

# **Design and simulation of a cortical control architecture for object recognition and representational learning**

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

**2005**

**Preprint:**

This is an accepted article published in Proceedings of the 3rd HRI International Workshop. The final authenticated version is available online at:  
[https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

# DESIGN AND SIMULATION OF A CORTICAL CONTROL ARCHITECTURE FOR OBJECT RECOGNITION AND REPRESENTATIONAL LEARNING

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

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

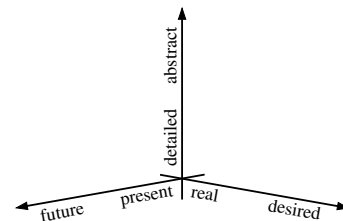
## ABSTRACT

In this work we present simulation results substantiating a previously proposed model of computation in neocortical architecture (Körner et al., *Neural Networks* 12:989-1005, 1999). For each stage of the cortical hierarchy we hypothesize three interacting columnar systems for (A) fast forward recognition, (B) refined recognition by feedback, and (C) behavior/prediction related processes, roughly corresponding to middle (IV), superficial (II/III), and deep (V/VI) cortical layers. In a first example we implement a simple system for word recognition. Focusing on the dynamics, we explore the interaction of the A and B systems in recognizing words, quenching out expected signals, and representing new words based on previously learned representations. In a second large-scale implementation of the visual and saccadic system we additionally demonstrate the learning of new object representations and the generation of predictions within the C system based on saccadic sequences.

## 1. INTRODUCTION

Recognizing and categorizing previously experienced objects and scenes as well as learning new representations is still one of the hardest problems in artificial intelligence and related disciplines such as computer vision or robotics. 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 at each level of representation. Thus, any solution has to detail (i) how prior knowledge is to be integrated (top-down) with the actual (bottom-up) stream of sensory data in a meaningful way, and (ii) when and how new representations are to be created and stably integrated into the previously learned knowledge hierarchy. Because biological organisms are currently the only systems capable of solving these tasks to a satisfactory degree, we want to use the growing knowledge about the anatomy and physiology of the brain to incorporate this knowledge in biologically inspired models.

In this work, we present first simulation results substantiating a previously proposed model of computation in neocortical architecture [1]. This model gives a detailed func-



**Fig. 1.** Cortical processes must be able to distinguish between different representational modi. E.g., representational states may refer to present, future, reality, wish, signal, or symbol. We have the idea that, at least for higher mammals, a single cortical column can represent several modi at the same time, and that different modi relate to different cortical layers.

tional interpretation of the well-known six-layered columnar cortical architecture and related subcortical (thalamic) structures. It hypothesizes three different but interfering processing systems at each stage of the cortical hierarchy: The A-system (including the middle cortical layers IV and III) 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 II and III) the initial hypothesis is refined by slower iterative processes, involving horizontal and vertical exchange of information. Finally, the C-system (deep layers V and VI) 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. Recognized or predicted input signals are suppressed at an early cortical stage, and only differences between predicted and actual signals can reach the next higher level. 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 hippocampus.

These ideas are illustrated by two example implementations of the model. The first example is a very simple (but instructive) system for word recognition consisting of three cortical levels representing letters, syllables, and words. Focusing on the dynamics, we show how the different process-

ing systems interact in order to quench out expected signals and accomplish symbolic recognition of words, and how representations for new words can be constructed based on old representations. The second example is a large-scale implementation of the visual system, involving several primary and higher visual cortical “areas” (denoted as V1, V2, V4, V6, IT) as well as parts of the hippocampal formation (EC, HC), and further subcortical structures involved in generating eye saccades. With this model we can demonstrate object classification and learning of new object representations based on the incremental refinement of an object hypothesis during a saccadic sequence.

## 2. FUNCTIONAL MODEL OF A CORTICAL COLUMN

Although, it is well known for a long time that neocortical anatomy exhibits a 6-layered structure, modelers have often neglected this fact when modeling a cortical patch by a single “monolithic” neuron population (e.g., [2, 3, 4]). This may be attributable to the wish to focus on a single layer or the lack of adequate computational resources to simulate more detailed models, but also to doubting or underestimating the functional significance of discrete within- or between-layer synaptic connections which appear to have a rather “fuzzy” character [5, 6].

