

# Synchronous Chaos in Highdimensional Modular Neural Networks \*

Thomas Wennekers  
Abteilung Neuroinformatik  
Universität Ulm, D-89069 Ulm  
thomas@neuro.informatik.uni-ulm.de

Frank Pasemann  
Institut für Biologische Informationsverarbeitung  
Forschungszentrum Jülich, D-52425 Jülich  
f.pasemann@kfa-juelich.de

## Abstract

The relationship between certain types of highdimensional neural networks and low-dimensional prototypical equations (neuromodules) is investigated. The high-dimensional systems consist of finitely many pools containing identical, dissipative and nonlinear single-units operating in discrete time. Under the assumption of random connections inside and between pools, the system can be reduced to a set of only a few equations, which - asymptotically in time and system size - describe the behavior of every single unit arbitrarily well. This result can be viewed as synchronization of the single units in each pool. It is stated as a theorem on systems of nonlinear coupled maps, which gives explicit conditions on the single unit dynamics and the nature of the random connections. As an application we compare a 2-pool network with the corresponding 2-dimensional dynamics. The bifurcation diagrams of both systems become very similar even for moderate system size ( $N=50$ ) and large disorder in the connection strengths (50% of mean), despite the fact, that the systems exhibit fairly complex behavior (quasiperiodicity, chaos, coexisting attractors).

---

\*in: *International Journal of Bifurcation and Chaos*, **6**, 2055 -2067, 1996.

# 1 Introduction

In recent years research interests have focussed more strongly on the dynamical properties of information processing in the brain. Although dynamical aspects of brain activities have been discussed in earlier theoretical papers [Harth *et al.*, 1970; Wilson & Cowan, 1972; an der Heiden, 1980; Guevara *et al.*, 1983; Harth, 1983; Ermentrout, 1984], the new impulse came from the observation of complex dynamics in the olfactory bulb [Skarda & Freeman, 1987], and of coherent oscillations of neuron groups in response to specific visual stimuli [Eckhorn *et al.*, 1988; Gray *et al.*, 1989; Eckhorn *et al.*, 1993]. More abstract theoretical arguments why dynamical aspects should be relevant for neural information processing are built on the fact that brains have a highly recurrent structure, on the level of small functional groups of neurons as well as on the level of larger functionally segregated areas. Since it has to be expected that in these feedback loops excitatory as well as inhibitory connections are involved, the appearance of complex dynamical behavior is in general unavoidable. This was suggested, for example, by Palm, who considered networks of simple threshold units and mixed excitatory and inhibitory connections [Palm 1982; Palm 1987; Harth *et al.*, 1970]. By means of biologically plausible parameter values he computed logistic map-like unimodal iteration curves for the probability of neurons to fire. The firing threshold acts as the control parameter in this case. Other examples of chaos in artificial neural networks can be found in the references [Cessac *et al.*, 1994; Doyon *et al.*, 1993; Sompolinski *et al.*, 1988; Pasemann & Nelle, 1993; Pasemann, 1995]. Thus complex behavior seems to be an intrinsic property of such networks.

If it is inevitable, one may ask whether or not complex dynamical behavior contributes to the high-level information capabilities of brains. Chaotic behavior in fact seems to be a feature of some processes generated in different areas of the brain. It has been observed recently in various physiological and psycho-physical studies [Duke & Pritchard, 1991; Elbert *et al.*, 1994; Richards *et al.*, 1994], but its contribution to higher-order brain processing is still being discussed [Babloyantz & Lourenço, 1994; Skarda & Freeman, 1987; Hansel & Sompolinski, 1992; Tani & Fukumura, 1995].

Following the *modular dynamics approach* to cognitive systems as outlined in reference [Pasemann, 1995], a cognitive process is represented by the global dynamics of a neural system, which is the result of the cooperative dynamics of the neuromodules responding to a particular set of input stimuli and motor actions. From this point of view chaotic neuromodules are good candidates for basic elements of cognitive systems. They do not only provide chaotic dynamics for some distinguished parameter regimes (perhaps

used when functionally not active) but also a large variety of periodic and quasi-periodic dynamics for different parameter domains (addressed during interaction).

Let us also mention that complex behavior, although still seldomly considered in the context of information processing, does not necessarily contradict standard ideas about neural networks. The concept of associative memories, for example, can be extended to nonstationary operating modes [Wennekers *et al.*, 1995; Erb & Aertsen, 1992]. Such networks display several features in favour of the attractor network paradigm, i.e. pattern segmentation and (very) fast reaction to changing environments. Thus associative properties can be included in neuromodules with complex time behavior.

Here we study the relationship between certain classes of large neural networks and low-dimensional neuromodules as described earlier [Pasemann & Nelle, 1993, Pasemann, 1995]. We view this study as a test for the relevance of qualitative results derived for the dynamical properties of small neural networks (neuromodules) with respect to the behavior of large systems with biologically motivated architectures. In Sec. 2 we define these *modular neural networks* consisting of large populations of cells with identical properties (pools) that are mutually coupled by random connections. Under appropriate conditions on the local single unit dynamics and the probability distributions of the coupling strengths, the temporal behavior of the modular network can be described arbitrarily well by one equation per pool (asymptotically in time and system size). In other words the neurons in every pool of the large system synchronize and the whole network dynamics effectively reduces to that of a low dimensional neuromodule (in the above sense). A theorem stating rigorous conditions for synchronization is given in Sec. 3. The theorem is related to recent results about synchronized chaos and globally coupled maps [Pecora & Carroll, 1990, 1991; Kaneko, 1990; Pikovsky & Kurths, 1994a, 1994b]. Since this correspondence between modular neural networks and neuromodules is asymptotically exact, the rich behavior usually found in the lowdimensional systems [Pasemann & Nelle, 1993, Pasemann, 1995] *must* also be exhibited by modular neural nets containing many neurons per pool. Even stronger, *all* behavior that is asymptotically found in the large systems *must* necessarily be of the same type as that found in the corresponding lowdimensional systems. This is demonstrated numerically in Sec. 4 for models consisting of two pools of cells. The behavior of the equivalent low-dimensional system - called the *2-module* - appears fairly complex; it contains fixed points, limit cycles, chaos, coexisting attractors and generalized hysteresis between these attractors. The same complexity is in detail recovered in the highdimensional case. The bifurcational structure of both models becomes very similar even for moderate system size ( $N=50$

per pool) and widely scattering coupling strengths (50% standard deviation). We finally discuss the relevance of our assumption of random connectivity with respect to biological neural networks.

## 2 Neurodynamics

### 2.1 Single-pool dynamics

We choose the standard description for the discrete activation dynamics of neurons: The state  $x_i(t)$  of neuron  $i$  is given by

$$x_i(t+1) = \lambda \cdot x_i(t) + I(t) + I_i^{net}(t), \quad 0 \leq \lambda < 1. \quad (1)$$

The first term on the right side defines the local dynamics of our cells and represents a time-discrete low-pass reflecting the dissipative properties of real neurons [Kandel & Schwartz, 1981].  $I(t)$  is an external input to cell  $i$  and  $I_i^{net}(t)$  the recurrent input received from neurons of the network.

