A Neural Network Model of Topographic Reorganization Following Cortical Lesions

Joseph Sirosh and Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin, Austin, TX–78712.
email: sirosh,risto@cs.utexas.edu

Abstract

A neural network model for the simultaneous self-organization of topographic receptive fields and lateral interactions in cortical maps is presented. The afferent and lateral connection weights in the network are initially random, but self-organize based on external input to form topographic maps. The maps are in dynamic equilibrium with the input, and can reorganize in response to lesions in the network. During reorganization, the area of functional loss resulting from the lesion first increases as lateral connections adapt, and then decreases as afferent connections reorganize to compensate. The reorganizing behavior closely matches experimental observations on cortical lesions and stroke. The model shows how lateral interactions produce dynamic receptive fields and predicts that adapting lateral interactions are fundamental to cortical reorganization. Based on the model, two techniques to accelerate recovery from stroke and cortical surgery are suggested.

1 Introduction

Until recently, it was believed that the structure of the cerebral cortex is essentially static after a critical period of early development. Recent results, however, 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 (for review see [4; 8; 2]). 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.

Previous models of cortical development and plasticity concentrated on the self-organization of the afferent connections to the cortex [9; 11; 3]. Lateral interactions

1

within the cortex were assumed to be predetermined and fixed, and the models were aimed at explaining how the afferent connection weights organize. However, recent studies show that lateral interactions in the cortex are highly plastic—the long-range lateral connections in the cortex change structure in response to input manipulations such as sensory deprivation and strabismus and adapt together with the afferent connections [1; 5; 6; 7]. Several aspects of cortical plasticity, such as the reorganization of the map in response to cortical lesions, involve significant adaptation of the lateral connections as well, and cannot be explained by these previous models.

A new model of cortical self-organization called LISSOM (Laterally Interconnected Synergetically Self-Organizing Map: [13; 14; 12]) was developed to explain how afferent and lateral connections could self-organize cooperatively and simultaneously to form topographic maps. The maps formed by LISSOM are continuously adapting structures in a dynamic equilibrium, and susceptible to changes in the distribution of external and intrinsic inputs. As a result, the model can account not only for the plasticity due to reorganizing afferent synapses, but also plasticity due to adapting lateral connections. This article (1) demonstrates how receptive fields in the cortex can be maintained dynamically by such lateral interactions, (2) how the self-organizing process accounts for the reorganization of the cortex after cortical lesions and (3) suggests techniques to accelerate recovery following cortical surgery and stroke.

2   The LISSOM Model

The LISSOM network is a sheet of neurons interconnected by short-range excitatory lateral connections and long-range inhibitory lateral connections (figure 1). Neurons receive input from a receptive surface or "retina" through the afferent connections. These connections come from overlapping patches on the retina called anatomical receptive fields, or RFs. The patches are distributed with a given degree of randomness. The $N \times N$ network is projected on the retina of $R \times R$ receptors, and each neuron is assigned a receptive field center (c_1, c_2) randomly within a radius $\rho * R$ ($\rho \in [0, 1]$) of the neuron's projection. Through the afferent connections, the neuron receives input from receptors in a square area around the center with side s. Depending on its location, the number of afferents to a neuron could vary from $\frac{1}{2}s \times \frac{1}{2}s$ (at the corners) to $s \times s$ (at the center).

The afferent and lateral weights are organized through an unsupervised learning process. At each training step, neurons start out with zero activity. The initial response η_{ij} of neuron (i, j) is based on the scalar product

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

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

2

the neuron, μ_{ij,r_1r_2} is the corresponding afferent weight, and σ is a piecewise linear approximation of the familiar sigmoid activation function. The response evolves over time through lateral interactions. At each time step, the neuron combines retinal activation with lateral excitation and inhibition:

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

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 $\eta_{kl}(t-\delta t)$ is the activity of neuron (k,l) during the previous time step. The constants γ_e and γ_i are scaling factors on the excitatory and inhibitory weights and determine the strength of the lateral interactions.

