

Cell Assembly Dynamics in Detailed and Abstract Attractor Models of Cortical Associative Memory

Anders Lansner, Erik Fransén, and Anders Sandberg

Studies of Artificial Neural Systems (SANS), Dept. of Numerical Analysis and Computer Science, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

Address for correspondence: Anders Lansner, Royal Institute of Technology, SE-100 44, Stockholm, Sweden

Phone: +46-8-7906210, Fax: +46-8-7900930, Email: ala@nada.kth.se

Received: November 22, 2002; accepted: December 15, 2002

Key words: Biophysical compartmental neuron model, hypercolumns, minicolumns, forgetting, incremental learning, reaction time

Summary: During the last few decades we have seen a convergence among ideas and hypotheses regarding functional principles underlying human memory. Hebb's now more than fifty years old conjecture concerning synaptic plasticity and cell assemblies, formalized mathematically as attractor neural networks, has remained among the most viable and productive theoretical frameworks. It suggests plausible explanations for Gestalt aspects of active memory like perceptual completion, reconstruction and rivalry.

We review the biological plausibility of these theories and discuss some critical issues concerning their associative memory functionality in the light of simulation studies of models with palimpsest memory properties. The focus is on memory properties and dynamics of networks modularized in terms of cortical minicolumns and hypercolumns. Biophysical compartmental models demonstrate attractor dynamics that support cell assembly operations with fast convergence and low firing rates. Using a scaling model we obtain reasonable relative connection densities and amplitudes. An abstract attractor network model reproduces systems level psychological phenomena seen in human memory experiments as the Sternberg and von Restorff effects.

We conclude that there is today considerable substance in Hebb's theory of cell assemblies and its attractor network formulations, and that they have contributed to increasing our understanding of cortical associative memory function.

The criticism raised with regard to biological and psychological plausibility as well as low storage capacity, slow retrieval etc has largely been disproved. Rather, this paradigm has gained further support from new experimental data as well as computational modeling.

Introduction

We are still in an early state with regard to understanding the fundamental principles underlying the function of the human brain. Yet, during the last few decades we have seen a convergence of ideas and hypotheses proposed, though we are still far from consensus. Hebb's now fifty years old conjecture concerning synaptic plasticity and cell assemblies [Hebb, 1949] as mental representations has remained among the most viable

and productive theoretical frameworks. By now Hebbian synapses are well established experimentally [Levy and Steward, 1979]. Hebb's hypothesis has been elaborated and modified considerably since its original formulation [Braitenberg, 1978; Fuster, 1995; Milner, 1957; Palm, 1982; Willshaw and Longuet-Higgins, 1969]. The attractor type of recurrent neural networks formalized by Hopfield [Hopfield, 1982] and further investigated by many workers [Amit, 1995; Hertz et al., 1991] may, in fact, be regarded as mathematical instantiations of Hebb's original idea. This general theoretical framework can provide explanations for Gestalt aspects of active memory like perceptual completion, reconstruction and rivalry, as well as polymodal sensory fusion if cell assemblies are assumed to extend over large parts of cortex.

More recently, evidence from various sources have accumulated demonstrating how artificial neural networks (ANN) of the attractor networks can account for findings ranging from synaptic plasticity and single cell recordings to high level cognitive functions [Eichenbaum, 1993; Pulvermuller, 1999; Quinlan, 1991]. Thus, there is today a substantial body of data that supports this paradigm.

In this paper, we review a number of issues related to the biological plausibility and functionality of cell assembly and attractor network theories of cortical associative memory. We describe and discuss simulations done using network models comprised of biophysically detailed model neurons as well as abstract attractor network models.

Attractor networks embedded in the neocortex

If we are to believe that cell assemblies and attractors have some biological relevance a critical question is: How can an attractor network model be mapped to the structure and function of a piece of neocortex? An attractor network in its generic form has all-to-all connectivity, which is also symmetric, in order to allow for fix-point dynamics. Are there synapses and recurrent circuitry of sufficient abundance, distribution and magnitude in real neocortex and is it in any sense symmetric? Can real neurons operating on a timescale of milliseconds with spike frequencies of some tens of Hz support memory retrieval within psychologically observed response times of a few hundred milliseconds?

Extensive recurrent circuitry can certainly be found in the neocortex, in the intrinsic local connections between nearby cells as well as in the long-range connectivity between cortical areas. Intracortical connections are relatively dense on a local scale and sparse on a more global scale [Gilbert et al., 1990; Palm, 1982]. The horizontal cortical fiber systems extend over large distances and are responsible for up to 80 % of the synapses on some cortical pyramidal cells [LeVay and Gilbert, 1976]. In the primary visual cortex cells with the same orientation selectivity are often connected [Hirsch and Gilbert, 1991]. These connections should be symmetric, if not on a neuron-to-neuron level, at least statistically between e.g. orientation minicolumns. Projections between neocortical areas are extensive [Scannell et al., 1995] and often reciprocal. Thus, the connectivity exists, but to find out if it is sufficiently potent to support cell assembly's dynamics we need more quantitative tools like simulations.

At the time of the first attempts to test Hebb's cell assembly theory by means of computer simulations the prototypical neuron was the spinal motor neuron. Using very simple pulse generating motor neurons with modifiable synapses Rochester et al. [Rochester et al., 1956] failed to demonstrate the formation of Hebbian assemblies. Later MacGregor and McMullen [MacGregor and McMullen, 1978] used a more realistic spiking motor neuron model. They found spike synchronization within

populations of neurons with recurrent excitation, but no signs of afteractivity of the kind hypothesized by Hebb [Hebb, 1949] (page 74).

One early argument against the biological relevance of Hebb's cell assembly theory was this failure to demonstrate afteractivity. However, in some of our own early investigations, using a cell model very similar to the one of MacGregor and McMullen, afteractivity could indeed be produced provided that the model motor neurons were replaced by pyramidal cell type neurons [Lansner, 1982; Lansner and Fransén, 1992]. This result established an intriguing possible relation between cortical neuronal firing properties and global network dynamics.

Compartmental model neurons

In a series of studies we examined recurrent networks employing more biophysically detailed compartmental model neurons. As the starting point we used an attractor neural network with a Bayesian learning rule storing sparse random patterns [Lansner and Ekeberg, 1989]. By replacing the units of the attractor network with model pyramidal cells and inhibitory interneurons we wanted to find out whether or not the network would still operate properly as an associative memory.

