

# Syntactic Sequencing in Hebbian Cell Assemblies\*

Thomas Wennekers<sup>1</sup> and Günther Palm<sup>2</sup>

<sup>1</sup> Corresponding author

Centre for Theoretical and Computational Neuroscience  
University of Plymouth, PL4 8AA Plymouth, United Kingdom

Tel: +44 1752 23 3593

Fax: +44 1752 23 3349

thomas.wennekers@plymouth.ac.uk

<sup>2</sup> Department of Neural Information Processing

University of Ulm, 89069 Ulm, Germany

Email: guenther.palm@uni-ulm.de

August 31, 2009

---

\*Journal of Cognitive Neurodynamics, in press

## Abstract

Hebbian cell assemblies provide a theoretical framework for the modeling of cognitive processes that grounds them in the underlying physiological neural circuits. Recently we have presented an extension of cell assemblies by operational components which allows to model aspects of language, rules, and complex behaviour. In the present work we study the generation of syntactic sequences using operational cell assemblies timed by unspecific trigger signals. Syntactic patterns are implemented in terms of hetero-associative transition graphs in attractor networks which cause a directed flow of activity through the neural state space. We provide regimes for parameters that enable an unspecific excitatory control signal to switch reliably between attractors in accordance with the implemented syntactic rules. If several target attractors are possible in a given state, noise in the system in conjunction with a winner-takes-all mechanism can randomly choose a target. Disambiguation can also be guided by context signals or specific additional external signals. Given a permanently elevated level of external excitation the model can enter an autonomous mode, where it generates temporal grammatical patterns continuously.

**Keywords:** Cell Assemblies; Attractor Networks; Grammar; Language; Behaviour

# 1 Introduction

Hebbian Cell Assemblies have been proposed as a model for cognitive processes that on one hand aims at functional explanations of psychological phenomena, but on the other hand tries to link them to physiological processes going on in neural circuits in the brain [25]. Related studies have mostly modeled aspects of object memory, technically realised by storing patterns in attractor neural networks [4, 5, 44]. Recent experimental findings support the basic framework of assemblies, see, e.g., [24, 58].

More complex spatio-temporal patterns and processes as compared to simple fixed point attractors have also received attention. Lashley [37] early pointed out the need for neural mechanisms that allow for complex temporal sequencing in neural activity in order to explain behaviour and language. These ideas have been further developed conceptually by Wickelgren, von Neumann, and others [3, 25, 43, 44, 47, 65, 66]. Central to computational models of sequences are hetero-associative connections between pools of neurons that can evoke a wave-like directed propagation of activity along specific synaptic pathways. In one such type of models, so-called “synfire-chain” models, hetero-associative connections alone suffice to implement sequences. Such studies date back to Abeles [1] and aim at explaining highly precise temporal patterns in neural firing times [2, 8, 14, 26, 60].

A second class of sequence models starts from attractor networks where a number of patterns are stored auto-associatively, but additional hetero-associative connections are used to allow for transitions between patterns, see, e.g., [30, 44, 55, 69]. These models often employ additional mechanisms to destabilise attractor patterns and to evoke autonomous transitions along the hetero-associatively stored sequences of attractors. For example, Horn and Usher [30] used adapting neurons to cause the transitions. They also provide a mean-field type analysis of such networks with adapting neurons. Rehn and Lansner [49] present a network where synaptic depression causes pattern transitions. Their model explains participant’s performance in free-recall tasks of ordered and random lists of items.

Although the physiological reality of precise hetero-associative firing patterns and synfire chains is still being discussed, some experiments lend support to the idea of a more or less deterministic directed activation flow through neural circuits as compared to merely stochastic firing sequences [13, 22, 31]. Specific sequential spatio-temporal firing patterns have especially been observed in the song production system of birds [22] where they apparently correspond to representations of syllables and their sequencing into songs [20, 21, 33].

It is still an open problem how cell assembly networks can support language and behaviour, both domains requiring the rule-based syntactic concatenation of temporal elements into more complex compounds. The respective mechanisms have to take interactions with the environment into account (or rather with the environment’s representation

in the brain). Some relevant physiology-inspired models for language and behaviour have been built, in part based on the concept of “mirror neurons” [6, 7, 18, 47, 48, 50]. This includes advanced models in robotics [34, 38, 71].

We have proposed a generic approach addressing mechanisms for complex behaviour and language in cell assembly networks by extending them by rule-based “operational” principles [61], see also [63, 64]. Operational cell assemblies combine attractor networks with rules and finite state automata in a simple and plausible manner by allowing for input-dependent transitions between attractors. This results in an intuitive picture of brain processes as occurring in a state space with a graph-like structure of transitions, where states are attractors of a local neural network dynamics and edges are hetero-associative transitions between attractors that are triggered if certain input events arrive. Such systems would describe processes going on locally in an area or column. The full complexity of cognition would then arise from many such modules interacting. Operational cell assemblies of this kind have been used to implement language and behavioural components on robots [17, 34, 35, 38–40]. They can also well be implemented in modern neural hardware [32, 53, 67] and may therefore form a programming paradigm for future computing hardware [44, 61].

Our previous work considered only systems, where transitions between assemblies were purely determined by specific input sequences. Either different syntactic input patterns (“words”) were recognised based on the order of arriving “syllables”, or different objects in a visual input uniquely triggered the generation of syntactically structured spatiotemporal outputs (“object naming”) [61].

In contrast, in the present work we consider the generation of syntactic spatio-temporal output patterns in cell assemblies without specific inputs. This may apply to the generation of free speech, free thinking, or other autonomous behaviours.

One practical difficulty in this context relates to the fact that the speed of speech and behaviour can be modulated in wide ranges. It is not obvious at all how neural circuits can support this. Earlier models for sequencing and timing exist, but the issue is far from being resolved, see [70] for a review. Sommer and Wennekers [55] proposed a switching mechanism to explicitly trigger transitions in a sequence network comprising conductance based neurons. The triggering signal was entirely unspecific with respect to any information stored in the networks but rather aimed at setting the pace of retrieval. We here propose a similar solution in the broader framework of operational cell assembly networks, where syntactic rules are implemented in form of hetero-associative transitions between assemblies as indicated earlier [61]. These can, for instance, form word networks or sentence networks. We then study how unspecific excitatory input can trigger transitions between attractors in such networks. Such a trigger would again not contain any information about the sequences to be generated; rather, it would serve as a threshold control that at some pace evokes transitions along the embedded syntactic

pathways. Valentino Braitenberg has suggested such a “pump of thought” [9]; we here show a possible realisation. The idea of triggered transitions separates the problem of regulating the speed of transitions from the problem of creating the structural properties of possible transitions in the network.

## 2 State switching by unspecific excitatory input

In the present section we study the most simple setup of state switching by unspecific excitatory input between latched attractor states. We assume that a single chain of attractor patterns is embedded in a network and linked by hetero-associative connections into a cyclic chain. Transitions between patterns can be caused at (almost) arbitrary times by entirely unspecific inputs that use no extra information beyond that stored in the hetero-associative connections. Analytical conditions for stable switching are derived.

### 2.1 Basic Model

We consider a network of  $n$  excitatory and  $n$  inhibitory graded response neurons. The excitatory neurons are described by low-pass membrane potentials  $x_i$  with time-constant  $\tau$  and an intrinsic cell-adaptation  $a_i$  per cell with time-constant  $\tau_a$  that decreases a cell’s excitability during periods of elevated firing. The neurons receive an input  $I_i$  and their output is computed by a sigmoid firing rate function as  $z_i = f(x_i - a_i)$ . A possible choice for  $f$  is the unit-step or Heaviside function,  $f(x) = 1$  if  $x \geq 0$  and  $f(x) = 0$  else, but other sufficiently steep functions work as well. The inhibitory cells ( $y_i$ ) are assumed to be linear for simplicity; they have a membrane time constant  $\tau_2$  but do not adapt. The dynamic equations of the model read

$$\tau \dot{x}_i = -x_i + I_i - \frac{1}{n} \sum_{j=1}^n D_{ij} y_j + \frac{1}{m} \sum_{j=1}^n (A_{ij} + H_{ij} + R_{ij}) z_j \quad (1)$$

$$\tau_a \dot{a}_i = -a_i + b z_i \quad (2)$$

$$\tau_2 \dot{y}_i = -y_i + \frac{1}{m} \sum_{j=1}^n C_{ij} z_j \quad (3)$$

$$z_i = f(x_i - a_i) \quad (4)$$

In (1)  $A = (A_{ij})$  and  $H = (H_{ij})$  are auto- and hetero-associative coupling matrices, respectively.  $D = (D_{ij})$  and  $R = (R_{ij})$  in (1) and  $C = (C_{ij})$  in (3) are matrices of random synapses. Their entries are assumed to be independent and identically distributed random variables with means  $E[R_{11}] = r$ ,  $E[C_{11}] = c$ ,  $E[D_{11}] = d$  and variances  $D[R_{11}] = \sigma_r^2$ ,  $D[C_{11}] = \sigma_c^2$  and  $D[D_{11}] = \sigma_d^2$ .

