

# **A cell assembly based model for the cortical microcircuitry.**

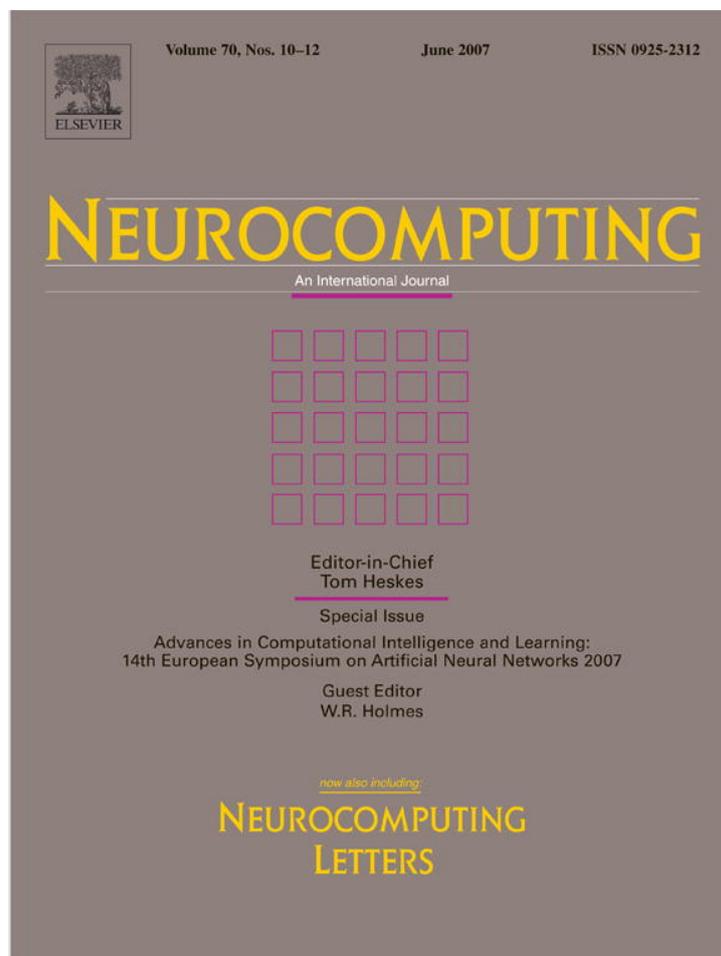
**Andreas Knoblauch, Rüdiger Kupper, Marc-Oliver Gewaltig, Ursula Körner, Edgar Körner**

**2007**

**Preprint:**

This is an accepted article published in Neurocomputing. The final authenticated version is available online at: [https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

Provided for non-commercial research and educational use only.  
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

# A cell assembly based model for the cortical microcircuitry

Andreas Knoblauch<sup>\*</sup>, Rüdiger Kupper, Marc-Oliver Gewaltig, Ursula Körner, Edgar Körner

*Honda Research Institute Europe GmbH, Carl-Legien-Str.30, D-63073 Offenbach/Main, Germany*

Available online 9 November 2006

## Abstract

This work presents first simulation results substantiating a previously proposed conceptual model of computation in neocortical architecture [E. Körner, M.-O. Gewaltig, U. Körner, A. Richter, T. Rodemann, A model of computation in neocortical architecture, *Neural Networks* 12 (1999) 989–1005]. This model gives a detailed functional interpretation of the six-layered columnar cortical architecture and related subcortical structures. It hypothesizes three interacting processing systems at each stage of the cortical hierarchy: The A-system (middle cortical layers) accomplishes fast bottom-up processing where the first spike wave traveling up the cortical hierarchy can activate a coarse initial hypothesis at each level. In the B-system (superficial layers) the initial hypothesis is refined by slower iterative processes involving feedback. Finally, the C-system (deep layers) represents the local hypothesis of a macrocolumn which is fed back to the B-system of a lower level inducing expectations and predictions for the present and future input signals. These ideas are illustrated by an example implementation of the microcircuitry in a single cortical macrocolumn based on cell assemblies and associative memories. In a second step we have integrated our model at the level of V4 into a large scale implementation of the visual system involving several primary and higher visual cortical areas as well as parts of the hippocampal formation, and subcortical structures involved in generating eye saccades. With this model we can demonstrate object classification and the learning of new object representations.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Cortical control architecture; Columns; Cortical learning; Associative memory; Visual system; Predictive coding