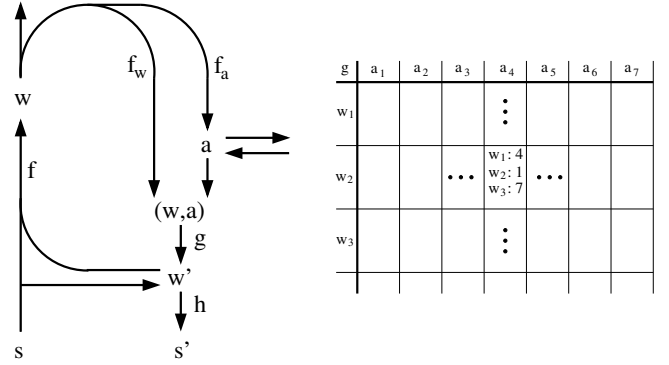
In accordance with ideas developed earlier in [1] (cf., [7, 8]), we assume that the *basic function of a cortical column* is to adequately represent and predict (or generate) its sensory inputs. To achieve this in a self-organizing, autonomous way, it is necessary to have access to different representational modi such as actual vs predicted sensory input. We propose that different representational modi of the same entity (e.g., orientation at a particular position in the visual field) are located in different layers within the same column rather than monolithically in different columns or areas (see Fig. 1).

What does such a *generative model* look like? We can assume that the model represents external states that produce the observed sensory inputs. Thus, at each time  $t$  the model must represent a state  $v$  from the state space  $V$  (or more generally, a probability distribution on the state space describing in which state the columnar system “believes” to be in). Then the system should be able to use sensory input  $s$  to update the state  $v$  according to a function  $f$ ,

$$v(t + \Delta t) = f(v(t), s(t)) \quad (1)$$

where  $\Delta t$  can be interpreted either as a fixed small simulation step size (for continuous models) or as a variable time interval between two events (for discrete models driven by events such as the onsets of stimuli presented in a sequence).

It makes sense to divide the state variable  $v = (w, a)$  into two rather independent entities, one variable  $w$  describ-



**Fig. 2. Left:** Basic functional circuit of a cortical column. Sensory input  $s$  is used to update the current world state  $w$ . This is used to choose an appropriate action  $a$ . World state and action can be used to predict the next world state  $w'$  and next sensory input  $s'$ . **Right:** A simple histogram representation of the conditional probability density  $g(w, a)$  on (discrete) states and actions: The example shows the density of the predicted “world” states  $w'$  when being in state  $w_2$  and performing action  $a_4$ . Learning is accomplished by simply incrementing bin  $k$  of the histogram in row  $i$ , column  $j$  whenever evidence  $(w_i, a_j) \rightarrow w_k$  is experienced. Our model essentially implements such a histogram representation where the states or state combinations  $(w_i, a_j)$  are “coded” with distributed cell assemblies in order to relieve combinatorial problems and reduce the number of required neuronal units [9, 10, 11, 12, 4].

ing “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  $s'$ ,

$$w(t + \Delta t) = f(w'(t), s(t)) \quad (2)$$

$$a(t) = f_a(w(t), \dots) \quad (3)$$

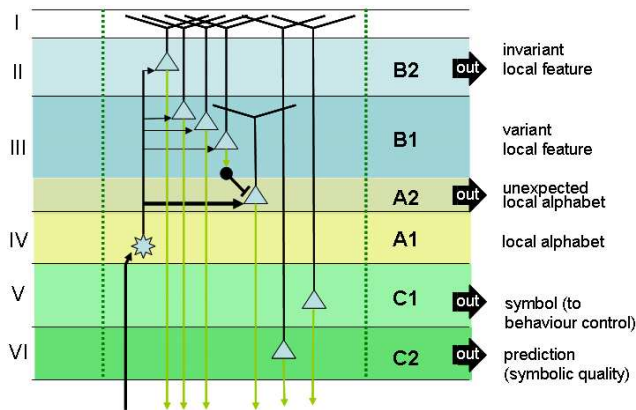
$$w'(t) = g(w(t), a(t)) \quad (4)$$

$$s'(t) = h(w'(t)) \quad (5)$$

We will also refer to  $f$  as the “forward recognition function”, to  $g$  as the “predictive function”, and  $h$  as the “backward function”.

How can we learn the recognition, prediction, and backward functions? In general, updating the model will involve two phases: (i) finding the most probable  $w$ , (ii) assuming fixed world and actor states  $(w, a)$ , the functions can simply be learned by counting up experienced evidence. For example, the prediction function  $g$  is essentially a conditional probability (the probability to get to state  $w'$  given state  $w$  and action  $a$ ) which, for discrete states and actions, can be represented by evidence histograms for each combination  $(w, a)$  (see Fig. 2).