The auto- and hetero-associative connections are set up between neurons in a set of binary memory patterns  $\{\xi^\alpha, \alpha = 1, \dots, P\}$  where the  $\xi_i^\alpha \in \{0, 1\}$  have ones at random positions and zeroes else. We assume that each pattern has exactly  $m$  ones.  $A$  and  $H$  are given by

$$A_{ij} = a \sum_{\alpha=1}^P \xi_i^\alpha \xi_j^\alpha \quad (5)$$

$$H_{ij} = h \sum_{\alpha=1}^P \xi_i^{\alpha+1} \xi_j^\alpha. \quad (6)$$

In (6) and elsewhere we implicitly assume that the hetero-associative chain of patterns is cyclically closed, i.e.,  $\xi^{P+1} = \xi^1$ . Parameters  $a$  and  $h$  set the relative strength of the auto- and hetero-associative connections. Note that here and throughout the paper upper Greek indexes denote patterns  $(1, \dots, P)$  and lower Roman indexes units  $(1, \dots, N)$ . Where they are obvious we often don't provide summation ranges in sums.

We define the overlap matrix between patterns as  $U^{\alpha\beta} := \frac{1}{m} \sum_i \xi_i^\alpha \xi_i^\beta$ . Because the patterns are random and each has  $m$  ones we have  $U^{\alpha\beta} \approx \delta^{\alpha\beta} + \mathcal{O}(\frac{m}{n})$ . The second term reflects overlaps between patterns and can be made small by using sparse patterns with  $p := m/n \ll 1$ .

## 2.2 Derivation of meanfield equations

We further define the following pattern specific averages of the model variables

$$h^\alpha := \frac{1}{m} \sum_{i=1}^n \xi_i^\alpha x_i \quad (7)$$

$$m^\alpha := \frac{1}{m} \sum_{i=1}^n \xi_i^\alpha z_i \quad (8)$$

$$u^\alpha := \frac{1}{m} \sum_{i=1}^n \xi_i^\alpha y_i \quad (9)$$

$$w^\alpha := \frac{1}{m} \sum_{i=1}^n \xi_i^\alpha a_i \quad (10)$$

The  $m^\alpha$  are often called ‘‘overlaps’’ because they measure how close the firing pattern  $z$  is to a stored pattern  $\xi^\alpha$ ;  $m^\alpha$  will take its maximum value of 1 only if all ones in  $\xi^\alpha$  are also fully active in  $z$ .

We now derive dynamic equations for these averaged variables from the original equations (1) to (4). For this we use the approximation

$$z_i \approx \sum_{\beta} m^\beta \xi_i^\beta. \quad (11)$$

Equation (11) states that the individual neuron activations can be written as a superposition of the memory patterns weighted by the current activation strength  $m^\beta$  of each pattern. This approximation reflects that the dynamics of the system will either stay in fully excited memory states (auto-associative mode) or move between such patterns (switched mode).

For the pattern specific adaptation we then get

$$\tau_a \dot{w}^\alpha = -w^\alpha + bm^\alpha \quad (12)$$

If we define the total activity as  $s := \sum_\beta m^\beta$  the inhibitory dynamics results in

$$\tau_2 \dot{u}^\alpha = -u^\alpha + \frac{1}{m^2} \sum_{i,j,\beta} \xi_i^\alpha C_{ij} m^\beta \xi_j^\beta \quad (13)$$

$$\approx -u^\alpha + \frac{c}{m} \sum_{j,\beta} m^\beta \xi_j^\beta = -u^\alpha + c \sum_\beta m^\beta = -u^\alpha + cs. \quad (14)$$

Note that the right hand side in (14) defines the same low-pass dynamics for all  $\alpha$ . Therefore, the pattern-specific inhibition variables all become similar asymptotically in time after initial transients have died out. (They may even become identical if we would further assume  $C_{ij} = c = \text{constant}$  or in the asymptotic limit of infinite pattern size,  $m \rightarrow \infty$ .) We could therefore just have started with a single globally acting linear inhibitory neuron. This reflects that the main purpose of the inhibitory neurons is to measure the total activity of the excitatory cells and provide a proportional dynamic threshold control.

The averaged excitatory potentials become

$$\tau \dot{h}^\alpha = -h^\alpha + I^\alpha - \frac{1}{mn} \sum_{i,j} \xi_i^\alpha D_{ij} u_j \quad (15)$$

$$+ \frac{1}{m^2} \sum_{i,j,\gamma} \xi_j^\gamma \left[ \sum_\beta (a \xi_i^\alpha \xi_i^\beta \xi_j^\beta + h \xi_i^\alpha \xi_i^{\beta+1} \xi_j^\beta) + \xi_i^\alpha R_{ij} \right] m^\gamma \quad (16)$$

$$= -h^\alpha + I^\alpha - \frac{1}{mn} \sum_{i,j} \xi_i^\alpha D_{ij} u_j + \frac{a}{m^2} \sum_{i,j,\beta,\gamma} \xi_j^\gamma \xi_i^\alpha \xi_i^\beta \xi_j^\beta m^\gamma \quad (17)$$

$$+ \frac{h}{m^2} \sum_{i,j,\beta,\gamma} \xi_j^\gamma \xi_i^\alpha \xi_i^{\beta+1} \xi_j^\beta m^\gamma + \frac{1}{m^2} \sum_{i,j,\beta,\gamma} \xi_i^\alpha \xi_j^\gamma R_{ij} m^\gamma \quad (18)$$

$$\approx -h^\alpha + I^\alpha - du^\alpha + a \sum_{\beta,\gamma} U^{\gamma\beta} U^{\alpha\beta} m^\gamma + h \sum_{\beta,\gamma} U^{\gamma\beta} U^{\alpha\beta+1} m^\gamma + rs \quad (19)$$

Now, remember that  $U^{\alpha\beta} \approx \delta^{\alpha\beta} + \mathcal{O}(p)$  where  $p = m/n \ll 1$ . Inserting this into (19) leads to

$$\tau \dot{h}^\alpha \approx -h^\alpha + I^\alpha - du^\alpha + rs \quad (20)$$

$$+ a \sum_{\beta, \gamma} \left( \delta^{\gamma\beta} \delta^{\alpha\beta} + \mathcal{O}(p)(\delta^{\gamma\beta} + \delta^{\alpha\beta}) \right) m^\gamma \quad (21)$$

$$+ h \sum_{\beta, \gamma} \left( \delta^{\gamma\beta} \delta^{\alpha\beta+1} + \mathcal{O}(p)(\delta^{\gamma\beta} + \delta^{\alpha\beta+1}) \right) m^\gamma + \mathcal{O}(p^2) \quad (22)$$

$$= -h^\alpha + I^\alpha - du^\alpha + am^\alpha + hm^{\alpha-1} + rs + \mathcal{O}(p)2(a+h)s . \quad (23)$$

Equation (23) shows that the pattern averaged potentials will follow a low-pass dynamics.  $I^\alpha$  is the averaged input and  $du^\alpha$  the effect of the inhibitory pool of neurons. Clearly, pattern  $\alpha$  is only influenced by effective auto-associative connections  $am^\alpha$  from neurons within the pattern and hetero-associative connections from patterns in the previous pattern,  $hm^{\alpha-1}$ . The expressions  $rs$  and  $\mathcal{O}(p)2(a+h)s$  in (24) are noise due to random synapses between excitatory cells and cross-talk noise between patterns, respectively. They are both proportional to  $s$  and should be small. Because  $a$ ,  $h$ , and  $r$  are all positive we may integrate their effect into a single constant  $r \leftarrow r + 2(a+h)$  that accounts for the maximum impact of the noise sources.

$$\tau \dot{h}^\alpha = -h^\alpha + I^\alpha - du^\alpha + am^\alpha + hm^{\alpha-1} + rs \quad (24)$$

Note that  $u^\alpha$  by means of (14) is a low-passed version of the total activity  $s$  as well. It should therefore be possible to balance the mean impact of the noise ( $rs$ ) by a proper choice of the inhibitory coupling strength  $d$ .

## 2.3 Analysis of triggered transitions

We assume that a sequence of patterns is stored in the synaptic connections. For proper parameters the model supports stable auto-associative retrieval of individual patterns, meaning that after its initial excitation and without further input the network activity stays in an attractor corresponding with the retrieved pattern. Such a state is characterised by one of the overlaps  $m^\alpha$  being close to one and all others being small, i.e., of the order of the mean overlap between patterns.

The activated pattern, say pattern number  $\alpha$ , feeds into the next pattern,  $\alpha + 1$ , by means of the hetero-associative synaptic connections in matrix  $H$ . If the whole network is in an attractor state this input should not raise the neurons in  $\alpha + 1$  above threshold.

