Neural Networks 37 (2013) 165–171
Text

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Neural associative memories and sparse coding

Günther Palm

University of Ulm, Institute of Neural Information Processing, D-89069 Ulm, Germany

ARTICLE INFO

Keywords:
Associative memory
Sparse coding
Cortical networks

ABSTRACT

The theoretical, practical and technical development of neural associative memories during the last 40 years is described. The importance of sparse coding of associative memory patterns is pointed out. The use of associative memory networks for large scale brain modeling is also mentioned.

1. Introduction

Associative memory has been an active topic of research for more than 50 years and is still investigated both in neuroscience and in artificial neural networks. The workings of associations in human memory have probably first been addressed in psychology and even philosophy (David Hume has already stated rules of association).

The basic observation of association occurs when we try to find a specific piece of information in our memory and we do not retrieve it immediately. In such cases we notice that the present state of our mind or brain which presumably contains aspects of the present situation and contextual information pointing at the missing piece (momentarily however not sufficient to find it), starts a sequential process of associations from one item to the next (possibly governed by semantic similarity) that eventually ends with the missing piece. Once this piece of information is there, we immediately recognize it as what we have been searching for, since it somehow *fits perfectly* into the context that triggered our mental search. So there seems to be a kind of information system in our brain (see Fig. 1) that associates a new output to a given input depending on contextual information (and perhaps its own previous state).

From a technical point of view there are two different mechanisms that are needed in this process of association: *hetero-association*, that leads from one pattern to the next, and *auto-association* from one pattern to itself, that is useful for the recognition of one pattern as best fitting, or also for slight correction or completion of this pattern, and thereby ending the chain of (hetero-) associations. There is another technical type of associative memory that is often mentioned (and, in principle, could be regarded as a special case of auto-association), namely *bidirectional association* that goes back and forth between two patterns *A* and *B*. Simple graphical representations of these three types of associative memories are shown in Fig. 2.

Neuroscientists are typically not content with a mere phenomenological description of the process of association in the mind on a cognitive level, they want to relate it to concrete neurophysiological mechanisms in the brain. The first concrete hypothesis in this direction goes back to the psychologist Donald Hebb (1949) who formulated a rule for *synaptic plasticity* that postulates an increase in synaptic connection strength induced by coincident activity in the two neurons connected by the synapse. This idea has led to a lot of very fruitful experimental investigations that eventually confirmed Hebb's ideas (see Caporale & Dan, 2008 for a recent review). On the technical side this has led to the development of *Neural Associative Memory* (NAM) models based on matrix calculus, where a memory storage matrix $C = (c_{ij})$ is formed that contains the weights c_{ij} of synaptic connections between neurons i and j. In hetero-association these connections connect an input pool of neurons (containing neuron i) to an output pool of neurons (containing neuron j). In auto-association they typically connect one pool of neurons back to itself in a recurrent fashion. The roots for this "basic NAM formalism" can be found in engineering (Steinbuch, 1961), in particular in holographic memory implementations (Gabor, 1969; Longuet-Higgins, Willshaw, & Buneman, 1970), and in early neural network modeling (e.g. Amari, 1972; Anderson, 1968; Anderson, Silverstein, Ritz, & Jones, 1977; Dunin-Barkowski & Larionova, 1985; Little, 1974; Marr, 1969; Wigström, 1975). The first systematic overviews were given by Kohonen (1977) and Palm (1980).

2. Basic NAM formalism

Given a set of (pairs of) patterns (x^μ, y^μ) to be stored in the matrix C, the process of storage (matrix formation (1)) and retrieval (activity propagation (2)) can usually be described by the following equations

$$c_{ij} = \sum_\mu x_i^\mu y_j^\mu \quad \text{or} \quad c_{ij} = \max_\mu x_i^\mu y_j^\mu \qquad (1)$$
$$\text{(additive rule)} \qquad \text{(binary rule)}$$

$$u = xC \quad \text{and} \quad y_j = \begin{cases} 1 & \text{if } u_j \geq \theta \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

E-mail addresses: palm@neuro.informatik.uni-ulm.de, guenther.palm@uni-ulm.de.

0893-6080/$ – see front matter
doi:10.1016/j.neunet.2012.08.013

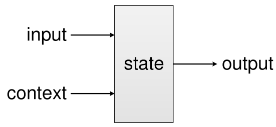


Fig. 1. Module that can be used in the process of association.

Here θ is an appropriately chosen threshold.

Considering a sequential memory storage process it is natural to describe the formation of the matrix C as a sequence of memory or learning steps in which one more pair (x^μ, y^μ) is added to the memory. In each learning step (at time t) the change Δc_{ij} of the entry c_{ij} of the matrix C depends only on the product $x_i^\mu y_j^\mu$ of the i-th coordinate of the input x and the j-th coordinate of the output y, i.e. only on the presynaptic activity x_i and the postsynaptic activity y_j at the synapse modeled by c_{ij} at time t. Thus the synaptic changes can be computed locally in space and time, and Eq. (1) is called a *local learning rule* (Palm, 1982, 1992). Local learning rules are biologically plausible and computationally simple; in particular they are useful for parallel implementation.

In NAM systems the output of the neurons is usually considered as binary. In the past either $\{-1, 1\}$ or $\{0, 1\}$ have been used as the binary values of the stored and retrieved patterns. The use of $\{-1, 1\}$ goes back to John Hopfield (1982). His version of an additive NAM, the Hopfield memory model, has turned out to be quite inefficient as a memory. This is due to two factors:

1. The "Hopfield learning rule" $\Delta c_{ij} = x_i y_j$ for $x_i, y_j \in \{-1, 1\}$ changes every entry c_{ij} of the matrix C in every learning step.
2. The changes go in both directions (up and down), so they can cancel each other.

This is actually quite different for the "Hebb learning rule" $\Delta c_{ij} = x_i y_j$ for $x_i, y_j \in \{0, 1\}$ which has been considered in earlier investigations of NAM going back to Steinbuch (1961) with a first asymptotic analysis given by Willshaw, Buneman, and Longuet-Higgins (1969), see also Willshaw (1971).

