Self-Organization, Plasticity, and Low-level Visual Phenomena in a Laterally Connected Map Model of the Primary Visual Cortex

Risto Miikkulainen, James A. Bednar, Yoonsuck Choe, and Joseph Sirosh
Department of Computer Sciences
The University of Texas at Austin, Austin, TX 78712
{risto,jbednar,yschoe,sirosh}@cs.utexas.edu

Abstract

Based on a Hebbian adaptation process, the afferent and lateral connections in the RF-LISSOM model organize simultaneously and cooperatively, and form structures such as those observed in the primary visual cortex. The neurons in the model develop local receptive fields that are organized into orientation, ocular dominance, and size selectivity columns. At the same time, patterned lateral connections form between neurons that follow the receptive field organization. This structure is in a continuously-adapting dynamic equilibrium with the external and intrinsic input, and can account for reorganization of the adult cortex following retinal and cortical lesions. The same learning processes may be responsible for a number of low-level functional phenomena such as tilt aftereffects, and combined with the leaky integrator model of the spiking neuron, for segmentation and binding. The model can also be used to verify quantitatively the hypothesis that the visual cortex forms a sparse, redundancy-reduced encoding of the input, which allows it to process massive amounts of visual information efficiently.

1 Introduction

The primary visual cortex, like many other regions of the neocortex, is a topographic map, organized so that adjacent neurons respond to adjacent regions of the visual field. In addition, neurons are responsive to particular features in the input, such as lines of a particular orientation or ocularity. All neurons in a vertical column in the cortex typically have the same feature preferences. Vertical groups of neurons with the same orientation preference are called orientation columns and vertical groups with the same eye preference are called ocular dominance columns. The feature preferences gradually vary across the surface of the cortex in characteristic spatial patterns that constitute the cortical maps.

Cortical maps are shaped by visual experience. Altering the visual environment can drastically change the organization of ocular dominance and orientation columns (Hubel and Wiesel 1962, 1968, 1977, 1974). The animal is most susceptible during a critical period of early life, typically a few weeks. For example, if a kitten is raised with both eyes sutured shut, the cortex does not develop a normal organization, and ocular dominance and orientation columns do not form. Even if the eye is opened after a few weeks, the animal remains blind, even though the eye and the LGN are perfectly normal. Similarly, if kittens are raised in environments containing only vertical or horizontal contours, their ability to see other orientations suffers significantly. In the cortex,

Psychology of Learning and Motivation, in press.

most cells develop preferences for these particular orientations, and do not respond well to the other orientations (Hirsch and Spinelli 1970; Blakemore and Cooper 1970; Blakemore and van Sluyters 1975). Such experiments indicate that visual inputs are crucial to form a normal cortical organization, and suggest that the cortex tunes itself to the distribution of visual inputs.

The discovery by von der Malsburg (1973; see also Amari 1980 and Grossberg 1976) that simple computational rules could drive the development of oriented receptive fields from visual input, raised the hope that much of the structure and development of V1 could be understood in terms of very simple neuronal behavior. However, since then substantial new discoveries have changed our understanding of the primary visual cortex. New evidence indicates that cells in V1 are coupled by highly-specific long-range lateral connections (Gilbert et al. 1990; Gilbert and Wiesel 1983; Schwark and Jones 1989). These connections are reciprocal and far more numerous than the afferents, and they are believed to have a substantial influence on cortical activity. They grow exuberantly after birth and reach their full extent in a short period. During subsequent development, they get automatically pruned into well-defined clusters. Pruning happens at the same time as the afferent connections organize into topographic maps (Callaway and Katz 1990, 1991,Burkhalter et al. 1993; Katz and Callaway 1992; Luhmann et al. 1986). The final clustered distribution corresponds closely to the distribution of afferent connections in the map. For example, in the mature visual cortex, lateral connections primarily run between areas with similar response properties, such as neurons with the same orientation or eye preference (Gilbert et al. 1990; Gilbert and Wiesel 1989; Löwel and Singer 1992).

Several observations indicate that the lateral connection structure is not defined genetically, but depends on the visual input: (1) When the primary visual cortex (of the cat) is deprived of visual input during early development, lateral connectivity remains crude and unrefined (Callaway and Katz 1991). (2) The pattern of lateral connection clusters can be altered by changing the input to the developing cortex. The resulting patterns reflect correlations in the input (Löwel and Singer 1992). (3) In the mouse somatosensory barrel cortex, sensory deprivation (by sectioning the input nerve) causes drastic decreases in the extent and density of lateral connections (McCasland et al. 1992). These observations suggest that the development of lateral connections, like that of afferent connections, depends on cortical activity driven by external input.

New discoveries have also changed the notion that the afferent structures and lateral connections are essentially static after a critical period of early development. Recent results show that the adult cortex can undergo significant, often reversible, reorganization in response to various sensory and cortical manipulations such as lesions in the receptive surface and the cortex (Gilbert 1992; Kaas 1991; Merzenich et al. 1990; Kapadia et al. 1994; Pettet and Gilbert 1992).

Based on the above results, a new theory of the visual cortex has started to emerge. The cortex appears to be a continuously-adapting structure in a dynamic equilibrium with both the external and intrinsic input. This equilibrium is maintained by cooperative and competitive lateral interactions within the cortex, mediated by lateral connections (Gilbert et al. 1990). The afferent and lateral connection patterns develop synergetically and simultaneously, based on the same underlying process. The primary function of the afferent and lateral structures is to form a sparse, redundancy-reduced encoding of the visual input (Barlow 1972; Field 1994). By integrating information over large portions of the cortex, lateral connections assist in grouping simple features into perceptual objects (Singer et al. 1990; von der Malsburg and Singer 1988), and may be responsible for low-level visual phenomena such as tilt illusions and aftereffects.

The exact mechanisms of such self-organization, plasticity, and function are still not completely understood. Computational models can play a fundamental role in this research. With the advent of massively parallel computers in the last five years, it has become possible to simulate large numbers of neural units and their connections. At the same time, neurobiological techniques for mapping the response properties and connectivity of neurons have become sophisticated enough to constrain and validate such models. This technological confluence provides a timely opportunity to test hypotheses about cortical mechanisms through large-scale computational experiments.

One of the most powerful computational abstractions of biological learning is the Hebb rule (Gustafsson and Wigström 1988; Hebb 1949), where synaptic efficacies are adjusted based on coincident pre- and postsynaptic activity. If two neurons are active at the same time, their connection is deemed useful and is strengthened. Hebbian learning is usually coupled with normalization so that the efficacies do not grow without bounds but only their relative strengths change (Miller 1994b). A network of such units and connections may develop a globally ordered structure where units represent specific inputs, such as lines of different orientation. Such a process is called self-organization: there is no global supervisor directing the process, but learning is based on a local rule and driven by the input.

Several computational models have already shown how receptive fields and their global organization in the cortical network can develop through Hebbian self-organization of afferent synapses (Erwin et al. 1995; Goodhill 1993; Kohonen 1982; Miller 1994a; Miller et al. 1989; Obermayer et al. 1990b; von der Malsburg 1973). Some of these models have also shown that aspects of cortical plasticity, such as remapping of cortical topography following peripheral lesions, can be explained with similar mechanisms (Obermayer et al. 1990a; Ritter et al. 1992). However, these models have not taken the lateral interactions between cells into account, or have assumed that they are preset and fixed and have a regular profile. Therefore, the simultaneous self-organization of lateral connections and afferent structures, and many aspects of cortical plasticity such as reorganization of the map in response to cortical lesions and reshaping of mature receptive fields in response to retinal lesions, cannot be explained by these models.

This article shows that Hebbian self-organization in a large recurrent network of simple neural elements can provide a unified account of self-organization, plasticity, and low-level function in the visual cortex. The model explains computationally (1) how the receptive fields develop selectivity to orientation, ocular dominance, and size, (2) how such receptive fields organize into intertwined columnar areas, (3) how the lateral connections develop synergetically with the afferent connections and follow their global organization, (4) how such structures are maintained in a dynamic equilibrium with the input, resulting in reorganization after retinal and cortical lesions, and (5) how fast adaptation of lateral connections can be responsible for functional phenomena such as tilt aftereffects and segmentation. The model also suggests a functional role for the lateral connections: during development, they learn the activity correlations between cortical neurons, and during visual processing, filter out these correlations from cortical activity to form a redundancy-reduced sparse coding of the visual input.

2 The Receptive Field LISSOM (RF-LISSOM) model

RF-LISSOM, or Receptive Field Laterally Interconnected Synergetically Self-Organizing Map (Sirosh 1995; Sirosh and Miikkulainen 1994, 1996, 1997; Sirosh et al. 1996; figure 1), was designed to give a computational account for the observed self-organization, plasticity, and low-level functional phenomena in the primary visual cortex. The cortical architecture has been simplified to the minimum necessary configuration to account for the observed phenomena. Because the focus is on the two-dimensional organization of the cortex, each "neuron" in the model corresponds to a vertical column

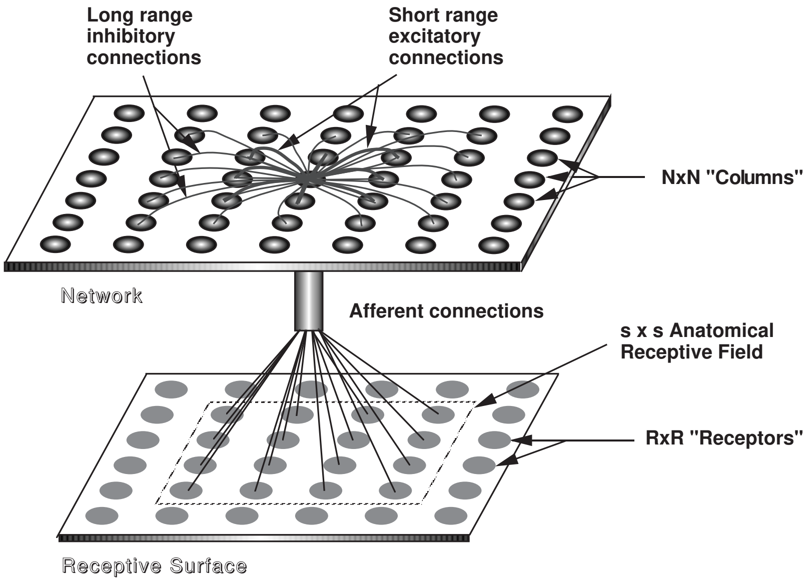


Figure 1: **The RF-LISSOM model.** The lateral excitatory and lateral inhibitory connections of a single neuron in the network are shown, together with its afferent connections. The afferents form a local anatomical receptive field on the retina.

of cells through the six layers of the cortex. The transformations in the LGN are also bypassed for simplicity.

In RF-LISSOM the cortical network consists of a sheet of interconnected neurons. Through afferent connections, each neuron receives input from a receptive surface, or "retina". In addition, each neuron has reciprocal excitatory and inhibitory lateral connections with other neurons. Lateral excitatory connections are short-range, connecting only close neighbors. Lateral inhibitory connections run for long distances, and may even implement close to full connectivity between neurons in the network. Such a lateral interaction profile is intended to establish local cooperation and global competition among the neurons, thereby making self-organization possible.

Neurons receive afferent connections from broad overlapping patches on the retina called anatomical receptive fields (RFs). The $N \times N$ network is projected on the retina of $R \times R$ receptors (for example the neuron in the top left corner of the network is projected to the top left of the retina, the center neuron to the center of the retina), and each neuron is assigned a square region of receptors of side s centered on its projection as its RF. Typically s is about half the side of the retina. Depending on the location of the neuron, its RF thereby consists of $\frac{1}{2}s \times \frac{1}{2}s$ (at the corners) to $s \times s$ (at the center) receptors.

Both afferent and lateral connections have positive synaptic weights. The weights are initially set to random values, and organized through an unsupervised learning process. At each training step, neurons start out with zero activity. An input activation pattern is introduced on the retina, and the activation propagates through the afferent connections to the cortical network. The initial response η_{ij} of neuron (i, j) is calculated as a weighted sum of the retinal activations:

$$\eta_{ij} = \sigma \left(\sum_{r_1, r_2} \xi_{r_1, r_2} \mu_{ij, r_1 r_2} \right), \tag{1}$$

where ξ_{r_1,r_2} is the activation of a retinal receptor (r_1, r_2) within the receptive field of the neuron,

μ_{ij,r_1r_2} is the corresponding afferent weight, and σ is a piecewise linear approximation of the sigmoid activation function.

The response evolves over time through lateral interaction. At each time step, each cortical neuron combines the above afferent activation $\sum \xi\mu$ with its lateral excitation and inhibition:

$$\eta_{ij}(t) = \sigma\left(\sum_{r_1,r_2} \xi_{r_1,r_2}\mu_{ij,r_1r_2} + \gamma_e \sum_{k,l} E_{ij,kl}\eta_{kl}(t-\delta t) - \gamma_i \sum_{k,l} I_{ij,kl}\eta_{kl}(t-\delta t)\right), \tag{2}$$

where $E_{ij,kl}$ is the excitatory lateral connection weight on the connection from neuron (k,l) to neuron (i,j), $I_{ij,kl}$ is the inhibitory connection weight, and $\eta_{kl}(t-\delta t)$ is the activity of neuron (k,l) during the previous time step. In other words, the retinal activity stays constant while the cortical response settles. The scaling factors γ_e and γ_i determine the strength of the lateral excitatory and inhibitory interactions. The activity pattern starts out diffuse and spread over a substantial part of the map, and converges iteratively into stable focused patches of activity, or activity bubbles.

The settling process determines the neighborhood around the initial response where the adaptation will occur. After the activity has settled, typically in a few iterations of equation 2, the connection weights of each active neuron are modified. Both afferent and lateral weights adapt according to the same mechanism: the Hebb rule, normalized so that the sum of the weights is constant:

$$w_{ij,mn}(t+1) = \frac{w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}}{\sum_{mn}\left[w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}\right]}, \tag{3}$$

where η_{ij} stands for the activity of the neuron (i,j) in the settled activity bubble, $w_{ij,mn}$ is the afferent or the lateral connection weight (μ_{ij,r_1r_2}, $E_{ij,kl}$ or $I_{ij,kl}$), α is the learning rate for each type of connection (α_A for afferent weights, α_E for excitatory, and α_I for inhibitory) and X_{mn} is the presynaptic activity (ξ_{r_1,r_2} for afferent, η_{kl} for lateral). Afferent inputs, lateral excitatory inputs, and lateral inhibitory inputs are normalized separately.

