

Chapter 7

Constructing Complex Systems Via Activity-Driven Unsupervised Hebbian Self-Organization

James A. Bednar

Abstract How can an information processing system as complex and as powerful as the human cerebral cortex be constructed from the limited information available in the genome? Answering this scientific question has the potential to revolutionize how computing systems for manipulating real-world data are designed and built. Based on an extensive array of physiological, anatomical, and imaging data from the primary visual cortex (V1) of mammals, we propose a relatively simple biologically based developmental architecture that accounts for most of the demonstrated functional properties of V1 neurons. Given the overall similarity between cortical regions, and the absence of V1-specific circuitry in the model architecture, we expect similar principles to apply throughout the cerebral cortex. The architecture consists of a network of simple artificial V1 neurons with initially unspecific connections that are modified by Hebbian learning and homeostatic plasticity, driven by input patterns from other neural regions and ultimately from the external world. Through an unsupervised developmental process, the model neurons begin to display the major known functional properties of V1 neurons, including receptive fields and topographic maps selective for all of the major low-level visual feature dimensions, realistic specific lateral connectivity underlying surround modulation and adaptation such as visual aftereffects, realistic behavior with visual contrast, and realistic temporal responses. In each case these relatively complex properties emerge from interactions between simple neurons and between internal and external drivers for neural activity, without any requirement for supervised learning, top-down feedback or reinforcement, neuromodulation, or spike-timing dependent plasticity. The model also unifies explanations of a wide variety of phenomena previously considered distinct, with the same adaptation mechanisms leading to both long-term development and short-term plasticity (aftereffects), the same subcortical lateral interactions providing both gain control and accounting for the time course of neural responses,

J. A. Bednar (✉)

Institute for Adaptive and Neural Computation, The University of Edinburgh,
10 Crichton St, Edinburgh EH8 9AB, UK
e-mail: jbednar@inf.ed.ac.uk

and the same cortical lateral interactions leading to complex cell properties, map formation, and surround modulation. This relatively simple architecture thus sets a baseline for explanations of neural function, suggesting that most of the development and function of V1 can be understood as unsupervised learning, and setting the stage for demonstrating the additional effects of higher- or lower-level mechanisms. The architecture also represents a simple, scalable approach for specifying complex data-processing systems in general.

1 Introduction

Current technologies for building complex information processing machines, particularly for dealing with continuous and noisy real-world data, remain far behind those of animals and particularly humans. Artificial neural networks originally inspired by biological nervous systems are in wide use for some tasks, and indeed have been shown to be able to perform any Turing-computable function in theory [48]. However, actually specifying and constructing a network capable of performing a particular complex task remains an open problem. With this in mind, it is important to study how such networks are specified and developed in animals and humans, to give us clues for how to build similarly powerful artificial systems. Understanding how systems as complex as the human brain can be built has the potential to revolutionize how computing systems for manipulating real-world data are designed and constructed.

The cerebral cortex of mammals is a natural starting point for study, since the cortex is the largest part of the human brain and serves a wide variety of sensory and motor functions in mammals, yet has a relatively uniform structure. The cortical surface can be divided into anatomically distinguishable cortical areas, of which there are dozens in humans, but perhaps the most commonly studied is the primary visual cortex (V1). After processing by circuitry in the retina, visual information travels from the Retinal Ganglion Cells (RGCs) of the eye to the lateral geniculate nucleus (LGN) of the thalamus, and from the thalamus goes directly to cells in V1. This simple, direct pathway, along with the unparalleled ease with which visual patterns can be controlled in the laboratory, means that V1 provides a unique opportunity for running well-controlled experiments for determining neural function.

Like the rest of the cerebral cortex, V1 has a laminar and columnar structure, i.e., it is composed of multiple thin layers (typically numbered 1–6) parallel to the cortical surface, with neurons at corresponding locations in each layer forming “columns” that have similar functional properties, while more distant columns tend to differ in their properties. This predominantly two-dimensional functional organization is often measured using experimental imaging techniques that record activity across the cortical surface in response to a visual image, which obscures any differences in neural activity across the layers but does provide a useful large-scale measure of neural organization. Complementary information about visual responses of individual neurons or small groups (typically measured as firing rates, i.e., spikes per

second) can be measured using microelectrodes placed nearby, providing detailed temporal responses but necessarily sampling from only a few neurons.

Studies using these techniques in monkeys, cats, ferrets, and tree shrews over the past half century have established a wide range of properties of V1 neurons that relate to their function in visual processing (reviewed in [7, 31]):

1. V1 neurons respond selectively, in terms of their average firing rate, to specific low-level visual features such as the position, orientation, eye of origin, motion direction, spatial frequency, interocular disparity, or color of a small patch of an image.
2. V1 neurons in most laboratory-animal species are organized into smooth topographic maps for some or all of these visual features, with specific patterns of feature preference variation (e.g. in orientation preference) across the cortical surface, and specific interactions between these maps.
3. V1 neurons in these maps are laterally connected with connection strengths and probabilities that reflect their selectivities (e.g. with stronger connections between neurons preferring similar orientations).
4. Due in part to these lateral connections, V1 neuronal responses depend on activities of both neighboring and more distant V1 neurons, yielding complex but systematic visual surround modulation effects.
5. V1 neurons exhibit contrast-invariant tuning for the features for which they are selective, such that selectivity is preserved even for strong input patterns. This property rules out most simple (linear) models of visual feature selectivity.
6. Many V1 neurons have complex spatial pattern preferences that cannot be characterized using a simple template of their preferred pattern, e.g. responding to similar patterns with some tolerance to the exact retinal position of the pattern.
7. Response properties of V1 neurons exhibit long-term and short-term plasticity and adaptation, measurable psychophysically as visual aftereffects, which suggests ongoing dynamic regulation of responses.
8. V1 neuron responses to changes in visual inputs exhibit a stereotyped temporal pattern, with transiently high responses at pattern onset and offset and a lower sustained response, which biases neural responses towards non-static stimuli.

These properties suggest a wide range of specific roles for V1 neurons in visual processing, and explaining how V1 neurons come to behave in this way would be a significant step towards understanding the cerebral cortex in general.

One way to account for the above V1 properties would be to build specific mathematical models for each property or a small set of them, and indeed many such models have been devised to account for aspects of V1 visual processing in adult animals. However, such an approach is unlikely to lead to a general explanation for cortical function, both because such models cannot easily be combined into a full visual processing system, and also because the models are specifically tailored by the modeller to account for that particular function, and thus do not generalize to arbitrary data processing tasks.

This chapter outlines and reviews results from an alternative “system building” approach, focusing on providing a general domain-independent explanation for how

all of the above properties of V1 could arise from a biologically plausible and initially unspecific cortical circuit. Specifically, my colleagues and I have developed closely interrelated models of V1 accounting for each property using only a small set of plausible principles and mechanisms, within a consistent biologically grounded framework:

1. Single-compartment (point neuron) firing-rate (i.e., non-spiking) retinal ganglion cell, lateral geniculate nucleus, and V1 model neurons (see Fig. 1),
2. Hardwired subcortical pathways to V1, including the main LGN or RGC cell types that have been identified,
3. Initially roughly retinotopic topographic projections from the eye to the LGN and from the LGN to V1, connecting corresponding areas of each region,
4. Initially roughly isotropic (i.e., radially uniform) local connectivity to and between neurons in layers in V1, connecting neurons non-specifically to their local and more distant neighbors,
5. Natural images and spontaneous subcortical input activity patterns that lead to V1 responses,
6. Hebbian (unsupervised activity-dependent) learning with normalization for synapses on V1 neurons,
7. Homeostatic plasticity (whole-cell adaptation of excitability to keep the average activity of each V1 neuron constant), and
8. Various modeller-determined parameters associated with each of these mechanisms, eventually intended to be set through self-regulating mechanisms.

Properties and mechanisms not necessary to explain the phenomena listed above, such as spiking, spike-timing dependent plasticity, detailed neuronal morphology, feedback from higher areas, neuromodulation, reinforcement learning, and supervised learning have been omitted, to clearly focus on the aspects of the system most relevant to those phenomena. The overall hypothesis is that much of the complex structure and properties observed in V1 emerges from interactions between relatively simple but highly interconnected computing elements, with connection strengths and patterns self-organizing in response to visual input and other sources of neural activity. Through visual experience, the geometry and statistical regularities of the visual world become encoded into the structure and connectivity of the visual cortex, leading to a complex functional cortical architecture that reflects the physical and statistical properties of the visual world.

