PAPER *Special Section on Nonlinear Theory and its Applications*

Blinking Long-Range Connections Increase the Functionality of Locally Connected Networks

Martin HASLER[†a)], ***Member and*** **Igor BELYKH**[†b)], ***Nonmember***

SUMMARY Information processing with only locally connected networks such as cellular neural networks is advantageous for integrated circuit implementations. Adding long range connections can often enhance considerably their performance. It is sufficient to activate these connections randomly from time to time (*blinking connections*). This can be realized by sending packets on a communication network underlying the information processing network that is needed anyway for bringing information in and out of the locally connected network. We prove for the case of multi-stable networks that if the long-range connections are switched on and off sufficiently fast, the behavior of the blinking network is with high probability the same as the behavior of the time-averaged network. In the averaged network the blinking connections are replaced by fixed connections with low (average) coupling strength.

key words: *networks, CNN, blinking connection, multistable system, averaging*

1. Introduction

Analog circuits can speed up computations considerably. They are used e.g. for massively parallel processing of visual information. Often they consist of 1-d or 2-d arrays of simple circuit modules (cells) that are interconnected by wires. Examples are cellular neural networks where usually first order dynamical systems are placed in a regular 2-d array and next nearest neighbors are connected by wires [1].

Many functions can be computed by arrays of locally connected dynamical systems, but others require long range interactions of the cells for efficient computation. Hardwired all-to-all connections of $n \times n$ cells would require n^4 wires which is not realistic in most cases. In paper [2] just a few long range connections were added in order to form a small-world interaction graph. Even though in this case relatively few connections were used, one would still have to hardwire them which is not convenient.

In this paper we show that connections can actually be switched on and off randomly in such a way that with high probability the computational function the network performs is the same as that of a corresponding non-switched system, the *averaged system*. We call the switched connections *blinking* connections. The conditions for highly similar behavior is that the (temporal) mean strength of each blinking interaction between two cells is the same as in the non-switched network and that the switching is rapid enough. The blinking connections can be realized by information packages routed on a communication network that is needed anyway to transport information to and from the cells.

The switching is performed at times $0, \tau, 2\tau, \cdots$ Each blinking connection is turned on and off randomly, independently in each time interval and independently of other connections. The fact that the rapidly switched system has the same behavior as the averaged system seems obvious, but in fact there are rare exceptions and therefore a careful proof of the property is needed which shows on what parameters the occurrence of the exceptions depends. While averaging is a classical technique in the study of nonlinear oscillators [3], averaging for blinking systems needs some special mathematical techniques for obtaining rigorous convergence proofs. Such techniques have been used in [4] for synchronization of blinking networks of chaotic dynamical systems.

In this paper we consider multi-stable rather than chaotic systems. The information to be processed by a multi-stable network determines the initial state and the information processing takes place by convergence to one of the stable equilibria. Therefore, ideally the blinking network should have the same basins of attraction of the equilibria as the averaged network. We will show that with high probability the main part of the basins of attraction is indeed the same in the averaged and in the blinking network. The details of the proof are different from those in [4] but the basic idea is the same.

2. Example

In order to illustrate the topic, we use the example of a Winner-take-all (WTA) 2-dimensional CNN (cellular neural network). We do not claim that this is a particularly good example from an application point of view, but it illustrates well the subject of the paper. A WTA network is designed to find, using the network dynamics, the largest among a set of numbers. In the realization we refer to, each number is associated with a cell of the network. More precisely, the initial state of each cell of the CNN is set to the value of the corresponding number. The time-evolution of the winner is such that the output of the cell corresponding to the largest number converges to 1 and all other outputs converge to -1, in normalized variables. It is not difficult to see that in an only locally connected CNN it is not possible to realize the WTA

Manuscript received January 27, 2005.
Final manuscript received June 23, 2005.

[†]The authors are with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 14, CH-1015 Lausanne, Switzerland.
a) E-mail: martin.hasler@epfl.ch
b) E-mail: igor.belykh@epfl.ch

DOI: 10.1093/ietfec/e88–a.10.2647

function, at least not in this way. Indeed, suppose that the initial state of a loclly connected CNN has two local maxima, at cell i and at cell j and these maxima are sufficiently far apart. Suppose that at cell i the maximum is also global. If this network performs the WTA function correctly there must be a stable equilibrium for which the output of the i-th cell is +1 and all other outputs are −1. However, when all cells are in saturation, the j-the cell and the i-th cell do not interact. Then there will be another stable equilibrium where in addition to the i-th cell the j-th cell has output +1, and again all other cells have output −1. Such an equilibrium point is not compatible with the WTA function.

In [5], the design equations for a globally connected WTA CNN are given. Actually, the point is made in [5] that one does not need all-to-all connections, but only an additional sum cell which has inputs and outputs that are connected to all cells. This reduces the number of wired connections from n^4 to $2n^2$, which is even less than the $5n^2$ connections in a nearest neighbor connected CNN. However most of these $2n^2$ connections go across a large part of the circuit and thus still pose problems of realization. In our blinking WTA CNN the $n^4 - 5n^2$ switched connections can just as well be reduced, but for the simplicity and generality of exposition, we do not want to take advantage of the peculiarities of the example.

Following [5], we consider the piecewise linear dynamical system, whose state equations are given in (1), for $i = 1, \ldots, N = n^2$.

$$\frac{dx_i}{dt} = -x_i + (1+\delta)y_i - \alpha \sum_{j=1}^{n} y_i + \boldsymbol{\kappa}$$
$$y_i = f(x_i) = \begin{cases} 1 & \text{for} \quad x_i > 1 \\ x_i & \text{for} \quad |x_i| \le 1 \\ -1 & \text{for} \quad x_i < -1 \end{cases} \tag{1}$$

