REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Institute for Brain and Neural Systems		Personnel and Training Research Programs Office of Naval Research (Code 1142PT)

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Brown University Providence, Rhode Island 02912	800 North Quincy Street Arlington, VA 22217-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
		N00014-91-J-1316

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Organization of Receptive Fields in Networks with Hebbian Learning: The Connection Between Synaptic and Phenomenological Models

12 PERSONAL AUTHOR(S)
Harel Shouval and Leon N. Cooper

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM _____ TO _____		

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Hebbian Learning, Phenomenological Models Receptive Field, Organization
05	08		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this paper we address the question of how does the lateral interaction affect the formation and organization of receptive fields in a network composed of interacting neurons with Hebbian type learning. We will show how to partially decouple the single cell effects from the network effects, and how some phenomenological models can be seen as approximations to these learning networks. We show that the interaction effects the structure of receptive fields. We also demonstrate how the organization of different receptive fields across the cortex is influenced by the interaction term, and that the type of singularities depend on the symmetries of the receptive fields.

DTIC QUALITY INSPECTED 3

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Joel Davis	(703) 696-4744	

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
                              All other editions are obsolete.

Organization of Receptive Fields in Networks with Hebbian Learning: The Connection Between Synaptic and Phenomenological Models

Harel Shouval and **Leon N Cooper**
Department of Physics, The Department of Neuroscience and
The Institute for Brain and Neural Systems
Box 1843, Brown University
Providence, R. I., 02912
hzs@cns.brown.edu

May 30, 1995

Abstract

In this paper we address the question of how does the lateral interaction affect the formation and organization of receptive fields in a network composed of interacting neurons with Hebbian type learning. We will show how to partially decouple the single cell effects from the network effects, and how some phenomenological models can be seen as approximations to these learning networks. We show that the interaction effects the structure of receptive fields. We also demonstrate how the organization of different receptive fields across the cortex is influenced by the interaction term, and that the type of singularities depend on the symmetries of the receptive fields.

1 Introduction: Synaptic Plasticity and the structure of the Visual Cortex

Recently Erwin et. al. (95) performed a comparison between predictions of many models, which attempt to explain organization of receptive fields in the visual cortex, and cortical maps obtained by optical imageing (Blasdel, 1992; Bonhoeffer and Gringald, 1991). In this paper we will consider two classes of such models; detailed synaptic models (Miller, 1992; Miller, 1994; Linsker, 1986a), and iterative phenomenological models (Swindale, 1982; Cowan and Friedman, 1991). Erwin et. al. (94) have examined these classes of models claim that the phenomenological models fit the experimental observations much better than the synaptic models.

Detailed synaptic models are fairly complex systems, with a large number of parameters. The experience of the last couple of decades in physics of complex systems has shown that phenomenological models are extremely important in explaining properties of complex systems. For complex systems of many interacting particles it is typically hopeless to try and solve the exact microscopic theory of these interacting systems, however in many cases coarse grained phenomenological models, that do not take into account most of the microscopic detail, do manage to predict the properties

1

of such systems. It seems that if this is the case for the relatively simple complex systems of the physical sciences, then it might also be the case for the much more complex systems of neurobiology.

In this paper we show how the detailed synaptic models can be approximated by much simpler models that are related to some of the phenomenological models. These simpler models can be used in order to gain a better understanding of the properties of the complex synaptic models and possibly to understand where reported the differences in their properties (Erwin et al., 1995) arise from.

We analyze a cortical network of learning neurons, each of the neurons in this network forms plastic bonds with the sensory projections, and fixed bonds with other cortical neurons. The learning rule assumed herein is a normalized Hebbian rule(Oja, 1982; Miller and MacKay, 1994). The type of network architecture assumed here is similar to that used in many other synaptic models of cortical plasticity (von der Malsburg, 1973; Nass and Cooper, 1975; Linsker, 1986b; Miller et al., 1989; Miller, 1994, etc.). Rather then presenting a new model we have attempted, in this paper, to analyze the behavior of these existing models, and map them onto much simpler models which we might be able to understand better When this analysis is be performed, it becomes clear that in order to understand the behavior of such networks, it is first necessary to understand how single neurons using the same learning rule behave.

The purpose of this paper is to start giving answers to two related questions about such networks.

(a) How do the lateral interactions affect the receptive fields of single cortical neurons, and in particular can cortical interactions produce oriented receptive fields, where the single cell model produces radially symmetric receptive fields.

(b) How do the cortical interactions affect the organization of different receptive fields across the cortex.

A similar analysis has been used by Rubinstein (94) in order to answer the second question. His analysis differences form our in that it concentrates on a periodic homogenous network, and does not examine the effect of the lateral interactions on receptive field formation.

2   A self organizing network model

We assume a network of interconnected neurons, which are connected with modifiable synapses to the input layer. The learning rule examined is a normalized Hebbian rule which is sensitive only to the second order statistics.

The output $c(\mathbf{r})$, of each neuron, in this model is defined as

$$c(\mathbf{r}) = \sum_{x,\mu} I(\mathbf{r} - \mathbf{x}) A(\mu - \beta(\mathbf{x})) d(\mu) m(\mathbf{x}, \mu).$$

where β is the center of the symmetric arbor function A, and is a function of $\mathbf{x}$. I is the cortical interaction function, m are the weights, d the input vectors, μ denotes points in input space, $\mathbf{r}$ and $\mathbf{x}$ in output space.

The energy function H, we will try to minimize is now

$$H \quad = -\sum_r c^2(\mathbf{r}) - \gamma(\mathbf{m}) \tag{1}$$

2

$$= -\sum_{r,x,x'} I(\mathbf{r} - \mathbf{x})I(\mathbf{r} - \mathbf{x}')$$

$$\sum_{\mu,\mu'} A(\mu - \beta(\mathbf{x}))A(\mu' - \beta(\mathbf{x}'))m(\mathbf{x},\mu)m(\mathbf{x}',\mu')d(\mu)d(\mu') - \gamma(\mathbf{m})$$

where γ is an abstract constraint on the allowed values of $\mathbf{m}$.