The parameters of our model neurons were adjusted using data from regular spiking pyramidal cells (RS) and local fast spiking inhibitory interneurons (FS) of layer 2/3 of neocortex [Connors and Gutnick, 1990; McCormick et al., 1985]. The role of the latter was to take care of the negative connections in the ANN by converting excitatory synapses to inhibitory action. The active cell properties included Na, K, Ca, and Ca dependent channels K (K_{Ca}). The RS-RS and RS-FS glutamate synapses were of a mixed fast AMPA/kainate and slower NMDA type with about equally large peak amplitude postsynaptic potentials (PSP) for each component alone. The FS-RS inhibitory synapse was assumed to be GABA_A-ergic. We took into account the modulatory effects of endogenous neuromodulators such as e.g. serotonin on the K_{Ca} -channels [Wallén et al., 1989] by changing the corresponding conductance in the model¹.

Structure of the network model

The basic unit in our model was the cortical minicolumn, following the view held by Mountcastle [Mountcastle, 1998]. The assumption that about a hundred neurons in the minicolumn operate as a unit may seem unintuitive. A possible reason may be that a single neuron is not by itself able to sustain the extensive connectivity with other units that are required for proper system function.

To limit simulation time the network of minicolumns was subsampled, i.e. each minicolumn had twelve pyramidal cells and three inhibitory interneurons. A small network comprised of fifty minicolumns was studied in detail, but a larger version with 169 minicolumns was later simulated to examine the scaling properties of the model.

Within each minicolumn the RS cells were connected among themselves (about 50% density) thus forming a local recurrent excitatory network. The local FS cells had inhibitory synapses on the soma of the RS cells in the same minicolumn. Intercolumnar, long-range connectivity was exclusively excitatory and much sparser than the local connectivity (about 7% density). Communication latencies were calculated assuming that the minicolumns were spread out over a 7x7 mm patch with 1 mm spacing. The axonal conduction velocity was 1 m/s which gave latency in the range 2 – 11 ms.

¹ These neuromodulators also have other synaptic and cellular effects not included here.

A. Lansner et al.

Values for the peak conductances of synapses between the network units (minicolumns) were adopted from an attractor ANN with 50 units (one for each minicolumn) that had been trained with 8 random patterns with 8 active units in each.

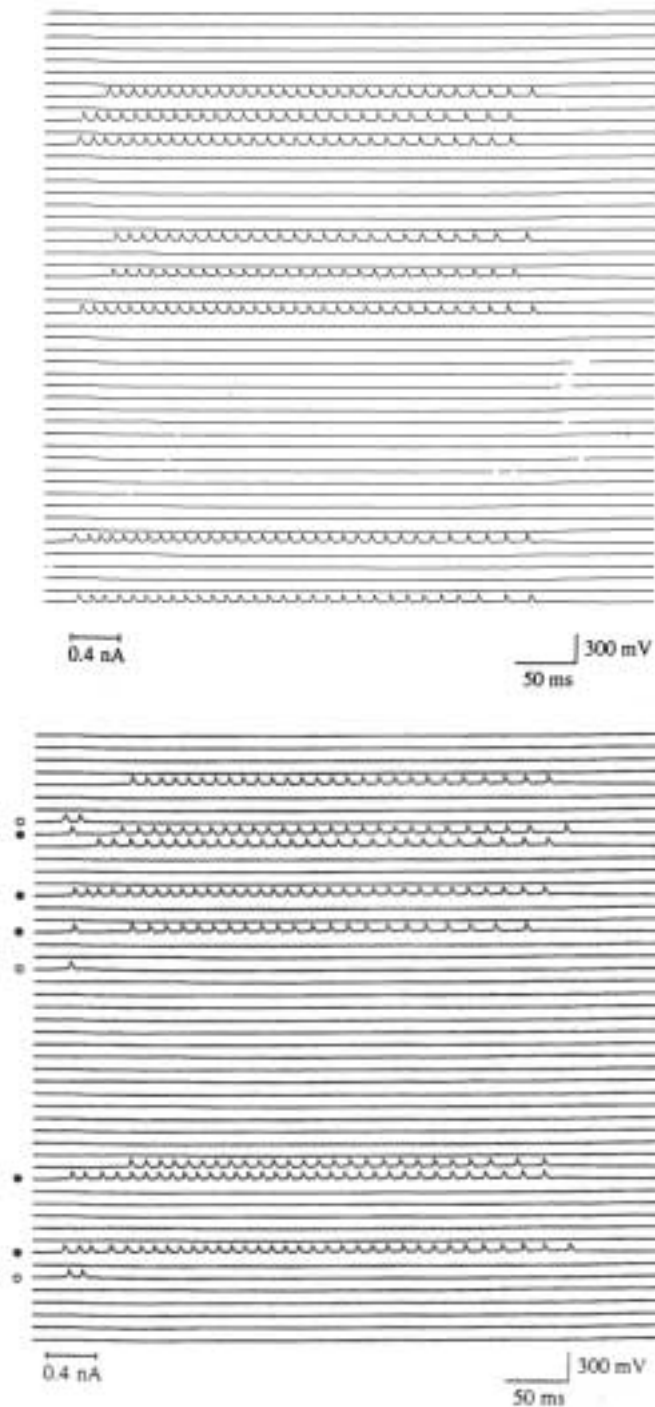


Fig. 1. Pattern completion and rivalry between cell assemblies. Soma membrane potential for 50 RS cells is shown, each one representative of one of 50 minicolumns Upper: Pattern completion and afteractivity in an assembly with five cells stimulated. These get active early on and activate their companion cells indirectly. Lower: Rivalry between two assemblies. The stimulated cells in the two different patterns are marked with open and closed circles (From [Lansner and Fransén, 1992]).

Synaptic noise was present in all cells. In the absence of input, this produced a background spiking activity of some 2-3 Hz in the network. The main point of this was to show that the network could operate at the background noise found in cortex, both for the low state representing no input as well as the high state representing an active assembly. The role of noise was further discussed in Fransén and Lansner (1998).

Simulation results

Simulation of a single minicolumn demonstrated that it was capable of operating as a functional unit in which stimulated RS could activate their non-stimulated companions within some 20-40 ms. No afteractivity was observed for a single active minicolumn.