The activation state of a neuron does not directly influence other cells, but via a nonlinear transfer ('rate' or 'squashing') function  $f$  that, with respect to biological experiments [Kandel & Schwartz, 1981], often is chosen in the sigmoidal form of the logistic (fermi) function

$$f(x) = 1/(1 + \exp(-x)). \quad (2)$$

With the usual weighted sum the recurrent input to neuron  $i$  then reads

$$I_i^{net}(t) = \frac{1}{N} \sum_{j=0}^N w_{ij} f(x_j(t)) \quad (3)$$

In Eqs. (1) - (3) local properties of single units are given by  $\lambda, I$  and  $f$ , which for convenience are assumed to be identical for all cells. The global network structure on the other side is determined by the couplings  $w_{ij}$ . Note that the local dynamics of single units, i.e.  $x_i(t+1) = \lambda \cdot x_i(t)$ , is purely linear. Nonlinearities and thus the possibility for complex autonomous behavior only enters through the output functions (2) in the coupling term (3) of Eq. (1).

The factor  $\frac{1}{N}$  in Eq. (3) ensures, that solutions of (1), (2) and (3) remain finite in the limit  $N \rightarrow \infty$ . The often chosen scaling  $1/\sqrt{N}$  leads to meaningful results only when  $Ew_{ij} = 0$  (this contains the strongly diluted case, see below) [Amari *et al.*, 1977; Cessac *et al.*, 1994; Sompolinski *et al.*, 1988]. But a vanishing average synaptic strength need not be a valid assumption in biological networks, because real neurons are known to often express only

synapses of a fixed sign ('Dales principle' [Eccles, 1984]). Networks containing neurons of different type therefore should be modelled as two (or more) pools of cells, as it is done in the next section.

## 2.2 Modular neural networks

We define a *modular neural network* (MNN) as a composition of  $P$  mutually interconnected subnetworks (pools) of the type described in the previous section [Amari, 1971; Amari, 1972] for a similar approach). Neurons within one pool are assumed to be identical with respect to their local properties ( $\lambda, I$  and  $f$ ), but may differ between pools. With  $N_\alpha$  counting the number of neurons in pool  $\alpha$ ,  $\vec{F}(\vec{x}) = (f(x_1), \dots, f(x_N))^T$  and  $\mathbf{W} = (w_{ij})$  we write for pool  $\alpha$  using vector notation

$$\vec{x}^\alpha(t+1) = \lambda^\alpha \cdot \vec{x}^\alpha(t) + \vec{I}^\alpha(t) + \sum_{\beta=0}^P \frac{1}{N_\beta} \mathbf{W}^{\alpha\beta} \vec{F}^\beta(\vec{x}^\beta(t)). \quad (4)$$

In general solutions of this system of coupled maps will be arbitrarily complex. In particular, if the network consists of several pools each containing many individual neurons the system can only be treated numerically. To gain some analytical insight into the typical behavior of very large neural networks statistical methods have often been used. The goal then is to derive simpler equations describing the development of macroscopic quantities like average activities or correlations [Harth *et al.*, 1970; Wilson & Cowan, 1972; Sompolinski *et al.*, 1988; Amari, 1971; Amari, 1972; Amari *et al.*, 1977; Geman, 1982; Rozonoér, 1969]. Such methods usually rely on certain assumptions concerning parameters of the model. In most typical cases connections are assumed to be random. This is, what we also require below. We constrain the possible connection matrices and assume that all couplings are randomly generated. This means the entries in any matrix  $W^{\alpha\beta}$  are independent and identically distributed random variables subject to a common density  $p^{\alpha\beta}(w)$ . (Upper indices  $\alpha, \beta$  indicate that this density can be different for different pairs of pools.) In doing so, the focus of attention shifts from individual networks to the study of the typical behavior of a whole ensemble of networks. A specific net can then be viewed as a representative of this ensemble. Now, under the randomness assumption, the typical behavior of system (4) can be described (asymptotically exact) by a low-dimensional deterministic set of equations, that only contains expectation values of the coupling constants. The reduced system has dimension  $P$ , and thus is much easier to study, theoretically as well as numerically.

The theorem below proves an even stronger result: in almost all sample

networks not only the average activation (macroscopic state) but also (almost) all microscopic variables will follow the reduced dynamical equations. They do this so much the better the more cells are contained in the network. This can be interpreted as synchronization of the individual neurons in each pool.

### 3 Reduction of MNNs to Neuromodules

In this section explicit conditions are given for the reduction of system (4) to a lowdimensional set of equations.

Pioneering work on related time discrete systems has been done by Rozonoér [1969] and Amari [1971; 1972; Amari *et al.*, 1977]. Considering random networks of binary threshold units these works derived macroscopic laws for average network quantities from the underlying microscopic state transition equations. Rozonoér pointed out the relationship between such nets and classical high-dimensional complex systems. He first applied the classical methods in a systematic manner to random threshold networks with different topologies [Rozonoér, 1969]. Amari extended this work considerably. In particular he presented a mathematical validation of the approach [Amari *et al.*, 1977] and intensively analyzed the bifurcation structure of the resulting mean-field equations [Amari, 1971; 1972] without noticing the chaotic dynamics we observe in Sec. 4.

The theorem below explicitly excludes threshold neurons since we require continuous transfer functions. Thus the results of Rozonoér and Amari are complementary to ours: propositions analogous to B1-B3 in theorem 1 can be deduced as special cases of theorems in [Amari *et al.*, 1977].

Randomly coupled differential equations with continuous sigmoid output functions have also been studied by Amari [1972]. Here dynamical equations for average activities and correlations have been derived utilizing a gaussian approximation. These calculations *ad hoc* assume statistical independence of all microscopic variables and stochastic network parameters ('local chaos hypothesis'). A rigorous validation of this has later been given by Geman [1982] developing a method on which also this work is based. Theorem 1 essentially represents a time-discrete version of [Geman, 1982].

For ease of notation we only consider the case of a single pool in the sequel. Thus upper indices  $\alpha, \beta$  are omitted. The generalization to any finite number of pools is obvious.

**Theorem 1** Consider the system of nonlinear coupled maps

$$x_i(t+1) = \lambda \cdot x_i(t) + I(t) + \frac{1}{N} \sum_{j=1}^N w_{ij} f(x_j(t)), \quad i = 1 \cdots N \quad (5)$$

satisfying

**A0**  $x_i(0) = x_{i0} \in \mathbb{R}, 1 \leq i \leq N,$

**A1**  $|\lambda| < 1,$

**A2**  $I(t)$  is a bounded sequence in  $\mathbb{R},$

**A3**  $f$  is bounded and fulfils a uniform Lipschitz-condition,

**A4**  $w_{ij}, 1 \leq i, j \leq N$  are independent and identically distributed random variables with expectation value  $\overline{w_{11}} = w$  and characteristic function  $\overline{\exp(i\nu w_{11})}$  analytic in  $\nu = 0.$

Define

$$\Phi(x, t) : \mathbb{R} \times \mathbb{N} \rightarrow \mathbb{R}, \quad \Phi(x, t) = \lambda \cdot x + I(t) + w f(x)$$

and  $X^{(N)}(t) = 1/N \sum_{j=1}^N x_j(t).$  Then under conditions A0 - A4 the following holds

**B1**  $\lim_{N \rightarrow \infty} \overline{\lim_{t \rightarrow \infty}} \sup_{1 \leq i, j \leq N} |x_i(t) - x_j(t)| = 0 \quad a.s.$

**B2**  $\lim_{N \rightarrow \infty} \overline{\lim_{t \rightarrow \infty}} \sup_{1 \leq i \leq N} |x_i(t+1) - \Phi(x_i(t), t)| = 0 \quad a.s.$

**B3**  $\lim_{N \rightarrow \infty} \overline{\lim_{t \rightarrow \infty}} |X^{(N)}(t+1) - \Phi(X^{(N)}(t), t)| = 0 \quad a.s.$

A sketch of the proof is given in appendix A as a special case of a more general theorem including certain types of nonlinear local dynamics and retarded interactions which will be published elsewhere.

