

Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations

Anders Lansner

Department of Computational Biology, School of Computer Science and Communication, Stockholm University and Royal Institute of Technology, 114 21 Stockholm, Sweden
Stockholm Brain Institute, Karolinska Institutet, 171 77 Stockholm, Sweden

The second half of the past century saw the emergence of a theory of cortical associative memory function originating in Donald Hebb's hypotheses on activity-dependent synaptic plasticity and cell-assembly formation and dynamics. This conceptual framework has today developed into a theory of attractor memory that brings together many experimental observations from different sources and levels of investigation into computational models displaying information-processing capabilities such as efficient associative memory and holistic perception. Here, we outline a development that might eventually lead to a neurobiologically grounded theory of cortical associative memory.

Introduction

Memory is an important and integral part of our cognitive functions. It comes in many different forms such as semantic, episodic and procedural. Learning and the underlying neural plasticity (i.e. activity-dependent changes in the neural tissue leading to a changed behaviour) are processes closely linked to memory. Experimental investigation of the mechanisms involved is extensive and produces data from many different sources, from the molecular and biochemical to the cognitive and behavioural levels. As in many other fields today, the use of computers for modelling and simulation is an increasingly important tool to organize and synthesize a coherent understanding from such massive amounts of experimental data [1–3].

Here, we provide an overview of the state of the art in cortical memory modelling grounded in neurobiology (i.e. neural-network models composed of biophysically reasonably detailed models of neurons and synapses). However, the development and study of more abstract connectionist models* is also important for progress in our understanding because they emphasize information-processing aspects, focus on key elements and mechanisms, and sometimes enable mathematical analysis. Therefore, the foundations in and relations to abstract models are highlighted.

One of the earliest and most viable conceptual theories of cortical associative memory is Hebb's theory of cell assemblies [4]. It has been the main source of inspiration

for today's attractor-memory models of cortex, mathematically instantiated in the form of, for instance, the Hopfield model [5]. Although alternative paradigms for understanding cortical function exist [6,7], the attractor-memory paradigm is today the best developed and most extensively investigated, and it is our main concern here.

Basics of cell assemblies and attractor-memory networks

The fundamental idea of formation and dynamics of cell assemblies proposed by Hebb relies on the existence of Hebbian synapses as a substrate for formation of such assemblies (Box 1). Excitatory synapses strengthened by coincident pre- and postsynaptic activity will connect neurons co-activated by the same stimulus. Cell assemblies so formed will subsequently serve as mental representations of their respective generating object. If a stimulus similar to one of those memorized later appears, perhaps in a fragmentary or otherwise somewhat distorted form, pattern completion will activate the entire cell assembly, which then reverberates in the cortex for a fraction of a second. This neural dynamics would explain, for example, holistic visual perception in the form of figure-background segmentation and perceptual completion[†]. Lateral inhibition via inhibitory interneurons would mediate perceptual rivalry[‡] in cases of ambiguous stimuli.

Basics of attractor memory

Hebb's theories, in themselves entirely conceptual and qualitative in nature, triggered some early biological neural-network-simulation studies [8]. The main question addressed was whether a group of neurons connected by enhanced excitatory synapses would display reverberating activity outlasting the stimulus as predicted by Hebb. These first studies used abstract model neurons and did not support the prediction.

[†] Figure-background segmentation refers to our tendency to see a figure in front of a background. Perceptual completion occurs when we perceive an object as a complete entity even when occluded or from otherwise partial cues. An example is the classical Dalmatian dog picture photographed by R. C. James (see http://en.wikipedia.org/wiki/Gestalt_psychology).

[‡] Perceptual rivalry refers to the phenomenon that perceptions of two pictured objects, though equally probable, are mutually exclusive. Perception occasionally alternates from one interpretation to the other. An example of an ambiguous picture is E. G. Boring's old lady/young woman (<http://mathworld.wolfram.com/YoungGirl-OldWomanIllusion.html>).

Corresponding author: Lansner, A. (ala@kth.se)

* A simplified network model of neural information processing built with abstract neural units and connections..

Box 1. Hebbian synapses and cell assemblies

Donald Hebb (1904–1985) made important early contributions to our understanding of memory [88,89]. He is well known for his hypothesis that synaptic long-term plasticity is a key to memory formation. In his book *The Organization of Behaviour* [4], Hebb stated 'When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased' (pp. 62). A second proposal was that such processes are instrumental in the formation of cell assemblies, that is, groups of repeatedly co-activated cells becoming wired together by enhanced excitatory synapses. Such 'mental objects' would support distributed memory representation and reverberatory (persistent) activity and would influence the dynamics of subsequent perception and memory recall. Hebb hypothesized that this could produce perceptual 'holistic' Gestalt-phenomena such as figure-background segmentation, pattern completion and rivalry and make possible memory recall from fragmentary information. An important extension of Hebb's theory was later made by Milner who brought lateral inhibition into the picture [90]. The third concept proposed by Hebb was the phase sequence. It was hypothesized that some synapses might operate with a wide temporal coincidence window thus connecting stimuli occurring in temporal proximity, for example, cell assemblies active in succession. This would enable storage and recall of the sequences in what could loosely be related to association chains.

Long-term potentiation (LTP) was first reported in 1973 [91], and its Hebbian nature was later verified [92]. Today, Hebbian synaptic plasticity is established as a key element in learning and memory processes [93]. Experimental data concerning these processes are now massive and many different forms are known, several of them non-Hebbian [94]. However, there is still much uncertainty, for instance, with respect to the quantitative relations that guide such plasticity and the spectrum of variation of the underlying learning rules, for example with regard to dependence on timing of spiking in pre- and postsynaptic neurons.

This work was soon followed by investigations of the properties and relevance of recurrent neural network models for understanding associative memory, which were more or less in line with Hebb's proposal [9–16]. In 1982, building on previous work by Little [17], John Hopfield published his proposal about the analogies between neural dynamics and collective phenomena studied in materials physics [5] (Box 2). This triggered a surge of interest in

Box 2. Attractor network models of memory

Hebb's theory of cell assemblies and related work inspired the neural network models mathematically formulated in the early 1980s by John Hopfield [5,95]. It was based on the analogy between a neural network with binary neurons and magnetic spin systems or 'spin glasses'. The theoretical and analytical methods developed by theoretical physicists to describe the collective stochastic dynamics of the latter were applied to the dynamics of a recurrently connected neural network.

A Hopfield network comprises a set of units connected by all-to-all connections (Figure 1). When used as a model for associative memory, the network connections are trained by some Hebbian learning rules to form a (symmetric) connection matrix.