If a complete assembly (a full pattern of 8 minicolumns) was stimulated, a transient peak of activity occurred, lasting about 100 milliseconds. Activity terminated due to the spike frequency adaptation of the RS cells. A more marked afteractivity for 400-500 ms resulted when the AHP amplitude of the cells was reduced by decreasing the conductance of the K_{Ca} -channel to simulate neuromodulator application [Fransén and Lansner, 1995]. These observations fit nicely with the existence of diffuse monoamine projections of subcortical origin to neocortex that have been implicated in e.g. the regulation of overall activity levels and attention.

The mutual excitation between minicolumns in the cell assembly provides the network with a capability for pattern completion (fig. 1 left). When a subset of the minicolumns in one assembly is stimulated the remaining ones become activated via the intercolumnar excitatory connections. The FS cell activity is essentially the complement of the RS cell activity on the minicolumn level which is consistent with the observation that in the prefrontal cortex closely adjacent pyramidal and non-pyramidal neurons with similar preferences show opposite activity modulation [Goldman-Rakic, 1995].

The spread of activation via overlap between the memorized patterns is prohibited by lateral inhibition between assemblies. This inhibition is essential also when parts of two assemblies are stimulated with conflicting/ambiguous input (fig. 1, right). Eventually one of the assemblies wins and completes its pattern, thus shutting down the activity in the other one. This kind of pattern rivalry may be a physiological correlate of some perceptual rivalry effects.

As can be seen in these simulations, a complete pattern is normally obtained within 30-60 ms. This may thus be accomplished during the initial activity peak without afteractivity. This short time is compatible with reaction times seen in psychophysical experiments [Thorpe et al., 1996; Thorpe and Imbert, 1989], that find a lower limit of 100-150 ms, measured by event-related potentials, for the identification of familiar visual objects.

The powerful positive feedback provided by the mutual excitation in attractor networks as studied here could be expected to drive the neurons to unnaturally high firing rates. Yet, in cortical *in vivo* recordings one rarely sees rates higher than 20-60 Hz [Fuster and Jervey, 1982], although a pyramidal cell can fire up to some 300 Hz or more when stimulated. It has been suggested that fast inhibitory feedback and noise driven activation could explain the low firing rates seen in cortical pyramidal cells [Amit and Tsodyks, 1991]. We found that another possible contributing mechanism might be saturating postsynaptic receptors and a mix of fast AMPA/kainate type receptors and slower NMDA type glutamate receptors [Fransén and Lansner, 1995].

Different types of oscillatory activity and spike synchronization has been observed in many cortical and sub-cortical structures [Freeman and Skarda, 1985; Gray and Singer, 1989; Traub et al., 1996]. Besides the basic pattern processing capabilities, the simple network simulated here also displayed at least two qualitatively different modes of

operation, i.e. bursting and sustained firing with oscillating average frequency [Fransén and Lansner, 1995]. For the sustained firing mode there was a weak tendency for spike synchronization, but for the burst firing mode, the cells were essentially unsynchronized. The tendency to spike synchronize was enhanced at low neuromodulator (e.g. serotonin) levels giving a pronounced AHP, and with short latencies and proximal placement of synapses [Lansner and Fransén, 1992]. Such oscillations and spike synchrony may support associative memory operation by itself or may complement and enhance such functionality already inherent in a graded unit attractor ANN [Sommer and Wennekers, 2001; Yen et al., 1999].

Scaling up the size of the network model

The relative proportions of pyramidal cells and inhibitory interneurons in our model as well as its dense local and sparse global connectivity are in qualitative agreement with neocortical anatomy. However, the way the cell model was matched to single cell data implied that a certain level of synaptic current was needed to drive a cell. This together with the small number of cells and synapses in the model made it necessary to exaggerate synaptic conductances as well as connection densities.

Two important questions then arise: Is it possible that cells in a network of a size corresponding to a macroscopic piece of cortex may see the same total input currents as in our small model network, although via much lower amplitude postsynaptic potentials and larger number of active synapses?

To answer these questions we formulated a number of scaling equations and examined their implications [Fransén, 1996; Fransén and Lansner, 1998]. We demonstrated that our small simulated network model can in fact be scaled up to macroscopic size and that it then predicts reasonable values for numbers and proportions of synapses as well as EPSP and IPSP amplitudes in postsynaptic neurons. The conduction delays of the small network were already matched to a 7x7 mm piece of cortex. Thus, it is not unlikely that the dynamics, e.g. the time to reach an attractor, of the scaled up network is of the same order as that of the smaller simulated ones. The scaling convergence time of a BCPNN with Furthermore, long-range cortico-cortical fibers have much higher conduction velocity than 1m/s so it is in fact possible that cell assemblies could extend over several centimeters. These results strengthen our trust in the relevance of our simulations done with the subsampled models and motivated further studies of the memory functionality of this type of network model.

Associative memory in abstract attractor network models

When examining the cell assembly paradigm, there are further issues that need to be addressed in addition to its compatibility with cortical morphology, neurophysiology and the resulting dynamics. For instance, it has often been claimed that the storage capacity of attractor networks is not high enough, so that even a network of the size of a brain could not hold as many memories as our brain evidently can. Catastrophic forgetting is another issue, i.e. as the number of memorized patterns passes beyond maximum capacity, interference becomes overwhelming and all useful memory properties are lost. These and similar issues have mostly been addressed in the context of more abstract network models. We will in the following review some of these results and also touch upon the relation between detailed compartmental models and reduced, abstract ones.

The BCPNN network architecture and learning rule

We turn to a specific instance of a simulated abstract attractor network based on the Bayesian Confidence Propagation Neural Network, BCPNN [Sandberg et al., 2002]. This model is based on a probabilistic perspective on learning and retrieval, with input and output unit activities representing confidence of feature observation and posterior probabilities of outcomes, respectively [Lansner and Holst, 1996; Sandberg et al., 2002]. Discrete valued attributes are represented as activity patterns in the network in a straight-forward manner. Continuous valued attributes can also be represented if they are first discretized by some form of interval coding², a principle abundantly used in the nervous system.

A BCPNN network will be seen as representing a network somewhat analogous to e.g. a cortical area in the brain. A hypercolumn represents a discrete attribute and there is one minicolumn for each possible value of that attribute (figure 2). Each hypercolumn consists of a central inhibitory (basket) cell population surrounded by minicolumns. A minicolumn unit has graded output between 0 and 1, corresponding to a normalized population frequency. The hypercolumn normalizes the activity so that the outputs of all minicolumns in the network sum to one, i.e. it imposes a soft winner-take-all operation.

