Scaling self-organizing maps to model large cortical networks

James A. Bednar, Amol Kelkar, and Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712
jbednar, amol, risto@cs.utexas.edu

Abstract

Self-organizing computational models with specific intracortical connections can explain many functional features of visual cortex, such as topographic orientation and ocular dominance maps. However, due to their computational requirements, it is difficult to use such detailed models to study large-scale phenomena like object segmentation and binding, object recognition, tilt illusions, optic flow, and fovea–periphery interaction. This paper introduces two techniques that make large simulations practical. First, a set of general linear scaling equations for the RF-LISSOM self-organizing model is derived and shown to result in quantitatively equivalent maps over a wide range of simulation sizes. This capability makes it possible to debug small simulations and then scale them up to larger simulations only when needed. The scaling equations also facilitate the comparison of biological maps and parameters between individuals and species with different brain region sizes. Second, the equations are combined into a new growing map method called GLISSOM, which dramatically reduces the memory and computational requirements of large self-organizing networks. With GLISSOM it should be possible to simulate all of human V1 at the single-column level using existing supercomputers, making detailed computational study of large-scale phenomena possible.

1 Introduction

Computational models of the self-organization in the visual cortex have shown that input-driven development can explain much of its topographic organization, such as retinotopy, orientation preference, and ocular dominance, as well as many of its functional properties, such as short-range contour segmentation and binding (Grossberg 1976; Kohonen 1989; von der Malsburg 1973; see Erwin et al. 1995; Swindale 1996 for review). However, other important phenomena have remained out of reach because they require too much computation time and memory to simulate. These phenomena, such as orientation interactions between spatially separated stimuli and long-range visual contour and object integration, are thought to arise out of specific lateral interactions between large numbers of neurons over a wide cortical area (Gilbert et al. 1996). Simulating such behavior requires an enormous number of specific, modifiable connections. Currently-practical methods can only model intracortical interactions abstractly (e.g. SOM, Erwin et al. 1992; Kohonen 1989; Obermayer et al. 1990), and thus cannot be used for such investigations.

In this paper we present two interrelated techniques for making detailed large-scale simulations practical. First, we derive a set of linear scaling equations that, when given a small-scale simulation, make it possible to determine the parameter settings necessary to perform a large-scale simulation. The original and scaled simulations have quantitatively-equivalent map-level and neuron-level organization; the larger map will just have more detail. Such a correspondence makes it possible to develop a small-scale simulation first using available hardware, then scale it up to study specific phenomena that require a larger map. The scaling equations can also help tie parameters from small models to experimental measurements in larger systems, help determine simulation sizes needed for realistic simulations, and allow comparison of species or individuals with brain regions of different sizes.

Second, we present a modeling approach called GLISSOM that allows much larger networks to be simulated in a given computation time and in a given amount of memory. The simulations begin with a small network, which is gradually scaled up as it self-organizes. This approach is effective for two reasons: (1) pruning-based self-organizing models tend to have peak computational and memory requirements at the beginning of training, and (2) self-organization tends to proceed in a global-to-local fashion, with large-scale order established first, followed by more detailed local self-organization (as found in experimental animals; Chapman et al. 1996). Thus small maps, which are much quicker to simulate and take less memory, can be used to establish global order, with larger maps used only to achieve more detailed structure.

Although the primary motivation for GLISSOM is computational, the scaling process is also well-motivated biologically, since it represents the integration of new neurons into an existing region during development. Recent experimental results suggest that new neurons continue to be added even in adulthood in many areas of primate cortex (Gould et al. 1999). Moreover, many of the neurons in the immature cortex corresponding to GLISSOM's early stages have not yet begun to make functional connections, having only recently migrated to their final positions (Purves 1988). Thus the scaleup

Preprint version. Available as Technical Report AI-2000-285 August 2001
http://www.cs.utexas.edu/users/nn/

procedure in GLISSOM corresponds to the gradual process of incorporating those neurons into the partially-organized map.

In the next section the model used in these simulations is introduced, and in section 3 scaling equations for it are derived and shown to achieve matching results over a wide range of simulation sizes. In section 4 the GLISSOM scaleup procedure is introduced and shown to greatly reduce simulation time and memory requirements while achieving results similar to the original model. Section 5 shows calculations that suggest that with GLISSOM it should be possible to simulate all of human V1 at the single-column level using existing supercomputers. The remaining sections discuss how the scaling equations relate to biological systems and how they can be used to simulate larger, more realistic systems that would otherwise be intractable.

2 RF-LISSOM model of the visual cortex

The scaling equations and GLISSOM are based on the RF-LISSOM (Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map) computational model of cortical maps. RF-LISSOM has been successfully used to model the development of ocular dominance and orientation maps, as well as low-level visual phenomena in the adult, such as tilt aftereffects and short-range segmentation and binding (Bednar and Miikkulainen 2000; Choe and Miikkulainen 1998; Miikkulainen et al. 1997; Sirosh and Miikkulainen 1994; Sirosh et al. 1996). We will first describe the architecture of the RF-LISSOM model, and then later present our extensions that allow scaling the network.

RF-LISSOM focuses on the two-dimensional organization of the cortex, so each "neuron" in the model cortex corresponds to a vertical column of cells through the six layers of the primate cortex. The cortical network is modeled with a sheet of interconnected neurons and the retina with a sheet of retinal ganglion cells (figure 1). Neurons receive afferent connections from broad overlapping circular patches on the retina. (Since the lateral geniculate nucleus (LGN) accurately reproduces the receptive fields of the retina, it has been bypassed to simplify the model.) The $N \times N$ network is projected on to the $R \times R$ retinal ganglion cells, and each neuron is connected to ganglion cells in an area of radius r_{A} around its projection. Thus, neurons at a particular cortical location receive afferents from a corresponding location on the retina, i.e. its anatomical receptive field (RF). Additional ganglion cells are included around the borders so that every neuron will have a complete set of afferent connections. For an example set of weights, see figure 9*a-c* in section 4.1.

In addition to the afferent connections, each neuron has reciprocal excitatory and inhibitory lateral connections with itself and other neurons. Lateral excitatory connections are short-range, connecting each neuron with itself and its close neighbors. Lateral inhibitory connections run for comparatively long distances, but also include connections to the neuron itself and to its neighbors.[1]

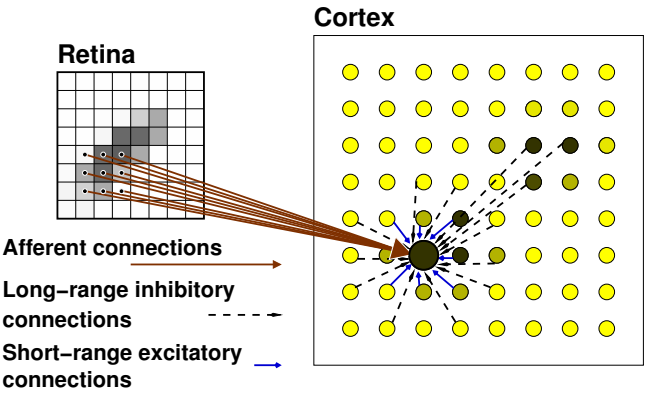


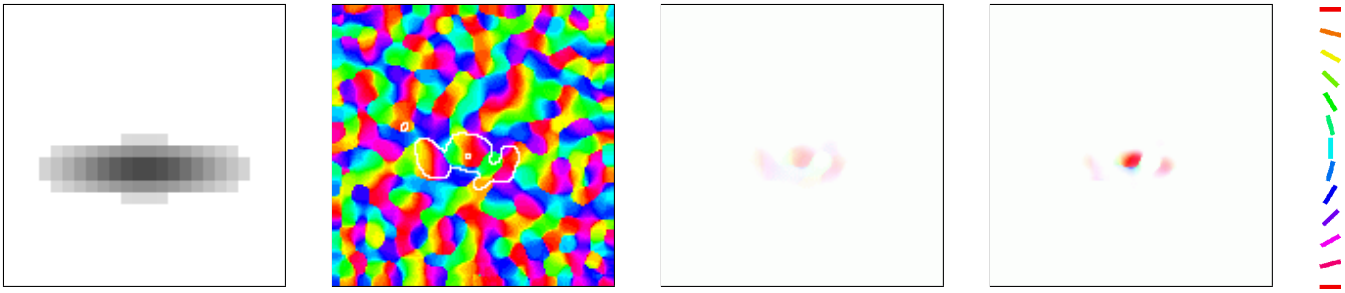
Figure 1: **Architecture of the RF-LISSOM network.** A small RF-LISSOM network and retina are shown, along with connections to a single neuron (shown as the large circle). The input is an oriented Gaussian activity pattern on the retinal ganglion cells (shown by grayscale coding); the LGN is bypassed for simplicity. The afferent connections form a local anatomical receptive field (RF) on the simulated retina. Neighboring neurons have different but highly overlapping RFs. Each neuron computes an initial response as a scalar (dot) product of its receptive field and its afferent weight vector, i.e. a sum of the product of each weight with its associated receptor. The responses then repeatedly propagate within the cortex through the lateral connections and evolve into activity "bubbles". After the activity stabilizes, weights of the active neurons are adapted using a normalized Hebbian rule.

The afferent weights are initially set to random values, and the lateral weights are preset to a smooth Gaussian profile. The connections are then organized through an unsupervised learning process. For an orientation map, the input for the learning process consists of 2-D ellipsoidal Gaussian patterns representing retinal ganglion cell activations (figure 2*a*); each pattern is presented at a random orientation and position. At each training step, neurons start out with zero activity. The initial response η_{ij} of neuron (i, j) is calculated as a weighted sum of the retinal activations:

$$\eta_{ij} = \sigma\left(\sum_{a,b} \xi_{ab}\mu_{ij,ab}\right), \qquad (1)$$

where ξ_{ab} is the activation of retinal ganglion (a, b) within the receptive field of the neuron, $\mu_{ij,ab}$ is the corresponding afferent weight, and σ is a piecewise linear approximation of the sigmoid activation function. The response evolves over

[1] For high-contrast inputs, long-range interactions must be inhibitory for proper self-organization to occur (Sirosh 1995). Optical imaging and electrophysiological studies have indeed shown that long-range column-level interactions in the cortex are inhibitory at high contrasts, even though individual long-range lateral connections between neurons are primarily excitatory (Grinvald et al. 1994; Hirsch and Gilbert 1991; Weliky et al. 1995). The model uses explicit inhibitory connections for simplicity since all inputs used are high-contrast, and since it is such inputs that primarily drive adaptation in a Hebbian model.