We now perform the standard procedure, of turning this into a correlational model, by taking the ensemble average over the input environment, and also taking the sum over $\mathbf{r}$. Thus

$$H = -\sum_{x,x'} \tilde{I}(\mathbf{x} - \mathbf{x}') \tag{2}$$

$$\sum_{\mu,\mu'} A(\mu - \beta(\mathbf{x}))A(\mu' - \beta(\mathbf{x}'))m(\mathbf{x},\mu)m(\mathbf{x}',\mu')Q(\mu - \mu') - \gamma(\mathbf{m}).$$

Where $\tilde{I}(\mathbf{x} - \mathbf{x}') = \sum_r I(\mathbf{r} - \mathbf{x})I(\mathbf{r} - \mathbf{x}')$ is the new effective interaction term, and $Q(\mu - \mu') = < d(\mu)d(\mu') >$ is the correlation function.

From now on, in order to simplify matters, we assume a step arbor function of the form,

$$A(\mathbf{x}) = \begin{cases} 1 & \text{for } |\mathbf{x}| \leq \mu_{max} \\ 0 & \text{for } |\mathbf{x}| > \mu_{max} \end{cases} \tag{3}$$

Using this function is justified since it can produce oriented receptive fields in both network models (Miller, 1994) and single cell models (Liu, 1994).

The gradient descent dynamics implied by this Hamiltonian are now

$$\frac{dm(\mathbf{x},\mu)}{dt} = \begin{cases} \displaystyle\sum_{\mathbf{x}'} \sum_{|\mu' - \beta(x')| < \mu_{max}} \tilde{I}(\mathbf{x} - \mathbf{x}')m(\mathbf{x}',\mu')Q(\mu - \mu') & \text{for } |\mathbf{x} - \beta(\mathbf{x})| < \mu_{max} \\ 0 & \text{for } |\mathbf{x} - \beta(\mathbf{x})| > \mu_{max} \end{cases} \tag{4}$$

This equation is the same as Millers (1994) synaptic plasticity equations for the step arbor function.

The method we use now, is to represent the weights m in terms of the complete set, i.e $m(\mu,\mathbf{x}) = \sum_{ln} a_{ln}(\mathbf{x})m_{ln}(\mu,\mathbf{x})$. We have chosen this complete set to be a solution of the non interacting eigenvalue problem.

$$\sum_{|\mu' - \alpha|} m_{ln}(\mu',\mathbf{x})Q(|\mu - \mu'|) = \lambda_{ln}m_{ln}(\mu - \alpha, \mathbf{x}), \tag{5}$$

In a radially symmetric environment the eigenfunctions will take the form

$$m_{ln}(\mu,\mathbf{x}) = m_{ln}(\mu, \phi_{ln}(\mathbf{x})) = e^{il\phi_{ln}(\mathbf{x})}U_{ln}(\mu)$$

where $\phi_{ln}(\mathbf{x})$ is an arbitrary phase which is undetermined, due to the radially symmetric nature of the correlation function.

Using this expansion, and taking sum over μ the Hamiltonian takes the form,

3

$$H = -\sum_{x,x'}\sum_{lnl'n'} \tilde{I}(\mathbf{x} - \mathbf{x}')\lambda_{ln}a_{ln}(\mathbf{x})a_{l'n'}(\mathbf{x}')$$
$$O\left[l,n,l',n',\phi_{ln}(\mathbf{x}),\phi_{ln}(\mathbf{x}'),\beta(\mathbf{x}),\beta(\mathbf{x}')\right] - \gamma(a) \qquad (6)$$

Where O the overlap between two receptive fields, is

$$O\left[l,n,l',n',\phi_{ln}(\mathbf{x}),\phi_{ln}(\mathbf{x}'),\beta(\mathbf{x}),\beta(\mathbf{x}')\right]$$
$$= \sum_{\mu'} m_{ln}\left(\mu' - \beta(\mathbf{x}),\phi_{ln}(\mathbf{x})\right) m_{l'n'}\left(\mu' - \beta(\mathbf{x}'),\phi_{l'n'}(\mathbf{x}')\right). \qquad (7)$$

It is important to notice that in general, O depends on the distance between the two arbor functions, as well as on the shapes and phases of the two eigen-states.

3 Analysis of the Common Input model:

We will now make the simplification that $\beta = \alpha = 0$ that is that all neurons have an arbor function with the same center point, this simplification makes analysis simpler, and is a good approximation when the number of output neurons is much greater than the dimensionality of the input.

It is important to notice though that the conclusions which we deduce from this Common input model, may not be applicable to more complex models. In particular it is important to notice that in other models, O may change signs as a function of $|\mathbf{x} - \mathbf{x}'|$ and thus turn an excitatory model, i.e: one in which $I \geq 0$ to one in which there is effectively an inhibitory portion to the interaction.

For this choice of arbor function the overlap O takes a very simple form,

$$O\left[l,n,l',n',\phi_{ln}(\mathbf{x}),\phi_{ln}(\mathbf{x}')\right] = \delta_{ll'}\delta_{nn'}\exp^{il[\phi_{ln}(\mathbf{x})-\phi_{ln}(\mathbf{x}')]}.$$

Thus for the common input model the Hamiltonian takes the form

$$H = -\sum_{xx'lm} \tilde{I}(\mathbf{x} - \mathbf{x}')\lambda_{ln}a_{ln}(\mathbf{x})a_{ln}(\mathbf{x}')e^{il[\phi_{ln}(\mathbf{x})-\phi_{ln}(\mathbf{x}')]} - \gamma(a). \qquad (8)$$

When observing equation 8 it is easy to see that the eigenvalues of the correlation function, play a major role in the resulting Hamiltonian. They could serve as an importance parameter by which approximations can be made. Hence the approximation proposed here is to neglect all terms except those multiplying the largest eigenvalues, these approximations will be termed herein simplified models.