B1 means that all microscopic variables  $x_i$  asymptotically become identical; in other words - they synchronize. (For  $P > 1$  the cells in any pool synchronize, but not necessarily between pools). B2 states the microscopic dynamical law governed asymptotically by any individual cell. Note, that B2 only contains expectation values  $w$  but not individual synaptic strengths  $w_{ij}.$  Thus it is a *deterministic* equation valid for almost all sample networks of the underlying statistical ensemble of networks. The same holds for B3, which

describes the macroscopic state of a network, i.e. its average activation  $X$ . The particular form of this ‘mean-field’ equation trivially follows from the synchronization property of the microdynamics.

At this point a note on the weight scaling in the basic Eqs. (4) seems to be worthwhile. In (4) we have chosen the ‘Ansatz’:  $u_i(t) = 1/N \sum_{j=1}^N w_{ij} f(x_j(t))$  where the  $w_{ij}$  are of order 1 (that is independent of the system size  $N$ ). In this case under condition A4 we have a strong law of large numbers for  $z_i = 1/N \sum_j w_{ij}$  and - as the theorem shows - also for the  $u_i(t)$  at any (large) fixed time  $t$ .

More generally one could start with  $u'_i(t) = \sum_{j=1}^N w_{ij} f(x_j(t))$  where only  $E|w_{ij}| = O(1/N)$  is required [Amari *et al.*, 1977]. Here the probability distribution of the  $w_{ij}$  may depend on the system size in an a priori unrestricted manner. (This for example contains ‘strongly diluted’ models, where almost all but a finite number of synapses of a neuron are zero and the nonzero are of order 1. More examples can be found in [Amari *et al.*, 1977]. Then, in general we will have (at most) a central limit theorem for  $z'_i = \sum_j w_{ij}^{(N)}$  and also the  $x_i$  will no longer satisfy B1 and B2. Instead of converging to a constant for fixed time  $t$  they will obey a more complex limiting distribution. Nevertheless a mean-field equation comparable to B3 may still hold. This is what has been shown by Amari for threshold neurons [Amari *et al.*, 1977]. The more general case contains our scaling as a special case. Thus B1 and B2 can be recovered for threshold neurons too. The mathematical method in [Amari *et al.*, 1977] is different from our derivation. Continuous squashing functions and the  $O(1/N)$  scaling enable a direct proof of theorem 1 (see appendix). The more general ‘Ansatz’ requires a more complicated procedure: In [Amari *et al.*, 1977] it is first shown that the joint distribution function of any finite subset of microscopic variables  $x_i$  converges (in probability) to the distribution function of an appropriately defined stochastic (Markovian) network. Then mean-field equations are derived for the Markovian approximation.

Let us now relate Theorem 1 briefly to results on ‘synchronous chaos’ given by Pecora and Carroll [1990; 1991], which have shown that certain subsystems of chaotic systems are able to synchronize if they are driven by common signals. A necessary condition for this is the negativity of all Lyapunov exponents restricted to the subsystem along the synchronized solution (conditional Lyapunov exponents [Pecora & Carroll, 1991]). In our case the subsystems are linear and dissipative cells whereas the driving signal is the feedback from other cells. These driving signals asymptotically become the same for all neurons in a particular pool. Since damped linear subsystems always have negative conditional Lyapunov exponents, it follows that for our

systems, the synchronization condition of Pecora and Carroll is not only local and necessary, but also globally valid for arbitrary initial conditions.

Our work may also be seen in relation to globally coupled chaotic maps [Kaneko, 1990; Pikovski & Kurths, 1994a, 1994b]. These are characterized by  $P = 1$ ,  $w_{ij} = w \forall i, j$ , but local *chaotic* dynamics, for example logistic maps in a chaotic regime. In this case perfect synchronization is in general impossible and the global dynamics cannot be described by a simple equation analogous to B2 (compare [Pikovsky & Kurths, 1994b]).

In comparison with globally coupled chaotic maps the systems considered here exhibit a rather simple type of local dynamics. On the other side, we can have a finite number of pools with arbitrary complex feedback schemes. Thus if we observe complex dynamical behavior in a modular neural network, as for example chaos in the two-pool case of the next section, this behavior is necessarily a consequence of the *macroscopic* structure and not of the micro-dynamics. In other words - the subunits evolve synchronously and chaotically due to the coarse connectivity structure; when isolated they only display trivial fixed points. This may have important implications for the interpretation of brain signals generated by electrical activity of whole pools of cells like local field potentials, EEG, current-source densities used in electrode studies or MEG. In cases where our theorem applies, that is when some sort of averaging takes place in a local patch of cortical tissue, those signals might reflect classes of cells with different general properties and their gross circuitry. This is an often made physiological assumption, but none at all obvious (cf. discussion).

## 4 Numerical Comparison of a 2-Pool Network and a 2-Module

As an application of Theorem 1 we now consider a modular neural network of two pools, one of which containing  $N$  excitatory cells, the other  $N$  inhibitory. Each pool is connected to itself and the other. The output function of every neuron is the fermi function (2). We set  $\lambda = 0$  and choose constant external inputs. The full network equations then read:

$$\begin{aligned}\vec{x}^1(t+1) &= \frac{1}{N}\mathbf{W}^{11} \cdot \vec{\sigma}^1(t) + \frac{1}{N}\mathbf{W}^{12} \cdot \vec{\sigma}^2(t) + \vec{I}^1 \\ \vec{x}^2(t+1) &= \frac{1}{N}\mathbf{W}^{21} \cdot \vec{\sigma}^1(t) + \frac{1}{N}\mathbf{W}^{22} \cdot \vec{\sigma}^2(t) + \vec{I}^2\end{aligned}\tag{6}$$

The 2-module equations are obtained from system (6) by replacing vectors by scalars and the coupling matrices  $\frac{1}{N}\mathbf{W}^{\alpha\beta}$  by expectation values  $w^{\alpha\beta} = E[W_{11}^{\alpha\beta}]$ . The couplings in the highdimensional case are chosen according to a gaussian distribution with expectation values  $w^{\alpha\beta}$  and standard deviations  $\sigma_{\alpha\beta} = 0.5 \cdot w^{\alpha\beta}$ . If not stated otherwise parameter values are  $N = 50$ ,  $w^{11} = 8$ ,  $w^{12} = -12$ ,  $w^{21} = 12$ ,  $w^{22} = -8$ ,  $I^1 = 4$  und  $I^2 = 4$ .

With these values system 6 can be viewed as composed of two self-interacting subsystems, one excitatory, the other inhibitory. If uncoupled ( $w^{12} = w^{21} = 0$ ) the excitatory subsystem exhibits a global fixed point attractor, the inhibitory one a global period-2 orbit attractor [Pasemann, 1993]. The result of the mutual coupling is not at all obvious and indeed leads to a rich spectrum of dynamical behavior [Pasemann & Nelle, 1993].

The bifurcation structure of threshold networks similar to (6) has been studied for the first time by Amari [1971]. The mean-field equations [Amari, 1971] are essentially the same as in our case. Amari reports conditions for stable fixed points and bifurcations to limit cycles. Though he failed to observe more complex dynamical behavior.