A discrete attribute i can take on M_i different values, each one represented by a unit ii' . Given a certain set of attributes X we want to estimate the probability of an attribute y_j , comprised of individual components $y_{jj'}$. If we assume the N attributes x_i to be independent as well as conditionally independent, by using Bayes' rule and taking the logarithm we get the following unit dynamics (for details see [Sandberg et al., 2002]):

$$\tau_m \frac{dh_{jj'}(t)}{dt} = \beta_{jj'} + \sum_i^N \log \left(\sum_{i'}^{M_i} w_{ii',jj'} \hat{\pi}_{ii'}(t) \right) - h_{jj'}(t) \quad (1)$$

with weights and bias as:

$$\beta_{jj'} = \log P(y_{jj'}) \quad w_{ii',jj'} = \frac{P(y_{jj'} | x_{ii'})}{P(y_{jj'})}. \quad (2)$$

The variable h corresponds to unit potential and the different $P(\cdot)$ are estimates from the training set of activation rates of the pre- (P_i) and postsynaptic (P_j) unit respectively and their co-activation (P_{ij}). The network units have an exponential activation function and the activity within a hypercolumns is normalized, thus giving:

$$\pi_{ii'} = f(h_{ii'}) = e^{h_{ii'}} \quad \hat{\pi}_{ii'} = \frac{\pi_{ii'}}{\sum_{i'} \pi_{ii'}} \quad (3)$$

Here, $\hat{\pi}_{ii'}$ serves as our estimate of y_i . From the expression for the weights above (2) it can be seen that the weight matrix is symmetric, thus allowing for a well-behaved fixed-point attractor dynamics in a recurrent network..

² Sensory receptors and neurons are often maximally activated by one value of the variable sensed (orientation, light wavelength, sound frequency, knee joint angle, skin temperature) and partially activated by values in an interval surrounding it.

The computational structure resulting from these assumptions relate in an interesting way to cortical functional architecture. As already noted, the basic computational unit of the neocortex is most likely the minicolumn. However, the minicolumns tend to aggregate into larger modules, hypercolumns also referred to as macrocolumns or simply columns by different authors [Hubel and Wiesel, 1977; Mountcastle, 1957]. The hypercolumn has later been proposed as the minimal repetitive microcircuit of the neocortex [Mountcastle, 1998]. It has a diameter of some 500 μm and comprises on the order of some hundred minicolumns. The pyramidal and basket cells within a distance corresponding to the diameter of the hypercolumn form extensive reciprocal connections.

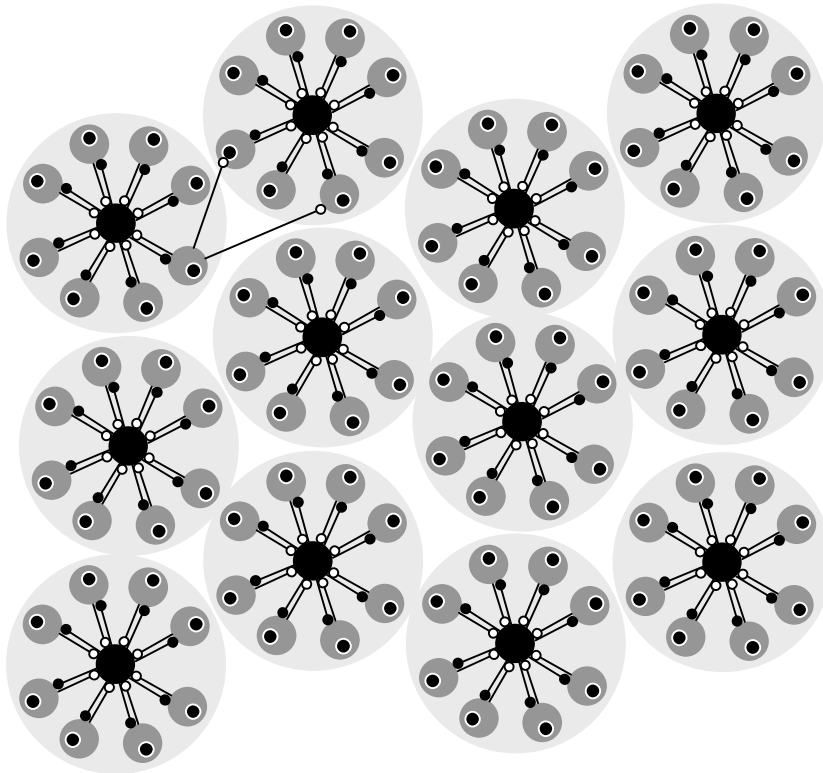


Fig. 2. Schematic structure of the BCPNN network. Each hypercolumn consists of a central inhibitory (basket) cell population (larger filled black circle) surrounded by minicolumns. Each filled gray circle with a small black-white dot represents a minicolumn with its pyramidal cell population (gray) and its local inhibitory interneurons (black-white). The basket cell population is reciprocally connected to all minicolumns in the hypercolumn. Connections within the hypercolumns are fixed, but each minicolumn connects to every other one except those within its own hypercolumn. A positive connection from one minicolumn to another is direct from pyramidal cell to pyramidal cell, whereas a negative connection goes via the local inhibitory interneurons (see upper left). A “connection” in this context represents a sparse synaptic connectivity between the neurons in the two minicolumns.

The precise functional implications of this architecture remain unclear. In the visual cortex, the so-called normalization models propose that the summed activity of the cells in a hypercolumn is normalized, possibly by shunting inhibition via basket cells [Carandini et al., 1997]. This is also how the BCPNN hypercolumn operates.

Storage capacity

The memory capacity of attractor networks in terms of the number of patterns that can be stored and reliably retrieved has been investigated in detail by many researchers. In short, given sparse activation the storage capacity in bits scales with the number of synapses, i.e. the size of available physical memory [Amari, 1989; Lansner and Ekeberg, 1989; Palm, 1982]. This is just about as good as one can get in any type of memory. The storage capacity of the hypercolumnar network described here behaves in the same way. On the order of one hundred minicolumns per hypercolumn gives an activity level of about 1%. If the theoretical relations obtained are scaled to networks with sizes of the order of the number of minicolumns in the human brain, we get estimates of some billion patterns. While this is a very crude estimate, it is of the same order as the number of seconds in a human lifetime and hence consistent as an upper bound for the number of experiences of an individual.