If the parameters α, β and κ are chosen suitably, there are exactly N asymptotically stable equilibrium points, one in each linear region where one of the output signals y_i is 1 and the other $N-1$ are −1. All other equilibrium points are unstable. The basin of attraction of the asymptotically stable equilibrium point with outputs

$$y_i = 1,\ y_j = -1,\ j \neq i \tag{2}$$

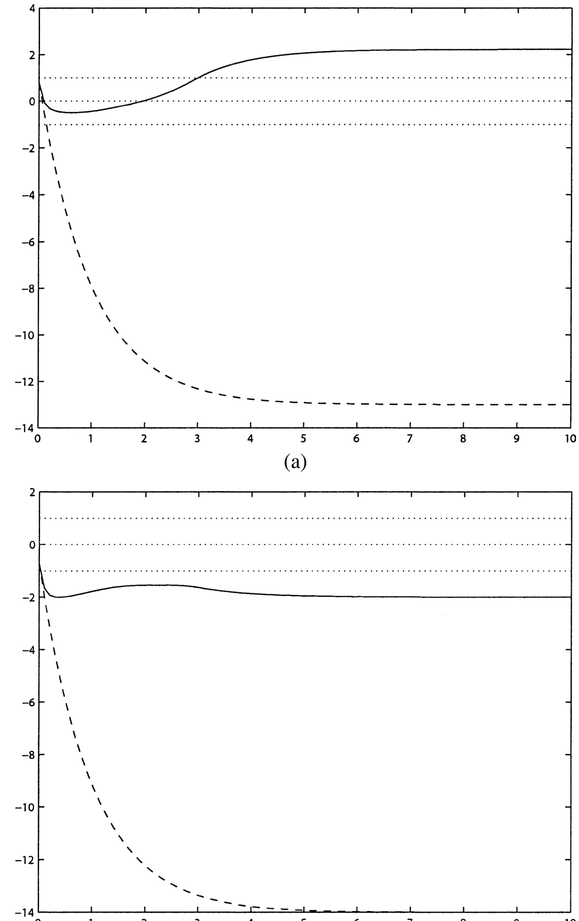
is composed of the state vectors $\boldsymbol{x}$ with

$$x_i > x_j \text{ for all } j \neq i \tag{3}$$

Hence, indeed the CNN realizes the WTA function by its internal dynamics. In Fig. 1, two components of the state trajectory $\boldsymbol{x}(t)$ of a 4×4 WTA CNN are shown, for the initial conditions.

$$\begin{matrix} 0.3644 & 0.3958 & 0.1871 & 0.2898 \\ -0.3945 & -0.2433 & -0.0069 & 0.6359 \\ 0.0833 & 0.7200 & 0.7995 & 0.3205 \\ -0.6983 & 0.7073 & 0.6433 & -0.3161 \end{matrix} \tag{4}$$

For the fully connected CNN (solid line), one can see that the state of the cell with largest initial condition (Fig. 1(a)) converges to a value higher than 1, whereas the state of any other cell (Fig. 1(b)) converges to a value lower than −1. Hence, asymptotically in time, the output of the cell with the largest initial state is 1 and the output of all other cells is −1. When only nearest neighbor connections are used (dashed line), the network fails to detect the largest initial value. In this case, all ouputs converge to −1.



(a)

(b)

Fig. 1 Trajectory component $x_i(t)$ for a 4×4 WTA CNN, with $a = 1, \delta = 1.11, \kappa = -13.89$ (solid line). (a) Component of the cell (3,3) whose initial condition has the maximal value. (b) Component of the cell (4,1). Dashed line: trajectory components of the CNN with the same parameters, but only nearest neighbor connections.

3. Convergence to an Asymptotically Stable Equilibrium

Consider a dynamical system of the more general form

$$\frac{dx_i}{dt} = -x_i + \sum_{i=1}^{N} A_{ij} y_j + I_i$$
$$y_i = f(x_i) \tag{5}$$

where f is continuously differentiable, and strictly increasing from −1 at −∞ to +1 at +∞ ("sigmoid function").

It is well-known that for a symmetric coupling matrix $\boldsymbol{A}$ almost all trajectories converge to an asymptotically stable equilibrium point. This fact is established using the Lyapunov function

$$W(\boldsymbol{y}) = \sum_{i=1}^{N} \int_{0}^{y_i} f^{-1}(z)dz - \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} y_i y_j - \sum_{i=1}^{N} I_i y_i \tag{6}$$

Along any solution of (5), we obtain, after some calculations

$$\begin{aligned} \frac{d}{dt}W(\boldsymbol{y}) &= -\sum_{i=1}^{N} \frac{df}{dx}\left(f^{-1}\left(y_i\left(t\right)\right)\right)\left[\frac{\partial W}{\partial y_i}\left(\boldsymbol{y}\left(t\right)\right)\right]^2 \\ &= -\sum_{i=1}^{N}\left[\frac{df}{dx}\left(f^{-1}\left(y_i\left(t\right)\right)\right)\right]^{-1}\left[\frac{dy_i}{dt}\right]^2 \\ &\leq 0 \end{aligned} \tag{7}$$

which implies the convergence property. Actually, the CNN with the piecewise-linear nonlinearity (1) does not fit exactly into this framework, because f is not invertible. The reasoning could be adapted to this case, but for the sake of simplicity, for the analysis we stick to the case with sigmoid f, whereas the simulations are performed using the piecewise linear output function.

4. Blinking Connections

We divide the time-axis into intervals of length τ and number them consecutively. Thus, interval k is defined by $(k-1)\tau \leq t < k\tau$. During each time interval, the circuit connections remain constant, but for different time intervals, the circuit connections are different. In the spirit of CNN's, however, nearest neighbor connections remain constant all the time. For each non-nearest-neighbor pair of cells i, j we introduce a discrete time binary signal s_{ij}^{k} whose value is 1 if the connection is switched on during the k-th time interval, and 0 otherwise. We call these signals *switching sequences*. We extend the switching sequences to continuous time switching signals by setting $s_{ij}(t) = s_{ij}^{k}$ in the k-th time interval.

Given a switching signal, we can write the equations of the blinking CNN as

$$\begin{aligned} \frac{dx_i}{dt} &= -x_i(t) + \sum_{j \text{ nn of } i} A_{ij} y_j(t) + \sum_{j \text{ not nn of } i} B_{ij} s_{ij}(t) y(t) + I_i \\ y_i &= f\left(x_i\right) \end{aligned} \tag{8}$$

The choice of the switching sequences is performed at random. More precisely, we consider the set of identically distributed independent random variables

$$S_{ij}^{k},\ i \text{ and } j \text{ not nearest neighbors},\ k = 1, 2, \ldots \tag{9}$$

which take the value 1 with probability p and the value 0 with probability $1-p$. The switching sequences are instances of this stochastic process.