Simulation results for different mutual coupling strengths are displayed in Fig. 1. Shown are sections of the  $(w^{21}, w^{12})$ -parameter subspaces of (A) the 2-module and (B) the 2-pool modular net. Note that even for  $N=50$  and 50% standard deviation in the coupling strengths the two plots display the same qualitative behavior, although a small 'scaling', especially of the inhibitory connection  $w^{12}$ , can be recognized.

We observe large domains corresponding to periodic attractors, marked in the figures by the respective periods. Furthermore there are two separated bands indicating attractors of higher periods and chaotic attractors. At the border of the fixed point domain and the horizontal band we observe bifurcations from stable fixed points to quasi-periodic attractors. Around the (black) chaotic band embedded in the period 3-domain a stable period-3 attractor coexists with all types of attractors appearing in the period doubling route to chaos (see also Fig. 3). For large coupling values both systems show attractors of period three.

Figure 2 exhibits the activation dynamics in more detail. In Fig. 2 the inhibitory connection  $w^{12}$  is varied across the horizontal bands of Fig. 1 while keeping the excitatory connection fixed at  $w^{21} = 12$  (compare the vertical dashed lines in Fig. 1). The vertical axis in Figs. 2A and 2B denotes the mean output of the networks at time  $t$  given by  $\text{mean}(o) := \sum_{i=1}^N (o_i^1 + o_i^2) / 2N$  with  $N=1$  respectively 50. Again it is revealed, that the qualitative behavior of the two systems is the same. The scaling of the inhibitory connection  $w^{12}$  becomes now more apparent giving a factor of approximately 1.06. This

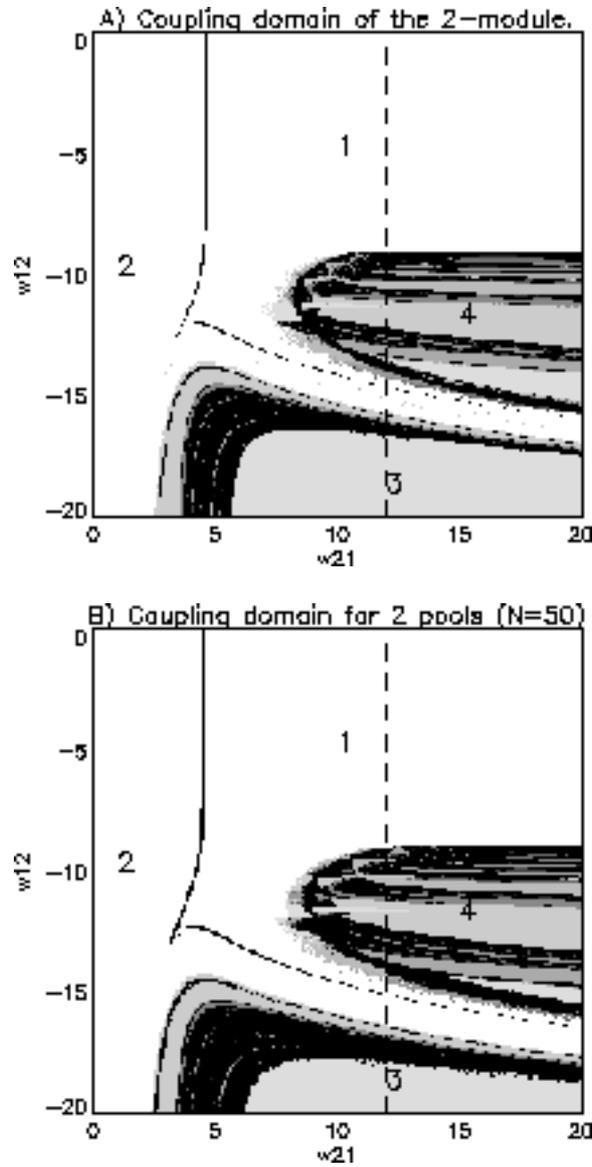


Figure 1: Subspace of the parameter space of the 2-module (A) and the 2-pool-network (B) spanned by the mutual couplings  $w^{12}$  and  $w^{21}$ . Greyscale and numbers refer to periods of orbits. Bifurcation diagrams along the dashed vertical lines are shown in Fig. 2.

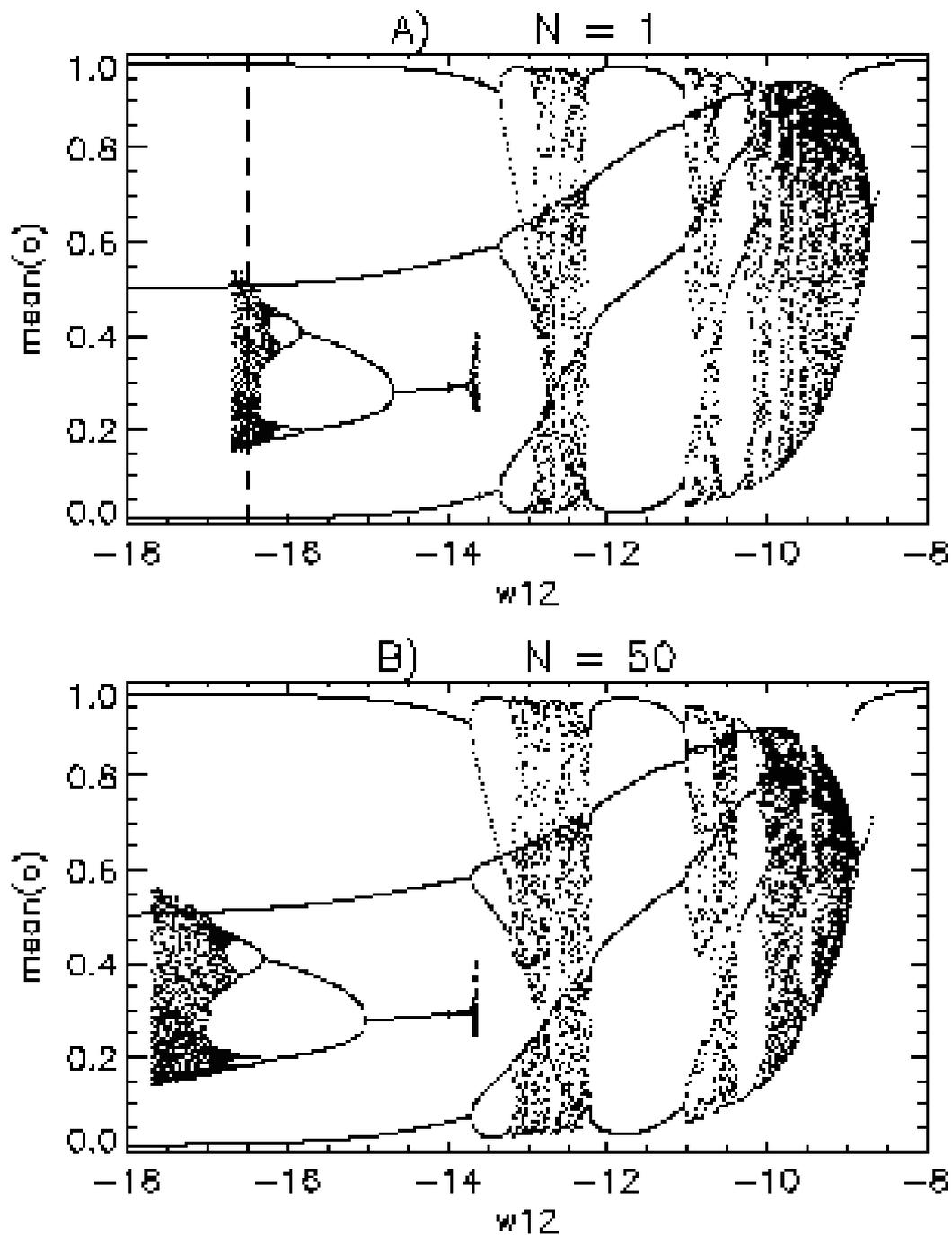


Figure 2: Bifurcation diagrams corresponding to the dashed lines in Fig. 1 for the 2-module (A) and the 2-pool network (B). The vertical axis denotes the mean output of the module respectively the 2-pool network (see text). Observe the same period-doubling routes to chaos and coexisting attractors in both network models. A state space plot for  $w_{12} = -16.5$  - dashed line in A - is shown in Fig. 3.

