Towards a mesoscopic-level canonical circuit definition for visual cortical processing

Georg Layher
Institute of Neural Information Processing
Ulm University
Ulm, Germany
georg.layher@uni-ulm.de

Tobias Brosch
Institute of Neural Information Processing
Ulm University
Ulm, Germany
tobias.brosch@uni-ulm.de

Heiko Neumann
Institute of Neural Information Processing
Ulm University
Ulm, Germany
heiko.neumann@uni-ulm.de

ABSTRACT

The mammalian cortex is organized into different layers which are characterized by different cell clusterings and prominent lateral fiber connections. As a primary organizational principle driving signal inputs enter mainly at layer IV, feedforward output activations leave at superficial, while modulating feedback signals leave at deep layers. Likewise, modulatory input signals from other areas mainly enter in superficial layers. Based on such compartmental structure we suggest an abstract formulation of composite structural elements which form building blocks to define canonical elements for columnar computation in cortex. There is evidence that the brain makes use of an operational set of canonical computations, like, e.g., sensory signal filtering, reentrant response gain enhancement, noise suppression, and normalization. As a further abstraction, we define a dynamical three-stage model of processing in a cortical column for processing that allows to investigate their dynamic response properties. Some examples of the analysis of the dynamics are presented together with some simulation results that highlight properties of visual processing in model cortices. Finally, we demonstrate how learning mechanisms that adapt the connection weights can be incorporated in such a scheme capable to form feature representations and functional sub-networks in an adaptive fashion.

Keywords

visual cortex, canonical circuit model, cortical dynamics, cortical column, unsupervised learning

1. INTRODUCTION

Evidence suggests that the primate cortex employs a set of canonical operations that serve as structured components for a modular computational design of brain function. Such components are identified, for example, for the driving feedforward input signal filtering for feature detection [19, 20], the activity gain control via reentrant signals that deliver contextual information [37], and activity normalization in the visual pathway [24]. The signal flows of feedforward and feedback processing define a system of counterstreams [44] which are combined at the level of individual cortical columns [26, 27]. Different theoretical frameworks have been defined for the recurrent stream interaction, namely *predictive coding* and *biased competition*. Based on such counterstream interactions, the interactive effects of contextual (down-) modulation are observed throughout the cortex. These effects have been conceptualized already in several previous studies to arrive at specific roles of such contextual interaction, namely activity normalization to generate feature selectivity in vision [8, 7, 25] as well as in other sensory modalities, decision-making and cognitive control [38].

In this paper, we show that such functional principles can be linked to patterns of cortical circuit elements that define chunks which are composed into larger circuits. Such component circuits are depicted in Fig. 1 which will be detailed in Section 2. Based on these computational elements we derive a simplified canonical circuit model that provides an abstract view of the columnar computations over different compartments. The model scheme consists of three hierarchically organized stages which are formally described by first-order dynamic equations to characterize the rate of change in activation over a spatially organized feature representation. This, in turn, allows to investigate the dynamics of such systems and qualitative changes of their behavior.

2. COMPARTMENTAL STRUCTURE OF VISUAL CORTEX AND DEFINITION OF A CANONICAL CIRCUIT MODEL

2.1 Layered structure and compartments

The structure of mammalian cortices appears to be organized into different layers organized into cortical columns [32]. Six layers have been identified that differ in appearance of cell types and their densities, the pattern of lateral connectivity, and the connection of input and output from/to different areas.

Based on experimental findings, we suggest that the layered structure can be described in terms of compartments which allows to link the collection of cells and lateral interactions to a function related to input processing and feature extrac-

BICT 2015, December 03-05, New York City, United States

DOI 10.4108/eai.3-12-2015.2262448

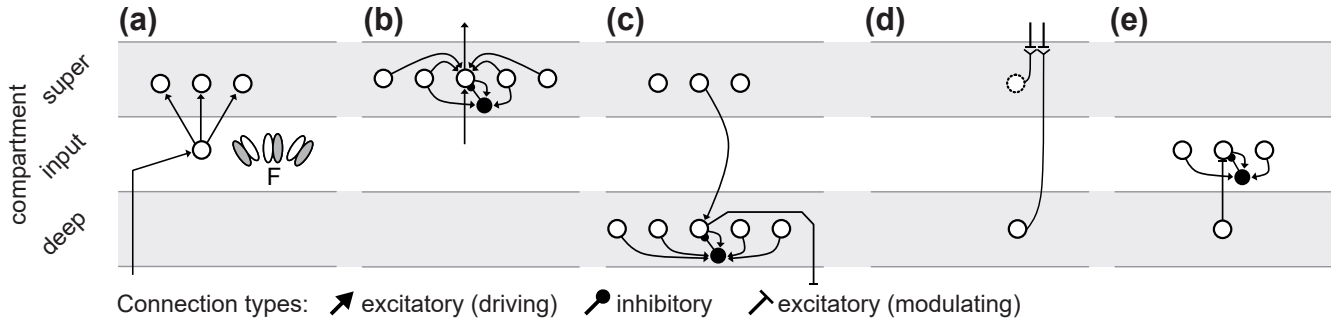

Figure 1: Component circuit elements defined over different compartments in visual cortex. The vertical columnar structure that runs orthogonal to the compartments assigns a spatial localization to the computations and the localized interactions operating upon the feature maps. The different patterns are defined by (a) the driving feedforward signal input entering mainly in the input compartment with the feature selective filtering (denoted by F) and the subsequent forward projection to the superficial compartment cell representations, (b) lateral interaction between filter responses to link related activities in the feature maps and generate output that is fed to other areas, (c) resulting superficial cell activities project to the deep compartment at the same position (within a column) to drive cells which, in turn, compete with cells in a larger spatial pool of cells (defined over a neighborhood in the space-feature domain(s)) and send modulatory feedback connections to different areas (situated lower in the hierarchy), (d) cells in the superficial and deep compartments receive modulatory excitatory input signals which are integrated from dendritic terminations in the superficial compartments, (e) deep compartment cells also modulate activations that arrive in the input compartment and close an intra-cortical loop of activation.

tion. Like in [41] we distinguish the input compartment, where the driving input signals from previous stages mainly terminate. The superficial and the deep level compartments build the other two substructures. In each of these compartments lateral interactions between cells serve to implement different types of functional links between activities in the space-feature domain. The vertical organization into columns allows to localize the resulting feature representation in space (Fig. 1).