As a result of this process, both inhibitory and excitatory lateral connections strengthen by correlated activity. At long distances, very few neurons have correlated activity and therefore most long-range connections eventually become weak. The weak connections are eliminated (i.e. pruned) periodically, modeling connection death during early development in animals (Burkhalter et al. 1993; Dalva and Katz 1994; Fisken et al. 1975; Gilbert 1992; Katz and Callaway 1992; Löwel and Singer 1992). Through the weight normalization, the remaining inhibition concentrates in a closer neighborhood of each neuron. The radius of the lateral excitatory interactions starts out large, but as self-organization progresses, it is decreased until it covers only the nearest neighbors. Such a decrease is necessary for global topographic order to develop and for the receptive fields to become well-tuned at the same time (for theoretical motivation for this process, see Kohonen 1982, 1989, 1993; Obermayer et al. 1992; Sirosh and Miikkulainen 1997; for neurobiological evidence, see Dalva and Katz 1994; Hata et al. 1993.) Together the pruning of lateral connections and decreasing excitation range produce activity bubbles that are gradually more focused and local. As a result, weights change in smaller neighborhoods, and receptive fields become better tuned to local areas of the retina. Let us next turn to simulations that illustrate this process.

3 Self-Organization

In this section, three self-organizing experiments with the RF-LISSOM model are presented. The experiments show how the observed organization of feature detectors and lateral connections in the primary visual cortex could form based on activity-dependent self-organization, driven by the regularities in the input. In the first experiment, the input patterns consist of elongated Gaussian spots of light, and the model develops orientation maps. In the second experiment, a second retina is included in the model, and ocular dominance columns appear on the map. In the third, Gaussian spots of different sizes are used as the input, resulting in size-selective columns. In all these cases, the lateral connectivity patterns are found to follow the receptive field properties, as has been observed in the cortex.

3.1 Development of Orientation Columns and Lateral Connections

In this experiment, the inputs to the network consisted of simple images of multiple elongated Gaussian spots on the retinal receptors. The activity ξ_{r_1,r_2} of receptor (r_1, r_2) inside a spot is given by

$$\xi_{r_1,r_2} = exp(-\frac{((r_1 - x_i)cos(\alpha) - (r_2 - y_i)sin(\alpha))^2}{a^2} - \frac{((r_1 - x_i)sin(\alpha) + (r_2 - y_i)cos(\alpha))^2}{b^2}), \quad (4)$$

where a^2 and b^2 specify the length along the major and minor axes of the Gaussian, α specifies its orientation (chosen randomly from the uniform distribution in the range $0 \leq \alpha < \pi$), and (x_i,y_i): $0 \leq x_i, y_i < R$ specifies its center.

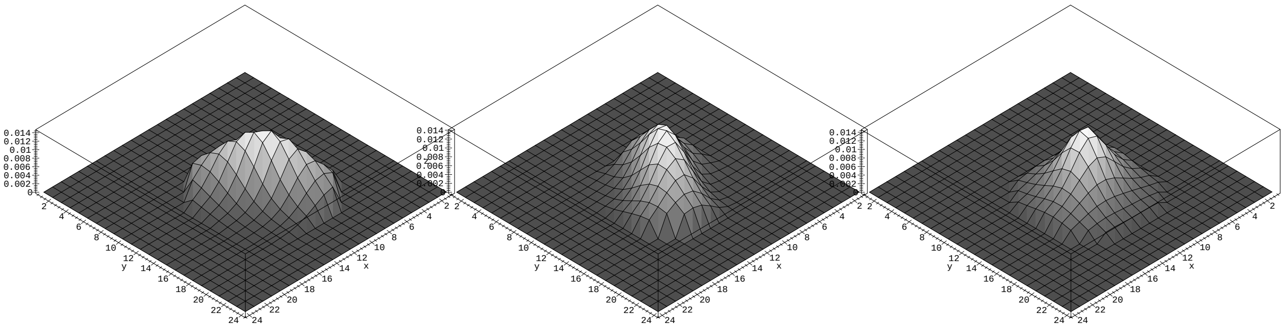
The model consisted of an array of 192×192 neurons, and a retina of 24×24 receptors. The anatomical receptive field of each neuron covered 11×11 receptors. The initial lateral excitation radius was 19 and was gradually decreased to 1. The lateral inhibitory radius of each neuron was 47, and weak inhibitory connections were pruned away at 30,000 iterations.[1]. The network had approximately 400 million connections in total, and took 8 hours to simulate on 64 processors of the Cray T3D at the Pittsburgh Supercomputing Center.

The self-organization of afferents results in a variety of oriented receptive fields similar to those found in the visual cortex (figure 2). Some are highly selective to inputs of a particular orientation, others unselective. The global organization of such receptive fields can be visualized by labeling each neuron by the preferred angle and degree of selectivity to inputs at that angle. The resulting orientation map (figure 3) is remarkably similar in structure to those observed in the primary visual cortex by recent imaging techniques (Blasdel 1992; Blasdel and Salama 1986) and contains structures such as pinwheels, fractures and linear zones.[2] The results strongly suggest that Hebbian self-organization of afferent weights, based on recurrent lateral interactions, underlie the development of orientation maps in the cortex.

The lateral connection weights self-organize at the same time as the orientation map forms. Initially, the connections are spread over long distances and cover a substantial part of the network (figure 3*a*). As lateral weights self-organize, the connections between uncorrelated regions

[1]The lateral connections were pruned if their strength was less than 0.00025. The afferent weights were initially random (chosen from a uniform distribution). In this particular experiment, the lateral inhibitory connections were initially set to a Gaussian distribution with $\sigma = 100$, and the lateral excitatory connections to a Gaussian with $\sigma = 15$ to speed up learning; uniform random values can be used as well. The widths of the oriented Gaussian input spots were $a = 7.5$ and $b = 1.5$. The lateral excitation γ_e and inhibition strength γ_i were both 0.9. The learning rate α_A decreased from 0.007 to 0.0015, α_E from 0.002 to 0.001 and α_I was a constant 0.00025. The lower and upper thresholds of the sigmoid increased from 0.1 to 0.24 and from 0.65 to 0.82. Small variations of these parameters produce roughly equivalent results. Similar parameters were used in other experiments described in this article.

[2]The similarity was measured by comparing Fourier transforms, autocorrelation functions, and correlation angle histograms of experimental and model maps. See (Erwin et al. 1995) for a discussion of these methods.



(*a*) RF sharply tuned to 60° (*b*) RF sharply tuned to 127.5° (*c*) Unselective RF

Figure 2: **Self-organization of oriented afferent receptive fields.** The afferent weights of three neurons at different locations in the network are shown, plotted on the retinal surface. The first two are strongly oriented, whereas the third is unoriented and symmetric.

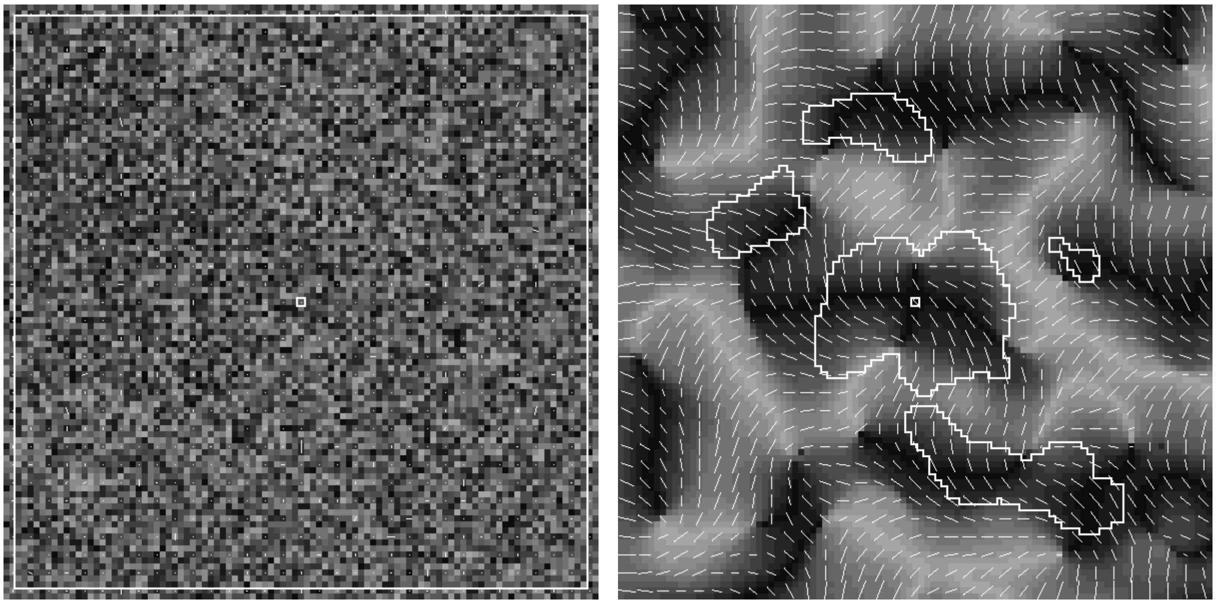
become weaker, and after pruning, only the strongest connections remain (figure 3*b*). The surviving connections of highly-tuned cells, such as the one illustrated in figure 3*b*, link areas of similar orientation preference, and avoid neurons with the orthogonal orientation preference. Furthermore, the connection patterns are elongated along the direction that corresponds to the neuron's preferred stimulus orientation. This organization reflects the activity correlations caused by the elongated Gaussian input pattern: such a stimulus activates primarily those neurons that are tuned to the same orientation as the stimulus, and located along its length. At locations such as fractures, where a cell is sandwiched between two orientation columns of very different orientation preference, the lateral connections are elongated along the two directions preferred by the two adjacent columns. Finally, the lateral connections of unselective cells, such as those at pinwheel centers, connect to all orientations around the cell. Thus the pattern of lateral connections of each neuron closely follows the global organization of receptive fields, and represents the long-term activity correlations over large areas of the network.

Some of these results have already been confirmed in very recent neurobiological experiments[3] (Fitzpatrick et al. 1994). In the iso-orientation columns of the tree-shrew cortex, horizontal connections were found to be distributed anisotropically, extending farther and giving rise to more terminals along the preferred orientation of the neuron. Most of these terminals also connected to cells with the same orientation preference. The connection patterns at pinwheel centers and fractures have not been studied experimentally so far; our model predicts that they will have unselective and biaxial distributions, respectively.

3.2 Self-organization of ocular dominance and lateral connection patterns

In addition to responding to specific orientations, the neurons in the primary visual cortex are also selective to the eye from which the input originates. In the second experiment with RF-LISSOM

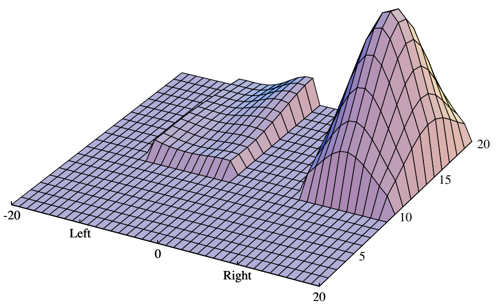
[3] Note that if the lateral connection patterns are observed on the cortex directly, it is very difficult to determine their orientation because of the log-polar mapping from the retina to the cortex. The cortical patterns would first have to be mapped back to the visual space. The model bypasses the log-polar transformation for simplicity, and the lateral connection patterns are directly observable.



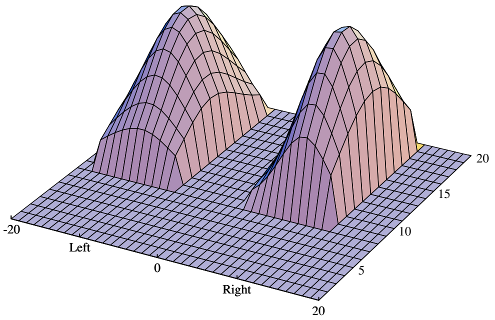
(a) Initial unordered map and connections

(b) Final orientation map and connections

Figure 3: **Self-organization of the orientation map and lateral connections.** Each neuron in this 100×100 central region of the map is shaded according to its orientation preference. Shades from dark to light represent continuously-changing orientation preference from 127.5°to 37.5°from the horizontal, and from light to dark preference from 37.5°to -52.5°, i.e. back to 127.5°. This gray-scale scheme was chosen so that the connections of the center neuron, which is tuned to 127.5°and identified by a small box in the figure, could be clearly plotted, and also so that the shading would be continuous through all angles (as is possible in color plots; see e.g. Sirosh et al. 1996). To disambiguate the shading, every third neuron in every third row is marked with a line that identifies the neuron's orientation preference. In addition, the length of the line indicates how selective the neuron is to its preferred orientation. The outlined areas indicate units from which the center unit has lateral connections. *(a)* Initially, the afferent weights of each neuron are random, and the receptive fields are randomly oriented and very unselective, as shown by the random shades and random and short lines (the line lengths were slightly magnified so that they could be seen at all). The lateral connections cover a wide area uniformly. *(b)* After several thousand input presentations, the receptive fields have organized into continuous and highly selective bands of orientation columns. The orientation preference patterns have all the significant features found in visuo-cortical maps: (1) *pinwheel centers*, around which orientation preference changes through 180°(e.g. the neuron eight lines from the left and four from the bottom), (2) *linear zones*, where orientation preference changes almost linearly (e.g. along the bottom at the lower left), and (3) *fractures*, where there is a discontinuous change of orientation preference (as in 7 lines from the left and 17 from the bottom). Most of the lateral connections have been pruned, and those that remain connect neurons with similar orientation preferences. The marked unit prefers 127.5°, and its connections come mostly from dark neurons. In the near vicinity, the lateral connections follow the twists and turns of the darkly-shaded iso-orientation column, and avoid the lightly-shaded columns representing the orthogonal preference. Further away, connections exist mostly along the 127.5°orientation, since these neurons tend to respond to the same input. All the long-range connections shown are inhibitory at this stage; there are excitatory connections only from the immediately neighboring units.



