# Semantic learning in autonomously active recurrent neural networks

Claudius Gros and Gregor Kaczor, *Institute for Theoretical Physics, Goethe University Frankfurt, 60054 Frankfurt/Main, Germany.*
*E-mail: gros07@itp.uni-frankfurt.de*

## Abstract

The human brain is autonomously active, being characterized by a self-sustained neural activity which would be present even in the absence of external sensory stimuli. Here we study the interrelation between the self-sustained activity in autonomously active recurrent neural nets and external sensory stimuli.

There is no a priori semantical relation between the influx of external stimuli and the patterns generated internally by the autonomous and ongoing brain dynamics. The question then arises when and how are semantic correlations between internal and external dynamical processes learned and built up?

We study this problem within the paradigm of transient state dynamics for the neural activity in recurrent neural nets, i.e. for an autonomous neural activity characterized by an infinite time-series of transiently stable attractor states. We propose that external stimuli will be relevant during the sensitive periods, *viz* the transition period between one transient state and the subsequent semi-stable attractor. A diffusive learning signal is generated unsupervised whenever the stimulus influences the internal dynamics qualitatively.

For testing we have presented to the model system stimuli corresponding to the bars and stripes problem. We found that the system performs a non-linear independent component analysis on its own, being continuously and autonomously active. This emergent cognitive capability results here from a general principle for the neural dynamics, the competition between neural ensembles.

*Keywords*: recurrent neural networks, autonomous neural dynamics, transient state dynamics, emergent cognitive capabilities

## 1 INTRODUCTION

It is well known that the brain has a highly developed and complex self-generated dynamical neural activity. We are therefore confronted with a dichotomy when attempting to understand the overall functioning of the brain or when designing an artificial cognitive system: A highly developed cognitive system, such as the brain [1], is influenced by sensory input but it is not driven directly by the input signals. The cognitive system needs however this sensory information vitally for adapting to a changing environment and survival.

In this context we then want to discuss two mutually interrelated questions:

- Can we formulate a meaningful paradigm for the self-sustained internal dynamics of an autonomous cognitive system?
- How is the internal activity influenced by sensory signals, *viz* which are the principles for the respective learning processes?

We believe that these topics represent important challenges for research in the field of recurrent neural networks and the modeling of neural processes. From an experimental point of view we note that an increasing flux of results from neurobiology supports the notion of quasi-stationary spontaneous neural activity in the cortex [1–6]. It is therefore reasonable

Vol. 18 No. 5, © The Author 2009. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
doi:10.1093/jigpal/jzp045 Advance Access published 14 September 2009

to investigate the two questions formulated above with the help of neural architectures centrally based on the notion of spontaneously generated transient states, as we will do in the present investigation using appropriate recurrent neural networks.

## *1.1 Transient-state and competitive dynamics*

Standard classification schemes of dynamical systems are based on their long-time behavior, which may be characterized, e.g., by periodic or chaotic trajectories [8]. The term 'transient-state dynamics' refers, on the other hand, to the type of activity occurring on intermediate time scales, as illustrated in Fig. 1. A time series of semi-stable activity patterns, also denoted transient attractors, is characterized by two time scales. The typical duration $t_{trans}$ of the activity plateaus and the typical time $\Delta t$ needed to perform the transition from one semi-stable state to the subsequent one. The transient attractors turn into stable attractors in the limit $t_{trans}/\Delta t \to \infty$.

Transient state dynamics is intrinsically competitive in nature. When the current transient attractor turns unstable the subsequent transient state is selected by a competitive process. Transient-state dynamics is a form of 'multi-winners-take-all' process, with the winning coalition of dynamical variables suppressing all other competing activities.

Humans can discern about 10-12 objects per second [7] and it is therefore tempting to identify the cognitive time scale of about 80-100ms with the duration $t_{trans}$ of the transient-state dynamics illustrated in Fig. 1. Interestingly, this time scale also coincides with the typical duration [4] of the transiently active neural activity patterns observed in the cortex [2, 3, 5, 6].

Several high-level functionalities have been proposed for the spontaneous neural brain dynamics. Edelman and Tononi [9, 10] argue that 'critical reentrant events' constitute transient conscious states in the human brain. These 'states-of-mind' are in their view semi-stable

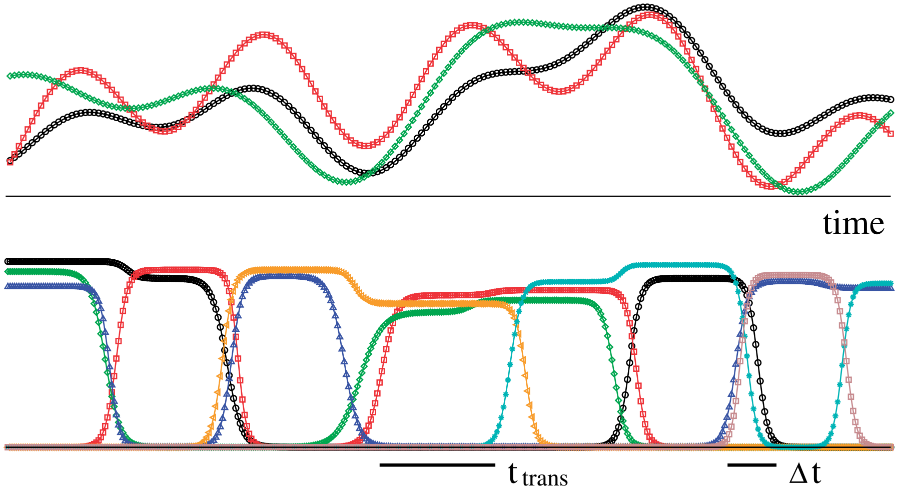

FIG. 1. (Colour online). Schematic illustration of general or fluctuating activity patterns (top) and of transient-state dynamics (bottom), which is characterized by typical time scales $t_{trans}$ and $\Delta t$ for the length of activity-plateau and of the transient period respectively.

global activity states of a continuously changing ensemble of neurons, the 'dynamic core'. This activity takes place in what Dehaene and Naccache [11] denote the 'global workspace'. The global workspace serves, in the view of Baars and Franklin [12], as an exchange platform for conscious experience and working memory. Crick and Koch [13] and Koch [14] have suggested that the global workspace is made-up of 'essential nodes', i.e. ensembles of neurons responsible for the explicit representation of particular aspects of visual scenes or other sensory information.

### *1.2 Autonomously active recurrent neural nets*

Traditional neural network architectures are not continuously active on their own. Feedforward setups are explicitly driven by external input [15] and Hopfield-type recurrent nets settle into a given attractor after an initial period of transient activities [16]. The possibilities of performing cognitive computation with autonomously active neural networks, the route chosen by nature, are however investigated increasingly [17]. In this context the time encoding of neural information, one of the possible neural codes [18], has been studied in various contexts. Two network architectures, the echo state network suitable for rate-encoding neurons [19], and the liquid state machine suitable for spiking neurons [20], have been proposed to transiently encode in time a given input for further linear analysis by a subsequent perceptron. Both architectures, the echo-state network and the liquid-state machine, are examples of reservoir architectures with fading memories, which however remain inactive in the absence of sensory input.