factor tends to 1.0 for increasing system size.

We observe a bifurcation from a stable fixed point to quasi-periodic attractors at the upper part of the horizontal band (in Fig. 1). Then for growing  $|w^{12}|$ , starting from a periodic orbit, chaotic attractors appear through period doubling bifurcations. In the interval  $-9.08 < w^{12} < -8.7$  they coexist with a stable fixed point (a generalized hysteresis effect). At the lower part of the horizontal band a period doubling sequence to chaos starts at  $w^{12} = -13.8$  from a stable fixed point. All these attractors coexist with a stable period-3 orbit up to  $w^{12} = -16.75$ .

A closer inspection of the system reveals, that there are also parameter regimes where at least two chaotic attractors coexist with a stable periodic orbit. For example at  $w^{21} = 8, w^{12} = -11.95$  two chaotic attractors coexist with a stable fixed point (data not shown). We have verified the existence of chaotic and quasiperiodic attractors also by calculation of the Lyapunov exponents.

Figure 3 displays an example of coexisting attractors in the  $(x_1, x_2)$ -state space of the 2-module ( $w_{12} = -16.5$ ). Crosses represent the points of a period-3 attractor that for large coupling strengths is a global attractor, and which in the range  $-16.7 \leq w_{12} \leq -13.8$  coexists with a period doubling route to chaos (cf. Fig. 2). A coexisting chaotic attractor is shown in the center of the plot. Also plotted are the scattering regions of single units around the period-3 orbit for a simulation with  $N = 50$  neurons per pool. Displayed are the principal ellipses computed from the estimated correlation matrices of elements in each of the three points. The centers of the ellipses represent the averages over the  $2 \times 50$  coordinates. Note that they are in rather good agreement with the periodic points (e.g.  $N = 1$ ). Thus the average activation of the modular network follows the prototypical 2-module quite well. Figure 4 shows the scaling of the scatter in dependence of the system size. The vertical axis measures the total standard deviation, that is the square root of the trace of the correlation matrices, computed in the three states of the periodic attractor and averaged over 10 trials. As has to be expected this scales down like  $1/\sqrt{N}$  with increasing system size [Amari, 1972].

The scatter around the point  $P_2 = (3.7, 3.8)$  always is much lower than for the other two states. This relates to the properties of the dynamics along the periodic orbit, which in  $(x_1, x_2)$  coordinates reads  $P_1 = (-12.4, -3.8), P_2 = (3.7, 3.8), P_3 = (-4.4, 7.7)$ . Output in state  $P_1$  is  $(0.00, 0.02)$ . The predecessor of  $P_2$  therefore is a state with (strong) negative activation and thus nearly no output. That means the activation in  $P_2$  is hardly influenced by any of the recurrent random connections, but mainly given by the external input values ( $P_2 \approx (I_1, I_2) = (4, 4)$ ). Since the input is the same for all cells, scatter

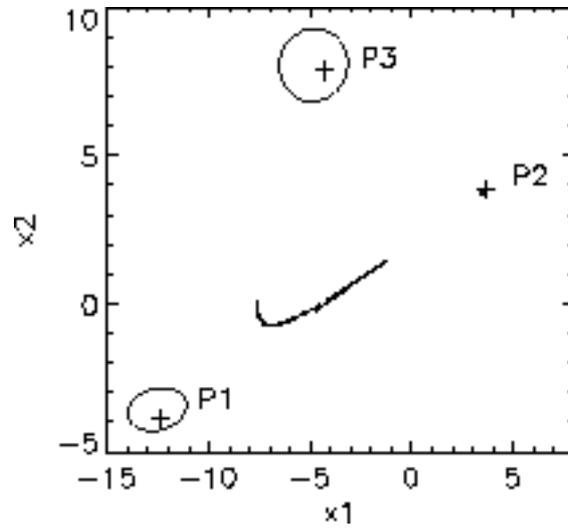


Figure 3: Coexisting period 3-attractor (crosses) and chaotic attractor for  $N=1$ . Ellipses indicate the scattering regions of single elements around the periodic attractor for  $N=50$  (see text). Scatter in P2 is beyond the resolution (cf. fig.4).

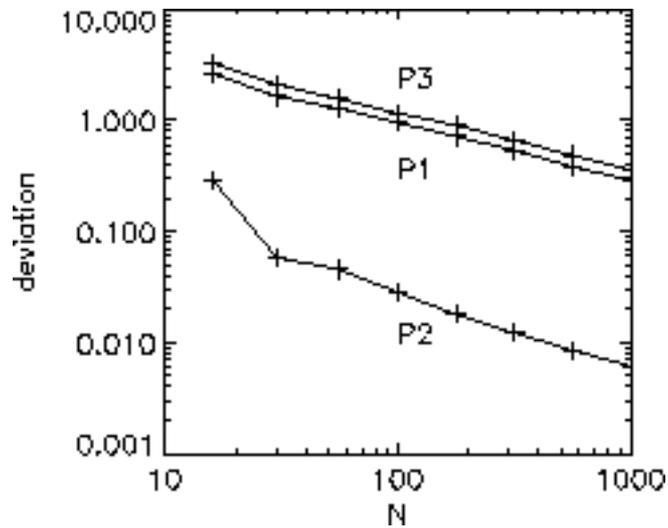


Figure 4: Scaling of scattering ranges around the points of the 3-cycle (cf. text)

necessarily is low. The same reasoning explains, why the standard deviation in state  $P_3$  is larger than in  $P_1$ . In the former case all four feedback paths contribute to the disorder, in the latter only two.

If we finally look at the period-3 orbit in output coordinates  $(o_1, o_2)$ , which approximately reads  $(0,0)$ ,  $(1,1)$ ,  $(0,1)$ , we see, that the periodic orbit obviously starts from a state with (nearly) no output signals ( $P_1$ ), which means, there is no interaction between neurons and thus no computation. Then the external input activates both cell classes ( $P_2$ ) and computation might take place. But due to the stronger inhibitory couplings in relation to the excitatory ones, the activation in the excitatory pool is again switched off ( $P_3$ ). Finally the selfinteraction in the inhibitory pool leads back to the totally silent state ( $P_1$ ). Altogether there is a periodic activation of output signals, which in biological networks would correspond to the rhythmical and coherent firing of the cells in consideration. This might be related to oscillations found in different sensory cortices of cat and monkeys [Eckhorn *et al.*, 1988; Gray *et al.*, 1989; Eckhorn *et al.*, 1993]. In fact, comparable mechanisms for the explanation of those cortical phenomena, based on interacting excitatory and inhibitory pools of cells have already been stated in the literature [Skarda & Freeman, 1987; Erb & Aertsen, 1992; Wennekers *et al.*, 1995]. Interestingly our results point at the possibility that more complex behavior might be found in experiments, when parameters are chosen in appropriate regimes (e.g. intermediate input strengths). Further studies of the possible dynamic modes in our networks seem to be worthwhile.