Towards the end of the 1980s it became clear that $\{0, 1\}$ and the corresponding Hebb learning rule should be used in practical applications and that *sparseness* of the stored patterns is most important for an effective use of NAMs for information storage and retrieval (Palm, 1988, 1990; Tsodyks & Feigelman, 1988). The importance of sparseness was already implicit in the early analysis of Willshaw (1971), but it was only made explicit a few years later by myself (Palm, 1980, 1985, 1987). Sparseness is the basis for the efficiency of technical applications and VLSI realizations of NAMs (see Palm & Bonhoeffer, 1984, US Patent No. 4777622 (1988) and Palm & Palm, 1991).

3. Information capacity and critical capacity

In order to demonstrate the importance of *sparseness* in associative memory patterns and to prove the efficiency of sparse NAMs it was necessary to develop a clean definition of the *information capacity* of NAMs. Such a definition is best given in terms of information theory, considering the total amount of information that can effectively be stored in and retrieved from an associative memory matrix of a given size (Palm, 1980, 1992; Palm & Sommer, 1992, 1995). Using proper definitions it was possible to show that an asymptotical (large memory matrices) optimal capacity of $\ln 2 \approx 0.69$ bit per matrix entry can be achieved for sparse memory patterns with the binary storage version (Palm, 1980), and $1/(2 \ln 2) \approx 0.72$ bit per matrix entry can be achieved with the additive storage version (Palm, 1988, 1990). The difference between these two values is quite small; in the binary version, however, one clearly needs just one hardware bit for one matrix entry, whereas one needs somewhat more hardware bits per entry in the additive version. These results were actually calculated for hetero-association; they also hold for bidirectional association (Sommer & Palm, 1999); for auto-association the information capacity is just half as large, corresponding also to the symmetry of the memory matrix C. Many more results concerning information capacity have been summarized in my earlier review article (Palm, 1991).

Initiated by the paper of Hopfield (1982) many theoretical physicists became interested in associative memory and applied methods from the theory of spin-glasses to the analysis of feedback auto-associative memories (Fig. 2(b) or (c)) as dynamical systems with a nice natural energy function

$$H(x) = -xCx^T$$

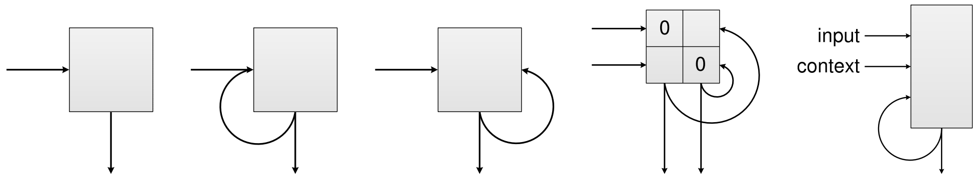
governing the asymptotic behavior, resulting (for symmetric C) in an attractor dynamics towards the minima of $H(x)$ as fixed points (Amit, 1989; Amit, Gutfreund, & Sompolinsky, 1987; Domany, van Hemmen, & Schulten, 1991; Hopfield, 1982). Concerning the use of these systems as practical associative memories, the most important questions are

1. Can we construct the matrix C in such a way that a predescribed set M of memory vectors become fixed points?
2. How large is M as compared to the matrix dimension or network size n (the *critical capacity* $\alpha = M/n$)?
3. How large are the "basins of attraction", i.e. how many errors in an input pattern can be corrected by the feedback retrieval dynamics?

The first two questions were studied extensively. In a nutshell the two most important results are that $\alpha = 2$ can be reached asymptotically in principle (Gardner, 1987, 1988), but there is no local rule to construct the appropriate matrix C from the memory set M and the entries of the matrix C need to be stored with high accuracy. And secondly, $\alpha \approx 0.14$ can be reached with the additive Hopfield rule (Eq. (1)) (Amit et al., 1987; Hopfield, 1982). Also the binary (or "clipped") storage version has been considered, leading to considerably lower values for α (e.g. $\alpha \approx 0.83$ for the non-constructive binary case was found by Krauth & Mézard, 1989).

The third question has also been studied (e.g. Horner, Bormann, Frick, Kinzelbach, & Schmidt, 1989; Nadal & Toulouse, 1990; Opper, Kleinz, Kohler, & Kinzel, 1989). However, satisfactory answers were at first only found for the non-local construction of C and only for α-values that are considerably lower than the critical capacity (for the binary memory matrix reasonable error correction is possible for α-values up to about 0.3 Opper et al., 1989), allowing a possible correction of about 4% of the entries. In this case ($\alpha = 0.3$) the *information capacity* is about 0.06. Numerical simulations of the Hopfield memory show that the basins of attraction are surprisingly small, corresponding to an *information capacity* of less than 0.04. Perhaps the most important reason for this is the large number of spurious (i.e. unwanted) fixed points that are created by the Hopfield rule. Their number seems to increase exponentially with the size of M.

From the application perspective the results for the Hopfield rule are far below the results achieved with the Hebb rule (e.g. 0.14 vs. 0.69). On top of that, there are two additional reasons why the *information capacity* (total information divided by n^2) is usually considerably smaller than the *critical capacity* of the same memory: First the information that can be extracted from a fixed point is always less and usually much less than n bits. Thus the information capacity will be much less than the critical capacity. Secondly, for sparse binary patterns with a probability p for a 1-entry, the total information content of one pattern is roughly $-np \log p$ which is again less than n. The second effect has been incorporated into the definition of the *critical capacity* α for sparse (or biased) memory patterns (e.g. Gardner, 1987).



(a) Hetero-association. (b) Auto-association. (c) Same as (b). (d) Bidirectional association. (e) Associative memory for Fig. 1.