An example of a continuously active recurrent network architecture is the winnerless competition based on stable heteroclinic cycles [21]. In this case the trajectory moves along heteroclines from one saddle point to the next approaching a complex limiting cycle. Close to the saddle points the dynamics slows down leading to well defined transiently active neural activity patterns.

## 2 CLIQUE ENCODING IN RECURRENT NETWORKS

In order to study the issues raised in the introduction, the notion of autonomous neural activity and its relation to the sensory input, we will consider a specific model based on clique-encoding recurrent nets. The emphasis will be on the discussion of the general properties and of the underlying challenges. We will therefore present here an overview of the algorithmic implementation, referring in part to the literature for further details.

### *2.1 Cliques, attractors and transient states*

Experimental evidence indicates that sparse neural coding is an important operating principle in the brain, as it minimizes energy consumption, maximizes storage capacity and contributes to make information encoding spatially explicit [22]. A powerful form of sparse coding is multi-winners-take-all encoding in the form of cliques. The term cliques stems from network theory and denotes subgraphs which are fully interconnected [8], a few examples are given in Fig. 2. Cliques are fully interconnected subgraphs of maximal size, in the sense

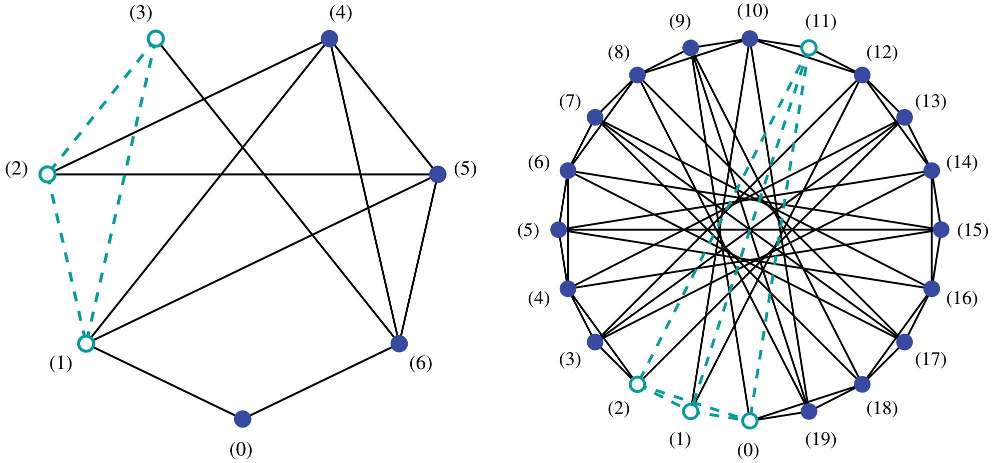

Fig. 2. (Colour online). Illustration of clique encoding. Note that cliques are fully connected subgraphs of maximal size. The dashed lines are highlighted examples of specific cliques. Left: A 7-site network with cliques (1,2,3), (3,6), (1,2,4,5), (4,5,6), (0,6) and (0,1). Right: A regular 20-site networks containing 10 four-site cliques.

that they are not part of another fully interconnected subgraph containing a larger number of vertices.

Clique encoding is an instance of sparse coding with spatially overlapping memory states. The use of clique encoding is in fact motivated by experimental findings indicating a hierarchical organization of overlapping neural clique assemblies for the real-time memory representation in the hippocampus [23]. In the framework of a straightforward auto-associative neural network the cliques are defined by the network of the excitatory connections, which are shown as lines in Fig. 2, in the presence of an inhibitory background [24, 25]. In this setting all cliques correspond to attractors of the network, *viz* to spatially explicit and overlapping memory representations.

One can transform the attractor network with clique encoding into a continuously active transient-state network, by introducing a reservoir variable for every neuron. In this setting the reservoir of a neuron is depleted whenever the neuron is active and refilled whenever the neuron is inactive. Via a suitable local coupling between the individual neural activity and reservoir variables a well defined and stable transient state dynamics is obtained [24, 25]. When a given clique becomes a winning coalition, the reservoirs of its constituting sites are depleted over time. When fully depleted the winning coalition becomes unstable and the subsequent winning coalition is activated through a competitive associative process, leading to an ever ongoing associative thought process. The resulting network architecture is a dense and homogeneous associative network (dHan) [24]. An illustrative result of a numerical simulation is given in Fig. 3.

For the isolated system, not coupled to any sensory input, this associative thought process has no semantic content, as the transient attractors, the cliques, have none. The semantic content can be acquired only by coupling to a sensory input and by the generation of correlation between the transient attractors and patterns extracted from the input data stream.

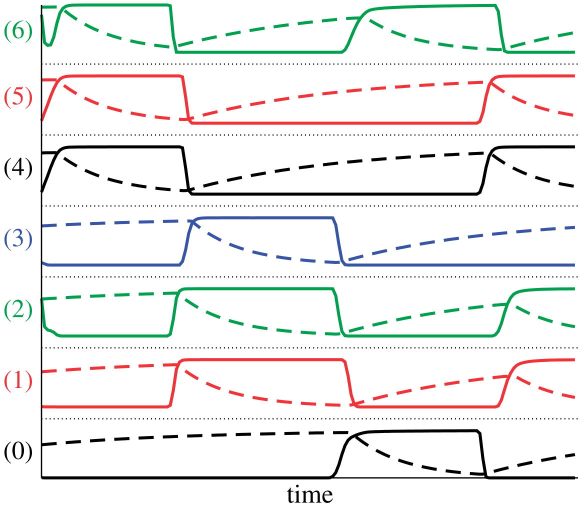

FIG. 3. (Colour online). Transient state dynamics of the 7-site network illustrated in Fig. 2. Shown are, vertically displaced, the time developments of the respective neural activities (solid lines) and of the neural reservoirs (dashed lines). The time series of spontaneously generated transient states are the cliques $(4,5,6) \to (1,2,3) \to (0,6) \to (1,2,4,5)$ (from the left to the right).

## *2.2 Competitive dynamics and sensitive periods*

To be definite, we utilize a continuous-time formulation for the dHan architecture, with rate-encoding neurons, characterized by normalized activity levels $x_i \in [0,1]$. One can then define, via

$$\frac{d}{dt}x_i = \begin{cases} (1-x_i)\,r_i & (r_i > 0) \\ x_i\,r_i & (r_i < 0) \end{cases} \tag{1}$$

the respective growth rates $r_i$. Representative time series of growth rates $r_i$ are illustrated in Fig. 4. When $r_i > 0$, the respective neural activity $x_i$ increases, approaching rapidly the upper bound; when $r_i < 0$, it decays to zero. The model is specified [24, 25], by providing the functional dependence of the growth rates with respect to the set of activity-levels $\{x_j\}$ of all sites and on the synaptic weights, as usual for recurrent or auto-associative networks.

During the transition periods many, if not all, neurons will enter the competition to become a member of the new winning coalition. The competition is especially pronounced whenever most of the growth rates $r_i$ are small in magnitude, with no subset of growth rates dominating over all the others. Whether this does or does not happen depends on the specifics of the model setup. In Fig. 4, two cases are illustrated. In the first case (lower graph) the competition for the next winning coalition is restricted to a subset of neurons, in the second case (upper graph) the competition is network-wide. When most neurons participate in the competition process for a new winning coalition the model will have ‘sensitive periods’ during the transition times and it will be able to react to eventual external signals.