We study if it is possible to reliably switch to the next pattern by means of an *unspecific* input pulse into *all* excitatory neurons. The hetero-associative input into  $\alpha + 1$  together with that extra input may then switch  $\alpha + 1$  on. This in turn would increase the level of inhibition which introduces competition between activated patterns. Because pattern  $\alpha$  does not see hetero-associative input as  $\alpha + 1$  does and its cells may furthermore be in an adapted state, the activation of  $\alpha + 1$  can subsequently lead to an inactivation of  $\alpha$ . This way unspecific excitatory pulses into the network may be used to trigger

specific transitions imprinted in the hetero-associative connections. We study unspecific triggering as a worst case scenario – it would of course be possible to specifically trigger transitions to the next pattern by activating it directly and inhibiting all others. This, however, would require the triggering control network to already possess the information in the stored pattern sequence that we want to retrieve.

We observe the population activity in three subsequent populations  $\xi^\alpha \rightarrow \xi^{\alpha+1} \rightarrow \xi^{\alpha+2}$  during the transition from  $\xi^\alpha$  to  $\xi^{\alpha+1}$  and distinguish three phases:

**Phase 1:**  $\xi^\alpha$  stays active;  $\xi^{\alpha+1}$  remains inactive without external input (attractor state)

**Phase 2:**  $\xi^\alpha$  stays active;  $\xi^{\alpha+1}$  grows with external input  $I$  into all cells

**Phase 3:**  $\xi^\alpha$  deactivates;  $\xi^{\alpha+1}$  is completed and stays activated without external input  $I$

In all three phases  $\xi^{\alpha+2}$  and other populations beside  $\xi^\alpha$  and  $\xi^{\alpha+1}$  should not get activated.

Collecting the simplified equations for the pattern specific activities derived in section 2.2 we have for any pattern  $\alpha$ :

$$\tau \dot{h}^\alpha = -h^\alpha + I^\alpha - du^\alpha + am^\alpha + hm^{\alpha-1} + rs \quad (25)$$

$$\tau_a \dot{w}^\alpha = -w^\alpha + bm^\alpha \quad (26)$$

$$\tau_2 \dot{u}^\alpha = -u^\alpha + cs \quad (27)$$

$$m^\alpha = f(h^\alpha - w^\alpha) \quad (28)$$

The neurons are modelled with a sigmoid rate function  $f$  which increases steeply from 0 to a constant value  $\bar{f}$  beyond a threshold  $\vartheta$ . Without loss of generality we assume  $\bar{f} = 1$ . One choice of rate function is the threshold- or step-function with threshold  $\vartheta$ . We also set  $d = 1$ .

**Phase 1 (attractor state):** During this phase activation remains stationary such that  $h^\alpha$  is above threshold but  $h^{\alpha+1}$  and all other potentials stay below. Accordingly,  $m^\alpha$  is close to one and all other overlaps are close to zero. Only patterns  $\alpha$  and  $\alpha + 1$  receive specific synaptic input via auto- or hetero-associative connections from neurons in the active pattern  $\alpha$ . Neurons in other patterns only receive random input from  $\alpha$  via random or cross-talk synapses. There is no unspecific external input,  $I_0 = 0$ .

Stationarity implies for pattern  $\alpha$

$$0 = -h^\alpha - u^\alpha + a + r \quad (29)$$

$$0 = -w^\alpha + b \quad (30)$$

$$0 = -u^\alpha + c \quad (31)$$

In order to stay active  $h^\alpha > \vartheta + w^\alpha = \vartheta + b$  is required which implies

$$h^\alpha = a + r - c > \vartheta + b . \quad (32)$$

According to the discussion after (23) the parameter  $r$  accounts for random synapses between cells as well as cross-talk between patterns. These are random effects that can vary between zero and some positive value characterised by  $r$ . To stay on the safe side and ensure stable attractors even for those patterns where  $r$  is vanishing, we assume instead of (32) the stronger condition

$$h^\alpha = a - c > \vartheta + b . \quad (33)$$

The case  $r = 0$  is indeed relevant if orthogonal patterns and no additional random synapses are assumed.

For pattern  $\alpha + 1$  we get  $0 = -h^{\alpha+1} + r - c + h$  and  $0 = -w^{\alpha+1}$ . Pattern  $\alpha + 1$  should not get activated such that  $h^{\alpha+1} < \vartheta$  is required. This implies

$$h^{\alpha+1} = h + r - c < \vartheta . \quad (34)$$

Pattern  $\alpha + 1$  is the only one that receives hetero-associative input from  $\alpha$ . Therefore, if it stays stable, no other pattern can get activated either (beside the already active one,  $\alpha$ ).

**Phase 2:** During this phase the untuned input into all neurons is activated such that pattern  $\alpha$  stays on but pattern  $\alpha+1$  increases its potentials  $h^{\alpha+1}$  close to a new equilibrium above  $\vartheta$ . Simultaneously  $m^{\alpha+1}$  increases from 0 to 1. The average adaptation  $w^{\alpha+1}$  of neurons in  $\alpha + 1$  will also slowly increase towards its maximal value  $b$ . During this phase we further require that  $\alpha+2$  and any other pattern remain below threshold. The stimulus is switched on for a time of the order of the larger of the membrane time-constants ( $\tau, \tau_2$ ) and the adaptation constant ( $\tau_a$ ) to allow for these changes of variables to take place.

For simplicity we assume that  $\tau_2 \ll \tau$ , meaning that the inhibition follows excitation quasi instantaneously,  $u^\alpha = cs$  for all  $\alpha$ . Because  $m^\alpha$  is requested to stay (near) one we have for patterns  $\alpha$  to  $\alpha + 2$

$$\tau \dot{h}^\alpha = -h^\alpha + I_0 + (r - c)(1 + m^{\alpha+1}) + a \quad (35)$$

$$\tau \dot{h}^{\alpha+1} = -h^{\alpha+1} + I_0 + (r - c)(1 + m^{\alpha+1}) + am^{\alpha+1} + h \quad (36)$$

$$\tau \dot{h}^{\alpha+2} = -h^{\alpha+2} + I_0 + (r - c)(1 + m^{\alpha+1}) + hm^{\alpha+1} \quad (37)$$

For a constant  $m^{\alpha+1}$  the asymptotic potentials approached would be

$$h^\alpha = I_0 + (r - c)(1 + m^{\alpha+1}) + a \quad (38)$$

$$h^{\alpha+1} = I_0 + (r - c)(1 + m^{\alpha+1}) + am^{\alpha+1} + h \quad (39)$$

$$h^{\alpha+2} = I_0 + (r - c)(1 + m^{\alpha+1}) + hm^{\alpha+1} . \quad (40)$$

We require patterns  $\alpha$  and  $\alpha + 1$  to be super-threshold, but  $\alpha + 2$  to stay below. This is requested during the whole phase 2 where  $m^{\alpha+1}$  grows from 0 to 1. More precisely, consider pattern  $\alpha + 1$ : At the beginning of phase 2,  $h^{\alpha+1}$  is small and certainly smaller than the right hand side in (36). It will therefore grow towards the asymptotic value in (39) as long as the right hand side in (36) stays positive due to the unspecific input. We need to guarantee that the right hand side is positive for all  $m^{\alpha+1} \in [0, 1]$ . By means of linearity it is sufficient to check this only at  $m^{\alpha+1} = 0$  and 1, namely at the beginning and the end of phase 2, respectively. Analogous arguments for pools  $\alpha$  and  $\alpha + 2$  lead to the following inequalities:

$$I_0 + a + r - c > \vartheta + b \quad (41)$$

$$I_0 + a + 2(r - c) > \vartheta + b \quad (42)$$

$$I_0 + h + r - c > \vartheta + \delta \quad (43)$$

$$I_0 + a + h + 2(r - c) > \vartheta + b \quad (44)$$

$$I_0 + r - c < \vartheta \quad (45)$$

$$I_0 + h + 2(r - c) < \vartheta \quad (46)$$

In (43)  $\delta$  is some slack variable that reflects the possibility that the adaptation of  $\alpha + 1$  is not entirely relaxed towards zero at the beginning of phase 2. This would mean a higher input is required to overcome the firing thresholds. It is possible to just consider the worst case  $\delta = b$  and also to again assume  $r = 0$  in the first four inequalities. This makes (41) and (44) redundant relative to (42) and we keep

$$I_0 + a - 2c > \vartheta + b \quad (47)$$

$$I_0 + h - c > \vartheta + b \quad (48)$$

$$I_0 + r - c < \vartheta \quad (49)$$

$$I_0 + h + 2(r - c) < \vartheta \quad (50)$$

**Phase 3:** During this phase the external input is switched off again. At the beginning of phase 3 pattern  $\alpha + 1$  sees auto-associative input from itself and hetero-associative input from pattern  $\alpha$ ; pattern  $\alpha$  sees auto-associative input from itself but no hetero-associative input; pattern  $\alpha + 2$  sees hetero-associative input from pattern  $\alpha + 1$  only; all other patterns see at best random or cross-talk input. Therefore, pattern  $\alpha + 1$  has the best chance to survive the competition induced by inhibition. All patterns receive inhibition, but in contrast to phase 1 this now stems from two activated patterns,  $\alpha$  and  $\alpha + 1$ , and can therefore deactivate pattern  $\alpha$  if it is strong enough. High adaptation in pattern  $\alpha$  may in addition support this pattern to die out. The average adaptation  $w^\alpha$  of neurons in  $\alpha$  will then decay to zero during phase 3 and the adaptation of  $w^{\alpha+1}$  will approach and stay around the equilibrium value  $b$ . Thus at the end of phase 3 we are again in the situation of phase 1 with pattern  $\alpha + 1$  activated instead of  $\alpha$ .