In order to make things concrete, we assume that $\lambda_{01} = \lambda_0$, and $\lambda_{11} = \lambda_1$ are the largest eigenvalues and considerably larger then the other eigen-values. This is not an artificial choice, since in many circumstances this indeed seems to be the case (MacKay and Miller, 1990; Liu and Shouval, 1994; Shouval and Liu, 1995), although in other regimes other eigen-states will dominate.

Under this approximation the Hamiltonian will take the form.

$$H = -\sum_{\mathbf{rr}'} \tilde{I}(\mathbf{r} - \mathbf{r}')\left\{\lambda_0 a(\mathbf{r})a(\ r') + \lambda_1 b(\mathbf{r})b(\mathbf{r}')cos[\phi(\mathbf{r}) - \phi(\mathbf{r}')]\right\} - \gamma(a,b)$$

4

Where $a = a_{01}$ and $b = a_{11}$. Different types of constraints can be considered. Given the constraint that the weights are square normalized, i.e that $\sum_\mu m(\mu, \mathbf{x})^2 = 1$,[1] which for the approximate model takes the form $a^2(\mathbf{r}) + b^2(\mathbf{r}) = 1$, the model is equivalent to the non isotropic Heisenberg spin model embedded in two dimensions. This can be made more evident when this is rewritten in two equivalent forms. If we set $a(\mathbf{r}) = S_z(\mathbf{r}) = cos(\theta(\mathbf{r}))$, $b(\mathbf{r}) = sin(\theta(\mathbf{r}))$, $S_x(\mathbf{r}) = b(\mathbf{r})cos(\phi(\mathbf{r}))$ and $S_y(\mathbf{r}) = b(\mathbf{r})sin(\phi(\mathbf{r}))$. then the Hamiltonian takes the form,

$$H = -\sum_{\mathbf{rr'}} \tilde{I}(\mathbf{r} - \mathbf{r'}) \left\{ \lambda_0 S_z(\mathbf{r}) S_z(\mathbf{r'}) + \lambda_1 [S_x(\mathbf{r}) S_x(\mathbf{r'}) + S_y(\mathbf{r}) S_y(\mathbf{r'})] \right\} \tag{9}$$

Subject to the constraint that $S_x^2 + S_y^2 + S_z^2 = 1$.

This can also be put in the closed form,

$$
\begin{aligned}
H &= \sum_{\mathbf{r,r'}} I(\mathbf{r} - \mathbf{r'}) \{ \lambda_0 \cos(\theta(\mathbf{r})) \cos(\theta((\mathbf{r'})) \\
&+ \lambda_1 \sin(\theta(\mathbf{r})) \sin(\theta(\mathbf{r'})) [\cos(\phi(\mathbf{r})) \cos(\phi(\mathbf{r'})) + \sin(\phi(\mathbf{r})) \sin(\phi(\mathbf{r'}))] \}.
\end{aligned}
\tag{10}
$$

Which is exactly the nonisotropic Heisenberg model of Statistical Mechanics.

Thus it can be seen that each neuron is associated with only two dynamical variables ϕ and θ, thus this approximation has resulted in a major simplification in the complexity of the model.

If $\lambda_1 >> \lambda_0$ and we can discard of the terms multiplying the λ_0 this becomes an XY model and is equivalent to the phenomenological model proposed by Cowan and Friedmann (Cowan and Friedman, 1991). Thus in this case the Hamiltonian takes the simple form

$$H = -\sum_{\mathbf{rr'}} I(\mathbf{r} - \mathbf{r'}) \cos(\phi(\mathbf{r}) - \phi(\mathbf{r'})) , \tag{11}$$

and the dynamics are

$$\dot{\phi}(\mathbf{r}) = \eta \sum_{\mathbf{r'}} I(\mathbf{r} - \mathbf{r'}) \sin(\phi(\mathbf{r}) - \phi(\mathbf{r'})) . \tag{12}$$

Another, well known, phenomenological models is the Swindale model (82) for the development of orientation selectivity, we will show that it is similar to the Cowan Friedmann model. Swindale defines the complex variable $Z(\mathbf{r}) = S_x(\mathbf{r}) + iS_y(\mathbf{r})$ or $Z(r) = b(\mathbf{r}) \exp(i\phi(\mathbf{r}))$ [2].

For which he defines the dynamics

$$\dot{Z}(\mathbf{r}) = \sum_{\mathbf{r'}} I(\mathbf{r} - \mathbf{r'}) Z(\mathbf{r'})(1 - |Z(\mathbf{r})|). \tag{13}$$

Thus for the variables $\phi(\mathbf{r}) = tg^{-1}(S_x/S_y)$ and $b(\mathbf{r}) = |Z(\mathbf{r})|$, the dynamics are

$$\dot{\phi}(\mathbf{r}) = \eta(1 - b(r)) \sum_{\mathbf{r'}} I(\mathbf{r} - \mathbf{r'}) \sin(\phi(\mathbf{r}) - \phi(\mathbf{r'})) \tag{14}$$

---

[1] This is the constraint to which the Oja (82) rule converges

[2] This angle is different than the one Swindale chooses for the orientation angle, He defines the orientation angle ϕ_s as $\phi_s = \frac{1}{2}\phi$.

and

$$\dot{b}(\mathbf{r}) = \eta\,(1 - b(\mathbf{r}))\,b(\mathbf{r}) \sum_{\mathbf{r}'} I(\mathbf{r} - \mathbf{r}') \cos\left(\phi(\mathbf{r}) - \phi(\mathbf{r}')\right) \qquad (15)$$

Comparing equations 12 and 14 we see that for $b(\mathbf{r}) < 1$ the dynamics of the angle ϕ in the Swindale model are parallel to those in the Cowan model. However, when $b(\mathbf{r}) = 1$ the Swindale dynamics terminate, and when $b(\mathbf{r}) > 1$ they are anti parallel to the Cowan and Friedmann dynamics. The dynamics for $b(\mathbf{r})$ when $b(\mathbf{r}) < 1$ are such that if $h(\mathbf{r}) = \sum_{\mathbf{r}'} I(\mathbf{r} - \mathbf{r}') \sin(\phi(\mathbf{r}) - \phi(\mathbf{r}')) > 0$ then $b(\mathbf{r}) \to 1$, this would be the common case since the dynamics for $\phi(\mathbf{r})$ try to maximize $h(\mathbf{r})$ however in some cases, such as near orientation singularities this may not be possible and in these cases $b(\mathbf{r}) \to 0$.