Due to the random choice of switching, the trajectories of (8) become also random processes. Intuitively, if the switching is much faster than the intrinsic time constants of the trajectories of a corresponding non-switched system, we expect the trajectories of the blinking system to follow closely the trajectories of the non-switched system whose connections have the average value with respect to (8), the *averaged system*:

$$\begin{aligned} \frac{d\xi_i}{dt} &= -\xi_i(t) + \sum_{j \text{ nn of } i} A_{ij}\eta_j(t) + \sum_{j \text{ not nn of } i} B_{ij} p \eta_j(t) + I_i \\ \eta_i &= f\left(\xi_i\right) \end{aligned} \tag{10}$$

Since we would like the blinking system to have the same behavior as the fully connected system (5), we have to set

$$B_{ij} = A_{ij}/p \tag{11}$$

In particular, we expect that the trajectories of the blinking system with (11) gets close to the same asymptotically stable equilibrium of the averaged system, as $t \to +\infty$, when both start from the same initial condition. We shall prove that by choosing the switching time sufficiently small, we can make the probability that this is not the case arbitrarily small.

Note that the blinking system does not have the same equilibrium points as the averaged system. This implies that the trajectories of the blinking system can only come close to the stable equilibrium points of the averaged system without converging to them. However, from a practical point of view this is sufficient. Indeed, in our example, as soon as the solution gets sufficiently close to a stable equilibrium point of the averaged system, the system decides for the corresponding largest initial condition. In the piecewise linear system, such a decision can be taken as soon as one output value is +1 and the others −1.

5. Asymptotic Behavior of the Connections of the Blinking Multistable System

In order to limit the exposition to the essentials, we consider a more general blinking system of equations:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}\left(\boldsymbol{x}\left(t\right), \boldsymbol{s}\left(t\right)\right) \tag{12}$$

where $\boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{F}: \mathbb{R}^{N+M} \to \mathbb{R}^N$. As above, the binary signal $\boldsymbol{s}(t)$ is supposed to be constant in each time interval of length τ, starting at 0 and the value of $s_i(t)$ in the k-th time interval is denoted by s_i^k. As general properties of (12) we require that

- $\boldsymbol{F}$ is continuously differentiable.

- There is a compact set C such that all solutions of (12) join C in finite time and thereafter remain in C for any choice of the binary signal $\boldsymbol{s}(t)$.

All interesting dynamics take place in C and therefore we restrict the analysis to solutions that lie in C. In C, we have a bound

$$\|\boldsymbol{F}(\boldsymbol{x},\boldsymbol{s})\| \le B_F \quad \text{for} \quad \boldsymbol{x} \in C, \boldsymbol{s} \in \{0,1\}^M \tag{13}$$

and a Lipschitz constant

$$\|\boldsymbol{F}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{F}(\boldsymbol{y},\boldsymbol{s})\| \le L_F \|\boldsymbol{x}-\boldsymbol{y}\| \quad \text{for} \quad \boldsymbol{x},\boldsymbol{y} \in C, \boldsymbol{s} \in \{0,1\}^M \tag{14}$$

As before, we suppose that the binary signal $\boldsymbol{s}(t)$ is generated by a random process $\boldsymbol{S}(t)$ and that the random variables S_i^k that determine the value of s_i^k are all independent and identically distributed according to

$$P\left(S_i^k = 1\right) = p, \quad P\left(S_i^k = 0\right) = 1-p \tag{15}$$

Consider the averaged equation

$$\frac{d\boldsymbol{\xi}}{dt} = \boldsymbol{\Phi}(\boldsymbol{\xi}(t)) \tag{16}$$

where $\boldsymbol{\xi} \in \mathbb{R}^N, \boldsymbol{\Phi} : \mathbb{R}^N \to \mathbb{R}^N$ and

$$\begin{aligned}\boldsymbol{\Phi}(\boldsymbol{x}) &= \mathrm{E}\left(\boldsymbol{F}\left(\boldsymbol{x},\boldsymbol{S}^k\right)\right) \text{ where } \boldsymbol{S}^k = \left(S_1^k,\ldots,S_M^k\right) \\ &= \sum_{\boldsymbol{s}\in\{0,1\}^M} \boldsymbol{F}(\boldsymbol{x},\boldsymbol{s})\,P(\boldsymbol{s}) \\ &= \sum_{\boldsymbol{s}\in\{0,1\}^M} \boldsymbol{F}(\boldsymbol{x},\boldsymbol{s}) \prod_{i=1}^{M} (ps_i + (1-p)(1-s_i))\end{aligned} \tag{17}$$

Now, suppose $\bar{\boldsymbol{\xi}}$ is an asymptotically stable equilibrium point of (16). Furthermore, we suppose there is a Lyapunov function $W : \mathbb{R}^N \to \mathbb{R}$ with the following properties:

a) W is continuously differentiable
b) W has a local minimum in $\bar{\boldsymbol{\xi}}$, for convenience $W\left(\bar{\boldsymbol{\xi}}\right) = 0$
c) If $\boldsymbol{\xi}(t)$ is a solution of (16) then $\frac{d}{dt}W(\boldsymbol{\xi}(t)) \le 0$

In addition, suppose there are two constants $0 < V_0 < V$ and a constant $\gamma > 0$ such that

e) The level set $\{\boldsymbol{\xi} \,|\, W(\boldsymbol{\xi}) \le V\}$ is contained in C and thus compact.
f) The connected component U of $\{\boldsymbol{\xi} \,|\, W(\boldsymbol{\xi}) \le V\}$ that contains $\bar{\xi}$ is contained in the basin of attraction of $\bar{\boldsymbol{\xi}}$.
g) Within U, but outside of $U_0 = \{\boldsymbol{\xi} \,|\, \boldsymbol{\xi} \in U, W(\boldsymbol{\xi}) \le V_0\}$, we have along any solution $\boldsymbol{\xi}(t)$ of (16) the inequality
$$\frac{d}{dt}W(\boldsymbol{\xi}(t)) \le -\gamma. \tag{18}$$

The meaning of the last two conditions is to consider the connected part U of a level set of W that fills out as much as possible of the basin of attraction of $\bar{\boldsymbol{\xi}}$, but that has still some distance from the basin boundary where an unstable equilibrium point of could sit and therefore the derivative of W could become 0 (Fig. 2). Within U, the connected part U_0 of a level set of W should be chosen as small as possible around $\bar{\boldsymbol{\xi}}$. Excluding U_0 from U allows guaranteeing that W decreases at a speed that cannot become smaller than a certain positive constant (γ).