The primary effect of lateral interaction is to sharpen the contrast between areas of high and low activity. The activity pattern starts out diffuse and spread over a substantial part of the map, but within a few iterations of equation 2, converges into a stable focused patch of activity, or activity bubble. After the activity has settled, the connection weights of each neuron are modified. Both afferent and lateral connection weights adapt according to the same mechanism: the Hebb rule, normalized so that the sum of the weights is constant:

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

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. The larger the product of the pre- and post-synaptic activity $\eta_{ij} X_{mn}$, the larger the weight change. Therefore, connections between areas with correlated activity are strengthened the most; normalization then redistributes the changes so that the sum of each weight type for each neuron remains constant.

3   Development of Topographic Maps and Lateral Interaction

The LISSOM network was simulated with Gaussian spots of "light" on the retina as input. At each presentation, the activation ξ_{r_1,r_2} at the receptor (r_1, r_2) was given by:

$$\xi_{r_1,r_2} = \sum_{i=1}^{n} exp(-\frac{(r_1-x_i)^2 + (r_2-y_i)^2}{a^2}), \quad (4)$$

3

where n is the number of spots, 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.

Figures 2—4 illustrate the self-organization of the LISSOM network. The afferent connections from the retina were initially ordered topographically, but their synaptic weights were completely random (figure 2a). Also, because several neurons connected to the same area of the retina, there was considerable overlap in anatomical receptive fields, and the initial topographic map was locally disordered (figure 3a). During self-organization, the initial rough pattern of afferent weights of each neuron evolved into a hill-shaped profile (figure 2b). As the afferent weight profiles of neurons peaked over different parts of the retina, their center of gravities (calculated in retinal coordinates) formed a precise topographical map (figure 3b).

The lateral connections evolve together with the afferents. By the normalized Hebbian rule (equation 3), the lateral connection weights of each neuron are distributed according to how well the neuron's activity correlates with the activities of the other neurons. As the afferent receptive fields organize into a uniform map (figure 3), these correlations fall off with distance approximately like a Gaussian, with strong correlations to near neighbors and weaker correlations to more distant neurons. The lateral excitatory and inhibitory connections acquire the Gaussian shape, and the combined lateral excitation and inhibition becomes an approximate difference of Gaussians (or a "Mexican hat"; figure 4).

Even after self-organization, the afferent and lateral connections in the network are not static. Each time an input is presented, the synaptic weight patterns adapt. As long as the distributions of afferent and lateral inputs seen by each neuron are stable, the map remains in a dynamic equilibrium and the weight patterns fluctuate around their self-organized state. However, when either of the distributions change, this equilibrium is altered, and the network reorganizes to compensate. Below, it is shown how cortical lesions change this self-organized state, and how it may be possible to accelerate the network's compensating reorganization.

4 Self-Organization After a Cortical Lesion

To study the effects of cortical lesions, a small set of neurons in the organized network were made unresponsive to input. Three phases of reorganization were observed, as in neurobiological studies [8]. Initially, the lesion reduces the inhibition of the perilesion neurons, and unmasks previously suppressed input activation. Therefore, immediately after the lesion, perilesion neurons begin responding to a broader range of inputs than before and their receptive fields appear to expand. The expansion is symmetric about the original receptive field centers, and the centers themselves do not shift. As a consequence of the expansion, perilesion receptive fields overlap to a greater degree with the receptive fields of the lesioned neurons. In effect, the neurons

4

right outside the lesioned area 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 5b).

The lesion disrupts the dynamic equilibrium of the network, and both lateral and afferent connections of the active neurons adapt to forge a new balance. 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 in the boundary neurons (figure 6). As a result, the lateral inhibition very rapidly becomes strong outside the lesion, and the activity that was previously unmasked is partly suppressed (figure 5c). Because the suppression is strongest at the boundary of the lesion, the receptive fields of boundary neurons appear to move outward. The functional loss is exacerbated, and there appears to be a regression from the initial recovery phase.