Commonly used equations for updating the membrane activation (difference between membrane potential and resting potential) and the output rate (normalized to $[-1, 1]$) of the units are:

$$C_j \frac{du_j}{dt} = \sum_i w_{ij} v_i - \frac{u_j}{R_j} + i_j$$

$$v_j = \tanh\left(\frac{u_j}{T}\right)$$

where u and v denote membrane activation and output rate, respectively, i and j index the pre- and postsynaptic units, C and R represent

membrane capacitance and resistance, w_{ij} is a synaptic weight, i is input current and T controls the steepness of the sigmoidal unit input–output relation (Figure 2). Given a set of P patterns to store in a memory with N units a standard Hebbian learning rule prescribes:

$$w_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{N} \sum_{\mu=1}^P \xi_i^{\mu} \xi_j^{\mu} & \text{otherwise} \end{cases}$$

where N is the number of units and μ indexes patterns, and ξ_k represents the output (± 1) of unit k in training pattern μ .

The connection matrix of the Hebbian learning rule is symmetric ($w_{ij} = w_{ji}$), which has been proven to give a simple and well-behaved network dynamics. This dynamics embodies some of the key characteristics suggested by Hebb and Milner, such as persistent activity, pattern completion and rivalry.

The most important performance criterion for associative memories is their capacity, quantified either as the number of patterns possible to store and retrieve reliably in a network of a certain size or the number of bits of stored information per connection [86,87]. Theory indicates that an attractor network modified to match, for example, the hippocampus, given the number of units, degree of connectivity and activity level estimated from biology, is an efficient associative memory (i.e. the amount of information possible to store and retrieve scales with the number of storage elements [synapses]).

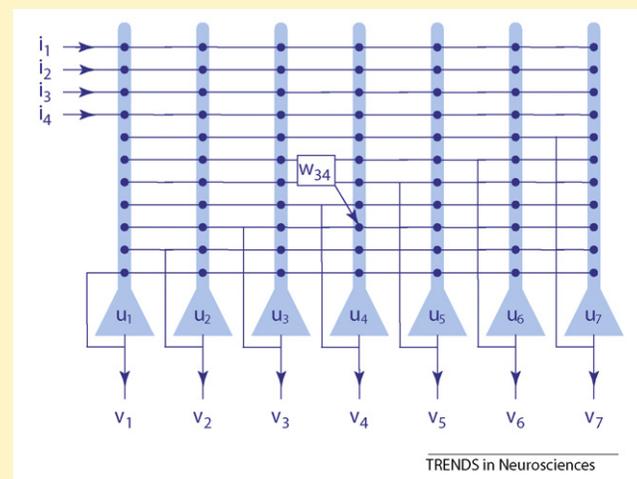


Figure 1. Schematic drawing of a recurrent Hopfield-type neural network. Neural units are in light blue, connections are in darker blue. For notation refer to the text in the box.

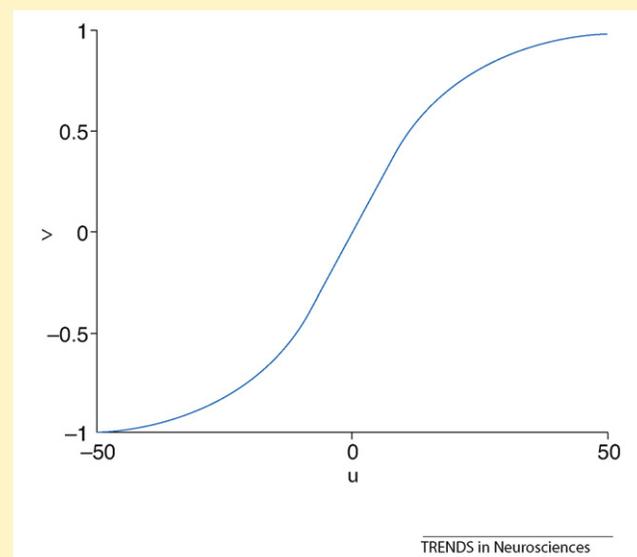


Figure 2. Sigmoidal input–output relation of neural units in a Hopfield-type neural network. Input, membrane activation (u); output, normalized rate (v).

Review

collective dynamical phenomena in the brain and inspired more theoretical physicists to analyze the properties of recurrent neural network models (for examples, see Refs. [18,19]).

The Little–Hopfield model had the basic features of a recurrently connected neuronal network with excitatory and inhibitory interactions between its units. In addition to being capable of performing the elementary perceptual and memory functions proposed by Hebb, including working memory in the form of reverberation, perceptual figure-background segmentation, pattern completion and rivalry, attractor networks were also found to be efficient content-addressable memories (i.e. memories that allow recall from some fragment of its content) from an information theoretical point of view.

The olfactory system and the hippocampus were proposed to be prototypical biological attractor networks [20,21]. When applied to the neocortex, the basic attractor-network model has mainly been seen as representing the predominantly horizontally connected layers 2/3 and 5 of higher order cortical areas with their memory and attention [22] and categorical perception capabilities (i.e. the perception of different stimuli as being qualitatively, or categorically, different) [23,24].

The standard Hopfield model is still abstract and far from neurobiology. It has been criticized on several points: (i) it requires biologically unrealistic all-to-all symmetric connectivity; (ii) it violates Dale's law, which states that each neuron releases only one neurotransmitter at all of its synapses [25]; (iii) activity levels are too high; (iv) the number of patterns possible to store is too small; (v) activity gets stuck in attractor states; and (vi) recall with real low-rate spiking neurons would probably be much slower than what is seen experimentally. We address these arguments and show how the standard model has, in fact, been modified and realized with more biorealistic elements, such that the end product is a neurobiologically plausible network that maintains the functionally appealing properties of the original model. It is illustrated how attractor-memory models have come closer to biological reality in terms of their constituent model neurons and synapses in addition to network architecture.

Early-spiking attractor-memory models

The very first neural-network simulations employing more biologically realistic spiking-model neurons with conductance-based synapses and spike-frequency adaptation were done in the 1970s [26]. They simulated a randomly connected recurrent network of excitatory and inhibitory neurons with firing characteristics reminiscent of spinal motoneurons. As in the first studies, Hebb's assumption of sustained reverberations was not supported. Instead a strong tendency for rhythmic firing and spike synchronization at a frequency of ~30–50 Hz was seen. Spike synchrony and gamma oscillations of a similar kind were experimentally observed ten years later in the cat visual cortex [27].

Despite these early negative results regarding reverberating cell assemblies, later studies have demonstrated that a recurrently connected network of excitatory and inhibitory neurons could indeed produce sustained reverbera-

tions [28–30]. This was typically achieved by fitting the excitatory-model neurons to cortical pyramidal cells rather than spinal motoneurons and by including *N*-methyl-D-aspartate (NMDA)-receptor-gated channels with their relatively slow and voltage-dependent excitatory postsynaptic potential (EPSP). This demonstrates clearly how properties at the level of neurons and synapses might qualitatively change macroscopic network dynamics.

The build-up of excitation in this kind of recurrent networks is rapid – almost explosive. However, the reverberatory state is transient. Its termination is promoted by slow processes such as spike-frequency adaptation and the activation of delayed hyperpolarizing currents. The calcium entering through NMDA channels activate potassium channels causing a gradually increasing afterhyperpolarization (AHP). After some hundred milliseconds, this AHP overtakes the recurrent excitation thus terminating abruptly the ongoing reverberation. This is like the 'fatigue process' Hebb suggested was acting to terminate an active cell assembly. Today, such 'fatigue' might be more broadly related to processes such as AHP of different kinds and time-courses and to synaptic depression [31,32].