## 5 Discussion

Our main result, theorem 1, assumes random connections between neurons. This surely is a strong biological simplification, since it does not explicitly take functional correlations into account that may arise due to synaptical learning [Hebb, 1949; Kandel & Schwartz, 1981]. Structural correlations, like different cell classes and/or a coarse connection scheme might at least be considered in form of an appropriate choice of pools.

Nevertheless, the assumption is partly supported by experimental results, suggesting that the topology of horizontal connections on the scale of millimeters to a high degree is of probabilistic nature [Braitenberg & Schüz, 1991; Schüz & Palm, 1989; Hellwig *et al.*, 1994]. Furthermore not all cell classes are able to learn, so this topological randomness is not changed functionally for those that do not learn. At least for those classes randomness seems not to be a too crude assumption.

Another argument why complete randomness might be a reasonable ap-

proximation comes out from the fact, that cortical cells usually exhibit either positive (excitatory) or negative (inhibitory) connections [Eccles, 1964]. This so-called 'Dale's Law' in our description means, that all entries in a coupling-matrix should be of equal sign (or zero), which in turn leads to a global network behavior that mainly is determined by the global connectivity structure between the pools and less by the microscopic level within pools:

In fact the numerical results suggest, that the prediction of theorem 1 is rather robust. Even for small systems the behavior in wide parameter regimes is well described by the low-dimensional prototypical equations. This is related to general properties of the spectrum of random matrices with non-vanishing mean. Under quite general conditions this typically shows (in the limit of infinite  $N$ ) a continuous part around zero of the order  $\sigma\sqrt{N}$  and a single eigenvalue of size  $wN$  (with  $w$  and  $\sigma$  the mean and standard deviation of the matrix entries). The large eigenvalue requires the scaling  $\frac{1}{N}$  in Eq. 4 and corresponds to an approximate eigenvector  $(1, 1, \dots, 1)^T$ . But since the latter exactly represents the synchronous mode, this will ever dominate if  $N$  becomes large enough. It is reasonable to assume, that theorem 1 is approximately valid (that is, the global dynamical behavior of the full system is qualitatively of the same type than that of the small one) if the continuous part of the spectrum is much smaller than the singular eigenvalue. This leads to the simple estimate  $N \gg (\sigma/w)^2$  or, with the values of our example,  $N \gg 1$ . From this point of view, the good agreement between the 2-module and the 2-pool-network with  $N = 50$  is no longer surprising. It is clear that infinitesimal correlations only weakly perturbate the typical properties of the limiting spectral distribution. Even for certain types of strong correlations it can be shown that the stated separation of a continuous part and an absolutely larger single eigenvalue remains valid [Sommers *et al.*, 1988; Edwards & Jones, 1976; Silverstein, 1986]. Nevertheless, it surely is not a general property of random matrices with nonvanishing mean and arbitrary correlations (counter-examples can be easily constructed).

In biological neural networks the variance of coupling matrices is determined by two factors, the probability that two cells are connected and the variance of realized synapses. On the base of physiological data, we estimate - still under the randomness hypothesis - that a number of  $N$  much larger than several hundred should suffice for the dominance of the synchronous eigenvector. At first glance, this number seems relatively small in comparison with the average number of neighboring cells a cortical pyramidal cell might reach (which is of the order of  $10^6$ ) and even with the number of synapses a neuron typically carries (coarsely  $10^4$ ). Nevertheless it only gives a lower estimate that has to be corrected to higher values for several reasons. One are the already mentioned correlations. At least as important is the complex

dynamics of real neurons. Those do not interact continuously (rate code) but via the emission of 'action potentials', short all-or-none events sharply localized in time [Kandel & Schwartz, 1981]. Since those are relatively sparse in space and time, the conclusion is drawn, that only a fraction of all cells projecting on a particular neuron contributes to its input in a short time-interval. Thus, in areas with relative high activation, we might conclude that there in fact is an effective reduction in dynamical complexity due to the signed interaction of the participating cell-classes. But for low activation and also for situations that require the modelling of many pools the question seems to be more difficult to answer.

We also should mention, that our conclusion is implicitly contained in many models of the cortex or other neural systems, in particular if whole cell classes - usually assumed to be excitatory or inhibitory - are combined and described *a priori* by interacting differential equations [an der Heiden, 1980; Geman, 1982; Amari 1972]; or if inhibitory cells are either neglected at all or only taken into account in form of a globally acting activation control [Palm, 1987; Palm 1982; Wennekers *et al.*, 1995; Erb & Aertsen, 1992]. The point becomes even more clear in field models of cortical tissue, which often assume translation-invariant and isotropic cell classes and connections [Wilson & Cowan, 1972; an der Heiden, 1980]. These assumptions can only be valid, if spatially separated cortical sites in principle show the same response properties. But as Cessac *et al.* have shown this cannot be expected in randomly connected networks with vanishing mean  $E[w_{11}]$  [Cessac *et al.*, 1994; Doyon *et al.*, 1993]. In a network similar to ours with one pool they found a period doubling route to chaos, when the disorder was enlarged above a certain critical value. These chaotical solutions are *not* determined by the global network structure; the attractors can be arbitrarily embedded in (the high dimensional) phase space depending on the sample realization of the connections. If this result is physiologically relevant, it would mean that the typical behavior of localized cortical patches should be expected to vary in space according to the microscopic details of connectivity. Furthermore, also the assumption of scalar interactions between different sites would appear to be questionable, such that spatial homogenous field equations could hardly be appropriate descriptions, but in turn more complex schemes would have to be derived. Actually we cannot decide which of the descriptions is more suitable. Perhaps both are appropriate in different situations, say different cortical areas, or different activation states.

To summarize: the main purpose of this work was to establish a relationship between the dynamics of highdimensional modular neural networks with several pools of randomly connected neurons and corresponding types of low-dimensional model equations (neuromodules). We have shown that the

rich dynamical behavior observed in large parameter regimes of neuromodules can be expected to be realized also for large modular networks. On the other hand, as our theorem demonstrates, the detailed study of such large networks - and in particular of interconnected subsystems of them - can be effectively reduced to the study of low-dimensional equations, still complex in behavior, but much easier to analyze. This means, the investigation even of extremely small networks can provide insight into the properties of much larger systems. Since the coarse feedback topology of our modular neural networks might be arbitrarily complex, the concept can be adapted to a broad range of biological situations, from the layered structure of cortical tissue, via its columnar structure, up to whole interconnected areas. Of course, since randomness in the connections is assumed, in any case only properties of the *global* activation dynamics can be inferred; the microscopic details of information processing still have to be explored with the aid of more elaborated models.

## 6 Acknowledgements

T.Wennekers acknowledges support by the Deutsche Forschungsgemeinschaft under grant Pa 268/8-1.