Catastrophic forgetting and incremental learning

The standard correlation based learning rules for attractor ANN typically suffer from catastrophic forgetting. To cope with this situation, so called marginalist learning [Nadal et al., 1986] and learning within bounds [Hopfield, 1982] have been proposed. An incremental version of the BCPNN learning rule that uses leaky counters gives the same effect as will be shown [Lansner and Sandberg, 2001; Sandberg et al., 2002]:

$$\tau_L \frac{d\hat{P}_i}{dt} = \kappa(\pi_i - \hat{P}_i) \quad \tau_L \frac{d\hat{P}_j}{dt} = \kappa(\pi_j - \hat{P}_j) \quad \tau_L \frac{d\hat{P}_{ij}}{dt} = \kappa(\pi_i \pi_j - \hat{P}_{ij}) \quad (4)$$

This can also be turned into a more elaborate form with two more state variables computed as running averages thus giving a more flexible and possibly asymmetric measure of simultaneity and allowing for a delayed print-now signal [Lansner and Sandberg, 2001]. A slightly modified version of this learning rule also matches quantitatively experimental data on spike timing dependent synaptic plasticity [Wahlgren, 2000].

Simulation results

This network architecture and learning rule has been introduced in a recurrent attractor network to evaluate its performance as well as learning and forgetting dynamics [Sandberg and Lansner, 2002; Sandberg et al., 2002; Sandberg et al., 2001]. The incremental learning rule avoids catastrophic interference by gradually forgetting the oldest patterns while recently learned patterns remain accessible. At the same time it retains the same maximal capacity as the non-incremental version (in the limit of infinite τ_L). The forgetting dynamics could be demonstrated by varying the parameter τ_L the (figure 3, left).

Using this learning rule, as more patterns are learned, the basins of attraction of older patterns get smaller and more “shallow” which results in a decrease in convergence speed. Between just a few learned patterns and a maximal memory load there is a linear increase in median convergence time up to a doubling. Interestingly, this reproduces quantitatively Sternberg’s classical finding of a reaction time dependence on list length [Sternberg, 1966] in experiments on human subjects. Thus, the effects seen can be explained by a parallel associative retrieval process in an attractor network, provided it

has a fast Hebbian learning capability, for which there is already some experimental support [Bao et al., 1997].

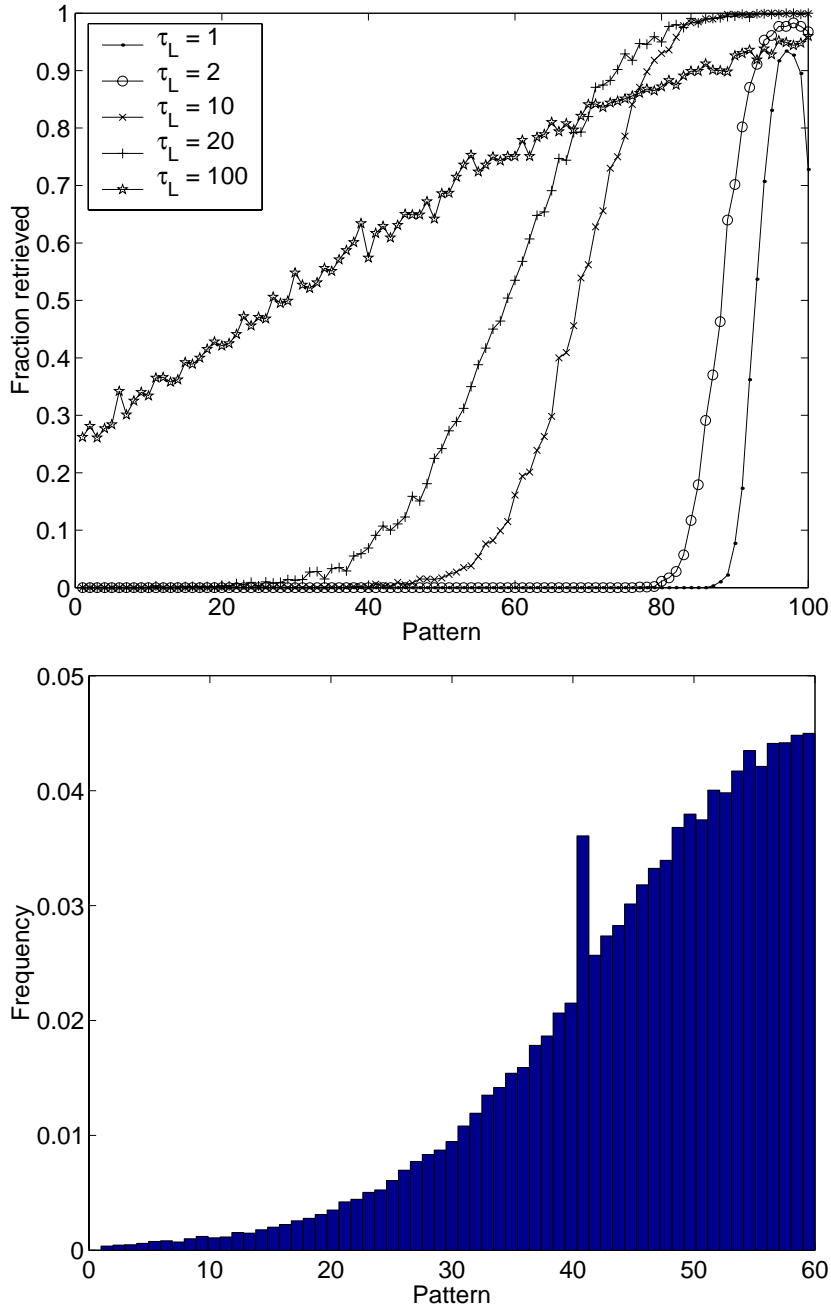


Fig. 3. Forgetting curve after memory modulation. Left: Pattern recall as a function of list position for different values of α (see legend) in a network with 100 units and 10 hypercolumns (a 10x10 architecture). Recall is estimated as the frequency of retrieval with overlap greater than 0.85 after a presentation of a pattern where the activation of two hypercolumns had been randomly changed. Right: 60 patterns were stored in the same network. For pattern number 41 the print now signal κ was double that for the other patterns. This represents the “isolate” in the von Restorff terminology.