(a) Monocular RF



(b) Binocular RF

Figure 4: **Receptive fields with varying degrees of eye-preference.** In (a), the final afferent weights of a neuron at position $(42, 39)$ in a 60×60 network are shown. This particular neuron is monocular with strong connections to the right eye, and weak connections to the left. In (b), the weights of a binocular neuron at position $(38, 23)$ are shown. This neuron has appoximately equal weights to both eyes.

self-organization, the development of eye selectivity, or ocular dominance, was simulated. A second retina was added to the model and the afferent connections were set up exactly as for the first retina, with local receptive fields and topographically ordered RF centers. Multiple symmetric Gaussian spots were presented in each eye as input, where the activity inside each spot is given by

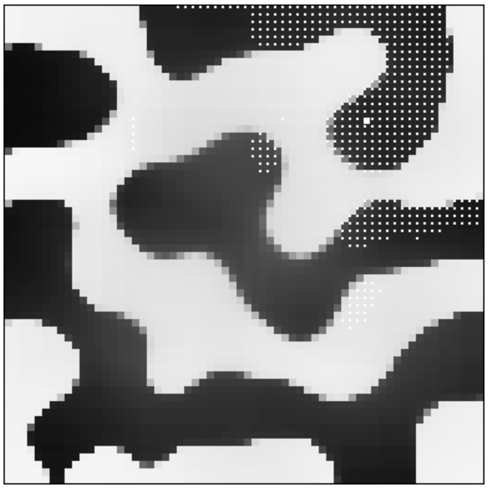
$$\xi_{r_1,r_2} = exp(-\frac{(r_1 - x_i)^2 + (r_2 - y_i)^2}{a^2}), \tag{5}$$

where a^2 specifies the width of the Gaussian, and the spot centers (x_i,y_i): $0 \leq x_i, y_i < R$, were chosen randomly. Because of cooperation and competition between inputs from the two eyes, groups of neurons developed strong afferent connections to one eye or the other, resulting in patterns of ocular dominance in the network (cf. von der Malsburg 1990; Miller et al. 1989).

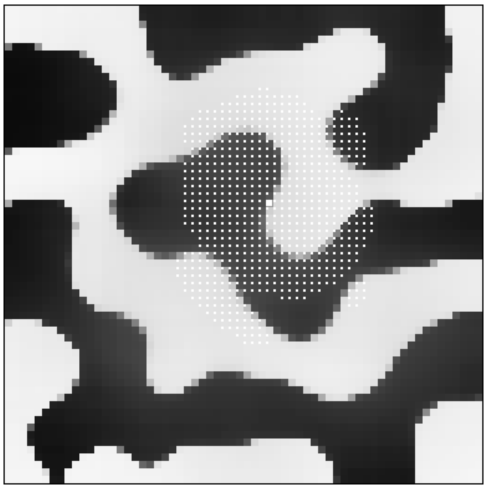
The self-organization of the network was studied with varying between-eye correlations. At each input presentation, one spot is randomly placed at (x_i,y_i) in the left retina, and a second spot within a radius of $c \times R$ of (x_i, y_i) in the right retina. The parameter $c \in [0, 1]$ specifies the spatial correlations between spots in the two retinas, and can be adjusted to simulate different degrees of correlations between images in the two eyes. Multi-spot images can be generated by repeating the above step: the simulations below used two-spot images in each eye.

Figure 4 shows the final afferent receptive fields of two typical neurons in a simulation with $c = 1$. In this case, the inputs were uncorrelated, simulating perfect strabismus (i.e. the inputs from the two eyes cannot be matched to form a single percept). In the early stages of the simulation, some of the neurons randomly develop a preference for one eye or the other. Nearby neurons will tend to share the same preference because lateral excitation keeps neural activity partially correlated over short distances. As self-organization progresses, such preferences are amplified, and groups of neurons develop strong weights to one eye. Figure 4a shows the afferent weights of a typical monocular neuron.

The extent of activity correlations on the network determines the size of the monocular neuronal groups. Farther on the map, where the activations are anticorrelated due to lateral inhibition, neurons will develop eye preferences to the opposite eye. As a result, alternating ocular dominance



(a) Connections of a Monocular Neuron



(b) Connections of a Binocular Neuron

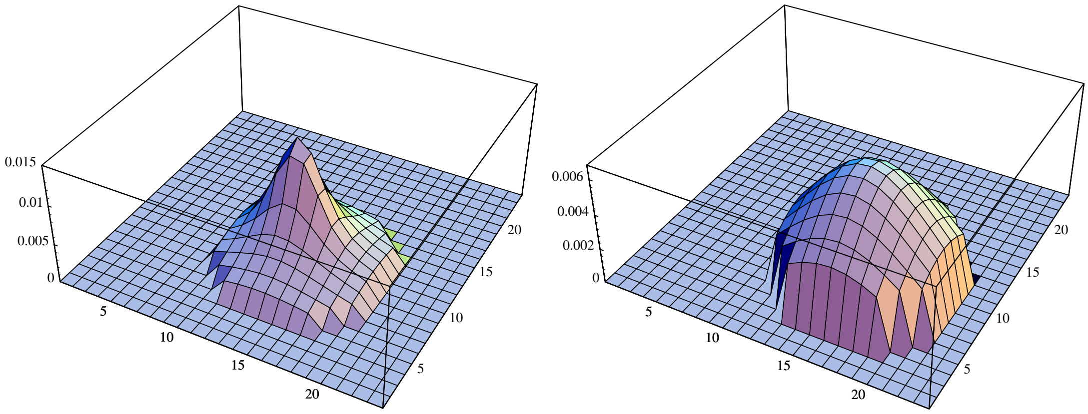
Figure 5: **Ocular dominance and patterned long-range lateral connections.** Each neuron is colored with a grey-scale value (*black* $\rightarrow$ *white*) that represents continuously changing eye preference from exclusive left through binocular to exclusive right. Most neurons are monocular, so white and black predominate. Small white dots indicate the strongest lateral input connections to the neuron marked with a big white dot. Only the long range inhibitory connections are shown. The excitatory connections link each neuron only to itself and to its eight nearest neighbors. (a) The lateral connections of a left monocular neuron predominantly link areas of the same ocular dominance. (b) The lateral connections of a binocular neuron come from both eye regions. In this simulation, the parameters were: network size $N = 64$; retinal size $R = 24$; afferent field size $s = 9$; $\delta = 0.1$; $\beta = 0.65$; spot width $a = 5.0$; excitation radius $d = 1$; inhibition radius=31; scaling factors $\gamma_e = 0.5$, $\gamma_i = 0.9$; learning rates $\alpha_{\mathrm{A}} = \alpha_{\mathrm{E}} = \alpha_{\mathrm{I}} = 0.002$; number of training iterations=$35,000$. The anatomical RF centers were slightly scattered around their topographically ordered positions (radius of scatter=0.5), and all connections were initialized to random weights.

patches develop over the map, as shown in figure 5.[4] In areas between ocular dominance patches, neurons will develop approximately equal strengths to both eyes and become binocular, like the one shown in figure 4b.

The lateral connection patterns closely follow ocular dominance organization (figure 5). As neurons become better tuned to one eye or the other, activity correlations between regions tuned to the same eye become stronger, and correlations between opposite eye areas weaker. As a result, monocular neurons develop strong lateral connections to regions with the same eye preference, and weak connections to regions of opposite eye preference. The binocular neurons, on the other hand, are equally tuned to the two eyes, and have activity correlations with both ocular dominance regions. Their lateral connection weights are distributed more or less symmetrically around them and include neurons from both eye-preference columns.

The normal case (simulated with $c = 0.4$), looks otherwise similar to figure 5, but the ocular dominance stripes are narrower and there are more ocular dominance columns in the network. Most neurons are neither purely monocular nor purely binocular and few cells have extreme values of ocular dominance. Accordingly, the lateral connectivity in the network is only partially determined by ocular dominance. However, the lateral connections of the few strongly monocular neurons follow the ocular dominance patterns like in the strabismic case. In both cases, the spacing between the lateral connection clusters matches the stripe-width.

[4]For a thorough treatment of the mathematical principles underlying the development of ocular dominance columns, see (Goodhill 1993; Miller et al. 1989; von der Malsburg and Singer 1988).



(a) Small RF: neuron $(78, 109)$ (b) Large RF: neuron $(69, 124)$

Figure 6: **Size-selective self-organized receptive fields.** The afferent weights of neurons at two different locations in a 192×192 network are shown after self-organization. Initially the weights are random, but after self-organization, a smooth hill-shaped weight profile develops. Though the anatomical RFs are the same, the afferent weights are organized into a variety of sizes from narrow, highly peaked receptive fields to large and broad ones.

The patterns of lateral connections and ocular dominance shown above closely match observations in the primary visual cortex. Löwel and Singer (1992) observed that when between-eye correlations were abolished in kittens by surgically induced strabismus, long-range lateral connections primarily linked areas of the same ocular dominance. However, binocular neurons, located between ocular dominance columns, retained connections to both eye regions. The ocular dominance stripes in the strabismics were broad and sharply defined (Löwel 1994). In contrast, ocular dominance stripes in normal animals were narrow and less sharply defined, and lateral connection patterns overall were not significantly influenced by ocular dominance. The receptive field model reproduces these experimental results, and also predicts that the lateral connections of strongly monocular neurons would follow ocular dominance even in the normal case. The model therefore confirms that patterned lateral connections develop based on correlated neuronal activity and demonstrates that they can self-organize cooperatively with ocular dominance columns.

3.3 Development of size selectivity and lateral connections

In their first recordings from the primary visual cortex of the cat, Hubel and Wiesel (1959, 1962) reported that cortical cells were more selective to the width of patterns than were retinal cells. They noted that cortical cells would give no response to a bar covering the whole receptive field, whereas in the retina and the LGN, cells would typically respond to such patterns. Subsequently, detailed studies by Campbell et al. (1969), De Valois et al. (1982) and others showed that cortical cells are narrowly tuned to the spatial frequency of inputs, and had typical bandpass responses, responding only to inputs in a specific frequency range. A continuum of spatial frequencies from low to high were represented in the cortex (Silverman et al. 1989), and cells in each range of spatial frequency were organized into distinct spatial frequency columns (Tootell et al. 1981, 1988). In essence, cortical cells exhibited an organization of spatial frequency selectivity similar to ocular dominance (OD) and orientation (OR) columns.

Modeling selectivity to spatial frequency would require much larger retinal and cortical networks than can currently be simulated. However, it is possible to test a special case of the hypothesis:

whether selective receptive fields, columnar organization, and lateral connection patterns form when the size of the Gaussian light spot is the main dimension of variation in the input.

In the third RF-LISSOM self-organization experiment, therefore, inputs of a variety of sizes were presented to the network. The light spots were similar to those in the ocular dominance experiment (equation 5), except that the width a was chosen uniformly randomly in the range [0.75,8.0]. The retinal activity vector was normalized to constant length, because without normalization, larger-sized spots would produce stronger activation. A total of $25,000$ training steps were used, with the network and simulation parameters similar to those in the orientation column simulation (section 3.1).

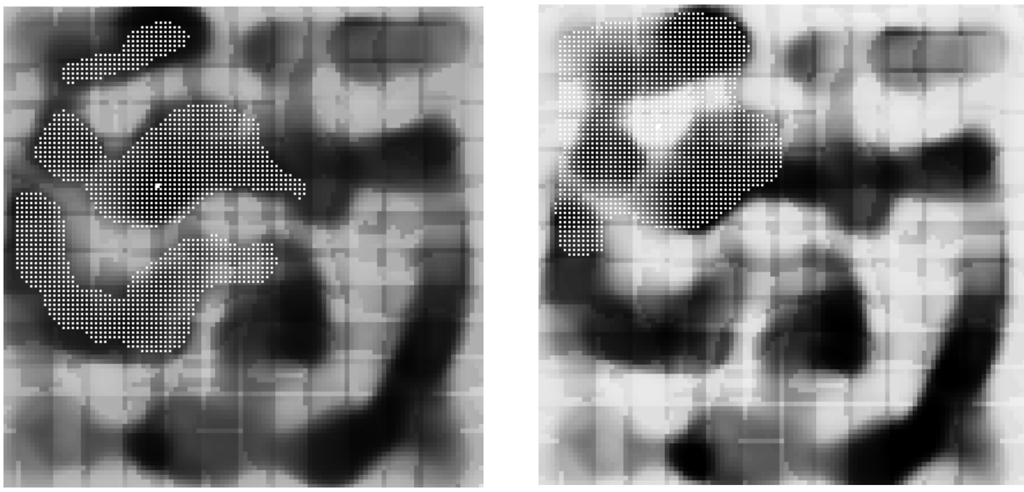
The self-organization of afferents results in smooth, hill-shaped RFs. A variety of RFs of different sizes are produced, some narrow and tuned to small stimuli, others large and most responsive to large stimuli (figure 6). Simultaneously with the RFs, each neuron's lateral connections evolve, and by the Hebbian mechanism, are distributed according to how well the neuron's activity correlates with the activities of the other neurons. Let us examine the nature of such activity correlations. The inputs vary in size from $a = 0.75$ to $a = 8.0$, and are normalized. Therefore, the smallest inputs produce very bright activity in a few receptors. They are also smaller than the size of each anatomical receptive field. Therefore, these inputs predominantly stimulate neurons with small receptive fields and having anatomical RFs in the same position as the spot. Such neurons will have strong activity correlations with other small receptive field neurons, but little correlation with neurons having broader receptive fields[5]. The global organization of size preferences and lateral connections can be visualized by labeling each neuron with a color that indicates the width of its RF, and plotting the patterns of lateral connections on top. As figure 7a shows, the RF organization has the form of connected, intertwined patches, similar to OD columns, and the lateral connections of neurons connect to regions of the same size preference.

Neurons with larger receptive fields have a slightly different pattern of activity correlations. The larger spots are not localized within the anatomical RF as are the smaller inputs, and extend beyond it. They produce activity over a wider area in the network than the smaller, localized spots. As a result, the inputs that best stimulate larger RF neurons also cause activity in large parts of the network. Therefore, the activity correlations of such neurons are not as strongly determined by size as that of small RF neurons. Therefore, the lateral connections of neurons with larger RFs often link to smaller RF neurons also. In the cortex, neurobiologists have not yet studied how the patterns of lateral connections relate to either size or spatial frequency preferences.