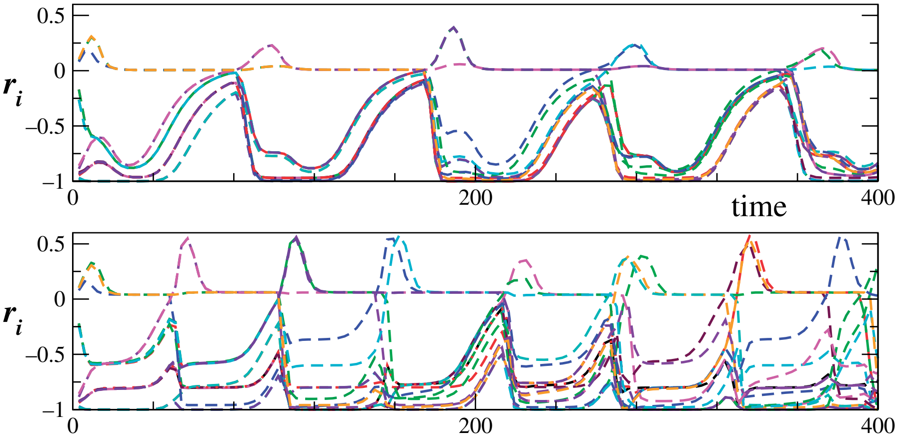

FIG. 4. (Colour online). The growth rates $r_i(t)$ generating an internal transient state-dynamics via Eq. (1). The two examples differ in the functional dependence of the $r_i(t)$ on the $x_j(t)$. The top graph corresponds to a system having sensitive periods, the bottom graph to a system without distinctive sensitive periods.

## *2.3 Sensitive periods and learning*

So far we have discussed in general terms the properties of isolated models exhibiting a self-sustained dynamical behavior in terms of a never-ending time series of semi-stable transient states, as illustrated in Figs. 3 and 4, using the dHan architecture with continuous-time and rate-encoding neurons.

The importance of sensitive periods comes in when the network exhibiting transient-state dynamics is coupled to a stream of sensory input signals. It is reasonable to assume, that external input signals will contribute to the growth rates $r_i$ via

$$r_i \;\equiv\; r_i^{dHan} + \Delta r_i\Big(1-\Theta(r_i^{dHan})\Theta(-\Delta r_i)\Big)\;. \qquad (2)$$

Here the $\Delta r_i$ encode the influence of the input signals and we have denoted now with $r_i^{dHan}$ the contribution to the growth rate a neuron in the dHan layer receives from the other dHan neurons. The factor $(1-\Theta(r_i^{dHan})\Theta(-\Delta r_i))$ in Eq. (2) ensures that the input signal does not deactivate the current winning coalition as we will discuss further below. Let us here assume for a moment, as an illustration, that the input signals are suitably normalized, such that

$$\Delta r_i \;\simeq\; \begin{cases} 0.5 & \text{(active input)} \\ \;\;0 & \text{(inactive input)} \end{cases}\;, \qquad (3)$$

in order of magnitude. For the simulations presented further below a qualitatively similar optimization will occur homeostatically. For the transient states the $r_i^{dHan}\approx -1$ for all sites not forming part of the winning coalition and the input signal $\Delta r_i$ will therefore not destroy the transient state, compare Figs. 4 and 5. With the normalization given by Eq. (3) the total

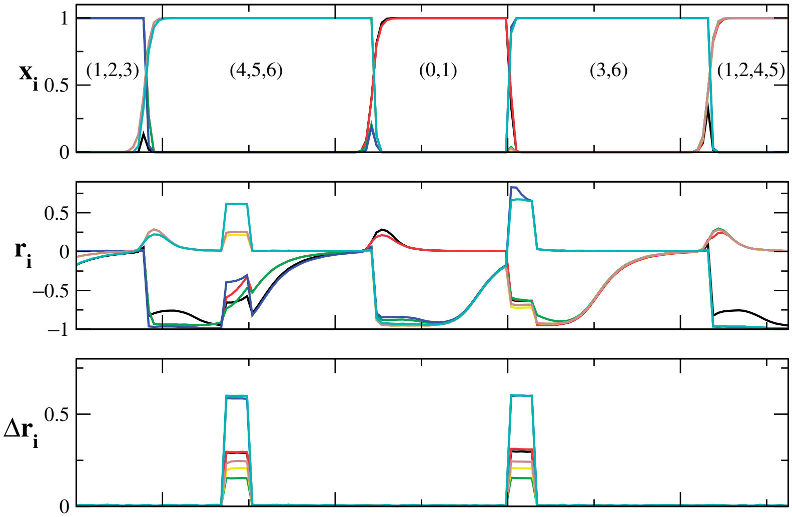

FIG. 5. (Colour online). Results for the activities $x_i(t)$, the total growth rates $r_i(t)$ and the input signals $\Delta r_i(t)$ from a simulation of the 7-site system shown in Fig. 2. The time series of winning coalitions is given. The first input signal does not change the composition of the winning coalition, whereas the second does.

growth rate $\sim(\Delta r_i - 1)$ will remain negative for all inactive sites and the sensory input will not be able to destroy the current winning coalition. The input signal will however enter the competition for the next winning coalition during a sensitive period, providing an additional boost for the respective neurons.

This situation is exemplified in Fig. 5, where we present simulation results for the 7-site system shown in Fig. 2, subject to two sensory inputs $\Delta r_i(t)$. The self-generated time series of winning coalitions is not redirected for the first sensory input. The second stimulus overlaps with a sensory period and its strongest components determine the new winning coalitions. The simulation results presented in Fig. 5 therefore demonstrate the existence of well defined time-windows suitable for the learning of correlations between the input signal and the intrinsic dynamical activity. The time windows, or sensitive periods, are present during and shortly after a transition from one winning coalition to the subsequent. A possible concrete implementation for this type of learning algorithm will be given further below.

Let us now come to the factor $(1-\Theta(r_i^{dHan})\Theta(-\Delta r_i))$ in Eq. (2), containing the Heaviside-step functions $\Theta(x)$. For vertices $i$ of the current winning coalition the intra dHan-layer growth rates are positive, $r_i^{dHan}>0$. Therefore, the above factor ensures, that a suppressive $\Delta r_i<0$ has no effect on the members of the current winning coalition. The contribution $\Delta r_i$ from the input may therefore alter the balance in the competition for the next winning coalition during the sensitive periods, but not suppress the current active clique.

Let us note, that the setup discussed here allows the system also to react to an occasional strong excitatory input signal having $\Delta r_i>1$. Such a strong signal would suppress the current transient state altogether and impose itself. This possibility of rare strong input signals is evidently important for animals and would be, presumably, also helpful for an artificial cognitive system.

## *2.4 Diffusive learning signals*

Let us return to the central problem inherent to all systems reacting to input signals and having at the same time a non-trivial intrinsic dynamical activity. Namely, when should learning occur, i.e. when should a distinct neuron become more sensitive to a specific input pattern and when should it suppress its sensibility to a sensory signal.