(*a*) Input pattern (*b*) Orientation map (*c*) Initial activity (*d*) Settled activity

Figure 2: **Orientation map activation in RF-LISSOM (*color figure*).** The orientation color key at the far right applies to all of the plots in (*b-d*), and to all similar plots in this paper. After being trained on inputs like the one in (*a*) with random positions and orientations, a 144×144 RF-LISSOM network developed the orientation map shown in (*b*), which is similar to those found in experimental animals (Blasdel 1992). Each neuron in the map is colored according to the orientation it prefers. The white outline shows the extent of the patchy self-organized lateral inhibitory connections of one neuron (marked with a white square), which has a horizontal orientation preference. The strongest long-range connections of each neuron are extended along its preferred orientation and link columns with similar orientation preferences, avoiding those with very different preferences. The brightness of the colors in (*c*,*d*) shows the strength of activation for each neuron to pattern (*a*). The initial response of the organized map is spatially broad and diffuse (*c*), like the input, and its cortical location around a horizontal line near the center of the cortex indicates that the input is horizontally extended near the center of the retina. The response is patchy because the neurons that encode orientations far from the horizontal do not respond (compare *c* and *d* to *b*). After the network settles through lateral interactions, the activation is much more focused, but the activated neurons continue to match the position and orientation of the input. See appendix A for the parameter values used in this and later simulations. Animated demos of these figures can be seen at `http://www.cs.utexas.edu/users/nn/pages/research/visualcortex.html`.

a very short time scale through lateral interaction. At each settling time step, the neuron combines the above afferent activation $\sum \xi\mu$ with lateral excitation and inhibition:

$$\eta_{ij}(t) = \sigma\left(\sum \xi\mu + \gamma_{\mathrm{E}} \sum_{k,l} E_{ij,kl}\eta_{kl}(t-1) - \gamma_{\mathrm{I}} \sum_{k,l} I_{ij,kl}\eta_{kl}(t-1)\right), \quad (2)$$

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 $\eta_{kl}(t-1)$ is the activity of neuron (k,l) during the previous time step. The scaling factors γ_{E} and γ_{I} determine the relative strengths of excitatory and inhibitory lateral interactions.

While the cortical response is settling, the retinal activity remains constant. The cortical activity pattern starts out diffuse and spread over a substantial part of the map (as in figure 2*c*), but within a few iterations of equation 2, converges into a small number of stable focused patches of activity, or activity bubbles (figure 2*d*). After an input is presented, and the activity has settled, the connection weights of each 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:

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

where η_{ij} stands for the activity of neuron (i,j) in the final activity bubble, $w_{ij,mn}$ is the afferent or lateral connection weight (μ, E or I), 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 for afferent, η for lateral). At long distances, very few neurons have correlated activity and therefore most long-range connections eventually become weak. The weak connections are eliminated periodically, resulting in patchy lateral connectivity similar to that observed in the visual cortex.

3 Scaling RF-LISSOM simulations

The RF-LISSOM algorithm is computationally intensive and requires a large amount of memory to store the specific lateral connections, which makes large simulations impractical. One straightforward way to model larger maps is to develop the initial model using a small network, and then scale up to more realistic cortical sizes once the behaviour is well-understood and the influence of the various parameters is clear. This way only a few runs will be needed at the larger size; if necessary the larger simulation can be performed on a remote supercomputer. In the following subsections we describe the general approach to systematic scaling, show how a self-organizing model can achieve similar results across simulations with different initial conditions, and derive scaling equations that make use of this property to develop similar maps with networks of different sizes.

3.1 The scaling approach

There are two general types of scaling transformations: a change in the total cortical and retinal area simulated, and a change in neuron or ganglion density for a fixed area. A change in the area corresponds to modeling a larger portion of the visual space, e.g. a larger part of V1 and of the eye. A change in density corresponds to modeling a given area at a

finer resolution (of either cortical neurons or retinal ganglia), as well as modeling a species, individual, or brain area that devotes more neurons or ganglia to the representation of a fixed amount of visual space.

Varying the density or area over a wide range is difficult in a complex dynamical system like RF-LISSOM. Parameter settings that work for one size will need to be very different to work properly with other sizes. To make such scaling feasible, we derived a set of size scaling equations that apply to RF-LISSOM networks and can be adapted for most other models with similar connectivity. The equations were derived by treating a cortical network as a finite approximation to a continuous map composed of an infinite number of units (as in Amari 1980 and Roque Da Silva Filho 1992). Under such an assumption, networks of different sizes represent coarser or denser approximations to the continuous map, and any given approximation can be transformed into another by (conceptually) reconstructing the continuous map and then resampling it.

Given an existing retina and cortex of some fixed size, the scaling equations provide the parameter values needed for a smaller or larger retina and cortex to form a functionally-equivalent map. The equations are linear, but despite the non-linearities in the model they have been been experimentally verified to result in maps with similar properties and map organization, for a wide range of network sizes. All of the equations can be applied before the simulation begins, and as described in section 4, the GLISSOM method allows the density to be changed dynamically during the self-organizing process.

3.2 Prerequisite: Insensitivity to initial conditions

One desirable property of a scaling algorithm is that it results in identical final maps even when the scaling steps are different. Therefore, it is crucial that the map organization not depend on the random initial weights, since those will vary between networks of different sizes. In this section we will show that the RF-LISSOM algorithm has exactly this property. Figure 3 shows that the map pattern depends primarily on the stream of inputs seen during self-organization, and not on the initial weights, as long as the initial weights are drawn from the same random distribution.

The insensitivity to initial weights follows from three features of the RF-LISSOM model: (1) the scalar product input response function, (2) lateral excitation between neurons, and (3) the initial period with high learning rates. First, a neuron's initial response to an input pattern is determined by the sum of the product of each retinal receptor with its corresponding weight value (equation 1). For smoothly-varying input patterns and large enough r_{A}, this sum will have a very similar value regardless of the specific values of the weights to each receptor. Thus, until the weights have self-organized into a smooth, spatially non-uniform distribution, the input response of each neuron will be largely insensitive to the weight values. Second, settling due to lateral excitation (equation 2) causes nearby neurons to have similar final activity levels, which further reduces the contribution of each random afferent weight value. Third, Hebbian learning depends on the final settled activity levels resulting from an input (equation 3), and with a high enough learning rate, the initial weight values are soon overwritten by the responses to the input patterns. Figure 3*d*,*e* shows that the large-scale map features develop similarly even before the initial weight values have been overcome, and thus that the Hebbian process of self-organization is driven by the input patterns rather than the initial weights.

The net result is that as long as the initial weights are generated from the same distribution, their precise values do not significantly affect map organization. Similar invariance to the initial weights should be found in other Hebbian models that compute the scalar product of the input and a weight vector, particularly if they include lateral excitation and use a high learning rate in the beginning of self-organization. If a model does not have such invariance, scaling equations can still generate functionally-equivalent maps, but they will not be visually identical like the density-scaled maps presented in later sections.

3.3 Scaling the area

The following subsections will present the scaling equations for area, retinal ganglion density, and cortical density, along with results from simulations of each type. Since the equations are linear they can also be applied together to change both area and density simultaneously. First, we consider the simplest case, changing the area of the visual space simulated. To change the area, both the cortex width N and the retina width R must be scaled by the same proportion k relative to their initial values N_o and R_o. In addition, to ensure that the resulting network has the same amount of learning per neuron per iteration, the average activity per input receptor needs to remain constant. Thus when using discrete input patterns (such as the oriented Gaussians used in these simulations), the average number of patterns $\bar{\imath}$ per iteration must be scaled with the retinal area. Consequently, the equations for scaling the area by a factor k are:

$$N = kN_o, \quad R = kR_o, \quad \bar{\imath} = k^2\bar{\imath}_o \tag{4}$$

See figure 4 for an example of scaling the area.

3.4 Scaling retinal ganglion density

The parameter changes required when changing retinal receptor or cortical neuron density are less obvious than those for area. To scale the retinal receptor density without disrupting the functional behavior of the map, the ratio between the afferent connection radius and R must be constant, i.e. r_{A} must scale with R. (Otherwise, the visual field area processed by each neuron would change when the retinal density changes.)

When the connection radius increases, the total number of afferent connections per neuron also increases. Because the