Even after the lateral connections reorganize, inputs that were previously stimulating the lesioned zone activate the boundary neurons. Driven by the Hebbian self-organizing mechanism, the afferent weights reorganize so that neurons respond better to these inputs. Gradually, receptive fields shift back inwards and representation of the receptive surface within the lesion zone is taken over by the neurons around it (figure 7). As a result, the cortical lesion is partly compensated for, as observed after stroke.

5   Discussion

The LISSOM model suggests two techniques to accelerate recovery following surgery or stroke in the sensory cortices. Normally, the recovery time after cortical surgery would include some immediate recovery, a phase of regression due to the reorganization of inhibition, and gradual and slow compensation afterward. The regression phase could be ameliorated if a transient blocker of inhibitory neurotransmitters were applied locally around the surgical area. Neurons around this area would then fire intensively because of reduced inhibition, and afferent connections would adapt rapidly to compensate for the lesion. By the time the blocker goes away, a substantial number of afferent receptive fields would have shifted and compensated for the lesion. Though the inhibition would strengthen when the blockade disappears, the pace of recovery would have been hastened.

Secondly, the receptive fields of perilesion neurons could be forced to shift and the topographic map reorganized as in figure 7 even before surgery. This could be

achieved by intensive and repetitive stimulation of the area expected to lose sensation and by sensory deprivation of its surroundings. Driven by the excessive stimulation, neurons oustside the surgical zone would shift receptive fields inward. Then, after surgery, the receptive fields would have to move much less to reach their 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 [10]. LISSOM provides a computational explanation of why such dynamics occur, and illustrates the primary role of lateral interactions in cortical plasticity.

6   Conclusion

The LISSOM model demonstrates that not only the self-organization of topographic maps, but also many aspects of cortical lesion plasticity can be explained based on the simultaneous adaptation of afferent and lateral connections. 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 suggests that functional recovery after cortical surgery may be hastened by blocking lateral inhibition locally in the cortex and by forced presurgical reorganization of cortical topographic maps.

References

[1]   R. A. Fisken, L. J. Garey, and T. P .S. Powell. The intrinsic, association and commissural connections of area 17 of the visual cortex. *Proceedings of the IEEE*,

61:268–278, 1975.

[2] C. D. Gilbert. Horizontal integration and cortical dynamics. *Neuron*, 9:1–13, July 1992.

[3] Granger G. Sutton III, James A. Reggia, Steven L. Armentrout, and C. Lynne D'Autrechy. Cortical map reorganization as a competitive process. *Neural Computation*, 6:1–13, 1994.

[4] Jon H. Kaas. Plasticity of sensory and motor maps in adult animals. *Annual Review of Neuroscience*, 14:137–167, 1991.

[5] Lawrence C. Katz and Edward M. Callaway. Development of local circuits in mammalian visual cortex. *Annual Review of Neuroscience*, 15:31–56, 1992.

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

[7] James S. McCasland, Kerry L. Bernardo, Karl L. Probst, and Thomas A. Woolsey. Cortical local circuit axons do not mature after early deafferentation. *Proceedings of the National Academy of Sciences, USA*, 89:1832–1836, March 1992.

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

[9] K. Obermayer, H. J. Ritter, and K. J. Schulten. 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, 1990. Elsevier.

[10] Mark W. Pettet and Charles D. Gilbert. Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences, USA*, 89:8366–8370, 1992.

[11] Helge J. Ritter, Thomas Martinetz, and Klaus J. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, Reading, MA, 1992.

[12] Joseph Sirosh and Risto Miikkulainen. How lateral interaction develops in a self-organizing feature map. In *Proceedings of the IEEE International Conference on Neural Networks* (San Francisco, CA), Piscataway, NJ, 1993. IEEE.