The columnar organization does not develop in small networks. Simulations show that, for a given variance of the stimuli size, the ratio of neurons in the network to receptors in the retina (the magnification factor) has to be greater than a threshold value for a stable columnar organization to appear. Below the threshold, smooth RFs and an ordered topographic map develop, but all the RFs tend to have the same size, corresponding to the average width of the input stimulus. Above the threshold, symmetry breaking occurs, producing a variety of RF sizes for each location in the retina. Such symmetry breaking is similar to that of the Self-Organizing Map (Kohonen 1982, 1989, 1995), where an input feature is represented in the network only if its variance is greater than a threshold proportional to the magnification factor (Obermayer et al. 1992).

It is not known whether the long-range lateral connections in the cortex are organized according to size or spatial frequency selectivity. So far, the lateral connection patterns have only been studied in relation to the organization of OD and OR preference (Malach et al. 1993; Löwel and

[5]Note that even small spots produce quite widespread activity in the network, because each retinal receptor connects to a large number of cortical neurons.



(a) Columns and lateral connections: Small RF

(b) Columns and lateral connections: Large RF

Figure 7: **Size-selective columns and lateral connection patterns.** Each neuron in the network is labeled with a grey-scale value (*black* $\rightarrow$ *white*) that represents continuously-changing size preference from small values to large values. Small white dots indicate the lateral input connections to the neuron marked with the big white dot. The size preferences are organized systematically across the network into connected, intertwined patches, and the strongest lateral connections predominantly link areas of the same size selectivity.

Singer 1992; Gilbert and Wiesel 1989). However, considerable psychophysical and neurobiological evidence indicates that selective lateral interactions exist between neurons tuned to different spatial frequencies (De Valois and Tootell 1983; Bauman and Bonds 1991). As in the RF-LISSOM model, these interactions are also known to be largely inhibitory (De Valois and Tootell 1983; Vidyasagar and Mueller 1994). The RF-LISSOM model is a first step towards modeling spatial frequency selectivity, and suggests that the long-range lateral connections could be the anatomical substrate for inhibition between spatial frequency channels. The model further predicts that the patterns of lateral connections in the cortex would be influenced not only by OD and OR preference, but also by selectivity to spatial frequency.

4 Functional role of the self-organized maps and lateral connections

The results on self-organization of OD, OR, and size-selectivity maps and lateral connections suggest that a single Hebbian mechanism produces the receptive fields and lateral interactions in the primary visual cortex. If so, what could be the functional role of these self-organized structures in visual processing?

The information processing role of the afferent RFs is best seen by analogy with Self-Organizing Maps (Kohonen 1982, 1989, 1995). The afferent connections self-organize in a similar fashion in both models. When presented with high-dimensional inputs, the self-organizing map selects the set of feature dimensions along which inputs vary the most and represents them along the dimensions of the map. For example, if the inputs lie mostly along the diagonal plane of a hypercube, the self-organized map (and hence the RFs) will spread out along this diagonal. If there is some input variance in the dimension perpendicular to this diagonal, receptive fields will be distributed along this direction as well, and the map will “fold” in that direction. If there are many such feature dimensions, a subset of them will be represented by the folds of the map in the order of their input variance (Obermayer et al. 1992).

The images in the visual world could be varying the most along the dimensions of ocular dominance, orientation preference and spatial frequency, and if so, the self-organized RFs will represent these dimensions. During visual processing, the cortex then projects incoming visual inputs onto these dimensions. As shown by Field (1994), such a projection produces a *sparse coding* of the input, minimizing the number of active neurons and forming a more efficient representation, which is well suited for the detection of suspicious coincidences, associative memory and feature grouping (Barlow 1972, 1985; Field 1987, 1994). Projecting onto the dimensions of maximum variance also achieves minimal distortion and minimal spurious conjunctions of features.

What would the role then be for the lateral connections? Through Hebbian self-organization, the lateral connections learn correlations between the feature detectors in the network—the stronger the correlation between two cells, the larger the connection strength between them. However, these long-range connections are inhibitory. Therefore, the strongly correlated regions of the network inhibit each other—in other words, the lateral connections *decorrelate* (Barlow 1972, 1989). Decorrelation is useful in producing efficient representations. If the connection between two cells is strong, then the response of one can be predicted to a large extent by knowing the response of the other. Therefore, the activity of the second cell is redundant, and a more efficient representation (in an information-theoretic sense) can be formed by eliminating the redundancy. Decorrelation filters out the learned redundancies and produces an efficient encoding of the visual input. Thus, the visual knowledge that lateral connections learn is used to filter out the already-known correlations between cortical cells, leaving only novel information to be passed on to higher levels of processing. The RF-LISSOM architecture demonstrates how decorrelation mechanisms could be implemented in the primary visual cortex.

To demonstrate sparse coding and decorrelation in RF-LISSOM, the representations in the orientation selectivity model of section 3.1 were analyzed in more detail. It was confirmed that (1) the network forms a sparse coding of the input, (2) the coding reduces redundancies, and (3) and to get these effects, it is crucial that the lateral connections are self-organized.

Sparseness can be measured by the kurtosis (i.e. peakedness) of the network response. A small number of strongly-responding neurons, that is, a sparse coding, will result in high kurtosis. In figure 8, the kurtosis measures of four different networks are compared: (1) a network without any lateral interactions at all (i.e. the initial response of the network), (2) a network with self-organized lateral weights, (3) one with fixed random lateral weights, and (4) one with lateral weights that have a fixed Gaussian profile (as assumed in some of the early self-organizing models). In each case, the amount of contrast in the input was varied: A constant pattern of several elongated Gaussian light spots was presented to the retina, and the height of the Gaussians was systematically increased.

The main observation is that the kurtosis of the self-organized lateral interactions is substantially higher than that of the other three networks. By Hebbian self-organization, the long-range lateral connections learn to encode correlations between the feature-selective units. Because the long-range connections are inhibitory, strongly correlated regions of the network inhibit each other. At the same time, the short-range lateral excitation locally amplifies the responses of active units. Together, the recurrent excitation and inhibition focuses the activity to the units best tuned to the features of the input stimulus, thereby producing high kurtosis, that is, a sparse coding of the input.

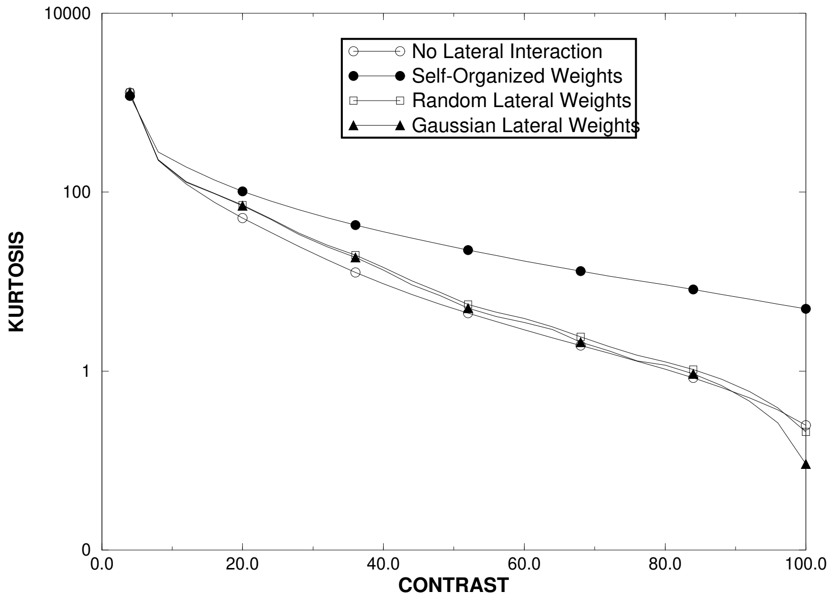
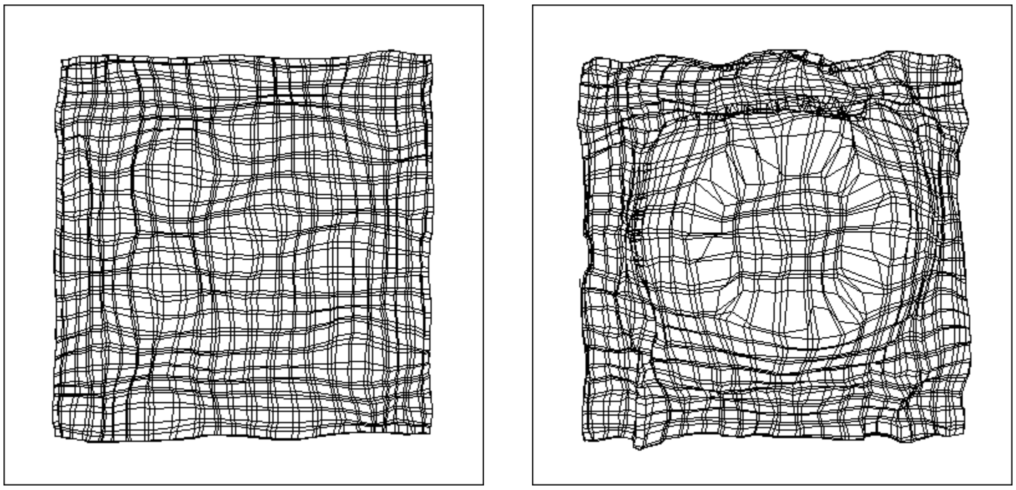


Figure 8: **Sparse coding in the RF-LISSOM model.** With self-organized lateral connections, the kurtosis (i.e. sparse coding) is significantly higher than with fixed lateral connections or without lateral connections at all. The difference is larger at higher intensities (measured by the intensity of the light spots, in percentage), suggesting that the self-organized lateral connections perform redundancy reduction on the response.

Second, the difference in kurtosis increases with contrast. As the light spots become more intense, there is more input activation to the network. More units will be activated, and kurtosis will decrease. This extra activity is mostly redundant, because the input pattern is still the same. The self-organized lateral connections are able to remove such redundant activation, and kurtosis decreases much slower than in the other cases. Third, such redundancy reduction takes place only when the lateral connections are self-organized. The kurtosis with fixed lateral weights decreases at the same rate as the kurtosis of the network without lateral weights.

In sum, the RF-LISSOM model suggests that the cortex performs two different computations during sensory processing: First, the inputs are projected onto the principal feature dimensions represented by the afferent receptive field structure. Then, the redundancies are filtered out by recurrent lateral interactions. The result is an efficient, redundancy-reduced sparse coding of the visual input which is then passed on to higher processing levels. This prediction can be verified experimentally by using information theory to analyze the optical images of cortical activity patterns produced in response to simple retinal images. If confirmed, it would constitute a major step in understanding the function of the observed primary visual cortex structures.



(a) Map before scotoma (b) Reorganized map

Figure 9: **Reorganization of the topographic map after a retinal scotoma**. The RF centers of every third neuron in the network are plotted in the retinal space, and the neighboring centers are connected by lines. (a) Before the scotoma, the centers are organized into a topographic map of the retina. (b) After the scotoma, neurons whose receptive fields were entirely covered by the scotoma remained unstimulated, and retained their old receptive fields. However, the surrounding neurons reorganized their afferent weights, and their receptive fields moved out into the periphery of the scotoma. Such reorganization produces dynamic expansions of receptive fields and an inward shift of the response at the edge of the scotoma.

5 Plasticity of the Adult Visual Cortex

So far we have demonstrated how the laterally connected cortex could self-organize in response to the external input, and what the self-organized structures mean. The model can also be used to study dynamic phenomena in the adult cortex. The first experiment below shows how the network reorganizes after a retinal lesion, giving a computational explanation to the phenomenon of dynamic receptive fields. The second experiment shows how the network reorganizes in response to cortical lesions. These experiments suggest that the same processes that are responsible for the development of the cortex also operate in the adult cortex, maintaining it in a dynamic equilibrium with the input.

5.1 Reorganization after Retinal Lesions

Dynamic receptive fields are observed in response to temporary artificial scotomas in the retina. Such a lesion prevents part of the cortex from receiving input, producing a corresponding cortical scotoma. However, if the surround of the scotoma is stimulated for a period of several minutes, and the scotoma is then removed, the receptive fields of the unstimulated neurons are found to have expanded (Pettet and Gilbert 1992). The expansion is largest along the preferred orientation of each neuron. Psychophysical experiments further show that after removing the scotoma, a stimulus at the edge of the scotoma appears to have shifted towards the center (Kapadia et al. 1994). Prima facie, such dynamic expansion of receptive fields, and the perceptual shift accompanying it, is incompatible with Hebbian self-organization (which are based on coincident activity), and has been difficult to explain.

The RF-LISSOM model suggests an explanation. Figure 9 shows how the orientation map of section 3.1 reorganizes when a retinal scotoma is introduced and inputs are presented in the surrounding area. The receptive fields of the central, unstimulated neurons remain in the same location as before. Neurons that have part of their receptive fields outside the scotoma are stimulated by the surrounding input, and by Hebbian adaptation, reorganize their afferent weights into the periphery of the scotoma. As a result, these neurons become insensitive to the central region. If the scotoma is now removed, and an input is presented in the scotoma region, only the previously unresponsive neurons respond vigorously to the new input; the surrounding ones do not. Therefore, there is considerably less lateral inhibition from the surrounding neurons to the central neurons, and responses that were previously suppressed by lateral inhibition are unmasked. Therefore, when the RF sizes of the central neurons are measured (based on network activity), they appear to have increased. The expansion is greatest along the preferred orientation because the strongest afferent weights lie in this direction (figure $2a,b$), and any decrease of inhibition unmasks responses mainly in that direction.

Such a reorganization can account for the psychophysical experiments as well. The neurons whose receptive fields have moved outward now respond to inputs farther from the center than before. Therefore, an input at the edge of the retinal scotoma stimulates many neurons inside the cortical scotoma that previously would not have responded, and the response pattern is shifted inward, producing the perceptual shift. After the scotoma is removed and the normal stimulation reestablished, the reorganized RFs gradually return to the normal state (of figure $9a$), and the shift disappears. The model thus shows how the same self-organizing processes and lateral interactions that sculpt the receptive fields during early development could, in the adult, maintain them in a continuously adapting, dynamic equilibrium with the visual environment.

5.2 Reorganization after Cortical Lesions

To simulate effects of cortical lesions the RF-LISSOM network was first organized with symmetric Gaussian patterns such as those used in the ocular dominance simulation of section 3.2. This resulted in regular Gaussian-shaped receptive fields, retinotopic global order, and a smooth "Mexican hat", or difference of Gaussians, lateral interaction profiles.