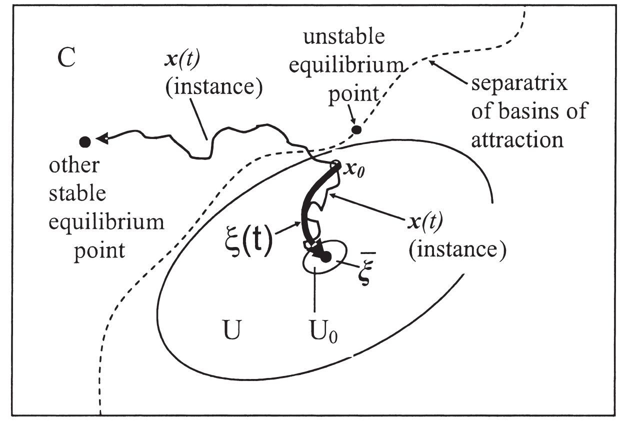


Fig. 2 Schematic representation of the state space with the compact regions C, U and U_0, the solution of the averaged system (bold) and two instances of the solution of the blinking system. One of them reaches U_0, whereas the other leaves U and approaches another asymptotically stable equilibrium point.

Theorem 1: Consider system (12) with the above listed properties and fix an initial condition $\boldsymbol{x}_0$ in U. The probability P_f that the corresponding solution of the blinking system reaches the neighborhood U_0 of the equilibrium point $\bar{\boldsymbol{\xi}}$ of the averaged equation in finite time converges to 1 as the switching period converges to zero.

Proof

We consider a set of solutions $\boldsymbol{x}(t)$ of the blinking system (12), all starting from the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0 \in U$. The different solutions correspond to different switching sequences. Furthermore, we consider the unique solution $\boldsymbol{\xi}(t)$ of the averaged system (16) starting from the same initial condition $\boldsymbol{\xi}(0) = \boldsymbol{x}(0) = \boldsymbol{x}_0$. Then, by (18),

$$W(\boldsymbol{\xi}(t)) \le W(\boldsymbol{\xi}(0)) - \gamma t \tag{19}$$

Thus, the solution $\boldsymbol{x}(t)$ of (16) remains in U and reaches U_0 at the latest at time

$$T = \frac{V - V_0}{\gamma} \tag{20}$$

We want to show that with high probability the solution $\boldsymbol{x}(t)$ of the blinking model also remains in U and reaches U_0 in finite time.

Since W is continuously differentiable and C is compact, it has a Lipschitz constant L_W in C and thus

$$|W(\boldsymbol{x}(t)) - W(\boldsymbol{\xi}(t))| \le L_W \|\boldsymbol{x}(t) - \boldsymbol{\xi}(t)\| \tag{21}$$

On the other hand

$$\|\boldsymbol{x}(t)-\boldsymbol{\xi}(t)\| \le \left\| \int_0^t \boldsymbol{F}(\boldsymbol{x}(\rho),\boldsymbol{s}(\rho))\, d\rho - \int_0^t \boldsymbol{\Phi}(\boldsymbol{\xi}(\rho))\, d\rho \right\|$$
$$\le \left\| \int_0^t [\boldsymbol{F}(\boldsymbol{x}(\rho),\boldsymbol{s}(\rho)) - \boldsymbol{F}(\boldsymbol{x}(0),\boldsymbol{s}(\rho))]\, d\rho \right\| + \left\| \int_0^t [\boldsymbol{F}(\boldsymbol{x}(\rho),\boldsymbol{s}(\rho)) - \boldsymbol{\Phi}(\boldsymbol{x}(0))]\, d\rho \right\| + \left\| \int_0^t [\boldsymbol{\Phi}(\boldsymbol{\xi}(\rho)) - \boldsymbol{\Phi}(\boldsymbol{\xi}(0))]\, d\rho \right\| \tag{22}$$

Using (13) and (14) we find the following bound for the first term on the RHS of (22):

$$\left\| \int_0^t [\boldsymbol{F}(\boldsymbol{x}(\rho),\boldsymbol{s}(\rho)) - \boldsymbol{F}(\boldsymbol{x}(0),\boldsymbol{s}(\rho))]\, d\rho \right\| \le L_F \int_0^t \|\boldsymbol{x}(\rho)-\boldsymbol{x}(0)\|\, d\rho = L_F \int_0^t \left\| \int_0^\rho \boldsymbol{F}(\boldsymbol{x}(\Theta),\boldsymbol{s}(\Theta))\, d\Theta \right\| d\rho \le L_F B_F \int_0^t \int_0^\rho d\Theta d\rho = L_F B_F \frac{t^2}{2} \tag{23}$$

Since $\boldsymbol{\Phi}$ is a mean value of $\boldsymbol{F}$, its bound and its Lipschitz constant on C is not larger than those of $\boldsymbol{F}$, and the third term has the same bound as the first:

$$\left\| \int_0^t [\boldsymbol{\Phi}(\boldsymbol{\xi}(\rho)) - \boldsymbol{\Phi}(\boldsymbol{\xi}(0))]\, d\rho \right\| \le L_F B_F \frac{t^2}{2} \tag{24}$$

For the second term, we suppose that t is a multiple of τ:

$$t = K\tau \tag{25}$$

Then,

$$\left\| \int_0^t [\boldsymbol{F}(\boldsymbol{x}(0),\boldsymbol{s}(\rho)) - \boldsymbol{\Phi}(\boldsymbol{x}(0))]\, d\rho \right\| = \left\| \sum_{k=1}^K \tau \boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{s}^k\right) - t\boldsymbol{\Phi}(\boldsymbol{x}(0)) \right\| \tag{26}$$

and the probability that this term grows linearly with K can be made arbitrarily small for large K. Indeed,

$$P\left[\left\| \int_0^t [\boldsymbol{F}(\boldsymbol{x}(0),\boldsymbol{s}(\rho)) - \Phi(\boldsymbol{x}(0))]\, d\rho \right\| > \lambda t\right] = P\left[\left\| \sum_{k=1}^K \tau \boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - tE\left[\boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\| > \lambda t\right] = P\left[\left\| \sum_{k=1}^K \boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - KE\left[\boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\| > \lambda K\right] \tag{27}$$

The probability $P(\lambda, K)$ can be bounded by the Chebychev inequality [6] as follows:

$$\begin{aligned} P(\lambda,K) &= P\left[\left\| \sum_{k=1}^K \boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - K\,\mathrm{E}\left[\boldsymbol{F}\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\|^2 > \lambda^2 K^2\right] \\ &\le P\left[\sum_{i=1}^N \left\| \sum_{k=1}^K F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - K\,\mathrm{E}\left[F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\|^2 > \lambda^2 K^2\right] \\ &\le \sum_{i=1}^N P\left[\left\| \sum_{k=1}^K F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - K\,\mathrm{E}\left[F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\|^2 > \lambda^2 K^2\right] \\ &\le \frac{\sum_{i=1}^N \mathrm{Var}\left[\sum_{k=1}^K F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right]}{\lambda^2 K^2} \\ &= \frac{\sum_{i=1}^N \sum_{k=1}^K \mathrm{Var}\left[F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right]}{\lambda^2 K^2} \\ &= \frac{N}{\lambda^2 K} \sum_{i=1}^N \sum_{s\in\{0,1\}^M} \left[F_i(\boldsymbol{x}(0),\boldsymbol{s}) - \left[\sum_{s\in\{0,1\}^M} F_i(\boldsymbol{x}(0),\boldsymbol{s})P(\boldsymbol{s})\right]\right]^2 P(\boldsymbol{s}) \end{aligned} \tag{28}$$