## 1. Introduction

Understanding how previously experienced objects and scenes are recognized and categorized by the brain is still one of the hardest problems in neuroscience. Because sensory stimuli regularly include complex and ambiguous scenes it is necessary to use top-down prediction in order to reduce the tremendous number of possible interpretations on each level of representation. Thus, any solution has to detail (i) how prior knowledge is integrated (top-down) with the actual (bottom-up) stream of sensory data in a meaningful way, and (ii) when and how new representations are created and stably integrated into the previously learned knowledge hierarchy.

This simulation study substantiates a previously proposed model of computation in neocortical architecture [16,17] (Fig. 1; see also [10,23,9] for related ideas). This model gives a detailed functional interpretation of the six-layered columnar cortical architecture and associated subcortical (thalamic) structures (cf. [6,4,1]). In particular, it relates cortical architecture to the integration of (top-down) prior knowledge with the actual (bottom-up) stream of sensory data and to processes controlling the learning of new representations. In short, the model hypothesizes three different but interacting processing systems at each stage of the cortical hierarchy: The A-system (cortical layers IV and lower III) accomplishes fast bottom-up processing where the first spike wave traveling up the cortical hierarchy can activate a coarse initial hypothesis about the stimulus at each level. In the B-system (layers II and upper III) the initial hypothesis is refined by slower iterative processes involving horizontal and vertical exchange of information. Finally, the C-system (layers V/VI) represents the local hypothesis of a macrocolumn which is fed back to the B-system of a lower level inducing expectations and

<sup>\*</sup>Corresponding author. Tel.: +49 69 89011 750; fax: +49 69 89011 749.

*E-mail addresses:* [andreas.knoblauch@honda-ri.de](mailto:andreas.knoblauch@honda-ri.de) (A. Knoblauch), [ruediger.kupper@honda-ri.de](mailto:ruediger.kupper@honda-ri.de) (R. Kupper), [marc-oliver.gewaltig@honda-ri.de](mailto:marc-oliver.gewaltig@honda-ri.de) (M.-O. Gewaltig), [ursula.koerner@honda-ri.de](mailto:ursula.koerner@honda-ri.de) (U. Körner), [edgar.koerner@honda-ri.de](mailto:edgar.koerner@honda-ri.de) (E. Körner).

*URL:* <http://www.honda-ri.de>.

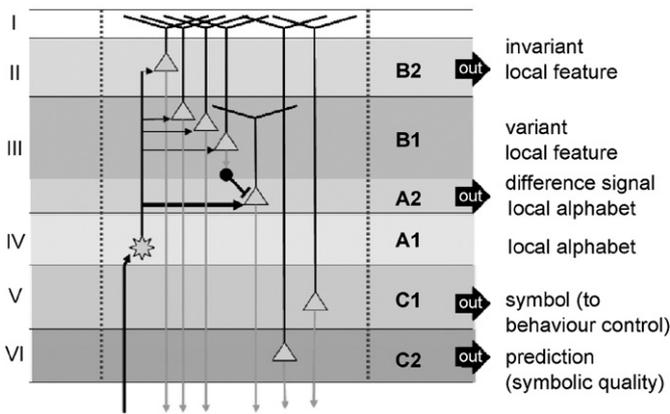


Fig. 1. Layered model of a cortical column as proposed in [16]. Three different processing subsystems corresponding to different vertical locations (i.e., layers) are intertwined within each cortical column. The A-system (middle layers) accomplishes fast bottom-up processing of sensory signals, the B-system (superficial layers) represents the input from the A-system in a refined way by activating more sparse and abstract representations and by exchanging information with neighboring columns. The C-system (deep layers) develops representations related to action/behavior and predictions used for feedback control. See text for more details. We have implemented parts of this model (see Fig. 2).

predictions for the present and future input signals. Predicted input signals are suppressed at an early cortical stage, and only differences between predicted and actual signals can reach the next higher level (cf. [18]). Learning of new representations is induced if the difference signal is too large and if the difference signal reaches the highest level of cortical integration, the hippocampal formation.