Driving feedforward signals mainly enter in the input compartment where the signals are filtered by specific filter functions (denoted by F; Fig. 1(a); e.g., orientation, movement). Lateral interaction in the superficial compartment (Fig. 1(b)) leads to, e.g., feature enhancement by cooperative interaction [2] or response normalization by targeting inhibitory inter-neurons in a local spatial and/or featural neighborhood [6, 8]. Cell activities in the superficial compartment project to deep layer competitive cells in an even larger pool. We suggest that this interaction scheme acts mainly in a divisive fashion upon the target cells' activation [6, 8]. Cells in the deep compartment send out feedback projections to areas lower in the sensory hierarchy which are modulatory (Fig. 1(c)) completing the main functional feedforward pathway [13].

Cells in the superficial and the deep compartments receive modulatory input signals which are integrated from terminations in the superficial compartment. Recipient cells reach out with their distal dendrites up to the upper section of the superficial compartment where their dendritic terminations contact the input fibers from other areas which are situated higher up in the hierarchy. These connections are mainly modulatory such that only paired signal activation from feedforward input and correlated feedback lead to response amplifications [27] (Fig. 1(d)). Together with the divisive pool inhibition defined in (c) this implements a mechanism of selective amplification while reducing activities at other locations in the space-feature domain (biased competition principle). Deep compartment cells not only send modulatory feedback projections to lower areas but also branch to close an intra-cortical loop by amplifying the activities generated by the filtering in the input compartment [30] (Fig. 1(e)). In order to prevent activities to grow without bounds we suggest that inhibitory cells within these compartments balance these filter activations. The definition follows the recent observation that a dense architecture of interneuron connectivity defines a canonical feature in the design of cortical microcircuits [36].

2.2 Mesoscopic level model of a canonical columnar cortical circuit

In the proposed abstraction of the model architecture, we have combined circuit elements (a) to (d) to define a three-stage mesoscopic level columnar model. Our model architecture utilizes representations consisting of two-dimensional sheets of visuotopically arranged model columns implemented as excitatory-inhibitory (E-I) pairs of single-voltage compartments (point-like) entities [18]. Model columns interact by lateral connectivity, feeding input signals as well as ascending and descending connections to and from other model areas. The 1st order dynamics of the membrane potential v for such model cells is captured by

$$C_m \dot{v} = -\alpha v + (E^{ex} - v) g_{ex} + (E^{in} - v) g_{in} \,, \qquad (1)$$

where we assume resting state $E^{leak} = 0$ with leak conductance $g_{leak} = \alpha$, excitatory and inhibitory input conductances g_{ex} and g_{in}, respectively, and C_m the membrane capacitance. A sketch of the specification of the canonical model cascade is depicted in Fig. 2(a).