Fig. 2. Different types of associative memories. Auto-association is typically used with feedback which can be depicted as in (b) or (c). (e) is a realization of the basic module in Fig. 1 in terms of hetero- and auto-association; it is also the basic building block of associative memory networks. Each drawing can be interpreted in three different ways: 1. as a (module for a) simple box-and-arrow diagram, 2. a graphical shorthand for vector–matrix multiplication (plus thresholding, see Eq. (2)), and 3. as a shorthand for a neural network architecture (see Palm, 1980, 1982 Appendix 2 or the cover of this journal).

The main reason for the large difference in performance is the sparseness of the memory patterns used with the Hebb rule in the so-called *Willshaw model* (Willshaw et al., 1969). This parameter range has not been well-treated in the early physics literature, probably due to the misleading symmetry assumption (symmetry with respect to sign change) that was imported from spin-glass physics. This prevented the use of binary $\{0, 1\}$ activity values and the corresponding Hebb rule and the discovery of sparseness. It also led to unrealistic neural models, both concerning neural activity (an active neuron carries more information than a passive one) and connectivity (a synaptic weight cannot cross the boundary between excitatory, positive and inhibitory, negative). Only in 1988, it was the contribution of Mischa Tsodyks that brought $\{0, 1\}$ activity modeling, the Hebb rule for synaptic plasticity and sparseness to a broader recognition in the theoretical physics community (Tsodyks & Feigelman, 1988). He showed that for sparse Hebbian associative memories $\alpha = 1/(2\ln 2) \approx 0.72$, corresponding to an information capacity (for auto-associative pattern completion) of about 0.18 (Palm, 1988, 1991; see also Schwenker, Sommer, & Palm, 1996). The corresponding older result for the sparse binary Willshaw model is $\alpha = \ln 2$ resulting in an *information capacity* of $\frac{1}{4}\ln 2 \approx 0.17$ (Palm, 1991; see also Palm & Sommer, 1992; Schwenker et al., 1996).

4. Sparse coding

In technical applications of NAM, efficiency is clearly an important issue. This involves not only storage capacity or storage efficiency, but also fast retrieval of the stored patterns. Since retrieval is simply done by vector–matrix multiplication with entries in $\{0, 1\}$, this is reduced to counting, followed by thresholding, so it is very fast. If the input patterns are sparse it is even faster, since the number of operations is simply proportional to the number of 1-entries in an input pattern. So it is practically important to use sparse binary patterns. This raises the problem of *sparse coding*. At first sight this appears as no big problem. If one wants a sparse representation for a fixed number M of items, for example in terms of binary vectors that each contain k 1-entries and n–k 0-entries, then there are $\binom{n}{k}$ such vectors and one can easily map the M items into such patterns if $M \le \binom{n}{k}$. Another possibility is to use a concatenation of several 1-out-of-n codes to create a k-out-of-kn code (if $M \le n^k$).

However, if one wants to make use of the nice property of NAMs that they respect pattern similarity (in the sense of overlap, inner product, or Hamming distance of binary patterns), then one has to represent semantically similar items by similar binary vectors (Baum, Moody, & Wilczek, 1988; Palm, Schwenker, & Sommer, 1994; Palm, Schwenker, Sommer, & Strey, 1997). This problem of *similarity based sparse coding* has already been treated systematically by Stellmann (1992) by investigating methods that can generate roughly similarity preserving sparse binary code vectors from a given similarity matrix. Also, in many practical applications there is a natural way of generating sparse code vectors: In many pattern recognition or classification tasks with a moderately large number of classes (e.g. in spoken word recognition, face recognition, written letter recognition) it is usual practice to output a 1-out-of-n binary vector representation of the n classes. In more complex applications with very many classes to be distinguished one often uses a more structured approach that describes each class by a large number of binary features, which often are sparse again. If these binary-feature-based representations are not sparse enough, it often makes sense to combine two or more of those features into one (creating a 1-out-of-kn representation from a 1-out-of-k and a 1-out-of-n representation). Of course, this does not make sense (in terms of similarity) for any arbitrary combination of features. These more practical issues of creating sparse codes with natural similarity have recently been rediscovered in practical applications such as visual object recognition (Ganguli & Sompolinsky, 2012; Kavukcuoglu, Lecun, & LeCun, 2010; Kreutz-Delgado et al., 2003; Lee, Battle, Raina, & Ng, 2007; Szlam, Gregor, & LeCun, 2012).

Even on the level of sensor outputs, signals are often sparse, because only changes of the output are reported as temporally separated events. Of course, this strongly depends on the type of sensor. In video signals, for example, the most common compression codes are essentially based on the sparseness of signal differences, both in time and visual (2d) space. The same principle is also working in the human or animal visual system resulting in center-surround antagonistic activation of retinal ganglion cells and sparse activity of edge-detecting cells in the primary visual cortex (Field, 1987; Olshausen, 2003b; Olshausen & Field, 1996a, 1996b; Vinje & Gallant, 2000).

5. The sparseness principle

Also outside the context of associative memory sparseness seems to be a useful principle in machine learning and signal processing (Candes & Romberg, 2007; Coulter, Hillar, Isley, & Sommer, 2010; Donoho & Elad, 2003; Hillar & Sommer, 2011; Hoyer, 2004; Hoyer & Hyvärinen, 2002; Hurley & Rickard, 2009; Kavukcuoglu et al., 2010; Kreutz-Delgado et al., 2003; Szlam et al., 2012), so that one can often expect sparse representations as a result of this kind of processing. In machine learning, in particular in unsupervised or semisupervised learning one tries to create useful compact representations of the data to be learned by autoencoding networks or component analysis (PCA, ICA). In this context one often uses techniques of regularization to obtain robust representations and avoid overfitting. Here again sparseness constraints have turned out to be extremely useful, leading to overcomplete sparse representations. The reasons for this common observation are currently not yet well understood although sparse sensory representations have been investigated since the late 1990s, for example by Bruno Olshausen and others (e.g. Carlson, Rasquinha, Zhang, & Connor, 2011; Coulter et al., 2010; Földiák & Young, 1998; Ganguli & Sompolinsky, 2012;

Hromádka, DeWeese, & Zador, 2008; Hurley & Rickard, 2009; Hyvärinen, 2010; Kreutz-Delgado et al., 2003; Lee, Ekanadham, & Ng, 2008; Olshausen, 2003a, 2003b; Olshausen & Field, 1996a, 1996b, 1997).