## References

- [1] van der Heiden, U. [1980] *Analysis of Neural Networks*, Lecture Notes in Biomathematics 35 (Springer Verlag, Berlin).
- [2] Amari, S.I. [1971] "Characteristics of randomly connected threshold-element networks and network systems", *Proc. IEEE* **59**, 35-47.
- [3] Amari, S.I. [1972] "Characteristics of random nets of analog neuron-like elements", *IEEE Trans. SMC* **2**, 643-657.
- [4] Amari, S.I., Joshida K. & Kanatani K.I. [1977] "A mathematical foundation for statistical neurodynamics", *SIAM J. Appl. Math.* **33**, 95-126.
- [5] Babloyantz, A. & Lourenço, C. [1994] "Computation with chaos: A paradigm for cortical activity", *Proc. Natl. Acad. Sci. USA* **91**, 9027-9031.
- [6] Braitenberg, V. & Schüz, A. [1991] *Anatomy of the Cortex*, Studies of Brain Functions Vol.18 (Springer, Berlin).
- [7] Cessac, B., Doyon, B., Quoy, M. & Samuelidis, M. [1994] "Mean-field equations, bifurcation map and route to chaos in discrete type neural networks", *Physica* **D74**, 24-44.
- [8] Doyon, B., Cessac, B., Quoy, M. & Samuelidis, M. [1993] "Control of the transition to chaos in neural networks with random connectivity", *Int. J. Bif. Chaos* **3**, 279-291.
- [9] Duke, W. & Pritchard, W.S. (eds) [1991] *Proceedings of the Conference on Measuring Chaos in the Human Brain* (World Scientific, Singapore).
- [10] Eccles, J.C. [1964] *Physiology of Synapses* (Springer, Berlin).
- [11] Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H.J. [1988] "Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analysis in the cat", *Biol.Cybern.* **60**, 121-130.
- [12] Eckhorn, R., Frien, A., Bauer, R., Woelbern, T. & Kehr, H. [1993] "High frequency (60-90 Hz) oscillations in primary visual cortex of awake monkey", *NeuroReport* **4**, 243.
- [13] Edwards, S.F. & Jones, R.C. [1976] "The eigenvalue spectrum of a large symmetric random matrix", *J. Phys.* **A9**, 1595-1603.

- [14] Elbert, T., Ray, W.J., Kowalik, Z.J., Skinner, K.E., Graf, K.E. & Birbaumer, N. [1994] "Chaos and physiology: Deterministic chaos in excitable cell assemblies", *Physiological Review* **74**, 1-47.
- [15] Erb, M. & Aertsen, A. [1992] "Dynamics of activity in biology-oriented neural network models: Stability at low firing rates", in: Aertsen, A. & Braitenberg, V. (eds.) [1992] *Information Processing in the Cortex: Experiments and Theory* (Springer, Berlin).
- [16] Ermentrout, B. [1984] "Period doublings and possible chaos in neural models", *SIAM J. Appl. Math.* **44**, 80-95.
- [17] Geman, S. [1982] "Almost sure stable oscillations in a large system of randomly coupled equations", *SIAM J. Appl. Math.* **42**, 695-703.
- [18] Gray, C.M., König, P., Engel, A.K. & Singer, W. [1989] "Oscillatory responses in cat visual cortex exhibit intercolumnar synchronisation which reflects global stimulus properties", *Nature* **338**, 334-337.
- [19] Guevara, M.R., Glass, L., Mackey, M.C. & Shrier, A. [1983] "Chaos in neurobiology", *IEEE Trans. SMC* **13**, 790-798.
- [20] Hansel, D. & Sompolinski, H. [1992] "Synchronization and computation in a chaotic neural network", *Pys.Rev.Lett.* **68**, 718-721.
- [21] Harth, E. [1983] "Order and chaos in neural systems: An approach to the dynamics of higher brain functions", *IEEE Trans. SMC* **13**, 782-798.
- [22] Harth, E., Csermely, T.J., Beek, B. & Lindsay, R.D. [1970] "Brain functions and neural dynamics", *J. Theo. Biol.* **26**, 93-120.
- [23] Hebb, D.O. [1949] *The Organization of Behavior. A Neuropsychological Theory* (Wiley, New York).
- [24] Hellwig, B., Schüz, A. & Aertsen, A. [1994] "Synapses on axon collaterals of pyramidal cells are spaced at random intervals: a Golgi study in the mouse cerebral cortex", *Biol.Cybern.* **71**, 1-12.
- [25] Kandel, E.R. & Schwartz, J.H. [1981] *Principles of Neural Science* (Elsevier, New York).
- [26] Kaneko, K. [1990] "Globally coupled chaos violates the law of large numbers but not the central-limit theorem", *Phys. Rev. Lett.* **65**, 1391.

- [27] Palm, G. [1982] *Neural Assemblies. An Alternative Approach to Artificial Intelligence* (Springer, Berlin).
- [28] Palm, G. [1987] "Comments on the paper 'Threshold regulation - a key to the understanding of the combined dynamics of EEG and event-related potentials' by T.Elbert and B.Rockstroh", *J.of Psychophysiology* **4**, 335.
- [29] Pasemann, F. [1993] "Dynamics of a single model neuron", *Int. J. Bif. Chaos* **2**, 271-278.
- [30] Pasemann, F. & Nelle, E. [1993] "Elements of nonconvergent neurodynamics", in: Andersson, S., Andersson, A. & Ottoson, U. (eds.) *Dynamical Systems - Theory and Application* (World Scientific, Singapore) pp. 167-201.
- [31] Pasemann, F. [1995] "Neuromodules: A dynamical systems approach to brain modelling", in: Herrmann, H.J., Wolf, D.E. & Pöppel, E. (eds.) *Supercomputing in Brain Research: From Tomography to Neural Networks* (World Scientific, Singapore) pp. 331-348.
- [32] Pecora, L.M. & Carroll, T.L. [1990] "Synchronization in chaotic systems", *Phys. Rev. Lett.* **64**, 821-824.
- [33] Pecora, L.M. & Carroll, T.L. [1991] "Driving systems with chaotic signals", *Phys. Rev.* **A44**, 2374-2383.
- [34] Pikovsky, A.S. & Kurths, J. [1994a] "Do globally coupled maps really violate the law of large numbers?", *Phys. Rev. Lett.* **72**, 1644-1664.
- [35] Pikovsky, A.S. & Kurths, J. [1994b] "Collective behavior in ensembles of globally coupled maps", *Physica* **D76**, 411-419.
- [36] Richards, W., Wilson, H.R. & Sommer, M.A. [1994] "Chaos in percepts?", *Biol. Cybern.* **70**, 345-349.
- [37] Rozonoér, L.I. [1969] "Random logical nets, I, II, III", *Automation and Remote Control* (English translation of the Russian Journal) **5**, 773-781; **6**, 922-931; **7**, 1130-1139.
- [38] Schüz, A. & Palm, G. [1989] "Density of neurons and synapses in the cerebral cortex of the mouse", *J. Comp. Neurol.* **268**, 442-455.