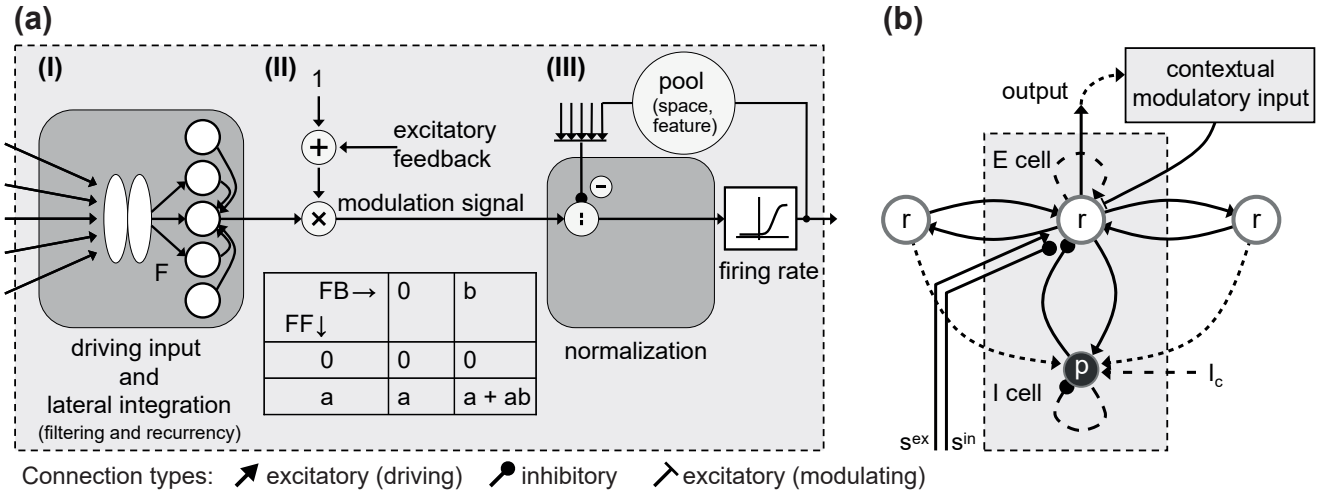

Figure 2: (a) Three-stage cascade of the canonical columnar circuit architecture. (I) Driving input signals (from the previous stage of processing) are filtered (F) such that a sampled (multi-dimensional) feature space representation is constructed; (II) activity in spatially organized feature maps can be amplified by modulatory signals from other areas, e.g., feedback from representations higher up in the hierarchy (with an asymmetry in the computational role of feedforward (FF) and feedback (FB) signals, see table); (III) down-modulation of activity through a competitive interaction between the current state amplitude and the integrated (weighted) output activation (firing rates) of cells in a surrounding pool that is defined over the spatial and the (multi-dimensional) feature domain. (b) Reduced model architecture of the columnar three-stage model. The model is condensed to an E-I circuit by lumping the stage of filtering and lateral integration as well as the stage of contextual modulation by reentrant signals into the E-cell and the pool for the activity normalization into the I-cell (see text for details; adapted from [3]).

The first stage of the columnar model cascade (Fig. 2(a) I) combines the driving feedforward input signal integration and filtering F that generates distributed and sampled feature representations (circuit patterns (a) and (b) shown in Fig. 1). These representations are spatially integrated to form a space-feature map in which the activation strength denotes the likelihood for the presence of a specific feature in the sensory map representation. The resulting activities communicate via lateral recurrent connections. For example, co-aligned contrast cells with similar orientation preference mutually strengthen their activities according to the saliency of connected contour or boundary outlines [23, 2, 31, 33, 16]. For clarity of the schematic outline we have omitted in Fig. 2(a) to explicitly depict a layer of inhibitory cells which are fed by the excitatory cells in the superficial compartment (which was made explicit in Fig. 1(b)). The activation that is fed to the next stage of the canonical columnar circuit is calculated by

$$A_{i\theta}^{FF+lat} \propto \left(s_{i\theta}^{F} + \gamma_{lat} \cdot \left\{ \sum_{k\phi} g_r(r_{k\phi}) \cdot \Lambda^{+}_{ik,\theta\phi} \right\} \right), \quad (2)$$

with i and θ denoting the locations in the spatial and the feature domain, respectively; $s_{i\theta}^{F}$ defines the filter responses and $g_r(r_{k\phi})$ is the firing-rate of a lateral cell (with $g(\cdot)$ as the sigmoidal firing-rate function); the lateral connectivity is defined by a kernel $\gamma_{lat} \cdot \Lambda^{+}_{ik,\theta\phi}$ to specify the selective weighting in the spatial and the feature domain (c.f. [3]).

The second stage of the columnar model cascade (Fig. 2(a) II) formalizes the influence of excitatory reentry signals, e.g., feedback signals from areas higher up in the processing hierarchy. Such signals enter the superficial compartment level [1] and preferentially contact the apical dendrites of pyramidal cells in layers 2/3 and 5 (compare the iconized sketch in Fig. 1(d)). It has been demonstrated that signals of the feedforward sensory stream are combined with reentrant signal stream activation at the level of individual columns and even individual pyramidal cells [27]. The latter can also arise from disinhibitory feedback interaction [4]. Since the cells in superficial compartments directly project to the cells in the deep compartment in their respective column we have lumped the excitatory influence of the reentrant signals on the feeding activity and formally represent them by a single computational stage. Evidence suggests that the influence of the reentry is modulatory and enhances coincident feeding input [21]. In order to account for this asymmetry in the role of the bottom-up/reentrant pathways [42] we utilize a tonic bias that can be combined with a reentrant signal net^{mod} (that can originate from, for example, areas higher up in the processing hierarchy; impact modulated by λ)

$$A_{i\theta}^{modFF} \propto A_{i\theta}^{FF+lat} \cdot (1 + \lambda \cdot net_{i\theta}^{mod}), \quad (3)$$

which yields the combined contingency table as depicted in Fig. 2(a) II. A similar mechanism has been proposed to model the effects of synchronizing activities across multiple areas (by linking fields, [10]) and selective attention [40].

Finally, the third stage of the columnar model cascade (Fig. 2(a) III) summarizes the output computation of the columnar circuit. Most importantly, this stage realizes a competitive interaction between target cells and a large spatial pool

of surrounding cells with different feature selectivities. We suggest that the circuit pattern depicted in Fig. 1(c) enables a division of the target activations by the pool activity via shunting inhibitory interaction. This implements the nonlinear response normalization of the overall local energy (over the surround pool of responses) as reported in [17, 6]. The generic pool integration is formalized by a weighted integration of activities over a spatial neighborhood and the feature domain(s) such that

$$A_{i\theta}^{pool} \propto \sum_{k,\phi} g_r(r_{k,\theta}) \cdot \Lambda^{-}_{ik,\theta\phi}\,, \tag{4}$$

with the firing-rate function $g_r(\cdot)$ and Λ^{-} denotes a kernel of weights that considers spatial distances $(i-k)$ as well distance metrics in the feature domain $(\theta-\phi)$. Such a model definition allows the parametric shaping of tuned and untuned feature selectivities for the pool integration [25].

The output that is generated to feed to subsequent stages (driving) and those activations that are sent back to previous stages in the cortical hierarchy (modulating) is generated in different compartments, as sketched in Fig. 1 [30]. From a functional point of view on cortical computation we only need to distinguish the influence different signals have at their termination contacts. Therefore, we define only one single output signal via the firing-rate of cells in a model column. The interaction from the grid of inhibitory cells is assumed to incorporate shunting inhibition in accordance to the basic dynamics denoted in Eq. 1. We assume that the integrated pool activation exerts a mainly divisive inhibition on the target response (see Fig. 2(a) III). The compartmentalization of the columnar circuit model allows the influence of inhibitory activity across different cortical layers, as, e.g., utilized in the dynamic attention model proposed by [45].