At present, many of the results have been obtained independently in a wide variety of separate projects performed with different collaborators at different times. However, all of the models share the same underlying principles outlined above, and all are implemented using the same simulator and a small number of underlying components. See [7] for an overview of each of the different models and how they fit together; here we focus on two representative models that account for the bulk of the above properties. First, we present a simple example of a single-V1-layer GCAL (gain-control adaptive laterally connected) model of the development of orientation preferences and orientation maps for a single eye (Fig. 1). Second, we present some results for a larger model that includes motion direction and ocular dominance as

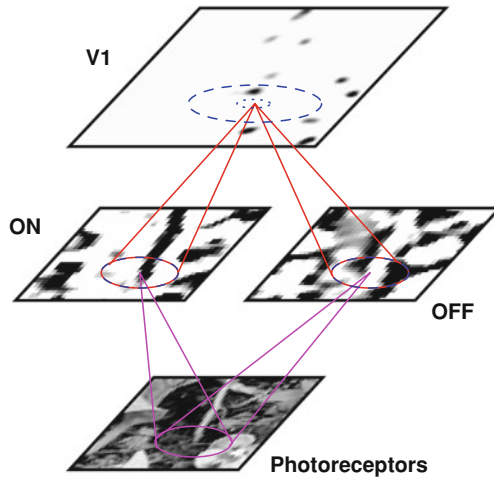


Fig. 1 *Basic GCAL model architecture.* In the simplest case, GCAL consists of a *greyscale* matrix representing the photoreceptor input, a pair of neural sheets representing the ON-center and OFF-center pathways from the photoreceptors to V1, and a single sheet representing V1. Each sheet is drawn here with a sample activity pattern resulting from one natural image patch. Each projection between sheets is illustrated with an oval showing the extent of the connection field in that projection, with lines converging on the target of the projection. Lateral projections, connecting neurons within each sheet, are marked with *dashed* ovals. Projections from the photoreceptors to the ON and OFF sheets, and within those sheets, are hardwired to mimic a specific class of response types found in the retina and LGN, in this case monochromatic center-surround neurons with a fixed spatial extent. Connections to and between V1 neurons adapt via Hebbian learning, allowing initially unselective V1 neurons to exhibit the range of response types seen experimentally, by differentially weighting each of the subcortical inputs (from the ON and OFF sheets) and inputs from neighboring V1 neurons

well (Fig. 2). Results for color, disparity, spatial frequency, complex cells, and surround modulation require still larger models not discussed here, but implemented using similar principles [3, 5, 7, 38, 40]. The goal for each of these models is the same—to explain how a cortical network can start from an initially undifferentiated state, to wire itself into a collection of neurons that behave, at a first approximation, like those in V1. Because such a model starts with no specializations (at the cortical level) specific to vision and would organize very differently when given different inputs, it also represents a general explanation for the development and function of any sensory or motor area in the cortex.

2 Architecture

All of the models whose results are presented here are implemented in the Topographica simulator, and are freely available along with the simulator from www.topographica.org. Both the basic and large models are described using the same

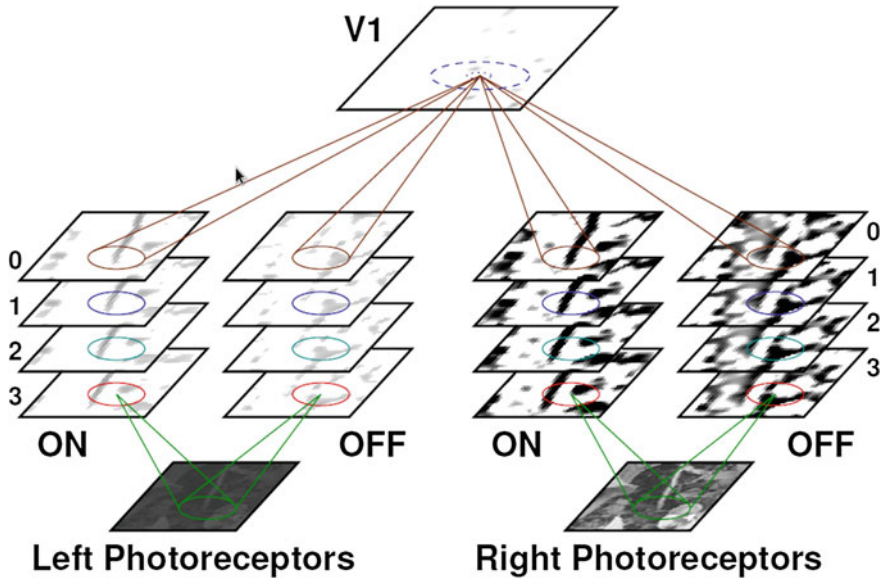


Fig. 2 Larger GCAL model architecture. Single-V1-sheet GCAL model for orientation, ocular dominance, and motion direction. This model consists of 19 neural sheets and 50 separate projections between them. Again, each sheet is drawn with a sample activity pattern resulting from one natural image, and all sheets below V1 are hardwired to cover a range of response types found in the retina and LGN. In this model, V1 neurons can become selective along many possible dimensions by various weightings for the incoming activity patterns. Other GCAL models add additional subcortical inputs (e.g. for color cone types) or additional populations and layers in V1. Reprinted from [13]

equations shown below, previously presented in Refs. [7, 29] but here extended to include temporal calibration from the TCAL model [51]. The model is intended to represent the visual system of the macaque monkey, but relies on data from studies of cats, ferrets, tree shrews, or other mammalian species where clear results are not yet available from monkeys.

2.1 Sheets and Projections

Each Topographica model consists of a set of sheets of neurons and projections (sets of topographically mapped connections) between them. A model has sheets representing the visual input (as a set of activations in photoreceptor cells), sheets implementing the transformation from the photoreceptors to inputs driving V1 (expressed as a set of ON and OFF LGN cell activations), and sheets representing neurons in V1. The simple GCAL model (Fig. 1) has 4 such sheets, the larger one (Fig. 2) has 19, and the complete unified model described in [7] has 29, each representing different topographically organized populations of cells in a particular region.

Each sheet is implemented as a two-dimensional array of firing-rate neurons. The Topographica simulator allows parameters for sheets and projections to be specified in measurement units that are independent of the specific grid sizes used in a particular run of the simulation. To achieve this, Topographica sheets provide multiple spatial coordinate systems, called *sheet* and *matrix* coordinates. Where possible, parameters (e.g. sheet dimensions or connection radii) are expressed in sheet coordinates, expressed as if the sheet were a continuous neural field rather than a finite grid. In practice, of course, sheets are always simulated using some finite matrix of units. Each sheet has a parameter called its density, which specifies how many units (matrix elements) in the matrix correspond to a length of 1.0 in continuous sheet coordinates, which allows conversion between sheet and matrix coordinates. When sizes are scaled appropriately [8], results are independent of the density used, except at very low densities or for simulations with complex cells, where complexity increases with density [4]. Larger areas can be simulated easily [8], but require more memory and simulation time.

A projection to an $m \times m$ sheet of neurons consists of m^2 separate *connection fields*, one per target neuron, each of which is a spatially localized set of connections from the neurons in one input sheet that are near the location corresponding topographically to the target neuron. Figures 1 and 2 show one sample connection field (CF) for each projection, visualized as an oval of the corresponding radius on the input sheet (drawn to scale), connected by a cone to the neuron on the target sheet (if different). The connections and their weights determine the specific properties of each neuron in the network, by differentially weighting inputs from neurons of different types and/or spatial locations. Each of the specific types of sheets and projections is described in the following sections.