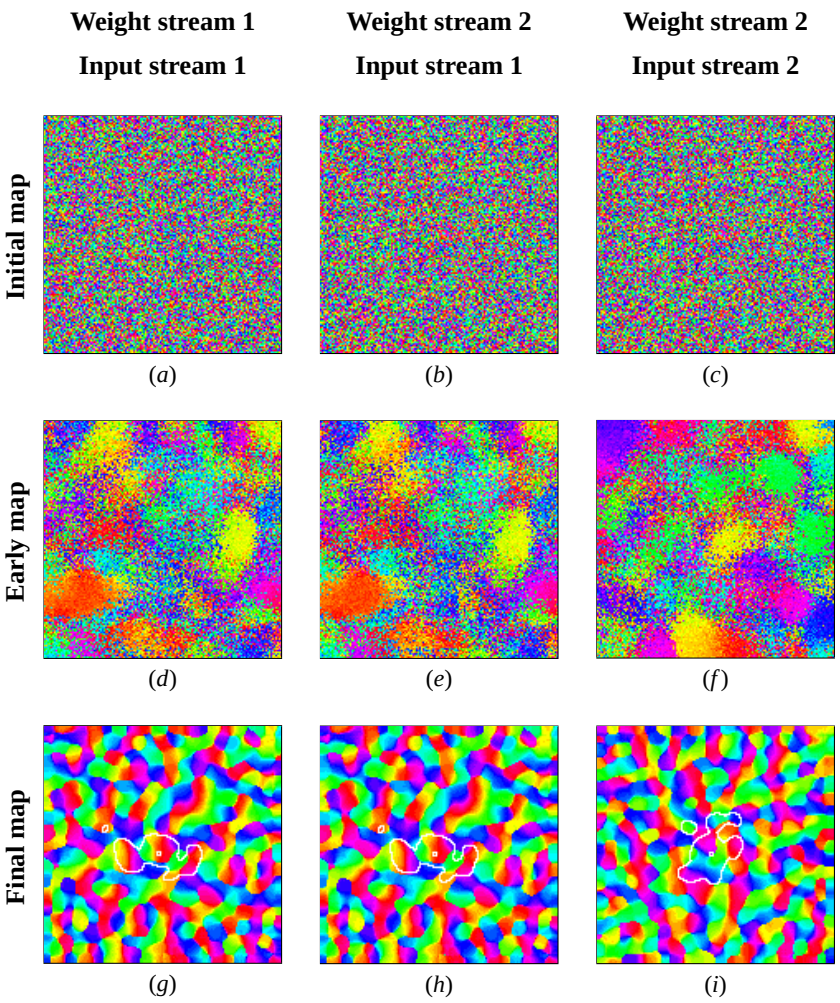


Figure 3: **Input stream determines map pattern in RF-LISSOM (*color figure*).** There are two sources of random fluctuations that can affect the final map pattern in a self-organizing system: the random initial values of the weights, and the random stream of inputs presented. Each of these sources is controlled by an independent pseudo-random number generator that can be given different seeds to produce different streams of values. Using a different seed for the weight generator results in different initial orientation maps (*a* and *b*), but has almost no effect on the final self-organized maps (compare *g* to *h*). (These plots and those in subsequent figures use the color key from figure 2.) This invariance occurs because the development of large-scale map features is insensitive to the initial weight values. (Compare maps *d* and *e* measured at iteration 100; the same large-scale features are emerging in both maps despite different patterns of local noise caused by the different initial weights.) In contrast, changing the input stream produces very different early and final map patterns (compare *e* to *f* and *h* to *i*), even when the initial weight patterns (and therefore the initial orientation maps) are identical (*b* and *c*). Note that regardless of the seed values, the overall properties of the final maps are similar (e.g. hypercolumn spacing, number of pinwheel centers), because the statistical properties of the random input distributions are the same. In animals, different input streams correspond to differing early visual experiences or different patterns of spontaneous activity. Thus the RF-LISSOM model predicts that the precise map pattern in an adult animal depends primarily on the order and type of activity patterns seen by the cortex in early development, and not on the details of the initial connectivity.

learning rate α_{A} specifies the amount of change per connection and not per neuron (equation 3), it must be scaled down so that the average total weight change per neuron per iteration remains constant. (Otherwise, a given input pattern would cause more total change in the weight vectors of each neuron in the scaled network than in the original.) So the afferent learning rate α_{A} must scale inversely with the number of afferent connections to each neuron, which in the continuous plane corresponds to the area enclosed by the afferent radius. Thus, α_{A} scales by the ratio $\frac{{r_{\mathrm{A}_o}}^2}{{r_{\mathrm{A}}}^2}$.

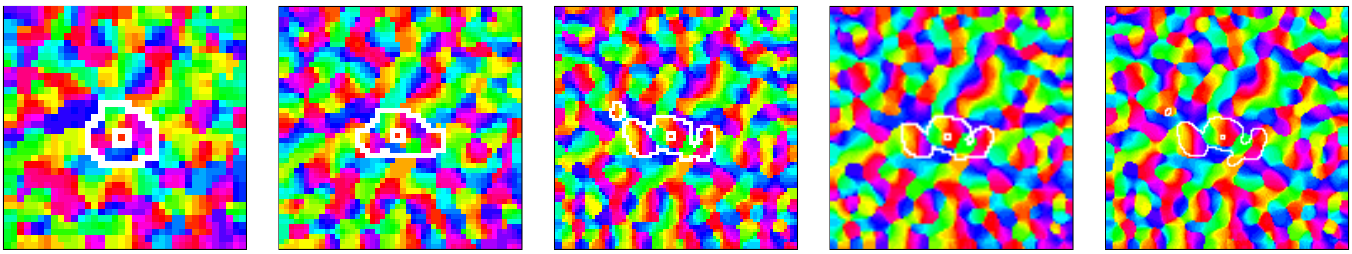
To keep the average activity per iteration constant, the size of the input features must also scale with R, keeping the ratio between the feature width and R constant. For Gaussian inputs this means keeping the ratio between the Gaussian σ and R constant. Thus the retinal density scaling equations are:

$$r_{\mathrm{A}} = \frac{R}{R_o} r_{\mathrm{A}_o} \quad \alpha_{\mathrm{A}} = \frac{{r_{\mathrm{A}_o}}^2}{{r_{\mathrm{A}}}^2} \alpha_{\mathrm{A}_o}, \quad \sigma_x = \frac{R}{R_o} \sigma_{x_o}, \quad \sigma_y = \frac{R}{R_o} \sigma_{y_o} \tag{5}$$

Figure 5 shows how this set of equations can be used to generate functionally-equivalent orientation maps using different retinal receptor densities.

3.5 Scaling cortical neuron density

Changing the cortical density is analogous to changing retinal receptor density, but the intracortical connection sizes and associated learning rates also need to be scaled. The lateral connection radii r_{E} and r_{I} should be scaled with N so that the ratio between each radius and N remains constant. For simulations that shrink the lateral excitatory radius, the final radius r_{E_f} must also be scaled. Like α_{A} in retinal density scaling, α_{E} and α_{I} must be scaled so that the average total weight change per neuron per iteration remains constant despite changes in the number of connections. Finally, the absolute weight level D_{I} below which lateral inhibitory connections are killed at a given iteration must be scaled when the total number of lateral connections changes. (The weight normalization in equation 3 ensures that when there are fewer weights each one is stronger, and thus the value of each weight is proportional to the number of connections.) Thus,



(*a*) 36×36
0.17 hours, 2.0MB

(*b*) 48×48
0.32 hours, 5.2MB

(*c*) 72×72
0.77 hours, 22MB

(*d*) 96×96
1.73 hours, 65MB

(*e*) 144×144
5.13 hours, 317MB

Figure 6: **Scaling the cortical density in RF-LISSOM (*color figure*).** Five RF-LISSOM orientation maps from networks of different sizes are shown; the parameters for each network were calculated using equations (6), and then each network was trained independently on the same pseudorandom stream of input patterns. The size of the network in the number of connections ranged from 2×10^6 to 3×10^8 (2 megabytes to 317 megabytes of memory), and the simulation time ranged from ten minutes to five hours on the same machine, a single-processor 600MHz Pentium III with 1024 megabytes of RAM. Much larger simulations on the massively-parallel Cray T3E supercomputer perform similarly. Despite this wide range of simulation scales, the final organized maps are both qualitatively and quantitatively similar, as long as their size is above a certain minimum (here about 64×64). Larger networks take significantly more memory and simulation time; this difficulty is addressed by GLISSOM (section 4).

the cortical density scaling equations are:

$$\begin{array}{lll} r_{\mathrm{E}} = \frac{N}{N_o} r_{\mathrm{E}_o}, & \alpha_{\mathrm{E}} = \frac{{r_{\mathrm{E}_o}}^2}{{r_{\mathrm{E}}}^2} \alpha_{\mathrm{E}_o}, & D_{\mathrm{I}} = D_{\mathrm{I}_o} \frac{{r_{\mathrm{I}_o}}^2}{{r_{\mathrm{I}}}^2} \\ r_{\mathrm{I}} = \frac{N}{N_o} r_{\mathrm{I}_o}, & \alpha_{\mathrm{I}} = \frac{r_{\mathrm{I}_o}}{r_{\mathrm{I}}^2} \alpha_{\mathrm{I}_o}, & \end{array} \tag{6}$$

Figure 6 shows examples of using the cortical density scaling equations to generate RF-LISSOM orientation maps of different sizes. For band-limited input patterns like those we use here for clarity, results at different sizes match closely, since even the larger networks are representing the same quantities represented by the smaller ones. If the input were much more complicated, as are natural images, then different behavior would be seen for each larger network, as it developed smaller-scale structure to represent the less-prominent features of the input. However, the large-scale features should still remain constant over scaling.

3.6 Limitations on scaling

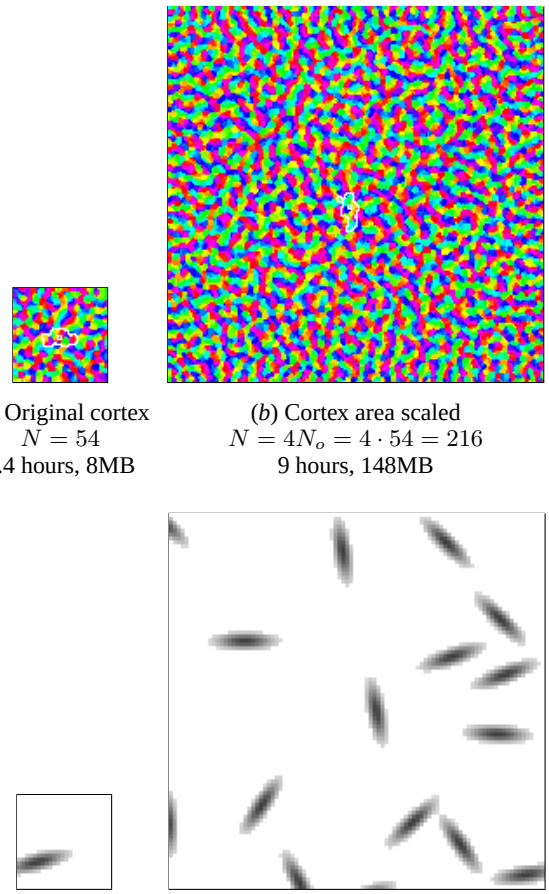
The scaling equations essentially allow any desired cortex and retina size to be simulated without a search for the appropriate parameters. They make it simple to trade off density for area to study specific phenomena given fixed computational resources, and to scale up to larger maps when required. Matching results will be achieved when the size is increased while keeping the complexity of the input patterns constant; in this case, qualitatively and quantitatively similar maps develop regardless of size. The equations can also be used to construct a smaller network, but at some lower bound the network will no longer have enough units to represent a similar map, and thus e.g. only position will be represented and not orientation. Similarly, if the retinal resolution is reduced so much that the retinal images become unrecognizable, the network will no longer have similar functional properties.

For a given simulation, the lower bound on cortex density is generally determined by the minimum excitatory radius (r_{E_f}); with r_{E_f} close to 1.0, the simulation will not give results similar to larger ones. Units are laid out on a rectangular grid, and thus the smallest neuron-centered radius that includes at least one other neuron center is 1.0. If the excitatory radius is allowed to drop below 1.0, the map will no longer have any local topographic ordering, since there will be no local excitation between neurons. Conversely, if the final radius is held at a minimum of 1.0 while the map size continues to be reduced, the lateral spread of excitation will take over a larger and larger portion of the map, causing the hypercolumn width of the resulting map to increase. To avoid such complications, in the rest of this paper we assume that the starting size is above this lower bound on the r_{E_f} and that sizes are being increased.[2]

4 Growing LISSOM (GLISSOM)

The scaling equations make it practical to do many simulations using widely-available hardware, but the scaled up simulations still require so much simulation time and memory that supercomputers are often necessary. Supercomputers generally have restricted access, are located at inconvenient remote sites, have limited auxiliary tool support, and in general are less practical to use than workstations. To address this problem we developed GLISSOM, an extension of the RF-LISSOM model that allows much larger networks to be simulated on the same hardware in the same amount of computation time. GLISSOM consists of successively scaling up an RF-LISSOM network during self-organization using the cortical density equations (6), while smoothly interpolating

[2]Radius values smaller than 1.0 could be approximated using a technique similar to antialiasing in computer graphics. Before a weight value is used in equation 2 at each iteration, it would be scaled by the proportion of its corresponding pixel's area that is included in the radius. This technique should permit smaller networks to be simulated faithfully even with a discrete grid. However, since our focus is on methods for simulating larger networks, we have not investigated such an extension.