2.3 Canonical circuit model and its analysis

In order to define the dynamics of the canonical circuit model from the components as specified in Eqs. 2, 3, and 4, we arrive at a 1st order 2-D system

$$\begin{aligned}
\tau_r \dot{r}_{i\theta} =& -\alpha_r r_{i\theta} + (\beta_r - r_{i\theta}) \cdot P_{i\theta}^{ex} - (\delta + r_{i\theta}) \cdot s_{i\theta}^{in} \\
& - (\eta + \gamma \cdot r_{i\theta}) \cdot g_p(p_i)\,, \\
\tau_p \dot{p}_{i\theta} =& -\alpha_p p_{i\theta} + \beta_p \cdot A_{i\theta}^{pool} + (I_c)_i \\
=& -\alpha_p p_{i\theta} + \beta_p \cdot \left\{ \sum_{k,\phi} g_r(r_{k,\phi}) \cdot \Lambda^{-}_{ik,\theta\phi} \right\} + (I_c)_i\,,
\end{aligned} \tag{5}$$

with excitatory net input defined by the initial filtering and the lateral recurrent integration

$$\begin{aligned}
P_{i\theta}^{ex} =& A_{i\theta}^{modFF} \\
=& \left(s_{i\theta}^{ex} + \gamma_{lat} \cdot \left\{ \sum_{k} g_r(r_k) \cdot \Lambda^{+}_{ik,\theta} \right\} \right) \\
& \cdot \left(1 + \lambda \cdot net_{i\theta}^{mod} \right)\,.
\end{aligned} \tag{6}$$

$g_{p,r}(\cdot)$ denote gain functions that map the potentials p_i and $r_{i\theta}$ to output firing-rates. The constants τ_r, $\tau_p > 0$ define the membrane constants for an excitatory unit and the inhibitory pool, respectively. $\Lambda^{+/-}$ denote the input weights to define the lateral interaction and the extent of the pool (e.g. with Gaussian weights). An additional input I_c is considered for the pool inhibition to later account for unspecific raises or lowerings of the inhibitory gain. The parameters $\alpha, \beta, \delta, \eta, \gamma, \lambda \geq 0$ are constants to denote the activity decay, the saturation levels, the relative strength of the pool inhibition and the FB strength (see [3] for more details).

The dynamic properties of the model architecture have been analyzed in order to evaluate different influences of specific components in the model definition. In order to do so, we investigated a reduced columnar model composed of an excitatory-inhibitory (E-I) pair for a single feature [47]. The components of such E-I pair are shown in Fig. 2(b) in which the core component is located within the dashed box in the center. The excitatory E component is represented by r activations in Eq. 5 while the inhibitory I component is represented by the pool activation p. In order to account for the influence laterally connected model cortical columns have, self-excitation is included for the E-cell component as well as for the I-cell component. Additional inhibitory action can be considered by the extra input denoted by I_c (see Fig. 2(b)).

3. ANALYSIS OF DYNAMIC PROPERTIES AND COMPUTATIONAL EXAMPLES

3.1 Analysis of dynamic properties

We investigate the dynamic properties of the model framework by reducing it to a single location with $s^{in} = 0$ and with $P^{ex} = s + \gamma_{SE} \cdot g_r(r)$ where the modulatory input $1 + \lambda \cdot net^{mod}$ effectively scales s and $\gamma_{SE} \cdot g_r(r)$. We thus omitted here the scaling term which yields the 2D system in the plane (compare [3])

$$\begin{aligned}
\tau_r \dot{r} =& f(r,p) = -\alpha_r r + (\beta_r - r) \cdot (s + \gamma_{SE} \cdot g_r(r)) \\
& - (\eta_r + \gamma \cdot r) \cdot g_p(p)\,, \\
\tau_p \dot{p} =& g(r,p) = -\alpha_p p + \beta_p \cdot g_r(r) + I_c - \eta_p \cdot g_p(p)\,.
\end{aligned} \tag{7}$$

This simplified 2D system was the basis to investigate the stability properties of such columnar circuit. An important characterization of the dynamic circuit properties was found to depend on the lateral integration of excitatory activity in the r-layer, which, in the single dimension analysis, is expressed by the self-excitation γ_{SE}. If the strength of γ_{SE} is small given that $\partial_r\ f(\bar{r},\bar{p}) < 0$ then the system is stable. For moderate γ_{SE} strength given that $0 \leq \partial_r\ f(\bar{r},\bar{p}) < |\partial_p\ g(\bar{r},\bar{p})|$ then the system can be either stable or unstable (which needs to be analyzed case-by-case). If the strength of γ_{SE} is large given that $|\partial_p\ g(\bar{r},\bar{p})| \leq \partial_r\ f(\bar{r},\bar{p})$ then the system leads to instabilities (see [3]; with $\partial_a \equiv \partial/\partial a$).
We further investigated the structure of the respective phase plane diagram for given pairs of γ_{SE} and driving inputs $s \equiv I$ and η, $\eta_p = 0$. The results of this computational analysis are displayed in Fig. 3. We identified the different zones with the qualitative properties of a single stable equilibrium (green) and regimes with excitatory nullclines that also have a positive slope, i.e. can form inhibition stabilized regimes [35, 22]. In such a regime, increased input can paradoxically lead to smaller responses and has recently been reported to occur in real cortical cells and small networks. In this regime, we identified two cases. First, that of a single stable equilibrium (blue). Second, a case in which three equilibria with two stable equilibria separated by an unstable equilibrium (red) occur. As the excitatory nullclines (red curves in the phase space) get increasingly more pronounced (as in the red region) a bifurcation eventually occurs at the blue-red