Thus it can be expected that if the Swindale simulations are started at small values of $b(\mathbf{r})$ and the learning rate is slow enough, then the fixed points of the orientation maps would be similar to the Cowan and Friedmann fixed points.

In synaptic space the constraint imposed in the Swindale model is equivalent to stopping the update of each neuron once it's synaptic weight becomes normalized. This seems a non local constraint, although it could possibly be made local by a procedure similar to the one Oja has used to attain square normalization. The final state of a Swindale type model can depend on the initial conditions, and on the learning rate η, thus the Swindale model has extra (spurious) fixed points, which the Cowan and Friedmann model does not possess.

In the following sections we will use the simplified model, as defined in equations 9 to 10 to answer the questions posed in the introduction.

4   The Excitatory case $I \geq 0$

In the Excitatory case, we know that the global minimum of the common input energy function (eq:8) as well as in the approximation (eq:10), occurs when all receptive fields are equivalent to the principal component, and have the same orientation.

Since we know what the global minim is, it is interesting to see how simulations of the simplified model manage to reach this minimum. Doing this is a test of our procedures, but it also has interesting implications for simulations of detailed synaptic models of the Miller, Linsker type.

In Figure 1 we show the time evolution of the excitatory simplified model, when $\lambda_1/\lambda_0 = 1.3$, i.e: the oriented solution has a higher eigenvalue. In these simulations we used a Gaussian interaction function, with a standard deviation $\sigma = 2.5$. We have defined one iteration in this network as the number of random updates, required for updating each neuron on average once.

We can see that the global minimum is indeed reached, however it takes several hundred iterations to reach the global minimum for the parameters we have used, which we have tried to optimize. The neurons attain their maximal selectivity quickly, most of them after 30 iterations, however it takes a very long time in comparison for the neurons to become parallel.

Simulations of the detailed synaptic models which are much more complex, were typically run with a smaller number of iteration (Miller, 1994). In the Miller model the learning rule used was unstable, therefore saturation limits were set on the synapses. When synapses reached these limits they were frozen, these type of bounds we term sticky bounds. When a high fraction of the synapses become saturated, the simulations are terminated. Since in our simulations neurons their maximal degree of orientation selectivity long before they became parallel, it implies that if the
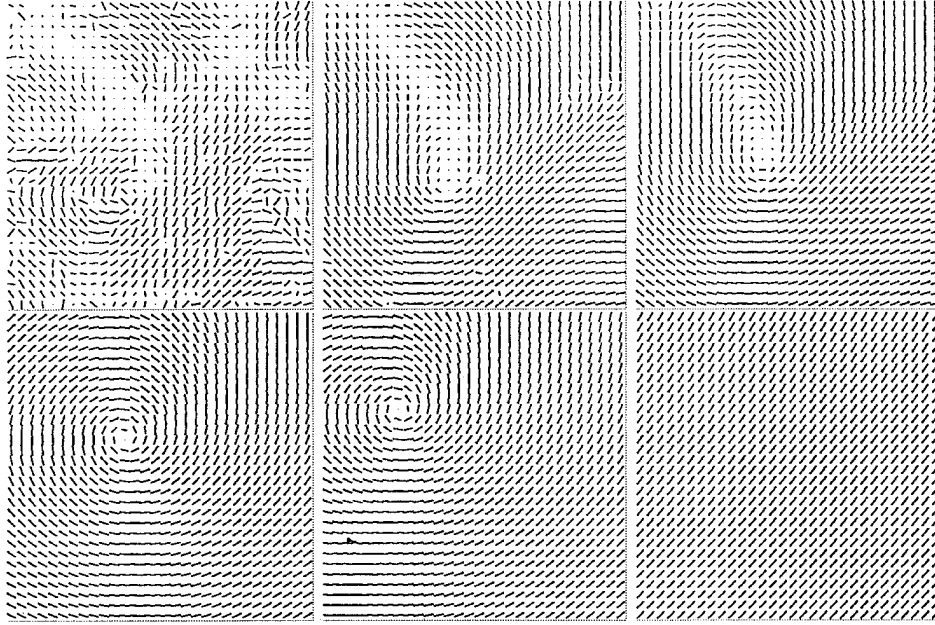
6

Figure 1: *Time evolution of orientation in an excitatory network. The images show the orientations in a 30 by 30 network of neurons, which were run for 5,15,30,100,400 and 600 iterations, displayed from left to right, top to bottom. The direction of the line codes orientation, and it's intensity codes for the degree of orientation selectivity. A great majority of the cells have already attained a high degree of selectivity, at 100 iterations, but it takes several hundred more iterations for them to become completely parallel.*

sticky bounds were not used by Miller (Miller, 1994) and if simulations were run for more iterations, then the resulting orientation maps may be quite different. Thus sticky boundaries are more than a computational convenience. and they should either have biological motivation, or be omited from simulations.

Another difference in the way these simulations, and the detailed synaptic simulations have been performed is that we used random sequential dynamics, whereas parallel dynamics have been used in the simulations of the complete synaptic models (Miller, 1994; Linsker, 1986b). This implies that all synapses in the network are updated synchronously, i.e: at once. This assumption does not seem biologically plausible, furthermore although parallel dynamics are computationally convenient, they may alter the final state the network will reach and in particular they can prevent the model from obtaining the real energy minima.

5   Interactions with Inhibition

In the case where there is some inhibition in the network, the situation is more complex, and more interesting. In this case we do not know what the global minimum is, we will therefore use simulations of the simplified models in order to examine numerically what these minima are.