The framework of competitive dynamics developed above allows for a straightforward solution of this central issue: Learning should occur exclusively when the input signal makes a qualitative difference, *viz* when the input signal deviates the transient-state process. For illustration let us assume that the series of winning coalitions is

$$(4,5,6) \overset{[a]}{\longrightarrow} (0,1) \overset{[a]}{\longrightarrow} (1,2,4,5)\ ,$$

where the index $[a]$ indicates that the transition is driven by the autonomous internal dynamics and that the series of winning coalitions take the form

$$(4,5,6) \overset{[a]}{\longrightarrow} (0,1) \overset{[s]}{\longrightarrow} (3,6)\ ,$$

in the presence of a sensory signal $[s]$, as it is the case for the data presented in Fig. 5. Note, that a background of weak or noisy sensory input could be present in the first case, but learning should nevertheless occur only in the second case. A reliable distinction between these two cases can be achieved via a suitable diffusive learning signal[1] $S(t)$. It is activated whenever any of the input contributions $\Delta r_i$ changes the sign of the respective growth rates during the sensitive periods,

$$\frac{d}{dt}S \ \to \ \begin{cases} \Gamma^{+}_{diff} & (r_i > 0) \text{ and } (r_i^{dHan} < 0) \\ -\Gamma^{-}_{diff} & \text{otherwise} \end{cases}\ , \qquad (4)$$

*viz* when it makes a qualitative difference. Let us remember that the $r_i^{dHan}$ are the internal contributions to the growth rate, i.e. the input a dHan neuron receives via recurrent connections from the other dHan neurons. The diffusive learning signal $S$ is therefore increasing in strength only when a neuron is activated externally, but not when activated internally, with the $\Gamma^{\pm}_{diff} > 0$ denoting the respective growth and decay rates. The diffusive learning signal $S(t)$ is a global signal and a sum $\sum_i$ over all dynamical variables is therefore implicit on the right-hand side of Eq. (4).

## *2.5 The role of attention*

The general procedure for the learning of correlation between external signals and intrinsic dynamical states for a cognitive system presented here does not rule out other mechanisms. Here we concentrate on the learning algorithm which occurs automatically, one could say sub-consciously. Active attention focusing, which is well known in the brain to potentially shut off a sensory input pathway, or to enhance sensibility to it, may very well work in parallel to the continuously ongoing mechanism investigated here.

[1] The name 'diffusive learning signal' [8] stems from the fact, that many neuromodulators are released in the brain in the intercellular medium and then diffuse physically to the surrounding neurons, influencing the behavior of large neural assemblies.

We note, however, that the associative thought process within the dHan carries with it a dynamical attention field [24]. Neurons receiving both positive and negative contributions from the winning coalition will need smaller sensory input signals in order to be activated than neurons receiving only negative contributions. To put it colloquially: When thinking of the color blue it is easier to spot a blue car in the traffic than a white one.

## 3 COMPETIVE LEARNING

So far we have described, in general terms, the system we are investigating. It has sensitive periods during the transition periods of the continuously ongoing transient-state process, with learning of input signals regulated by a diffusive learning signal. The main two components are therefore the dHan layer and the input layer, as illustrated in Fig. 6.

### *3.1 Input data-stream analysis*

The input signal acts via Eq. (2) on the dHan layer, with the contribution $\Delta r_i$ to the growth rate of the dHan neuron $i$ given by

$$\Delta r_i = \sum_j v_{ij} y_j, \qquad \Delta s_i = \sum_j v_{ij}(1-y_j), \tag{5}$$

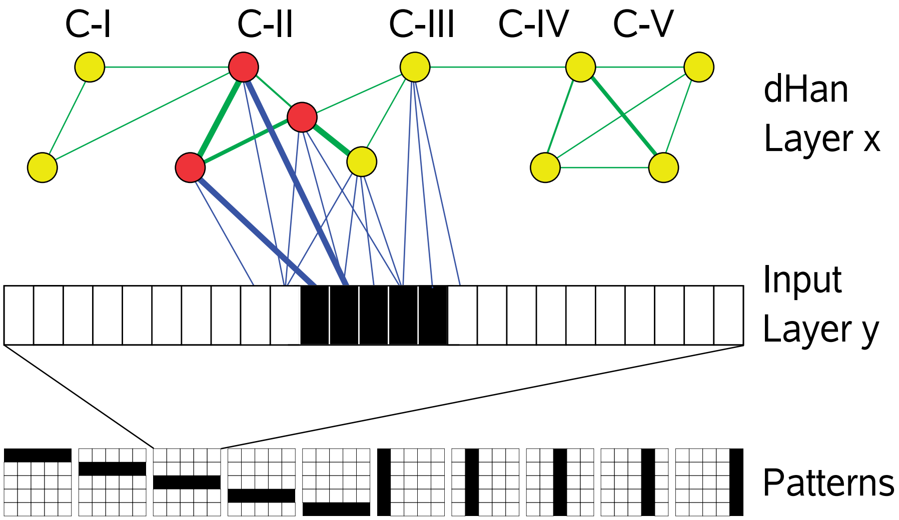

Fig. 6. (Colour online). Schematic representation of the information flow from a raw pattern (bottom) via the input layer (middle) to the dHan layer (top). The synaptic strengths $v_{ij}$ connecting the input with the dHan layer are adapted during the learning process (illustrated selectively by respective thick/thin blue lines in the graph). The dHan layer consists of active and inactive neurons (dark/light circles) connected by intra-layer synaptic weights. The topology shows five cliques (denoted C-I to C-V in the graph) of which C-II is active, as emphasized by the dark circles.

where we have denoted with $y_j \in [0,1]$ the activity-levels of the neurons in the input layer. For subsequent use we have defined in Eq. (5) an auxiliary variable $\Delta s_i$, which quantifies the influence of inactive input-neurons. The task is now to find a suitable learning algorithm which extracts relevant information from the input-data stream by mapping distinct input-patterns onto selected winning coalitions of the dHan layer. This setup is typical for an independent component analysis [26].

The multi-winners-take-all dynamics in the dHan module implies individual neural activities to be close to 0/1 during the transient states and we can therefore define three types of inter-layer links $v_{ij}$ (see Fig. 6):

- active (‘*act*’)
  Links connecting active input neurons with the winning coalition of the dHan module.
- orthogonal (‘*orth*’)
  Links connecting inactive input neurons with the winning coalition of the dHan module.
- inactive (‘*ina*’)
  Links connecting active input neurons with inactive neurons of the dHan module.

The orthogonal links take their name from the circumstance that the receptive fields of the winning coalition of the target layer need to orthogonalize to all input-patters differing from the present one. Note that it is not the receptive field of individual dHan neurons which is relevant, but rather the cumulative receptive field of a given winning coalition.