The inequality in the brackets can only hold if at least one term exceeds $\frac{\lambda^2 K^2}{N}$. Thus

$$\begin{aligned} P(\lambda,K) &\le \sum_{i=1}^N P\left[\left\| \sum_{k=1}^K F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right) - K\,\mathrm{E}\left[F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right] \right\|^2 > \frac{\lambda^2 K^2}{N}\right] \\ &\le \frac{\sum_{i=1}^N \mathrm{Var}\left[\sum_{k=1}^K F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right]}{\lambda^2 K^2} \\ &= \frac{N \sum_{i=1}^N \sum_{k=1}^K \mathrm{Var}\left[F_i\left(\boldsymbol{x}(0),\boldsymbol{S}^k\right)\right]}{\lambda^2 K^2} \\ &= \frac{N}{\lambda^2 K} \sum_{i=1}^N \sum_{s\in\{0,1\}^M} \left[F_i(\boldsymbol{x}(0),\boldsymbol{s}) - \left[\sum_{s\in\{0,1\}^M} F_i(\boldsymbol{x}(0),\boldsymbol{s})P(\boldsymbol{s})\right]\right]^2 P(\boldsymbol{s}) \end{aligned} \tag{29}$$

Actually, much better bounds on this probability can be derived for particular forms of $\boldsymbol{F}$. We now conclude that with probability at least $1 - P(\lambda, K)$ we have the bound

$$\begin{aligned} W(\boldsymbol{x}(t)) &\leq W(\boldsymbol{\xi}(t)) + |W(\boldsymbol{x}(t)) - W(\boldsymbol{\xi}(t))| \\ &\leq W(\boldsymbol{\xi}(0)) - \gamma t + L_W\left[L_F B_F t^2 + \lambda t\right] \\ &= W(\boldsymbol{x}(0)) - \gamma t + L_W L_F B_F t^2 + L_W \lambda t \end{aligned} \tag{30}$$

Choosing

$$\lambda = \frac{\gamma}{4L_W}, \quad t = \frac{\gamma}{4L_W L_F B_F} \tag{31}$$

we obtain

$$W(\boldsymbol{x}(t)) \leq W(\boldsymbol{x}(0)) - \frac{\gamma}{2}t \tag{32}$$

Strictly speaking, we would have to choose t as a multiple of τ, but since the theorem makes a statement for small τ, this would change t only by a small amount and therefore, we disregard this detail in the sequel of the proof.

For the binary signals which do not satisfy (30), we can give the bound

$$\begin{aligned} W(\boldsymbol{x}(t)) &\leq W(\boldsymbol{x}(0)) + \int_0^t \frac{\partial W}{\partial \boldsymbol{x}}(\boldsymbol{x}(\rho), \boldsymbol{s}(\rho))\frac{d\boldsymbol{x}}{dt}d\rho \\ &\leq W(\boldsymbol{x}(0)) + L_W B_F t \end{aligned} \tag{33}$$

By time invariance, we can write (32) and (33) replacing $t = 0$ by t and t by $t + \Delta t$, where now Δt is given by (31).

In summary, starting from a state $\boldsymbol{x}(t) \in U \backslash U_0$ after the lapse of time $\Delta t = \frac{\gamma}{4L_W L_F B_F}$ the Lyapunov function W has decreased at least the amount $\frac{\gamma}{2}\Delta t$ with probability at least $1 - P(\lambda, K)$ and it has increased by at most $L_W B_F \Delta t$ with probability at most $P(\lambda, K)$.

As proved in the appendix, it follows from this that the probability P_f that the solution reaches U_0 in finite time is bounded by

$$P_f = 1 - P_\infty > 1 - \left(2\frac{P(\lambda, K)}{1 - P(\lambda, K)}\right)^{\frac{1}{J}} \tag{34}$$

where J is given by (A· 1) and λ by (31). Since by (28) $P(\lambda, K)$ converges to zero when K goes to infinity and since our choices of λ, J and Δt are independent of τ, whereas $K = \Delta t/\tau$, P_f converges to 1 as τ goes to zero. This concludes the proof of Theorem 1.

Remark:

Actually, inequality (28) combined with Theorem 2 of the appendix would allow us to give an explicit lower bound on P_f. However, this bound can be improved substantially for specific forms of $\boldsymbol{F}$, as we will show in the example. For this reason, we do not pursue the general bound any further.

6. Back to the Example

The various conditions can be verified for our example system, with a sigmoid nonlinearity. In Fig. 3 we show the trajectory for the blinking system (with piecewise nonlinearity), starting from the same initial conditions as in (4). Indeed, the blinking network performs the WTA function correctly for this switching sequence. According to the theorem, only a very small portion of the switching sequences could lead to a misidentification of the largest initial value.

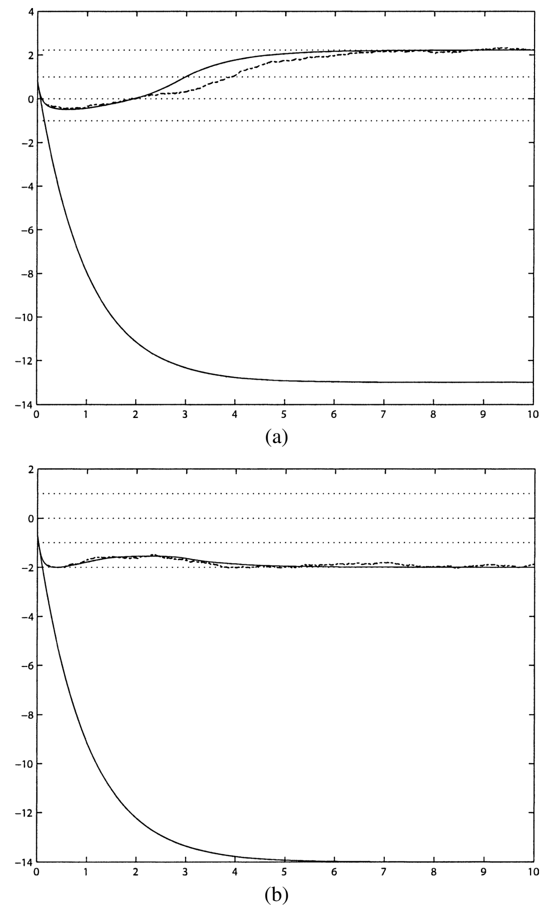


Fig. 3 Same figure as Fig. 1, but in addition trajectory of the blinking network with a switching time $\tau = 0.0001$ and $p = 0.1$.