By comparison with well known anatomical facts we can match our functional model (Fig. 2) with the layered organization of neocortex ([1, 13, 6]; see Fig. 3). We believe that the forward recognition function ( $f$ ) is located in the middle and upper layers, while the remaining functionality, related to behavior and predictions, is located in the lower layers (see also introduction). Furthermore, we believe that  $f$  is split up into two subsystems, one for fast bottom-up recognition (A system) and another for refined recognition employing feedback (B system).



**Fig. 3.** Layered model of a cortical column as proposed in [1]. 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. 7, cf. Fig. 2).

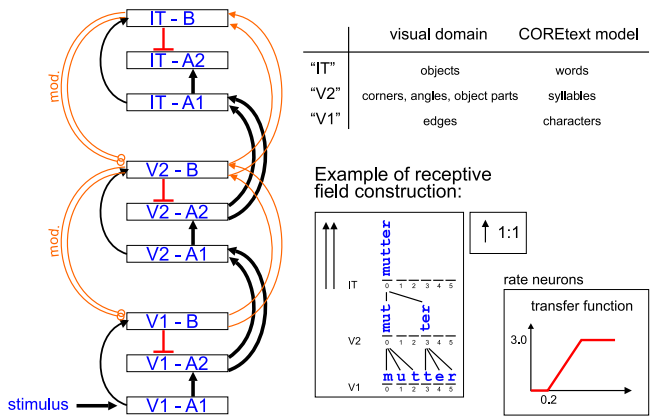
### 3. CORETEXT MODEL

In our first model example we investigate the basic activation dynamics within the columnar A and B subsystems. For this we have implemented a model that has been reduced to the minimum of necessary subsystems and connections, that still maintain the desired function.

It consists of three cortical stages, which are internally identically structured (Fig. 4). Just for convenience, they are denoted as V1, V2, and IT, but the setup is completely generic and may refer to any place in the cortical hierarchy.

In order to have concrete examples for features and receptive field properties, and concentrate on the principle signal flow inside the columns, we chose an example of *word*

recognition. We assume that text and images have similar structural properties. E.g., text consists of words, consisting of syllables which again are constructed from characters. For the principle purpose described here, this is sufficiently similar to visual scenes, consisting of objects, which again are constructed from features such as corners and edges (despite the reduction to one-dimensional space; cf. [14]).



**Fig. 4.** The “CoreText” model implements three cortical levels (denoted V1, V2, IT for convenience) including the columnar A and B subsystem to explore the neural activation dynamics and the “switching-off” mechanism (inhibition from B to A2) as proposed in [1]. We use word recognition as a simplified example, giving exact rules for the construction of receptive fields.

In our current implementation, area V1 represents letters, area V2 syllables, and area IT words. Each area consists of three populations A1, A2, and B corresponding to the cortical layers of the A and B subsystem (see Fig. 3). Each population consists of simple threshold units. For simplicity, each cell codes a single entity (letter, syllable, word). Synaptic connections from A1 to A2 are one-to-one (i.e., A2 is essentially a copy of A1). Connections from A2 to B are such that “symbolic” representations in B classify entities corresponding to the next higher level from the lower-level inputs A2 (e.g., in area V2, layer B classifies words from syllable based representations in A2).

Currently, the model is hard-wired without any synaptic plasticity where we focus on the activation dynamics, in particular the *switching-off* of familiar signals in layer A2 induced by inhibition from B [1]. Results from the CoreText model are illustrated by two simulation examples (Figs. 5, 6).

In the following, we describe results from a complementary large scale model which aims to implement the full functionality of a cortical column and focusses in particular on the interaction of the A,B subsystems with the behavior and prediction related C system.

```

t=10
-----
IT-A1: 0:      IT-A2: 0:      IT-B: 0: mut-ter( 1.3) göt-t
IT-A1: 1:      IT-A2: 1:      IT-B: 1:
IT-A1: 2:      IT-A2: 2:      IT-B: 2:
IT-A1: 3:      IT-A2: 3:      IT-B: 3:
IT-A1: 4:      IT-A2: 4:      IT-B: 4:
IT-A1: 5:      IT-A2: 5:      IT-B: 5:

V2-A1: 0:      V2-A2: 0:      V2-B: 0: mut( 3.0) göt( 2.2)
V2-A1: 1:      V2-A2: 1:      V2-B: 1:
V2-A1: 2:      V2-A2: 2:      V2-B: 2: cher( 0.4) sten( 0.
V2-A1: 3:      V2-A2: 3:      V2-B: 3: te( 3.0) tern( 3.0)
V2-A1: 4:      V2-A2: 4:      V2-B: 4: er( 0.4)
V2-A1: 5:      V2-A2: 5:      V2-B: 5:

V1-A1: 0: m( 1.0)      V1-A2: 0:      V1-B: 0: m( 3.0)
V1-A1: 1: u( 1.0)      V1-A2: 1:      V1-B: 1: u( 3.0)
V1-A1: 2: t( 1.0)      V1-A2: 2:      V1-B: 2: t( 3.0)
V1-A1: 3: t( 1.0)      V1-A2: 3:      V1-B: 3: t( 3.0)
V1-A1: 4: e( 1.0)      V1-A2: 4:      V1-B: 4: e( 3.0)
V1-A1: 5: x( 1.0)      V1-A2: 5:      V1-B: 5: x( 3.0)