Also in sensor fusion sparse representations (often coming from these sources) can be very useful, but here one can perhaps give a hint of the reason. When we want to fuse two sparse binary representations x and y of two types of sensor data coming from the same object, we can learn to associate the positive (nonzero) features in y with those in x. If these features are sparse, then the co-occurrence of a feature in x with a feature in y is much more significant, i.e. much more unlikely to happen by chance, compared to the non-sparse case. Thus we are learning or associating less, but more significant, events, which is likely to result in a better performance.

In the neurosciences today it is commonplace that spiking neural activity is mostly sparse (e.g. Carlson et al., 2011; Földiák & Young, 1998; Franco, Rolls, Aggelopoulos, & Jerez, 2007; Hahnloser, Kozhevnikov, & Fee, 2002; Koulakov & Rinberg, 2011; Vinje & Gallant, 2000; Wolfe, Houweling, & Brecht, 2010). Considering spike-trains of single neurons the argument is simply that the duration of a spike is typically less than a millisecond, which would allow for up to 1000 spikes per second, whereas the observed spike frequencies are much lower; even a very active neuron hardly produces more than 100 spikes per second. A more qualitative observation is that observed spike frequencies tend to go down when we move from sensory or motor systems to more central brain regions like the cerebral cortex. It is not easy to say what the average spike frequency of a cortical neuron may be during a typical day, but it is most likely not more than about 5 spikes per second. This observed sparsity of neural spiking may of course be due to an energy saving principle (Lennie, 2003), but it may also be related to the working of associative synaptic plasticity in the cortex, i.e. to storage efficiency. Concerning the functional role of sparse activity in the cortex, there are several theoretical ideas, in particular for the learning and generation of sparse representations (e.g. Földiák, 1990; Hyvärinen, 2010; Lee et al., 2007; Olshausen, 2003a; Perrinet, 2010; Rehn & Sommer, 2006, 2007; Rozell, Johnson, Baraniuk, & Olshausen, 2008; Zetzsche, 1990).

Also on a cognitive psychological level sparseness appears to be very natural. Most of our mental concepts (or the words for them) occur sparsely. This becomes immediately obvious when we consider negation. We cannot really imagine something like a non-car or a non-table, because this would encompass essentially everything and cannot be conceived. Thus our usual concepts signify rather small, compact and rare constellations of sensations.

6. Technical realization of NAMs

During the 1990s some serious attempts were made to develop technical hardware realizations for massively parallel computing of NAM functionalities (Gao & Hammerstrom, 2003; Heittmann, Malin, Pintaske, & Rückert, 1997; Heittmann & Rückert, 1999, 2002; Zhu & Hammerstrom, 2002). The basic idea is to use a large number of conventional RAM chips storing columns of the storage matrix with parallel counting and thresholding (Palm & Bonhoeffer, 1984; Palm & Palm, 1991).

Another essential idea is to use an address bus to communicate the NAM activity patterns between processors. This is important because generally the inter-process communication is always the bottleneck in massively parallel computing architectures. Here the use of sparse activity patterns makes it possible to save on transmission rate by just transmitting the addresses of the few active neurons in a NAM population. This idea has always been used in our own parallel architectures (e.g. Palm & Palm, 1991; Palm et al., 1997) and later it has been widely adopted in the parallel implementation of spiking or pulse-coded neural networks (e.g. Mahowald, 1994) and even received an acronym, AER, i.e. address-event representation. Now these ideas are used in several larger projects aimed at hardware implementations of large-scale spiking neural networks for technical applications or for brain simulations (De Garis, Shuo, Goertzel, & Ruiting, 2010; Djurfeldt et al., 2008; Fay, Kaufmann, Knoblauch, Markert, & Palm, 2005; Johansson & Lansner, 2007; Markram, 2006; Seiffert, 2004; Zhu & Hammerstrom, 2002).

In such application oriented projects there also occurred the idea of developing efficient software implementations in terms of sparse matrix operations using pointers to the nonzero elements (Bentz, Hagstroem, & Palm, 1989). In these applications there already appeared some indications that the limit of $\ln 2$ retrievable bits per hardware bit, i.e. the efficiency limit of $\ln 2$ can be surpassed. Only recently we were able to show that this is indeed the case (Knoblauch, 2011; Knoblauch, Palm, & Sommer, 2010). By optimizing the quotient of the information storage capacity and the information contained in the storage matrix itself, we found a regime of ultra-sparse memory patterns, where the storage matrix is also a sparse matrix and the quotient, i.e. the efficiency of the memory approaches 1.

7. More detailed modeling of NAMs inspired by neuroscience

Following the early ideas of Donald Hebb (1949), the stepwise formation of an associative memory matrix is understood as a rule for the change of synaptic efficacies in the synaptic connectivity matrix connecting the neurons of the NAM. Thus associative learning is realized by *synaptic plasticity*, which was a subject of intensive investigation in the neurosciences. In neuroscience and recently also in neuromorphic engineering there has been increasing interest in spiking neuron models and the role of spike synchronization (e.g. Jin, Furber, & Woods, 2008; Knoblauch & Palm, 2001; Mehrtash et al., 2003; Palm & Knoblauch, 2005; Plana et al., 2011; Serrano-Gotarredona et al., 2008). In this context the mechanisms for Hebbian synaptic plasticity has been analyzed and modeled on a finer time-scale as a long-term process of synaptic modification that is triggered by the coincidence or close temporal succession of single pre- and postsynaptic spikes (or sometimes two or three of these spikes). These more detailed versions of the general idea of Hebbian synaptic plasticity were observed in neurophysiology (e.g. Bi & Poo, 1998; Caporale & Dan, 2008; Froemke, Poo, & Dan, 2005; Markram, Lübke, Frotscher, & Sakmann, 1997; Song, Miller, & Abbott, 2000) and are called *spike-timing dependent plasticity* (STDP).