It is well known that endogenous processes activated by experience can modulate memory strength in terms of recall probability [Martinez et al., 1991; McGaugh, 2000]. The novelty of a stimulus also plays an important role. Some of these effects can be interpreted as a relevance modulation of the “print-now” signal regulating memory encoding and synaptic plasticity [Wickens and Kötter, 1995]. This relates to the von Restorff effect that shows up as an improved memory of a distinctly different item (the isolate) in a set, while the other items are less well retained [von Restorff, 1933]. This effect can be reproduced in simulations by modulating the print-now signal κ in our incremental learning rule (figure 3, right).

In biological neurons, cellular adaptation and synaptic depression mechanisms with time constants in the hundred-millisecond range can serve to make fixpoints metastable as shown in previous simulations [Cartling, 1993; Cartling, 1995; Fransén and Lansner, 1995; Lansner and Fransén, 1992]. In the present abstract model, we added one more building block, i.e. a fast modifying recurrent BCPNN projection. Its weight is subtracted from the learned connection strength to give the effective weight. This adaptation had a time constant τ_L of 160 ms, corresponding to the decay rate of the slow AHP in the previous biophysically detailed pyramidal cell model.

This somewhat abstract implementation of adaptation is motivated by the urge to reuse the same building blocks as before, to avoid introducing additional equations and parameters in the model. The net result is that the summed adaptation and depression regulates network dynamics in a highly balanced fashion [Sandberg et al., 2001].

As shown in figure 4 (left), a freely running network visits patterns one by one, staying in each for a couple of hundred milliseconds. Preliminary results show that patterns are represented in free recall proportionally to how strongly they have been imprinted [Sandberg et al., 2001].

To demonstrate the robustness of the operation of this network and to move somewhat closer to biology we performed the same simulation as in figure 4 (left) but with spiking network units (figure 4, right). Despite very low frequency spiking³, network operation maintained qualitatively the same behavior as with graded output units. Moreover, the h variable corresponding to the intracellular potential of a neuron in the minicolumn looked intriguingly similar to what is seen in intracellular potential recordings from pyramidal cells.

Conclusions and Discussion

We have surveyed research aimed at demonstrating how the cell assembly theory and its mathematical formulations as attractor networks contribute to our understanding of cortical associative memory. We have found that the criticisms raised with regard to biological and psychological plausibility as well as theoretical shortcomings with regard to storage capacity, speed of retrieval etc. have largely been disproved. Rather, this paradigm has gained further support from new experimental data as well as computational modeling.

We have demonstrated that biologically detailed models show attractor dynamics that can reproduce cell activities like those seen e.g. in prefrontal working memory experiments both *in vivo* and *in vitro*. Experimentally measured connection densities and relative cell and synapse counts can be incorporated into a scaling model to predict e.g. magnitudes of postsynaptic potentials that are close to experimental estimates.

³ Since a minicolumn comprises some hundred neurons its effective spiking frequency should be in the order of a hundred times that of a single neuron.

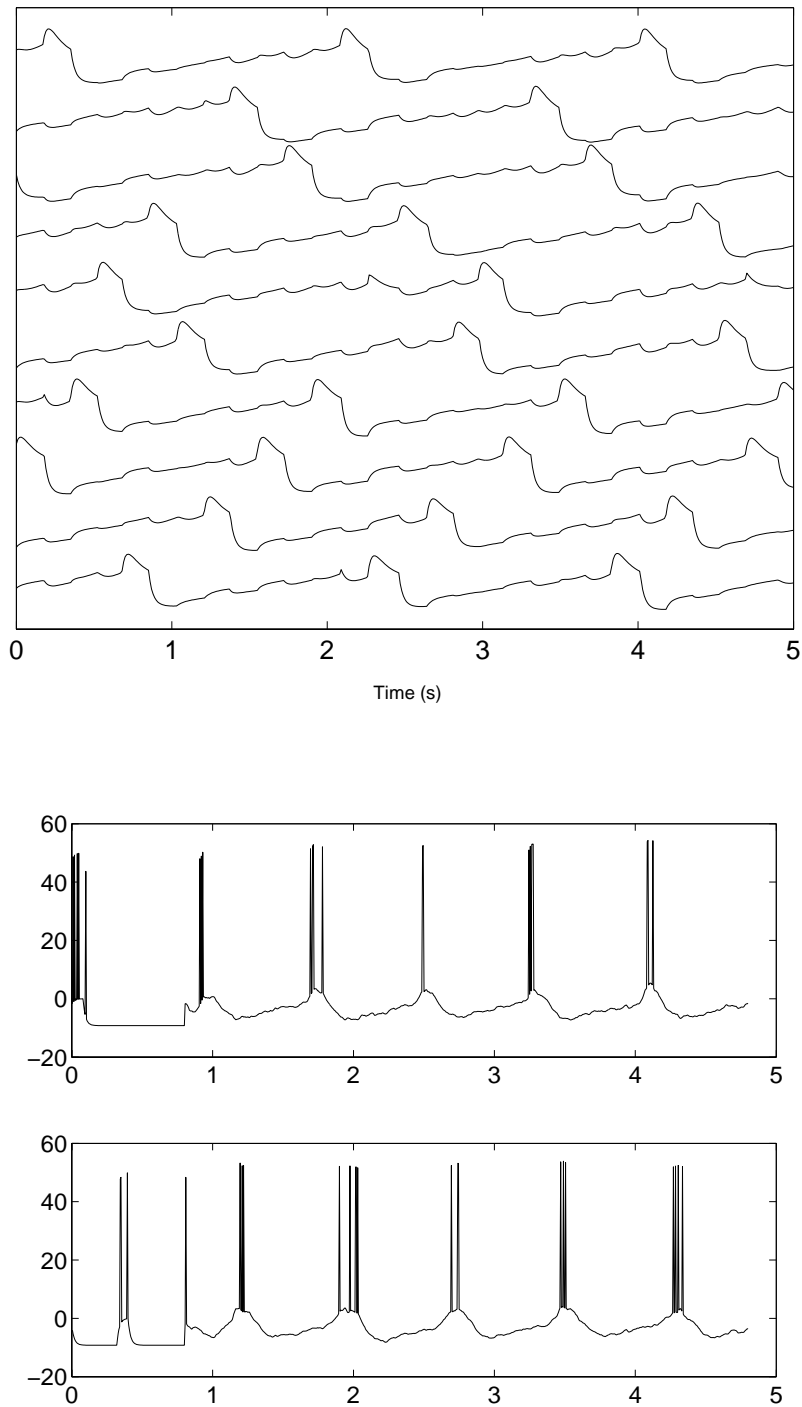


Fig. 4. Dynamics of metastable attractors. Left: A 10x10 network was trained with 10 orthogonal patterns. It was then allowed to run freely for 5 seconds. The traces show the h variable (corresponding to intracellular potential) of one unit from each pattern (all units in one pattern behaved the same). The traces have been shifted vertically for visibility. Right: Output from two different units in a spiking implementation of the same network. The output variable was used as parameter in a spike generating Poisson process. The spikes have been pasted on top of the h variable of each unit.