```

**Fig. 5.** Stimulation example using a known word “mut-ter”. The stimulus is decomposed into known parts in the B-system, which shows self-supportive activation. The activation in the B-system switches off all known parts from the signal stream in the A2 system. Since the word can be fully explained from knowledge, no activation remains at all in the A2-system.

```

t=8
-----
IT-A1: 0:      IT-A2: 0:      IT-B: 0: va-ter(1.5) va-ters
IT-A1: 1:      IT-A2: 1:      IT-B: 1:
IT-A1: 2:      IT-A2: 2:      IT-B: 2:
IT-A1: 3:      IT-A2: 3:      IT-B: 3:
IT-A1: 4:      IT-A2: 4:      IT-B: 4:
IT-A1: 5:      IT-A2: 5:      IT-B: 5:

V2-A1: 0:      V2-A2: 0:      V2-B: 0: va(3.0) hat(3.0) ta
V2-A1: 1:      V2-A2: 1:      V2-B: 1:
V2-A1: 2:      V2-A2: 2:      V2-B: 2: ter(3.0) ters(3.0)
V2-A1: 3:      V2-A2: 3:      V2-B: 3:
V2-A1: 4:      V2-A2: 4:      V2-B: 4:
V2-A1: 5:      V2-A2: 5:      V2-B: 5:

V1-A1: 0: v(1.0)      V1-A2: 0:      V1-B: 0: v(3.0)
V1-A1: 1: a(1.0)      V1-A2: 1:      V1-B: 1: a(3.0)
V1-A1: 2: t(1.0)      V1-A2: 2:      V1-B: 2: t(3.0)
V1-A1: 3: x(1.0)      V1-A2: 3: x(0.9)      V1-B: 3: x(0.3)
V1-A1: 4: x(1.0)      V1-A2: 4:      V1-B: 4: x(3.0)
V1-A1: 5: x(1.0)      V1-A2: 5: x(0.9)      V1-B: 5: x(0.3)

```

**Fig. 6.** Stimulation example of an unknown word that is similar to the known word “vater”, but contains two distorted characters. The stimulus is decomposed into known parts in the B-system, which shows self-supportive activation. Known parts are switched off from the signal stream in the A2 system. The two unknown characters remain activated in the A2 system, indicating the new parts of the stimulus that have to be learned.

#### 4. MODEL OF SACCADIC OBJECT RECOGNITION

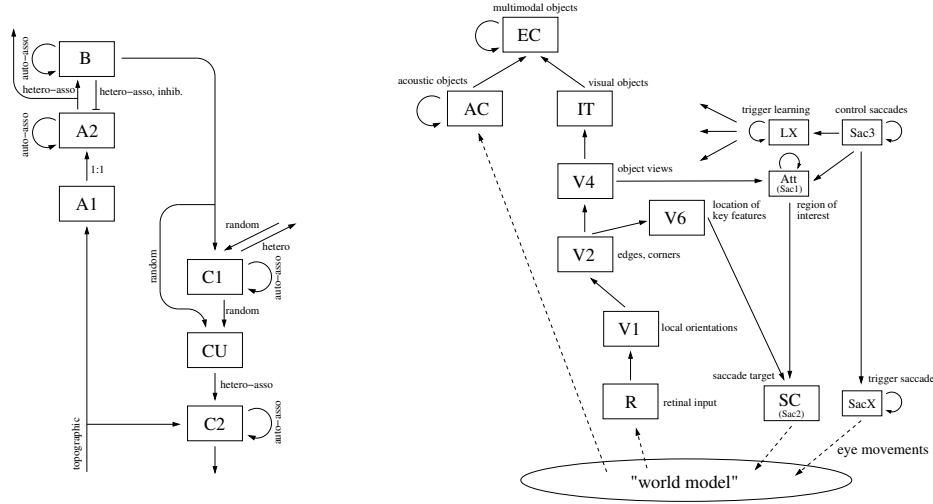
Modeling the behavioral and predictive functionality of the C system requires a more dynamic scenario which covers at least some of the functional implications hypothesized by our model. For this we have designed a model of visual object recognition involving saccadic eye movements which allows the incremental refinement of an object hypothesis during a saccadic sequence. Here the first object view initiates an object hypothesis by “holistic” 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 strengthened or weakened.