Their analysis (e.g. Izhikevich & Desai, 2003; Kempter, Gerstner, & Van Hemmen, 1999; Pfister & Gerstner, 2006) has led to some controversial arguments regarding the consistency between the observed requirement of temporal ordering of pre- and postsynaptic spikes (pre- slightly before post-) and the formation and stabilization of recurrent auto-associations (Clopath, Büsing, Vasilaki, & Gerstner, 2010; Knoblauch & Hauser, 2011). In many STDP models exactly synchronized spikes in two neurons result in a weakening of the synapse connecting them. So synchronized spiking activity in a recurrent assembly would destroy the assembly. However, already a slight temporal jittering of these synchronized spikes in the range of milliseconds can reverse this effect and lead to a strengthening of the synapse. Now the current debate is, what is more likely to happen in a biologically realistic parameter range and in neurobiological reality (Knoblauch, Hauser, Gewaltig, Körner, & Palm, 2012). Currently this interesting discussion can certainly not be interpreted as a conclusive falsification of the assembly ideas, it rather shows new technical possibilities, in particular if one moves from a local

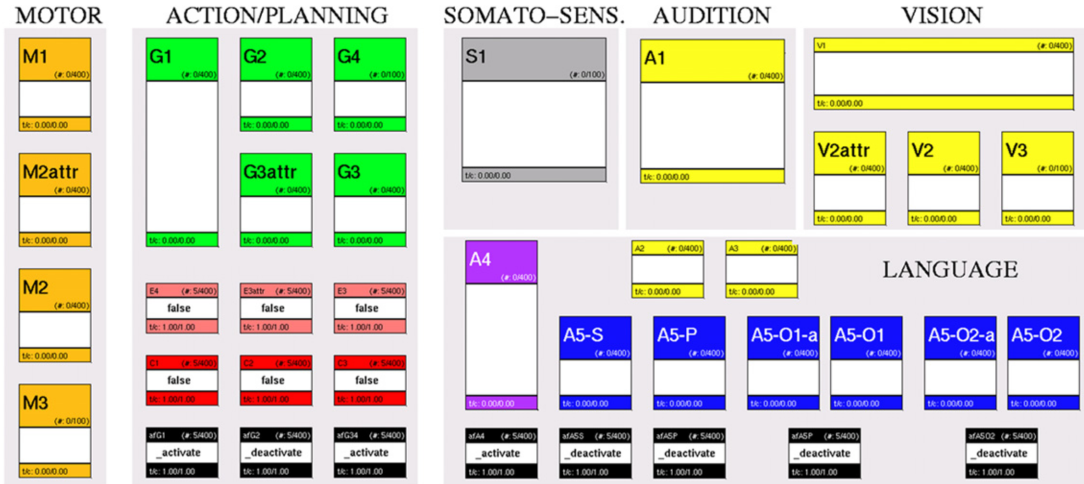


Fig. 3. Cortical architecture involving several interconnected cortical areas or modules corresponding to auditory, grammar, visual, goal, and motor processing. Interconnections are not shown. Each module is realized by a combination of hetero- and auto-association as in Fig. 2(d). Additionally the model comprises evaluation fields and activation fields.
Source: Adapted from Fay et al. (2005).

interpretation of assemblies to a more global systemic one, where the network of the whole cortex is considered as a machine for learning and organization of behavior.

In such a more constructive fashion it is easily possible to create larger systems of several cortical modules based on hetero- and auto-associative connectivity structures that work with spiking neuron models producing biologically plausible single neuron and population dynamics and that can interact in a functionally meaningful fashion to generate computationally demanding behavior which may be called cognitive (Fransén & Lansner, 1998; Lansner, 2009; Lansner & Fransén, 1992; Sandberg, Tegner, & Lansner, 2003). We produced an example of such a system (see Fig. 3) that contains about 30 cortical modules or areas and demonstrates the understanding of simple command sentences by controlling a robot to perform the appropriate actions (Fay et al., 2005; Knoblauch, Markert, & Palm, 2005; Markert, Knoblauch, & Palm, 2007). The basic idea of such an approach is to model the cerebral cortex as a network of associative memory modules. I have developed this idea already in my book 'Neural Assemblies' (Palm, 1982). It was strongly influenced by intense discussions with Valentino Braitenberg and by his analysis and interpretation of the anatomical cortical connectivity (see Braitenberg, 1977, 1978; Braitenberg & Schüz, 1998). Valentino also had the rather cute idea to codify the basic concept of a recurrent associative memory module (see Fig. 2(e) and Palm, 1980) in a widely visible logo, namely the logo of the Springer book series 'Studies of Brain Function' which started in 1977. Incidentally this logo also has some similarity to the logo of 'Neural Networks'.

8. Conclusion

Theoretically it is no problem to show computational Turing universality of binary or spiking neural network systems. The first results on this topic go back to McCulloch and Pitts' paper and to early work in computer science, notably by Kleene. Later this topic was taken up again in a wider context by Wolfgang Maass and others (e.g. Funahashi & Nakamura, 1993). The same is of course also true for associative memory networks (Wennekers & Palm, 2007), which may be used for a higher-level psychologically more plausible implementation of thought processes or human problem solving capabilities. This type of higher level brain modeling based on networks of larger modules, each containing several populations of (thousands of) neurons, may eventually bring us closer to the goal of early neuroscientists like Donald Hebb or Warren McCulloch, namely to bridge the huge gap between lower level neuroscientific analysis of brain activity and higher level synthetic psychological descriptions of human cognition, by creating an additional process description (in terms of Hebbian cell assemblies and cortical modules) at an intermediate level that is amenable to interpretations from both sides. In the recent literature one can find several projects or schools that may be associated with such a program (e.g. Edelman & Tononi, 2000, Hawkins & Blakeslee, 2004 and Hecht-Nielsen, 2007 and of course also Steve Grossberg and John Taylor) and most of them are entertaining ideas that are based on, closely related to, or at least easily translatable into associative memory models.