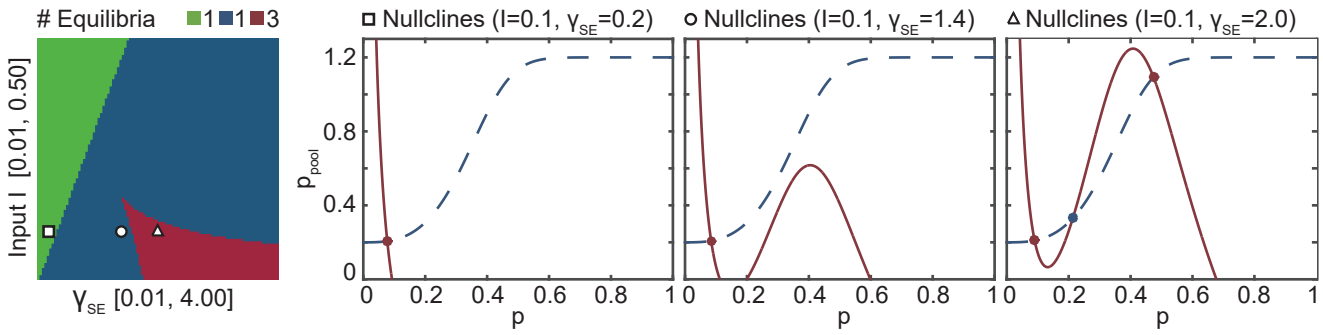

Figure 3: Numerical investigation of stability/instability regions as a function of different strengths of (lateral) self-excitation γ_{SE} and different driving input strengths $s \equiv I$. Three regions with qualitatively different response characteristics have been identified (first column) and corresponding phase spaces are displayed. Parameter regimes exist (green) in which a single stable equilibrium occurs, and regimes with a excitatory nullcline that has a positive slope, i.e. that can result in ISNs. Here, two cases can be distinguished, one with a single stable equilibrium (blue), and regimes with three (two stable and one unstable) equilibria (red). Associate phase space diagrams show the shapes of excitatory (red) and inhibitory (dashed blue) nullclines (for details, see text; $\eta_r = \eta_p = 0$, $\alpha_r = \alpha_p = \beta_r = \beta_p = \gamma = 1$, $I_c = 0.2$).

boundary. The bifurcation creates a new equilibrium which splits into a pair of unstable and stable equilibria. This is the first time that qualitatively different stability criteria have been identified for such canonical circuit definition and also that bifurcations may occur as a function of intrinsic network properties. The conditions of ISN (inhibition stabilized network) occurrence depend on the particular shape of the excitatory nullclines given by

$$p = f(r)\,, \\ = g_p^{-1}\left(\frac{-\alpha_r r + (\beta_r - r)\cdot(s + \gamma_{SE}\cdot g_r(r))\cdot(1 + \lambda\ net^{mod})}{\gamma \cdot r}\right). \tag{8}$$

We refer to Fig. 3 for examples of the different nullcline shapes (see [3] for analytical details).

3.2 Example simulation results

The computational properties of larger networks composed of two-dimensional layers of laterally connected canonical columns are shown for different model simulations. First, we demonstrate an example of contour enhancement of an early layer by reentry of a higher-level representation (corresponding to, e.g., visual cortex V1 and V2). The effects of the modulatory enhancement are evaluated by using methods of signal detection theory, namely the z-measure that considers the differences between the mean activity of the target (contour) against all neurons (scaled by the responses' standard deviation) and the r-measure, that is the ratio between such mean values. In the example shown in Fig. 4(a) an improvement of these measures of $(\Delta r, \Delta z) \approx (1.1, 0.5)$ is achieved. In addition, the responses at the contours' ends are enhanced by the interaction of modulatory feedback and normalization which leads the contours' representations to be enhanced at the line ends, and the surrounding regions to be suppressed.

In a second example we demonstrate the necessity of a modulatory excitatory enhancement in the case of texture boundary detection for search targets to be detected. The detectability of region boundaries composed of homogeneous texture items depends on the change of gradient along the feature dimension [34]. In case that the orientation change within the background and the target regions is low the boundary is easily detected for small to moderate orientation contrasts based on feedforward processing only. For noisy and cluttered stimuli, however, i.e. when the orientation change is increased within the homogeneous parts (shown here is background noise of 20°) additional feedback is required to stabilize the boundary formation. The model predicts that V4 → V2 feedback is of particular importance in a cascaded model architecture with model areas V1, V2, and V4, each of which is bidirectionally coupled by driving and reentrant connections and composed of the canonical columnar circuits. If the V4-V2 recurrency is selectively removed from the model, significant drops in the response ratios occur (see Fig. 4(b)).

4. MECHANISMS OF LEARNING

In the proposed model, the initial stage performs feature selective filtering of the input (Fig. 1). In many cases of modeling, e.g., oriented contrast detection, such filters have been implemented by a convolution operation. If the filter selectivities become more specific, pre-programming might no longer be suitable and a learning mechanism should be employed to let such weighting functions self-organize from exposition of sets of training data (see e.g. [5] for a biologically inspired learning algorithm for recurrent networks). Here, we consider a complementary learning mechanism to build certain selectivities at an input compartment in a model architecture. In other words, we investigate learning the connection weights for the fan-in structure of the filter mechanisms being involved to build feature selective representations. This is combined with learning the fan-out connections afterwards.