2.2 Images and Photoreceptor Sheets

The basic GCAL model (Fig. 1) has one input sheet, representing responses of photoreceptors of one cone class in one retina, while the larger GCAL model considered here has two, adding an additional set from another eye (Fig. 2). The full unified GCAL model of all the input dimensions includes six input sheets (three different cone types in each eye; not shown or analyzed further here). For the larger model, input image pairs (left, right) were generated by choosing one image randomly from a database of single calibrated images, selecting a random patch within the image, a random direction of motion translation with a fixed speed (described in Ref. [10]), and a random brightness difference between the two eyes (described in Ref. [31]). These modifications are intended as a simple model of motion and eye differences, to allow development of direction preference, ocular dominance, and disparity maps, until suitable full-motion stereo calibrated-color video datasets of natural scenes are available. Simulated retinal waves can also be used as inputs, to provide initial receptive-field and map structure before eye opening, but are not required for receptive-field or map development in the model [11].

2.3 *Subcortical Sheets*

The subcortical pathway from the photoreceptors to the LGN and then to V1 is represented as a set of hardwired subcortical cells with fixed connection fields (CFs) that determine the response properties of each cell. These cells represent the complete processing pathway to V1, including circuitry in the retina (including the retinal ganglion cells), the optic nerve, the lateral geniculate nucleus, and the optic radiations to V1. Because the focus of the model is to explain cortical development given its thalamic input, the properties of these RGC/LGN cells are kept fixed throughout development, for simplicity and clarity of analysis.

Each distinct RGC/LGN cell type is grouped into a separate sheet, each of which contains a topographically organized set of cells with identical properties but responding to a different topographically mapped region of the retinal photoreceptor input sheet. Figure 1 shows the two main different spatial response types used in the GCAL models illustrated here, ON (with an excitatory center) and OFF (with an excitatory surround). All of these cells have Difference-of-Gaussian (DoG) receptive fields, and thus perform edge enhancement at a particular size scale. Additional cell classes can easily be added as needed for spatial frequency (with multiple DoG sizes) or color (with separate cone types for the center and surround Gaussians) simulations. Figure 2 shows additional ON and OFF cell classes with different delays, added to allow development of motion preferences.

For the ON and OFF cells in the larger model, there are multiple copies with different delays from the retina. These delays represent the different latencies in the lagged versus non-lagged cells found in cat LGN [44, 59], and allow V1 neurons to become selective for the direction of motion. Many other sources of temporal delays would also lead to direction preferences, but have not been tested specifically.

2.4 *Cortical Sheets*

The simulations reported in this chapter use only a single V1 sheet for simplicity, but in the full unified model, V1 is represented by multiple cortical sheets representing different cell types and different V1 layers [3, 7]. In this simplified version, cells make both excitatory and inhibitory connections (unlike actual V1 neurons), and all cells receive direct input from LGN cells (unlike many V1 neurons). Even so, the single-sheet V1 can demonstrate most of the phenomena described above, except for complex cells (which can be obtained by adding a separate population of cells without direct thalamic input [4]) and contrast-dependent surround modulation effects (which require separate populations of inhibitory and excitatory cells [3]).

The behavior of the cortical sheet is primarily determined by the projections to and within it. Each of these projections is initially non-specific other than the initial rough topography, and becomes selective only through the process of self-organization (described below), which increases some connection weights at the expense of others.

2.5 Activation

The model is simulated in a series of discrete time steps with step size $\delta t = 0.05$ (roughly corresponding to 12.5 ms of real time). At time 0.0, the first image is drawn on the retina, and the activation of each unit in each sheet is updated for the remaining 19 steps before time 1.0, when a new pattern is selected and drawn on the retina (and similarly until the last input pattern is drawn at time 10,000). Each image patch on the retina represents one visual fixation (for natural images) or a snapshot of the relatively slowly changing spatial pattern of spontaneous activity (such as the well-documented retinal waves [60]). Thus the training process consists of a constant retinal activation, followed by recurrent processing at the LGN and cortical levels. For one input pattern, assume that the input is drawn on the photoreceptors at time t and the connection delay (constant for all projections) is defined as 0.05. Then at $t + 0.05$ the ON and OFF cells compute their responses, and at $t + 0.10$ the thalamic output is delivered to V1, where it similarly propagates recurrently through the intracortical projections to the cortical sheet(s) every 0.05 time steps. As described in Sect. 3.4, a much smaller step size of $\delta t = 0.002$ allows replication of the detailed time course of responses to individual patterns, but this relatively coarse step size of 0.05 is more practical for simulations of long-term processes like neural development.

Images are presented to the model by activating the retinal photoreceptor units. The activation value $\Psi_{i,p}$ of unit i in photoreceptor sheet P is given by the brightness of that pixel in the training image.

For each model neuron in the other sheets, the activation value is computed based on the combined *activity contributions* to that neuron from each of the sheet's incoming projections. The activity contribution from a projection is recalculated whenever its input sheet activity changes, after the corresponding connection delay. For each unit j in a target sheet and an incoming projection p from sheet s_p , the activity contribution is computed from activations in the corresponding connection field F_{jp} . F_{jp} consists of the local neighborhood around j (for lateral connections), or the local neighborhood of the topographically mapped location of j on s_p (for a projection from another sheet); see examples in Figs. 1 and 2. The activity contribution C_{jp} to j from projection p is then a dot product of the relevant input with the weights in each connection field:

$$C_{jp}(t + \delta t) = \sum_{i \in F_{jp}} \eta_i(t) \omega_{ij,p} \quad (1)$$

where unit i is taken from the connection field F_{jp} of unit j , $\eta_i(t)$ is the activation of unit i , and $\omega_{ij,p}$ is the connection weight from i to j in that projection. Across all projections, multiple direct connections between the same pair of neurons are possible, but each projection p contains at most one connection between i and j , denoted by $\omega_{ij,p}$.

For a given cortical unit j , the activity $\eta_j(t + \delta t)$ is calculated from a rectified weighted sum of the activity contributions $C_{jp}(t + \delta t)$:

$$\eta_{jV}(t + \delta t) = \lambda f \left(\sum_p \gamma_p C_{jp}(t + \delta t) \right) + (1 - \lambda) \eta_{jV}(t) \quad (2)$$

where f is a half-wave rectifying function with a variable threshold point (θ) dependent on the average activity of the unit, as described in the next subsection, and V denotes one of the cortical sheets. λ is a time-constant parameter that defines the strength of smoothing of the recurrent dynamics in the network, chosen to match the transient behaviour of V1 neurons; here $\lambda = 1$ throughout except that $\lambda = 0.01$ for the simulations of the detailed time course of responses (Sect. 3.4).

Each γ_p is an arbitrary multiplier for the overall strength of connections in projection p . The γ_p values are set in the approximate range 0.5–3.0 for excitatory projections and -0.5 to -3.0 for inhibitory projections. For afferent connections, the γ_p value is chosen to map average V1 activation levels into the range 0–1.0 by convention, for ease of interconnecting and analyzing multiple sheets. For lateral and feedback connections, the γ_p values are then chosen to provide a balance between feedforward, lateral, and feedback drive, and between excitation and inhibition; these parameters are crucial for making the network operate in a useful regime.

RGC/LGN neuron activity is computed similarly to Eq. 2, except to add divisive normalization and to fix the threshold θ at zero:

$$\eta_{jL}(t + \delta t) = \lambda f \left(\frac{\sum_p \gamma_p C_{jp}(t + \delta t)}{\gamma_S C_{jS}(t + \delta t) + k} \right) + (1 - \lambda) \eta_{jL}(t) \quad (3)$$

where L stands for one of the RGC/LGN sheets. Projection S here consists of a set of isotropic Gaussian-shaped lateral inhibitory connections (see Eq. 6, evaluated with $u = 1$), and p ranges over all the other projections to that sheet. k is a small constant to make the output well-defined for weak inputs. The divisive inhibition implements the contrast gain control mechanisms found in RGC and LGN neurons [2, 3, 17, 23]. Here again $\lambda = 1$ throughout except that $\lambda = 0.03$ for the detailed simulations in Sect. 3.4.

Each time the activity is computed using Eq. 2 or 3, the new activity values are sent to each of the outgoing projections, where they arrive after the projection delay. The process of activity computation then begins again, with a new contribution C computed as in Eq. 1, leading to new activation values by Eq. 2 or 3. Activity thus spreads recurrently throughout the network, and can change, die out, or be strengthened, depending on the parameters.