The incremental BCPNN architecture with palimpsest properties described here suggests the possibility to unify memories on different timescales, from long-term to

working memory and it provides a handle to the important influences of reward systems and neuromodulation on learning. Moreover, it reproduces systems level psychological phenomena seen in human memory experiments as the Sternberg and von Restorff effects. These results lend further support to attractor network models of the neocortex.

Obviously, the simulated networks are not just much smaller but also much less complex and intricate compared to the neocortex. They still lack many important aspects of cortical memory processes. Important issues for future research include temporal sequence learning and processing capability. Another equally important capability is that of experience dependent self-organization of hypercolumnar structure and unit response properties and the ability to reorganize in case of damage [Merzenich et al., 1984]. Theoretically and neurobiologically well founded work in this direction is already ongoing [Miller, 1995; Olshausen and Field, 1997]. Other issues that call for further investigation include, for instance: Is there always one active global assembly or can there be several at the same time? Can the dynamics be decoupled into independent subcomponents? Can there be meaningful attractor states that were never explicitly laid down as memories? How can invariant representation and perception emerge in this kind of networks? These and other open issues calls for further investigation.

The complexity is further raised significantly as we turn to multi-network structures that build systems specialized for functions such as perception, motor control, and behavior selection that also are capable of interacting in a synergistic fashion. Only when we reach such a level of sophistication can we approach constructs of a complexity and capacity on par with that of real brains.

Acknowledgements

Support for this research from the Swedish Natural Science Foundation (NFR), the Swedish Research Council for Engineering Sciences (TFR), the Swedish Medical Science Foundation (MFR), and the Swedish Science Research Council (VR) is gratefully acknowledged.

References

- Amari, S. I. (1989). "Characteristics of sparsely encoded associative memory." *Neural Networks* 5(451-457).
- Amit, D. J. (1995). "The Hebbian paradigm reintegrated: Local reverberations as internal representations." *Behavioral and Brain Sciences* 18: 617-657.
- Amit, D. J. and M. V. Tsodyks (1991). "Quantitative study of attractor neural network retrieving at low spike rates: I Substrate — spikes, rates and neuronal gain." *Network: Computation in neural systems* 2: 259-273.
- Bao, J.-X., E. R. Kandel and R. D. Hawkins (1997). "Involvement of pre- and postsynaptic mechanisms in posttetanic potentiation at Aplysia synapses." *Science* 275: 969-973.
- Braitenberg, V. (1978). "Cell assemblies in the cerebral cortex". *Theoretical Approaches to Complex Systems*. R. Heim and G. Palm, Eds. Berlin, Springer: 171-188.
- Carandini, M., D. J. Heeger and J. A. Movshon (1997). "Linearity and normalization in simple cells of the macaque primary visual cortex." *J. of Neurosci.* 17: 8621-44.
- Cartling, B. (1993). "Control of the complexity of associative memory dynamics by neuronal adaptation." *Int. J. Neural Systems* 4: 129-141.
- Cartling, B. (1995). "Autonomous neuromodulatory control of associative processes." *Network: Computation in Neural Systems* 6(247-260).
- Connors, B. W. and M. J. Gutnick (1990). "Intrinsic firing patterns of diverse neocortical neurons." *TINS* 13: 99-104.
- Eichenbaum, H. (1993). "Thinking About Brain Cell Assemblies." *Science* 261: 993-994.

A. Lansner et al.

- Fransén, E. (1996). "Biophysical Simulation of Cortical Associative Memory", Royal Institute of Technology, Stockholm, Sweden, Dept. of Numerical Analysis and Computing Science.
- Fransén, E. and A. Lansner (1995). "Low spiking rates in a population of mutually exciting pyramidal cells." *Network: Computation in Neural Systems* **6**: 271-288.
- Fransén, E. and A. Lansner (1998). "A model of cortical associative memory based on a horizontal network of connected columns." *Network: Computation in Neural Systems* **9**: 235-264.
- Freeman, W. and C. Skarda (1985). "Spatial EEG patterns, non-linear dynamics and perception: the neo-Sherringtonian view." *Brain Res. Rev.* **10**: 147-175.
- Fuster, J. and J. P. Jervey (1982). "Neuronal firing in the inferotemporal cortex of the monkey in a visual task." *J. Neurosci.* **2**: 361-375.
- Fuster, J. M. (1995). "Memory in the Cerebral Cortex". Cambridge, Massachusetts, The MIT Press.
- Gilbert, C. D., J. A. Hirsch and T. N. Wiesel (1990). "Lateral interactions in the visual cortex". *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press. **LV**: 663-676.
- Goldman-Rakic, P. S. (1995). "Cellular basis of working memory." *Neuron* **14**: 477-485.
- Gray, C. M. and W. Singer (1989). "Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex." *Proc. Natl. Acad. Sci. USA* **86**: 1698-1702.
- Hebb, D. O. (1949). "The Organization of Behavior". New York, John Wiley Inc.
- Hertz, J., A. Krogh and R. G. Palmer (1991). "Introduction to the theory of neural computation". Redwood City, California, Addison-Wesley Publishing Company.
- Hirsch, J. A. and C. D. Gilbert (1991). "Synaptic physiology of horizontal connections in the cats visual cortex." *J. Neurosci.* **11**(6): 1800-1809.
- Hopfield, J. J. (1982). "Neural Networks and physical systems with emergent collective computational properties." *Proceedings of the National Academy of Sciences, USA* **81**: 3088-3092.
- Hubel, D. and T. N. Wiesel (1977). "The functional architecture of the macaque visual cortex. The Ferrier lecture." *Proc. Royal. Soc. B* **198**: 1-59.
- Lansner, A. (1982). "Information processing in a network of model neurons. A computer simulation study". Tech. rep., Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, (TRITA-NA-8211).
- Lansner, A. and Ö. Ekeberg (1989). "A one-layer feedback artificial neural network with a Bayesian learning rule." *Int. J. Neural Systems* **1**: 77-87.
- Lansner, A. and E. Fransén (1992). "Modeling Hebbian cell assemblies comprised of cortical neurons." *Network: Computation in Neural Systems* **3**: 105-119.
- Lansner, A. and A. Holst (1996). "A higher order Bayesian neural network with spiking units." *Int. J. Neural Systems* **7**(2): 115-128.
- Lansner, A. and A. Sandberg (2001). "Functionality and Performance of Brain-Inspired Neural Networks". NOLTA 2001, Zao, Sendai, Japan: 501-504.
- LeVay, S. and C. D. Gilbert (1976). "Laminar patterns of geniculocortical projections in the cat." *Brain Res.* **113**(1-19).
- Levy, W. B. and O. Steward (1979). "Synapses as associative memory elements in the hippocampal formation." *Brain. Res.* **175**: 233-245.
- MacGregor, R. J. and T. McMullen (1978). "Computer simulations of diffusely connected neuronal populations." *Biol. Cybernetics* **28**: 121-127.
- Martinez, J. L., G. Schulteis and S. B. Weinberger (1991). *Learning and Memory: A Biological View*. R. P. Kesner, Ed, Academic Press: 149-198.
- McCormick, D. A., B. W. Connors and J. W. Lighthall (1985). "Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex." *J. Neurophysiol.* **54**: 782-806.
- McGaugh, J. L. (2000). "Memory -- a Century of Consolidation." *Science* **287**: 248-251.
- Merzenich, M. M., R. J. Nelson, M. P. Stryker, M. S. Cynander, A. Schoppmann and J. M. Zook (1984). "Somatosensory map changes following digit amputation in adult monkey." *J. Comp. Neurol.* **224**: 591-605.
- Miller, K. D. (1995). "Receptive Fields and Maps in the Visual Cortex: Models of Ocular Dominance and Orientation Columns". *Models of Neural Networks III*, Eds. (, NY), pp. 55-78. K. Schulten, Ed. New York, Springer-Verlag: 55-78.
- Milner, P. M. (1957). "The cell assembly: Mark II." *Psychological Review* **64**: 242-252.
- Mountcastle, V. B. (1957). "Modality and topographic properties of single neurons of cat's somatic sensory cortex." *J. Neurophysiol.* **20**: 408-434.
- Mountcastle, V. B. (1998). "Perceptual neuroscience: The Cerebral Cortex". Cambridge, Massachusetts, Harvard University Press.