A biologically plausible learning mechanism should employ a generically unsupervised scheme which might be augmented

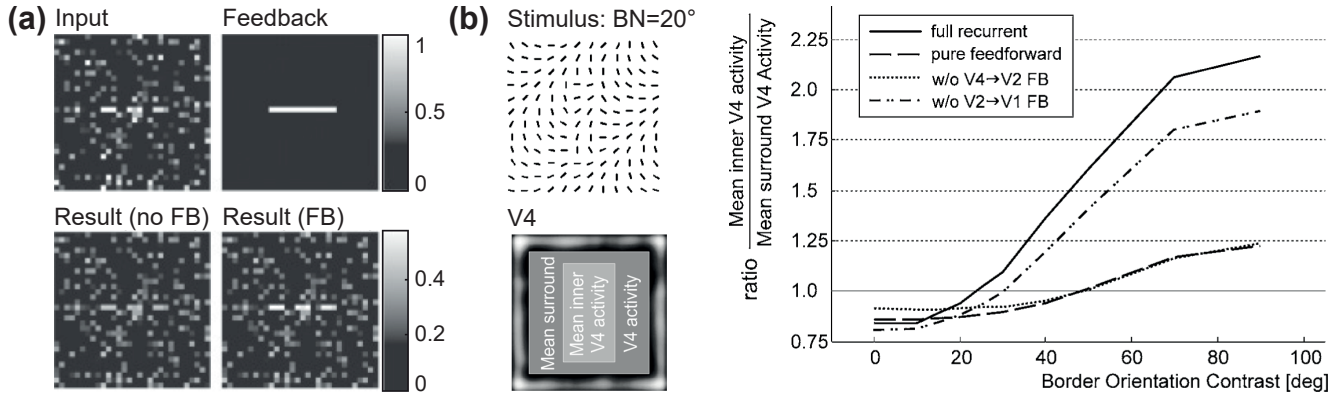

Figure 4: (a) Processes of contour enhancement (for details, see text; the result is replicated from [3]).(b) Processes of texture boundary formation in visual search target detection (for details, see text; the result is replicated from [43]).

by some mechanisms of self-supervision and/or reinforcement which is controlled by the interaction of the learning systems (as part of an agent) who interacts with its sensory environment. In general, such learning incorporates a mechanism to adapt the connection weights $\dot{w}_{ij} = F(u_i, v_j, w_{ij}, \text{success})$ which considers correlated pre- and postsynaptic activity, the current weight (as a state variable), and possibly, a measure of a given network activity's success related to the current task, or goal. In the model mechanism sketched below, we consider adaptation of feedforward as well as feedback connections in a recurrently connected hierarchical model network (c.f. [5]). We consider weight changes that occur at fixed discrete learning steps, formally described as $w_{ij}^t = w_{ij}^{t-1} + \Delta w_{ij}^t$. The weight adaptation defined by $F(\cdot)$ does not use an external teacher signal, such as in, e.g., back-propagation [29], but rather operates in an unsupervised fashion using modified correlative learning. To adapt the feedforward connection weights we suggest a biologically plausible competitive learning mechanism which combines two basic principles of learning: (a) a modified version of Hebbian correlation learning [14] and (b) a trace rule of activation in which the Hebbian learning mechanism is stabilized by considering a temporal trace of the pre- or post-synaptic neuron [12, 46]. The discretized weight adaptation reads

$$\Delta w_{ji}^{FF} = \eta_{FF} \cdot z_i^{post} \cdot (u_j^{pre} - z_i^{post} \cdot w_{ji}^{FF}) \,. \tag{9}$$

The trace mechanism is realized by temporal low-pass filtering of the post-synaptic activity at the fan-in convergence site, $\dot{z} = G(z, v^{post}) = -\alpha z + \beta v^{post}$. A simple one-step discretization and proper parameter setting yields the exponential moving average equation used in several learning mechanisms, $z_i^t = (1-\lambda) z_i^{t-1} + \lambda v_i^t$ $(0 < \lambda < 1)$ in which $z \equiv \bar{v}^{post}$ [46]. In the Hebbian-style learning scheme Δw_{ji}^{FF} represents the discretized rate of change in the efficacy of the connection weights, η is the learning rate, and the variables $u^{pre} = g(x)$ and $v^{post} = g(y)$ are the firing rates generated by the membrane potentials of the pre- and post-synaptic cells, respectively. The equation employed for learning connection weights along the feedforward projection is stabilized by the difference term $u - z \cdot w$ such that the weight vector is scaled by the (trace of the) postsynaptic activity. It can be shown that the weight adaptation converges to representing a normalized fan-in weight vector, i.e. $\|\boldsymbol{w}|_{t\to\infty}\| = 1$ which eliminates potential biases from the input stimulus energy. In addition, the trace (or temporal low-pass) of the post-synaptic node establishes a smooth graduation of the learning that helps generating categorical representations in model cortex.

We combine the feedforward learning with learning of the feedback connections. Here, the weight adaptation uses a slightly different format

$$\Delta w_{ji}^{FB} = \eta_{FB} \cdot z_i^{pre} \cdot (u_j^{post} - w_{ji}^{FB}) \,, \tag{10}$$

in which the weight in the difference term $u - w$ tends to approach u without further scaling by the post-synaptic activation v, i.e. $\boldsymbol{w}_i^{FB} = \boldsymbol{u}^{post}$. In other words, the structure of the fan-out weights approaches the activity distribution u_j^{post} so that the units with activity $z^{pre} \equiv \bar{v}^{pre}$ memorize an expected pattern of the local driving input signals. The learning constants $\eta_{FF,FB}$ and the duration of the temporal low-pass, with trace length parameter λ, steer the control of establishing neural representations of visual categories. We have empirically observed that a ratio of $\eta_{FF}/\eta_{FB} < 1$ stabilizes the learning such that the representation is invariant against strong variations in the input configuration. We have successfully demonstrated the capabilities of such a learning architecture to build form and motion representations as well as sequence-selective representations for the visual analysis of articulated motion sequences [28].