We can then formulate three simple rules for the respective link-plasticity. Whenever the new winning coalition in the dHan layer is activated by the input layer, *viz* whenever there is a substantial diffusive learning signal, i.e. when $S_{diff}$ exceeds a certain threshold $S_{diff}^c$, the following optimization procedures should take place:

- active links
  The sum over active afferent links should take a large but finite value $r_v^{act}$,

$$\sum_j v_{ij}\, y_j \Big|_{x_i \,\text{active}} \;\to\; r_v^{act} \;.$$

- orthogonal links
  The sum over orthogonal afferent links should take a small value $s_v^{orth}$,

$$\sum_j v_{ij}\, (1-y_j) \Big|_{x_i \,\text{active}} \;\to\; s_v^{orth} \;.$$

- inactive links
  The sum over inactive links should take a small but non-vanishing value $r_v^{ina}$,

$$\sum_j v_{ij}\, y_j \Big|_{x_i \,\text{inactive}} \;\to\; r_v^{ina} \;.$$

The $r_v^{act}$, $r_v^{ina}$ and $s_v^{orth}$ are the target values for the respective optimization processes. In order to implement these three rules we define three corresponding contributions to the link

Table 1. The set of parameters entering the time-evolution equations of the links connecting the input to the dHan layer, with $\Gamma^{+}_{\mathit{diff}}=4.0$ and $\Gamma^{-}_{\mathit{diff}}=0.15$, used in the actual simulations.

| $\Gamma_v^{act}$ | $\Gamma_v^{orth}$ | $\Gamma_v^{ina}$ | $r_v^{act}$ | $s_v^{orth}$ | $r_v^{ina}$ | $x_v^{act}$ | $x_v^{ina}$ | $S^c_{\mathit{diff}}$ |
|---|---|---|---|---|---|---|---|---|
| 0.002 | 0.001 | 0.001 | 0.8 | 0.2 | 0.2 | 0.4 | 0.2 | 0.25 |

plasticities:

$$\begin{aligned} c_i^{act} &= \Gamma_v^{act}\,\Theta(x_i - x_v^{act})\,\mathrm{Sign}(r_v^{act} - \Delta r_i) \\ c_i^{orth} &= \Gamma_v^{orth}\,\Theta(x_i - x_v^{act})\,\mathrm{Sign}(s_v^{orth} - \Delta s_i) \\ c_i^{ina} &= \Gamma_v^{ina}\,\Theta(x_v^{ina} - x_i)\,\mathrm{Sign}(r_v^{ina} - \Delta r_i) \end{aligned} \tag{6}$$

where the inputs $\Delta r_i$ and $\Delta s_i$ to the dHan layer are defined by Eq. (5). For the sign-function $\mathrm{Sign}(x)=\pm 1$ is valid, for $x>0$ and $x<0$ respectively, $\Theta(x)$ denotes the Heaviside-step function. In Eq. (6) the $\Gamma_v^{act}$, $\Gamma_v^{orth}$ and $\Gamma_v^{ina}$ are suitable optimization rates and the $x_v^{act}$ and $x_v^{ina}$ the activity levels defining active and inactive dHan neurons respectively. A suitable set of parameters, which has been used for the numerical simulations, is given in Table 1.

Using these definitions, the link plasticity may be written as

$$\frac{d}{dt} v_{ij} \;=\; \Theta(S_{\mathit{diff}} - S^c_{\mathit{diff}})\Big[c_i^{act}\, y_j + c_i^{orth}\,(1-y_j) + c_i^{ina}\, y_j\Big]\,, \tag{7}$$

where $S^c_{\mathit{diff}}$ is an appropriate threshold for the diffusive learning signal. The inter-layer links $v_{ij}$ cease to be modified whenever the total input is optimal, *viz* when no more ‘mistakes’ are made [27].

We note, that a given interlayer-link $v_{ij}$ is in general subject to competitive optimization from the three processes (act/orth/ina). Averaging would occur if the respective learning rates $\Gamma_v^{act}/\Gamma_v^{orth}/\Gamma_v^{ina}$ would be of the same order of magnitude. It is therefore necessary, that $\Gamma_v^{act} \gg \Gamma_v^{orth}$ and $\Gamma_v^{act} \gg \Gamma_v^{ina}$.

### *3.2 Homeostatic normalization*

It is desirable that the interlayer connections $v_{ij}$ neither grow unbounded with time (runaway-effect) nor disappear into irrelevance. Suitable normalization procedures are therefore normally included explicitly into the respective neural learning rules and are present implicitly in Eqs. (6) and Eq. (7).

The strength of the input-signal is optimized by Eq. (7) both for active as well as for inactive dHan neurons, a property referred to as fan-in normalization. Eqs. (6) and (7) also regulate the overall strength of inter-layer links emanating from a given input layer neuron, a property called fan-out normalization.

Next we note, that the time scales for the intrinsic autonomous dynamics in the dHan layer and for the input signal could in principle differ substantially. Potential interference problems can be avoided when learning is switched-on very fast. In this case the activation and decay rates $\Gamma^{\pm}_{\mathit{diff}}$ for the diffusive learning signal are large and the corresponding characteristic time scales $1/\Gamma^{\pm}_{diff}$ are smaller than both the typical time scales of the input and of the self-sustained dHan dynamics.

## 4 THE BARS PROBLEM

A cognitive system needs to extract autonomously meaningful information about its environment from its sensory input data stream via signal separation and feature extraction. The identification of recurrently appearing patterns, i.e. of objects, in the background of fluctuation and of combinations of distinct and noisy patterns, constitutes a core demand in this context. This is the domain of the independent component analysis [26] and blind source separation [28], which seeks to find distinct representations of statistically independent input patterns.

In order to test our system made-up by an input layer coupled to a dHan layer, as illustrated in Fig. 6, we have selected the bars problem [29, 30]. The bars problem constitutes a standard non-linear reference task for the feature extraction via a non-linear independent component analysis for an $L\times L$ input layer. Basic patterns are the $L$ vertical and $L$ horizontal bars. The individual input patterns are made-up of a non-linear superposition of the $2L$ basic bars, containing with probability $p=0.1$ any one of them, as illustrated in Fig. 7.

### *4.1 Simulations and setup*

For the simulations we presented to the system about $N_{patt}\approx 5\times 10^3$ randomly generated $5\times 5$ input patterns of the type shown in Fig. 7. The bars pattern are black/white with the $y_i=1/0$ for active/inactive sites, irrespectively of possible overlaps of vertical and horizontal bars. The individual patterns lasted $T_{patt}=20$ with about $T_{inter}=100$ for the time between two successive input signals. These time scales are to be compared with the time scale of the autonomous dHan dynamics illustrated in the Figs. 4 and 5, for which the typical stability-period for a transient state is about $t_{trans}\approx 70$. We also note that there is no active training for the system. The associative thought process continuous in the dHan layer, at no time are the neural activities reset and the system restarted. All that happens is that the ongoing associative thought process is influenced from time to time by the input layer and that then the synaptic strengths $v_{ij}$ connecting the input layer to the dHan layer are modified via Eq. (7).

The results for the simulation are presented in Fig. 8. For the geometry of the dHan network we used a regular 20-site star containing 10 cliques, with every clique being composed

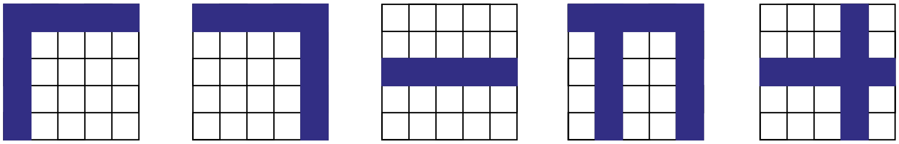

FIG. 7. (Colour online). Examples of typical input patterns for a $5\times 5$ bars problem with a probability $p=0.1$ for the occurrence of the individual horizontal or vertical bars. The problem is non-linear since the pattern intensity is not enhanced when an elementary horizontal and vertical bar overlap each other.

of four neurons, see Fig. 2. In Fig. 8 we present the response

$$R(\alpha,\beta) \;=\; \frac{1}{S(C_\alpha)}\sum_{i\in C_\alpha,j} v_{ij}\,y_j^\beta, \qquad\qquad \begin{array}{l}\alpha=1,..,10\\ \beta=1,..,10\end{array} \tag{8}$$

of the 10 cliques $C_\alpha$ in the dHan layer to the 10 basic input patterns $\{y_j^\beta, j=1,..,25\}$, the isolated bars. Here the $C_\alpha \in \{\text{C-I,..,C-X}\}$ denotes the set of sites of the winning-coalition $\alpha$ and $S(C_\alpha)$ its size, here $S(C_\alpha)\equiv 4$. The response $R(\alpha,\beta)$ is equivalent to the clique averaged afferent synaptic signals $\Delta r_i$, compare Eq. (5), in the presence of an elementary bar in the sensory input field.