[13] Joseph Sirosh and Risto Miikkulainen. Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71(1):66–78, 1994.

[14] Joseph Sirosh and Risto Miikkulainen. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. submitted to *Neural Computation*, 1994.

Figure 1: **The LISSOM architecture.** The afferent and lateral connections of a single neuron in the network are shown. All connection weights are positive.

Figure 2: **Self-organization of the afferent input weights.** The afferent weights of five neurons (located at the center and at the four corners of the network) are superimposed on the retinal surface in this figure. The retina had 21×21 receptors, and the receptive field radius was chosen to be 8. Therefore, neurons could have anywhere from 8×8 to 17×17 afferents depending on their distance from the network boundary. (a) The anatomical RF centers were topographically ordered, and the weights were initialized randomly. There are four concentrated areas of weights slightly displaced from the corners, and one larger one in the middle. At the corners, the profiles are taller because the normalization divides the total afferent weight among a smaller numer of connections. (b) As the self-organization progresses, the weights organize into smooth hill-shaped profiles. In this simulation, each input consisted of 3 randomly-located Gaussian spots with $a = 2.0$. The lateral interaction strengths were $\gamma_e = \gamma_i = 0.9$, with total lateral excitation = total inhibition = 1.0. The learning rates were $\alpha_A = \alpha_E = \alpha_I = 0.002$, and the upper and lower thresholds of the sigmoid were 0.65 and 0.1 respectively. Only the parameters γ_e and the sigmoid's upper threshold were somewhat sensitive.

Figure 3: **Self-organization of scattered receptive fields into a topographic map.** The center of gravity of the afferent weight vector of each neuron in the 64×64 network is projected onto the receptive surface (represented by the square). Each center of gravity point is connected to those of the four immediately neighboring neurons by a line. The resulting dark grid depicts the topographical organization of the map. In (a), the anatomical RF centers were topographically ordered, but because the afferent weights were initially random, the center of gravities are locally scattered. As the self-organization progresses, the network unfolds and the weight vectors spread out to form a regular topographic map of the receptive surface, such as shown in (b).
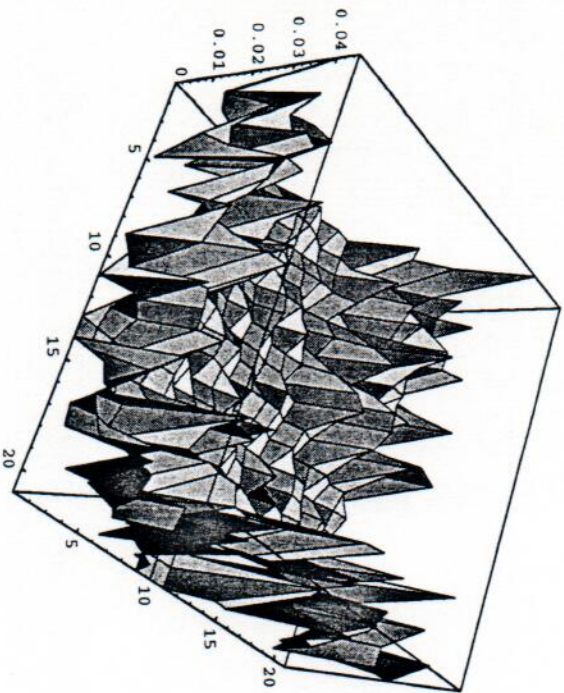
Figure 4: **Self-organization of the lateral interaction.** The lateral interaction profile for a neuron at position $(32, 32)$ in the 64×64 network is plotted. The excitation and inhibition weights are initially randomly distributed within radii 3 and 18. The combined interaction is the sum of the excitatory and inhibitory weights and illustrates the total effect of the lateral connections. The sums of excitation and inhibition were chosen to be equal, but because there are fewer excitatory connections, the interaction has the shape of a rough plateau with a central peak (a). During self-organization, smooth patterns of excitatory and inhibitory weights evolve, resulting in a smooth "Mexican hat" shaped lateral interaction profile (b).