To study the effects of cortical lesions, a small set of neurons in the organized network were then made unresponsive to input. Three phases of reorganization wereobserved, like in the somatosensory cortex (Merzenich et al. 1990). Immediately after the lesion, the receptive fields (RFs) of neurons in the perilesion zone enlarge. The lesion reduces the inhibition of the perilesion neurons, and unmasks previously suppressed input activation. In effect, the perilesion neurons immediately take over representing part of the input to the lesioned region, and the apparent loss of receptive surface representation is smaller than expected based on the prelesion map (figure 10b).

The lesion disrupts the dynamic equilibrium of the network, and both lateral and afferent connections of the active neurons adapt to compensate for the lesion. Neurons close to the lesion boundary encounter a large imbalance of lateral interaction in their neighborhood, with no lateral activation from inside the lesion and normal activation from outside. As a result, the lateral connection weights to the lesioned area decrease to zero, and by Hebbian adaptation and normalization, all the lateral weights rapidly redistribute to the the lesion's periphery. Neurons at the lesion boundary have the largest number of inhibitory connections from the lesioned zone; therefore, the reorganization of inhibition is especially pronounced at the boundary. As a result, in the second phase the lateral inhibition very rapidly becomes strong outside the lesion, and the previously unmasked activity is partly suppressed (figure 10c). This produces an apparent outward shift of

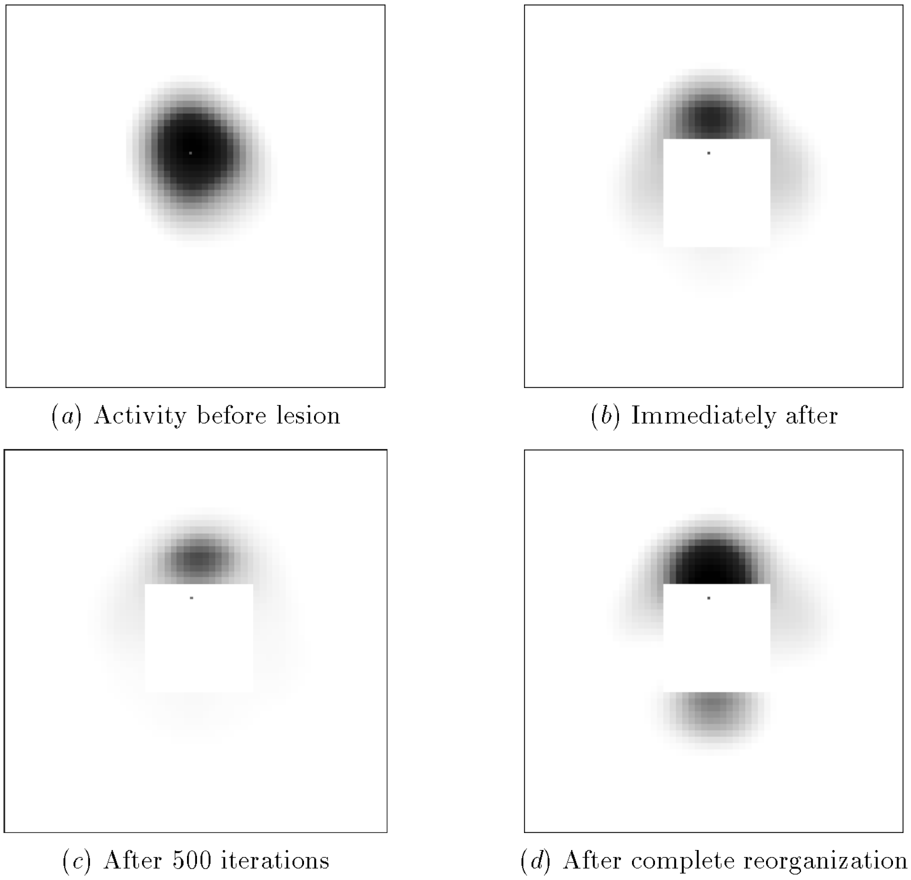
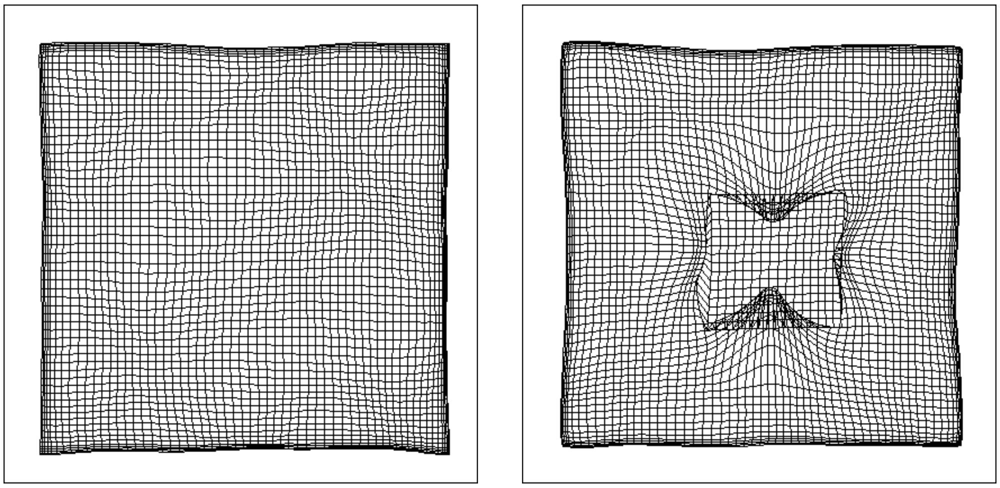


Figure 10: **Changing response patterns after a cortical lesion.** The activity of neurons across the network are shown for the same input before the lesion and at several stages after it. The lesioned area is seen as a white square with no activity, and the black dot inside the square indicates the maximally responding neuron before the lesion. Immediately after the lesion, the activity spreads out to neurons that were previously inactive and therefore, the functional loss appears less severe than expected. As lateral connections reorganize, the unmasked activity decreases because of increased lateral inhibition. Several thousand adaptation steps after the lesion, afferent weights of the perilesion neurons have spread out into the area previously represented by the lesioned neurons (figure 11*b*). Though lateral inhibition is still stronger in the perilesion area, the input activation after reorganization overcomes the inhibition, and neurons at the boundary of the lesion become more responsive to inputs previously stimulating lesioned neurons.

perilesion receptive fields.

Even after the lateral connections reorganize, the remaining unmasked input activation causes an imbalance in the network. Such activation forces the afferent weights to reorganize and respond better to inputs that were previously stimulating the lesioned zone. Gradually, the representation of the receptive surface within the lesion zone is taken over by the neurons around it (figure 11), and the cortical lesion is partly compensated for (figure 10*d*). The RF-LISSOM model predicts that for full compensation to occur, the lesion must be small enough so that neurons across the lesion are connected with excitatory connections. In that case they can act as neighbors on the map and the gap (such as that in figure 11*b*) can be closed.

The results with the RF-LISSOM model suggest two techniques to accelerate recovery following surgery in the sensory cortices. Normally, the recovery time after cortical surgery would include an initial period of regression due to the reorganization of inhibition, and gradual and slow compensa-



(a) Map before lesion (b) Reorganized map

Figure 11: **Reorganization of receptive fields after a cortical lesion.** Initially the centers of the Gaussian receptive fields are regularly and uniformly distributed over the retinal area. After lesion, the receptive fields of neurons near the periphery of the lesion move towards the center, partly compensating for the loss of function.

tion afterward. The first phase of regression could be ameliorated if a transient blocker of inhibitory neurotransmitters were applied locally around the surgical area. Neurons around the surgical area would then fire intensively because of reduced inhibition, and afferent connections would adapt rapidly to compensate for the lesion. Though the inhibition would strengthen when the blockade goes away, the pace of recovery would have been hastened. Secondly, the topographic map could be shifted (as in figure 11) even before surgery. This preshifting could be achieved by intensive and repetitive stimulation of the area expected to lose sensation and by sensory deprivation of its surroundings. The receptive fields would then have to move less to reach the final state, and the recovery would be faster.

The model shows that receptive fields are maintained dynamically by excitatory and inhibitory interactions within the cortex. The combined effect of afferent input, lateral excitation and lateral inhibition determine the responses of neurons. When the balance of excitation and inhibition is perturbed, neuronal response patterns change dynamically, and receptive fields appear to expand or decrease in size rapidly. If the perturbations are transient, they produce only transient changes in synaptic weight patterns and the topographic map does not shift much. However, if the perturbation persists for long, synaptic weight changes accumulate, and the topographic map reorganizes substantially. Such receptive field dynamics has been recently observed in the visual cortex (Pettet and Gilbert 1992). RF-LISSOM provides a computational explanation of why such dynamics occur, and illustrates the primary role of lateral interactions in cortical plasticity.

6 Modeling Low-Level Visual Functions

The lateral connections in the visual cortex are believed to be mediating several low-level visual phenomena such as tilt illusions, tilt aftereffects, segmentation and binding (Tolhurst and Thompson 1975; von der Malsburg and Singer 1988; Singer et al. 1990). In this section, two experiments are presented that rely on adaptation of connections after the network has self-organized. In the first

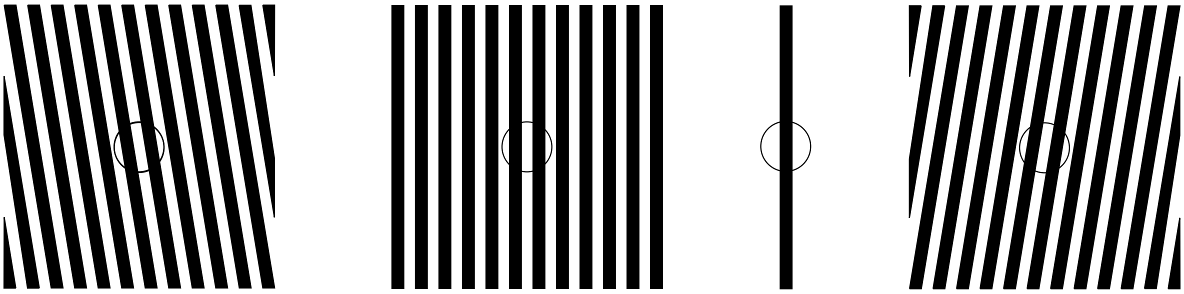


Figure 12: **Tilt aftereffect patterns.** Fixate upon the circle inside the square at the left or the right for at least thirty seconds, moving the eye slightly inside the circle to avoid developing strong afterimages. Now fixate upon the line or square at the center; they should appear tilted in the direction opposite to the previous pattern. (Adapted from Campbell and Maffei 1971.)

one, tilt aftereffects are shown to result from increased inhibition between neurons responsive to similar orientations. In the second, after extending the RF-LISSOM architecture with a more realistic spiking model of the neuron, fast adaptation of lateral connections is shown to establish synchronization between neurons representing the same input object, forming a basis for binding and segmentation.

6.1 Tilt Aftereffects

The tilt after-effect (TAE, Gibson and Radner 1937) is an intriguing visual phenomenon that has long been studied as an indication of the underlying mechanisms of the visual system. After staring at a pattern of tilted lines or gratings, subsequent lines appear to have a slight tilt in the opposite direction. Figure 12 demonstrates the effect. The effect resembles an afterimage from staring at a bright light, but it causes changes in orientation perception rather than color or brightness perception.

Several explanations have been proposed for the TAE over the years. Gibson and Radner (1937) hypothesized that the TAE results from perceptual normalization, where the subjective vertical and horizontal norms are modified by visual experience. This theory does not explain why the effect occurs also for vertical and horizontal adaptation angles. Köhler and Wallach (1944) postulated that activation of cortical sensory areas results in local electrical fields that spread electrotonically to nearby areas. Activated areas experience satiation, which results in displacement of subsequent perceptions to unsatiated areas. Such electrotonic mechanisms have not been found, and a more modern version of that theory was devised based on feature-detector fatigue: if neurons with orientation preferences close to the adaptation figure become fatigued as a result of activation, a different set of neurons will be activated for the test figure, accounting for the perceptual shift (Coltheart 1971; Sutherland 1961). This theory has also been discounted because such fatigue mechanisms have not been found, and cell firing has been shown neither necessary nor sufficient for adaptation (Vidyasagar 1990).

The prevailing theory for these effects attributes them to lateral interactions between orientation-specific feature-detectors in the primary visual cortex (Tolhurst and Thompson 1975). The lateral inhibitory connection strengths between activated neurons are believed to increase temporarily while an input pattern is inspected, causing changes in the perception of subsequent orientations. This occurs because the detectors are broadly tuned, and detectors for neighboring orientations also adapt somewhat. When a subsequent line of a slightly different orientation is presented, the most strongly responding units are now the ones with orientation preferences further from the adapting line, resulting in a change in the perceived angle.

Although the inhibition theory was proposed in the 1970s (in a slightly different form), it is computationally very expensive to simulate and has not been tested in a detailed model of cortical function. RF-LISSOM is such a model, and it is computationally feasible to test it on the orientation network of section 3.1. The results suggest that tilt aftereffects are not flaws in an otherwise well-designed system, but an unavoidable result of a self-organizing process that aims at producing an efficient, sparse encoding of the input through decorrelation.

To compare with results from human subjects, the tilt aftereffect was measured in the orientation map model as a function of the angular separation of the inducing and test figures. For adaptation on the inducing stimulus, the x and y coordinates of the center of a vertical Gaussian on the retina were fixed at the center of the retina. The learning rates α_I, α_A, and α_E were all set to 0.000005. All other parameters remained as in the self-organization of the orientation map.

To obtain a quantitative estimate of the aftereffect in the model, the perceived orientation in the model was measured as an activity-weighted sum of the pre-adaptation orientation preference of all active units. Perceived orientation was computed separately for each possible orientation of a test Gaussian at the center of the retina, both before and after adaptation. For a given angular separation of the adaptation stimulus and the test stimulus, the computed magnitude of the tilt aftereffect is the difference between the initial perceived angle and the one perceived after adaptation.

Figure 13 plots these differences after adapting to a vertical training line. For comparison, the figure also shows results from the most detailed recent data available for the tilt aftereffect in human foveal vision (Mitchell and Muir 1976). The results from the model are clearly consistent with those seen in human observers.