Using the simplified model has the great advantage of being much less complex, and therefore requiring much less computation, furthermore definitions such as orientation selectivity, and ori-

entation angle have a transparent definition. On the other hand it runs the risk of not being a valid approximation. This would happen if when running the full model, the final receptive fields would have strong components of the less prominent eigenfunctions which are not represented in the approximation.

In order to examine the amount of single cell symmetry breaking, we define an order parameter, $rms(a)$, which is defined as $rms(a) = \sqrt{<a^2>}$ where the average is taken over all network sites, this order parameter shows how much the single cell receptive fields in the network deviate from non interacting single cells. Since in the all excitatory case the values of $rms(a)$ should be like those in the non interacting case, we can asses the effect of the interaction by comparing the values of $rms(a)$ for interactions with inhibition, to those which have only excitation. Strong effects of the interaction may also imply that the approximation may break down.

In our simulations, we used an interaction function of the form,

$$\tilde{I}(\mathbf{r}) = \exp[-(r^2/2\sigma^2)] - IN \cdot exp[-(r^2/2(2\sigma)^2)],$$

where we have fixed $\sigma = 2.5$. Hence for simplicity we have varied only the value of IN. Cross sections of the different interaction terms used are shown in figure: 2.

The cases we have examined are **Excitatory** $IN = 0.0$, **Balanced** $IN = .25$, where the integral of the interaction function is zero, **Net Inhibitory** $IN = .5$ in which the is more inhibition then excitation, and **Net Excitatory** $IN = .125$ where there is more excitation then inhibition. We have run all of our simulations for at least 1500 iterations and tested the stability of the final state, in order to make sure that we have indeed reached a stable fixed point.
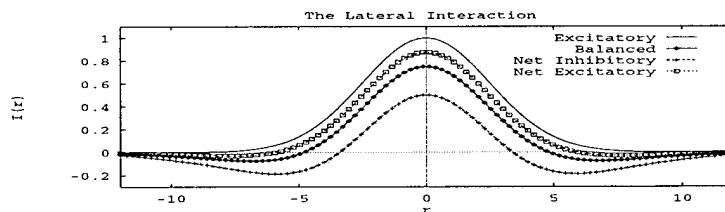

Figure 2: *This image presents the different types of interaction functions of the form $I(r) = \exp(-r^2/\sigma^2) - IN * \exp(-r^2/(2\sigma)^2)$.* **Excitatory** *$IN = 0.0$, i.e there is only excitation,* **Balanced** *$IN = .25$ i.e the inhibition and excitation are balanced,* **Net Inhibitory** *$IN = .5$ more excitation then inhibition, and* **Net Excitatory** *$IN = .125$ more excitation then inhibition.*

As can be seen figure 3 the interaction can indeed break the symmetry, although only mildly since under the most conditions investigated most receptive fields remain dominated by the first principal component. In figure 4, we chose to display the network, in an extreme case of symmetry breaking, even here most receptive fields have a strong, radially symmetric component.

The effect of the interaction on the global ordering is more pronounced. In figure 5, we can see that when the radially symmetric principal component is dominant, the effect of the interaction would be a staggered type of final state.

When the oriented solution dominates, the emerging structure has bands with similar orientations, and some orientation singularities as can be seen in figure: 6.

These singularities however, are of the $S = 1$ type which means that, around a singularity, the orientation angle changes by 360 deg. It has been claimed (Erwin et al., 1995) that in the cortex and
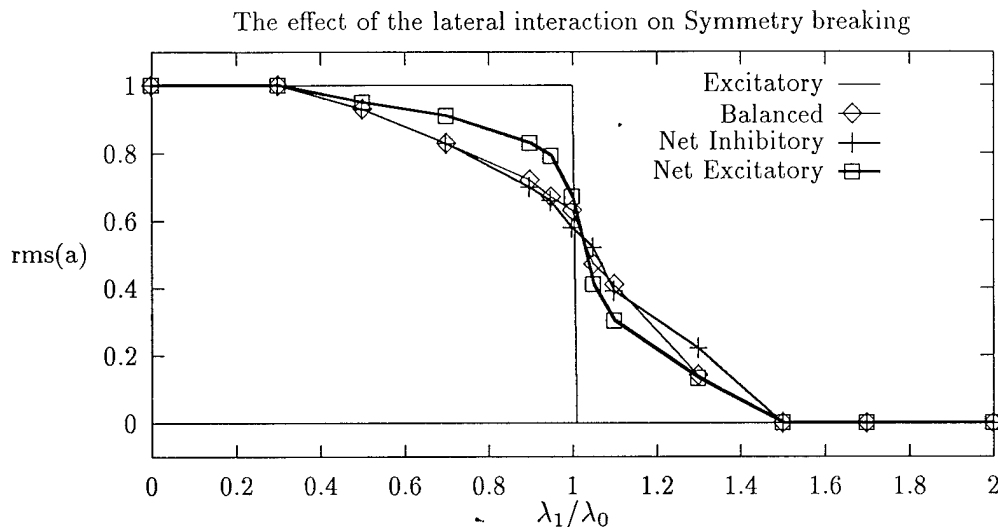
8

The effect of the lateral interaction on Symmetry breaking

Figure 3: *This image presents the rms value of a as a function of λ_1/λ_0 for the different types of interactions as defined above. Inhibitory interactions can create orientation selective cells in regions in which single cells are radially symmetric. The different types of interaction we have used have the form $I(r) = \exp(-r^2/\sigma^2) - IN * \exp(-r^2/(2\sigma)^2)$. and differ in the value of IN, Excitatory $IN = 0.0$, i.e there is only excitation, Balanced $IN = .25$ i.e the inhibition and excitation are balanced, Net Inhibitory $IN = .5$ more excitation then inhibition, and Net Excitatory $IN = .125$ more excitation then inhibition.*

in many models such as in the Swindale (92) model and in the Miller (94) model, the singularities are of the $S = \frac{1}{2}$ type. i.e the orientation angle changes by 180 deg, examples of such singularities are illustrated in figure 7.