Figure 5: **How response patterns change after a cortical lesion.** The activity of neurons across the network are shown for the same input before the lesion (a), immediately after (b) and a few hundred adaptation steps later (c). The lesioned area is seen as a white square with no activity in figure (b). 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 (figure 6), the unmasked activity decreases because of increased lateral inhibition.
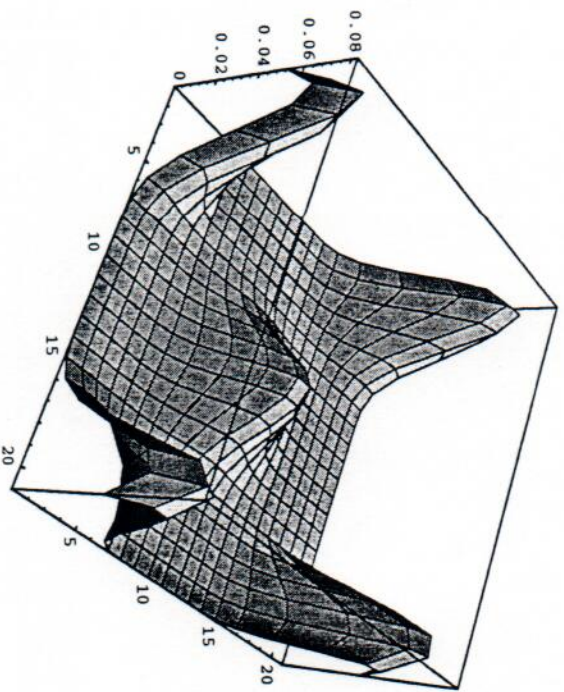
Figure 6: **Reorganization of lateral inhibition at the lesion boundary.** The inhibitory connections of a neuron at the boundary of the lesion are shown. The neuron has 40×40 connections, and the prelesion inhibition is circularly symmetric around the neuron (a). Shortly after the lesion, the inhibitory weights from the lesioned neurons decrease to zero. Because the total inhibitory weight is kept constant by weight normalization, the inhibition concentrates in the connections outside the lesioned zone, and the trough becomes deeper (b).

Figure 7: **Topography and activity in the reorganized network.** 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. 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.