In order to substantiate these ideas we have designed and implemented a model of the cortical microcircuitry based on cell assemblies and associative memories [22,15,12]. To provide an adequate functional scenario, the model is integrated into a large scale implementation of the visual system including several primary and higher visual cortical areas as well as parts of the hippocampal formation, and subcortical structures involved in generating eye saccades. With this model we can demonstrate object classification and the learning of new object representations.

## 2. Model of the cortical microcircuitry

In accordance with ideas developed earlier in [16] (cf. [10,23,9]), we assume as a working hypothesis that the *basic function of a cortical column* is to adequately represent and predict its sensory inputs. To achieve this in a self-organizing, autonomous way, it is necessary to have access to different representational modi (such as reality vs. prediction) presumed to be located in different cortical layers. Thus, a cortical column has to represent external states  $v$  producing the observed input signals  $s$ . We found it reasonable to divide the state variable  $v = (w, a)$  into two rather independent components, one variable  $w$  describing “external” entities from the outside world and another variable  $a$  describing a local “internal actor”. In addition

to updating a state, the system should also be able to predict a future state  $w'$  (without accessing sensory input) and sensory inputs  $s'$ ,

$$w(t + \Delta t) = f(w'(t), s(t)), \tag{1}$$

$$a(t) = f_a(w(t), \dots), \tag{2}$$

$$w'(t) = g(w(t), a(t)), \tag{3}$$

$$s'(t) = h(w'(t)). \tag{4}$$

Fig. 2 illustrates this functional model and our current implementation employing simplified spiking associative networks [22,15,12].

Our model for the cortical microcircuitry consists of six neuron populations. Input layer A1 receives bottom-up “forward” input from a hierarchically lower area. A1 cells have feedforward and lateral inhibition with a “soft” winner-takes-all (WTA) characteristics. A1 cells project in a one-to-one manner to population A2. A2 has the same size as A1 but is modeled as a simplified spiking associative memory [15,12] including recurrent excitatory connections used to store auto-associatively prototypical representations of the input signals (corresponding to the “world states”  $w$  of the functional model). Similarly, population B is also modeled as spiking associative memory but the auto-associatively learned patterns have a more symbolic character since they are sparser and consist of rather randomly selected cells. Population B receives hetero-associative input from A2, where backprojections of B can inhibit A2 thereby suppressing familiar or expected signals. The representation of B can bias the selection of behaviorally relevant representations in population C1 (corresponding to the “action state”  $a$  of the functional model) which is also realized as spiking associative memory. An additional soft-WTA neuron population CU is used to represent conjunctions  $(w, a)$  of “world states” and “actions”. CU receives inputs from B and C1 via

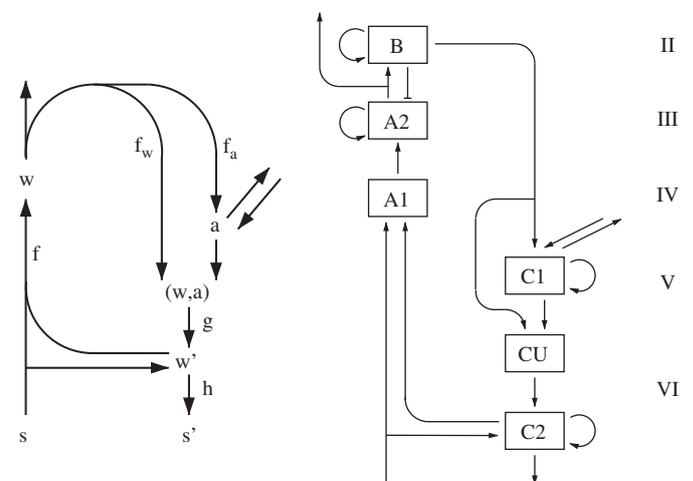


Fig. 2. Left: Basic functional circuit of a cortical column. Right: Current cell assembly based implementation of the cortical microcircuitry. Each box corresponds to a neuron population implemented either by a simplified spiking associative memory [15,12] or by a soft-WTA network.

random connections which is useful for hetero-associatively linking conjunctions  $(w, a)$  to predicted states  $w'$  represented in population C2 which, again, is modeled as spiking associative memory. Activity in A2 and B is transmitted forward to hierarchically higher areas, while predictive activity in C2 is fed back to lower areas and to the middle layers of the same column to bias relevant and suppress irrelevant representations.