In addition to sustained activity, a recurrent network of this kind has been shown to operate as an attractor associative memory that can retrieve a set of memorized patterns stored in a connection matrix formed by a Hebbian learning rule [28]. The negative component of the connection matrix was 'inverted' via an inhibitory interneuron (Figure 1a) to comply with Dale's law. Such a network was able to perform pattern completion and rivalry on timescales compatible with that observed in psychophysical experiments [28,33].

In summary, despite their simplicity and limited complexity and size, these early-spiking attractor-memory models displayed some of the fundamental neuronal activity patterns in addition to functional perceptual and memory properties experimentally observed in the real cortex.

Does an abstract computational unit correspond to a real neuron?

To fully appreciate the contribution of connectionist models it is crucial to understand the mapping between their computational units and real neurons and cortical circuitry. It is often taken for granted that a unit in, for example, a Hopfield network, corresponds to a single real neuron. But this is, in fact, not the only or even the most plausible assumption (Figure 2). It is more likely that such a unit represents a small population of neurons with similar response properties, like, for example, an orientation column in the cat primary visual cortex [34] or some functional subpopulation within a barrel of the rat somatosensory cortex [35]. The cortical minicolumn, which is most prominent in primates and humans, is yet another candidate as a computational unit [36–40]. Consequently, a connection from an abstract unit to another would correspond to several synaptic connections between such local neuron populations. A functional minicolumnar structure could generate the patchy connectivity or 'daisies' reported for many cortical areas and species (notably, daisies and

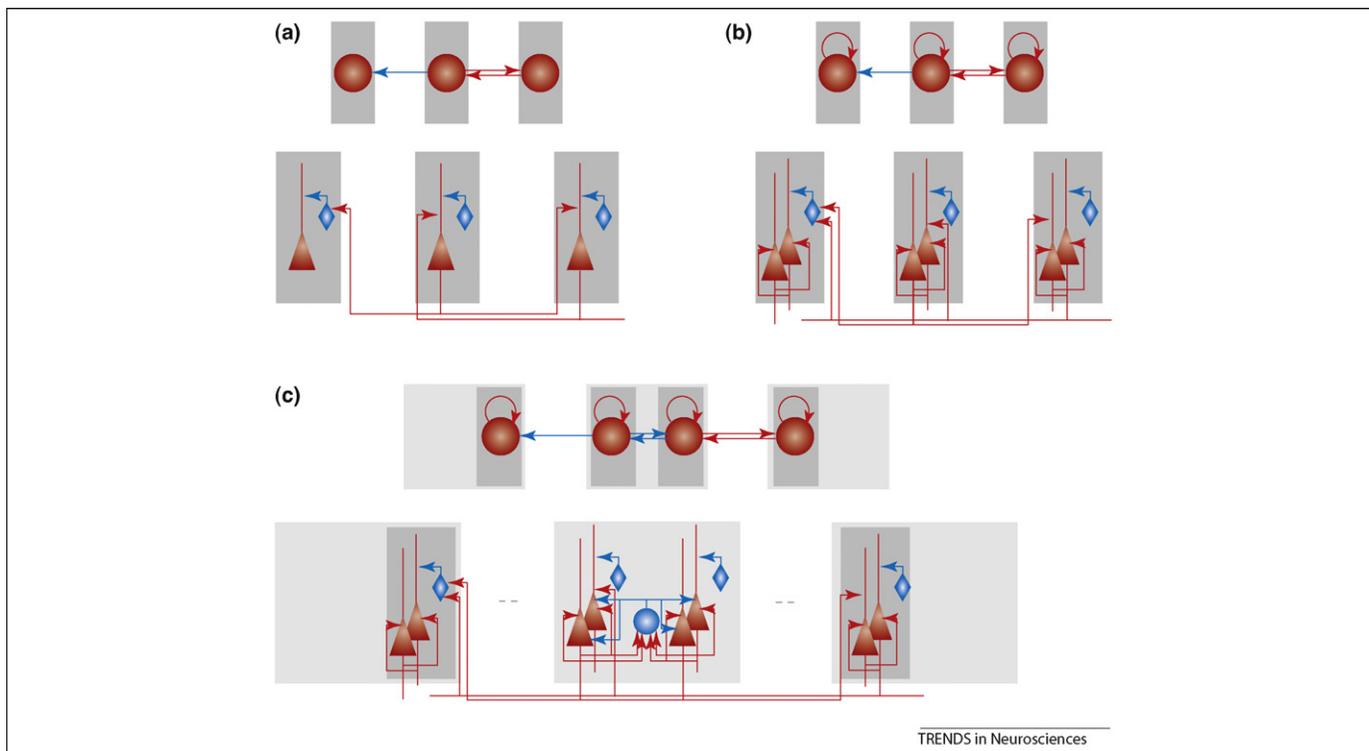


Figure 1. Example of the gradual development of a cortical layer 2/3 model over time. Red indicates excitatory cells and synaptic connections; blue indicates inhibitory ones. Gray indicates minicolumnar (darker) and hypercolumnar (lighter) modules. Pyramidal cells (red), basket cells (blue circle) and RSNP cells (blue diamond) are shown. In each panel, the network motif is shown at the top and the model architecture below. The middle module is in the same pattern as the right one (recurrent excitation) and in a different pattern than the left one. Lateral inhibition is mediated by basket and RSNP cells. Network with **(a)** one excitatory and one inhibitory cell in each module [28], **(b)** 12 excitatory and three inhibitory cells in each module [44], and **(c)** 30 pyramidal cells and two RSNP cells in each minicolumn (module), and nine minicolumns and eight basket cells per hypercolumn [59]. The largest simulated version of the latter network had almost 7000 hypercolumns with 100 minicolumns each [62].

minicolumns are not prominent in rodent neocortex) [41]. An alternative view is to consider a larger cortical module, a hypercolumn or macrocolumn, as constituting a local recurrent network, akin to a standard attractor network

and operating as a local associative memory module [42,43].