Acknowledgments

I would like to thank Miriam Schmidt and Martin Schels for helping me with the final preparation of the manuscript.

References

Amari, S. I. (1972). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, *C-21*, 1197–1206.
Amit, D. (1989). *Modeling brain function: the world of attractor neural networks* (1st ed.). Cambridge University Press.
Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1987). Statistical mechanics of neural networks near saturation. *Annals of Physics*, *173*, 30–67.
Anderson, J. A. (1968). A memory storage model utilizing spatial correlation functions. *Biological Cybernetics*, *5*, 113–119.
Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychological Review*, *84*, 413–451.
Baum, E. B., Moody, J., & Wilczek, F. (1988). Internal representations for associative memory. *Biological Cybernetics*, *59*, 217–228.
Bentz, H. J., Hagstroem, M., & Palm, G. (1989). Information storage and effective data retrieval in sparse matrices. *Neural Networks*, *2*, 289–293.
Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons. *Journal of Neuroscience*, *18*, 10464–10472.
Braitenberg, V. (1977). *On the texture of brains*. Berlin, Heidelberg, New York: Springer.
Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim, & G. Palm (Eds.), *Theoretical approaches to complex systems* (pp. 171–188). Berlin, Heidelberg, New York: Springer.
Braitenberg, V., & Schüz, A. (1998). *Neural assemblies. An alternative approach to artificial intelligence*. Berlin, Heidelberg, New York: Springer.
Candes, E., & Romberg, J. (2007). Sparsity and incoherence in compressive sampling. *Inverse Problems*, *23*, 969–985.
Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, *31*, 25–46.