Learning occurs in three different subsystems of the cortical columns. First, statistical (PCA-like) learning within populations A1 and A2 defining basis vectors or the coordination system of the macrocolumn. Second, clustering-like learning of prototypes in populations A2 and B. Third, predictive learning on “symbolic” states in populations B, C1, CU, and C2. Here, the short-cut link of the external forward input to C2 is used to learn the prediction  $(w, a) \rightarrow w'$  which exploits (e.g., by means of STDP) the fact that the response of CU to a new forward input signal has a longer latency than the C2 cells driven by short-cut inputs. Currently, only the latter two learning systems are implemented.

### 3. Integration into a large scale model for saccadic object recognition

Modeling the behavioral and predictive functionality of the C system requires an adequate functional scenario. For this we have designed a model of visual object recognition involving saccadic eye movements which allows the incremental refinement of an object hypothesis during a saccadic sequence. Here the first object view initiates an object hypothesis by “first glance” recognition which can be used to make saccades to other object parts and to predict the corresponding new object views. Then, by comparing the actual object view with the predicted object view, the initial hypothesis can be refined.

We have implemented a large-scale model of several primary (V1, V2) and higher visual cortical areas from the “what” (V4, IT) and “where” path (V6) as well as parts of the hippocampal formation (EC), and subcortical structures involved in generating eye saccades and triggering learning (see Fig. 3). The primary visual areas do standard preprocessing (orientations, edges), area V6 represents locations of key features (e.g., edges), whereas the superior colliculus (SC) integrates V6 input with an attentional bias from another area (Att/Sac1) to determine saccade targets. Area V4 represents fixation specific object views, whereas the central areas (EC, AC, IT) are used to learn modality-independent object categories (see [14] for more details).

We have modeled area V4 with the detailed columnar model described before (Fig. 2), while all other areas are modeled either as simple soft-WTA populations or as simplified spiking associative memories [15,12]. V4-A1 receives topographically organized inputs from V2 representing key features (e.g., corners) of simple line drawings of different classes of buildings (houses, churches, etc.). During execution of explorative saccades, V4-A2 and V4-B

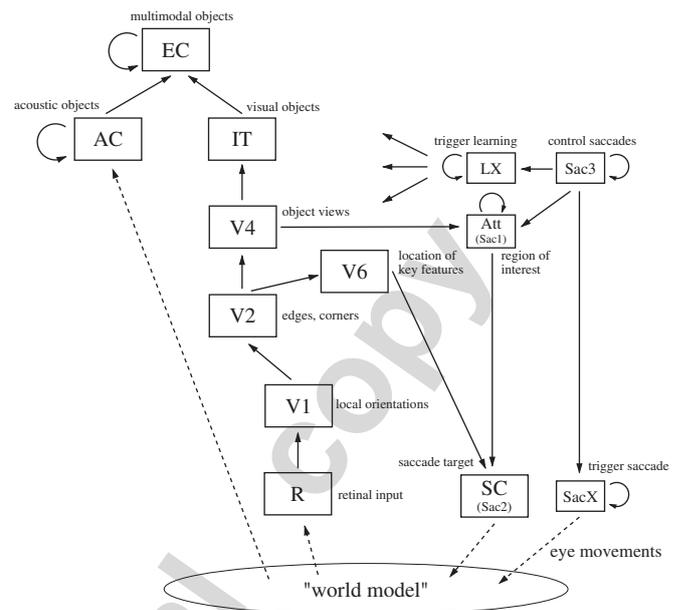


Fig. 3. Visual model for saccadic object recognition. The model consists of various visual areas (R, V1, V2, V4, V6, IT), auditory areas (AC), hippocampal areas (EC), saccade related areas (SC, Sac1-3), and some auxiliary areas triggering learning and the execution of saccades (LX, SacX). Currently only visual area V4 implements the full columnar model as shown in Fig. 2.