With typical parameters that lead to realistic topographic map patterns, initially blurry patterns of afferent-driven cortical activity are sharpened into well-defined “activity bubbles” through locally cooperative and more distantly competitive lateral interactions [31]. Nearby neurons are thus influenced to respond more similarly, while more distant neurons receive net inhibition and thus learn to respond to different input patterns. The competitive interactions “sparsify” the cortical response into

patches, in a process that can be compared to the explicit sparseness constraints in non-mechanistic models [26, 36], while the local facilitatory interactions encourage spatial locality so that smooth topographic maps will be developed.

As described in more detail below, the initially random weights to cortical neurons are updated in response to each input pattern, via Hebbian learning. Because the settling (sparsification) process typically leaves only small patches of the cortical neurons responding strongly, those neurons will be the ones that learn the current input pattern, while other nearby neurons will learn other input patterns, eventually covering the complete range of typical input variation. Overall, through a combination of the network dynamics that achieve sparsification along with local similarity, plus homeostatic adaptation that keeps neurons operating in a useful regime, plus Hebbian learning that leads to feature preferences, the network will learn smooth, topographic maps with good coverage of the space of input patterns, thereby developing into a functioning system for processing patterns of visual input without explicit specification or top-down control of this process.

2.6 Homeostatic Adaptation

For this model, the threshold for activation of each cortical neuron is a very important quantity, because it directly determines how much the neuron will fire in response to a given input. Mammalian neurons appear to regulate such thresholds automatically, a process known as homeostatic plasticity or homeostatic adaptation [54] (where homeostatic means to keep similar and stable). To set the threshold automatically, each neural unit j in V1 calculates a smoothed exponential average of its own activity ($\bar{\eta}_j$):

$$\bar{\eta}_j(t) = (1 - \beta)\eta_j(t) + \beta\bar{\eta}_j(t - 1) \quad (4)$$

The smoothing parameter ($\beta = 0.999$) determines the degree of smoothing in the calculation of the average. $\bar{\eta}_j$ is initialized to the target average V1 unit activity (μ), which for all simulations is $\bar{\eta}_{jA}(0) = \mu = 0.024$. The threshold is updated as follows:

$$\theta(t) = \theta(t - 1) + \kappa(\bar{\eta}_j(t) - \mu) \quad (5)$$

where $\kappa = 0.0001$ is the homeostatic learning rate. The effect of this scaling mechanism is to bring the average activity of each V1 unit closer to the specified target. If the average activity of a V1 unit moves away from the target during training, the threshold for activation is thus automatically raised or lowered in order to bring it closer to the target.

2.7 Learning

Initial connection field weights are random within a two-dimensional Gaussian envelope. E.g., for a postsynaptic (target) neuron j located at sheet coordinate $(0, 0)$, the weight $\omega_{ij,p}$ from presynaptic unit i in projection p is:

$$\omega_{ij,p} = \frac{1}{Z_{\omega p}} u \exp\left(-\frac{x^2 + y^2}{2\sigma_p^2}\right) \quad (6)$$

where (x, y) is the sheet-coordinate location of the presynaptic neuron i , u is a scalar value drawn from a uniform random distribution for the afferent and lateral inhibitory projections ($p = A, I$), σ_p determines the width of the Gaussian in sheet coordinates, and Z_{ω} is a constant normalizing term that ensures that the total of all weights $\omega_{ij,p}$ to neuron j in projection p is 1.0, where all afferent projections are treated together as a single projection so that their sum total is 1.0. Weights for each projection are only defined within a specific maximum circular radius r_p ; they are considered zero outside that radius.

Once per input pattern (after activity has settled), each connection weight ω_{ij} from unit i to unit j is adjusted using a simple Hebbian learning rule. (Learning could instead be performed at every simulation time step, but doing so would require significantly more computation time). This rule results in connections that reflect correlations between the presynaptic activity and the postsynaptic response. Hebbian connection weight adjustment for unit j is dependent on the presynaptic activity η_i , the post-synaptic response η_j , and the Hebbian learning rate α :

$$\omega_{ij,p}(t) = \frac{\omega_{ij,p}(t-1) + \alpha \eta_j \eta_i}{\sum_k (\omega_{kj,p}(t-1) + \alpha \eta_j \eta_k)} \quad (7)$$

Unless it is constrained, Hebbian learning will lead to ever-increasing (and thus unstable) values of the weights. The weights are constrained using divisive post-synaptic weight normalization (denominator of Eq. 7), which is a simple and well understood mechanism. All afferent connection weights from RGC/LGN sheets are normalized together in the model, which allows V1 neurons to become selective for any subset of the RGC/LGN inputs. Weights are normalized separately for each of the other projections, to ensure that Hebbian learning does not disrupt the balance between feedforward drive, lateral and feedback excitation, and lateral and feedback inhibition. Subtractive normalization with upper and lower bounds could be used instead, but it would lead to binary weights [32, 33], which is not desirable for a firing-rate model whose connections represent averages over multiple physical connections. More biologically motivated homeostatic mechanisms for normalization such as multiplicative synaptic scaling [54] or a sliding threshold for plasticity [15] could be implemented instead, but these have not been tested so far.

Note that some of the results below use the earlier LISSOM model [31], which follows the same equations but lacks gain control and homeostatic adaptation (equivalent

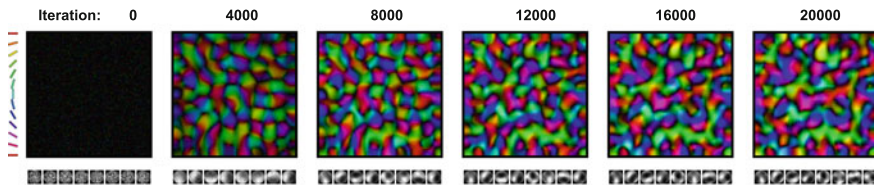


Fig. 3 *Development of maps and afferent connections.* Over the course of 20,000 input presentations, GCAL model V1 neurons develop selectivity for typical features of the input patterns. Here simulated retinal waves were presented for the first 6,000 inputs (modelling prenatal development), and monochromatic images of natural scenes were presented for the remainder (modelling postnatal visual experience). Connection fields to V1 neurons were initially random and isotropic (*bottom* of Iteration 0; CFs for 8 sample neurons are shown). Neurons were initially unselective, responding approximately equally to all orientations, and are thus *black* in the orientation map plot (where saturated colors represent orientation-selective neurons whose preference is labeled with the *color* indicated in the key). Over time, neurons develop specific afferent connection fields (*bottom* of remaining iterations) that cause neurons to respond to specific orientations. Nearby neurons respond to similar orientations, as in animal maps, and as a whole they eventually represent the full range of orientations present in the inputs. Reprinted from [29]

to setting $\gamma_S = 0$ and $k = 1$ in Eq. 3 and $\kappa = 0$ in Eq. 5). Without these automatic mechanisms, LISSOM requires the modeller to set the input strength and activation thresholds separately for each dataset and to adjust them as learning progresses. As long as these values have been set appropriately, previous LISSOM results can be treated equivalently to GCAL results, but GCAL is significantly simpler to use and describe, while being more robust to changes in the input distributions [29], so only GCAL is described here.

3 Results

The following sections outline a series of model results that account for anatomical, electrophysiology, imaging, psychophysical, and behavioral results from studies of experimental animals, all arising from the neural architecture and self-organizing mechanisms outlined in the previous section.