By replacing the single pyramidal cell with a local population of interconnected pyramidal cells in a model

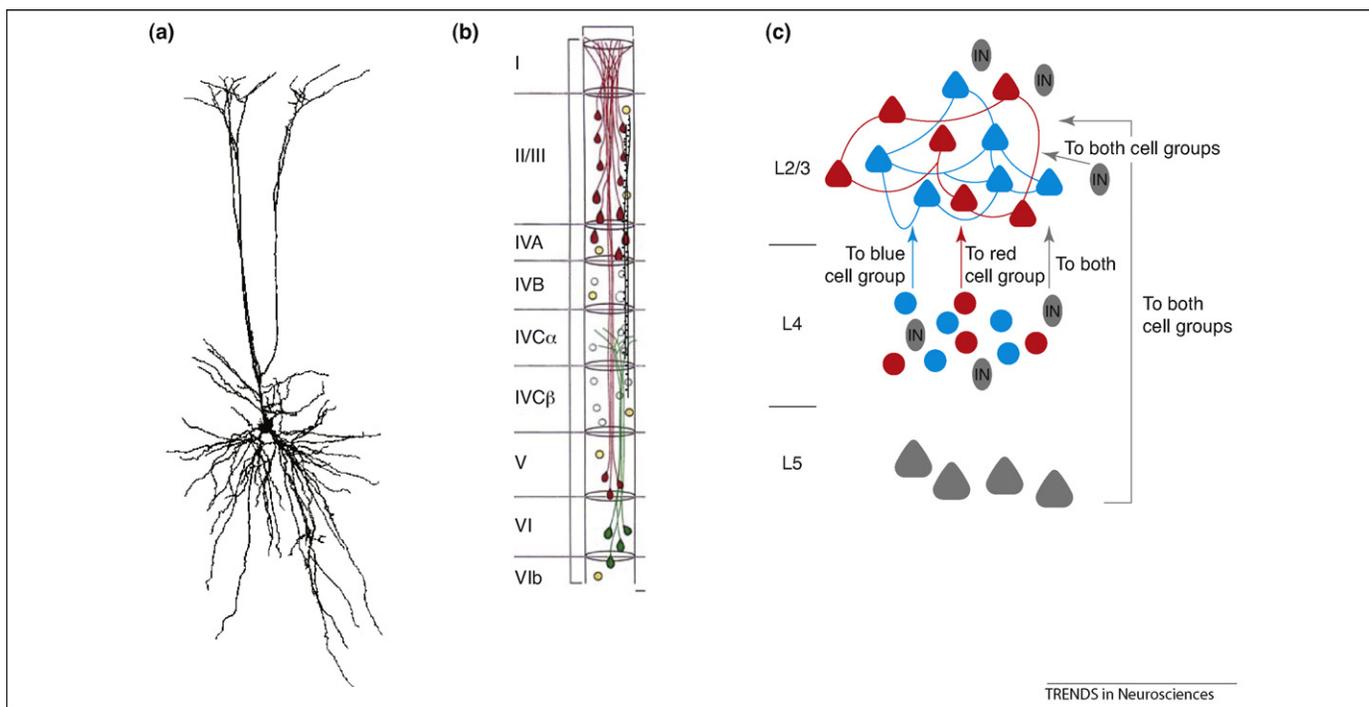


Figure 2. What is the computational unit in the cortex, corresponding to a unit in a recurrent connectionist network model? Here are three suggestions: **(a)** the single neuron, here a layer 5 pyramidal cell [96]; **(b)** the anatomical mimicolumn, here in macaque cortex [37]; the different cortical layers are indicated with roman numerals, sub-layers with letters; **(c)** a local distributed sub-network of neurons in the rat visual cortex [35]. Cortical layers 2/3 (L2/3), 4 (L4) and 5 (L5) are indicated. Red and blue show two different sub-networks, gray ovals marked IN indicate inhibitory interneurons. Reproduced, with permission, from Refs [35,37,96].

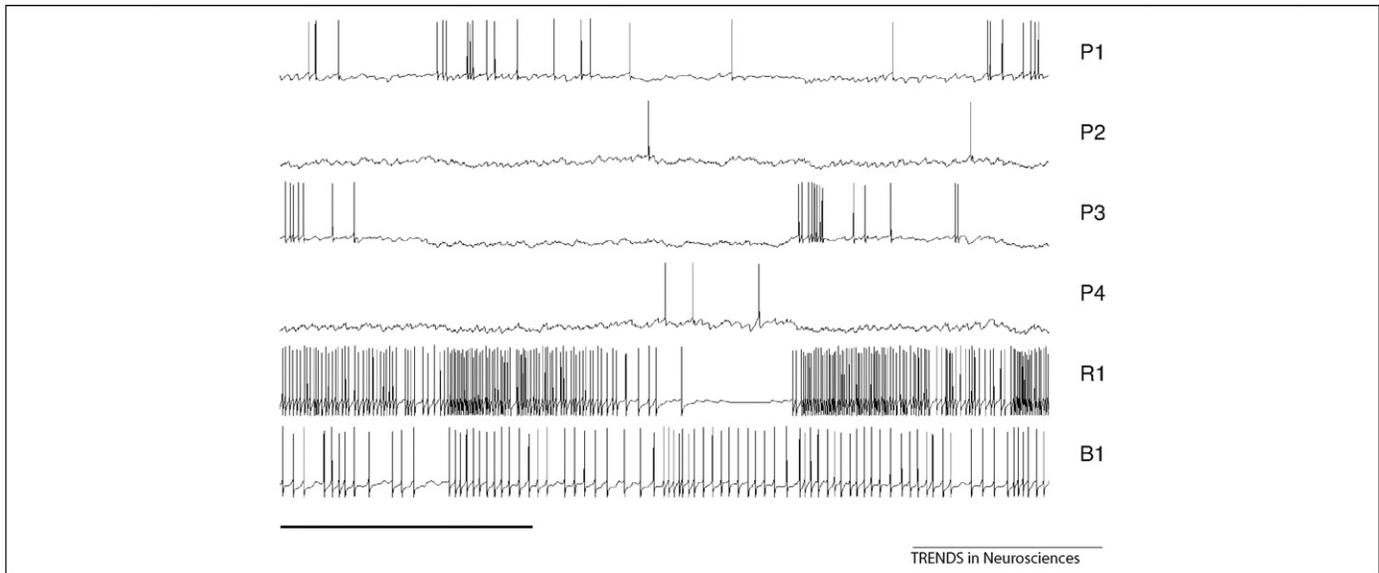


Figure 3. Samples of simulated intracellular potentials during three simulated seconds from a set of cells in the same hypercolumn of an attractor network model from Ref. [59]. This illustrates the diversity of responses among nearby cells depending on cell type and whether the cell is part of the active cell assembly or not. P1–P4 are pyramidal cells, R1 is an RSNP cell and B1 is a basket cell. During the middle second P4 and R1 are part of an active cell assembly and during the last second P3 is part of one. The RSNP cells typically show inverted selective activity relative to pyramidal cells as seen in P4–R1. Horizontal scale bar is 1 s.

of a piece of cortex, the correspondence with modular cortical connectivity sometimes observed could be improved [44] (Figure 1b). The original biologically implausible dense and symmetric connectivity then becomes sparse and asymmetric at the neuron-to-neuron level, although functionally symmetric at the minicolumnar level. Simulations showed that such a network could operate properly as an attractor associative memory. Perhaps surprisingly, the symmetry of the functional connectivity between minicolumns was sufficient to produce a well-behaved dynamics and proper network function.