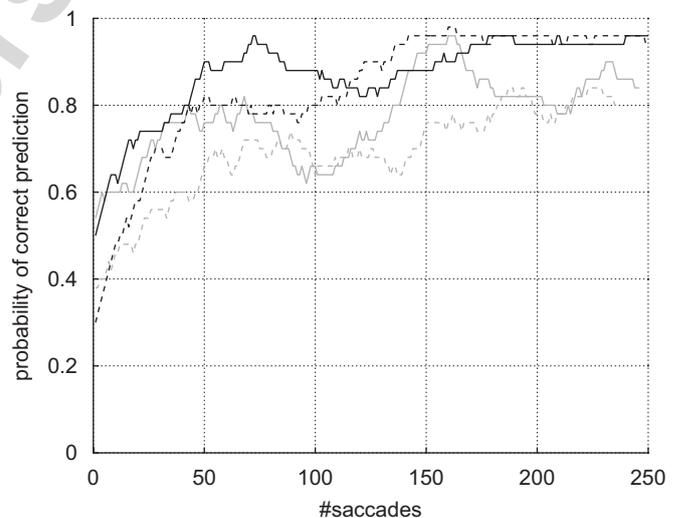


Fig. 4. Simulation results from area V4 (layer C2). The plot shows the probability of making a correct prediction about a future object view before executing a saccade. The curves show how the probability increases with the number of explorative saccades. The four lines correspond to different parameter sets. Gray: V4-CU consists of 100 neurons with high (solid) or low translation invariance in V2 (dashed). Black: high translation invariance, V4-CU consists of 400 (dashed) or 900 (solid) neurons. Here translation invariance is defined as the degree of convergence of V2 cells projecting to a V4 cell representing corresponding features.

learn symbolic prototypes of fixation-specific object views. V4-C1 can bias the direction of saccades, while V4-C2 predicts the outcome of the saccade. Fig. 4 illustrates the learning of the C2-predictions during exploration of a new

visual stimulus. One crucial parameter was the size of population V4-CU because this population is required to represent a large number of conjunctions of object views (V4-B) and saccade direction related representations (V4-C1). Another important factor was the degree of translation invariance of V4 representations (as measured by the convergence of V2 cells to a corresponding V4-A1 cell). This is because the precise fixation location can vary considerably between different saccades to the same key feature of an object view. More details of the model can be found in a technical report [14].

In summary, we have proposed a plausible model of the cortical columnar microcircuitry based on commonly accepted anatomical and physiological results. We have demonstrated its functioning by a simplified implementation integrated within a large scale model of the visual system.

#### 4. Summary and discussion

In this work we have given a functional interpretation to the well-known six-layered architecture of the cortex [16,6,4]. As explained in more detail in Section 2, we believe that the basic function of a cortical (macro-)column is to adequately represent and predict its sensory inputs (cf. [17,10,23,9,18]). This requires an ongoing comparison of sensory inputs with representations of previously experienced signals, more generally the integration of bottom-up and top-down processing, and the forming of new representations in case none of the existing hypothesis can explain the incoming signals. Specifically, we have hypothesized three interacting processing systems within each cortical column (see Fig. 1): the A system for fast bottom-up processing and the generation of difference signals (middle cortical layers); the B system for top-down driven refinement of bottom-up inputs (upper cortical layers); and the C system for controlling lower cortical areas as well as behavioral processes (cf. [9]).

Contrary to related models (such as ART networks [5]), our approach is based on the framework of cell assemblies and associative networks [11,27,3,21,15,12]. With this we are able to establish a much closer relation to real cortical networks, although we currently use only relatively abstract neuron models (such as soft-WTA populations or simplified spiking associative memories [15,12]). In contrast to more abstract cortex models, we explicitly address questions such as how to “allocate” new representational “nodes” (which corresponds in our framework to the synaptic learning of an additional local cell assembly), and which intra- and extra-columnar processes are required to support this (e.g., computing difference signals in the A system). Many related cell assembly based models implement only a single cortical layer or even identify a piece of cortex with only one or two associative networks (e.g. [3,22,8,2,7]). In contrast, we emphasize a more detailed layered structure of cortex [6,16] for segregating different representational modi and different

interareal information streams (see Sections 1 and 2). Moreover, we address combinatorial issues such as combining a large set of internal states  $w$  with another possibly large set of actions  $a$  (see Fig. 2) making use of *sparse* distributed representations where  $n$  neurons can represent almost  $n^2$  different entities [27,21,19,13].