- [39] Silverstein, J.W. [1986] "Eigenvalues and eigenvectors of large dimensional sample covariance matrices", in: Cohen, J.E. *et al.* (eds.) *Random Matrices and their Applications* (Amer.Math.Soc., Providence, Rhode Island).
- [40] Skarda, C.A. & Freeman, W. [1987] "How brains make chaos in order to make sense of the world", *Beh. Brain. Sci.* **10**, 161-195.
- [41] Sommers, H.J., Crisanti, A., Sompolinsky, H. & Stein, Y. [1988] "Spectrum of large random asymmetric matrices", *Phys. Rev. Lett.* **60**, 1895-1898.
- [42] Sompolinski, H., Crisanti, H. & Sommers, H.J. [1988] "Chaos in random neural networks", *Phys. Rev. Lett.* **61**, 259-262.
- [43] Tani, J. & Fukumura, N. [1995] "Embedding a grammatical description in deterministic chaos: an experiment in recurrent neural learning", *Biol. Cybern.* **72**, 365-370.
- [44] Wennekers, T., Sommer, F.T. & Palm, G. [1995] "Iterative retrieval in associative memories by threshold control of different neural models", in: Herrmann, H.J., Wolf, D.E. & Pöppel, E. (eds.) *Supercomputing in Brain Research: From Tomography to Neural Networks* (World Scientific, Singapore).
- [45] Wilson, H.R. & Cowan, J.D. [1972] "Excitatory and inhibitory interactions in localized populations of model neurons", *Biophysical Journal* **12**, 1-24.

## A Proof of the theorem

Here we outline the proof of theorem 1. The main steps follow the work of Geman [1982] with modifications due to the time-discrete nature of the dynamics. We define

$$u_i(t+1) = \lambda u_i(t) + I(t) + \frac{w}{N} \sum_j f(x_j(t)) \quad \text{with} \quad u_i(0) = 0, \quad i = 1, 2, \dots, N, \quad (7)$$

set  $\beta = \sup_{x \in \mathbb{R}} |f(x)|$ , and choose a Lipschitz-constant  $L$  for  $f$ .

First we show that the norm  $\|\vec{x}(t) - \vec{u}(t)\|$  remains bounded for arbitrary  $N$  and  $t$ . (We take the Euclidean norm for vectors and induced norm for matrices). With  $\mathbf{M} = \{w\}$ , the  $N \times N$  matrix with all components  $w = E[w_{11}]$ , we derive from eqns. (5) and (7) using the boundedness of  $f$ :

$$\|\vec{x}(t+1) - \vec{u}(t+1)\| \leq |\lambda| \|\vec{x}(t) - \vec{u}(t)\| + \left\| \frac{\mathbf{W} - \mathbf{M}}{\sqrt{N}} \right\| \cdot \beta, \quad (8)$$

which (proof by induction) leads to

$$\|\vec{x}(t) - \vec{u}(t)\| \leq \|\vec{x}(0)\| + \beta \cdot \left\| \frac{\mathbf{W} - \mathbf{M}}{\sqrt{N}} \right\| / (1 - |\lambda|) \stackrel{\text{def}}{=} C(\mathbf{W}, \vec{x}(0)). \quad (9)$$

Since  $\lim_{N \rightarrow \infty} \left\| \frac{\mathbf{W} - \mathbf{M}}{\sqrt{N}} \right\| \leq 2\sigma = 2 \cdot E[(w_{11} - w)^2]^{1/2}$  (almost surely, see [Geman, 1982]) and  $|\lambda| < 1$ , it follows that  $\|\vec{x}(t) - \vec{u}(t)\|$  remains bounded for all  $N$  and  $t$ . (Note that this step of the proof fails for autonomous (e.g. chaotic) single unit dynamics).

Given the above bound (eqn.9), we now show that the maximum difference of any  $x_i(t)$  and  $u_i(t)$  asymptotically tends to zero:

$$\begin{aligned} \sup_i |x_i(t+1) - u_i(t+1)| &\leq |\lambda| \cdot \sup_i |x_i(t) - u_i(t)| \\ &+ \left\| \frac{\mathbf{W} - \mathbf{M}}{N} \right\| \|\vec{F}(\vec{x}(t)) - \vec{F}(\vec{u}(t))\| \quad (10) \\ &+ \sup_i \left| \frac{1}{N} \sum_j (w_{ij} - w) f(u_j(t)) \right| \end{aligned}$$

Using the Lipschitz-condition on  $f$  and the bound (9) in the second term on the right hand side of (10), the boundedness of  $f$  and the fact that (by definition) all  $u_i(t)$  are identical in the third term, the iteration of (10) leads to

$$\begin{aligned} \sup_i |x_i(t) - u_i(t)| &\leq \sup_i |x_i(0)| \cdot |\lambda|^t + (L \cdot \left\| \frac{\mathbf{W} - \mathbf{M}}{N} \right\| C(\mathbf{W}, \vec{x}(0))) \\ &+ \beta \sup_i \left| \frac{1}{N} \sum_j (w_{ij} - w) \right| / (1 - |\lambda|) \quad (11) \end{aligned}$$

Because under the conditions given in the theorem  $\lim_{N \rightarrow \infty} \left\| \frac{\mathbf{W} - \mathbf{M}}{N} \right\| = 0$  and  $\lim_{N \rightarrow \infty} \sup_i \left| \frac{1}{N} \sum_j w_{ij} - w \right| = 0$  are valid (again almost surely, cp. [Geman, 1982]),  $|\lambda| < 1$  and  $C(\mathbf{W}, \vec{x}(0))$  is bounded, we finally have

$$\lim_{N \rightarrow \infty} \limsup_{t \rightarrow \infty} \sup_i |x_i(t) - u_i(t)| = 0 \quad a.s. \quad (12)$$

which directly leads to proposition B1 of theorem 1.

To proof proposition B2 we write

$$\begin{aligned} & \sup_i |x_i(t+1) - \{\lambda x_i(t) + w \cdot f(x_i(t)) + I(t)\}| \\ & \leq \sup_i |x_i(t+1) - u_i(t+1)| + \sup_i |u_i(t+1) - \{\cdot\cdot\cdot\}| \end{aligned} \quad (13)$$

$$\begin{aligned} & \leq \sup_i |x_i(t+1) - u_i(t+1)| + |\lambda| \sup_i |x_i(t) - u_i(t)| \\ & + L \cdot w \cdot \frac{1}{N} \sum_j \sup_{i,j} |(x_j(t) - x_i(t))| \end{aligned} \quad (14)$$

To derive (14) we inserted (7) in the second term on the right hand side of (13) and used the Lipschitz condition on  $f$ . From (14) proposition B2 of the theorem follows by applying B1 and equation (12).

B3 is a simple consequence of B1 and B2:

$$\begin{aligned} & |X^{(N)}(t+1) - \Phi(X^{(N)}(t), t)| \leq \frac{1}{N} \sum_i |x_i(t+1) - \Phi(X^{(N)}(t), t)| \\ & \leq \frac{1}{N} \sum_i (|x_i(t+1) - \Phi(x_i(t), t)| + |\Phi(x_i(t), t) - \Phi(X^{(N)}(t), t)|) \\ & \leq \sup_i |x_i(t+1) - \Phi(x_i(t), t)| + (\lambda + wL) \sup_i |x_i(t) - X^{(N)}(t)| \end{aligned} \quad (15)$$

In (15) we have used the fact that  $\lambda + wL$  is a Lipschitz constant for  $\Phi(x, t)$  in the first argument uniformly in the second. Now the first term in (15) asymptotically vanishes because B2 holds. For the second term we have:

$$\sup_i |x_i(t) - X^{(N)}(t)| \leq \frac{1}{N} \sum_j |x_i(t) - x_j(t)| \leq \sup_{i,j} |x_i(t) - x_j(t)|$$

Since the latter converges to zero by B1 the proof is complete.