Receptive Surface

Network

Long range
inhibitory
connections

Short range
excitatory
connections

Afferent connections

Receptors

Anatomical
Receptive Field

Neurons

1
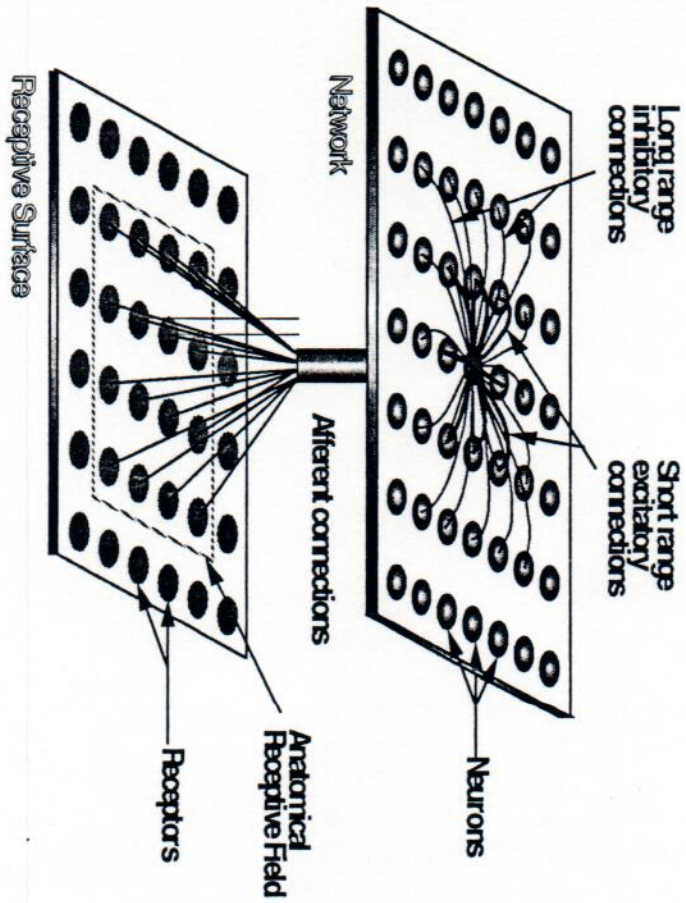
(a) Initial random weights

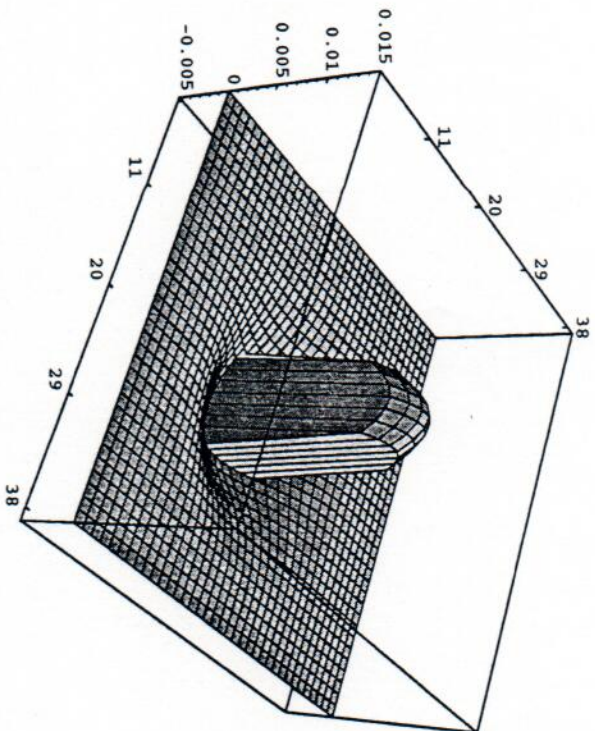(b) Final organized receptive fields