5. CONCLUSIONS

In this contribution we proposed a novel scheme of building blocks to describe the interaction of superficial and deep level compartments of cortical columns. The modular description at an intermediate level of detail enables the efficient and flexible usage of these building blocks. The decomposition into manageable entities additionally allows its mathematical analysis. By reducing the model to an E-I circuit, we showed conditions of stability and response bifurcations (c.f. [47]), regimes of inhibition stabilized net properties and regimes of multiple stable and unstable fixed points. Finally, we elucidated how online learning mecha-

nisms can be incorporated to allow the adaptation of connection weights for the bottom-up weights to the input compartment as well as those connections that allow signals to propagate backwards (from the deep compartment) to the neuronal representations of the incoming activities.

The proposed model circuit has similarities with the laminar architecture proposed by [15, 39] and the columnar microcircuit model proposed by [45]. Unlike the investigations of Grossberg we emphasize different *compartments* instead of individual layers of cortical structure enabling a compact mathematical description of the dynamics. This let us arrive at the E-I formulation with the qualitative analysis of the network dynamics. The columnar structure and explicit compartmental subdivision of excitatory and inhibitory circuit components corresponds to the work of Diesmann and colleagues [45]. Unlike their work, we put a stronger emphasize on the structural connectivity and different types of interactions (driving vs modulating) in a local population of cells. In addition, the proposed focus on compartments rather than individual layers allowed us to more easily derive an even more abstract dynamical description of the system.

So far, we did not discuss the implications of closing the intra-cortical loop by explicitly incorporating the modulating connections from deep compartment cells to cells in the input compartment. We suggest that such a closed loop (which would incorporate a modulatory feedback connection from the output stage to the filtering stage in the cascade model in Fig. 2(a)) allows to spatially enhance the activations for features at individual locations. In addition, we did not differentiate between different modulating inputs (as indicated in Fig. 1). There is evidence that different cells contact the upper level superficial inputs which originate from different sites in the frontal or from the lateral intraparietal cortex. For example, [45] proposed sharply focused connections to feature selective cells to exert feature-based attention effects. Such different attention signals can be incorporated in the proposed model as well. We suggest that feature attention signals mainly contact cells in the superficial compartment in a spatial and featural neighborhood such that cells with detailed individual feature selectivity can be enhanced. This might then also be the basis for feature-*selective* attention effects in which composite features need to be emphasized or de-emphasized before the grouping into categories occurs [9]. The subsequent normalization of responses further implements the nonlinear effects of activity normalization observed in several studies [7]. More recent investigations have discussed different selectivities in the featural surround inhibition [25]. In the proposed model framework such variants can be studied as well. Lateral interaction in deep compartments can be parameterized to account for these different versions of normalization.

Furthermore, we have suggested that a variant of Hebbian learning as shown in Eq. 10 allows to establish a kernel of weights that represents the expected input pattern that drives the recipient representation. To stabilize the online learning, we utilized a temporal trace mechanism that operates as a local decaying tag to enable the weight adaptation over a given temporal window. Such functional adaptation mechanism might be useful as well when different representations must be linked during the recurrent interaction in a network. In [11] the mechanism of neuronal group selection has been proposed in which functionally segregated maps can be linked through reentrant signals to enhance the correlation of their representations. Synaptic strengthening during learning is thus confronted with the problem to select the proper features and feature compositions that need to influence other item representations at different stages (corresponding to cortical areas). A neural computational framework as the one proposed here serves as a basis for further investigation of such fundamental questions.

6. ACKNOWLEDGMENTS

This work has been supported in part by grants from the Transregional Collaborative Research Center "A Companion-Technology for Cognitive Technical Systems" SFB/TRR62 funded by the German Research Foundation (DFG) and the "SenseEmotion" project funded by the Federal Ministry of Education and Research (BMBF).

7. REFERENCES

[1] H. Barbas and N. Rempel-Clower. Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, 7:635–46, 1997.

[2] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. *Journal of Neuroscience*, 17(6):2112–27, 1997.

[3] T. Brosch and H. Neumann. Computing with a Canonical Neural Circuits Model with Pool Normalization and Modulating Feedback. *Neural Computation*, 26(12):35–89, 2014.

[4] T. Brosch and H. Neumann. Interaction of Feedforward and Feedback Streams in Visual Cortex in a Firing–Rate Model of Columnar Computations. *Neural Networks*, 54:11–6, 2014.

[5] T. Brosch, H. Neumann, and P. R. Roelfsema. Reinforcement learning of linking and tracing contours in recurrent neural networks. *PLoS Computational Biology*, 11(10):e1004489, 2015.

[6] M. Carandini and D. J. Heeger. Summation and Division by Neurons in Primate Visual Cortex. *Science*, 264(5163):1333–6, 1994.

[7] M. Carandini and D. J. Heeger. Normalization as a Canonical Neural Computation. *Nature Reviews Neuroscience*, 13:51–62, 2012.

[8] M. Carandini, D. J. Heeger, and J. A. Movshon. Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. *Journal of Neuroscience*, 17(21):8621–44, 1997.

[9] L. Chelazzi, C. Della Libera, I. Sani, and E. Santandrea. Neural basis of visual selective attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(4):392–407, 2011.

[10] R. Eckhorn. Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Visual Cognition*, 6(3):231–65, 1999.

[11] G. Edelman. Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron*, 10:115–25, 1993.