(*a*) Original cortex
$N = 54$
0.4 hours, 8MB

(*b*) Cortex area scaled
$N = 4N_o = 4 \cdot 54 = 216$
9 hours, 148MB

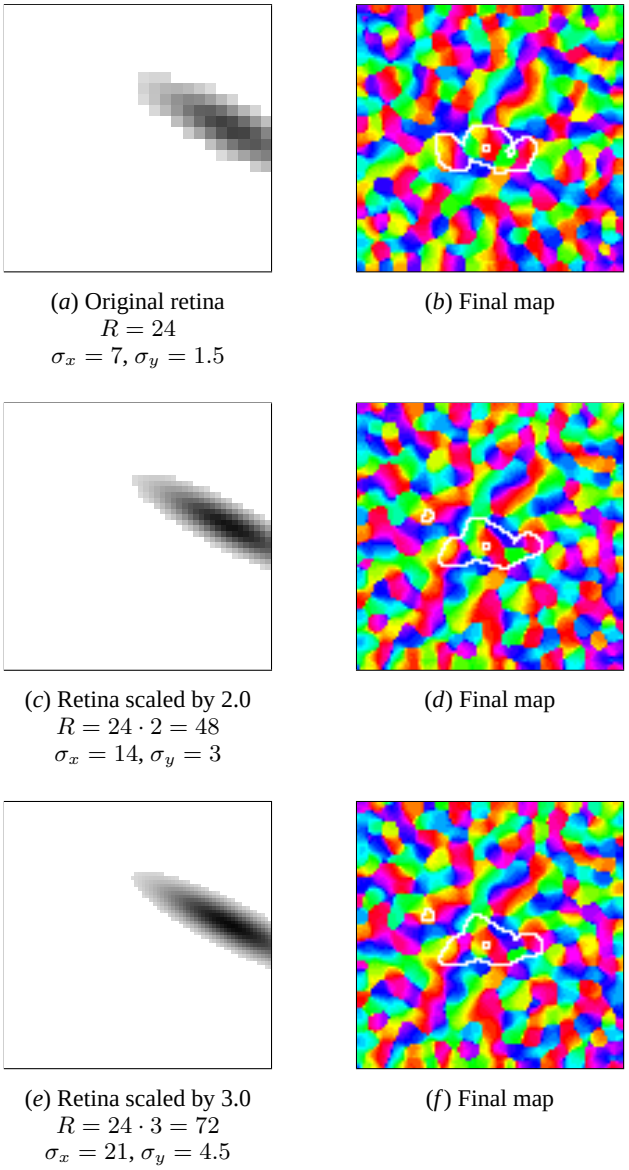
(*c*) Original retina
$R = 24$

(*d*) Retina area scaled
$R = 4R_o = 4 \cdot 24 = 96$

Figure 4: **Scaling the total area in RF-LISSOM (*color figure*).** Equations (4) can be used to scale the total area simulated, in order to simulate a larger portion of the cortex, retina, and visual field. For large N the number of connections and the simulation time scale approximately linearly with the area, and thus a network four times as wide and four times as tall (above) takes about sixteen times the memory and sixteen times as long to simulate. For discrete input patterns like these oriented Gaussians, larger areas require more input patterns to keep the total learning per neuron and per iteration constant. Because the inputs are generated randomly across the active surface of the retina, each map sees an entirely different stream of inputs, and so the final map patterns always differ when the area differs. The area scaling equations are most useful for testing a model with a small area and then scaling up to eliminate border effects and to simulate the full area of a corresponding biological preparation.

the existing afferent and lateral weights to create the new neurons and new connections for the larger network. This scaling allows neuron density to be increased while keeping the large-scale structural and functional properties constant, such as the organization of the orientation preference map. In essence, the large network is grown in place, thereby minimizing the computational resources required for simulation.

Figure 7demonstrates why such an approach is effective. Both the memory requirements and the computation time of



(*a*) Original retina
$R = 24$
$\sigma_x = 7$, $\sigma_y = 1.5$

(*b*) Final map

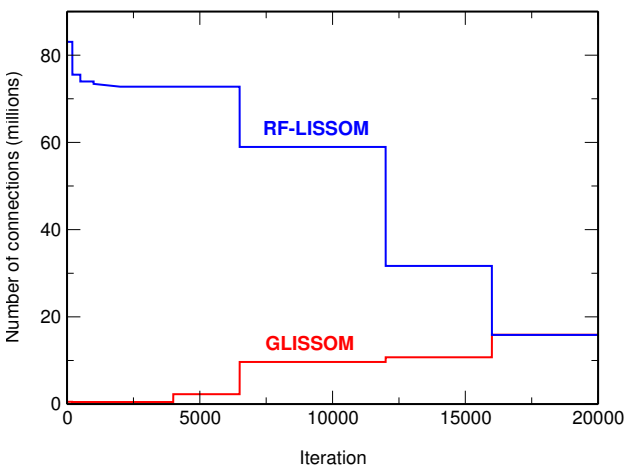
(*c*) Retina scaled by 2.0
$R = 24 \cdot 2 = 48$
$\sigma_x = 14$, $\sigma_y = 3$

(*d*) Final map

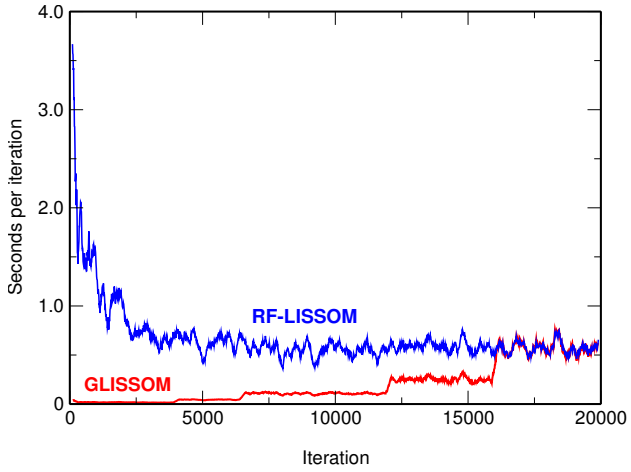
(*e*) Retina scaled by 3.0
$R = 24 \cdot 3 = 72$
$\sigma_x = 21$, $\sigma_y = 4.5$

(*f*) Final map

Figure 5: **Scaling the retinal density in RF-LISSOM (*color figure*).** Equations (5) can be used to scale the density of retinal ganglion cells simulated per unit of visual area, while keeping the area and the cortex density constant. Here we show RF-LISSOM orientation maps from three separate 96×96 networks that have retinas of different densities. The parameters for each network were calculated using equations (5), and then each network was trained independently on the same pseudorandom stream of input patterns. The size of the input pattern in retinal units increases as the retinal density is increased, but its size as a proportion of the retina remains constant. For spatial-frequency band-limited inputs like the ellipsoidal Gaussians shown above and used to train these orientation maps, the final map changes very little above a certain retinal density (here about $R = 48$). With natural images and other stimuli that have high spatial-frequency components, scaling the retinal density will allow the map to represent this higher-frequency data while preserving the large-scale map organization.



(*a*) Memory requirements at each iteration



(*b*) Training time per iteration

Figure 7: **Training time and memory requirements at each iteration.** Data is shown for one run of each algorithm with a final network size of 144×144; for GLISSOM the starting size was 36×36. (*a*) Each line shows the number of connections alive at a given iteration. Memory requirements of RF-LISSOM peak at early iterations, decreasing at first in a series of small drops as the lateral excitatory radius shrinks, and then later in a few large drops as long-range inhibitory weights are pruned at iterations 6500, 12,000, and 16,000. GLISSOM does similar radius shrinking and pruning, while also scaling up the network size at iterations 4000, 6500, 12,000, and 16,000. Since the GLISSOM map starts out small, memory requirements peak much later, and remain bounded because connections are pruned as the network is grown. By this means GLISSOM can keep its peak number of connections, which determines the simulation memory requirements, as low as the smallest number of connections occurring in RF-LISSOM. (*b*) Each line shows a 20-point running average of the time spent in training for one iteration, with a data point measured every 10 iterations. Only training time is shown; times for initialization, plotting images, pruning, and scaling networks are not included. Computational requirements of RF-LISSOM peak at early iterations, falling as the excitatory radius (and thus the number of neurons activated by a given pattern) shrinks and as neurons become more selective. In contrast, GLISSOM requires little computation time until the final iterations. Since the total training time is determined by the area under each curve, GLISSOM spends much less time in training overall.

RF-LISSOM (and other pruning-based models) peak at the start of self-organization, when all connections are active, none of the neurons are selective, and activity is spread over a wide area. As the neurons become selective and smaller regions of the cortex are activated by a given input, simulation time dramatically decreases, since only the active neurons need to be simulated in a given iteration. GLISSOM takes advantage of this process by approximating the map with a very small network early in training, then gradually growing the map as selectivity and specific connectivity are established.

4.1 GLISSOM methods

The GLISSOM weight scaling interpolates between existing values by treating the original weight matrices as a discrete set of samples of a smooth, continuous function. Under such an assumption, the underlying smooth function can be resampled at the higher density. Such resampling is equivalent to the smooth bitmap scaling done by computer graphics programs, as can be seen in the plots shown later in figure 9. This type of scaling always adds at least one whole row or column of neurons at once, as is done in more abstract growing self-organizing maps like the Growing Grid model (Fritzke 1995; Rodriques and Almeida 1990).