3.1 Maps and Connection Patterns

Figure 3 shows how orientation selectivity emerges in the basic GCAL model from Fig. 1, whose subcortical pathway consists of a single set of non-lagged monochromatic ON and OFF LGN inputs for a single eye. Over the course of development, initially unspecific connections become selective for specific patterns of LGN activity, including particular orientations. Hebbian learning ensures that each afferent

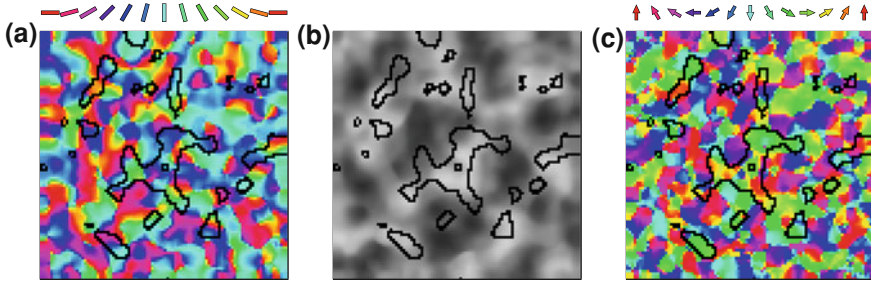


Fig. 4 *Lateral connections across maps.* GCAL and LISSOM model neurons each participate in multiple functional maps, but have only a single set of lateral connections. Connections are strongest from other neurons with similar properties, respecting each of the maps to the degree to which that map affects correlation between neurons. Maps for a combined orientation (OR), ocular dominance (OD), direction (DR) simulation LISSOM model are shown above, with the *black* outlines indicating the connections to the central neuron (marked with a small *square black* outline) that remain after weak connections have been pruned. Model neurons connect to other model neurons with similar orientation preference (a) (as in tree shrew, [18]) but even more strongly respect the direction map (c). This highly monocular unit also connects strongly to the same eye (b), but the more typical binocular cells have wider connection distributions. Reprinted from Ref. [13]

connection field shown represents the average pattern of LGN activity that has driven that neuron to a strong response; each neuron prefers a different pattern at a specific location on the retinal surface. Preferences from the set of all V1 neurons form a smooth topographic map covering the range of orientations present in the input patterns, yielding an orientation map similar to those from monkeys [16]. For instance, the map shows iso-feature domains, pinwheel centers, fractures, saddle points, and linear zones, with a ring-shaped Fourier transform. As in animals [46], orientation selectivity is preserved over a very wide range of contrasts, due to the effect of lateral inhibitory connections in the LGN and in V1 that normalize responses to be relative to activation of neighboring neurons rather than absolute levels of contrast [29].

Similar results are found for models including each of the other low-level features of images, with specific map patterns that match those found in animals. Figure 4 shows results from the larger orientation, ocular dominance, and motion direction simulation from Fig. 2; each neuron becomes selective for some portion of this multidimensional feature space, and together they account for the variation across this space that was seen during self-organization [13]. Other simulations not included here show how color, spatial frequency, and disparity preferences and maps can develop when appropriate information is made available to V1 through additional RGC/LGN sheets [5, 7, 38, 40]. As described in the original source for each model, the model results for each dimension have been evaluated against the available animal data, and capture the main aspects of the feature value coverage and the spatial organization of the maps [31, 38]. The maps simulated together (e.g. orientation and ocular dominance) also tend to intersect at right angles, such that high-gradient regions in one map avoid high-gradient regions in others [13].

These patterns primarily emerge from geometric constraints on smoothly mapping the range of values for the indicated feature, within a two-dimensional retinotopic map [31]. They are also affected by the relative amount by which each feature varies in the input dataset, how often each feature appears, and other aspects of the input image statistics [12]. For instance, orientation maps trained on natural image inputs develop a preponderance of neurons with horizontal and vertical orientation preferences, which is also seen in ferret maps and reflects the statistics of contours found in natural images [11, 20].

While the feature maps and afferent connections of neurons primarily represent a decomposition of the image into commonly recurring local features, lateral connections between these neurons store patterns of correlation between each neuron that represent larger-scale structure and correlations. Figure 4 shows the pattern of lateral connectivity for a neuron embedded in an orientation, ocular dominance, and motion direction map. Because the lateral connections are also modified by Hebbian learning, they represent correlations between neurons, and are thus strong for short-range connections (due to the shared retinotopic preference of those neurons) and between other neurons often coactivated during self-organization (e.g. those sharing orientation, direction, and eye preferences). The lateral connections are thus patchy and orientation and direction specific, as found in animals [18, 43, 50]. Neurons with low levels of selectivity for any of those dimensions (e.g. binocular neurons) receive connections from a wide range of feature preferences, while highly selective neurons receive more specific connections, reflecting the different patterns of correlation in those cases. These connection patterns represent predictions, as only a few of these relationships have been tested so far in animals. The model strongly predicts that lateral connection patterns will respect all maps that account for a significant fraction of the response variance of the neurons, because each of those features will affect the correlation between neurons.

Overall, where it has been possible to make comparisons, these models have been shown to reproduce the main features of the experimental data, using a small set of assumptions. In each case, the model demonstrates how the experimentally measured map can emerge from Hebbian learning of corresponding patterns of subcortical and cortical activity. The models thus illustrate how the same basic, general-purpose adaptive mechanism will lead to very different organizations, depending on the geometrical and statistical properties of that feature. Future work will focus on showing how all the results so far could emerge simultaneously in a single model (as outlined in Ref. [7]).

3.2 Surround Modulation

Given a model with realistically patchy, specific lateral connectivity and realistic single-neuron properties, as described above, the patterns of interaction between neurons can be compared with neurophysiological evidence for surround modulation— influences on neural responses from distant patterns in the visual field. These studies

can help validate the underlying model circuit, while helping understand how the visual cortex will respond to complicated patterns such as natural images.

For instance, as the size of a patch of grating is increased, the response of a V1 neuron typically increases at first, reaches a peak, and then decreases [45, 47, 55]. Similar patterns can be observed in a GCAL-based model orientation map with complex cells and separate inhibitory and excitatory subpopulations (figure from Ref. [3]; not shown). Small patterns initially activate neurons weakly, due to low overlap with the afferent receptive fields of layer 4 cells, but the response increases with larger patterns. For large enough patterns, lateral interactions are strong and in most locations net inhibitory, causing many neurons to be suppressed (leading to a subsequent dip in response). The model demonstrates that the lateral interactions are sufficient to account for typical size tuning effects, and also accounts for less commonly reported effects that result from neurons with different specific self-organized patterns of lateral connectivity. The model thus accounts both for the typical pattern of size tuning, and explains why such a diversity of patterns is observed in animals. The results from these studies and related studies of orientation-dependent effects [3] suggest both that lateral interactions may underlie many of the observed surround modulation effects, and also that the diversity of observed effects can at least in part be traced to the diversity of lateral connection patterns, which in turn is a result of the various sequences of activations of the neurons during development.

3.3 *Aftereffects*

The previous sections have focused on the network organization and operation after Hebbian learning can be considered to be completed. However, the visual system is continually adapting to the visual input even during normal visual experience, resulting in phenomena such as visual aftereffects [53]. To investigate whether and how this adaptation differs from long-term self-organization, we tested LISSOM and GCAL-based models with stimuli used in visual aftereffect experiments [9, 19]. Surprisingly, the same Hebbian equations that allow neurons and maps to develop selectivity also lead to realistic aftereffects, such as for orientation and color (Fig. 5). In the model, we assume that connections adapt during normal visual experience just as they do in simulated long-term development, albeit with a lower learning rate appropriate for adult vision. If so, neurons that are coactive during a particular visual stimulus (such as a vertical grating) will become slightly more strongly laterally connected as they adapt to that pattern. Subsequently, the response to that pattern will be reduced, due to increased lateral excitation that leads to net (disynaptic) lateral inhibition for high contrast patterns like those in the aftereffect studies. Assuming a population decoding model such as the vector sum [9], there will be no change in the perceived orientation of the adaptation pattern, but the perceived value of a nearby orientation will be repelled *away* from the adapting stimulus, because the neurons activated during adaptation now inhibit each other more strongly, shifting the population response. These changes are the direct result of Hebbian learning

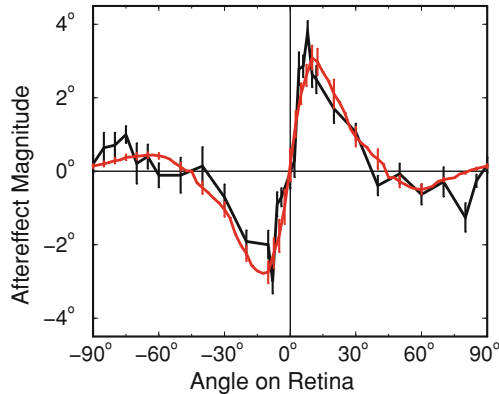


Fig. 5 *Tilt aftereffects from short-term self-organization.* If the fully organized network is repeatedly presented patterns with the same orientation, connection strengths are updated by Hebbian learning (as during development, but at a lower learning rate). The net effect is increased inhibition, which causes the neurons that responded during adaptation to respond less afterwards. When the overall response is summarized as a “perceived value” using a vector average, the result is systematic shifts in perception, such that a previously similar orientation will now seem very different in orientation, while more distant orientations will be unchanged or go in the opposite direction (*red line*, with error bars showing the standard error of measurement). These patterns are a close match to results from humans [34] (e.g. the subject from [34] plotted here as a black line, with error bars showing the standard error of measurement), suggesting that short-term and long-term adaptation share similar rules. Reprinted from Ref. [9] and replotted data from Ref. [34].

of intracortical connections, as can be shown by disabling learning for all other connections and observing no change in the overall behavior.