We now derive an explicit upper bound on the probability of misclassification. For this purpose, we have to give an upper bound for $P(\lambda, K)$ as defined in (27). The state equations of the blinking WTA system are

$$\frac{dx_i}{dt} = -x_i + (1+\delta)y_i - \alpha \sum_{\substack{j=1 \\ j \text{ nn of } i}}^{n^2} y_j - \frac{\alpha}{p}\sum_{\substack{j=1 \\ j \text{ not nn of } i}}^{n^2} s_{ij}(t)y_j + \kappa \tag{35}$$

Then

$$P(\lambda, K) = P\left[\left\|\sum_{k=1}^{K}\boldsymbol{F}\left(\boldsymbol{x}(0), \boldsymbol{S}^k\right) - K\,\mathrm{E}\left[\boldsymbol{F}\left(\boldsymbol{x}(0), \boldsymbol{S}^k\right)\right]\right\|^2 > \lambda^2 K^2\right]$$

$$= P\left[\frac{\alpha^2}{p^2}\sum_{i=1}^{n^2}\left(\sum_{k=1}^{K}\sum_{\substack{j=1\\ j \text{ not nn of } i}}^{n^2}\left(S_{ij}^k - p\right)y_j\right)^2 > \lambda^2 K^2\right]$$

$$= P\left[\sum_{i=1}^{n^2}\left(\sum_{k=1}^{K}\sum_{\substack{j=1\\ j \text{ not nn of } i}}^{n^2}\left(S_{ij}^k - p\right)y_j\right)^2 > \frac{\lambda^2 K^2 p^2}{\alpha^2}\right] \quad (36)$$

In order for the inequality in the brackets to hold, at least one of the terms indexed by i must be larger than $\frac{\lambda^2 K^2 p^2}{n^2\alpha^2}$. Therefore, and since $|y_j| \le 1$,

$$P(\lambda, K) \le \sum_{i=1}^{n^2} P\left[\left(\sum_{k=1}^{K}\sum_{\substack{j=1\\ j \text{ not nn of } i}}^{n^2}\left(S_{ij}^k - p\right)y_j\right)^2 > \frac{\lambda^2 K^2 p^2}{n^2\alpha^2}\right]$$

$$= \sum_{i=1}^{n^2} P\left[\left|\sum_{\substack{j=1\\ j \text{ not nn of } i}}^{n^2} y_j \sum_{k=1}^{K}\left(S_{ij}^k - p\right)\right| > \frac{\lambda K p}{n\alpha}\right]$$

$$\le \sum_{i=1}^{n^2} P\left[\sum_{\substack{j=1\\ j \text{ not nn of } i}}^{n^2}\left|\sum_{k=1}^{K}\left(S_{ij}^k - p\right)\right| > \frac{\lambda K p}{n\alpha}\right] \quad (37)$$

Again reasoning that at least one term in the sum over j must be larger than $\frac{\lambda K p}{n^2 n\alpha}$ we obtain

$$P(\lambda, K) \le \sum_{i,j=1}^{n^2} P\left[\left|\sum_{k=1}^{K}\left(S_{ij}^k - p\right)\right| > \frac{\lambda K p}{n^3\alpha}\right] \quad (38)$$

Applying now the Chernoff bounds for binary random variables [6], [7]:

$$P\left[\sum_{k=1}^{K} S_{ij}^k - p \ge (1+\delta) K p\right] \le e^{-\frac{\delta^2 p K}{3}}$$

$$P\left[\sum_{k=1}^{K} S_{ij}^k - p \le (1-\delta) K p\right] \le e^{-\frac{\delta^2 p K}{2}} \quad (39)$$

we obtain

$$P(\lambda, K) \le 2n^4 e^{-\frac{\lambda^2}{3n^6\alpha^2}K} \quad (40)$$

Applying now Theorem 2 of the appendix leads, for sufficiently small τ to the following bound for the probability that the blinking trajectory leaves the domain U and thus potentially leads to misclassification:

$$P_\infty \le 3n^4 e^{-\frac{\lambda^2 \Delta t}{3Jn^6\alpha^2}\frac{1}{\tau}} \quad (41)$$

where J is given by (A· 1) and $\Delta t = \frac{\gamma}{4L_W L_F B_F}$.

This inequality shows that once the switching time τ is small enough, then the probability that the blinking trajectory does not follow the averaged system trajectory goes exponentially fast to zero. Thus, there is a kind of threshold for the switching time. Of course, this is not the whole probability of misclassification, because we limited the analysis from the outset to a region that is smaller than the basin of attraction of the equilibrium point that represents the correct classification. Typically, we neglect initial state vectors with a component that is very close to the maximum. It is quite intuitive that in such a case, the blinking system is likely to lead to misclassification. However, choosing the parameter γ closer to zero allows getting closer to the border of the basin of attraction. The price to pay is a lower threshold for τ.

7. Conclusions

We have shown that in a network of interacting dynamical systems, long-range interactions can be realized by blinking connections, i.e. connections that are only switched on from time to time, in a random fashion. If the switching is fast enough, the behavior of the blinking system is close to the behavior of the averaged system, where the long range connections are fixed, but with an interaction strength that is the mean strength of the same connections in the blinking system. The blinking connections can be realized by sending packets on a communication network underlying the network of dynamical systems. Such a communication network is in any case needed to bring the information to and from the dynamical systems.

A rigorous analysis of the blinking system is carried out for the case where the averaged system is multi-stable. It is shown that for a large part of the basin of attraction of an asymptotically stable equilibrium point of the averaged system, the solutions of the blinking system approach the same equilibrium point with high probability. In fact, this probability converges to one when the switching period goes to zero. Rigorous lower bounds for this probability can be given. As an illustrative example, a winner-take-all neural network is used. For this example, we show that for large enough switching frequencies the probability converges to one exponentially fast as a function of the switching frequency. This causes a sort of a threshold for the switching frequency above which the trajectories of the blinking system follow those of the averaged system almost with probability one. Of course, whether or not there really is a threshold cannot be decided here, because we only have given bounds on the probabilities.