Lateral inhibition, activity normalization and hypercolumns

As more data on *in vivo* cortical firing patterns became available, it was evident that the early-spiking attractor-memory models typically displayed a spiking frequency that was too high and also unnaturally regular spiking patterns. It was found that strong negative-feedback inhibition from basket cells was crucial for models to reproduce the low-rate irregular firing of cells seen in the cortex *in vivo* [45–47]. With well-balanced excitation and inhibition, several models have thereafter been able to adequately reproduce *in vivo* spiking activity [48–50].

The basket cells are horizontally projecting over distances in the order of a millimetre, which is approximately the same size as that of thalamocortical and corticocortical axonal arbors [51,52] and of the basal dendritic tree of layer 2/3 pyramidal cells [53]. Cortical modules such as hypercolumns [34], macrocolumns or segregates [54] are also of this size. These data combined with the computational benefits of a modular structure have prompted theoreticians to propose models comprising ‘winner-take-all’ or activity-normalization modules [38,55–58].

Such a cortical modular structure has been incorporated in some recent models of the cortical layer 2/3 network [59]. Here, in addition to the basket cells, a further type of

inhibitory interneuron, that is, a regular spiking non-pyramidal (RSNP) cell [60], was included (Figure 1c). In contrast to the horizontally projecting and soma-targeting basket cells, RSNP cells are dendritic targeting and vertically projecting within a narrow vertical column. They seem to be ideally suited to ‘invert’ long-range excitation to di-synaptic inhibition as described earlier and as seen experimentally [61]. This more elaborate network model has demonstrated the same functionality as previous network models with regard to cell-assembly dynamics (Figures 3,4) and has reproduced the low-rate irregular spiking of the pyramidal cells seen *in vivo* in addition to population oscillations in beta and gamma frequency bands [62]. Such a model could also reproduce semi-quantitatively the complex cognitive phenomenon of attentional blink [59,63].

In summary, biophysically detailed modelling of the recurrently and horizontally connected cortical layers 2/3 have demonstrated that this sub-network might, indeed, operate as suggested by Hebb and formalized in the Little–Hopfield type of abstract attractor models. It is most likely that layer 5 with its prominent input from layers 2/3 is also involved in the long-range corticocortical interactions within this sub-network. As we briefly discuss next, layer 4 that feeds into layer 2/3 might be primarily responsible for transforming incoming information into a sparse and distributed internal representation suitable for the following holistic processing.

Layer 4, the temporal dimension and volatile memories

At the current stage of development, existing attractor-network models are limited in several important respects. Three of the main open questions concern how and where the required sparse and distributed internal representations emerge, how timescales beyond a few hundred milliseconds can be represented and processed and, finally, if and how short-term forms of memory can be realized in

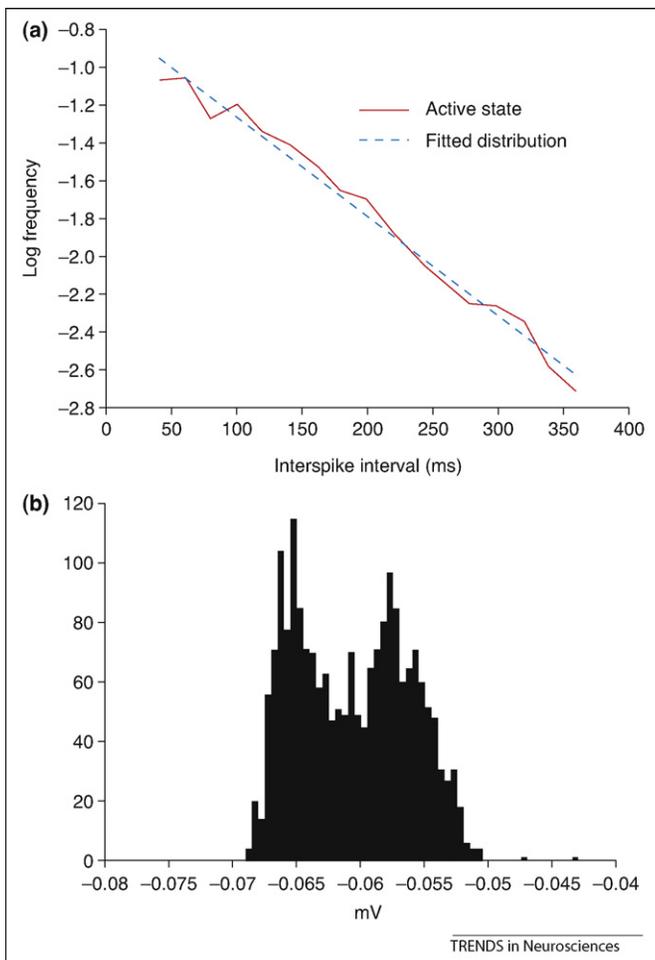


Figure 4. Statistics of spiking and intracellular potential of simulated pyramidal cells [62]. (a) Exponentially distributed interspike intervals of pyramidal cells in the active state of a simulated network. This shows that a spiking attractor network can, perhaps unexpectedly, produce the irregular and low-rate firing seen in experiments. (b) Histogram showing imodal distribution of intracellular potential of one of all the simulated pyramidal cells in the same simulation. Similar results have been obtained when measuring in the cat visual cortex (for example, Ref. [97]). Reproduced, with permission, from Ref. [62].

addition to ‘standard’ long-term memory. In this section, these issues are briefly considered and some possibilities for extensions of the current models are suggested.

Attractor networks are first and foremost seen as models of processing in the higher order sensory and association cortex, thus emphasizing the horizontal connectivity in layers 2/3 and 5 and their role in holistic-pattern processing. The input from different sensory receptor arrays is typically not decorrelated and sparse enough to be suitable as an input to such an attractor network. Therefore, it is often assumed that the necessary transformations of internal representation occur along the feed-forward processing stream passing from the thalamus to layers 4 and 2/3 in primary sensory cortices and continuing to higher order cortical areas. The formation of specific response properties of neurons is a key element in modelling experience-dependent development of sparse and distributed cortical representations. The synaptic plasticity rules commonly applied in the mathematical modelling of such ‘feature extraction’ typically embodies some form of competitive learning [64–69]. To make further progress, we need to integrate elements of holistic

processing and feature extraction into the same theoretical framework and network models.

Thus far, a little-studied aspect using computational memory models is temporal association, relating to Hebb’s proposal of ‘phase sequences’ and association chains. Some early attractor-network models investigated ways this could occur via plastic asymmetric synaptic connections [70] and later models have considered mechanisms of serial order and recall in recurrent networks [71,72]. Yet, considering the fundamental importance of order and timing in memory recall, there has been surprisingly little focus on modelling and theoretical analysis of spatiotemporal dynamics over longer time-spans. One reason might be that the underlying cellular and synaptic processes are likely to be slow and of low amplitude and, thus, hard to observe and measure experimentally.