Interestingly, for distant orientations, the human data suggests an attractive effect, with a perceived orientation shifted *towards* the adaptation orientation [34]. The model reproduces this feature as well, and provides the novel explanation that this indirect effect is due to the divisive normalization term in the Hebbian learning equation (Eq. 7). Specifically, when the neurons activated during adaptation increase their mutual inhibition, the normalization term forces this increase to come at the expense of connections to other neurons *not* (or only weakly) activated during adaptation. Those neurons are thus disinhibited, and can respond more strongly than before, shifting the response towards the adaptation stimulus.

Similar patterns occur for the McCollough Effect [30]; see Ref. [19]. Here the adaptation stimulus coactivates neurons selective for orientation, color, or both, and again the lateral interactions between all these neurons are strengthened. Subsequent stimuli then appear different in both color and orientation, in patterns similar to the human data. Interestingly, the McCollough effect can last for months, which suggests that the modelled changes in lateral connectivity can become essentially permanent, though the effects of short-term exposure typically fade in darkness or in subsequent visual experience.

Overall, the model suggests that the same process of Hebbian learning could explain both long-term development and short-term adaptation, unifying phenomena previously considered distinct. Of course, the biophysical mechanisms may indeed be distinct, potentially operating at different time scales and this short-term adaptation being largely temporary rather than the permanent changes found early in development. Even so, the results here suggest that both early development and adult adaptation may operate using similar mathematical principles. How mechanisms for long- and short-term plasticity may interact, including possible transitions from long- to short term plasticity during so-called “critical periods”, is an important area for future modelling and experimental studies.

3.4 Time Course of Neural Responses

Visual aftereffects reflect changes in responses over the course of seconds and minutes, but with a sufficiently short time step (i.e., neural connection delay), the detailed subsecond time course of GCAL LGN and V1 neurons can also be investigated, before adaptation takes effect. Due to the recurrent nature of the GCAL architecture, responses to inputs go through a stereotypical time course that serves to highlight temporal differences in input patterns, just as the mechanisms outlined above serve to highlight spatial differences (e.g. contrast edges). As part of an ongoing project to understand temporal aspects of cortical processing [51], the temporal response properties of the GCAL orientation map were adjusted to match experimental data from [24] for the LGN, and to a fit of experimental data from [1], using a time step size and projection delay of $\delta t = 0.002$ (roughly corresponding to 0.5 ms) instead of the previous $\delta t = 0.05$. Remarkably, even though the model was originally built only to study spatial processing, we were able to match the time course at both the LGN and V1 levels by adjusting only a single parameter in the model LGN and V1: λ , which controls temporal smoothing of activity values in Eq. 3. Figure 6 compares the time course of GCAL responses to a step change in the visual input to the experimental data.

At the LGN level, λ controls only how fast the neural response can change; the underlying trends in the time course reflect the recurrent processing in the LGN, i.e., there is initially a strong response due to the afferent connectivity, which is then reduced by the divisive lateral inhibition in the LGN, with some ringing for this particular λ value. Higher levels of damping (larger λ) would eliminate this ringing, as suggested by some other experimental studies [51], but it has been retained here to show the match to this study. These results are intriguing, because they show how detailed and realistic temporal properties can arise from a circuit with elements originally added for contrast gain control (i.e., the lateral inhibition in the LGN); transient responses emerge as a natural consequence and will serve to highlight temporally varying input.

The time courses of response at the V1 level are similar, and reflect both the time course of its LGN input, and also that due to its own recurrent lateral connections, again smoothed by a λ parameter (Eq. 2). The same conclusions also apply at the

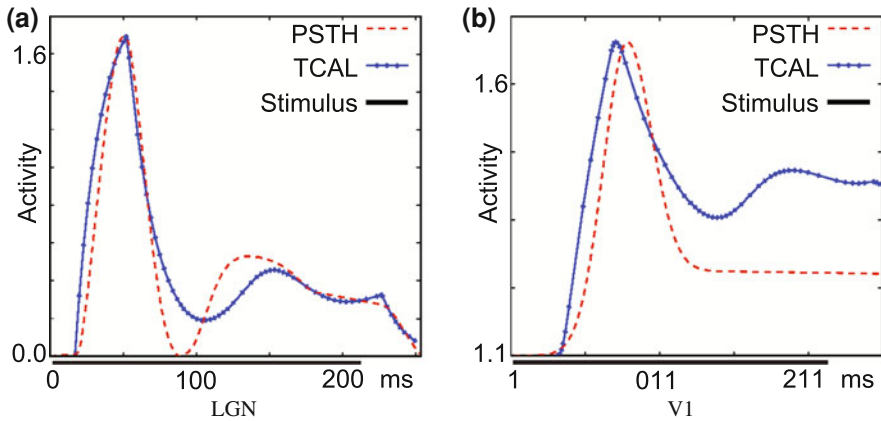


Fig. 6 *Model LGN and V1 temporal responses.* **a** The dashed red line shows experimental measurements of a cat LGN peristimulus time histogram (PSTH) in response to a step input [24], with a characteristic large onset response, some ringing, smaller sustained response, and eventual offset. The blue line shows the best fit from adjusting the activity smoothing in a GCAL orientation map; the fit is remarkably good considering that only a single GCAL parameter was varied (λ for the LGN). **b** The dashed red line shows results from a simple mathematical model of experimental measurements from monkey V1 [1], compared to the best fit response from GCAL from varying λ for V1 (and using the above fit at the LGN level). Again, the fit is remarkably good given the limited control provided by λ , indicating that the underlying dynamics of the model network are already a good match to neural circuits. Reprinted from Ref. [51]

V1 level, with V1 responses higher for changing stimuli than for sustained inputs, and reflecting the structure of the recurrent network in which neurons are embedded, rather than complex single-neuron temporal properties.

4 Discussion and Future Work

The results reviewed above illustrate a general approach to understanding the large-scale development, organization, and function of cortical areas, as a way of understanding processing for real-world data in general. The models show that a relatively small number of basic and largely uncontroversial assumptions and principles may be sufficient to explain a very wide range of experimental results from the visual cortex. Even very simple neural units, i.e., firing-rate point neurons, generically connected into topographic maps with initially random or isotropic weights, can form a wide range of specific feature preferences and maps via unsupervised normalized Hebbian learning of natural images and spontaneous activity patterns. The resulting maps consist of neurons with realistic spatial and temporal response properties, with variability due to visual context and recent history that explains significant aspects of surround modulation and visual aftereffects. The simulator and

example simulations are freely downloadable from www.topographica.org, allowing any interested researcher to build on this work.

Although all the results listed above were from V1, the cortical architecture contained no vision-specific elements at the start of the simulation, and is thus general purpose. Similar models have already been used for other cortical regions, such as rodent barrel cortex [57]. Combining the existing models into a single, runnable visual system is very much a work in progress, but the results so far suggest that doing so will be both feasible and valuable.

As previously emphasized, many of the individual results found with GCAL can also be obtained using other modelling approaches, which can be complementary to the processes modeled by GCAL. For instance, it is possible to generate orientation maps without any activity-dependent plasticity, through the initial wiring pattern between the retina and the cortex [37, 42] or within the cortex itself [25]. Such an approach cannot explain subsequent experience-dependent development, whereas the Hebbian approach of GCAL can explain both the initial map and later plasticity, but it is of course possible that the initial map and the subsequent plasticity occur via different mechanisms. Other models are based on abstractions of some of the mechanisms in GCAL [22, 35, 58, 61], operating similarly but at a higher level. GCAL is not meant as a competitor to such models, but as a concrete, physically realizable implementation of those ideas, forming a prototype of both the biological system and potential future artificial vision systems.