Carlson, E. T., Rasquinha, R. J., Zhang, K., & Connor, C. E. (2011). A sparse object coding scheme in area V4. *Current Biology*, *21*, 288–293.
Clopath, C., Büsing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience*, *13*, 344–352.
Coulter, W. K., Hillar, C. J., Isley, G., & Sommer, F. T. (2010). Adaptive compressed sensing—a new class of self-organizing coding models for neuroscience. In *ICASSP* (pp. 5494–5497). IEEE.
De Garis, H., Shuo, C., Goertzel, B., & Ruiting, L. (2010). A world survey of artificial brain projects, part I: large-scale brain simulations. *Neurocomputing*, *74*, 3–29.
Djurfeldt, M., Lundqvist, M., Johansson, C., Rehn, M., Ekeberg, O., & Lansner, A. (2008). Brain-scale simulation of the neocortex on the IBM blue gene/L supercomputer. *IBM Journal of Research and Development*, *52*, 31–41.
Domany, E., van Hemmen, J., & Schulten, K. (1991). *Models of neural networks*. Berlin, Heidelberg, New York: Springer.
Donoho, D., & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, *100*, 2197.
Dunin-Barkowski, W. L., & Larionova, N. P. (1985). Computer simulation of a cerebellar cortex compartment. II. An information learning and its recall in the Marr's memory unit. *Biological Cybernetics*, *51*, 407–415.
Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: how matter becomes imagination*. New York: Basic Books.
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 S. Wermter, G. Palm, & M. Elshaw (Eds.), *Lecture notes in computer science*: *vol. 3575*. *Biomimetic neural learning for intelligent robots* (pp. 118–143). Berlin, Heidelberg: Springer.
Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and Image Science*, *4*, 2379–2394.
Földiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, *64*, 165–170.
Földiák, P., & Young, M. P. (1998). Sparse coding in the primate cortex. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 895–898). Cambridge, MA, USA: MIT Press.
Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Jerez, J. M. (2007). Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biological Cybernetics*, *96*, 547–560.
Fransén, E., & Lansner, A. (1998). A model of cortical associative memory based on a horizontal network of connected columns. *Network: Computation in Neural Systems*, *9*, 235–264.
Froemke, R. C., Poo, M. M., & Dan, Y. (2005). Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature*, *434*, 221–225.
Funahashi, K., & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, *6*, 801–806.
Gabor, D. (1969). Associative holographic memories. *IBM Journal of Research and Development*, *13*, 156–159.
Ganguli, S., & Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, *35*, 485–508.
Gao, C., & Hammerstrom, D. (2003). Platform performance comparison of PALM network on Pentium 4 and FPGA. In *Proceedings of the international joint conference on neural networks* (pp. 995–1000).
Gardner, E. (1987). Maximum storage capacity in neural networks. *Europhysics Letters*, *4*, 481–485.
Gardner, E. (1988). The space of interactions in neural network models. *Journal of Physics A: Mathematical and General*, *21*, 257–270.
Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *419*, 65–70.
Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: Henry Holt and Company.
Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.
Hecht-Nielsen, R. (2007). *Confabulation theory: the mechanism of thought*. Berlin, Heidelberg, New York: Springer.
Heittmann, A., Malin, J., Pintaske, C., & Rückert, U. (1997). Digital VLSI implementation of a neural associative memory, In Klar H., König A., R.U. (Eds.), *Proceedings of the 6th international conference on microelectronics for neural network, evolutionary and fuzzy systems* (pp. 280–285).
Heittmann, A., & Rückert, U. (1999). Mixed mode VLSI implementation of a neural associative memory. In *Proceedings of the seventh international conference on microelectronics for neural, fuzzy and bio-inspired systems, 1999,* MicroNeuro'99 (pp. 299–306).
Heittmann, A., & Rückert, U. (2002). Mixed mode VLSI implementation of a neural associative memory. *Analog Integrated Circuits and Signal Processing*, *30*, 159–172.
Hillar, C., & Sommer, F. (2011). Ramsey theory reveals the conditions when sparse coding on subsampled data is unique.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2558.
Horner, H., Bormann, D., Frick, M., Kinzelbach, H., & Schmidt, A. (1989). Transients and basins of attraction in neural network models. *Zeitschrift für Physik B Condensed Matter*, *76*, 381–398.
Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, *5*, 1457–1469.
Hoyer, P. O., & Hyvärinen, A. (2002). A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, *42*, 1593–1605.
Hromádka, T., DeWeese, M., & Zador, A. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biology*, *6*, e16.
Hurley, N., & Rickard, S. (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory*, *55*, 4723–4741.
Hyvärinen, A. (2010). Statistical models of natural images and cortical visual representation. *Topics in Cognitive Science*, *2*, 251–264.
Izhikevich, E. M., & Desai, N. S. (2003). Relating STDP to BCM. *Neural Computation*, *15*, 1511–1523.
Jin, X., Furber, S., & Woods, J. (2008). Efficient modelling of spiking neural networks on a scalable chip multiprocessor. In *IEEE international joint conference on neural networks* (pp. 2812–2819).
Johansson, C., & Lansner, A. (2007). Towards cortex sized artificial neural systems. *Neural Networks*, *20*, 48–61.
Kavukcuoglu, K., Lecun, Y., & LeCun, Y. (2010). Fast inference in sparse coding algorithms with applications to object recognition. arXiv:1010.3467v1.
Kempter, R., Gerstner, W., & Van Hemmen, J. (1999). Hebbian learning and spiking neurons. *Physical Review E*, *59*, 4498–4514.
Knoblauch, A. (2011). Neural associative memory with optimal Bayesian learning. *Neural Computation*, *23*, 1393–1451.
Knoblauch, A., & Hauser, F. (2011). STDP, temporal coding, and anatomical connectivity patterns. *Technical report*. Honda Research Institute Europe. HRIEU Report.
Knoblauch, A., Hauser, F., Gewaltig, M. O., Körner, E., & Palm, G. (2012). Does spike-timing-dependent synaptic plasticity couple or decouple neurons firing in synchrony? *Frontiers in Computational Neuroscience*, *6*.
Knoblauch, A., Markert, H., & Palm, G. (2005). An associative cortical model of language understanding and action planning. In J. R. Alvarez, & J. Mira (Eds.), *LNCS*, *Artificial intelligence and knowledge engineering applications: a bioinspired approach*, IWINAC 2005 Part 2. (pp. 405–414). Springer.
Knoblauch, A., & Palm, G. (2001). Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Networks*, *14*, 763–780.
Knoblauch, A., Palm, G., & Sommer, F. T. (2010). Memory capacities for synaptic and structural plasticity. *Neural Computation*, *22*, 289–341.
Kohonen, T. (1977). *Associative memory*. Berlin: Springer.
Koulakov, A. A., & Rinberg, D. (2011). Sparse incomplete representations: a potential role for olfactory granule cells. *Neuron*, *72*, 124–136.
Krauth, W., & Mézard, M. (1989). Storage capacity of memory networks with binary couplings. *Journal de Physique France*, *50*, 3057–3066.
Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T. W., & Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation. *Neural Computation*, *15*, 349–396.
Lansner, A. (2009). Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. *Trends in Neurosciences*, *32*, 178–186.
Lansner, A., & Fransén, E. (1992). Modelling Hebbian cell assemblies comprised of cortical neurons. *Network: Computation in Neural Systems*, *3*, 105–119.
Lee, H., Battle, A., Raina, R., & Ng, A. (2007). Efficient sparse coding algorithms. *Advances in Neural Information Processing Systems (NIPS 2006)*, 801–808.
Lee, H., Ekanadham, C., & Ng, A. (2008). Sparse deep belief net model for visual area V2. In *NIPS* (pp. 873–880).
Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*, 493–497.
Little, W. A. (1974). The existence of persistent states in the brain. *Mathematical Biosciences*, *19*, 101–120.