We first consider the interpolation procedure for afferent connections. The afferent connection strength to a neuron X in a scaled map from a retinal ganglion G in its receptive field is calculated from the strengths of the connections from G to the *ancestors* of X. The ancestors of neuron X consist of up to four neurons X_1, X_2, X_3, and X_4 surrounding the *image* of X, i.e. the real-valued location in the original network to which the neuron would be mapped by linear scaling of its position (figure 8). In the middle, each neuron has four ancestors; at the corners, each has only one, and along the edges, each has two. Each ancestor X_i of X has an *influence* $\Im$ ranging from 0 to 1.0 on the computed weights of X, determined by its proximity to X:

$$\Im_{X_i} = 1.0 - \frac{d(X, X_i)}{d_m}, \tag{7}$$

where $d(X, X_i)$ represents the Euclidean distance between the image of X and its ancestor X_i in the original network, and d_m is the maximum possible distance between the image of a scaled neuron and any of its ancestors, i.e. the diagonal spacing between ancestors in the original network. The afferent connection strength w_{XG} is then a normalized proximity-weighted linear combination of the weights from G to the ancestors of X:

$$w_{XG} = \frac{\sum_i w_{X_i G} \Im_{X_i}}{\sum_i \Im_{X_i}}, \tag{8}$$

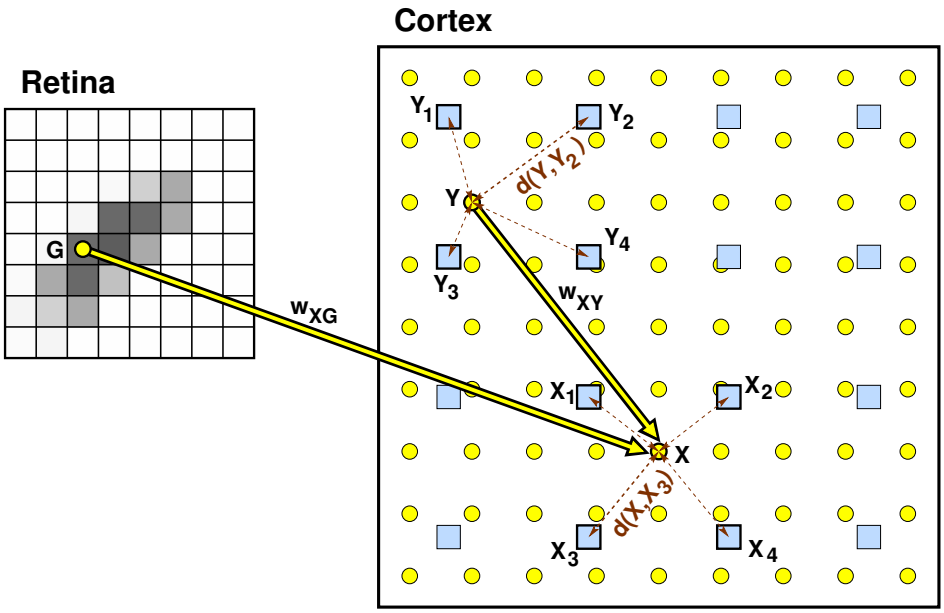


Figure 8: **GLISSOM connection strength scaling.** This example shows a cortex of size 4×4 being scaled to 9×9 during self-organization, with the retina size fixed at 8×8. Both cortical networks are plotted in the continuous plane representing a fixed area of cortex. The squares on the cortex represent neurons in the original network (i.e. before scaling) and circles represent neurons in the new network. G is a retinal ganglion cell and X and Y are neurons in the new network. Afferent connection strengths to neuron X in the new network are calculated from the connection strengths of the *ancestors* of X. The ancestors of X are X_1, X_2, X_3, and X_4, i.e. the neurons in the original network that surround the position of X. The new afferent connection strength w_{XG} from ganglion G to X is a normalized combination of the connection strengths w_{X_iG} from G to each ancestor X_i of X, weighted by the distance $d(X, X_i)$ between X_i and X. Lateral connection strengths from Y to X are calculated similarly, as a proximity-weighted combination of the connection strengths between the ancestors of those neurons. First the *contribution* of each X_i is calculated as a normalized combination of the connection strengths $w_{X_iY_j}$ from each Y_j to X_i, weighted inversely by the distance $d(Y, Y_j)$ between Y_j and Y. The connection strength w_{XY} is then the sum of the contributions of each X_i, weighted inversely by the distance $d(X, X_i)$. Thus the connection strengths in the scaled network consist of proximity-weighted combinations of the connection strengths in the original network.

where w_{X_iG} is the afferent connection weight from ganglion G to the ith ancestor of X. Because receptive fields are limited in size, not all ancestors will necessarily have connections to that ganglion, and only those ancestors with connections to G contribute to the sum.

Lateral connection strengths from neuron Y to neuron X in the scaled map are computed similarly, as a proximity-weighted linear combination of the connection strengths between the ancestors of Y and the ancestors of X. First, the *contribution* from the ancestors of Y to each X_i is calculated:

$$C_{X_iY} = \frac{\sum_j w_{X_iY_j} \Im_{Y_j}}{\sum_j \Im_{Y_j}}, \tag{9}$$

where $w_{X_iY_j}$ is the connection weight from the jth ancestor of Y to the ith ancestor of X, where defined. The new lateral connection strength w_{XY} is then the proximity-weighted sum of the contributions from all ancestors of X:

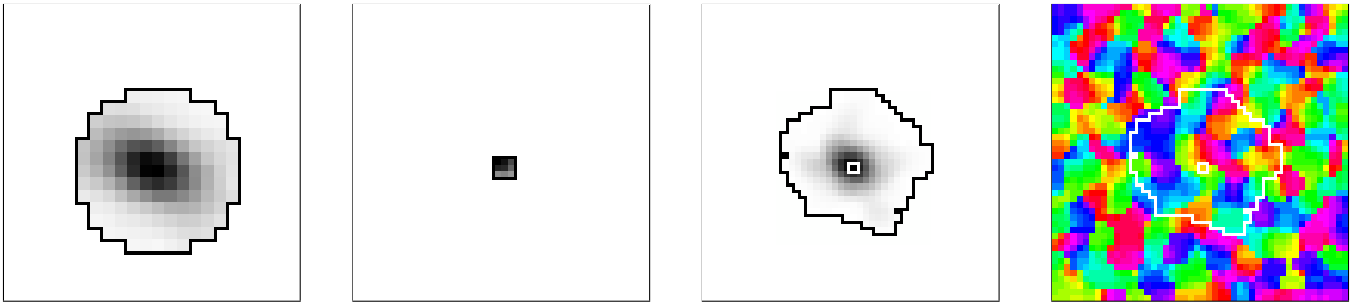
$$w_{XY} = \frac{\sum_i C_{X_iY} \Im_{X_i}}{\sum_i \Im_{X_i}}, \tag{10}$$

Figure 9 shows a scaling example for a partially-organized orientation map and the weights of one neuron in it.

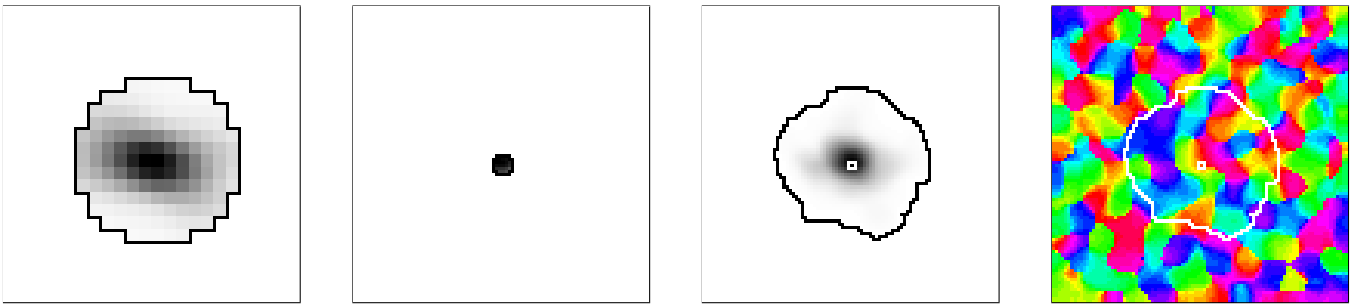
4.2 GLISSOM experiments

A GLISSOM simulation starts with a cortex that has a low cortical density. The scaling procedure is then used to gradually increase the density as the network self-organizes. At the same time, the scaling equations from section 3.5 are used to keep the other parameters functionally equivalent. Similar scaling could be used to increase the retinal density during self-organization, but since the retinal processing has less of an influence on the computation and memory requirements, retinal density scaling was not included in the simulations reported below.

The precise scaling schedule is not crucial as long as the starting point is large enough to represent the largest-scale features in the final map, without being so large that the speed benefits are minimal. A linear size increase from the starting size N_o to the final size N_f shortly before the end of training usually works well; scaling up more quickly increases simulation accuracy, while delaying scaling reduces training time at the expense of accuracy. Making a few large scaling steps gives similar results in less computation time than making many smaller steps, so usually only a few steps are used. The intermediate scaling steps in the following sections were



(*a*) Afferent weights Before scaling | (*b*) Excitatory weights Before scaling | (*c*) Inhibitory weights Before scaling | (*d*) 48×48 cortex Before scaling



(*e*) Afferent weights After scaling | (*f*) Excitatory weights After scaling | (*g*) Inhibitory weights After scaling | (*h*) 96×96 cortex After scaling

Figure 9: **GLISSOM: Scaling cortex density in place (*color figure*).** These figures represent a single large GLISSOM cortical density scaling operation, going from a 48×48 cortex to a 96×96 cortex in one step at iteration 10,000 out of a total of 20,000. (Usually much smaller steps are used, but the changes are exaggerated here to make them more obvious.) A set of weights for one neuron from the original 48×48 network is shown in (*a-c*), and the orientation map measured for the network in (*d*). The active inhibitory weights of that neuron are outlined in white. Similarly, a set of weights for one neuron from the 96×96 scaled network is shown in (*e-g*). The orientation map (*h*) measured from the scaled map is identical to (*d*) except that the resolution has been doubled smoothly; this network can then self-organize at the new density to represent finer details. Since each scaled neuron receives four times as many lateral weights as one from the original map, pruning is usually done during the scaling step to keep the total size of the network bounded; pruning was skipped here for simplicity.

computed using the following formula:

$$N = N_o + \kappa(N_o - N_f), \tag{11}$$

where κ was a constant whose values increased approximately linearly over training. Unless stated otherwise the simulations below use four scaling steps, κ =0.20 at iteration 4000, 0.47 at 6500, 0.67 at 12,000, and 1.0 at 16,000. Figure 10 shows examplesof the scaled maps at each iteration for one GLISSOM simulation, compared to the full-size RF-LISSOM simulation. The GLISSOM network passes through similar stages of self-organization, with the map size gradually approaching that of the RF-LISSOM map.