We thus assume that essentially only pattern  $\alpha$  will change its activation during phase 3.  $h^\alpha$  will move to a new equilibrium below  $\vartheta$  and  $m^\alpha$  will decrease from one to close to zero. Therefore:

$$\tau \dot{h}^\alpha = -h^\alpha + am^\alpha + (r - c)(1 + m^\alpha) \quad (51)$$

$$\tau \dot{h}^{\alpha+1} = -h^{\alpha+1} + am^{\alpha+1} + hm^\alpha + (r - c)(1 + m^\alpha) \quad (52)$$

$$\tau \dot{h}^{\alpha+2} = -h^{\alpha+2} + hm^{\alpha+1} + (r - c)(1 + m^\alpha) \quad (53)$$

For  $m^\alpha = 0$  at the end of phase 3 we recover the same inequalities as in phase 1 (because the situation is the same with  $\alpha + 1$  activated stably instead of  $\alpha$ ). At the beginning of phase 3 we obtain (using additional worst case assumptions as earlier)

$$a + 2(r - c) < \vartheta \quad (54)$$

$$a + h - 2c > \vartheta + b \quad (55)$$

$$h + 2(r - c) < \vartheta \quad (56)$$

Inequality (56) is redundant because (50) is more restrictive. This leaves us with 8 inequalities for the model parameters: (33), (34), (47), (48), (49), (50), (54), and (55). These are conditions on the (non-negative) parameters  $a, h, b, c, r, I_0, \vartheta$  that allow for stable switching between attractors.

$$a - c - b > \vartheta > a - 2c + 2r \quad (57)$$

$$a + h - 2c - b > \vartheta > h - c + r \quad (58)$$

$$I_0 + a - 2c - b > \vartheta > I_0 - c + r \quad (59)$$

$$I_0 + h - c - b > \vartheta > I_0 + h - 2c + 2r \quad (60)$$

Observe that if we find a set of parameters with  $b > 0$  which satisfies these conditions, the same set also satisfies them for all smaller  $b$ . This includes  $b = 0$  such that adaptation is apparently not necessary for triggering transitions. However, the above inequalities are not the most general – in some parameter regimes not covered by them adaptation of neurons in the currently active pattern may ease the transition towards the next unadapted and therefore more excitable pattern. Suppression of the current pattern via inhibitory competition will be easier in this case.

Moreover, for any valid parameter set with  $r \neq 0$  the same inequalities stay valid for all smaller  $r$ . We may therefore continue the analysis by studying the case  $r = b = 0$ . If we find parameters that satisfy the above inequalities under this assumption, any margin can be used to choose maximal  $b$  and  $r$  afterwards independently. For the same reason we can assume  $\vartheta = 0$ .

We therefore make the assumptions  $r = b = \vartheta = 0$  and  $a = 1$ , and introduce new parameters  $\mu, \kappa, \zeta$  by means of  $h = \mu a, c = \kappa a, I_0 = \zeta a$ . The inequalities (57) to (60) then

become

$$1 - \kappa > 0 > 1 - 2\kappa \quad (61)$$

$$1 + \mu - 2\kappa > 0 > \mu - \kappa \quad (62)$$

$$1 + \zeta - 2\kappa > 0 > \zeta - \kappa \quad (63)$$

$$\zeta + \mu - \kappa > 0 > \zeta + \mu - 2\kappa \quad (64)$$

Equation (61) immediately implies  $1/2 < \kappa < 1$  and (62) and (63) imply  $\mu < \kappa$  and  $\zeta < \kappa$ , respectively. The right inequality in (64) is then redundant but the left requires  $\mu + \zeta > \kappa$  or  $\mu > \kappa - \zeta$  and  $\zeta > \kappa - \mu$ . If we use the latter two conditions in (62) and (63) these inequalities reduce to  $1 - \zeta - \kappa > 0$  and  $1 - \mu - \kappa > 0$  which is the same as  $\zeta < 1 - \kappa$  and  $\mu < 1 - \kappa$ . Because  $\kappa$  is already constrained in the range  $1/2 < \kappa < 1$  we get  $\zeta < 1 - \kappa < \kappa$  and  $\mu < 1 - \kappa < \kappa$  such that the earlier derived conditions  $\mu < \kappa$  and  $\zeta < \kappa$  are obsolete. This finally leaves us with

$$\vartheta = 0, \quad a = 1, \quad 1/2 < \kappa < 1, \quad \zeta < 1 - \kappa, \quad \mu < 1 - \kappa, \quad \mu + \zeta > \kappa. \quad (65)$$

In words: For a vanishing threshold and an arbitrary but fixed auto-associative coupling constant, inhibition has to be relatively strong between 50% and 100% of the auto-associative coupling. The input and hetero-association strengths individually may only be as strong as up to 50% of the auto-associative coupling, but input and hetero-association together have to be stronger than the inhibition.

## 2.4 Examples for triggered and autonomous transitions

To satisfy (65) we may choose  $\mu = \zeta$  which results in  $\kappa/2 < \mu = \zeta < 1 - \kappa$  in addition to the earlier  $1/2 < \kappa < 1$ . This allows for solutions in the range  $1/2 < \kappa < 2/3$ , for instance,  $\kappa = 0.6, \mu = \zeta = 0.35$ . Given these values we can check the margins in inequalities (57) to (60), which we can then use to calculate maximal  $b$  and  $r$  values (as fractions of  $a = 1$ ) assuming vanishing thresholds,  $\vartheta = 0$ . We find as maximal values  $b^{\max} = r^{\max} = 0.1$ . Alternatively we can split the margin for  $b$  into a fixed positive threshold  $0 < \vartheta < b^{\max}$  and a new upper bound for the maximal adaptation strength, e.g.,  $\vartheta = 0.05$  and  $b = 0.05$ . Positive firing thresholds have the advantage that cells do not autonomously activate if there is no input and no already active pattern that can cause inhibitory competition in order to avoid firing of unwanted neurons.

Figures 1 and 2 display example simulations. The simulations implement (1) to (6) for  $n = 36$  excitatory neurons, and  $P = 6$  non-overlapping patterns with  $m = 6$  active units each. The rate function of the excitatory neurons is the Heaviside function (unit step).  $R$  is the zero matrix. There is only a single inhibitory neuron; synapses from excitatory cells to the inhibitory cell have identical values  $C_{1j} = c$ ; synapses from the inhibitory cell

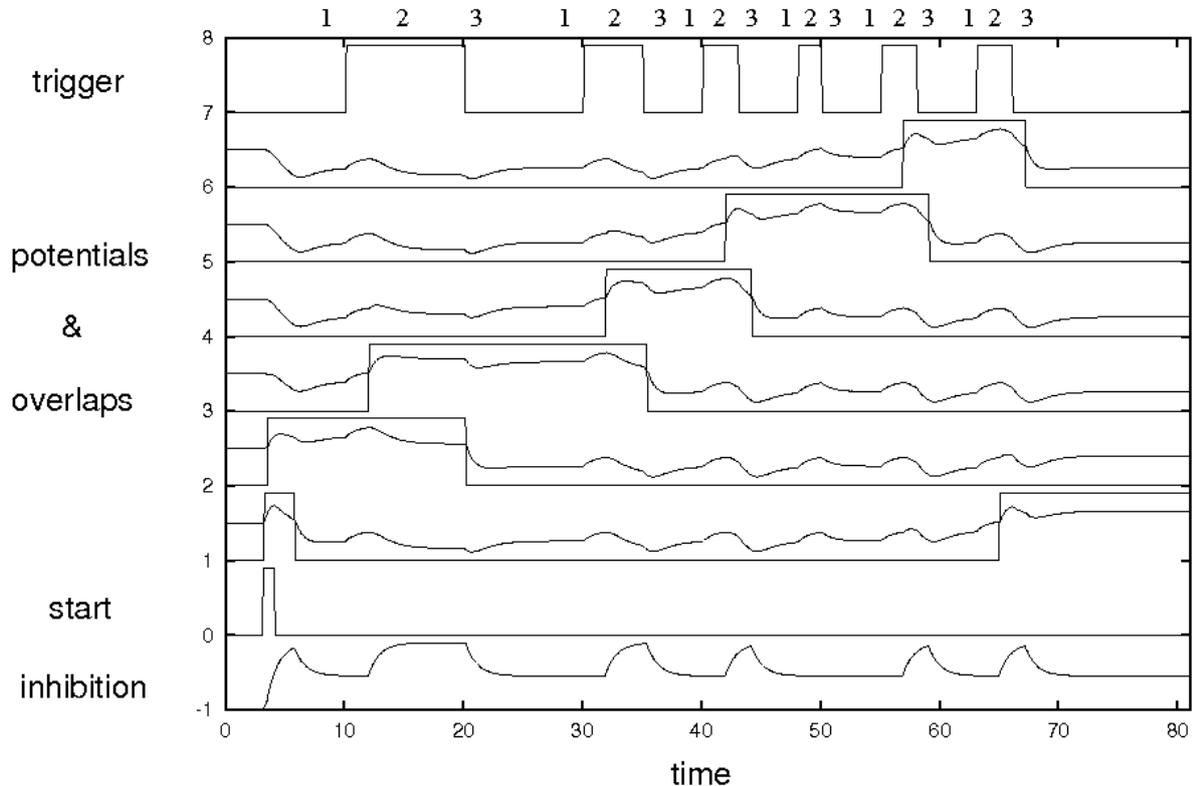


Figure 1: Transitions between patterns for various durations of an unspecific trigger. Trace 7 shows the ‘trigger’ signal; small numbers 1,2,3 on top indicate the three phases of a transition defined in section 2.3. Traces 1 to 6 show pattern overlaps (rectangular curves) and potentials of individual cells in each pattern (smooth curves). Trace -1 is the global inhibition and trace 0 indicates a specific excitatory driving input into the cells of the first pattern which ‘starts’ the sequence retrieval. Overlaps are between 0 and 1. Other quantities are in arbitrary units. Time is in multiples of the membrane time constants. For more explanations see text.