This stands in contradiction to the results in the simplified model, which therefore raises the question of what is the source of these differences?

In the Swindale model the orientations angles are defined to be half of the angle of the complex variable z, i.e $\phi_s = \frac{1}{2}\phi$, and thus when ϕ has a $S = 1$ singularity ϕ_s has a $S = \frac{1}{2}$ singularity. The differences between the simplified model and the Swindale model actually arise from a similar reason. We claim that when the dominant single cell solution has the angular component $\cos(m\theta)$, or in general when a rotation of $(360/n)$ deg leaves the receptive field as it was, then the resulting singularities will be of the type $S = \frac{1}{m}$. The heuristic reasoning for this claim is the following; suppose that we divide the interaction function into two terms, an inhibitory term and an excitatory term, $I = I_E + I_I$ such that $I_E > 0$ and $I_I < 0$. It is usually assumed that the excitatory term dominates the short range interactions. The excitatory term attempts to enforce continuity, and low curvature, i.e: a small as possible change between neighboring orientations. Thus a closed path around a singularity should exhibit the minimal curvature which retains continuity which for receptive fields with an angular dependence $\cos(m\theta)$, it is a rotation around the curve by $360/m$ degrees, i.e a singularity of type $S = \frac{1}{m}$. More explicitly we predict that when the single cell solution has a $\cos(2\theta)$ angular dependence, then the singularities exhibited will be of the type $S = \frac{1}{2}$. In
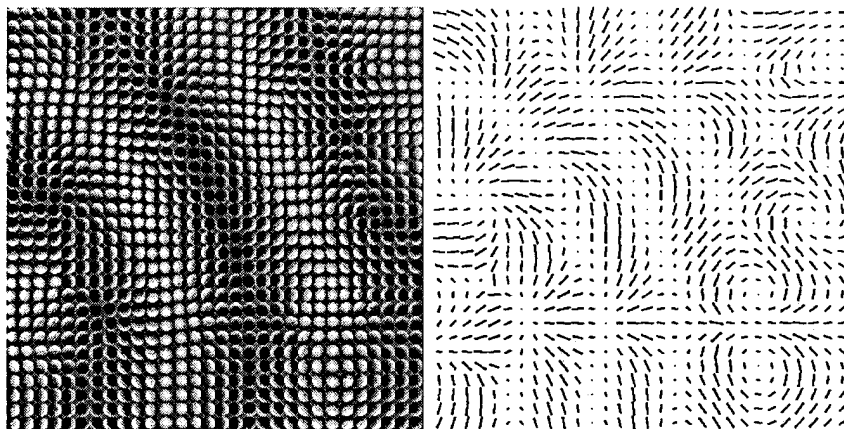
9

Figure 4: *Simulations results for $\lambda_1/\lambda_0 = 0.95$ and a balanced interaction. The receptive fields (left) and the orientation directions (right). The single cell solution is radially symmetric, but here some symmetry breaking does occur although many of the cells are still dominated by the radially symmetric component (shown by the short length of many of bars representing orientation). In order to represent the single receptive fields we have chosen an arbitrary normalized functions, one radially symmetric and one with an angular dependence of $\propto \cos(\theta)$*

figure 8 we have displayed what the simplified model produces when we make this assumption.

The receptive fields obtained in the $\cos(2\theta)$ model look far from plausible biologically, in a previous paper (Shouval and Liu, 1995) it had been shown that this can be corrected by the asymmetric portion of the correlation function, which we did not take into account here, however when we simulate the degenerate case ($\lambda_2 = \lambda_0$) as shown in figure 9 we obtain more plausible receptive fields even in the absence of the non symmetric portion of the correlation function. However, all the receptive fields obtained under the assumptions we have made in this paper have a broader tuning curve than typically found in the cortex.

We believe that the type of singularities exhibited in synaptic types of models, depend critically on the dominant eigen function of the single cell in the same parameter regime. Our belief is based on the simplifications obtained by the analysis performed in this paper, and this is but one example of the power and usefulness of such analysis. Therefore claims about synaptic models (Erwin et al., 1995) should be restricted to the parameters examined.

6 Related Models

When we do not make the common input assumption, the overlap function O takes on a different form.

The other extreme from the common input case, is the case where inputs to neighboring cells have no spatial overlap. In this case the overlap function takes the simple form
$O\left[1,1,1,1,\mathbf{x},(\mathbf{x'}),\beta(\mathbf{x}),\beta(\mathbf{x'})\right] = \delta(\mathbf{x}-\mathbf{x'})\delta_{ll'}\delta_{mm'}$ In which case, the Hamiltonian takes the form

$$H = \sum_x \sum_{ln} \tilde{I}(0)\lambda_{ln}a_{ln}^2 - \gamma(a).$$

Thus the Hamiltonian becomes a a non interacting Hamiltonian, and no effective interaction exists
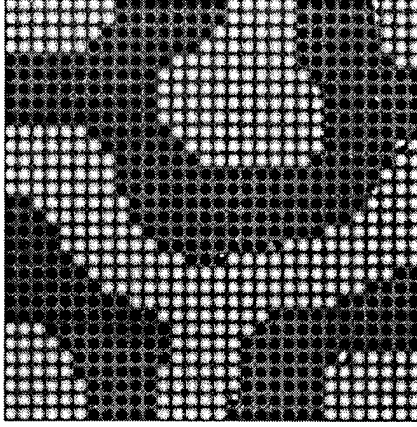
10

Figure 5: *Simulations results for $\lambda_1/\lambda_0 = 0.3$ Balanced interaction. The single cell solution is radially symmetric, and so are the network solutions. The effect of the inhibitory term in the lateral interaction has been to form staggered ordering.*

between neighboring neurons. Therefore the final state will be such that each neuron will settle in the single cell receptive fields, and the orientations will not be coupled. For this model no continuity of the orientation field should be expected. It should be noted that the outputs of the neurons $C(\mathbf{r})$ are correlated due to the interactions, just the angels of the feed forward component of the receptive fields become uncorrelated in this case.