Acknowledgments

This work has been financially supported by the grant nr. 2100-065268 of the Swiss National Science Foundation.

References

[1] L.O. Chua and T. Roska, Cellular Neural Networks and Visual Computing, Cambridge University Press, Cambridge, UK, 2002.

[2] G. Timar and D. Balya, "Regular small-world cellular neural networks: Key properties and experiments," Proc. IEEE International

Symposium on Circuits and Systems (ISCAS'04), vol.III, pp.369–372, Vancouver, Canada, May 2004.

[3] N.M. Bogoliubov and Y.A. Mitropolsky, Asymptotic Methods in the Theory of Non-Linear Oscillators, Hindustan Publishing, 1961.

[4] I. Belykh, V. Belykh, and M. Hasler, "Blinking model and synchronization in small-world networks with a time-varying coupling," Physica D, vol.195, no.1-2, pp.188–206, 2004.

[5] G. Seiler and J.A. Nossek, "Winner-take-all cellular neural networks," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol.40, no.3, pp.184–190, 1993.

[6] P. Brémaud, Markov Chains: Gibbs Fields, Monte Carlo Simulations and Queues, 2nd ed., Springer Verlag, New York, 2001.

[7] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," Ann. Math. Statist., vol.23, pp.493–507, 1952.

[8] T. Hagerup and C. Rub, "A guided tour of Chernoff bounds," Inf. Proc. Lett., vol.33, p.305, 1990.

Appendix

Theorem 2: Let

$$r = \frac{P(\lambda, K)}{1 - P(\lambda, K)}, \quad J = \left[\frac{2L_W B_F}{\gamma}\right] + 1 \qquad \text{(A·1)}$$

where [.] denotes the integer part and

$$\lambda = \frac{\gamma}{4L_W}, \quad K = \frac{\Delta t}{\tau} = \frac{\gamma}{4L_W L_F B_F \tau} \qquad \text{(A·2)}$$

If $r \leq 2^{-(J+1)}$ then the probability that a trajectory of the blinking system starting in U does not reach U_0 in finite time is bounded by $(2r)^{\frac{1}{J}}$.

Proof

Consider a trajectory of the blinking system starting at a point $\boldsymbol{x}_0 \in U$. More precisely, consider the set of trajectories starting at $\boldsymbol{x}_0$ generated by all possible binary signals $\boldsymbol{s}(t)$. Let P_∞ be the probability of the set of binary signals that generate trajectories that do not reach U_0 in finite time. We now define an auxiliary Markov chain which will allow deriving a bound for P_∞. The state space of the Markov chain is and the transition probabilities p_{ij} from state i to state j are

$$\begin{aligned} &p_{i\,i+N} = P(\lambda, K) \\ &p_{i\,i-1} = 1 - P(\lambda, K) \\ &p_{i\,j} = 0 \quad \text{for all other } i, j \end{aligned} \qquad \text{(A·3)}$$

A sample path of length Q starting at zero of the Markov chain is a function $z : \{0, \cdots, L\} \to \mathbb{Z}$ with

$$\begin{aligned} &z(0) = 0 \\ &z(q+1) - z(q) = J \text{ or } -1, \\ &\quad \text{for } q = 0, \ldots, Q-1 \end{aligned} \qquad \text{(A·4)}$$

Here, Q may also be infinite. To each trajectory of the blinking system, observed at the discrete time instants $t = q\Delta t = qK\tau, q = 0, 1, 2, \ldots$ we associate a sample path of the Markov chain of length Q that satisfies for $q = 0, 1, 2, \ldots, Q-1$

$$\left\| \sum_{k=1}^{K} \boldsymbol{F}\left(\boldsymbol{x}(q\Delta t), \boldsymbol{S}^k\right) - KE\left[\boldsymbol{F}\left(\boldsymbol{x}(q\Delta t), \boldsymbol{S}^k\right)\right] \right\| > \lambda K$$
$$\Rightarrow z(q+1) = z(q) + J$$
$$\left\| \sum_{k=1}^{K} \boldsymbol{F}\left(\boldsymbol{x}(q\Delta t), \boldsymbol{S}^k\right) - KE\left[\boldsymbol{F}\left(\boldsymbol{x}(q\Delta t), \boldsymbol{S}^k\right)\right] \right\| \leq \lambda K$$
$$\Rightarrow z(q+1) = z(q) - 1 \qquad \text{(A·5)}$$

where Q is such that

$$\begin{aligned} &\boldsymbol{x}(q\Delta t) \in U \backslash U_0 \text{ for } q = 0, 1, \ldots, Q-1 \\ &\quad \text{and, for finite } Q,\ \boldsymbol{x}(Q\Delta t) \notin U \backslash U_0 \end{aligned} \qquad \text{(A·6)}$$

Thus, at time $Q\Delta t$, the trajectory is either outside of U or inside U_0. If Q is infinite, the trajectory remains forever in $U \backslash U_0$. Note that since in general the sample path is only related to the trajectory of the blinking system for a finite time interval, there are infinitely many trajectories with the same associated sample path.

Because of the definition of the transition probabilities (A·3) of the auxiliary Markov chain, the set of all trajectories with a given associated sample path has the same probability in the blinking system as this sample path in the Markov chain. Furthermore, let $\boldsymbol{x}(t)$ be a trajectory of the blinking system and let $z(q), q = 1, \ldots, Q$ be its associated sample path in the Markov chain. Then

$$\begin{aligned} &W(\boldsymbol{x}(q\Delta t)) \leq W(\boldsymbol{x}(0)) + z(q)\frac{\gamma}{2}\Delta t \\ &\quad \text{for } q = 0, 1, \ldots, Q \end{aligned} \qquad \text{(A·7)}$$

This can be seen by induction. It clearly holds for $q = 0$. Suppose it holds for q. If $z(q+1) = z(q) + J$, then by (33) and time invariance we obtain

$$\begin{aligned} W(\boldsymbol{x}((q+1)\Delta t)) \ leq\ & W(\boldsymbol{x}(q\Delta t)) + L_W B_F \Delta t \\ \leq\ & W(\boldsymbol{x}(0)) + z(q)\frac{\gamma}{2}\Delta t + J\frac{\gamma}{2}\Delta t \\ =\ & W(\boldsymbol{x}(0)) + z(q+1)\frac{\gamma}{2}\Delta t \end{aligned} \qquad \text{(A·8)}$$

whereas if $z(q+1) = z(q) - 1$, by (32) and time invariance

$$\begin{aligned} W(\boldsymbol{x}((q+1)\Delta t)) \leq\ & W(\boldsymbol{x}(q\Delta t)) - \frac{\gamma}{2}\Delta t \\ \leq\ & W(\boldsymbol{x}(0)) + z(q)\frac{\gamma}{2}\Delta t - \frac{\gamma}{2}\Delta t \\ =\ & W(\boldsymbol{x}(0)) + z(q+1)\frac{\gamma}{2}\Delta t \end{aligned} \qquad \text{(A·9)}$$