A. Lansner et al.

- Nadal, J. P., G. Toulouse, J.-P. Changeux and S. Dehaene (1986). "Networks of formal neurons and palimpsests." *Europhys. Lett.* **1**: 535-42.
- Olshausen, B. A. and D. J. Field (1997). "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research* **37**: 3311-3325.
- Palm, G. (1982). "Neural Assemblies. An Alternative Approach to Artificial Intelligence". Berlin, Springer.
- Pulvermuller, F. (1999). "Words in the brain's language." *Behavioral and Brain Sciences* **22**: 280-336.
- Quinlan, P. (1991). "Connectionism and Psychology. A Psychological Perspective on Connectionist Research". New York, Harvester, Wheatsheaf.
- Rochester, N., J. H. Holland, L. H. Haibt and W. L. Duda (1956). "Tests on a cell assembly theory of the action of the brain, using a large digital computer." *IRE Trans. Information theory* **IT-2**: 80-93.
- Sandberg, A. and A. Lansner (2002). "Synaptic Depression as an Intrinsic Driver of Reinstatement Dynamics in an Attractor Network." *Neurocomputing* **44-46**: 615-622.
- Sandberg, A., A. Lansner, K.-M. Petersson and Ö. Ekeberg (2002). "Bayesian attractor networks with incremental learning." *Network: Computation in neural systems* **13**(2): 179-194.
- Sandberg, A., K. M. Petersson and A. Lansner (2001). "Selective Enhancement of Recall through Plasticity Modulation in an Autoassociative Memory." *Neurocomputing* **38-40**: 867-873.
- Scannell, J. W., C. Blakemore and M. P. Young (1995). "Analysis of connectivity in the cat cerebral cortex." *J. Neurosci.* **15**: 1463-1483.
- Sommer, F. T. and T. Wennekers (2001). "Associative memory in networks of spiking neurons." *Neural Networks* **14**: 825-834.
- Sternberg, S. (1966). "High-speed scanning in human memory." *Science* **153**: 652-654.
- Thorpe, S., D. Fize and C. Marlot (1996). "Speed of processing in the human visual system." *Nature* **381**: 520-522.
- Thorpe, S. and M. Imbert (1989). "Biological constraints on connectionist modelling". *Connectionism in Perspective*. R. Pfeiffer, Ed. Berlin, Springer-Verlag.
- Traub, R. D., M. A. Whittington, S. B. Colling, G. Buzsaki and J. G. R. Jeffreys (1996). "Analysis of gamma rhythms in the rat hippocampus in vitro and in vivo." *J. Physiol.* **493**: 471-84.
- Wahlgren, N. (2000). "A comparison between biological learning and Hebbian-Bayesian incremental learning". Dept. Numerical Analysis and Computing Science, Stockholm, (TRITA-NA-E200X).
- Wallén, P., J. Buchanan, S. Grillner, J. Christenson and T. Hökfelt (1989). "The effects of 5-hydroxytryptamine on the afterhyperpolarisation, spike frequency regulation and oscillatory membrane properties in lamprey spinal cord neurons." *J. Neurophysiol.* **61**: 759-768.
- Wickens, J. and R. Kötter (1995). "Cellular Models of Reinforcement". *Models of Information Processing in the Basal Ganglia*. J. C. Houk, J. L. Davis and D. G. Beiser, Eds. Cambridge, Massachusetts, The MIT Press: 187-214.
- Willshaw, D. J. and H. C. Longuet-Higgins (1969). "Associative memory models". *Machine Learning 5*. Meltzer and Michie, Eds. Edinburgh, Scotland, Edinburgh University Press.
- von Restorff, H. (1933). "Analyse von Vorgängen in Spurenfeld." *Psychol. Forschung* **18**: 299-342.
- Yen, S.-C., E. D. Menschik and L. H. Finkel (1999). "Perceptual grouping in striate cortical networks mediated by synchronization and desynchronization." *Neurocomputing* **26-27**(1-3): 609-616.