We have implement a 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 further subcortical structures involved in generating eye saccades and triggering learning (see Fig. 7, right panel). At this point we have used only a simple neuron model to simulate and explore the basic neuronal dynamics and columnar functions without too much computational expenses. Most neuron populations are modeled either as a simple k-WTA population (i.e., at each simulation step, the  $k$  most excited neurons are activated), or as simplified spiking associative memories ([4, 15, 16]; cf. [17]).

In retinal area R we represent binary images (size  $81 \times 81$ ) of simple line drawings of buildings (Fig. 8). The cells in primary visual area V1 receive input via oriented Gabor-like filter kernels (8 orientations per spatial location). Area V2 represents additionally corner parts (i.e., conjunctions of two edges, 120 per location). Area V6 averages over all corners types at a spatial location thereby representing the positions of *key features* in a visual scene (cf. Fig. 8). Area V4 represents invariant visual object views at a lower spatial resolution. Area IT (*inferior temporal*) represents visual objects. For learning new objects we activate a static representation in an auditory area AC to label the new object with a spoken word. Converging multimodal (visual, auditory) input to a further area EC (*entorhinal cortex*) of the hippocampal complex allows the binding of different object views to a single object representation. Furthermore there are a number of presumably subcortical auxiliary areas involved in generating saccades and triggering learning (areas Sac1,Sac2,Sac3,SacX,LX). In particular, we have implemented a simple model of the superior colliculus (SC/Sac2) representing target locations of saccades determined by the location of key features (input from V6) and an attentional window (region of interest represented in Sac1 and biased by V4). Additionally, area Sac3 represents “control sequences” to control the execution of saccades (area SacX) and the triggering of learning (area LX) in specific phases of a repeating “frame generating” oscillation.

Each area consists of one or several neuron populations to implement our ideas of the layered architecture of cortical columns. Fig. 7 (left) shows the architecture of the columnar model as implemented for area V4. The other areas implement only parts of our columnar model, mostly the fast forward processing A-system. Our columnar model as implemented for area V4 consists of 6 neuron populations. Input layer A1 receives input from area V2 and consists of  $120 \times 9$  cells corresponding to the 120 V2 features and 9 spatial fields (central plus 8 peripheral), each V4 cell pooling over the corresponding set of V2 cells. A1 cells have no local recurrent excitatory connections but feedforward and lateral inhibition with a “soft” winner-takes-all that is emulated by activating only the  $k = 13$  most excited cells (k-WTA). A1 cells project in a one-to-one manner to population A2 corresponding to lower layer III. A2



**Fig. 7. Left:** Columnar model as implemented for our current simulations. Boxes correspond to cell populations (corresponding to particular cortical layers, cf. Fig. 3), arrows to synaptic connections between the cell populations/layers. Populations with recurrent connections (A2,B,C1,C2) are implemented as simplified spiking associative memories [4, 15, 16]. The additional population CU (“combinatorial units”) is used to combine representations of B and C1 (cf. Fig. 2). **Right:** Layout of our visual model of 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, S1, S3), and some auxiliary areas triggering learning and the execution of saccades (LX, SX). Currently only area V4 implements the full columnar model.

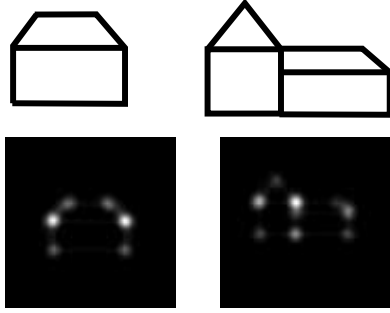
has the same size as A1 but is modelled as a spiking associative memory [4, 15, 16]. Thus A2 has recurrent excitatory connections used to store auto-associatively feature vector prototypes learned during saccadic object recognition. Similarly, population B (size 100) is also modelled as spiking associative memory but the auto-associatively learned patterns consist of randomly selected cells ( $k = 5$ ) and therefore have rather symbolic character. Population B receives hetero-associative input from A2, where backprojections of B can inhibit A2 thereby emulating the ideas of quenching off expected signals as described above (see section 3). The representation of B can be used to bias the selection of behaviorally relevant symbols in population C1 (corresponding to cortical layer V). C1 is also realized as spiking associative memory of 100 cells representing currently 9 possible target directions of saccadic eye movements (central plus 8 equispaced directions). During learning of new objects this population receives random inputs and thereby biases the production of random saccades to explore new visual scenes or objects. An additional neuron population CU is used to represent conjunctions  $(w, a)$  of “world states” and “actions” (see Fig. 2). The most simple way to do this is to model CU as a  $k$ -WTA population (CU has size 100 or 2500 and  $k = 13$ ) receiving inputs from B and C1 via random connections leading to nearly “uncorrelated” activity in CU for any combination of  $w$  and  $a$ . This is useful for hetero-associatively linking conjunctions  $(w, a)$  to pre-