This implies

$$\begin{aligned} &\boldsymbol{x}(Q\Delta t) \notin U \Rightarrow z(Q) > 0 \\ &z(Q) \leq 0 \Rightarrow \boldsymbol{x}(Q\Delta t) \in U_0 \\ &Q = \infty \Rightarrow \boldsymbol{x}(q\Delta t) \in U \backslash U_0 \\ &\quad \text{for all } q \geq 0 \end{aligned} \qquad \text{(A·10)}$$

In addition, since W is non-negative on U, we have

$$\boldsymbol{x}(q\Delta t) \in U \Rightarrow z(q) \geq -\frac{2\gamma W(\boldsymbol{x}(0))}{\Delta t} \quad \text{(A· 11)}$$

Thus, if Q is infinite, either the sample path is confined to the finite number of states between zero and the lower bound (A· 11), or it eventually reaches a positive state. It is not difficult to see that the set of all infinite length sample paths in the Markov chain that take values only in a finite set has probability 0. Therefore, if Q is infinite, then almost surely the sample path takes also positive values. This, together with (A· 10) allows us to conclude that the probability that a trajectory of the blinking system does not reach U_0 in finite time is bounded by the probability that a sample path reaches a positive state in the Markov chain, i.e.

$$P_\infty \leq P\{z | z(q) > 0 \text{ for some } q \geq 0\} \quad \text{(A· 12)}$$

We now analyze this probability in the Markov chain. Consider sample paths that start from any state. Define

$$h(i) = P\{z | z(0) = i, z(q) > 0 \text{ for some } q \geq 0\} \quad \text{(A· 13)}$$

Thus, (A· 12) becomes

$$P_\infty \leq h(0) \quad \text{(A· 14)}$$

and h satisfies

$$h(i) = P(\lambda, K)\, h(i+J) + (1 - P(\lambda, K))\, h(i-1) \quad \text{(A· 15)}$$

$$h \text{ is increasing}, \; h(i) = 1 \text{ for } i > 0, \quad \lim_{i \to -\infty} h(i) = 0 \quad \text{(A· 16)}$$

In fact h has the properties of absorption probabilities [6]. Define $d(i)$ by

$$d(i) = h(i) - h(i-1) \quad \text{(A· 17)}$$

and r by (A· 1). Then all $d(i)$ are non-negative and

$$\begin{aligned} &d(i) = r(d(i+1) + \cdots + d(i+J)) \\ &d(i) = 0 \quad \text{for} \quad i \geq 1, \\ &d(1) = 1 - h(1), \\ &\sum_{i=0}^{\infty} d(i) = h(0) \end{aligned} \quad \text{(A· 18)}$$

We are now looking for a bound on $d(i)$ of the form

$$d(i) \leq C\alpha^i \quad \alpha > 1 \quad \text{(A· 19)}$$

Suppose (A· 19) holds for $i = j+1, j+2, \ldots, j+J$. Then

$$\begin{aligned} d(j) &\leq Cr\left(\alpha^{j+1} + \cdots + \alpha^{j+J}\right) \\ &= Cr\alpha^{j+1}\frac{\alpha^J - 1}{\alpha - 1} \end{aligned} \quad \text{(A· 20)}$$

Therefore, if

$$r\alpha\frac{\alpha^J - 1}{\alpha - 1} \leq 1 \quad \text{(A· 21)}$$

then (A· 19) holds for all $i \leq j$ provided it holds for $i = j+1, j+2, \ldots, j+J$. Set

$$\alpha = \left(\frac{1}{2r}\right)^{\frac{1}{J}} \quad \text{(A· 22)}$$

Then because of $r \leq 2^{-(J+1)}$ we have $\alpha \geq 2$, $\frac{\alpha}{\alpha-1} \leq 1$ and (A· 21) is satisfied. Now, according to (A· 18), we start with $d(2) = \ldots = d(J) = 0$ and $d(1)$ arbitrary. Therefore, (A· 19) is satisfied for $i = 1, \ldots, J$ if $d(1) = C\alpha$. Then, again according to (A· 18)

$$h(0) \leq \frac{C}{1-\alpha^{-1}} = \frac{d(1)}{\alpha-1} = \frac{1-h(0)}{\alpha-1} \quad \text{(A· 23)}$$

which leads finally to

$$P_\infty \leq h(0) \leq \frac{1}{\alpha} = (2r)^{\frac{1}{J}} \quad \text{(A· 24)}$$

Martin Hasler received the Diploma in 1969 and the PhD degree in 1973 from the Swiss Federal Institute of Technology, Zurich, both in physics. He continued research in mathematical physics at Bedford College, University of London, from 1973 to 1974. At the end of 1974 he joined the Circuits and Systems group of the Swiss Federal Institute of Technology Lausanne (EPFL), where he is at present full professor. In 2002, he was acting Dean of the newly created School of Computer and Communication Sciences of EPFL. His current research interests are centered in nonlinear dynamics and information processing, both in engineering and in biological systems. In particular, he is interested in the engineering applications of complicated nonlinear dynamics, especially chaos. This also includes the modeling and identification of nonlinear circuits and systems. Very recently, he is interested in the qualitative behavior and modeling of networks composed of nonlinear systems, be they biological (brain) or technical (computer networks). He is a Fellow of the IEEE. He is currently Vice-President for Technical Activities of the IEEE CAS Society. He is a member of the Scientific Council of the Swiss National Science Foundation since Jan. 2000.

Igor Belykh received the Diploma in 1995 and the Ph.D. degree in 1999 from University of Nizhny Novgorod, both in physics and mathematics. He is currently a postdoctoral fellow in the Laboratory of Nonlinear Systems and a lecturer in the School of Computer and Communication Sciences of Swiss Federal Institute of Technology Lausanne (EPFL). From January 2006, he starts an Assistant Professor position in the Department of Mathematics and Statistics at Georgia State University, Atlanta, USA. His present interests include dynamical systems and ordinary differential equations, theory of bifurcations and chaos, coupled oscillators and synchronization, and neural networks.