4.3 GLISSOM results

The first requirement for GLISSOM is that the maps and connection weights it produces should be equivalent to those produced by RF-LISSOM. As shown in figures 11 and 12, for a sufficiently large starting map size, GLISSOM produces an orientation preference map and weight patterns that are qualitatively and quantitatively equivalent to those of RF-LISSOM.

The second requirement is that GLISSOM significantly reduce the overall computation time and memory requirements. This result can clearly be seen in figure 13. For example, for a final $N = 144$ on a 600MHz Pentium III workstation, RF-LISSOM takes 5.1 hours for 20,000 training iterations, while GLISSOM finishes in 1.6 hours (a speedup ratio of 3.1). For that simulation, RF-LISSOM requires 317MB of memory to store its connections, while GLISSOM requires only 60MB (a memory savings ratio of 5.2). Importantly, the speedup and memory savings increase as the network size increases (figures 13*c* and 13*d*), which means that GLISSOM can make it practical to simulate much larger networks.

These results validate our conjecture that a coarse approximation suffices for the early iterations in RF-LISSOM. Early in training, only the large-scale organization of the map is important; using a smaller map for this stage does not significantly affect the final results. Once the large-scale structure settles, individual neurons start to become more selective and to differentiate from their local neighbors; a denser map is then required so that this detailed structure can develop. Thus GLISSOM uses a map size appropriate for each stage in self-

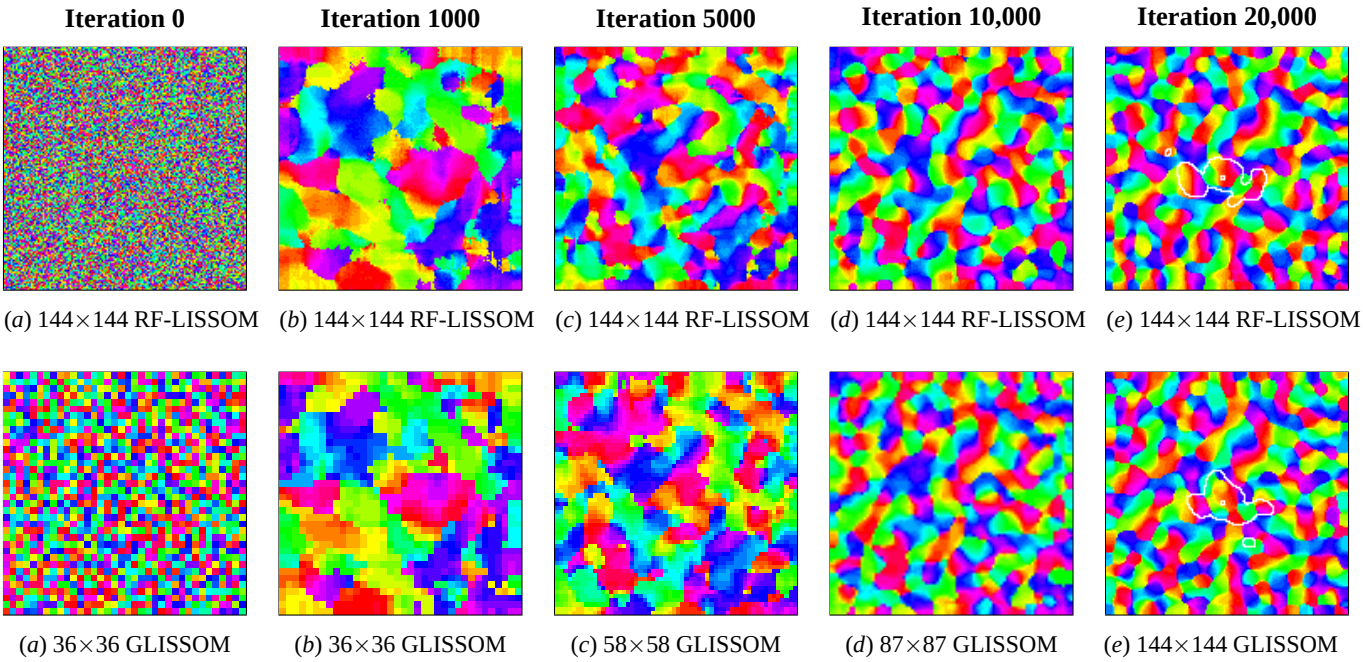


Figure 10: **Self-organization in RF-LISSOM and GLISSOM (*color figure*).** At each iteration, the emerging GLISSOM map features are similar to those of RF-LISSOM except for discretization differences, and the maps are gradually scaled so that by the final iteration the GLISSOM map has the same size as RF-LISSOM. To make the scaling steps more obvious, this example uses the smallest acceptable starting size; figure 12 shows that results match more closely for larger starting sizes.

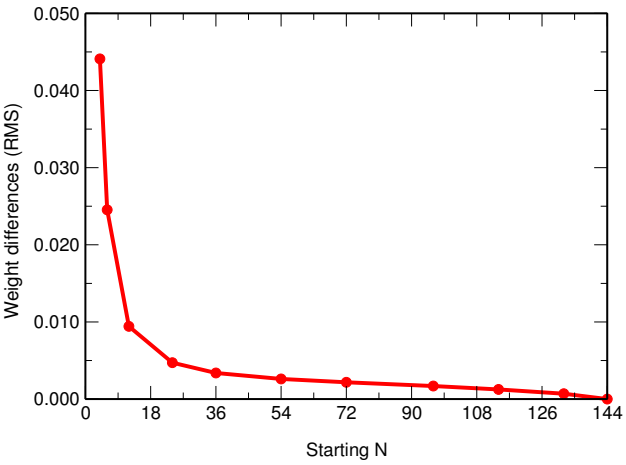
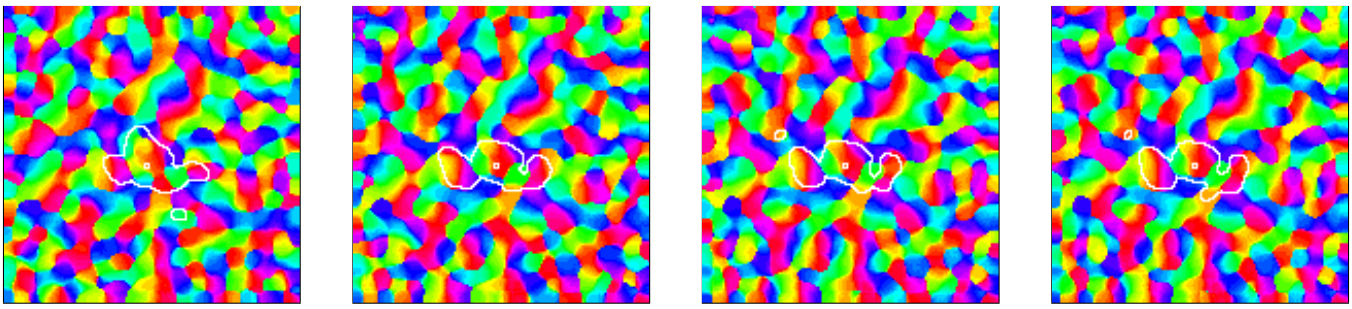


Figure 11: **For sufficiently large starting N, GLISSOM weights are a close match to those of RF-LISSOM.** Each point shows the root mean squared (RMS) difference between the final values of the corresponding weights of each neuron in two networks: a 144×144 RF-LISSOM map, and a GLISSOM map with an initial size shown on the x-axis and a final size of 144×144. Both maps were trained on the same stream of oriented inputs. The GLISSOM maps starting as large as $N = 96$ used four scaling steps, while the three starting points closer to the final size used fewer steps: $N = 114$ had one step at iteration 6500, $N = 132$ had one step at iteration 1000, and there were no scaling steps for $N = 144$. Low values of RMS difference indicate that the corresponding neurons in each map developed very similar weight patterns. The RMS difference drops quickly as the starting size increases, becoming negligible above 36×36. As described in section 3.6, this lower bound is determined by r_{E_f}, the minimum size of the excitatory radius.

organization, in order to model development faithfully while saving simulation time and memory.

5 Scaling to cortical dimensions

The maps in section 4 represent only a small region of cortex and have a limited range of connectivity. Combined with the scaling equations, these results make it possible to obtain a rough estimate of the resource requirements needed to approximate the full density, area, and connectivity of the visual cortex in a particular species. As discussed in the next section, with such a simulation it will be possible to study phenomena that require the entire visual field or the full cortical column density and connectivity. Calculating the full-scale parameter values is also valuable by itself, since it can help tie the parameters of a small model to physical measurements. For instance, once the relevant scaling factors are calculated, the connection lengths, receptive field sizes, retinal area, and cortical area to use in a model can all be derived directly from the corresponding quantities measured in a biological preparation. Conversely, where such measurements



(*a*) GLISSOM
Starting $N = 36$
1.63 hours, 61MB

(*b*) GLISSOM
Starting $N = 54$
1.94 hours, 60 MB

(*c*) GLISSOM
Starting $N = 72$
2.29 hours, 69 MB

(*d*) RF-LISSOM
Fixed $N = 144$
5.13 hours, 317 MB

Figure 12: **Orientation maps match between GLISSOM and RF-LISSOM (*color figure*).** Above the minimum 36×36 starting size, the final GLISSOM maps closely match those of RF-LISSOM, yet take much less time and memory to simulate. Computation time increases smoothly as the starting size is increased, allowing a tradeoff between accuracy and time; accuracy is high even for computation times substantially lower than those of RF-LISSOM. Memory requirements are bounded by the size of the final maps until the starting size approaches the final size, after which the requirements gradually approach those of RF-LISSOM.

are not available, model parameter values which produce realistic behavior constitute predictions for future experiments.