dicted states  $w'$  represented in population C2. C2 is also modelled as spiking associative memory (size 100,  $k = 13$ ). The short-cut link of the external sensory input (from V2) to C2 is used to learn the prediction  $(w, a) \rightarrow w'$  (see below). The prediction represented in C2 can be used to bias expectations in other (lower) areas of the cortical hierarchy and/or to narrowing the search space within the same cortical column (e.g., via a modulating input from C2 to A1,A2, or B).

Learning may occur in three different hypothesized subsystems of the cortical columns.

1. PCA-like basis vector learning system for cell populations A1 and A2.
2. Clustering-like learning algorithm in populations A2 and B.
3. Predictive learning on symbolic states in populations B, C1, CU, and C2.

The first learning system is concerned with generating appropriate basis vectors for adequately describing the sensory inputs (possibly only during a critical learning period). So far this learning system has not been implemented in our model, i.e., all sensory inputs to layer A1 are hard-wired. Instead, we have conducted some isolated simulation experiments (unpublished results) which suggest that a simple standard statistical learning procedures (such as essen-



**Fig. 8.** We used simple line drawings for stimulating the visual model, e.g., a house or a church (upper panel). Preprocessing in areas R,V1,V2,V6 essentially extracts key features (such as corners) as the basis for saccade generation and object view recognition in the higher areas. Lower panels show the V6 representations of the key features corresponding to the house and church stimuli.

tially additive Hebbian learning plus an adequate synaptic normalization procedure) will do the job similar to PCA or maximization of reconstruction quality and sparseness which can generate plausible receptive field properties [18, 19].

The second learning system involving populations A2 and B performs a kind of clustering on the space spanned by the basis vectors of A1/A2, quite similar to ART networks [20] but relying on cell assemblies instead of simple nodes. Essentially, it implements the operation

$$\text{IF } (|A2 - H \cdot B| > \Theta) \text{ THEN NEW SYMBOL}$$

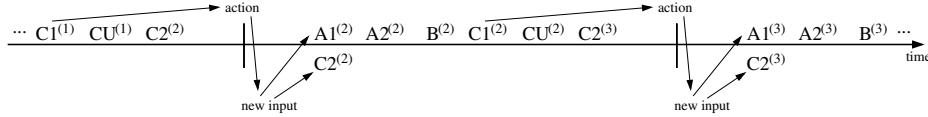
where A2 and B denote the activity vectors of the corresponding neuron populations, H is the matrix of hetero-associative inhibitory connections from B to A2, and  $|\dots|$  essentially sums over the residual activity after subtracting B from A2. This learning process involves the following steps. Input from A1 initiates a retrieval in the spiking associative memory A2, i.e. A1 will activate a prototype (or a mixture of several prototypes) in A2 that is most similar to A1. Then A2 will activate a corresponding cell assembly in the spiking associative memory B. This activation pattern is fed back via inhibitory hetero-associative connections to A2. Ideally, the two representations in A2 and B match each other and B will quench the activity in A2. However, if the activity vector in A2 is too far from a previously learned prototype then there will be considerable residual activity. This then initiates the learning of new “symbols”. For that, the original activity of A1 and A2 is stored auto-associatively in A2 and a new “symbolic” cell assembly (generated by noise) is stored in B and bidirectionally and hetero-associatively linked with the new A2 assembly. Additionally, a new cell assembly (of the same quality as A2) is stored in C2. We note that it may be challenging

to create a more detailed neurodynamic model which implements these discrete steps using realistic spiking neuron and plasticity models.

