the ones used to describe the acquisition of song by birds.[14] Furthermore, the method allows a detailed study of new phenomena in the collective behavior of neurons. Only a few of them have been reported in the present work.

This work is supported by the Deutsche Forschungsgemeinschaft (Bonn).

[1]J. J. Hopfield, Proc. Nat. Acad. Sci. U.S.A. **79**, 2554 (1982); **81**, 3088 (1984).

[2]J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).

[3]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **173**, 30 (1987).

[4]G. Toulouse, S. Dehaene, and J.-P. Changeux, Proc. Nat. Acad. Sci. U.S.A. **83**, 1695 (1986).

[5]J. L. van Hemmen and R. Kühn, Phys. Rev. Lett. **57**, 913 (1986).

[6]J. L. van Hemmen, Phys. Rev. A **36**, 1959 (1987).

[7]D. Kleinfeld, Proc. Nat. Acad. Sci. U.S.A. **83**, 9469 (1986).

[8]H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986).

[9]Further physiological evidence in favor of the delay mechanism was provided by D. Kleinfeld and H. Sompolinsky, J. Neurosci. (unpublished).

[10]A. C. Scott, *Neurophysics* (Wiley, New York, 1977).

[11]R. B. Griffiths, C. Y. Weng, and J. S. Langer, Phys. Rev. **149**, 301 (1966).

[12]In a technical sense, these transitions are Hopf bifurcations. The first is supercritical, the second subcritical.

[13]H. L. Swinney, Physica D **7**, 3 (1983).

[14]S. Dehaene, J.-P. Changeux, and J.-P. Nadal, Proc. Nat. Acad. Sci. U.S.A. **84**, 2727 (1987).