In the model, the amount of aftereffect increases with adaptation as connection weights gradually change. The S-shape of the curve in figure 13 results from the redistribution of inhibitory weights to connections linking active neurons. Since the most active neurons are those encoding vertical orientations, the response to vertical and nearly vertical lines decreases dramatically. This causes the *direct effect* of angle expansion at these angles ($0° - 35°$), just as predicted by the lateral inhibition theory of tilt aftereffects.

At the same time as the lateral inhibitory connections increase between active neurons, they must decrease between those neurons and inactive ones (i.e., those encoding distant orientations). This happens because the inhibitory weights are normalized to a constant sum in the model. The result is an *increase* in the response to lines from $45° - 90°$, or the *indirect effect*. Of the detectors activated for a test line in this range, those with preference closer to vertical are more active due to the reduction of inihibitory connections to them, while those encoding nearly orthogonal preference are unchanged. The perceived orientation thus shifts towards the training orientation. This result is a straightforward consequence of the lateral inhibition theory when the fact of limited synaptic resources is taken into account (cf. Purves 1988), which idea is a major contribution of the RF-LISSOM model.

The current results provide strong computational support for the theory that plastic lateral interactions are responsible for the tilt aftereffect as well as the initial self-organization of the orientation detectors. Through similar mechanisms, the model should also be able to explain simultaneous tilt illusions between overlapping or nearby stimuli (as proposed in Carpenter and Blakemore (1973)). However, the simple method of computing the perceived orientation as a

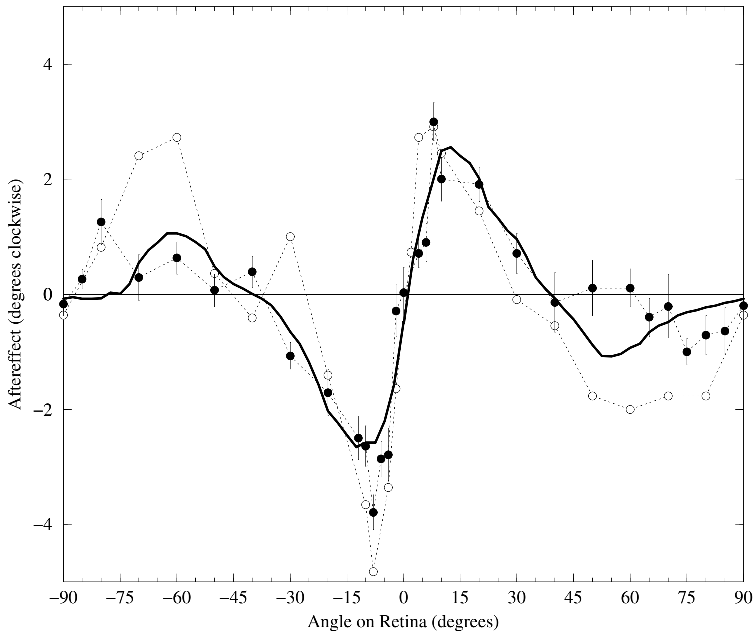


Figure 13: **Tilt aftereffect versus retinal angle.** The heavy line shows the magnitude of the tilt aftereffect in the RF-LISSOM model for each orientation after adapting to a vertical training line for 128 iterations. Positive values denote a clockwise change in the perceived orientation of the test line, and negative counterclockwise. The filled circles represent the average tilt aftereffect for a single human subject (DEM) from Mitchell and Muir (1976) over 10 trials; this subject had the most complete data of the four in the study. For each angle in each trial, the subject trained for 3 minutes on a sinusoidal grating of a given angle, then measured the tested the effect on a horizontal grating. Error bars indicate ±1 standard error of measurement from that study. The unfilled circles show the tilt affereffect under similar conditions for subject DWM from that study; this subject showed the largest indirect effect of the four from the study.

weighted average of activated units requires that the stimuli be spatially sufficiently separated so that their responses do not overlap. With the single-spot training inputs used for this version of the orientation map, connections do not develop over areas large enough to permit such separations, and thus the magnitude of the effect cannot be measured for the simultaneous case using this technique. To overcome this difficulty, it will be necessary to train the system with inputs that have longer-range correlations between similar orientations, such as sinusoidal gratings (representing objects with parallel lines). A version of the RF-LISSOM model trained on such patterns should be able to account for tilt illusions as well as tilt aftereffects. Although such experiments require even larger cortex and retina sizes, they should become practical in the near future.

In addition, many similar phenomena such as aftereffects of curvature, motion, spatial frequency, size, position, and color have been documented in humans (Barlow 1990). Since specific detectors for most of these features have been found, RF-LISSOM is expected to be able to account for them by the same process of decorrelation mediated by self-organizing lateral connections.

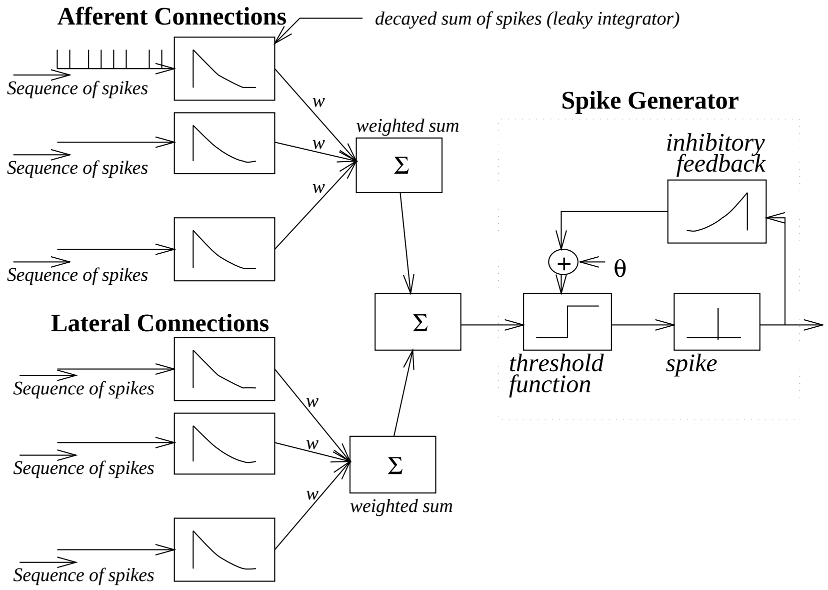


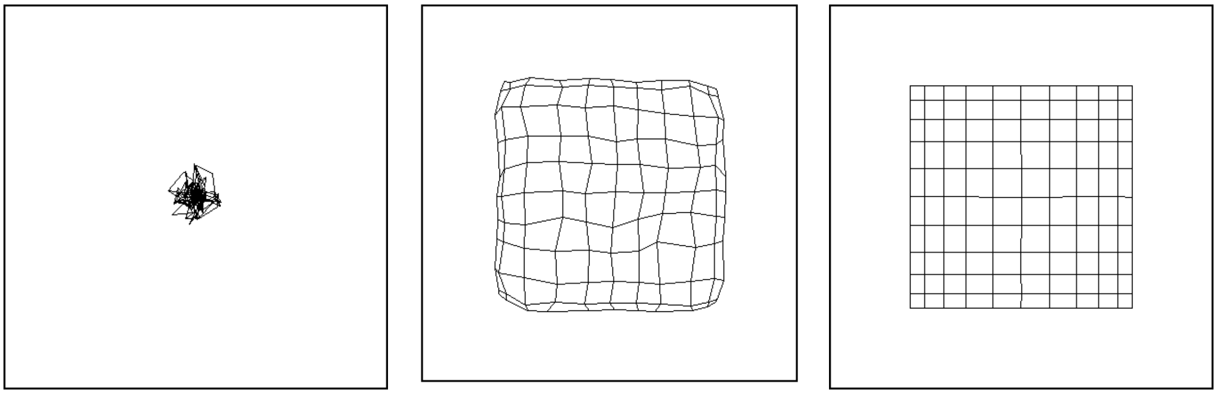
Figure 14: **The leaky integrator neuron in SLISSOM.** Leaky integrators at each synapse perform decayed summation of incoming spikes. The spike generator compares the weighted sum of the integrator outputs to a dynamic threshold, firing a spike if the sum is greater. Each spike increases the threshold, with exponential decay.

6.2 Segmentation and Binding

The RF-LISSOM model suggests that lateral connections play a central role in self-organization and in function of the visual cortex by establishing competitive and cooperative interactions between feature detectors. They may also mediate function at lower level by modulating the spiking behavior of neuronal groups. This way they could cause synchronization and desynchronization of spiking activity, thus mediating feature binding and segmentation. Such synchronization of neuronal activity emerges in the visual cortex of the cat when light bars of various length are presented (Gray and Singer 1987; Eckhorn et al. 1988; Gray et al. 1989). Several models have been proposed to explain this phenomenon (von der Malsburg 1987; von der Malsburg and Buhmann 1992; Eckhorn et al. 1990; Reitboeck et al. 1993; Wang 1996). The model of Reitboeck et al. (1993) is particularly interesting because of its sophisticated model of the neuron: the synapses are leaky integrators that sum incoming signals over time with exponential decay. A network of such neurons can segment multiple objects in a scene by synchronizing neuronal activity. Spikes of neurons representing the same object are synchronized, and those of neurons representing different objects are desynchronized.

The leaky integrator model of the spiking neuron can be integrated with the RF-LISSOM model of self-organization to model segmentation and binding. This extension of RF-LISSOM is called Spiking Laterally Interconnected Synergetically Self-Organizing Map, or SLISSOM. SLISSOM (1) forms a topological map from an initially random network through synergetic self-organization as before, and (2) generates synchronized and desynchronized neuronal activity that can be used for segmenting multiple objects in the scene.

Each connection in SLISSOM is a leaky integrator that performs decayed summation of incoming spikes, thereby establishing not only spatial summation, but also temporal summation of activity (figure 14). Each new spike is added to the sum of the previous ones, and the sum is exponentially



(a) Initial map (b) Final map (c) Ideal map

Figure 15: **Self-Organization of the SLISSOM Map.** (a) The afferent weights are initially randomized and their centers of gravity are about the same. (b) After 5500 iterations, the network forms a well-formed mapping of the input space, comparable to (c), the ideal grid where each node represents a gaussian receptive field located directly below the map unit.

decayed over time. The current sums are multiplied by the connection weight and added together [6] to form the net input to the neuron. The spike generator compares the net input to a threshold and decides whether to fire a spike. The threshold is a sum of two factors: the base threshold θ and the decayed sum of past spikes, formed by a similar leaky integrator as in the input synapses. Active spiking therefore increases the effective threshold, making further spiking less likely and keeping the activation of the system within a reasonable range.

Each connection has a queue that stores previous spikes. In calculating the postsynaptic potential, the latest spike has the value of 1.0 and older ones are decayed by $\frac{1}{e^{\lambda_q}}$, where λ_q is the decay parameter, as they are shifted through the queue. The inhibitory feedback loop in the spike generator (figure 14) is a similar queue that receives spikes from the spike generator itself, with decay $\frac{1}{e^{\lambda_s}}$.

The input to the network consists of squares of fixed size (3×3 in the experiments reported below), activated at random locations on the retina. Spikes are generated at the active retinal neurons and sent through the afferent connections to the cortical neurons. The net input σ_{ij} to the spike generator of the cortical neuron at location (i, j) at time t is calculated by summing the afferent and excitatory lateral contributions and subtracting the inhibitory lateral contributions:

$$\sigma_{ij}(t) = \gamma_{\mathrm{a}} \sum_{r_1,r_2} \xi_{r_1,r_2} \mu_{ij,r_1r_2} + \gamma_{\mathrm{e}} \sum_{k,l} \eta_{kl}(t-1) E_{ij,kl} - \gamma_{\mathrm{i}} \sum_{k,l} \eta_{kl}(t-1) I_{ij,kl}, \tag{6}$$

where γ_{a}, γ_{e}, and γ_{i} are the scaling factors for the afferent, excitatory, and inhibitory contributions, ξ_{r_1,r_2} is the decayed sum of the incoming queue from the retinal neuron at (r_1, r_2), μ_{ij,r_1r_2} is the corresponding afferent connection weight, $\eta_{kl}(t-1)$ is the decayed sum of the incoming queue from the map neuron at (k, l) at time $t-1$, and $E_{ij,kl}$ is the corresponding excitatory and $I_{ij,kl}$ the inhibitory lateral connection weight. The spike generator fires a spike if $\sigma_{ij} > \theta + \vartheta_{ij}$, where θ is the base threshold and ϑ_{ij} the output of the spike generator's leaky integrator.

[6] This differs from Eckhorn et al. (1990) and Reitboeck et al. (1993) who *multiplied* the weighted sums from afferent connections and those from lateral connections. Multiplying exerts better modulation on the neuronal activity, but disturbs self-organization by rapid fluctuation. In our experiments, modulation turned out to be possible with *additive* neurons as well.

Through equation 6, SLISSOM goes through a similar settling process as the RF-LISSOM network. The input is kept constant and the cortical neurons are allowed to exchange spikes. After a while, the neurons reach a stable rate of firing, and this rate is used to modify the weights. Both the afferent and the lateral weights are modified according to the Hebbian principle:

$$w_{ij,mn}(t) = \frac{w_{ij,mn}(t-1) + \alpha V_{ij} X_{mn}}{\mathcal{N}}, \tag{7}$$

where $w_{ij,mn}(t)$ is the connection weight between neurons (i,j) and (m,n), $w_{ij,mn}(t-1)$ is the previous weight, α is the learning rate (α_{a} for afferent, α_{E} for excitatory, and α_{i} for inhibitory connections), V_{ij} and X_{mn} are the average spiking rates of the neurons, and $\mathcal{N}$ is the normalization factor, $\sum_{mn}\left[w_{ij,mn}(t-1) + \alpha V_{ij} X_{mn}\right]^2$ for afferent connections and $\sum_{mn}\left[w_{ij,mn}(t-1) + \alpha V_{ij} X_{mn}\right]$ for lateral connections (cf. Sirosh and Miikkulainen 1994).

The SLISSOM experiment consists of two parts: (1) self-organization, and (2) object segmentation. During self-organization, single objects are shown to the network, and the lateral and afferent connection weights are adapted to form a topological map of the input space. After the network has stabilized, multiple objects are presented to the retina and the network segments the objects by temporally alternating the activity on the map.