to the excitatory cells are all 1. The latter is possible without loss of generality because the inhibition is purely linear; it is therefore enough if  $c$  can be varied. The differential equations are solved by the Euler method [46] with stepsize 0.1. Parameter values were  $\tau = \tau_2 = \tau_a = 1, r = 0, b = 0.05, \vartheta = 0.05, a = 1.0, \mu = 0.35, \zeta = 0.35, \kappa = 0.6$ .

Figure 1 shows switching between patterns given unspecific triggers of various duration. Note that we have chosen all time constants identical,  $\tau = \tau_2 = \tau_a = 1$ , although in section 2.3 we assumed that inhibition is fast in order to proof the stability conditions provided there. However, simulations show that this assumption is not very crucial. Even with identical time-constants as used here triggering works well.

In Figure 1 the first pattern is excited by an excitatory specific input into all neurons

of pattern 1 during time 3.0 to 4.0 (see trace 0). Note that this activates pattern 1 (trace 1) but also already pattern 2 (trace 2) via the hetero-associative couplings from pattern 1 to 2. Pattern 2 can fire because the inhibition (trace -1) starts from 0 and has to build up first in order to suppress patterns that only get hetero-associative input. This initialisation step is not covered by the stability considerations in section 2.3. Firing of pattern 2 could, for instance, be avoided by initialising the inhibition with an appropriate positive value or by providing excitatory input not only into the first pattern but also the inhibitory unit during the primary excitation of the starting pattern.

Observe that each triggering phase causes synchronous fluctuations in the potentials of all excitatory cells regardless of which pattern they belong to (smooth curves in traces 1 to 6). This reflects the unspecific nature of the trigger. A positive (excitatory) response is usually followed by an inhibitory slightly later component caused by the next pattern getting excited and thus suppressing its competitors. Neurons in successive patterns are super-threshold such that the respective overlaps are large. Because we use step-functions as rate functions of the excitatory neurons the overlaps switch in an almost rectangular manner between 0 and 1 (traces 1 to 6).

Small numbers on top of Figure 1 reflect the three different phases of a transition as defined in section 2.3. During phases 1 only a single pattern is activated and stays roughly in equilibrium. During phase 2 the unspecific input raises the subsequent pattern above firing threshold; other patterns do not fire because they do not get hetero-associative input. If two patterns are activated during phase 2 inhibition rises proportionally. This increased inhibition switches the previously active pattern off in phase 3 because that pattern only receives auto-associative input whereas its successor receives hetero- and auto-associative input and therefore stays above threshold. Afterwards signals relax back to equilibrium values until the next transition is initiated.

The duration of a triggering phase 2 can be arbitrarily long as long as it exceeds a certain minimum length. This basically was assumed in the stability considerations in section 2.3 where the only restriction on the duration of phase 2 was that it is long enough to allow for the transition to take place. In Figure 1 the durations start much longer than the neural time-constants ( $\tau = \tau_2 = \tau_a = 1$ ) but get successively shorter. As long as the unspecific input is present for long enough a time to raise the subsequent pattern above threshold the retrieval works well; if it gets too short retrieval fails. This happens around time 50 where the trigger is present only for 2 units of time and the network stays in attractor 5. Finally note that the sequence of stored patterns in the simulation is cyclic. At time 65 attractor 1 is reached again.

Figure 2 displays a simulation with adaptation too strong to guarantee stable switching. In this simulation the unspecific input is permanently on (trace 7). The figure shows, that above a certain value of the adaptation strength (model parameter  $b$ ) autonomous propagation of activity through the sequence of attractors results. As the adaptation gets

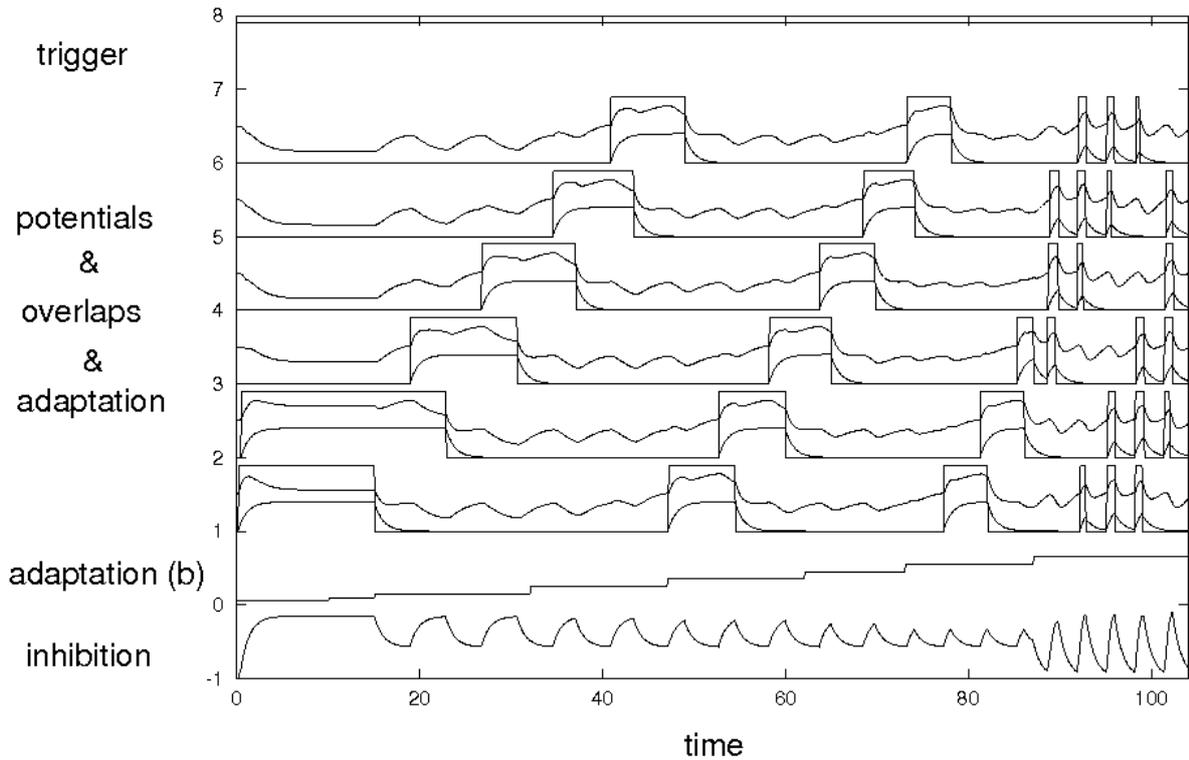


Figure 2: Autonomous propagation of activity if the unspecific input is persistently high (trace 7, ‘trigger’). The level of adaptation (model parameter  $b$ ) increases during the simulation in several steps from 0.05 to .55 (trace 0). Trace -1 displays the level of inhibition and traces 1 to 6 the pattern overlaps as well as potentials and adaptation variables of individual neurons in the different patterns. Pattern 1 is specifically activated during the first time unit. Observe autonomous propagation of activity, a shortening of retrieval periods with adaption strength (quicker sequence retrieval), and unstable propagation for high adaptation values ( $t > 85$ ).

stronger, the times spent in any single attractor gets shorter, and therefore the retrieval of the pattern sequence speeds up. A similar effect can also be reached by increasing the hetero-associative coupling strength, but not, for example, by changes in the level of unspecific input (not shown). At high adaptation, retrieval fails with more complex firing patterns emerging than linear sequences of states ( $t > 85$  in the figure).

Activity starts propagating through the combined action of adaptation and inhibition. During phase 2 of a retrieval cycle, characterised by the presence of the unspecific triggering stimulus, two patterns are ideally active, the current and the previous one. If the adaptation strength is too high it switches the previous pattern off because its auto-associative self-excitation together with the unspecific input will no longer be strong

enough to overcome the threshold plus adaptation minus the inhibition. As a consequence the inhibition is released such that the next pattern can become super-threshold, it fires, and after a brief relaxation time the system reaches the initial state again just one step forward in the sequence.