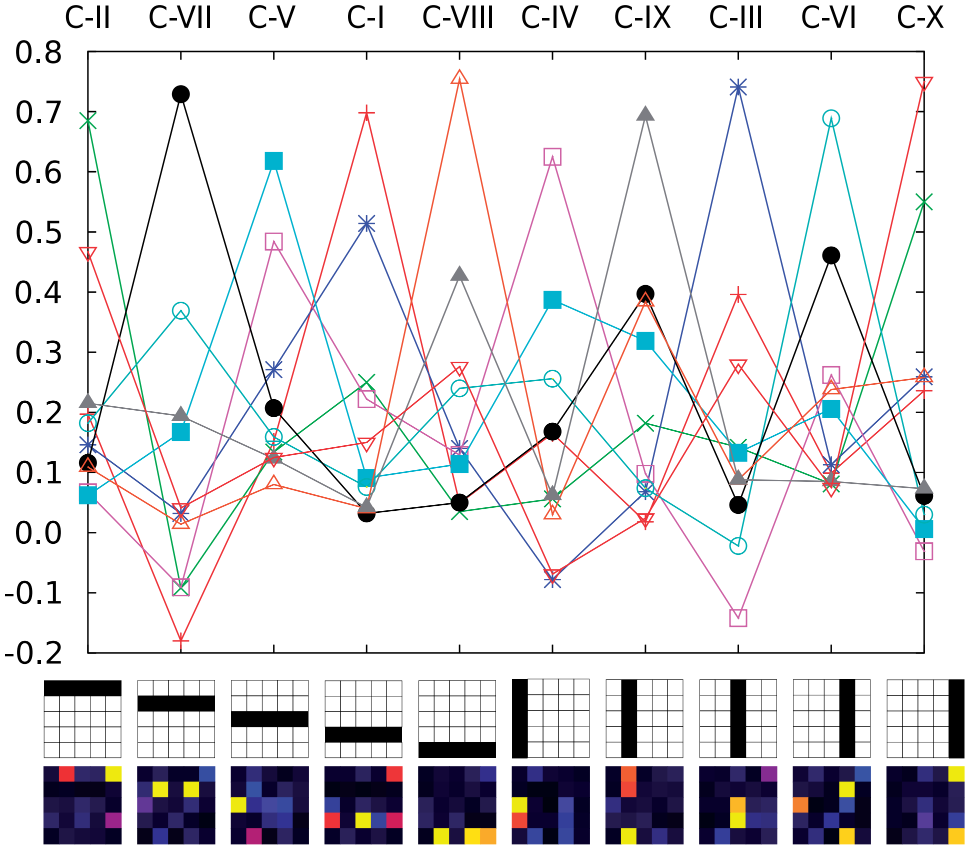

FIG. 8. (Colour online). For the 5 × 5 bars problem the response, as defined by Eq. (8), for the 10 winning coalitions in the dHan layer (compare Fig. 2 and Fig. 6) to the ten reference patterns, *viz* the 5 horizontal bars and the 5 vertical bars of the 5 × 5 input field. In the top row the numbering of the cliques C-I,..,C-X having the maximal response to the respective reference patterns is given. In the bottom row, below each of the 10 black/white reference patterns, the receptive fields, Eq. (9), for the winning coalitions C-I,..,C-X given in the top row. The strength of the receptive fields are colour-coded, with black/dark/white (black/blue/red/yellow online) denoting synaptic strengths of increasing intensities.

### *4.2 Semantic learning*

The individual potential winning coalitions, *viz* the cliques, have acquired in the course of the simulation, via the learning rule Eq. (7), distinct susceptibilities to the 10 bars, compare Fig. 8. At the start of the simulation the winning coalitions were just given by properties of the network typology, *viz* by the cliques, having no explicit semantic significance. The susceptibilities to the individual bars, which the cliques have acquired via the competition of the internal dHan dynamics with the sensory data input stream, can then be interpreted as a semantic assignment. The internal associative thought process of the dHan layer therefore becomes semantically meaningful via the coupling to the environment, corresponding to a sequence of horizontal and vertical bars. This learning paradigm is compatible with multi-electrode array studies of the visual cortex of developing ferrets [1], which indicate that the ongoing cortical dynamics is void of semantic content immediately after birth, acquiring semantic content however during the adolescence.

### *4.3 Competitive learning*

The winning coalitions of the dHan layer are overlapping and every link $v_{ij}$ targets in general more than one potential winning coalition in the dHan layer. This feature contrasts with the 'single-winner-takes-all' setup, normally used for standard neural algorithms performing an independent component analysis [26], for which the target neurons are physically separated. For the regular 20-site network used in the simulation every dHan neuron appertains to exactly two cliques, compare Fig. 2. The unsupervised learning procedure, Eq. (7), involves therefore a competition between the contribution $c_i^{act}$, $c^{orth}$ and $c^{ina}$, as given by Eq. (6). For the simulations we used a set of parameters, see Table 1, for which the contribution to $c_i^{act}$ is adapted at a much higher rate than the contributions to $c^{orth}$ and $c^{ina}$. The responses $R(\alpha,\beta)$ of the winning coalitions are therefore close to, but somewhat below, the optimal value $r_v^{act}=0.8$ used for the simulations, compare Fig. 8. The target value $r_v^{act}$ will not be reached even for extended simulations, due to the competition with the other optimization procedures, namely $c^{orth}$ and $c^{ina}$, compare Eq. (6).

### *4.4 Receptive fields*

The averaged receptive fields

$$F(\alpha,j) \;=\; \frac{1}{S(C_\alpha)}\sum_{i\in C_\alpha} v_{ij}, \qquad \alpha=1,..,10, \tag{9}$$

of the $\alpha=1,...,10$ cliques in the dHan layer with respect to the $j=1,...,25$ input neurons are also presented in Fig. 8. The inter-layer synaptic weights $v_{ij}$ can be both positive and negative and the orthogonalization procedure, Eq. (6), results in complex receptive fields. The time evolution equations for the inter-layer synaptic strengths (7) are optimizing, but not maximizing, the response of the winning coalition to a given input signal. The receptive fields retain consequently a certain scatter, since the optimization via Eq. (7) ceases whenever a satisfactory signal separation has been obtained. This behavior is consistent with the 'learning by mistakes' paradigm [27], which states that a cognitive system needs to learn in general only when committing a mistake.