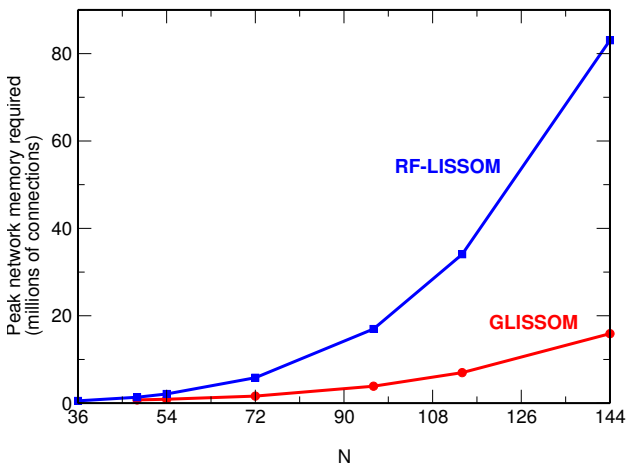
In this section we will compute resource requirements and several key RF-LISSOM parameters needed for a full-scale simulation of human visual cortex. We will first consider a full-density simulation of a small but typical simulation area, and then a full-density simulation of the total area of human V1. The parameters used in most of the simulations in the preceding sections construct a map whose global features match approximately a 5mm × 5mm (25mm^2) area of macaque cortex (compare figure 12 with Blasdel 1992). Many map models use an area of about this size because it corresponds to the experimental data available, and because it shows enough hypercolumns in each direction to study the overall map organization. Estimates for the number of neurons in a 25mm^2 area of human V1 range from 1.6×10^6 (Wandell 1995) to 9×10^6 (Rockel et al. 1980). Each cortical unit in RF-LISSOM represents one vertical column, and the number of neurons per vertical column in primate V1 has been estimated at 259 (Rockel et al. 1980). Thus a full-scale simulation of 25mm^2 would require from $6,000$ to $35,000$ column units total, which corresponds to RF-LISSOM parameter N ranging from 78 to 188.

Simulations with this number of units are practical; some of the workstation simulations presented in the previous sections were as large as $N = 144$, and similar simulations on a supercomputer have been as large as $N = 192$ (Bednar and Miikkulainen 2000). However, it is the number of lateral connections, not the number of columns, that is the crucial determinant of simulation time and memory requirements. Lateral connections in V1 are as long as 8mm (Gilbert et al. 1990), but previous RF-LISSOM simulations have underestimated lateral connection length for computational reasons. For the smallest realistic size $N = 78$, a simulation comparable to those in the previous sections would use a maximum inhibitory radius $r_{\mathrm{I}} = 18$, but simulating the full 8mm would require $r_{\mathrm{I}} = 124$.[3] Because the current RF-LISSOM implementation stores all connections possible within the full extent, it would need to store $78^2 \times 249^2 \approx 4 \times 10^8$ long-range lateral connections for the 25mm^2 area.[4] Thus a minimum of1.4 terabytes of RAM (assuming 4 bytes per connection) would be required to simulate the 25mm^2 area at full density in RF-LISSOM. Supercomputers will probably reach this range within the next few years but are not yet there. Since the total area of human V1 is about 2400mm^2 (Wandell 1995), the minimum realistic RF-LISSOM simulation size for all of V1 would be about 1000 times greater than for the 25mm^2 patch, or 135 terabytes. This figure is out of the range of even the largest supercomputers expected to be available in the near future.

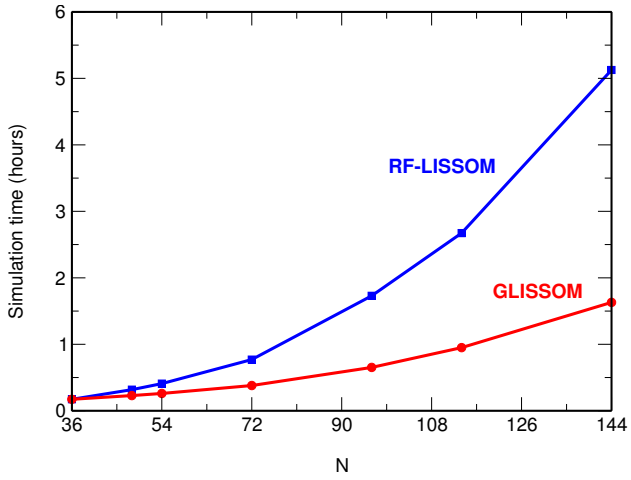
In contrast, since GLISSOM does not need to represent all of the possible final connections at the beginning of the simulation, it can make use of a sparse lateral connection storage format that takes much less memory, and correspondingly less computation time. The memory required depends on the number of connections active after self-organization, which in current GLISSOM simulations is about 15%. As the radius increases this percentage will drop with at least the square of the radius, since long-range connections extend only along the preferred orientation of the neuron and not in all directions (Bosking et al. 1997). Thus, for the full

[3] Since this range would provide greater than full connectivity for the 78×78 area simulated, to make the following calculations we assume that the area under discussion is part of a larger area also simulated. Note that $r_{\mathrm{I}} = 18$ requires only a tiny fraction of the memory needed to simulate $r_{\mathrm{I}} = 124$, since the memory requirements increase with the area enclosed by the radius.

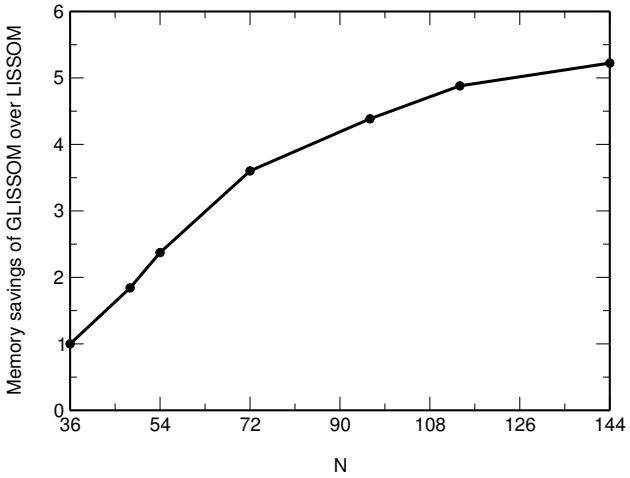
[4] For the large networks discussed here the number of long-range lateral inhibitory connections is far greater than the connections of other types, so the memory requirements will be calculated from the inhibitory connections alone.



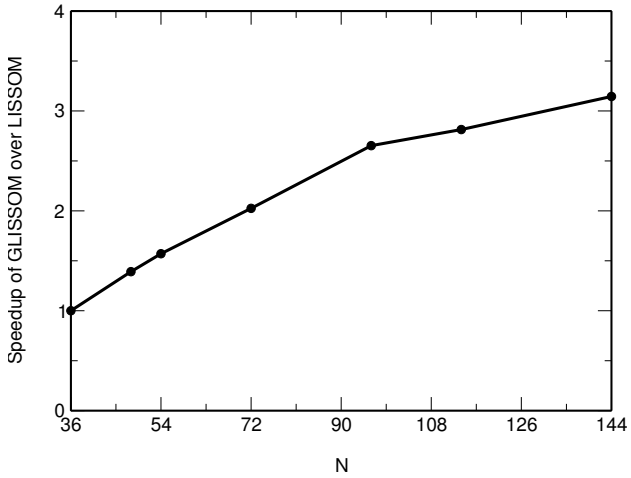
(*a*) Peak memory requirements vs. N



(*b*) Simulation time vs. N



(*c*) Memory savings vs. N



(*d*) Speedup vs. N

Figure 13: **Comparison of simulation time and memory requirements for GLISSOM and RF-LISSOM.** Summary of results from fourteen simulations on the same 600MHz Pentium III workstation, each running either RF-LISSOM (with the fixed N indicated in the x-axis) or GLISSOM (with a starting size of $N = 36$ and the final size N indicated in the x-axis). (*a*) The memory requirements consist of the peak number of network connections required for the simulation; this peak determines the minimum physical memory needed when using an efficient sparse format for storing weights. RF-LISSOM's memory requirements increase very quickly as N is increased, while GLISSOM is able to keep the peak number low so that much larger networks can be simulated on a given machine. For example (reading across a horizontal line through the graph), if a machine can hold 20 million connections, with RF-LISSOM it can only simulate $N = 96$ (9000 units), while with GLISSOM it can simulate $N = 144$ (21,000 units); GLISSOM's advantage increases for larger values of N. (*b*) The total simulation times for RF-LISSOM also increase dramatically for larger networks, because larger networks have many more connections to process. In contrast, since GLISSOM uses fewer connections for most of the self-organization process, its computation time increases only modestly for the same range of N. (Simulation time includes training time plus time for all other computations, including plotting, orientation map measurement, and GLISSOM's scaling steps.) (*c*,*d*) As the network size increases, GLISSOM results in greater memory savings (ratio between memory requirements of RF-LISSOM and GLISSOM) and a greater speed boost, which makes large networks much more practical to simulate. [The variance in simulation time and memory usage between simulation runs is very small (less than 1% even with different input and weight seeds), and thus for $N \geq 72$ the differences between GLISSOM and RF-LISSOM are highly statistically significant (Student's *t*-test, $P < 0.005$) after only two runs of each simulation. Since the results are so consistent, for simplicity all data points plotted here were measured from a single run of the indicated algorithm.]

scale simulation, about $15\% \cdot \frac{18^2}{124^2} = 0.3\%$ of the connections would remain active. Under these assumptions, the memory requirements for a fully-realistic simulation of a 25mm^2 area could be reduced to approximately 5 gigabytes, which is currently within the range of a large workstation. Even for a simulation of the full V1 area, GLISSOM using an efficient sparse connection storage format would require only about 450 gigabytes, which is within the reach of the largest existing supercomputers, e.g. San Diego Supercomputing Center's 584GB IBM RS/6000 SP. Thus a sparse implementation of GLISSOM should be able to simulate all of V1 at the single-column level with realistic lateral connectivity using currently-available computer hardware.

6 Discussion and future work

The RF-LISSOM simulation results in section 3 showed that the scaling equations are valid over a wide range of spatial scales. The GLISSOM results in section 4 showed that the equations can be used to significantly reduce simulation time and memory usage and thereby make the study of large-scale phenomena tractable. Similar equations should apply to most other models with specific intracortical connectivity, and can be adapted to those with more abstract connectivity such as a Mexican hat interaction function. The growth process of GLISSOM should provide similar performance and memory benefits to most other densely-connected models whose peak number of connections occurs early in training. Essentially, the GLISSOM procedure allows a fixed model to be adapted into one that grows in place, by using scaling equations and an interpolation algorithm.

The benefits of a GLISSOM approach will be somewhat lower for models that do not shrink an excitatory radius during self-organization, and therefore do not have a temporary period with widespread activation. For such models, it may be worthwhile to consider a related approach, whereby only the lateral connection density is gradually increased, instead of increasing the total number of neurons in the cortex. Such an approach would still keep the number of connections (and therefore the computational and memory requirements) low, while keeping large-scale map features such as the hypercolumn distance constant over the course of self-organization.

Apart from their application to simulations, the scaling equations give insight into how the corresponding quantities differ between individuals, between species, and during development. In essence, the equations predict how the biophysical correlates of the model parameters will differ between any two cortical regions that differ in size but perform similar computations. The discrepancy between the actual parameter values and those predicted by the scaling equations can give insight into the difference in function and performance of different brain regions, individuals and species.