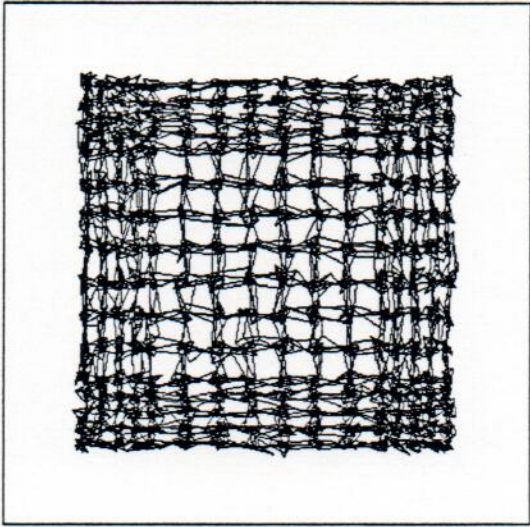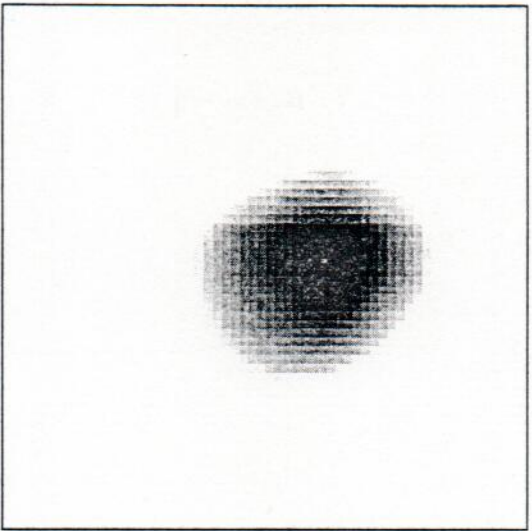
(a) Initial disordered map
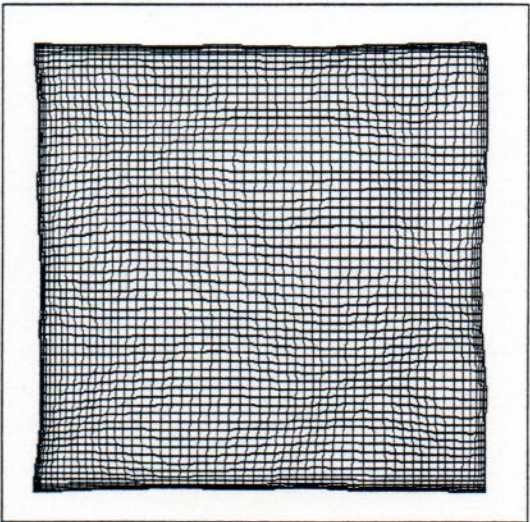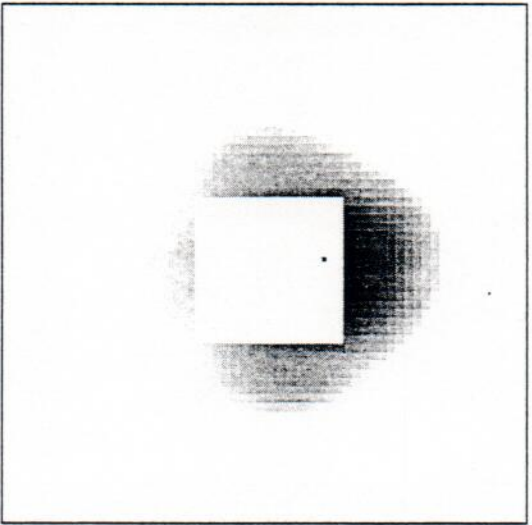
(b) Final ordered map
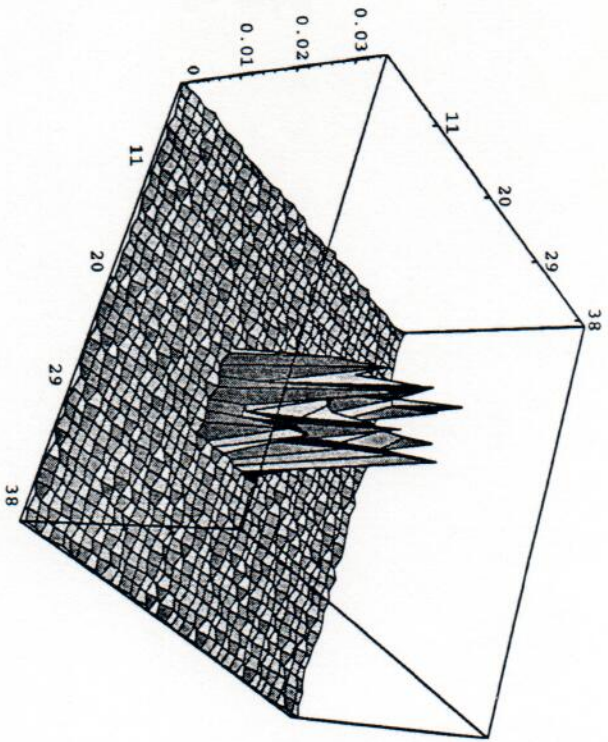
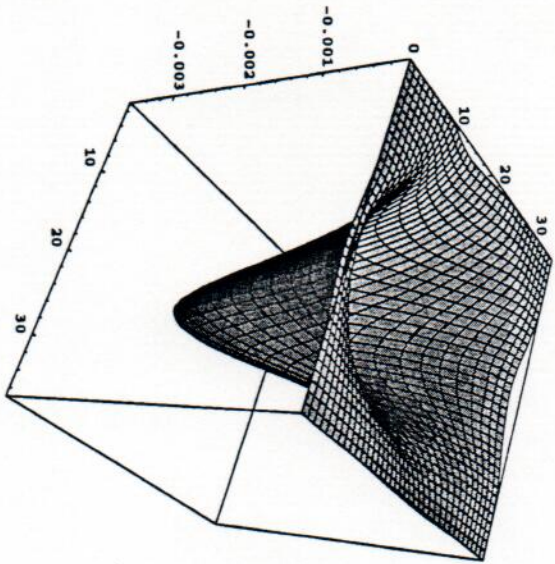(a) Initial rough interaction profile
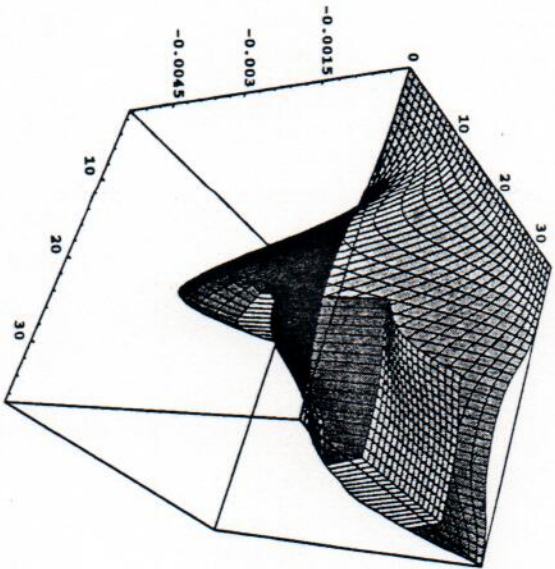