For increasing adaptation strength the adaptation times get shorter because threshold conditions are reached quicker. The inhibition can then no longer follow the excitatory activity changes fully, such that its amplitude changes decrease (trace -1 in Figure 2). This diminishing inhibition causes an instability to complex firing patterns at high adaptation strength where the inhibition can no longer suppress the firing of more than one or two patterns only ( $t > 85$  in the figure).

In summary the present section has presented simulations of proper retrieval of pattern sequences by unspecific external triggers consistent with the analysis in section 2.3. Such simulations reveal that the stability conditions derived in section 2.3 are actually quite robust. Nonetheless, if the conditions are broken, especially for too strong adaptation or hetero-association, more complex activation patterns can arise including an autonomous mode of retrieval of a stored sequence. These modes will be discussed in the next section in relation to generators and recognisers of more complex, syntactical time patterns than linear sequences.

## 2.5 Context and graph-like transition patterns

In traditional attractor networks static patterns are stored in auto-associative connections; they can be retrieved from partial or noisy cues [29, 68] by the recurrent network dynamics. Hetero-associative connections as shown in the previous section can further be used to extend the possible dynamic modes towards sequences [1, 30] thereby introducing (temporal) order. In an autonomously evolving setup they also provide a notion of “time”. Such network properties may underlie episodic memory and time perception [61]. Still, linear sequences are not yet sufficiently powerful models of spatio-temporal pattern processing appearing in real-world situations. Those usually reveal syntactic structure, for instance, describable by some kind of grammar.

We have earlier pointed out how attractor networks can be extended in a natural way towards the processing of syntactic structures [61, 63, 64]. Keys to this approach are, first, the introduction of several possible target states for each attractor coded in hetero-associative couplings, and second, specific external inputs, that disambiguate which transition is selected if there are several for a given attractor state. These extensions were partly motivated by von Neumann’s and Wickelgren’s earlier works, see, e.g., [43, 65, 66].

This extension of attractor networks leads from pure attractors via linear sequences (also called “synfire chains” [1]) to labelled graphs (“synfire graphs”) representing complex, input- and context-driven patterns of activation flow in a generalised attractor sys-

tem. Nodes in these graphs are the stored patterns (attractors) and the edge-labels represent the specific external inputs that cause a particular transition between patterns. This generic image indeed describes nothing but the state transition graph of some finite state automaton (FSA) [28]. A simple and plausible extension thus allows for a significant boost in computational capabilities of attractor networks. We have called the general framework “operational cell assemblies” because it concretises how classical Hebbian cell assemblies can systematically support procedural and rule-based processing, see [61]. Clearly, the different dynamic modes – fixed point attractors, linear flow of activity (sequences), and input dependent gating of activation flow in a graph-like structure (syntax) – can all be integrated in the same neuronal network; they just represent three different principles in an isolated and abstract way that can be studied independently.

In [61] the recognition of spatio-temporal syntactic patterns was studied as well as their autonomous generation. These simulations made use of “full disambiguation” of transitions by external inputs: If in any state several transitions were possible a specific input external to the network provided a bias that determined the next state uniquely (together with the state information). If there was no such input or if it did not fully determine the next state, the network behaviour was undefined. This is equivalent to the behaviour of a finite state automaton and thereby also a modern computer – ambiguous states do not occur.

Processes in the brain may well be different. However, an unspecific trigger causing the network to compute as in section 2.3 assumes that the next target is unambiguous. This is true for the linear sequences used in section 2.3 and it was also true in [61] where specific inputs disambiguated target states in a word recogniser. In a more general setup there may be uncertainties about the next network state, especially where syntactic patterns are generated and not recognised. Whereas external input can disambiguate activity flow in the second case, other mechanisms are necessary in the case of internal generation.

In neural structures disambiguation can have various causes not always clearly separable from each other. One possibility is specific input into an operational associative module as already mentioned. A second one is general context that remains stable for a certain time and constrains or biases some of the possible transitions in a network (e.g., [65, 66]). Instead of a single fixed transition graph this results in context-dependent mappings implemented in a recurrent operational cell assembly network. Such contextual selection may underlie rule- and task-set maps in frontal cortical areas and the switching between them [15, 36, 42, 51, 56].

It is also useful to consider that a local network will usually be embedded in a larger-scale super-network of operational modules. This situation reflects that cortical columns and areas are quite extensively connected with each other. Processes going on in the different modules mutually provide input to each other, which again may disambiguate local processes. Finally, there is certainly also the impact of noise as a way to choose

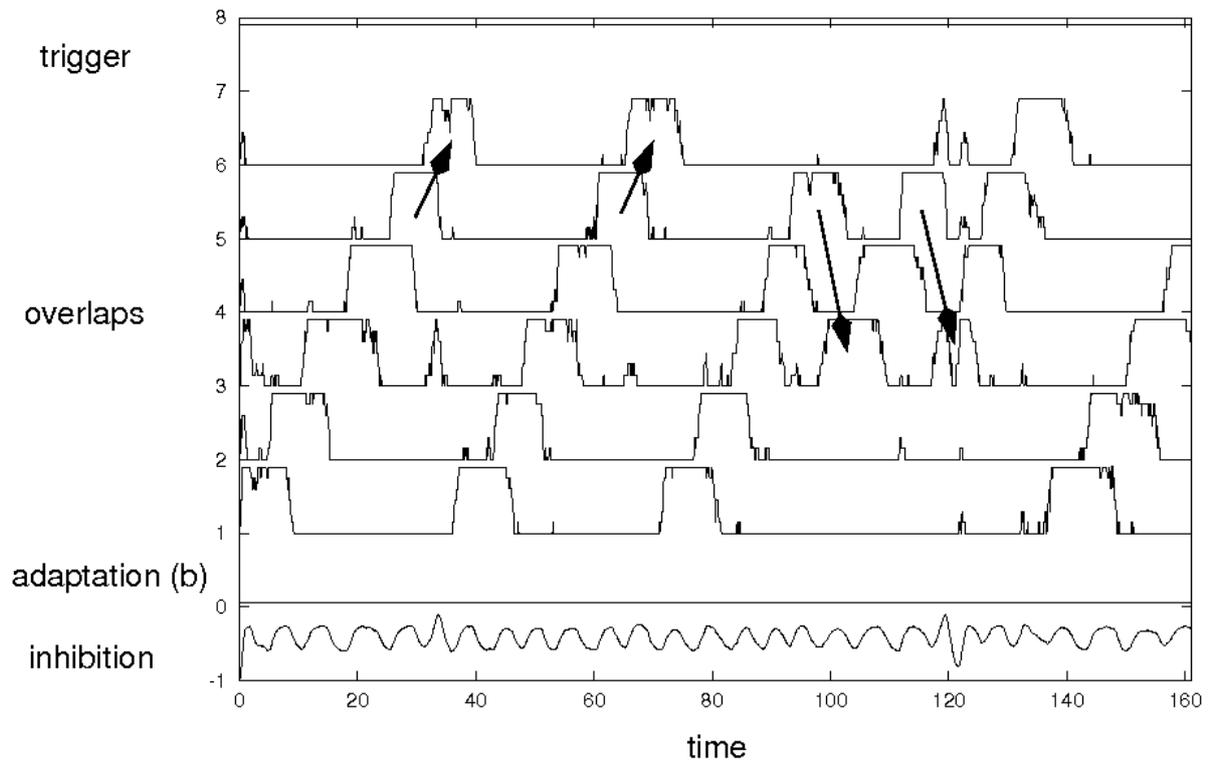


Figure 3: Example for random transitions in autonomous retrieval mode. The same network and parameters as in Fig. 2 (for  $b = 0.05$ ) are used but membranes receive noise input and beside the cyclic stored sequence an extra transition from pattern 5 to 3 has been embedded in the coupling matrix. Observe that the noise causes transitions from state 5 to either of its target states 6 and 3 (arrows).

between different possible targets. This requires in addition some kind of winner-take-all mechanism because a target, once chosen, needs to stabilise itself against the firing of other targets. The next section demonstrates disambiguation by noise in a simple example setup.

## 2.6 Example for random transitions in a network

Figure 3 shows an example of a network with random transitions. The model is the same as in section 2.4 including parameter values, but the neurons now receive independent Gaussian white noise inputs that allows for random transitions in case of ambiguities. In addition to the cyclic sequence of patterns 1 to 6 an additional hetero-associative transition has been added from pattern 5 to 3. All hetero-associative weights are the same, therefore the probability for transitions from pattern 5 to either 6 or 3 should be

about equal (neglecting effects of adaptation that may favour one or the other transition). Pattern 1 is externally activated at time zero. The unspecific input is permanently on which causes an autonomous network behaviour. As can be seen the network first cycles two rounds through the “regular” cyclic sequence and then does another two transitions along the second pathway (arrows).