Hebb’s early theories and related quantitative models have almost exclusively focused on long-term memory storage with synaptic plasticity as the underlying mechanism. However, several early Hebbian synaptic-plasticity processes are expressed almost instantly after an inducing stimulus [73–76]. An intriguing hypothesis is that short-term and working memory are crucially dependent on fast and volatile forms of Hebbian plasticity induced by one-shot learning and expressed within seconds [77,78]. Such synaptic learning dynamics relates closely to previous theoretical attractor-network models of palimpsest memory [58,79–81] and unify in an elegant way models of memory acting at multiple timescales within the paradigm of attractor memory.

Towards supercomputer simulation of full-scale cortex models

The computational power of our computers has been and is still growing according to Moore’s law (Figure 5). The supercomputers available 10–15 years from now will probably be capable of simulating in real-time a full-scale human brain model at a level of detail illustrated earlier. There are already examples of supercomputer simulations of very large neuronal-network models. For instance, a cortex model with 22-million neurons connected by 11-billion synapses (approaching the size of the mouse cortex [82]) was run on an IBM BlueGene/L cluster supercomputer with 8192 processors working in parallel [62]. The capacity to simulate large networks is important because the tiny network models possible to simulate on today’s standard computers owing to their small size often require compensations in terms of denser and stronger connectivity. This tends to distort network dynamics [29], which degrades the quality of the model.

Indeed, large-scale models enable the study of more elaborate network architectures that better match the complexity of the real brain. The brain-scale network-of-network architectures already under investigation using more abstract connectionist types of models will be candidates for simulation using biophysically detailed spiking neuron models [83,84]. From simulations of the mass activity of properly localized and oriented model neurons over seconds and minutes it will be possible to calculate, for instance, the corresponding electroencephalogram (EEG), magnetoencephalogram (MEG), voltage-sensitive dye im-

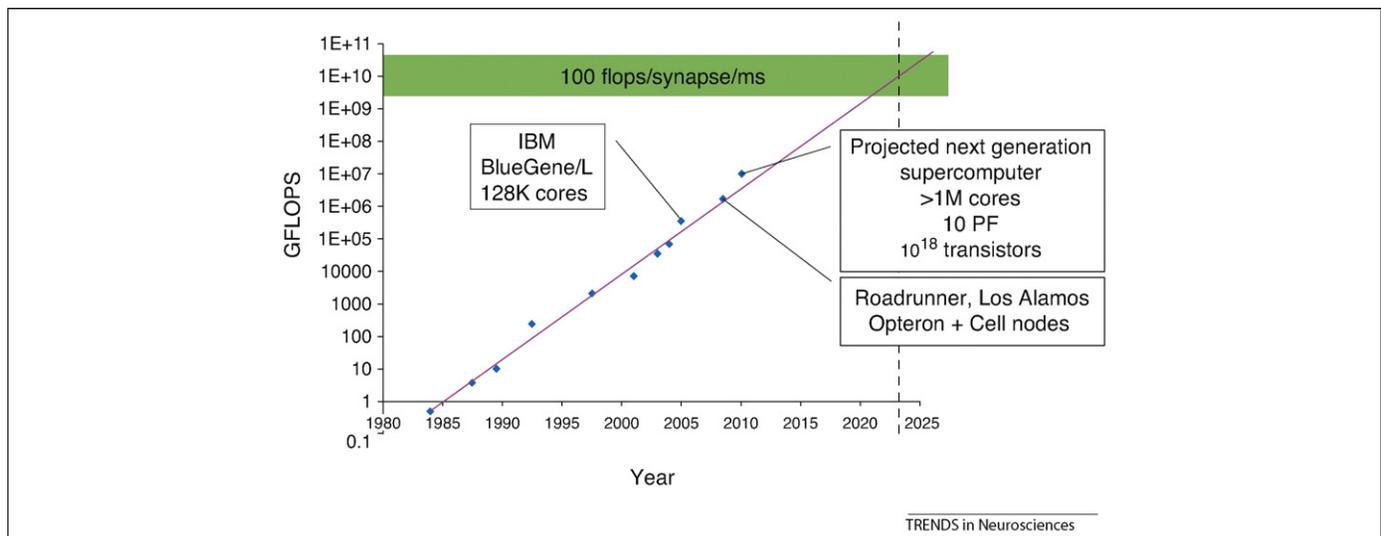


Figure 5. Trend of the development of supercomputer power: ‘Moore’s law’. The vertical axis shows data on supercomputing power (in GFLOPS) over time. The purple line represents Moore’s law. The prognosis is that this relation will hold at least until 2015, but thereafter uncertainty is higher. The blue marks represent several historical and one projected top supercomputer installations, the three latest of which are named. The green area indicates the region where the largest supercomputers will enable real-time simulation of human-brain sized models of the type described in this article. The vertical dashed line shows that this might happen around 2023. Abbreviations: GFLOPS, gigaflops (billion floating-point operations per second); PF, petaflops (1 million GFLOPS).

age (VSD) and blood-oxygenation-dependent contrast image (BOLD; measured in functional magnetic brain imaging) [85]. This will enable addressing in new ways the huge amounts of experimental data obtained with these techniques and will improve our understanding of how these measures relate to cellular- and synaptic-level processes. Furthermore, large-scale network models have the potential to display non-trivial information-processing functions and demonstrate how these might be implemented in the spiking neuronal networks of the brain. This will connect to experimental results at the behavioural, psychophysical and psychological level and expand further the amount of experimental data addressed by our models. Although large-scale brain simulation is in its infancy, it might provide the technology that eventually enables a quantitative understanding of all this complexity. That alone makes it worthwhile to continue efforts to develop the necessary tools in terms of computer hardware and software.

Concluding remarks

Like experimental memory research, modelling of memory function is continuing to develop rapidly. Today’s computational models have improved substantially compared with those developed and studied at the dawn of modern computing. Theoretical analysis and computer simulation of network models have already identified putative mechanisms behind associative memory and fundamental holistic processing capabilities of the cortex. Such biophysically detailed models are able to mimic this basic functionality and at the same time reproduce quantitatively, for example, membrane potential trajectories, spiking statistics and local field potentials measured *in vivo*. The attractor-memory paradigm conceptualized more than half a century ago is today investigated using supercomputer simulations of millions of neurons and billions of synapses. New models feature an increasing diversity in

terms of neuron types, physiology and architecture, and will thus come closer to real cortical microcircuitry, including its layered and modular organization. Of course, computational models can never do the job on their own. Experiments and animal models will continue to be necessary components in brain research including modelling for the foreseeable future. They will provide new data on all levels to constrain and improve our computational models. The computer models will, in turn, help to squeeze maximal information out of the data and promote more hypotheses-driven experimentation.

The use of supercomputers opens new important avenues for quantitative computational brain modelling. It already enables us to simulate and investigate biophysically detailed mouse-cortex-sized network models, and full-scale real-time simulation of human-cortex-sized models will most likely be feasible in a decade or two. This will enable investigation of multiple dynamically interacting cortical areas and sub-cortical structures connected via synaptic projections with realistic densities, magnitude and plasticity. Such models will be more adequate than the small-sized monolithic network models often used in the past. They will enable a detailed understanding of the macroscopic phenomenology of perception, cognition and behaviour in terms of dynamical processes occurring at the cellular, synaptic and molecular levels. Long before the grand challenge to understand brain function has been successfully completed, progress towards multi-scale mechanistic and dynamic simulation models of different aspects of brain function is likely to facilitate considerably the development of drugs and therapies for treatment of major psychiatric disorders and brain diseases.