## *4.5 Emergent cognitive capabilities*

The simulation results for the $5\times5$ bars problem presented in Fig. 8 may be generalized to larger systems. For comparison we discuss now the results for the $10\times10$ bars problem, for which there are ten horizontal and ten vertical elementary bars. For the dHan network we used a regular 40-site star with 20 cliques, a straightforward generalization of the regular star illustrated in Fig. 2. We used otherwise exactly the same set of parameters as previously for the $5\times5$ bars problem, in particular also the same number of input training patterns. No optimization of parameters has been performed. The respective responses $R(\alpha,\beta)$ and receptive fields $F(\alpha,j)$ (compare Eqs. (8) and (9)) are presented in Fig. 9 and 10.

The probability for any of the 20 bars to occur in a given input pattern, like the ones for the $5\times5$ bars problem illustrated in Fig. 7, is $p=0.1$ and any individual $10\times10$ input patterns contains on the average $\approx$2.3 bars superposed non-linearly. The separation of the 20 statistically independent components in the input data stream is therefore a non-trivial task. The results presented in Fig. 9 indicate that the system performs the source separation surprisingly well, but not perfectly. The respective receptive fields, shown in Fig. 10, are only in part self-evident. This is, again, due to the competitive nature of the unsupervised and local learning process, which has the task to optimize the input rates to the dHan layer and not to maximize the signal-to-noise ratio. We note in this context that the system contains no prior knowledge about the nature and statistics of the input signals.

In fact, the system has not been constructed in the first place to tackle the non-linear independent component task. The setup used here has been motivated by two simple guiding principles, the occurrence of self-sustained internal neural activity and the principle of

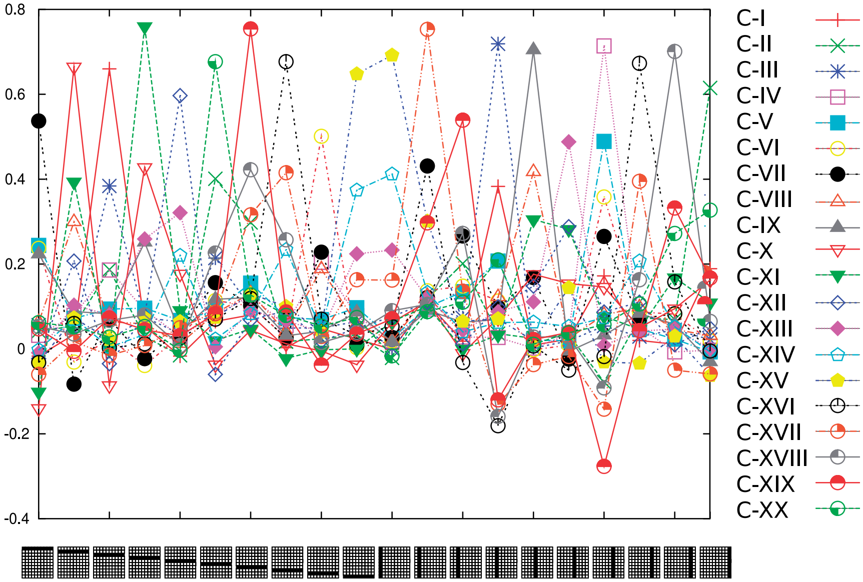

FIG. 9. (Colour online). For the $10\times10$ bars problem the response, as defined by Eq. (8), for the 20 winning coalitions C-I,..,C-XX in the dHan layer (compare Fig. 2 and Fig. 6) to the twenty reference patterns, *viz* the 10 horizontal bars and the 10 vertical bars of the $10\times10$ input field.

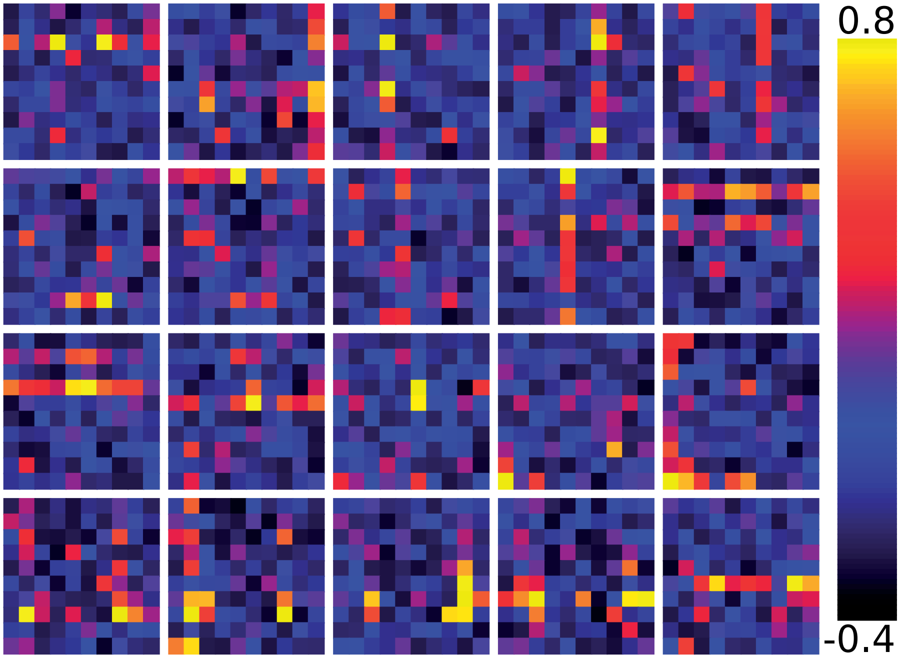

FIG. 10. For the $10 \times 10$ bars problem the colour-coded receptive fields, Eq. (9), for the 20 cliques C-I,..,C-XX of the dHan layer, compare Fig. 9, with black/dark/white (black/blue/red/yellow online) encoding synaptic strengths of increasing intensities.

competitive neural dynamics. These principles have been used in our study to examine the interplay of the self-sustained internal neural dynamics with the inflow of external information via a sensory data stream. One can therefore interpret, to a certain extend, the capability of the system to perform a non-linear independent component analysis as an example of an 'emergent cognitive capability'. This information processing capability emerges from general construction principles and does not result from the implementation of a specific neural algorithm.

## 5 DISCUSSION AND CHALLENGES

### *5.1 Discussion*

A standard approach in the field of neural networks is to optimize the design of a network such that a given cognitive or computational task can be tackled efficiently. This strategy has been very successful in the past with respect to technical applications like handwriting recognition [31] and regarding the modeling of initial feed-forward sensory information processing in cortical areas like the primary optical cortex [32]. Task-driven network design standardly results in input-driven neural networks, with cognitive computation coming to a standstill in the absence of sensory inputs.

Real-world cognitive systems like the human brain are however driven by their own internal dynamics and it constitutes a challenge to present and to future research in the field of neural networks to combine models of this self-sustained brain activity with the processing of sensory data. This challenge regards especially recurrent neural networks, since recurrency is an essential ingredient for the occurrence of spontaneous internal neural activities.

In this work we studied the interplay of self-generated neural states, the time-series of winning coalitions, with the sensory input for the purpose of unsupervised feature extraction. We proposed learning to be autonomously activated during the transition from one winning coalition to the subsequent one.