[12] P. Foeldiak. Learning invariance from transformation sequences. *Neural Computation*, 3:194–200, 1991.

[13] Y. Fregnac, M. Blatow, J.-P. Changeaux, et al. Group report: Neocortical microcircuits - UPs and DOWNs in cortical computation. In S. Grillner and A. Graybiel, editors, *Microcircuits – The interface between neurons and global brain function*, chapter 19, pages 393–433. MIT Press, 2006.

[14] S. Grossberg. Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11:23–63, 1987.

[15] S. Grossberg. How Does the Cerebral Cortex Work? Learning, Attention, and Grouping by the Laminar Circuits of Visual Cortex. *Spatial Vision*, 12:163–85, 1999.

[16] S. Grossberg and E. Mingolla. Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations. *Perception & Psychophysics*, 38(2):141–71, 1985.

[17] D. J. Heeger. Normalization of Cell Responses in Cat Striate Cortex. *Visual Neuroscience*, 9(2):191–97, 1992.

[18] A. V. M. Herz, T. Gollisch, C. K. Machens, and D. Jaeger. Modeling Single–Neuron Dynamics and Computations: A Balance of Detail and Abstraction. *Science*, 314(5796):80–5, 2006.

[19] D. H. Hubel and T. N. Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, 160(1):106–54, 1962.

[20] D. H. Hubel and T. N. Wiesel. Laminar and Columnar Distribution of Geniculo–cortical Fibers in the Macaque Monkey. *Journal Comparative Neurology*, 146:421–50, 1972.

[21] J.-M. Hupe, A. James, P. Girard, et al. Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, 85:134–45, 2001.

[22] M. P. Jadi and T. J. Sejnowski. Regulating Cortical Oscillations in an Inhibition–Stabilized Network. *Proceedings of the IEEE*, 102(5):830–42, 2014.

[23] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer. Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys. *Neuron*, 15(4):843–56, 1995.

[24] M. Kouh and T. Poggio. A Canonical Neural Circuit for Cortical Nonlinear Operations. *Neural Computation*, 20(6):1427–51, 2008.

[25] M. R. Krause and C. C. Pack. Contextual Modulation and Stimulus Selectivity in Extrastriate Cortex. *Vision Research*, 104:36–46, 2014.

[26] V. A. F. Lamme and P. R. Roelfsema. The Distinct Modes of Vision Offered by Feedforward and Recurrent Processing. *Trends in Neurosciences*, 23(11):571–9, 2000.

[27] M. Larkum. A Cellular Mechanism for Cortical Associations: An Organizing Principle for the Cerebral Cortex. *Trends in Neurosciences*, 36(3):141–51, 2013.

[28] G. Layher, M. Giese, and H. Neumann. Learning Representations of Animated Motion Sequences – A Neural Model. *Topics in Cognitive Science*, 6(1):170–82, 2014.

[29] Y. LeCun, B. Boser, J. S. Denker, et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4):541–51, 1989.

[30] C. C. Lee and S. M. Sherman. Modulator Property of the Intrinsic Cortical Projection from Layer 6 to Layer 4. *Frontiers in Systems Neuroscience*, 3(3):1–5, 2009.

[31] W. Li, V. Piech, and C. Gilbert. Contour saliency in primary visual cortex. *Neuron*, 50:951–62, 2006.

[32] V. B. Mountcastle. The Columnar Organization of the Neocortex. *Brain*, 120(4):701–22, 1997.

[33] H. Neumann and W. Sepp. Recurrent V1–V2 Interaction in Early Visual Boundary Processing. *Biological Cybernetics*, 81(5-6):425–444, 1999.

[34] H. Nothdurft. Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31(6):1073–8, 1991.

[35] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster. Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. *Neuron*, 62(4):578–92, 2009.

[36] A. Packer and R. Yuste. Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *Journal of Neuroscience*, 31(37):13260–71, 2011.

[37] W. A. Phillips, A. Clark, and S. M. Silverstein. On the Functions, Mechanisms, and Malfunctions of Intracortial Contextual Modulation. *Neuroscience and Biobehavioral Reviews*, 52:1–20, 2015.

[38] N. C. Rabinowitz, B. D. B. Willmore, J. W. H. Schnupp, and A. J. King. Contrast Gain Control in Auditory Cortex. *Neuron*, 70(6):1178–91, 2011.

[39] R. D. S. Raizada and S. Grossberg. Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cerebral Cortex*, 13:100–13, 2003.

[40] J. H. Reynolds and D. J. Heeger. The Normalization Model of Attention. *Neuron*, 61:168–85, 2009.

[41] M. Self, T. van Kerkoerle, H. Supèr, and P. Roelfsema. Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Current Biology*, 23:2121–9, 2013.

[42] M. W. Spratling. Cortical Region Interactions and the Functional Role of Apical Dendrites. *Behavioral and Cognitive Neuroscience Reviews*, 1(3):219–28, 2002.

[43] A. Thielscher and H. Neumann. Neural mechanisms of human texture processing: texture boundary detection and visual search. *Spatial Vision*, 18(2):227–57, 2003.

[44] S. Ullman. Sequence Seeking and Counter Streams: A Computational Model for Bidirectional Information Flow in the Visual Cortex. *Cerebral Cortex*, 5(1):1–11, 1995.

[45] N. Wagatsuma, T. Potjans, M. Diesmann, and T. Fukai. Spatial and feature-based attention in a layered cortical microcircuit model. *PLoS ONE*, 8(12):e80788, 2013.

[46] G. Wallis and E. Rolls. Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2):167–194, 1997.

[47] L. Zhaoping. Neural Circuit Models for Computations in Early Visual Cortex. *Current Opinion in Neurobiology*, 21(5):808–15, 2011.