Acknowledgements

This work was partly supported by grants from the Swedish Science Council (Vetenskapsrådet, VR-621-2004-3807; www.vr.se), the European Union (FACETS project, FP6-2004-IST-FETPI-015879; <http://>

facets.kip.uni-heidelberg.de) and the Swedish Foundation for Strategic Research (through the Stockholm Brain Institute; www.stockholmbrain.se). Comments on the manuscript from Erik Fransén, Örjan Ekeberg and David Silverstein are gratefully acknowledged.

References

- 1 MacGregor, R.J. (1987). In *Neural and Brain Modeling (Neuroscience)* (Thompson, R.F., ed.), Academic Press
- 2 Amit, D.J. (1998) Simulation in neurobiology: theory or experiment? *Trends Neurosci.* 21, 231–237
- 3 Koch, C. and Segev, I., eds (1998) *Methods in Neuronal Modeling: From Ions to Networks* (2nd edn), MIT Press
- 4 Hebb, D.O. (1949) *The Organization of Behavior*. John Wiley
- 5 Hopfield, J.J. (1982) Neural Networks and physical systems with emergent collective computational properties. *Proc. Natl. Acad. Sci. U. S. A.* 81, 3088–3092
- 6 Abeles, M. (1991) *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press
- 7 Maass, W. et al. (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560
- 8 Rochester, N. et al. (1956) Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Trans. Inf. Theory* IT-2, 80–93
- 9 Steinbuch, K. (1961) Die Lernmatrix. *Kybernetik (Biol. Cybern.)* 1, 36–45
- 10 Taylor, W.K. (1964) Cortico-thalamic organization and memory. *Proc. R. Soc. Lond. B. Biol. Sci.* 159, 466–478
- 11 Willshaw, D.J. et al. (1969) Non-holographic associative memory. *Nature* 222, 960–962
- 12 Amari, S-I. (1972) Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* C-21, 1197–1206
- 13 Kohonen, T. (1972) Correlation matrix memories. *IEEE Trans. Comput.* C-21, 353–359
- 14 Wigström, H. (1975) Associative recall and formation of stable modes of activity in neural network models. *J. Neurosci. Res.* 1, 287–313
- 15 Grossberg, S. and Levine, D. (1975) Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *J. Theor. Biol.* 53, 341–380
- 16 Anderson, J.A. et al. (1977) Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychol. Rev.* 84, 414–451
- 17 Little, W.A. (1974) The existence of persistent states in the brain. *Math. Biosci.* 19, 101–120
- 18 Amit, D. (1989) *Modeling Brain Function: The World of Attractor Neural Networks*. (1st edn), Cambridge University Press
- 19 Hertz, J. et al. (1991) *Introduction to the Theory of Neural Computation*. (1st edn), Addison-Wesley Publishing
- 20 Marr, D. (1971) Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81
- 21 Haberly, L.B. and Bower, J.M. (1989) Olfactory cortex: model circuit for the study of associative memory? *Trends Neurosci.* 12, 258–264
- 22 Rolls, E.T. and Deco, G. (2006) Attention in natural scenes: neurophysiological and computational bases. *Neural Netw.* 19, 1383–1394
- 23 Rotshtein, P. et al. (2004) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* 8, 107–113
- 24 Keri, S. (2003) The cognitive neuroscience of category learning. *Brain Res. Brain Res. Rev.* 43, 85–109
- 25 Eccles, J.C. (1964) *The Physiology of synapses*. (1st edn), Academic Press
- 26 MacGregor, R.J. and McMullen, T. (1978) Computer simulations of diffusely connected neuronal populations. *Biol. Cybern.* 28, 121–127
- 27 Gray, C.M. and Singer, W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 86, 1698–1702
- 28 Lansner, A. and Fransén, E. (1992) Modeling Hebbian cell assemblies comprised of cortical neurons. *Network: Computation in Neural Systems* 3, 105–119
- 29 Wilson, M. and Bower, J.M. (1992) Cortical oscillations and temporal interactions in a computer simulation of piriform cortex. *J. Neurophysiol.* 67, 981–995
- 30 Traub, R.D. et al. (1992) Computer simulation of carbachol-driven rhythmic oscillations in the CA3 region of the *in vitro* rat hippocampus. *J. Neurophysiol.* 451, 653–672
- 31 Sah, P. (1996) Ca²⁺-activated K⁺-currents in neurons: types, physiological roles and modulation. *Trends Neurosci* 19, 150–154
- 32 Markram, H. et al. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215
- 33 Thorpe, S. et al. (1996) Speed of processing in the human visual system. *Nature* 381, 520–522
- 34 Hubel, D.H. and Wiesel, T.N. (1977) The functional architecture of the macaque visual cortex. The Ferrier lecture. *Proc. R. Soc. Lond. B. Biol. Sci.* 198, 1–59
- 35 Yoshimura, Y. et al. (2005) Excitatory cortical neurons form fine-scale functional networks. *Nature* 433, 868–873
- 36 Mountcastle, V.B. (1978) An organizing principle for cerebral function: the unit module and distributed function. In *The Mindful Brain* (Edelman, G. and Mountcastle, V.B., eds), pp. 7–50, MIT Press
- 37 Peters, A. and Sethares, C. (1996) Myelinated axons and the pyramidal cell modules in monkey primary visual cortex. *J. Comp. Neurol.* 365, 232–255
- 38 Favorov, O.V. and Kelly, D.G. (1994) Minicolumnar organization within somatosensory cortical segregates: I. development of afferent connections. *Cereb. Cortex* 4, 408–427
- 39 Buxhoeveden, D.P. and Casanova, M.F. (2002) The minicolumn hypothesis in neuroscience. *Brain* 125, 935–951
- 40 DeFelipe, J. (2006) Double-bouquet cells in the monkey and human cerebral cortex with special reference to areas 17 and 18. *Prog. Brain Res.* 154, 15–31
- 41 Binzegger, T. et al. (2007) Stereotypical bouton clustering of individual neurons in cat primary visual cortex. *J. Neurosci.* 27, 12242–12254
- 42 Sommer, F.T. and Wennekers, T. (2001) Associative memory in networks of spiking neurons. *Neural Netw.* 14, 825–834
- 43 Amit, D.J. and Brunel, N. (1997) Model of global spontaneous activity and local structured delay activity during delay periods in the cerebral cortex. *Cereb. Cortex* 7, 237–252
- 44 Fransén, E. and 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
- 45 Abeles, M. et al. (1990) Firing patterns of single units in the prefrontal cortex and neural network models. *Network: Computation in Neural Systems* 1, 13–25
- 46 Mainen, Z.F. and Sejnowski, T.J. (1995) Reliability of spike timing in neocortical neurons. *Science* 268, 1503–1505
- 47 Rudolph, M. et al. (2007) Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. *J. Neurosci.* 27, 5280–5290
- 48 Amit, D.J. and Treves, A. (1989) Associative memory neural network with low temporal spiking rates. *Proc. Natl. Acad. Sci. U. S. A.* 86, 7871–7875
- 49 Hansel, D. and Sompolinsky, H. (1996) Chaos and synchrony in a model of a hypercolumn in visual cortex. *J. Comput. Neurosci.* 3, 7–34
- 50 van Vreeswijk, C. and Sompolinsky, H. (1998) Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10, 1321–1371
- 51 Jones, E.G. (1981) Functional subdivision and synaptic organization of the mammalian thalamus. *Int. Rev. Physiol.* 25, 173–245
- 52 Kakei, S. et al. (1996) Projection pattern of single corticocortical fibers from the parietal cortex to the motor cortex. *Neuroreport* 7, 2369–2372
- 53 Elston, G.N. and Rosa, M.G. (2000) Pyramidal cells, patches, and cortical columns: a comparative study of infragranular neurons in TEO, TE, and the superior temporal polysensory area of the macaque monkey. *J. Neurosci.* 20, 1–5
- 54 Favorov, O.V. and Diamond, M. (1990) Demonstration of discrete place-defined columns, segregates, in cat S1. *J. Comp. Neurol.* 298, 97–112
- 55 Kanter, I. (1988) Potts-glass models of neural networks. *Phys. Rev. A* 37, 2739–2742
- 56 Joubin, F. et al. (1996) A columnar model of somatosensory reorganizational plasticity based on Hebbian and non-Hebbian learning rules. *Biol. Cybern.* 74, 276–286
- 57 Lansner, A. and Holst, A. (1996) A higher order Bayesian neural network with spiking units. *Int. J. Neural Syst.* 7, 115–128