The third learning system essentially learns to make predictions on the symbolic states learned by the second learning system. Functionally, it implements the learning of the conditional probability density histograms  $p(w'|w, a)$  illustrated in Fig. 2. First a new unquenched sensory signal  $s$  enters the column. On the main path it will travel via A1, A2, B, C1 finally to C2 (Fig. 9). However, on the shortcut to cortical layer VI it will directly enter C2 and transiently activate, by hetero-association, a symbol corresponding to the new sensory input. Since at that time the old state symbols are still active in B, C1, and CU, a simple asymmetric STDP-like Hebbian learning rule will hetero-associatively link the old CU-representation to the new C2 symbol. Similar to the second learning system it will also be a challenge to implement this procedure employing more realistic neuron and plasticity models.

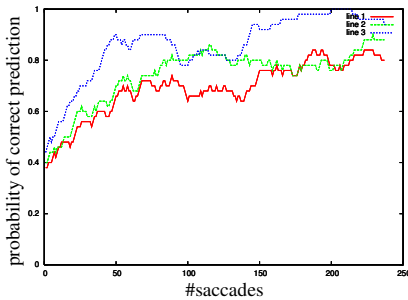
In order to test our model we have conducted simulations applying simple line stimuli as shown in Fig. 8. In each simulation run we presented one new stimulus or several new stimuli in a sequence. During the presentation the system will saccade on the key features (i.e., corners) of the stimulus. Eye movements are controlled by the saccade control system (areas Sac1,Sac2,Sac3,SacX,LX) as follows: In the first phase the system executes a saccade defined by the visual target map in area SC/Sac2 (superior colliculus). After the saccade up to three correction saccades follow in order to center the fixation on the location of the most salient feature. After that learning is enabled for 10 simulation steps by modulating input from area LX (which could correspond to some hippocampal and/or subcortical areas) to several targets: (i) within the A/B system of area V4 in order to learn a new object view in case the current object view differs too much from the views experienced previously; (ii) within the C system of area V4 in order to learn to predict the outcome of the saccade; (iii) between area IT and EC in order to associate the particular object view representation in V4 and IT with a static auditory representation in the area AC via entorhinal area EC. At the same time, population C1 of area V4 preselects a future saccade direction out of 9 possible directions ( $2\pi i/8$  for  $i = 0, 1, \dots, 7$  or “center”). Via area S1 this biases a particular region of the visual field in area SC. Since SC receives also the locations of the key features from area V6, the most active cells in the target map of SC can again select the location of the following saccade.

Fig. 10 shows some simulation results when stimulating with the house illustrated in Fig. 8. The plot shows that our model is able to rapidly learn new object views and to learn to predict the outcome of saccades under a certain object hypothesis. For the latter the system has to learn something like the prediction histogram of Fig. 2. Combining a



**Fig. 9.** Temporal dynamics of the activation of the different populations or layers in a cortical column within an action-perception cycle. An action (for example) causes new sensory input which reaches first A1 and via the input shortcut C2. Then the main stream proceeds by activating A1, B, C1, CU, and C2. The shortcut is used to learn the predictive hetero-association  $CU^{(i-1)} \rightarrow C2^{(i)}$ . Assuming the cell populations being located as illustrated in Fig. 3 this is consistent with neurobiological results (e.g., see [13]). Indices  $^{(i)}$  refer to the represented entity (i.e., the  $i$ -th object view).

large number of world states with a large number of possible behaviors/actions the complete histogram will be quite large. Here, we have compressed the representations by using cell assemblies where a state or state combination is not represented by a single cell or a conjunctive unit, but instead by distributed cell groups [9, 11, 2]. Theoretically, this could reduce the number of required neural units from  $M = \#states \times \#actions$  (if we use one conjunctive unit for each combination of world states and actions) to approximately  $\sqrt{M}$  for large networks of distributed cell assemblies [12, 16].



**Fig. 10.** Simulation results from area V4. The plot shows the probability of making a correct prediction about a future object view before executing a saccade as a function of the number of saccades when learning a new object (here the house of Fig. 8). Different lines correspond to different model parameters. Best results occur if population CU is large and the corresponding representations are uncorrelated.

## 5. CONCLUSION

We have described two initial implementations of a previously suggested model of the columnar organization of the cortex [1]. The first model for word recognition (CoreText; see section 3) focussed on the activation dynamics in the A and B columnar subsystems (see Figs. 3,7 left) which is used for fast bottom-up processing in the A system and subsequent top-down driven refinement in the B system. We

demonstrated the viability of previously presented ideas [1], in particular the quenching-off of familiar signals to be able to attend to residuals conveying potentially new information, and to compose new hierarchically structured representations from the old (already learned) ones.