A more complex case is the one which is examined in the Linsker and Miller models, in which $\beta(\mathbf{x}) = \mathbf{x}$, in this case the overlap function is complex, and depends on the exact details of the receptive fields, however we can obtain some understanding by looking qualitatively at the form of the overlap function. When $\beta(\mathbf{x}) = const$ an overlap exists only between same type receptive fields, when the inputs are shifted, interactions exists between different types of receptive fields as well, e.g between $\cos(\theta)$ and $\cos(2\theta)$ receptive fields. The interaction between same type receptive fields is also a function of the distance of their centers in the input plane, and in particular it can change signs as a function of $|\mathbf{x} - \mathbf{x}'|$. Some such abstract examples are shown in figure 10.

When this happens, a model which is purely excitatory can behave like a model with inhibition. To understand this lets assume for simplicity that $\lambda_{11} > \lambda_{ij}$ for every i and j, and that the cross term overlaps are very small as well. In such a case the Hamiltonian will take the form,

$$H \approx \sum_{\mathbf{x}\mathbf{x}'} \tilde{I}(\mathbf{x} - \mathbf{x}')O\left[\phi(\mathbf{x}), \phi(\mathbf{x}'), \beta(\mathbf{x}), \beta(\mathbf{x}')\right]. \tag{16}$$

Where we have omitted the $l = l' = 1$ and $n = n' = 1$ indexes for simplicity. In a purely excitatory case, $\tilde{I} > 0$, the aim will be to minimize the overlap functions O, as a function of the angles ϕ. When observing figure 10, it is obvious that for $Y = \beta(\mathbf{x}) - \beta(\mathbf{x}')$ small the two receptive fields would preferentially have the same phase angle ϕ, where as at larger distances it would be minimized when $\phi(\mathbf{x}) = \phi(\mathbf{x}') + \pi$. This observation enables us to understand that the common input approximation is sensible, If the width of the interaction term $\tilde{I}$ is smaller that the first zero crossing of the overlap term O.
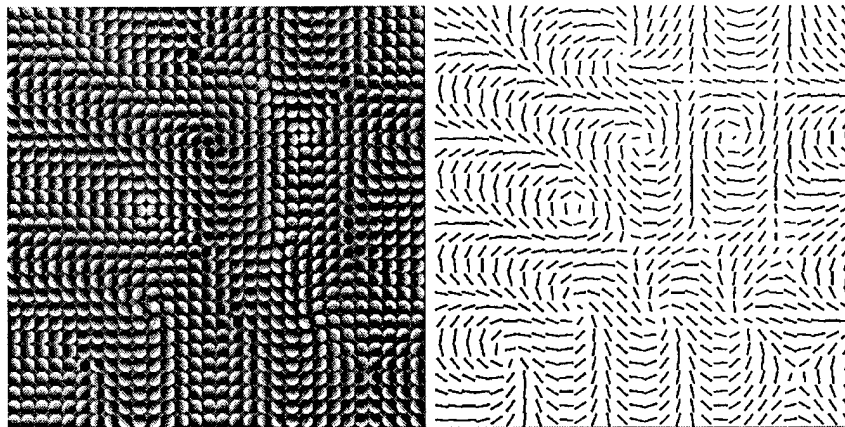
11

Figure 6: *When $\lambda_1/\lambda_0 = 1.3$ the final state shows a banded structure, and structures sometimes described in the biological literature as pinwheels, The map displayed here shows singularities of the $S = 1$ type.*

7 Discussion

We have investigated models of interacting Hebbian neurons and have shown that learning in such a network is equivalent to minimizing an energy function. We have further shown that under certain conditions this network can be approximated by continuous spin models, and is similar to some of the proposed phenomenological models (Cowan and Friedman, 1991; Swindale, 1982). In the common input case we have shown that for excitatory interactions there is no symmetry breaking, and that all receptive fields have the same orientation. We have also used this case to make the point that sticky boundary assumptions can drastically change the organization of receptive fields in the model, so that reported results are heavily dependent on when simulations are terminated.

We have also shown that inhibition effect both the structure of the receptive fields and the organization of different receptive fields across the cortex. In extreme cases we have shown that inhibition can create oriented receptive fields, where the single cell receptive fields are radially symmetric.

The global organization produced by models with inhibition, shows Iso-orientation patches, and pinwheel type of singularities. We demonstrated that the type of singularities depend critically on the on the symmetry of the receptive fields.

The major qualitative difference between the common input model, and shifted input models, is that even in the absence of inhibition, shifted input models can behave like models with inhibition. We have qualitatively demonstrated why this is the case.
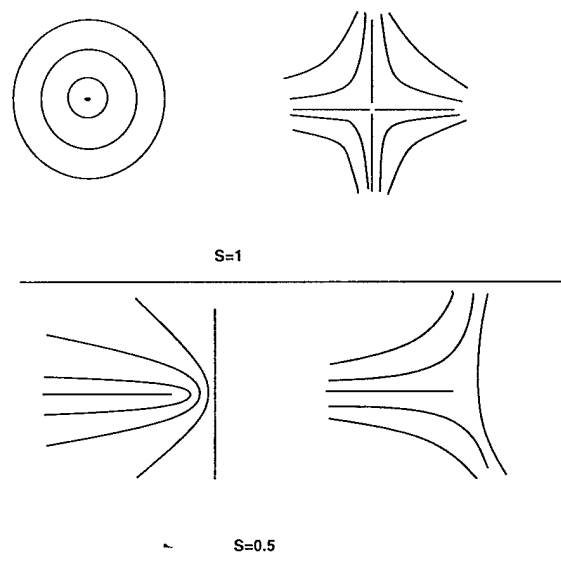
8 Acknowledgments

Figure 7: *A schematic drawing of different types of singularities , $S = 1$ are drawn above, $S = \frac{1}{2}$ below .*

References

Blasdel, G. G. (1992). Orientation selectivity,preference, and continuity in monkey striate cortex. *The Journal of Neuroscience*, 12(8):3139–3161.