The retina and the cortex both consisted of 11×11 units. Each neuron was fully connected to the retina. The afferent weights were initialized to be most sensitive to a 3×3 area on the retina, centered right below each neuron, and then 65% noise was added to their values. The lateral connection weights were randomly initialized. Inhibitory connections covered the whole map, and excitatory connections linked to a square area centered at each neuron, with initial radius of 8, gradually decreasing to 1. At the same time, the lateral inhibitory learning rate α_{i} gradually increased from 0.001 to 0.1. Slow adaptation in the beginning captures long-term correlations within the inputs, which is necessary for self-organization. Fast adaptation towards the end facilitates quick modulation of the activity necessary for segmentation.

During self-organization, single 3×3 square objects were presented to the network. The retinal neurons representing objects were spiking at each time step, and the settling consisted of 15 cycles of cortical activity update (equation 6). After settling, connection weights were modified according to equation 7. Each such presentation was counted as an iteration. After 5500 iterations, both the afferent and the lateral weights stabilized into smooth profiles. Afferent weights formed smooth Gaussian receptive fields most sensitive to input from the retinal neuron right below the map neuron, and the lateral weights formed smooth Mexican-hat profiles. Figure 15 shows the global organization of the map during the process. The final map (figure 15*b*) closely resembles the ideal map of the input space (15*c*).

Once the SLISSOM network had formed smooth and concentrated receptive fields and lateral interaction profiles, segmentation experiments were conducted on it. Several input objects (again, $3{\times}3$ squares) were presented to the retina at the same time, separated by at least 3 rows or columns of receptors. The objects constantly spiked on the retina for several hundred time steps, and the spike activity at the cortical layer was recorded. For each object, a separate 5×5 area on the map responded and the other areas remained silent. Segmentation is evident in the total number of spikes per time step (i.e. the multi-unit activity, or MUA; figure 16). The spikes within the same area are synchronized, and spikes across the different areas desynchronized. This result is very robust and works for different locations on the retina and for different numbers of objects.

Several studies have shown that fast adaptation of synaptic efficacy may mediate feature binding through temporal coding (von der Malsburg 1987; Wang 1996). Similarly in the experiments with SLISSOM, rapid adaptation of lateral weights was found crucial for oscillatory behavior. On the

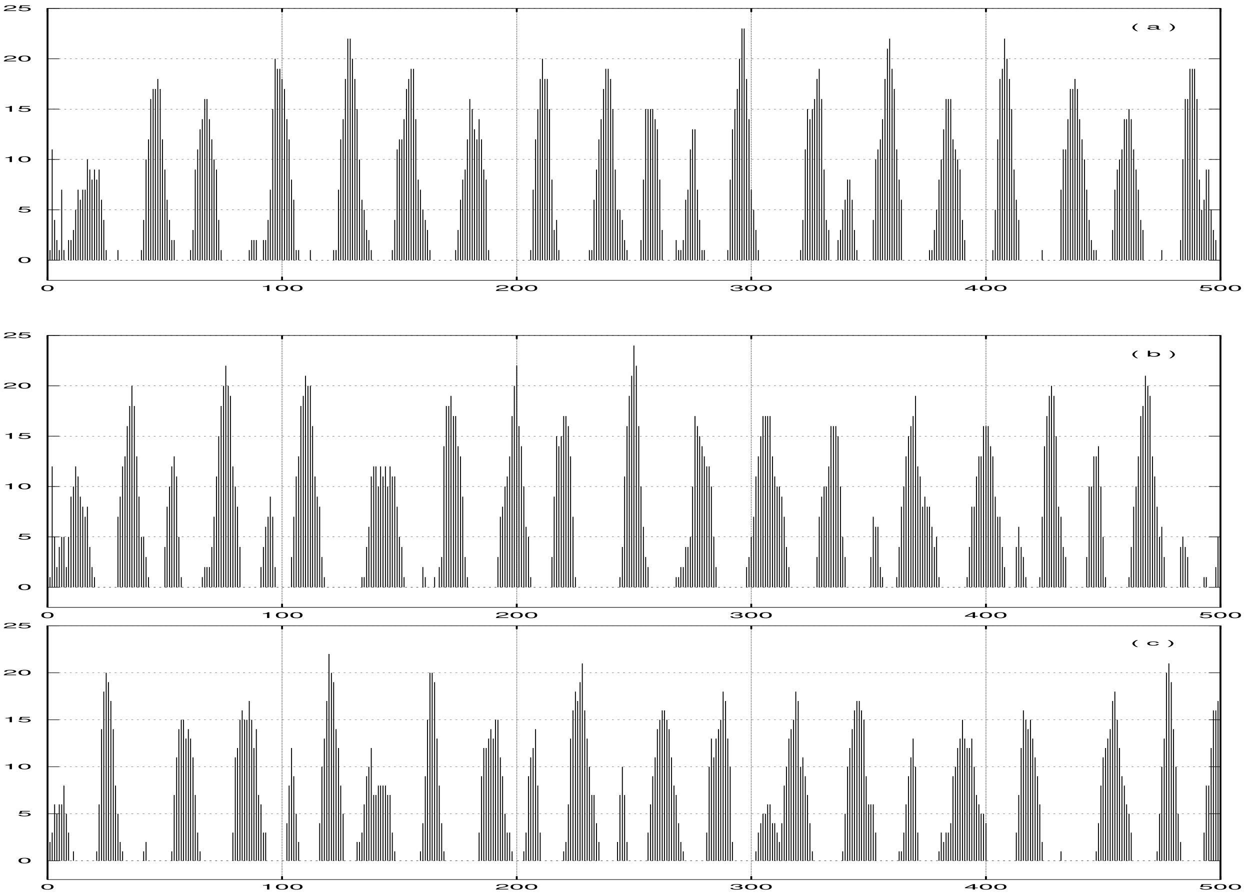


Figure 16: **The multi-unit activities of areas responding to three different objects.** The total number of spikes per time step in each of the three 5×5 areas are plotted over 500 time steps. Although initially there is simultaneous activity in all areas, they quickly desynchronize and activation rotates from one area to another.

other hand, self-organization requires slow adaptation so that long-term correlations can be learned. If the weights are initially random and change rapidly, they will fluctuate too much and an ill-formed map will result. One way to deal with this problem is to start out with a slow learning rate and gradually make learning faster. Such a change does not disturb the self-organization since the activity on the map becomes more consistent and predictable as the training goes on, and the need for keeping track of the long-term correlations disappears. The parameter α_{i} was increased to 0.1 at iteration 3500, allowing segmentation to occur, but that did not disrupt the already well-formed map (figure 15).

The MUAs show some overlap even when the input is successfully segmented (figure 16). This is due to the slightly overlapping receptive fields in the model. Gray et al. (1989) observed that in the cat visual cortex, strong phase-locking occurred when the receptive fields were clearly separate. Apparently when they overlap slightly, phase locking becomes less well defined at the edges. The overlap is unavoidable in the current small SLISSOM network, but could be reduced in larger-scale simulations. Simulations on a more detailed self-organized model of the visual cortex with orientation columns and patterned lateral connections might also account for perceptual phenomena such as Gestalt effects. For example, principles such as proximity, smoothness, and continuity of contours can be seen as regularities among image features. They would be encoded as correlations in the lateral connections, and could aid in determining object boundaries.

7 Discussion and Future Work

The RF-LISSOM model is based on three fundamental assumptions: First, vertical columns form the basic computational units in the visual cortex, and it is possible to approximate cortical function by modeling the 2-D layout of such columns instead of having to model each individual neuron separately. Since neurons in such columns are known to have similar response properties (Hubel and Wiesel 1959, 1965), this is a reasonable abstraction that makes computational modeling tractable.

Second, the lateral interactions between such columns are primarily excitatory in the short range and inhibitory in the long range. This assumption is more controversial since about 80% of the long-range connections are known to have excitatory synapses in the cortex (Gilbert et al. 1990). However, the connections in RF-LISSOM are intended to model the overall excitation/inhibition between columns, not the individual synapses. The long-range excitatory connections could link to inhibitory cells in simple local circuits, and the overall effect could be inhibitory. Optical imaging and electrophysiological studies indeed suggest that there is substantially more inhibition in the cortex than predicted by the synapse types (Grinvald et al. 1994; Hata et al. 1993; Hirsch and Gilbert 1991). The nature of long-range interactions remains controversial at this point. The RF-LISSOM model makes the computational prediction that the overall effect of long-range interactions on the column is inhibitory; otherwise self-organization does not occur in the model.

Third, the input patterns are sufficiently sparse so that distinct feature detectors may develop. As the number of spots in the images are increased, the organization of the map gradually degrades. When there is more overlap in the input more complex and random visual features are created, and the network attempts to represent these features as well. At first this seems incompatible with the dense nature of natural input images. However, there are two ways why such sparse input patterns might actually be realistic: (1) it is possible that much of the self-organization occurs prenatally based on traveling activity waves in the retina, such as those observed in the ferret and in the cat (Meister et al. 1991; Wong et al. 1993). Such waves appear to have the necessary sparse structure. (2) Although the natural images may be dense originally, after the edge-enhancement mechanisms in the retina, they also have a sparse structure, with linear features at the local level (Wandell 1995). If retinal preprocessing mechanisms are added, it is likely that the RF-LISSOM network can develop realistic orientation, ocular dominance, and size maps from natural images.

The main difference between RF-LISSOM and other recent models of cortical self-organization (Amari 1980; Bienenstock et al. 1982; Durbin and Mitchison 1990; Erwin et al. 1995; Goodhill 1993; Grossberg 1976; Kohonen 1982; Linsker 1990; Miller 1994a; Miller et al. 1989; Obermayer et al. 1992; Ritter et al. 1992; Shouval 1995; Stryker et al. 1990; Tanaka 1993; Willshaw and von der Malsburg 1976) is that in RF-LISSOM, the lateral connections between feature detectors are explicitly modeled. Therefore, the model can explain how not only the observed afferent structures but also the lateral connectivity patterns emerge as a result of self-organization. Several testable predictions can be made based on the model, such as the connectivity at the pinwheel centers and fractures. The explicit lateral connections also make it possible to test the decorrelation hypothesis, possible plasticity mechanisms, and the inhibitory theory of tilt aftereffects computationally.

The tilt aftereffect and segmentation and binding in RF-LISSOM are modeled using the same adaptation processes as the self-organization. However, these processes operate in different time scales, and most likely would have to be based on two different processes in the cortex. The mechanisms underlying perceptual phenomena may be a short-time-scale, temporary version of the self-organizing processes that capture longer-term correlations. These fast weights could act as a small additive or multiplicative term on a relatively static long-term weight. Each inhibitory weight w could be represented as $w_o + \Delta_w$, where the w_o portion would be comparatively static, keeping its value indefinitely in darkness and changing with a long time constant. The Δ_w term, on the other hand, would adapt and decay very quickly, presumably to represent the short-term correlations between image elements. The tilt aftereffect and the segmentation and binding phenomena would be a result of such short-term adaptation.

Future work on RF-LISSOM models focuses on three areas: scaling the model up to more realistic scales, extending the functional simulations, and modeling self-organization of hierarchical representations. The models presented in this article were simplified in two ways: (1) there were separate models for each experiment (although with very similar networks and parameters), and (2) simple light spots were used as inputs instead of realistic images. This way, it was possible to study each individual aspect separately, and show how cortical structure and function depends on the structure of the inputs. With complex natural images, it would have been difficult to perform such an analysis and identify the crucial parameters that are responsible for each effect. In the future, this work should be extended to more complex inputs, to study the combined organization of various features in the cortex. One problem is that more computational power will be needed for such simulations than is currently available. When more features have to be represented on the map at the same time, the maps need to be larger. Perhaps 512×512 units are needed for simultaneous orientation, ocular dominance, and size preferences to develop. The requirement will grow perhaps four-fold if gratings such as those used in tilt illusions in humans are to be used on the organized network. The leaky integrator model of the neuron in turn doubles or triples the resource requirements. This is beyond the current computational resources, but could be available in the next few years. It may also be possible to approximate such large networks with smaller ones and address issues in processing realistic input with realistic-size networks sooner.

Second main direction of future work consists of extending the functional simulations on the organized model. In addition to tilt aftereffects, other figural aftereffects and also simultaneous tilt illusions can be modeled, as outlined in section 6.1. Binding and segmentation can be studied with more complex images, with multiple, differently-shaped, and moving objects. Since the lateral connections in RF-LISSOM store long-range activity correlations, it may be possible to model perceptual grouping, or gestalt, effects such as continuity of contours, proximity, and smoothness.

The time course of the organization of representations could be matched with developmental data on young animals. The network develops coarse representations of average inputs at first and gradually makes the map more accurate. For example, when it becomes computationally feasible to form maps of spatial frequency selectivity, it should be possible to model early perceptual development, where infants first are aware of only low spatial frequencies, but develop sensitivity to higher frequencies during the first few months.

Fahle et al. (1995) and Weiss et al. (1993) found that performance in hyperacuity tasks, such as deciding whether two lines of same orientation are separated by a small perpendicular offset, improves with practice even without feedback. The effect is specific to position and orientation, but transfers between eyes to some degree. Shiu and Pashler (1992) reported similar results for orientation discrimination tasks, although they found that the effect also depends on cognitive factors. The RF-LISSOM model should be able to account for such psychophysical learning phenomena. The active feature detectors and lateral connections between them would adapt, resulting in representation and discrimination of smaller differences. Hyperacuity and discrimination phenomena might also form a good testbed for extensions of RF-LISSOM that include feedback from higher cortical levels.

Third, the RF-LISSOM model could be extended to learning of hierarchical visual representations, to discover increasingly complex structures in the input. Other self-organizing algorithms such as the self-organizing map (Kohonen 1982, 1989, 1995) do not directly lend themselves to such hierarchical networks. A higher level map will pick up essentially the same features of the input as the lower level map, and represent the same level of information. However, the lateral interactions in RF-LISSOM eliminate redundant activity, and the map encodes the input in terms of unique feature detectors. A higher-level RF-LISSOM network would learn the correlations between the first-level of feature detectors, and thereby form more complex detectors. Thus, by combining self-organization and decorrelation in one algorithm, a mechanism is obtained that can be applied in multiple stages to extract increasingly complex structures from the input. Therefore, LISSOM networks can possibly be cascaded in multiple levels to model the hierarchical representations in the visual cortex.

Large-scale computational models like RF-LISSOM are likely to play a significant role in future research on the structure and function of the visual cortex, because only computational models can be rigorous enough and detailed enough to verify ideas about how the visual cortex operates. At the current rate of technological progress, computers will be powerful enough to simulate the visual cortex at realistic details within a decade or two. Such models should provide a fundamental understanding of higher brain function and perception, and suggest novel algorithms for pattern recognition, computer vision and sensory processing.