5 GCAL as a Starting Point for Higher-Level Mechanisms

At present, all of the models reviewed contain feedforward and lateral connections, but no feedback from higher cortical areas to V1 or from V1 to the LGN, because such feedback has not been found necessary to replicate the features surveyed. However, note that nearly all of the physiological data considered was from anesthetized animals not engaged in any visually mediated behaviors. Under those conditions, it is not surprising that feedback would have relatively little effect. Corticocortical and corticothalamic feedback is likely to be crucial to explain how these circuits operate during natural vision [49, 52], and determining the form and function of this feedback is an important aspect of developing a general-purpose cortical model. By clearly establishing which V1 properties do not require such feedback, GCAL represents an excellent starting point for building and understanding models of these phenomena.

Eventually, such models will need to be trained using input that reflects the complete context in which an animal develops, rather than just the fixed and arbitrary stream of training images used so far. Ideally, a model of visual system development in primates would be driven by color, stereo, foveated video streams replicating typical patterns of eye movements, movements of an animal in its environment, and responses to visual patterns. Collecting data of this sort is difficult, and moreover cannot capture any causal or contingent relationships between the current visual input

and the current neural organization that can affect future eye and organism movements that will then change the visual input. In the long run, to account for more complex aspects of visual system development such as visual object recognition and optic flow processing, it will be necessary to implement the models as embodied, situated agents [39, 56] embedded in the real world or in realistic 3D virtual environments. Building such robotic or virtual agents will add significant additional complexity, however, so it is important first to see how much of the behavior of V1 neurons can be addressed by the present “open-loop”, non-situated approach.

As discussed throughout, the main focus of this modelling work has been on replicating experimental data using a small number of computational primitives and mechanisms, with a goal of providing a concise, concrete, and relatively simple explanation for a wide and complex range of experimental findings. A complete explanation of visual cortex development and function would go even further, demonstrating more clearly *why* the cortex should be built in this way, and precisely what information-processing purpose this circuit performs. For instance, realistic receptive fields can be obtained from “normative” models embodying the idea that the cortex is developing a set of basis functions to represent input patterns faithfully, with only a few active neurons [14, 26, 36, 41], maps can emerge by minimizing connection lengths in the cortex [28], and lateral connections can be modelled as decorrelating the input patterns [6, 21]. The GCAL model can be seen as a concrete, mechanistic implementation of these ideas, showing how a physically realizable local circuit could develop receptive fields with good coverage of the input space, via lateral interactions that also implement sparsification via decorrelation [31]. Making more explicit links between mechanistic models like GCAL and normative theories is an important goal for future work. Meanwhile, there are many aspects of cortical function not explained by current normative models. The focus of the current line of research is on first capturing those phenomena in a general-purpose mechanistic model, so that researchers can then build deeper explanations for why these computations are useful for the organism. The following section outlines the beginning of such an explanation, in the context of data processing in general.

6 Building Complex Systems

If one steps back from cortical modelling to consider what the underlying circuits in GCAL are doing, the simulations reported here suggest a relatively straightforward process for building a circuit or device to process real-world input data:

1. Make sure your input data is organized into a meaningful two-dimensional spatial arrangement. Such a representation comes for free with many types of input data, reflecting the spatiotemporal ordering of the physical world, but for other types of data (as in the olfactory system) the data must first be organized into a two-dimensional pattern in some space (as for the odorant maps in the olfactory bulb [27]). GCAL can perform such mapping itself for small-scale networks, but large

enough networks would require a very extensive set of connections, and thus establishing some initial mapping (as for retinotopy in animals and in GCAL) can be considered a prerequisite.

2. Given this spatial representation, decompose your input data to be processed by a large array of local processing units by mapping it to a set of simulated neurons with topographically local afferent connection fields, so that each can compute independently.
3. Connect your processing units laterally with a pattern that ensures that local patches of neurons respond to similar inputs, and that more distant neurons respond to different inputs. This type of network will generate “activity bubbles” in response to a strong input pattern, with bubbles in different locations depending on the input pattern, to achieve coverage of the input pattern features.
4. Allow neural excitability to vary via homeostatic plasticity, so that neurons adapt to the patterns of input strength over time.
5. Adjust all connections using Hebbian learning, to ensure that neurons become selective for particular local patterns, while others become selective for other patterns.

The resulting network will thus remap the inputs into a sparse representation that covers the ranges of variability in the input data, with lateral connectivity that makes neurons compete to represent a given input, while filling in expected patterns in cases of weak inputs. Short-term adaptation following similar rules as self-organization will make neurons respond most strongly to changes in the input statistics, again highlighting important events. At an even faster time scale, the temporal responses of these recurrently connected neurons will again highlight moment-to-moment changes in input patterns.

The resulting spatially and temporally decorrelated representation can then be available as a substrate for higher-level processing such as reinforcement or supervised learning, acting on sparse patterns that reflect the underlying sources of variability in the environment rather than the initial dense and highly redundant pattern of inputs that reflect the structure of the input receptors. In principle, this approach can be used for any type of input data, potentially offering a starting point for building complex systems for data processing in general.

7 Conclusions

The GCAL model results suggest that it will soon be feasible to build a single model visual system that will account for a very large fraction of the visual response properties, at the firing rate level, of V1 neurons in a particular species. Such a model will help researchers make testable predictions to drive future experiments to understand cortical processing, as well as determine which properties require more complex approaches, such as feedback, attention, and detailed neural geometry and dynamics. The model suggests that cortical neurons develop to cover the typical range

of variation in their thalamic inputs, within the context of a smooth, multidimensional topographic map, and that lateral connections store pairwise correlations and use this information to modulate responses to natural scenes, dynamically adapting to both long-term and short-term visual input statistics.

Because the model cortex starts without any specialization for vision, it represents a general model for any cortical region, and is also an implementation for a generic information processing device that could have important applications outside of neuroscience. By integrating and unifying a wide range of experimental results, the model should thus help advance our understanding of cortical processing and real-world information processing in general.

Acknowledgments Thanks to all of the collaborators whose modelling work is reviewed here, and to the members of the Developmental Computational Neuroscience research group, the Institute for Adaptive and Neural Computation, and the Doctoral Training Centre in Neuroinformatics, at the University of Edinburgh, for discussions and feedback on many of the models. This work was supported in part by the UK EPSRC and BBSRC Doctoral Training Centre in Neuroinformatics, under grants EP/F500385/1 and BB/F529254/1, and by the US NIMH grant R01-MH66991. Computational resources were provided by the Edinburgh Compute and Data Facility (ECDF).

References

1. D.G. Albrecht, W.S. Geisler, R.A. Frazor, A.M. Crane, Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *J. Neurophysiol.* **88**(2), 888–913 (2002)
2. H.J. Alitto, W.M. Usrey, Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron* **57**(1), 135–146 (2008)
3. J. Antolik, Unified developmental model of maps, complex cells and surround modulation in the primary visual cortex. Ph.D. thesis, School of Informatics, The University of Edinburgh, Edinburgh, UK, 2010
4. J. Antolik, J.A. Bednar, Development of maps of simple and complex cells in the primary visual cortex. *Frontiers Comput. Neurosci.* **5**, 17 (2011)
5. C.E. Ball, J.A. Bednar, A self-organizing model of color, ocular dominance, and orientation selectivity in the primary visual cortex. in *Society for Neuroscience Abstracts*. Society for Neuroscience, www.sfn.org, Program No. 756.9 (2009)
6. H.B. Barlow, P. Földiák, Adaptation and decorrelation in the cortex, in *The Computing Neuron*, ed. by R. Durbin, C. Miall, G. Mitchison (Addison-Wesley, Reading, 1989), pp. 54–72
7. J.A. Bednar, Building a mechanistic model of the development and function of the primary visual cortex. *J. Physiol.* (Paris, 2012 in press)
8. J.A. Bednar, A. Kelkar, R. Miiikkulainen, Scaling self-organizing maps to model large cortical networks. *Neuroinformatics* **2**, 275–302 (2004)
9. J.A. Bednar, R. Miiikkulainen, Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Comput.* **12**(7), 1721–1740 (2000)
10. J.A. Bednar, R. Miiikkulainen, Self-organization of spatiotemporal receptive fields and laterally connected direction and orientation maps. *Neurocomputing* **52–54**, 473–480 (2003)
11. J.A. Bednar, R. Miiikkulainen, Prenatal and postnatal development of laterally connected orientation maps. in *Computational Neuroscience: Trends in Research*, (2004), p. 985–992
12. J.A. Bednar, R. Miiikkulainen, Prenatal and postnatal development of laterally connected orientation maps. *Neurocomputing* **58–60**, 985–992 (2004)