This general principle may be implemented algorithmically in various fashions. Here we used a generalized neural net (dHan - dense homogeneous associative net) for the autonomous generation of a time series of associatively connected winning coalitions and controlled the unsupervised extraction of input-features by an autonomously generated diffusive learning signal.

We tested the algorithm for the bars problem and found good and fast learning and that the initially semantically void transient states acquired, through interaction with the data input stream, a semantic significance. Further preliminary results indicate that the learning algorithm retains functionality under a wide range of conditions and for various sets of parameters. We plan to extend the simulations to various forms of temporal inputs, especially to quasi-continuous input and to natural scene analysis, and to study the embedding of the here proposed concept within the framework of a full-fledged and autonomously active cognitive system.

### *5.2 The overall perspective*

There is a growing research effort trying to develop universal operating principles for biologically inspired cognitive systems, the rational being, that the number of genes in the human genome is by far too small for the detailed encoding of the fast array of neural algorithms the brain is capable off. There is therefore a growing consensus, that universal operating principles may be potentially of key importance also for synthetic cognitive and complex systems [33, 34]. The present work is motivated by this line of approach.

Universal operating principles for a cognitive system remain functionally operative for a wide range of environmental conditions. Examples are, universal time prediction tasks for the unsupervised extraction of abstract concepts and intrinsic generalized grammars from the sensory data input stream [8, 35, 36] and the optimization of complexity and information theoretical measures for closed-loop sensorimotor behavioral studies of simulated robots [37–39]. The present study is motivated by a similar line of thinking, investigating the consequences of a self-sustained internal neural activity in recurrent networks, being based on the notion of transient-state and competitive neural dynamics. The long-term goal of an autonomous cognitive system is pursued in this approach via a modular approach, with each module being based on one of the above mentioned general architectural and operational principles.

## References

[1] J. Fiser, C. Chiu and M. Weliky, *Small modulation of ongoing cortical dynamics by sensory input during natural vision*, Nature 431 (2004) 573–578.

[2] M. Abeles *et al.*, *Cortical activity flips among quasi-stationary states*, PNAS 92 (1995) 8616–8620.

[3] D.L. Ringach, *States of mind*, Nature 425 (2003) 912–913.

[4] T. Kenet, D. Bibitchkov, M. Tsodyks, A. Grinvald and A. Arieli, *Spontaneously emerging cortical representations of visual attributes*, Nature 425 (2003) 954–956.

[5] J.S. Damoiseaux, S.A.R.B. Rombouts, F. Barkhof, P. Scheltens, C.J. Stam, S.M. Smith and C.F. Beckmann, *Consistent resting-state networks across healthy subjects*, PNAS 103 (2006) 13848–13853.

[6] C.J. Honey, R. Kötter, M. Breakspear and O. Sporns, *Network structure of cerebral cortex shapes functional connectivity on multiple time scales*, PNAS 104 (2007) 10240–10245.

[7] R. VanRullen and C. Koch, *Is perception discrete or continuous?*, Trends in Cognitive Sciences 5 (2003) 207–213.

[8] C. Gros, *Complex and Adaptive Dynamical Systems, A Primer*, Springer 2008.

[9] G.M. Edelman and G.A. Tononi, *A Universe of Consciousness*, New York: Basic Books 2000.

[10] G.M. Edelman, *Naturalizing consciousness: A theoretical framework*, PNAS 100 (2003) 5520–5524.

[11] S. Dehaene and L. Naccache, *Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework*, Cognition 79 (2003) 1–37.

[12] B.J. Baars and S. Franklin, *How conscious experience and working memory interact*, Trends Coginitve Science 7 (2003) 166–172.

[13] F.C. Crick and C. Koch, *A framework for consciousness*, Nature Neuroscience 6 (2003) 119–126.

[14] C. Koch, *The Quest for Consciousness - A Neurobiological Approach*, Robert and Company 2004.

[15] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall 1994.

[16] J.J. Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, PNAS 79 (1982) 2554–2558.

[17] C. Gros, *Cognitive computation with autonomously active neural networks: An emerging field*, Cognitive Computation (2009, in press).

[18] J.J. Eggermont, *Is There a Neural Code?*, Neuroscience and Biobehavioral Reviews 22 (1998) 355–370.

[19] H. Jaeger and H. Haas, *Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication*, Science 304 (2004) 78–80.

[20] W. Maass, T. Natschlager and H. Markram, *Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations*, Neural Computation 14 (2002) 2531–2560.

[21] M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H.D.I. Abarbanel and G. Laurent, *Dynamical Encoding by Networks of Competing Neuron Groups: Winnerless Competition*, Physical Review Letters 87 (2001) 68102.

[22] B.A. Olshausen and D.J. Field, *Sparse coding of sensory inputs*, Current Opinion in Neurobiology 14 (2004) 481–487.

[23] L. Lin, R. Osan and J.Z. Tsien, *Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes*, Trends in Neurosciences 29 (2006) 48–57.

[24] C. Gros, *Self-Sustained Thought Processes in a Dense Associative Network*, in KI 2005, U. Furbach (Ed.), Springer Lecture Notes in Artificial Intelligence 3698 (2005) 366–379; also available as http://arxiv.org/abs/q-bio.NC/0508032.

[25] C. Gros, *Neural networks with transient state dynamics*, New Journal of Physics 9 (2007) 109.

[26] A. Hyvärinen and E. Oja, *Independent component analysis: Algorithms and applications*, Neural Networks 13 (2000) 411–430.

[27] D.R. Chialvo and P. Bak, *Learning from mistakes*, Neuroscience 90 (1999) 1137–1148.

[28] S. Choi, A. Cichocki, H.M. Park and S.Y. Lee, *Blind Source Separation and Independent Component Analysis: A Review*, Neural Information Processing 6 (2005) 1–57.

[29] P. Földiák, *Forming sparse representations by local anti-Hebbian learning*, Biological Cybernetics 64 (1990) 165–170.

[30] N. Butko and J. Triesch, *Learning Sensory Representations with Intrinsic Plasticity*, Neurocomputing 70 (2007) 1130–1138.

[31] G. Dreyfus, *Neural Networks: Methodology and Applications*, Springer, 2005.

[32] M.A. Arbib, *The Handbook of Brain Theory and Neural Networks*, MIT Press 2002.

[33] C. Müller-Schloer, C. von der Malsburg, und R.P. Würtz, '*Organic Computing*', Informatik Spektrum 27 (2004) 2–6.

[34] R.P. Würtz, *Organic Computing*, Springer Verlag, 2008.

[35] J.L. Elman, *Finding structure in time*, Cognitive Science 14 (1990) 179–211.

[36] J.L. Elman, *An alternative view of the mental lexicon*, Trends in Cognitive Sciences 8 (2004) 301–306.

[37] A.K. Seth and G.M. Edelman, *Environment and Behavior Influence the Complexity of Evolved Neural Networks*, Adaptive Behavior 12 (2004) 5–20.

[38] O. Sporns and M. Lungarella, *Evolving coordinated behavior by maximizing information structure*, in L. Rocha et al. (eds), Proceedings of Artificial Life X (2006) 3–7.

[39] N. Ay, N. Bertschinger, R. Der, F. Güttler and E. Olbrich, *Predictive information and explorative behavior of autonomous robots*, European Physical Journal B 63 (2008) 329–339.