In a complementary large scale model of the visual system (see section 4) we have focussed on the interaction of the columnar A and B subsystems with the C system which is required to influence behavior/actions and to make predictions about future sensory inputs [1]. This model is able to generate new representations for new object views and to learn predictions about the outcome of saccades applying the framework of cell assemblies [9, 10, 11, 12, 4]. To this end it was necessary to model a larger part of the visual system involving several visual cortical areas of the ventral (what) and parietal (where) stream as well as additional areas 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 predict future sensory states. 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. Consequently, this model example can also implement only a limited version of the switching-off demonstrated in the first (CoreText) model example.

In future work we will extend and unify our models of the visual system to include to 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 very simple neuron models to more realistic spike based models of neurons and synaptic plasticity.

## 6. REFERENCES

- [1] E. Körner, M.-O. Gewaltig, U. Körner, A. Richter, and T. Rodemann, "A model of computation in neocorti-



- cal architecture.,” *Neural Networks*, vol. 12, pp. 989–1005, 1999.
- [2] G. Palm, *Neural Assemblies. An Alternative Approach to Artificial Intelligence.*, Springer, Berlin, 1982.
- [3] R. Ritz, W. Gerstner, U. Fuentes, and J.L. van Hemmen, “A biologically motivated and analytically soluble model of collective oscillations in the cortex. II. Applications to binding and pattern segmentation.,” *Biol. Cybern.*, vol. 71, pp. 349–358, 1994.
- [4] A. Knoblauch and G. Palm, “Pattern separation and synchronization in spiking associative memories and visual areas.,” *Neural Networks*, vol. 14, pp. 763–780, 2001.
- [5] M. Abeles, *Corticonics: Neural circuits of the cerebral cortex.*, Cambridge University Press, Cambridge UK, 1991.
- [6] V. Braitenberg and A. Schüz, *Anatomy of the cortex. Statistics and geometry.*, Springer-Verlag, Berlin, 1991.
- [7] J. Hawkins, *On Intelligence.*, Times Books Henry Holt, New York, 2004.
- [8] R.P. Rao and D.H. Ballard, “Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects.,” *Nature Neuroscience*, vol. 2(1), pp. 79–87, 1999.
- [9] D.O. Hebb, *The organization of behavior. A neuropsychological theory.*, Wiley, New York, 1949.
- [10] D.J. Willshaw, O.P. Buneman, and H.C. Longuet-Higgins, “Non-holographic associative memory.,” *Nature*, vol. 222, pp. 960–962, 1969.
- [11] V. Braitenberg, “Cell assemblies in the cerebral cortex.,” in *Lecture notes in biomathematics (21). Theoretical approaches to complex systems.*, R. Heim and G. Palm, Eds., pp. 171–188. Springer-Verlag, Berlin Heidelberg New York, 1978.
- [12] G. Palm, “On associative memories.,” *Biological Cybernetics*, vol. 36, pp. 19–31, 1980.
- [13] R.J. Douglas and K.A.C. Martin, “Neuronal circuits of the neocortex.,” *Annu. Rev. Neurosci.*, vol. 27, pp. 419–451, 2004.
- [14] B.W. Mel and J. Fiser, “Minimizing binding errors using learned conjunctive features.,” *Neural Computation*, vol. 12, pp. 247–278, 2000.
- [15] A. Knoblauch, “Synchronization and pattern separation in spiking associative memory and visual cortical areas.,” *PhD thesis, Department of Neural Information Processing, University of Ulm, Germany*, 2003.
- [16] A. Knoblauch, “Neural associative memory for brain modeling and information retrieval.,” *Information Processing Letters*, vol. 95, pp. 537–544, 2005.
- [17] R. Fay, U. Kaufmann, A. Knoblauch, H. Markert, and G. Palm, “Combining visual attention, object recognition and associative information processing in a neurobotic system.,” in *Biomimetic Neural Learning for Intelligent Robots*, S. Wermter, G. Palm, and M. Elshaw, Eds., vol. 3575 of *Lecture Notes in Artificial Intelligence*, pp. 118–143. Springer-Verlag, Berlin Heidelberg, 2005.
- [18] E. Oja, “Simplified neuron model as a principal component analyzer.,” *Journal of Mathematical Biology*, vol. 15(3), pp. 267–273, 1982.
- [19] B.A. Olshausen and D.J. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?,” *Visison Research*, vol. 37, pp. 3311–3325, 1997.
- [20] G.A. Carpenter and S. Grossberg, “Adaptive resonance theory.,” in *The Handbook of Brain Theory and Neural Networks, Second Edition*, M.A. Arbib, Ed., pp. 87–90. MIT Press, Cambridge, MA, 2003.