This simple example is shown here for demonstration purposes. It should be clear that quite arbitrary transition structures can be implemented and that by choosing different synaptic weights hetero-associative transitions can be given different probabilities to be taken. Thereby a wide variety of Markov chains can be implemented. (Although not all possible ones, unless additional assumptions are made; the finite system noise gives every transition a finite chance to occur, even those that are not imprinted in the hetero-associative connections. Dynamic network properties may pose additional constraints – long-lasting adaptation, for instance, will induce temporal correlations beyond those described by Markov chains.)

The setup can apparently also be combined with the mechanisms described earlier: Specific external input can provide context that disambiguates situations by making some transitions more or less likely. An unspecific external input can still act as a trigger: The simulation in Figure 3 shows the autonomous generation of a syntactic time-pattern, but the same network works also in a triggered mode as in Figure 1 (not shown).

### 3 Discussion

In summary we have shown that syntactic patterns can be represented in operational cell assembly networks comprising attractor states, input dependend transitions between states, and an unspecific activation control that triggers actual transitions unless they evolve in an autonomous mode. Noise in the system can be used to disambiguate possible target states and give them different transition probabilities as reflected by different strengths of hetero-associative connections. Disambiguation can further be realised by contextual signals, either spatially or temporally (as for instance already suggested by Wickelgren [65]). In an abstract sense, such networks may be seen as just implementing some kind of Markov chains in neural hardware. However, by flexibly combining the various features mentioned with the well-known classical memory properties of attractor networks complex computational structures can result that exceed the usual world of Markov chains by far.

We here have provided parameter regimes for a proper retrieval of syntactic sequences using unspecific excitatory triggers and have shown simulations of autonomous and non-autonomous modes of retrieval. These parameter constraints should enable an implementation of temporal pattern generators and recognisers in software simulations and

potentially also in neural hardware [32, 53, 61].

In the present work we have used simplistic graded response neurons. The dynamics of spiking neuron networks can often be reduced to “mean-field” equations of similar structure [11, 12, 16, 30]. Single units in our model may be interpreted as representing pools of spiking neurons in an asynchronous regime. In [55] we have shown that triggered transitions can also be caused in networks of conductance based neurons, although the results presented here do not generalise to this work because of a different network structure.

We only considered linear and unspecific inhibition, i.e., the inhibitory neurons did not store information about the pattern sequences. The latter reflects that mainly only synapses between excitatory neurons are engaged in synaptic learning. The linearity assumption is made to simplify the analysis in the main text. It is not entirely crucial, but if relaxed, many of the conditions constraining the model parameters become nonlinear equations that can not be further analysed by hand. Linear inhibition informally means that in stationary states the total amount of inhibition is proportional to that of excitation, which is not unreasonable as some author’s even make the stronger claim that excitation and inhibition are balanced in physiological states [23, 59]. The inhibition is again assumed to reflect the average activity level of inhibitory neurons. Although individual neurons may have non-linear rate functions those may average to a close(r) to linear relationship in the mean.

The general framework of operational cell assemblies can be applied to model rule-based temporal behaviour. A primary target would be a model for syntactic aspects of language on its different hierarchical levels of phonemes, syllables, words, and sentences, because the use of attractor states allows for a quite arbitrary timing of triggers. Trigger signals can come at any time if the stability of attractors that hold information about the current state of parsing is guaranteed. The inflow of acoustic information can thus happen at flexible speed. Experimental results by Koechlin and Jubault [36] may be interpreted as evidence for triggered transitions. This work suggests the existence of a hierarchy of behavioural/language sequences in Broca’s area and it’s right homologue, as well as trigger-like signals occurring at segment boundaries at the various hierarchical levels. Strictly speaking this work does not show that the triggering signals are unspecific. However, our assumption of unspecific triggers is an extreme case; specificity would ease the sequencing problem by potentially providing information about possible target states.

On the shortest time-scale of phonemes and syllables, other than triggered transitions may be of importance, see, e.g., [10, 22]. On that level the flow of incoming information may cause a more continuous flow of activity through the neural recogniser circuits without discrete triggers. Such circuits may be more akin to “synfire chains” [1, 22, 60] or the untriggered autonomous mode of operation displayed in Figures 2 and 3. Clearly, any kind of mixture with some transitions triggered, others autonomously evolving is possible.

Previous sequencing models have often focused on autonomous retrieval. They usually require some mechanism for automatically destabilising attractors. Several such mechanisms have been proposed like adaptation, depressing synapses, delayed hetero-associative connections, and others [1, 27, 30, 49, 52]. Our main interest here was mainly on non-autonomous triggered transitions. Few works have considered those before, but see [55, 63]. Our analyses show that adaptation is not necessary to stably trigger transitions; however, as the simulations in section 2.4 show that it can cause autonomous transitions compatible with the mechanisms in [30]. It is possible to include effects of synaptic depression and facilitation in our analysis by assuming that the synaptic strengths approach stationary values during the three phases defined in section 2.3. In that case the parameters “a” and/or “h” for the auto- and hetero-association strengths in the stability inequalities have to be replaced by their adapted or maximum values, depending on whether a target pattern is requested to fire or stay below threshold, respectively. The form of the conditions remains similar as before, being slightly more restrictive. Stable triggering can be obtained at least for facilitating synapses or not too strongly adapting ones.

Our approach of triggered transitions precisely aims at implementing rule-like operations in cell assembly networks. In a linguistic context these rules may refer to entities at different levels, low-level phonetic features or higher level syntactic categories. Different levels can be combined into hierarchies such that linguistic knowledge can be represented efficiently (perhaps reflected by the experiments in [36]). We have worked out the main ideas in [61, 63]. The framework extends attractor networks from fixed point retrieval and linear sequences towards context-dependent rules. In these previous works, however, spatio-temporal “grammatical” sequences were recognised or generated such that the necessary state transitions either occurred fully autonomously or were driven by specific input events. The triggering proposed in the present work adds an additional principle of unspecific control.

An important question refers to memory capacities of associative networks. A common result is that under optimal conditions associative networks are quite efficient memories able to store a significant fraction of a bit per (binary) synapse [11, 12, 45]. Simulations using spiking neuron networks get close to the theoretical limits as shown in [49, 54]. The network studied in the present work assumes some stability margin expressed by positive threshold distances in the stability conditions. It will therefore not reach the absolute theoretical capacity limits, but as long as the margins are not too big, one would still expect an extensive memory capacity. Given the huge number of synapses in the cortex, a large number of associations can certainly be stored.

We should finally note that we haven’t touched on the difficult problem of how to learn the type of model described here in a self-organising way, that is, how such structures can emerge as a consequence of stimulus-driven synaptic plasticity. Machine learning techniques can be used to learn some grammars in artificial neural networks, but they are

often not too plausible biologically [57, 70]. The problem has been approached recently in more realistic settings, but with no answer yet: some of the proposed systems appear quite engineered [34, 41], others with more generic architectures are not yet able to learn more than the simplest grammatical temporal patterns [19]. We have earlier pointed out that grammar and rule-like behaviour may in principle result from a timing dependent synaptic learning rule that maximises a temporal generalisation of mutual information [62]. However, this result was derived in the quite abstract framework of Markov chains and needs further exploration in realistic contexts.

## Acknowledgement

The authors thank the anonymous reviewers for their extended and constructive comments that helped to improve the manuscript significantly. The main content of this paper has been prepared while G.P. was visiting researcher at the Centre for Theoretical and Computational Neuroscience in Plymouth U.K. This visit was generously funded by an Incoming Visitors Grant to T.W. from the Royal Society. Support by the Engineering and Physical Sciences Research Council to T.W. is also greatly acknowledged (grant EP/C010841/1, COLAMN - A Novel Computing Architecture for Cognitive Systems based on the Laminar Microcircuitry of the Neocortex).

## References

- [1] Abeles M (1991) *Corticonics: Neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge UK.
- [2] Abeles M, Bergman H, Margalit E, Vaadia E (1993) Spatio-temporal firing patterns in frontal cortex of behaving mokeys. *Jof Neurophysiol* 70:1629–1643
- [3] Amari SI (1972) Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers* 21:1197–1206
- [4] Amit D (1988) *Modeling Brain Function*. Cambridge University Press
- [5] Amit D (1995) The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioral and Brain Sciences* 18:617–657
- [6] Arbib M, Billard A, Iacoboni M, Oztop E (2000) Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks* 13:975–997
- [7] Arbib MA (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioural and Brain Sciences* 28:105–167