To demonstrate the viability of our ideas we have implemented and integrated our model of a cortical macrocolumn with a large scale model of the visual system (see Section 3). In order to provide an adequate functional scenario such as saccadic object recognition, the visual model had to comprise several cortical areas of the ventral (“what”) and parietal (“where”) paths as well as additional structures to control saccades and learning. Our implementation gives a plausible illustration how a cortical column can develop representations for its incoming sensory inputs in order to successfully predict future sensory states (Fig. 4). Moreover, our model is consistent with many neuroanatomical and neurophysiological findings (e.g. [6,9,20,18]). However, in this work we have not attempted to achieve recognition results comparable to technically optimized artificial neural networks (e.g. [25,26,24]).

So far we have endowed only some key areas (V4 and partially IT) with the complete set of properties of our columnar model while the other areas rather perform preprocessing of sensory input and control of the saccades and learning. In future work we will extend our visual model to include the full set of columnar functionality across any level of the cortical hierarchy. We believe that our approach scales well to an integrative implementation of the visual system and parts of the hippocampus. Another key challenge will be to translate our results which are based on the crude dynamics of simplified neuron population models to more realistic spike based models of neurons and synaptic plasticity.

#### References

- [1] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, Cambridge, UK, 1991.
- [2] A. Billard, G. Hayes, DRAMA, a connectionist architecture for control and learning in autonomous robots, *Adaptive Behav. J.* 7 (1) (1999) 35–64.
- [3] V. Braitenberg, Cell assemblies in the cerebral cortex, in: R. Heim, G. Palm (Eds.), *Lecture Notes in Biomathematics*, vol. 21, Theoretical Approaches to Complex Systems, Springer, Berlin, Heidelberg, New York, 1978, pp. 171–188.
- [4] V. Braitenberg, A. Schüz, *Anatomy of the Cortex. Statistics and Geometry*, Springer, Berlin, 1991.
- [5] G.A. Carpenter, S. Grossberg, Adaptive resonance theory, in: M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, second ed., MIT Press, Cambridge, MA, 2003, pp. 87–90.
- [6] R.J. Douglas, K.A.C. Martin, Neuronal circuits of the neocortex, *Annu. Rev. Neurosci.* 27 (2004) 419–451.
- [7] R. Fay, U. Kaufmann, A. Knoblauch, H. Markert, G. Palm, Combining visual attention, object recognition and associative information processing in a neurobotic system, in: S. Wermter, G. Palm, M. Elshaw (Eds.), *Biomimetic Neural Learning for Intelligent*

- Robots, vol. 3575, Lecture Notes in Artificial Intelligence, Springer, Berlin, Heidelberg, 2005, pp. 118–143.
- [8] E. Fransen, A. Lansner, A model of cortical associative memory based on a horizontal network of connected columns, *Network Comput. Neural Syst.* 9 (1998) 235–264.
- [9] R.W. Guillery, Branching thalamic afferents link action and perception, *J. Neurophysiol.* 90 (2003) 539–548.
- [10] J. Hawkins, *On Intelligence*, Times Books Henry Holt, New York, 2004.
- [11] D.O. Hebb, *The Organization of Behavior. A Neuropsychological Theory*, Wiley, New York, 1949.
- [12] A. Knoblauch, Neural associative memory for brain modeling and information retrieval, *Inf. Process. Lett.* 95 (2005) 537–544.
- [13] A. Knoblauch, On compressing the memory structures of binary neural associative networks, Internal Report HRI-EU 06-02, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany, April 2006.
- [14] A. Knoblauch, R. Kupper, M.-O. Gewaltig, U. Körner, E. Körner, Design and simulation of a cortical control architecture for object recognition and representational learning, in: H. Tsujino, K. Fujimura, B. Sendhoff (Eds.), *Proceedings of the third HRI International Workshop on Advances in Computational Intelligence*, Honda Research Institute, Wako, Japan, 2005.
- [15] A. Knoblauch, G. Palm, Pattern separation and synchronization in spiking associative memories and visual areas, *Neural Networks* 14 (2001) 763–780.
- [16] E. Körner, M.-O. Gewaltig, U. Körner, A. Richter, T. Rodemann, A model of computation in neocortical architecture, *Neural Networks* 12 (1999) 989–1005.
- [17] R. Kupper, A. Knoblauch, M.-O. Gewaltig, U. Körner, E. Körner, Simulations of signal flow in a functional model of the cortical column, *Neurocomputing*, this issue.
- [18] S.O. Murray, D. Kersten, B.A. Olshausen, P. Schrater, D.L. Woods, Shape perception reduces activity in human primary visual cortex, *Proc. Nat. Acad. Sci.* 99 (23) (2002) 15164–15169.
- [19] J.-P. Nadal, Associative memory: on the (puzzling) sparse coding limit, *J. Phys. A Math. Gen.* 24 (1991) 1093–1101.
- [20] K. Nakamura, C.L. Colby, Updating of the visual representation in monkey striate and extrastriate cortex during saccades, *Proc. Nat. Acad. Sci.* 99 (6) (2002) 4026–4031.
- [21] G. Palm, On associative memories, *Biol. Cybern.* 36 (1980) 19–31.
- [22] G. Palm, *Neural Assemblies. An Alternative Approach to Artificial Intelligence*, Springer, Berlin, 1982.
- [23] R.P. Rao, D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nat. Neurosci.* 2 (1) (1999) 79–87.
- [24] M. Riesenhuber, T. Poggio, Models of object recognition, *Nat. Neurosci.* 3 (2000) 1199–1204.
- [25] S. Ullman, S. Soloviev, Computation of pattern invariance in brain-like structures, *Neural Networks* 12 (7/8) (1999) 1021–1036.
- [26] H. Wersing, E. Körner, Learning optimized features for hierarchical models of invariant object recognition, *Neural Comput.* 15 (2003) 1559–1588.
- [27] D.J. Willshaw, O.P. Buneman, H.C. Longuet-Higgins, Non-holographic associative memory, *Nature* 222 (1969) 960–962.