Bonhoeffer, T. and Gringald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353:429–431.

Cowan, J. D. and Friedman, A. E. (1991). Simple spin models for the development of ocular dominance columns and iso-orientation patches. In *Advances in Neural Information Processing Systems 3*.

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominace in visual cortex. *Neural Computation*, 7.3:425–468.

Kosterlitz, M. J. and Thouless, D. J. (73). Ordering, metastability and phase transitions in two-dimensional systerms. *J. Phys.*, C6:1181.

Linsker, R. (1986a). From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. USA*, 83:8779–83.

Linsker, R. (1986b). From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. USA*, 83:7508–12, 8390–4, 8779–83.

Liu, Y. (1994). *Synaptic Plasticity: From Single Cell to Cortical Network*. PhD thesis, Brown University.

Liu, Y. and Shouval, H. (1994). Localized principal components of natural images - an analytic solution. *Network.*, 5.2:317–325.

MacKay, D. J. and Miller, K. D. (1990). Analysis of linsker's simulations of hebbian rules to linear networks. *Network*, 1:257–297.

Miller, K. D. (1992). Development of orientation columns via competition between on- anf off-center inputs. *NeuroReprot*, 3:73–76.
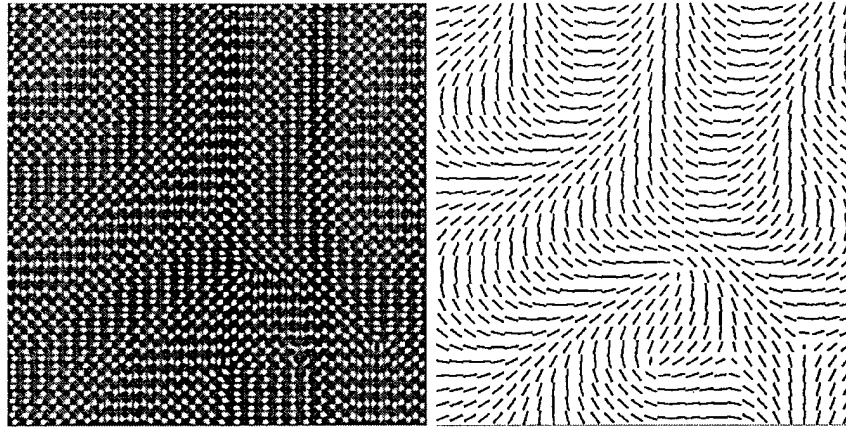
Figure 8: *When the dominant eigenfunction has a* $\cos(2\theta)$ *angular form, and respectively* $\lambda_2/\lambda_0 = 1.3$ *the final state shows* $S = \frac{1}{2}$ *type singularities.*

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *J. of Neurosci.*, 14.

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

Miller, K. D. and MacKay, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Computation*, 6:98–124.

Nass, M. N. and Cooper, L. N. (1975). A theory for the development of feature detecting cells in visual cortex. *Biol. Cyb.*, 19:1–18.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273.

Rubinstien, J. (1994). Pattern formation in associative neural networks with weak lateral interaction. *Biological Cybernetics*, 70.

Shouval, H. and Liu, Y. (1995). Principal component neurons in a realistic visual environment. Submitted.

Swindale, N. (1982). A model for the formation of orientation columns. *Proc. R. Soc. London*, B208:243.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in striata cortex. *Kybernetik*, 14:85–100.
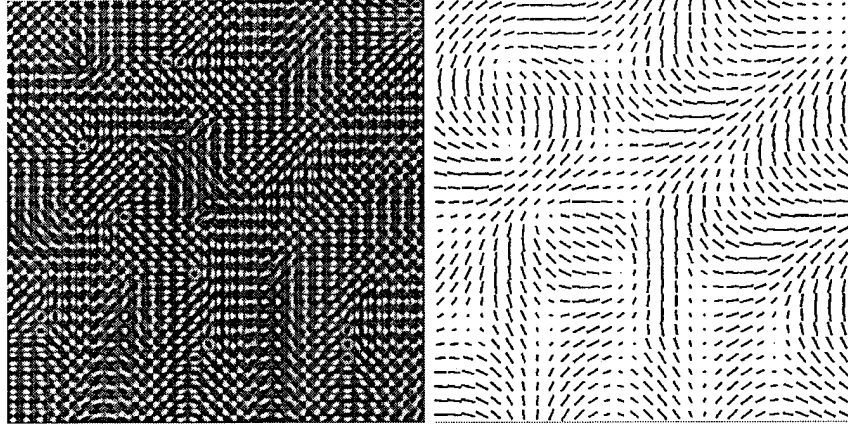
Figure 9: *In the degenerate case, when the eigen value of the radially symmetric, center surround, solution λ_0 is equal to that of the $\cos(2\theta)$ solution, the resulting receptive field structure seems biologically plausible, and the final state shows $S = \frac{1}{2}$ type singularities.*
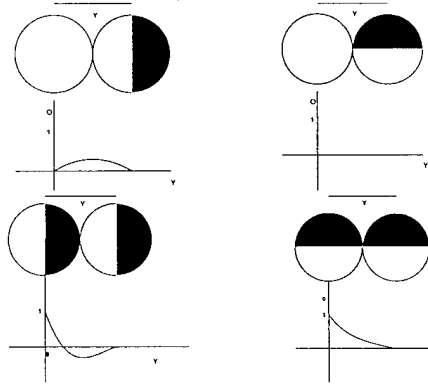

Figure 10: *A schematic drawing of some aspects of the overlap functions for shifted inputs. We have defined $Y = \beta(\mathbf{x}) - \beta(\mathbf{x}')$. Above cases of overlaps between different types of RF's are displayed, below between same kinds of RF's*