Longuet-Higgins, H. C., Willshaw, D. J., & Buneman, O. P. (1970). Theories of associative recall. *Quarterly Reviews of Biophysics*, *3*, 223–244.
Mahowald, M. (1994). The silicon optic nerve. In *The Kluwer international series in engineering and computer science*: *vol. 265*. *An analog VLSI system for stereoscopic vision* (pp. 66–117). New York: Springer.
Markert, H., Knoblauch, A., & Palm, G. (2007). Modelling of syntactical processing in the cortex. *BioSystems*, *89*, 300–315.
Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, *7*, 153–160.
Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, *275*, 213–215.
Marr, D. (1969). A theory of cerebellar cortex. *The Journal of Physiology*, *202*, 437–470.
Mehrtash, N., Jung, D., Hellmich, H. H., Schoenauer, T., Lu, V. T., & Klar, H. (2003). Synaptic plasticity in spiking neural networks (SP(2)INN): a system approach. *IEEE Transactions on Neural Networks*, *14*, 980–992.
Nadal, J. P., & Toulouse, G. (1990). Information storage in sparsely coded memory nets. *Network: Computation in Neural Systems*, *1*, 61–74.
Olshausen, B. (2003a). Learning sparse, overcomplete representations of time-varying natural images. In *IEEE international conference on image processing. vol. 1* (pp. i-41–i-44).
Olshausen, B. (2003b). Principles of image representation in visual cortex. In L. M. Chalupa, & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1603–1615). Bradford Books (Chapter 108).
Olshausen, B. A., & Field, D. J. (1996a). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
Olshausen, B. A., & Field, D. J. (1996b). Natural image statistics and efficient coding. In *Network: computation in neural systems*: *vol. 7* (pp. 333–339). UK: Informa UK Ltd.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, *37*, 3311–3325.
Opper, M., Kleinz, J., Kohler, W., & Kinzel, W. (1989). Basins of attraction near the critical storage capacity for neural networks with constant stabilities. *Journal of Physics A: Mathematical and General*, *22*, L407–L411.
Palm, G. (1980). On associative memories. *Biological Cybernetics*, *36*, 19–31.
Palm, G. (1982). *Neural assemblies. An alternative approach to artificial intelligence*. Berlin, Heidelberg, New York: Springer.
Palm, G. (1985). Associative networks and their information storage capacity. *Cognitive Systems*, *1*, 107–118.
Palm, G. (1987). Computing with neural networks. *Science*, *235*, 1227–1228.
Palm, G. (1988). On the asymptotic information storage capacity of neural networks. In R. Eckmiller, & C. von der Malsburg (Eds.), *Neural Computers* (pp. 271–280). New York: Springer-Verlag.
Palm, G. (1990). Local learning rules and sparse coding in neural networks. In R. Eckmiller (Ed.), *Advanced neural computers* (pp. 145–150). North-Holland, Amsterdam: Elsevier.
Palm, G. (1991). Memory capacities of local rules for synaptic modification. a comparative review. *Concepts in Neuroscience*, *2*, 97–128.
Palm, G. (1992). On the information storage capacity of local learning. *Neural Computation*, *4*, 703–711.
Palm, G., & Bonhoeffer, T. (1984). Parallel processing for associative and neuronal networks. *Biological Cybernetics*, *51*, 201–204.
Palm, G., & Knoblauch, A. (2005). Scene segmentation through synchronization. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 618–623). Elsevier.
Palm, G., & Palm, M. (1991). Parallel associative networks: the Pan-system and the Bacchus-chip. In U. Ramacher, U. Rückert, J.N. (Eds.), *Proceedings of the second international conference on microelectronics for neural networks* (pp. 411–416).
Palm, G., Schwenker, F., & Sommer, F. (1994). Associative memory networks and sparse similarity preserving codes. In V. Cherkassky, J. Friedman, & H. Wechsler (Eds.), *NATO-ASI series F*, *From statistics to neural networks: theory and pattern recognition applications* (pp. 283–302). Berlin, Heidelberg: Springer.
Palm, G., Schwenker, F., Sommer, F., & Strey, A. (1997). Neural associative memory. In A. Krikelis, & C. Weems (Eds.), *Associative processing and processors* (pp. 307–326). Los Alamitos, CA, USA: IEEE Computer Society.
Palm, G., & Sommer, F. (1992). Information capacity in recurrent McCulloch–Pitts networks with sparsely coded memory states. *Network: Computation in Neural Systems*, *3*, 177–186.
Palm, G., & Sommer, F. T. (1995). Associative data storage and retrieval in neural networks. In E. Domany, J. van Hemmen, & K. Schulten (Eds.), *Models of neural networks III: association, generalization and representation* (pp. 79–118). Springer.
Perrinet, L. U. (2010). Role of homeostasis in learning sparse representations. *Neural Computation*, *22*, 1812–1836.
Pfister, J. P., & Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience*, *26*, 9673–9682.
Plana, L. A., Clark, D., Davidson, S., Furber, S., Garside, J., Painkras, E., et al. (2011). Spinnaker: design and implementation of a GALS multicore system-on-chip. *Journal on Emerging Technologies in Computing Systems*, *7*, 17:1–17:18.
Rehn, M., & Sommer, F. (2006). Storing and restoring visual input with collaborative rank coding and associative memory. *Neurocomputing*, *69*, 1219–1223.
Rehn, M., & Sommer, F. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience*, *22*, 135–146.
Rozell, C. J., Johnson, D. H., Baraniuk, R. G., & Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, *20*, 2526–2563.
Sandberg, A., Tegner, J., & Lansner, A. (2003). A working memory model based on fast Hebbian learning. *Network: Computation in Neural Systems*, *14*, 789–802.
Schwenker, F., Sommer, F. T., & Palm, G. (1996). Iterative retrieval of sparsely coded associative memory patterns. *Neural Networks*, *9*, 445–455.
Seiffert, U. (2004). Artificial neural networks on massively parallel computer hardware. *Neurocomputing*, *57*, 135–150.
Serrano-Gotarredona, R., Serrano-Gotarredona, T., Acosta-Jimenez, A., Serrano-Gotarredona, C., Perez-Carrasco, J., Linares-Barranco, B., Linares Barranco, A., Jimenez-Moreno, G., Civit-Ballcels, A., et al. (2008). On real-time AER 2-D convolutions hardware for neuromorphic spike-based cortical processing. *IEEE Transactions on Neural Networks*, *19*, 1196–1219.
Sommer, F. T., & Palm, G. (1999). Improved bidirectional retrieval of sparse patterns stored by Hebbian learning. *Neural Networks*, *12*, 281–297.
Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3*, 919–926.
Steinbuch, K. (1961). Die Lernmatrix. *Biological Cybernetics*, *1*, 36–45.
Stellmann, S. (1992). Ähnlichkeitserhaltende Codierung. *Ph.D. thesis*. Universität Ulm.
Szlam, A., Gregor, K., & LeCun, Y. Fast approximations to structured sparse coding and applications to object classification. arXiv:1202.6384v1, 2012.
Tsodyks, M. V., & Feigelman, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters*, *6*, 101–105.
Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*, 1273–1276.
Wennekers, T., & Palm, G. (2007). Modelling generic cognitive functions with operational Hebbian cell assemblies. In M. Weiss (Ed.), *Neural network research horizons* (pp. 225–294). Nova Science Publishers.
Wigström, H. (1975). Associative recall and formation of stable modes of activity in neural network models. *Journal of Neuroscience Research*, *1*, 287–313.
Willshaw, D. J. (1971). Models of distributed associative memory. *Ph.D. thesis*. University of Edinburgh.
Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, *222*, 960–962.
Wolfe, J., Houweling, A. R., & Brecht, M. (2010). Sparse and powerful cortical spikes. *Current Opinion in Neurobiology*, *20*, 306–312.
Zetzsche, C. (1990). Sparse coding: the link between low level vision and associative memory. In G. Eckmiller, G. Hartmann, & G. Hauske (Eds.), *Parallel processing in neural systems and computers* (pp. 273–276). Amsterdam: Elsevier.
Zhu, S., & Hammerstrom, D. (2002). Simulation of associative neural networks. In *Proceedings of the 9th international conference on neural information processing 2002,* ICONIP 02 (pp. 1639–1643).