13. J.A. Bednar, R. Miikkulainen, Joint maps for orientation, eye, and direction preference in a self-organizing model of V1. *Neurocomputing* **69**(10–12), 1272–1276 (2006)
14. A.J. Bell, T.J. Sejnowski, The independent components of natural scenes are edge filters. *Vision Res.* **37**, 3327 (1997)
15. E.L. Bienenstock, L.N. Cooper, P.W. Munro, Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982)
16. G.G. Blasdel, Orientation selectivity, preference, and continuity in monkey striate cortex. *J. Neurosci.* **12**, 3139–3161 (1992)
17. V. Bonin, V. Mante, M. Carandini, The suppressive field of neurons in lateral geniculate nucleus. *J. Neurosci.* **25**, 10844–10856 (2005)
18. W.H. Bosking, Y. Zhang, B.R. Schofield, D. Fitzpatrick, Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**(6), 2112–2127 (1997)
19. J. Ciroux, Simulating the McCollough effect in a self-organizing model of the primary visual cortex. Master's thesis, The University of Edinburgh, Scotland, UK, 2005
20. D.M. Coppola, L.E. White, D. Fitzpatrick, D. Purves, Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proc. Nat. Acad. Sci. U.S.A.* **95**(5), 2621–2623 (1998)
21. D.W. Dong, Associative decorrelation dynamics: A theory of self-organization and optimization in feedback networks, in *Advances in Neural Information Processing Systems 7*, ed. by G. Tesauro, D.S. Touretzky, T.K. Leen (MIT Press, Cambridge, 1995), pp. 925–932
22. B.J. Farley, H. Yu, D.Z. Jin, M. Sur, Alteration of visual input results in a coordinated reorganization of multiple visual cortex maps. *J. Neurosci.* **27**(38), 10299–10310 (2007)
23. F. Felisberti, A.M. Derrington, Long-range interactions modulate the contrast gain in the lateral geniculate nucleus of cats. *Vis. Neurosci.* **16**, 943–956 (1999)
24. K. Funke, F. Wörgötter, On the significance of temporally structured activity in the dorsal lateral geniculate nucleus (LGN). *Prog. Neurobiol.* **53**(1), 67–119 (1997)
25. A. Grabska-Barwinska, C. von der Malsburg, Establishment of a scaffold for orientation maps in primary visual cortex of higher mammals. *J. Neurosci.* **28**(1), 249–257 (2008)
26. A. Hyvärinen, P.O. Hoyer, A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Res.* **41**(18), 2413–2423 (2001)
27. T. Imai, H. Sakano, L.B. Vosshall, Topographic mapping—the olfactory system. *Cold Spring Harb. Perspect. Biol. Med.* **2**(8), (2010)
28. A.A. Koulakov, D.B. Chklovskii, Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron* **29**, 519–527 (2001)
29. J.S. Law, J. Antolik, and J.A. Bednar. Mechanisms for stable and robust development of orientation maps and receptive fields. Technical report, School of Informatics, The University of Edinburgh, 2011. EDI-INF-RR-1404
30. C. McCollough, Color adaptation of edge-detectors in the human visual system. *Science* **149**(3688), 1115–1116 (1965)
31. R. Miikkulainen, J.A. Bednar, Y. Choe, J. Sirosh, *Computational Maps in the Visual Cortex* (Springer, Berlin, 2005)
32. K.D. Miller, A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *J. Neurosci.* **14**, 409–441 (1994)
33. K.D. Miller, D.J.C. MacKay, The role of constraints in Hebbian learning. *Neural Comput.* **6**, 100–126 (1994)
34. D.E. Mitchell, D.W. Muir, Does the tilt after effect occur in the oblique meridian? *Vision Res.* **16**, 609–613 (1976)
35. K. Obermayer, H. Ritter, K.J. Schulten, A principle for the formation of the spatial structure of cortical feature maps. *Proc. Nat. Acad. Sci. U.S.A.* **87**, 8345–8349 (1990)
36. B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996)
37. S.-B. Paik, D.L. Ringach, Retinal origin of orientation maps in visual cortex. *Nat. Neurosci.* **14**(7), 919–925 (2011)

38. C.M. Palmer, Topographic and laminar models for the development and organisation of spatial frequency and orientation in V1. Ph.D. thesis, School of Informatics, The University of Edinburgh, Edinburgh, UK, 2009
39. Z.W. Pylyshyn, Situating vision in the world. *Trends Cogn. Sci.* **4**(5), 197–207 (2000)
40. T. Ramtohl, A self-organizing model of disparity maps in the primary visual cortex. Master's thesis, The University of Edinburgh, Scotland, UK, 2006
41. M. Rehn, F.T. Sommer, A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.* **22**(2), 135–146 (2007)
42. D.L. Ringach, On the origin of the functional architecture of the cortex. *PLoS One* **2**(2), e251 (2007)
43. B. Roerig, J.P. Kao, Organization of intracortical circuits in relation to direction preference maps in ferret visual cortex. *J. Neurosci.* **19**(24), RC44 (1999)
44. A.B. Saul, A.L. Humphrey, Evidence of input from lagged cells in the lateral geniculate nucleus to simple cells in cortical area 17 of the cat. *J. Neurophysiol.* **68**(4), 1190–1208 (1992)
45. M.P. Sceniak, D.L. Ringach, M.J. Hawken, R. Shapley, Contrast's effect on spatial summation by macaque V1 neurons. *Nat. Neurosci.* **2**, 733–739 (1999)
46. G. Sclar, R.D. Freeman, Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.* **46**, 457–461 (1982)
47. F. Sengpiel, A. Sen, C. Blakemore, Characteristics of surround inhibition in cat area 17. *Exp. Brain Res.* **116**(2), 216–228 (1997)
48. H.T. Siegelmann, E.D. Sontag, Turing computability with neural nets. *Appl. Math. Lett.* **4**, 77–80 (1991)
49. A.M. Sillito, J. Cudeiro, H.E. Jones, Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci.* **29**(6), 307–316 (2006)
50. L.C. Sincich, G.G. Blasdel, Oriented axon projections in primary visual cortex of the monkey. *J. Neurosci.* **21**, 4416–4426 (2001)
51. J.-L. Stevens, A temporal model of neural activity and VSD response in the primary visual cortex. Master's thesis, The University of Edinburgh, Scotland, UK, 2011
52. A. Thiele, A. Pooremaeli, L.S. Delicato, J.L. Herrero, P.R. Roelfsema, Additive effects of attention and stimulus contrast in primary visual cortex. *Cereb. Cortex* **19**(12), 2970–2981 (2009)
53. P. Thompson, D. Burr, Visual aftereffects. *Curr. Biol.* **19**(1), R11–14 (2009)
54. G.G. Turrigiano, Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* **22**(5), 221–227 (1999)
55. C. Wang, C. Bardy, J.Y. Huang, T. FitzGibbon, B. Dreher, Contrast dependence of center and surround integration in primary visual cortex of the cat. *J. Vis.* **9**(1), 20.1–15, (2009)
56. J. Weng, J.L. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, E. Thelen, Autonomous mental development by robots and animals. *Science* **291**(5504), 599–600 (2001)
57. S.P. Wilson, J.S. Law, B. Mitchinson, T.J. Prescott, J.A. Bednar, Modeling the emergence of whisker direction maps in rat barrel cortex. *PLoS One*, **5**(1), (2010)
58. F. Wolf, T. Geisel, Universality in visual cortical pattern formation. *J. Physiol. Paris* **97**(2–3), 253–264 (2003)
59. J. Wolfe, L.A. Palmer, Temporal diversity in the lateral geniculate nucleus of cat. *Vis. Neurosci.* **15**(4), 653–675 (1998)
60. R.O.L. Wong, Retinal waves and visual system development. *Annu. Rev. Neurosci.* **22**, 29–47 (1999)
61. H. Yu, B.J. Farley, D.Z. Jin, M. Sur, The coordinated mapping of visual space and response features in visual cortex. *Neuron* **47**(2), 267–280 (2005)