**Andreas Knoblauch** is currently a senior scientist at the Honda Research Institute Europe in Offenbach. He received his Ph.D. from the University of Ulm, Germany, in computer science and neural information processing, 2003, and holds a master degree from the same university in computer science and electrical engineering, 2000. From 2003 to 2005 he did postdoctorates at the Neural Information Processing Department at the University of Ulm and

the Cognition and Brain Sciences Unit at the MRC Cambridge, UK. His research interests include modeling of cortical circuits, learning in associative networks, information retrieval, and the design of simulation software.



**Rüdiger Kupper** is a senior scientist at the Honda Research Institute Europe in Offenbach/Main, Germany. He has been a member of this institute since its foundation in 2003. Rüdiger received his Ph.D. in physics from the University of Marburg, Germany, for a neural simulation model of distance-invariant object recognition. His work at the Honda Research Institute is dedicated to the discovery and computer-simulation of the fundamental processing strategies and control architecture of the human cortex. His vision is a neural theory of intelligence.



**Marc-Oliver Gewaltig** was born in 1967 in Germany. He received his M.Sc. in physics from the Ruhr-University of Bochum in 1994. After associations with the Ruhr-University Bochum and the Albert-Ludwigs University Freiburg, he received his Ph.D. in 1999. In 1998, he started working for Honda R&D Europe (Deutschland) GmbH. Since 2003, he is a principal scientist at the Honda Research Institute Europe GmbH in Offenbach/Germany. His research interests are the neuronal principles of information processing in the primate brain as well as methods for simulation of large cortical network.



**Ursula Körner** is a principal scientist at the Honda Research Institute Europe in Offenbach, Germany. She received a degree in neurophysiology at the Moscow State University and obtained a Ph.D. in physiology at the Humboldt State University in Berlin. Ursula Körner's research interests concentrate on principles of information processing and learning in cortical structures.



**Edgar Körner** is president of the Honda Research Institute Europe GmbH. He studied electrical engineering, control engineering, and biomedical cybernetics at the Ilmenau Institute of Technology, Germany, where he was appointed professor in 1988. In 1992, he joined Honda R&D in Wako near Tokyo as chief scientist for brain-like computation research. In 1997, he moved to Honda R&D Europe, to head the Future Technology Research division, focussing on computational neuroscience, evolutionary technology, and cognitive robotics. Since 2003, Prof. Dr. Körner is president of the Honda Research Institute Europe GmbH. His research focus is on brain-like artificial neural systems and self-organization of knowledge representation for autonomous robots.