(b) Final smooth interaction

(a) Activity before lesion
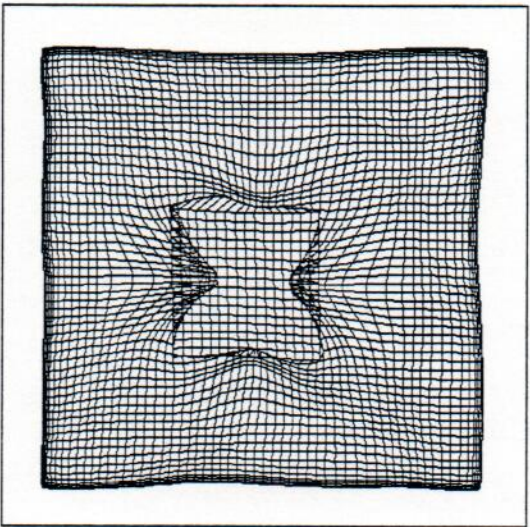
(b) Immediately after
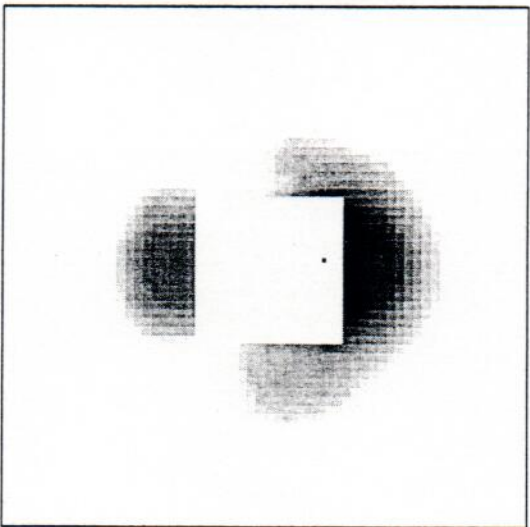
(c) After 500 iterations

(a) Inhibition before lesion

(b) After 500 steps

(a) Reorganized map

(b) Final activity