For instance, equations (6) and the simulation results suggest that learning rates per connection should scale with the total number of connections per neuron. Otherwise neurons from a more densely-connected brain area would have significantly greater total plasticity, which (to our knowledge) has not been demonstrated. Consequently, unless the number of synapses per neuron is constant, the learning rate must be regulated at the whole-neuron level rather than being a property of individual synapses. This principle conflicts with assumptions implicit in most neural-network models, including RF-LISSOM, that specify learning rates for individual connections directly. Future work will be needed to determine whether such whole-neuron regulation of plasticity does occur, and if not, whether more densely-connected regions do have a greater level of overall plasticity.

Similarly, equations (6) suggest that the connection strength pruning threshold D_{I} depends on the total number of connections to the neuron, rather than being an arbitrary fixed value. With divisive weight normalization, increasing the number of connections decreases the strength of each one; this procedure is motivated by *in vitro* findings of whole-cell regulation of excitability (Turrigiano et al. 1998). A consequence is that a fixed D_{I} that prunes e.g. 1% of the connections for a small cortex would prune all of the connections for a larger cortex. This finding provides independent theoretical support for experimental evidence that shows pruning to be a competitive process (Purves 1988).

The scaling equations also provide an effective tool for making cross-species comparisons, particularly between species with different brain sizes. In effect, the equations specify the parameter values that a network *should* use if it is to have similar behavior as a network of a different size. As pointed out by Kaas (2000), different species do *not* usually scale faithfully, probably due to geometrical, metabolic, and other restrictions. As a result, as V1 size increases, the lateral connection radii do not increase as specified in the cortical density scaling equations, and processing becomes more and more topographically local. Kaas (2000) proposes that such limitations on connection length may explain why larger brains such as human and macaque are composed of so many visual areas, instead of just expanding the area of V1 to support greater functionality. The scaling equations combined with LISSOM provide a concrete platform on which to test these ideas in future multi-region simulations.

The most important result of the scaling equations and the GLISSOM approach is that they will make much larger simulations feasible. Future work can use this power to study larger-scale and more complex phenomena that would otherwise be possible only in abstract models. For instance, the visual tilt illusion is thought to occur through orientation-specific lateral interactions between spatially-separated stimuli (Carpenter and Blakemore 1973). Memory constraints limited the lateral connection length in previous RF-LISSOM simulations, and thus such long-range interactions could not be studied. Existing computational models of the tilt illusion have been practical only because they ignore the spatial components of the input pattern and the cortical response (Mundel

et al. 1997; Spivey-Knowlton 1993), which is an unrealistic assumption since the tilt illusion is strongly dependent on spatial position (Gibson and Radner 1937). With GLISSOM it should be possible to integrate both the spatial and orientation aspects of the tilt illusion by self-organizing a network with a large area and the full range of specific lateral connectivity found in the cortex.

Visual contour integration, object binding, and object segmentation are similarly thought to depend on specific long-range lateral interactions (Gilbert et al. 1996). Existing models use fixed connectivity or support only short-range interactions (Choe and Miikkulainen 1998; Li 1998). It has not yet been possible to model the detailed self-organization of a large enough area to represent the multiple detailed objects necessary to study such phenomena realistically. With a GLISSOM approach such simulations should become practical. The results can then be extended to include multiple hierarchically-organized feedback networks so that large-scale object recognition can be studied. The computational requirements will increase when more maps are included, but the GLISSOM approximation techniques should apply to each of the maps.

Current laterally-connected map models also focus on a single stimulus dimension (such as orientation, ocularity, or spatial frequency) so that a low-density approximation to the cortical network suffices. More abstract models have been able to incorporate several such dimensions (typically orientation and ocular dominance) because their computational requirements are lower (Obermayer et al. 1991; Osan and Ermentrout 2002; Swindale 1992). However, such models cannot be used to study phenomena like the interocular transfer of the tilt aftereffect, which depends on specific lateral connections between binocular orientation-selective neurons (Bednar and Miikkulainen 2000; Gibson and Radner 1937). Since GLISSOM makes a higher-density network practical, future simulations can model such phenomena by using natural image training stimuli which vary along all of the stimulus dimensions represented in the cortex.

Many other important phenomena also require large maps, including visual attention, saccades between stimulus features, the interaction between the foveal and peripheral representations of the visual field, and how large-scale patterns of optic flow due to head movement influence self-organization. The scaling equations and the GLISSOM method should help make detailed models of these simulations practical as well. Modeling at this level will allow the components and parameters of the model to be tied directly to neural structures and experimental measurements, which will provide new predictions for biology as well as allow the models to be validated rigorously.

7 Conclusion

The scaling equations and the GLISSOM method should allow detailed laterally-connected cortical models like RF-LISSOM to be applied to much more complex, large-scale phenomena. Using the largest available supercomputers, it should even be possible to model all of V1 at the column level. These methods also provide insight into the cortical mechanisms at work in organisms with brains of widely different sizes. Thus the scaling equations and GLISSOM can help explain brain scaling in nature as well as helping to scale up computational simulations of the brain.

Acknowledgments

This research was supported in part by the National Science Foundation under grants IRI-9309273 and IIS-9811478.

A Parameter values and simulation details

All simulations were run using the same reference set of RF-LISSOM parameters adapted from Bednar and Miikkulainen (2000). The reference simulation used a cortex $N = 192$ and a retina $R = 24$. All parameters listed below are for this reference simulation only; they were scaled using the density and area scaling equations from section 3 to get the specific parameters for each of the other simulations.

The cortex was self-organized for $20,000$ iterations on oriented Gaussian inputs with major and minor axes with $\sigma = 7.5$ and 1.5, respectively. The afferent receptive field radius r_{A} was 6; the initial connections within that circular radius were drawn from a uniform random distribution. Where multiple input patterns were used, they were constrained to have centers at least $2.2r_{\mathrm{A}}$ apart. The initial lateral excitation radius r_{E} was 19 and was gradually decreased to 5.5. The lateral inhibitory radius r_{I} was 48. The lateral inhibitory connections were initialized to a Gaussian profile with $\sigma = 100$, and the lateral excitatory connections to a Gaussian with $\sigma = 15$, with no connections outside the circular radius. The lateral excitation γ_{E} and inhibition strength γ_{I} were both 0.9. The learning rate α_{A} was gradually 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 were increased from 0.1 to 0.24 and from 0.65 to 0.88, respectively. The number of iterations for which the lateral connections were allowed to settle at each training iteration was initially 9, and was increased to 13 over the course of training. Inhibitory connections below 0.0000007 were pruned at iteration 6500, those below 0.000035 at iteration 12,000, and those below 0.0002 at iteration 16,000.

All simulations were run on an unloaded single-processor 600MHz Pentium III Linux machine with 1024 megabytes of RAM. All timing results are user CPU times reported by the GNU `time` command; for these simulations CPU time is essentially the same as the elapsed wallclock time since the CPU utilization was always over 99%.

References

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

Anderson, J. A., and Rosenfeld, E., editors (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

Bednar, J. A., and Miikkulainen, R. (2000). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Computation*, 12(7):1721–1740.

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

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal con-

nections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6):2112–2127.

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

Chapman, B., Stryker, M. P., and Bonhoeffer, T. (1996). Development of orientation preference maps in ferret primary visual cortex. *Journal of Neuroscience*, 16(20):6443–6453.

Choe, Y., and Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, 21:139–157.

Erwin, E., Obermayer, K., and Schulten, K. (1992). Self-organizing maps: Ordering, convergence properties and energy functions. *Biological Cybernetics*, 67:47–55.

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.

Fritzke, B. (1995). Growing grid: A self-organizing network with constant neighborhood range and adaptation strength. *Neural Processing Letters*, 2:9–13.

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., Das, A., Ito, M., Kapadia, M., and Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences, USA*, 93:615–622.

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.

Gould, E., Reeves, A. J., Graziano, M. S. A., and Gross, C. G. (1999). Neurogenesis in the neocortex of adult primates. *Science*, 286:548–552.

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.

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

Kaas, J. H. (2000). Why is brain size so important: Design problems and solutions as neocortex gets bigger or smaller. *Brain and Mind*, 1:7–23.

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

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940.

Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (1997). Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In Goldstone, R. L., Schyns, P. G., and Medin, D. L., editors, *Perceptual Learning*, vol. 36 of *Psychology of Learning and Motivation*, 257–308. San Diego, CA: Academic Press.

Mundel, T., Dimitrov, A., and Cowan, J. D. (1997). Visual cortex circuitry and orientation tuning. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Advances in Neural Information Processing Systems 9*, 887–893. Cambridge, MA: MIT Press.

Obermayer, K., Blasdel, G. G., and Schulten, K. J. (1991). A neural network model for the formation and for the spatial structure of retinotopic maps, orientation- and ocular dominance columns. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Proceedings of the International Conference on Artificial Neural Networks 1991* (Espoo, Finland), 505–511. Amsterdam; New York: North-Holland.

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

Osan, R., and Ermentrout, B. (2002). Development of joint ocular dominance and orientation selectivity maps in a correlation-based neural network model. In Bower, J. M., editor, *Computational Neuroscience: Trends in Research, 2002*. New York: Elsevier. To appear.

Purves, D. (1988). *Body and Brain: A Trophic Theory of Neural Connections*. Cambridge, MA: Harvard University Press.

Rockel, A. J., Hiorns, R. W., and Powell, T. P. S. (1980). The basic uniformity in structure of the neocortex. *Brain*, 103:221–244.

Rodriques, J. S., and Almeida, L. B. (1990). Improving the learning speed in topological maps of patterns. In *Proceedings of the International Neural Networks Conference* (Paris, France), 813–816. Dordrecht; Boston: Kluwer.

Roque Da Silva Filho, A. C. (1992). *Investigation of a Generalized Version of Amari's Continuous Model for Neural Networks*. PhD thesis, University of Sussex at Brighton, Brighton, UK.

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. Technical Report AI95-237.

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

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.

Spivey-Knowlton, M. J. (1993). Simulating tilt illusions with lateral inhibition and a virtual axis. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, 953–958. Hillsdale, NJ: Erlbaum.

Swindale, N. V. (1992). A model for the coordinated development of columnar systems in primate striate cortex. *Biological Cybernetics*, 66:217–230.

Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network – Computation in Neural Systems*, 7:161–247.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:845–846.

von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, 15:85–100. Reprinted in Anderson and Rosenfeld 1988.

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

Weliky, M., Kandler, K., Fitzpatrick, D., and Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron*, 15:541–552.