8 Conclusion

The RF-LISSOM model demonstrates how a variety of phenomena in the visual cortex can be explained by local cooperation and global competition whose specific form is learned from correlations in the input. Both the feature detectors and the lateral interactions self-organize simultaneously and synergetically based on a single Hebbian learning mechanism, resulting in patterns of selectivity and lateral connectivity some of which have already been observed in the visual cortex, and others that constitute testable predictions of the model. Such self-organization stores long-range activity correlations between units into the lateral connections, and during visual processing, this information is used to eliminate redundancies and to form an efficient sparse coding of the input. Many aspects of cortical plasticity can be explained by the model, such as reorganization after retinal lesions and that after cortical lesions, and the model suggests ways to hasten recovery from such lesions. The simulated reorganizations are reversible, and demonstrate how a topographic map can be maintained in a dynamic equilibrium with extrinsic and intrinsic inputs.

The model demonstrates that not only cortical structure, but also many of its functional characteristics could be emergent properties driven by activity-dependent self-organization. Tilt aftereffects in the model result from the same decorrelating process that is responsible for the initial development of the orientation map. Combined with the leaky integrator model of the spiking neuron, adaptation in RF-LISSOM can also account for segmentation by synchronization. These results suggests that both the development and function of the primary visual cortex could eventually be understood computationally based on Hebbian adaptation of laterally interconnected leaky integrator neurons.

Acknowledgments

Thanks to Rob Goldstone, Philippe Schyns, and Jim Tanaka for insightful comments on an earlier draft of this paper. This research was supported in part by National Science Foundation under grant #IRI-9309273. Computing time for the simulations was provided by the Pittsburgh Supercomputing Center under grants IRI930005P and IRI940004P, and by a High Performance Computer Time Grant from the University of Texas at Austin.

Note

Various LISSOM software, visualizations, and papers are available at http://www.cs.utexas.edu/users/nn.

References

Amari, S.-I. (1980). Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, 42:339–364.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Barlow, H. B. (1985). The twelfth Bartlett memorial lecture: The role of single neuron in the psychology of perception. *Quarterly Journal of Experimental Psychology*, 37A:121–145.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

Barlow, H. B. (1990). A theory about the functional role and synaptic mechanism of visual aftereffects. In Blakemore, C., editor, *Vision: Coding and Efficiency*, 363–375. New York: Cambridge University Press.

Bauman, L., and Bonds, A. (1991). Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex. *Vision Research*, 31(6):933–944.

Bienenstock, E. L., N.Cooper, L., and W.Munro, P. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2:32–48.

Blakemore, C., and Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, 228:477–478.

Blakemore, C., and van Sluyters, R. C. (1975). Innate and environmental factors in the development of the kitten's visual cortex. *Journal of Physiology (London)*, 248:663–716.

Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *The Journal of Neuroscience*, 12:3139–3161.

Blasdel, G. G., and Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, 321:579–585.

Burkhalter, A., Bernardo, K. L., and Charles, V. (1993). Development of local circuits in human visual cortex. *Journal of Neuroscience*, 13:1916–1931.

Callaway, E. M., and Katz, L. C. (1990). Emergence and refinement of clustered horizontal connections in cat striate cortex. *Journal of Neuroscience*, 10:1134–1153.

Callaway, E. M., and Katz, L. C. (1991). Effects of binocular deprivation on the development of clustered horizontal connections in cat striate cortex. *Proceedings of the National Academy of Sciences, USA*, 88:745–749.

Campbell, F., Cooper, G., and Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology (London)*, 203:223–235.

Campbell, F. W., and Maffei, L. (1971). The tilt aftereffect: A fresh look. *Vision Research*, 11:833–840.

Carpenter, R. H. S., and Blakemore, C. (1973). Interactions between orientations in human vision. *Experimental Brain Research*, 18:287–303.

Coltheart, M. (1971). Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, 78(2):114–121.

Dalva, M. B., and Katz, L. C. (1994). Rearrangements of synaptic connections in visual cortex revealed by laser photostimulation. *Science*, 265:255–258.

De Valois, K. K., and Tootell, R. B. H. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *Journal of Physiology (London)*, 336:359–376.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22:545–559.

Durbin, R., and Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, 343:644–647.

Eckhorn, R., Bauer, R., Jordan, W., Kruse, M., Munk, W., and Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130.

Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Computation*, 2:293–307.

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural Computation*, 7(3):425–468.

Fahle, M., Edelman, S., and Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, 35:3003–3013.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 4:2379–2394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6:559–601.

Fisken, R. A., Garey, L. J., and Powell, T. P. S. (1975). The intrinsic, association and commissural connections of area 17 of the visual cortex. *Philosophical Transactions of the Royal Society of London Series B*, 272:487.

Fitzpatrick, D., Schofield, B. R., and Strote, J. (1994). Spatial organization and connections of iso-orientation domains in the tree shrew striate cortex. In *Society for Neuroscience Abstracts*, vol. 20, 837.

Gibson, J. J., and Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology*, 20:453–467.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9:1–13.

Gilbert, C. D., Hirsch, J. A., and Wiesel, T. N. (1990). Lateral interactions in visual cortex. In *Cold Spring Harbor Symposia on Quantitative Biology, Volume LV*, 663–677. Cold Spring Harbor Laboratory Press.

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3:1116–1133.

Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and cortico-cortical connections in cat visual cortex. *Journal of Neuroscience*, 9:2432–2442.

Goodhill, G. (1993). Topography and ocular dominance: A model exploring positive correlations. *Biological Cybernetics*, 69:109–118.

Gray, C. M., Konig, P., Engel, A., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337.

Gray, C. M., and Singer, W. (1987). Stimulus specific neuronal oscillations in the cat visual cortex: A cortical functional unit. In *Society of Neuroscience Abstracts*, vol. 13, 404.3.

Grinvald, A., Lieke, E. E., Frostig, R. D., and Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14:2545–2568.

Grossberg, S. (1976). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics*, 21:145–159.

Gustafsson, B., and Wigström, H. (1988). Physiological mechanisms underlying long-term potentiation. *Trends in Neurosciences*, 11:156–162.

Hata, Y., Tsumoto, T., Sato, H., Hagihara, K., and Tamura, H. (1993). Development of local horizontal interactions in cat visual cortex studied by cross-correlation analysis. *Journal of Neurophysiology*, 69:40–56.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory.* New York: Wiley.

Hirsch, H., and Spinelli, D. (1970). Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168:869–871.

Hirsch, J. A., and Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *The Journal of Neuroscience*, 11:1800–1809.

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148:574–591.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154.

Hubel, D. H., and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243.

Hubel, D. H., and Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, 158:267–294.

Hubel, D. H., Wiesel, T. N., and LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London [Biology]*, 278:377–409.

Kaas, J. H. (1991). Plasticity of sensory and motor maps in adult animals. *Annual Review of Neuroscience*, 14:137–167.

Kapadia, M. K., Gilbert, C. D., and Westheimer, G. (1994). A quantitative measure for short-term cortical plasticity in human vision. *Journal of Neuroscience*, 14:451–457.

Katz, L. C., and Callaway, E. M. (1992). Development of local circuits in mammalian visual cortex. *Annual Review of Neuroscience*, 15:31–56.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kohonen, T. (1989). *Self-Organization and Associative Memory.* Berlin; Heidelberg; New York: Springer. Third edition.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6:895–905.

Kohonen, T. (1995). *Self-Organizing Maps.* Berlin; Heidelberg; New York: Springer.

Köhler, W., and Wallach, H. (1944). Figural after-effects: an investigation of visual processes. *Proceedings of the American Philosophical Society*, 88:269–357.

Linsker, R. (1990). Perceptual neural organization: Some approaches based on network models and information theory. *Annual Review of Neuroscience*, 13:257–281.

Löwel, S. (1994). Ocular dominance column development: Strabismus changes the spacing of adjacent columns in cat visual cortex. *Journal of Neuroscience*, 14(12):7451–7468.

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255:209–212.

Luhmann, H. J., Martínez Millán, L., and Singer, W. (1986). Development of horizontal intrinsic connections in cat striate cortex. *Experimental Brain Research*, 63:443–448.

Malach, R., Amir, Y., Harel, M., and Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in the primate striate cortex. *Proceedings of the National Academy of Sciences, USA*, 90:10469–10473.

McCasland, J. S., Bernardo, K. L., Probst, K. L., and Woolsey, T. A. (1992). Cortical local circuit axons do not mature after early deafferentation. *Proceedings of the National Academy of Sciences, USA*, 89:1832–1836.

Meister, M., Wong, R., Baylor, D., and C.J.Shatz (1991). Synchronous bursts of action-potentials in the ganglion cells of the developing mammalian retina. *Science*, 252:939–943.

Merzenich, M. M., Recanzone, G. H., Jenkins, W. M., and Grajski, K. A. (1990). Adaptive mechanisms in cortical networks underlying cortical contributions to learning and nondeclarative memory. In *Cold Spring Harbor Symposia on Quantitative Biology, Vol. LV*, 873–887. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Miller, K. D. (1994a). 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. *Journal of Neuroscience*, 14:409–441.

Miller, K. D. (1994b). The role of constrains in Hebbian learning. *Neural Computation*, 6:98–124.

Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: Analysis and simulation. *Science*, 245:605–615.

Mitchell, D. E., and Muir, D. W. (1976). Does the tilt aftereffect occur in the oblique meridian? *Vision Research*, 16:609–613.

Obermayer, K., Blasdel, G. G., and Schulten, K. J. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Physical Review A*, 45:7568–7589.

Obermayer, K., Ritter, H. J., and Schulten, K. J. (1990a). Large-scale simulation of a self-organizing neural network. In *Proceedings of the International Conference on Parallel Processing in Neural Systems and Computers (ICNC)*. New York: Elsevier.

Obermayer, K., Ritter, H. J., and Schulten, K. J. (1990b). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences, USA*, 87:8345–8349.

Pettet, M. W., and Gilbert, C. D. (1992). Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences, USA*, 89:8366–8370.

Purves, D. (1988). *Body and brain: a trophic theory of neural connections*. Cambridge, MA: Harvard University Press.

Reitboeck, H., Stoecker, M., and Hahn, C. (1993). Object separation in dynamic neural networks. In *Proceedings of the IEEE International Conference on Neural Networks* (San Francisco, CA), vol. 2, 638–641.

Ritter, H. J., Martinetz, T., and Schulten, K. J. (1992). *Neural Computation and Self-Organizing Maps: An Introduction*. Reading, MA: Addison-Wesley.

Schwark, H. D., and Jones, E. G. (1989). The distribution of intrinsic cortical axons in area 3b of cat primary somatosensory cortex. *Experimental Brain Research*, 78:501–513.

Shiu, L.-P., and Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, 52:582–588.

Shouval, H. (1995). *Formation and organization of receptive fields, with an input environment composed of natural scenes.* PhD thesis, Department of Physics, Brown University.

Silverman, M. S., Grosof, D. H., De Valois, R. L., and Elfar, S. D. (1989). Spatial frequency organization in primate striate cortex. *Proceedings of the National Academy of Sciences, USA*, 86:711–715.

Singer, W., Gray, C., Engel, A., König, P., Artola, A., and Bröcher, S. (1990). Formation of cortical cell assemblies. In *Cold Spring Harbor Symposia on Quantitative Biology, Vol. LV*, 939–952. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sirosh, J. (1995). *A Self-Organizing Neural Network Model of the Primary Visual Cortex.* PhD thesis, Department of Computer Sciences, The University of Texas at Austin, Austin, TX.

Sirosh, J., and Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71:66–78.

Sirosh, J., and Miikkulainen, R. (1996). Self-organization and functional role of lateral connections and multisize receptive fields in the primary visual cortex. *Neural Processing Letters*, 3:39–48.

Sirosh, J., and Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*. In press.

Sirosh, J., Miikkulainen, R., and Bednar, J. A. (1996). Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

Stryker, M., Chapman, B., Miller, K., and Zahs, K. (1990). Experimental and theoretical studies of the organization of afferents to single orientation columns in visual cortex. In *Cold Spring Harbor Symposia on Quantitative Biology*, vol. LV, 515–527. Cold Spring Harbor Laboratory Press.

Sutherland, N. S. (1961). Figural after-effects and apparent size. *Quarterly Journal of Psychology*, 13:222–228.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262:685–688.

Tolhurst, D. J., and Thompson, P. G. (1975). Orientation illusions and aftereffects: Inhibition between channels. *Vision Research*, 15:967–972.

Tootell, R. B., Silverman, M. S., and De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex. *Science*, 214:813–815.

Tootell, R. B., Silverman, M. S., Hamilton, S. L., Switkes, E., and De Valois, R. L. (1988). Functional anatomy of macaque striate cortex. v. spatial frequency. *Journal of Neuroscience*, 8:1610–1624.

Vidyasagar, T. R. (1990). Pattern adaptation in cat visual cortex is a co-operative phenomenon. *Neuroscience*, 36:175–179.

Vidyasagar, T. R., and Mueller, A. (1994). Function of GABA inhibition in specifying spatial frequency and orientation selectivities in cat striate cortex. *Experimental Brain Research*, 98(1):31–38.

von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, 15:85–100.

von der Malsburg, C. (1987). Synaptic plasticity as basis of brain organization. In Changeux, J.-P., and Konishi, M., editors, *The Neural and Molecular Bases of Learning*, 411–432. New York: Wiley.

von der Malsburg, C. (1990). Network self-organization. In Zornetzer, S. F., Davis, J. L., and Lau, C., editors, *An Introduction to Neural and Electronic Networks*, chapter 22, 421–432. New York: Academic Press.

von der Malsburg, C., and Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, 67:233–242.

von der Malsburg, C., and Singer, W. (1988). Principles of cortical network organization. In Rakic, P., and Singer, W., editors, *Neurobiology of Neocortex*, 69–99. New York: Wiley.

Wandell, B. A. (1995). *Foundations of Vision.* Sunderland, Massachusetts: Sinauer Associates, Inc.

Wang, D. (1996). Synchronous oscillations based on lateral connections. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function.* Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

Weiss, Y., Edelman, S., and Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718.

Willshaw, D. J., and von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London B*, 194:431–445.

Wong, R., Meister, M., and Shatz, C. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, 11(5):923–938.