Review

- 58 Sandberg, A. *et al.* (2002) Bayesian attractor networks with incremental learning. *Network: Computation in Neural Systems* 13, 179–194
- 59 Lundqvist, M. *et al.* (2006) Attractor dynamics in a modular network model of the neocortex. *Network: Computation in Neural Systems* 17, 253–276
- 60 Kawaguchi, Y. (1995) Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex. *J. Neurosci.* 15, 2638–2655
- 61 Hirsch, J.A. and Gilbert, C.D. (1991) Synaptic physiology of horizontal connections in the cats visual cortex. *J. Neurosci.* 11, 1800–1809
- 62 Djurfeldt, M. *et al.* (2008) Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer. *IBM J. Res. Develop.* 52, 31–41
- 63 Marois, R. and Ivanoff, J. (2005) Capacity limits of information processing in the brain. *Trends Cogn. Sci.* 9, 296–305
- 64 von der Malsburg, C. (1973) Self organization of orientation selective cells in striate cortex. *Kybernetik* 14, 85–100
- 65 Oja, E. (1982) Simplified neuron model as a principal component analyzer. *J. Math. Biol.* 15, 267–273
- 66 Bienenstock, E.L. *et al.* (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48
- 67 Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69
- 68 Bell, A.J. and Sejnowski, T.J. (1995) An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159
- 69 Olshausen, B.A. and Field, D.J. (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* 37, 3311–3325
- 70 Kanter, I. and Sompolinsky, H. (1986) Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* 57, 2861–2864
- 71 Levy, W.B. and Wu, X. (1996) The relationship of local context codes to sequence length memory capacity. *Network: Computation in Neural Systems* 7, 371–384
- 72 Rehn, M. and Lansner, A. (2004) Sequence memory with dynamical synapses. *Neurocomputing* 58–60, 271–278
- 73 Gustafsson, B. *et al.* (1989) Onset characteristics of long-term potentiation in the guinea-pig hippocampal CA1 Region *in vitro*. *Eur. J. Neurosci.* 1, 384–393
- 74 Bagal, A.A. *et al.* (2005) Long-term potentiation of exogenous glutamate responses at single dendritic spines. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14435–14439
- 75 Vertes, R.P. (2005) Hippocampal theta rhythm: a tag for short-term memory. *Hippocampus* 15, 923–935
- 76 Irvine, E.E. *et al.* (2006) α CaMKII autophosphorylation: a fast track to memory. *Trends Neurosci.* 29, 459–465
- 77 Sandberg, A. *et al.* (2003) A working memory model based on fast Hebbian learning. *Network: Computation in Neural Systems* 14, 789–802
- 78 Romani, S. *et al.* (2008) Optimizing one-shot learning with binary synapses. *Neural Comput.* 20, 1928–1950
- 79 Nadal, J.P. *et al.* (1986) Networks of formal neurons and palimpsests. *Europhys. Lett.* 1, 535–542
- 80 Amit, D.J. and Fusi, S. (1994) Learning in neural networks with material synapses. *Neural Comput.* 6, 957–982
- 81 Henson, R. and Willshaw, D.J. (1995) Short-term associative memory. *Proceedings of the INNS World Congress on Neural Networks* 1, pp. 438–441, INNS Press
- 82 Braitenberg, V. and Schüz, A. (1998) *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer
- 83 Atallah, H.E. *et al.* (2004) Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. *Neurobiol. Learn. Mem.* 82, 253–267
- 84 Grossberg, S. (2007) Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition. *Prog. Brain Res.* 165, 79–104
- 85 Izhikevich, E.M. and Edelman, G.M. (2008) Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3593–3598
- 86 Schwenker, F. *et al.* (1996) Iterative retrieval of sparsely coded associative memory patterns. *Neural Netw.* 9, 445–455
- 87 Papp, G. *et al.* (2007) The CA3 network as a memory store for spatial representations. *Learn. Mem.* 14, 732–744
- 88 Cooper, S.J. (2005) Donald O Hebb's synapse and learning rule: a history and commentary. *Neurosci. Biobehav. Rev.* 28, 851–874
- 89 Fuster, J.M. (1995) *Memory in the Cerebral Cortex*. The MIT Press
- 90 Milner, P.M. (1957) The cell assembly: Mark II. *Psychol. Rev.* 64, 242–252
- 91 Bliss, T.V.P. and Lømo, T. (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* 232, 331–356
- 92 Levy, W.B. and Steward, O. (1979) Synapses as associative memory elements in the hippocampal formation. *Brain Res.* 175, 233–245
- 93 Barco, A. *et al.* (2006) Common molecular mechanisms in explicit and implicit memory. *J. Neurochem.* 97, 1520–1533
- 94 Nelson, S.B. *et al.* (2002) Rate and timing in cortical synaptic plasticity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1851–1857
- 95 Hopfield, J.J. (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U. S. A.* 81, 3088–3092
- 96 Contreras, D. *et al.* (1997) Intracellular and computational characterization of the intracortical inhibitory control of synchronized thalamic inputs *in vivo*. *J. Neurophysiol.* 78, 335–350
- 97 Anderson, J. *et al.* (2000) Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. *Nat. Neurosci.* 3, 617–621