- [8] Bienenstock E (1995) A model of the neocortex. *Network* 6:179–224
- [9] Braitenberg V (1978) Cell assemblies in the cerebral cortex. In: Heim R, Palm G (eds) *Theoretical Approaches to Complex Systems*, Springer, Berlin, pp 171–188
- [10] Buonomano DV (2003) Timing of neural responses in cortical organotypical slices. *Proceedings of the National Academy of Sciences (USA)* 100:4897–4902
- [11] Coolen A (2001) Statistical mechanics of recurrent neural networks i - statics. In: Moss F, Gielen S (eds) *Handbook of Biological Physics*, 4th edn, Elsevier Science B.V., chap 14, pp 531–596
- [12] Coolen A (2001) Statistical mechanics of recurrent neural networks ii - dynamics. In: Moss F, Gielen S (eds) *Handbook of Biological Physics*, 4th edn, Elsevier Science B.V., chap 15, pp 597–662
- [13] Cossart R, Anonov D, Yuste R (2003) Attractor dynamics of network UP states in the neocortex. *Nature* 423:283–288
- [14] Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402:529–533
- [15] Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar VL, Petersen SE (2006) A core system for the implementation of task sets. *Neuron* 50:799–812
- [16] Eggert J, van Hemmen J (2000) Unifying framework for neuronal assembly dynamics. *Physical Review E* 61:1855–1874
- [17] Fay R, Kaufmann U, Knoblauch A, Markert H, Palm G (2005) Combining visual attention, object recognition and associative information processing in a neurobotic system. In: Wermter S, Palm G, Elshaw M (eds) *Biomimetic Neural Learning for Intelligent Robots*, LNAI, vol 3575, Springer, Berlin, Heidelberg, pp 118–143
- [18] Feldman J, Narayanan S (2004) Embodied meaning in a neural theory of language. *Brain and Language* 89:385–392
- [19] Garagnani M, Wennekers T, Pulvermüller F (2008) A neuroanatomically grounded Hebbian leaning model of attention-language interactions in the brain. *European Journal of Neuroscience* 27:492–513
- [20] Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. *Nature* 424:669–674
- [21] Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440:1204–1207

- [22] Hahnloser R, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65–70
- [23] Haider B, Hämmerlin B, Duque A (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience* 26:4535–4545
- [24] Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsáki G (2003) Organization of cell assemblies in the hippocampus. *Nature* 424:552–555
- [25] Hebb D (1949) *The Organization of Behavior*. Wiley, New York
- [26] Herrmann M, Hertz J, Prugel-Bennet A (1995) Analysis of synfire chains. *Network* 6:403–414
- [27] Hertz J, Krogh A, Palmer R (1991) *Introduction to the theory of neural computation*. Addison Wesley
- [28] Hopcroft J, Ullman J (1969) *Formal Languages and Their Relation to Automata*. Addison-Wesley
- [29] Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79:2554–2558
- [30] Horn D, Usher M (1989) Neural networks with dynamical thresholds. *Physical Reviews A* 40:1036–1044
- [31] Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, Yuste R (2004) Synfire chains and cortical songs: Temporal modules of cortical activity. *Science* 304:559–564
- [32] Indiveri G (2007) Synaptic plasticity and spike-based computation in VLSI networks of integrate-and-fire neurons. *Neural Information Processing – Letters and Reviews* 11:135–146
- [33] Jarvis ED (2004) Learned birdsong and the neurobiology of language. *Ann NY Acad Sci* 1016:749–777
- [34] Knoblauch A, Markert H, Palm G (2005) An associative cortical model of language understanding and action planning. In: JR JM, Alvarez (eds) *Proceedings of IWINAC 2005, First International Work-Conference on the Interplay between natural and artificial computation*, Las Palmas de Gran Canaria, Spain, *Lecture Notes in Computer Science*, vol 3562, Springer, Berlin, New York, pp 405–414

- [35] Knoblauch A, Markert H, Palm G (2005) An associative model of cortical language and action understanding. In: Cangelosi A, Bugmann G, Borisyuk R (eds) *Modeling Language, Cognition and Action. Proceedings of the Ninth Neural Computation and Psychology Workshop, Plymouth 2004.*, World Scientific, pp 79–83
- [36] Koechlin E, Jubault T (2006) Broca’s area and the hierarchical organization of human behaviour. *Neuron* 50:963–974
- [37] Lashley K (1951) The problem of serial order in behavior. In: Jeffress L (ed) *Cerebral mechanisms in behavior*, New York: Wiley, pp 112–136
- [38] Markert H, Palm G (2006) An approach to language understanding and contextual disambiguation in human-robot interaction. In: *International Workshop on Neural-Symbolic Learning and Reasoning (NeSy 2006)*, pp 23–35
- [39] Markert H, Knoblauch A, Palm G (2005) Detecting sequences and understanding language with neural associative memories and cell assemblies. In: Wermter S, Palm G, Elshaw M (eds) *Biomimetic Neural Learning for Intelligent Robots, Lecture Notes in Artificial Intelligence*, vol 3575, Springer, Berlin, Heidelberg, pp 107–117
- [40] Markert H, Knoblauch A, Palm G (2007) Modelling of syntactical processing in the cortex. *BioSystems* 89:300–315
- [41] Markert H, Kayikci Z, Palm G (2008) Sentence understanding and learning of new words with large-scale neural networks. In: Prevost L, Marinai S, Schwenker F (eds) *Proc. Artificial Neural Networks in Pattern Recognition (ANNPR 2008)*, LNAI, vol 5064, Springer, Berlin, Heidelberg, pp 217–227
- [42] Mushiake H, Saito N, Sakamoto K, Itoyama Y, Tanji J (2006) Activity in lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron* 50:631–641
- [43] von Neumann J (1958) *The Computer and the Brain*. Yale University Press, New Haven, CT, USA
- [44] Palm G (1982) *Neural Assemblies: An Alternative Approach to Artificial Intelligence*. Springer-Verlag, Berlin
- [45] Palm G, Sommer FT (1992) Information capacity in recurrent mcculloch-pitts networks with sparsely coded memory states. *Network* 2:177–186
- [46] Press W, Teukolsky S, Vetterling W, Flannery B (1993) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, MA.

- [47] Pulvermüller F (1992) Constituents of a neurological theory of language. *Concepts in Neuroscience* 3:157–200
- [48] Pulvermüller F (2003) *The Neuroscience of Language: On Brain Circuits of Words and Serial Order*. Cambridge University Press
- [49] Rehn M, Lansner A (2004) Sequence memory with dynamical synapses. *Neurocomputing* 58–60:271–278
- [50] Rizzolatti G, Fogassi L, Gallese V (2002) Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology* 12:149–154
- [51] Rushworth M, Walton M, Kennerley S, Bannerman D (2004) Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences* 8:410–417
- [52] Russo E, Namboodiri VM, Treves A, Kropff E (2008) Free association transitions in models of cortical latching dynamics. *The New Journal of Physics* 10:015,008
- [53] Schemmel J, Meier K, Mueller E (2004) A new VLSI model of neural microcircuits including spike time dependent plasticity. In: *Proceedings IJCNN, IEEE Press*, pp 1711–1716
- [54] Sommer F, Wennekers T (2001) Associative memory in networks of spiking neurons. *Neural Networks* 14:825–834
- [55] Sommer FT, Wennekers T (2005) Synfire chains with conductance-based neurons: Internal timing and coordination with timed input. *Neurocomputing* 65–66:449–454
- [56] Stoet G, Snyder LH (2003) Executive control and task-switching in monkeys. *Neuropsychologia* 41:1357–1364
- [57] Sun R, Giles L (2000) *Sequence Learning: Paradigms, Algorithms, and Applications*. Springer, Berlin, New York
- [58] Tsien JZ (2001) Linking Hebb’s coincidence-detection to memory formation. *Current Opinion in Neurobiology* 10:266–273
- [59] van Vreeswijk C SH (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274:1724–1726
- [60] Wennekers T (2000) Dynamics of spatio-temporal patterns in associative networks of spiking neurons. *Neurocomputing* 32:597–602
- [61] Wennekers T (2006) Operational cell assemblies as a paradigm for brain-inspired future computing architectures. *Neural Information Processing - Letters and Reviews* 10:135–145

- [62] Wennekers T, Ay N (2005) Finite state automata resulting from temporal information maximization and a temporal learning rule. *Neural Computation* 17:2258–2290
- [63] Wennekers T, Palm G (2007) Modelling generic cognitive functions with operational Hebbian cell assemblies. In: Weiss M (ed) *Neural Network Research Horizons*, Nova Science Publishers, pp 225–294
- [64] Wennekers T, Garagnani M, Pulvermüller F (2006) Language models based on Hebbian cell assemblies. *Journal of Neuroscience (Paris)* 100:16–30
- [65] Wickelgren WA (1979) I liked the postcard you sent Abe and i: Context-sensitive coding of syntax and other procedural knowledge. *Bulletin of the Psychonomic Society* 13:61–63
- [66] Wickelgren WA (1992) Webs, cell assemblies, and chunking in neural nets. *Concepts in Neuroscience* 3:1–53
- [67] Wijekoon J, Dudek P (2008) Compact silicon neuron circuit with spiking and bursting behaviour. *Neural Networks* 21:524–534
- [68] Willshaw D, Buneman O, Longuet-Higgins H (1969) Non-holographic associative memory. *Nature* 222:960–962
- [69] Willwacher G (1982) Storage of a temporal pattern sequence in a network. *Biological Cybernetics* 43:115–126
- [70] Wörgötter F, Porr B (2005) Temporal sequence learning, prediction, and control: A review of different models and their relation to biological mechanisms. *Neural Computation* 17:245–319
- [71] Yamashita Y, Tani J (2008) Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment. *PLoS Computational Biology* 4:e1000,220