Linsker-type Hebbian Learning: A Qualitative Analysis On The Parameter Space *

JIANFENG FENG†

Mathematical Department
University of Rome "La Sapienza"
P. le A. Moro, 00185 Rome, Italy

HONG PAN

VWANI P. ROYCHOWDHURY

School of Electrical Engineering
1285 Electrical Engineering Building
Purdue University
West Lafayette, IN 47907-1285

PURDUE UNIVERSITY TECHNICAL REPORT TR-EE 95-12

*The work of V. P. Roychowdhury and H. Pan was supported in part by the General Motors Faculty Fellowship and by the NSF Grant No. ECS-9308814. J. Feng was partially supported by CNR of Italy and A. v. Humboldt Foundation of Germany.

†On leave from Department of Probability & Statistics, Peking University, Beijing 100871, P.R.China. Current address: Mathematisches Institut, Universität München, Theresienstr. 39, D-80333 München, Germany.

Contents

List of Tables

List of Figures

Abstract

We propose a novel rigorous approach for the analysis of the Linsker's unsupervised Hebbian learning network. The behavior of this model is determined by the underlying nonlinear dynamics that are parameterized by a set of parameters originating from the Hebbian rule and the arbor density of the synapses. These parameters determine the presence or absence of a specific receptive field (also referred to as a connection pattern) as a saturated fixed point attractor of the model. In this paper, we perform a qualitative analysis of the underlying nonlinear dynamics over the parameter space, determine the effects of the system parameters on the emergence of various receptive fields, and provide a rigorous criterion for the parameter regime in which the network will have the potential to develop a specially designated connection pattern. In particular, this approach analytically demonstrates, for the first time, the crucial role played by the synaptic arbor density. For example, our analytic predictions indicate that no structured connection pattern can emerge in a Linsker's network that is fully feedforward connected without localized synaptic arbor density. Our general theorems lead to a complete and precise picture of the parameter space that defines the relationships between the different sets of system parameters and the corresponding fixed point attractors, and yield a method to predict whether a given connection pattern will emerge under a given set of parameters without running a numerical simulation of the model. The theoretical results are corroborated by our examples (including center-surround and certain oriented receptive fields), and match key observations reported in Linsker's numerical simulation. The rigorous approach presented here provides a unified treatment of many diverse problems about the dynamical mechanism of a class of models that use the limiter function (also referred to as the piecewise linear sigmoidal function) as the constraint limiting the size of the weight or the state variables, and applies not only to the Linsker's network but also to other learning or retrieval models of this class.

Key Words: Unsupervised Hebbian learning, Network self-organization, Linsker's developmental model, Brain-State-in-a-Box model, Ontogenesis of primary visual system, Afferent receptive field, Synaptic arbor density, Correlations, Limiter function, Nonlinear dynamics, Qualitative analysis, Parameter space, Coexistence of attractors, Fixed point, Stability.

1 Introduction

For the purpose of understanding the self-organization mechanism of primary visual system, Linsker has proposed a multilayered unsupervised Hebbian learning network with random uncorrelated inputs and localized arborization of synapses between adjacent layers [20]. The simulation results reported in [20, 21, 22] have shown that for *appropriate* parameter regimes, several structured connection patterns (e.g., center-surround and oriented receptive fields) occur progressively as the Hebbian evolution of the weights is carried out layer by layer.

In addition to exhibiting some of the major features of early neural development in visual system, the significance of the Linsker's developmental model lies in the simplicity of the basic factors that the model is based on:

1. Hebbian learning rule for the evolution of the synaptic connection strengths. From a physics perspective, the Hebbian rule represents the mechanism for self-amplification of fluctuations in this self-organization model [27].

2. Constraints on the magnitude of the synaptic connection strengths; the imposition of these constraints introduces the nonlinearity into the model. Again, from a physics perspective [27], these constraints provide the competition mechanism among fluctuations due to the limitation of resources.

3. The localized arborization of synapses between adjacent layers. Thus, this model applies to the formation of afferent receptive fields (aRFs) [1] [27]. As observed in [20, 21, 22], and as rigorously shown in this paper, this factor plays a crucial role in the emergence of structured aRFs.

The behavior of the Linsker's model is determined by the underlying nonlinear dynamics that are parameterized by a set of parameters originating from the Hebbian rule and the arbor density of the synapses The simulation results in [20, 21, 22] show that different aRFs emerge under different sets of parameters; that is, for the same layer one can obtain different aRFs by varying the model parameters.

The purpose of this paper is to gain more insights into the dynamical mechanism of this multilayered Hebbian learning network by performing a rigorous analysis on its parameter space without any approximation. That is, our goal is to reveal the effects of the system parameters on the development of aRFs, and to characterize *precisely* within which parameter regime the network will have the potential to develop a specially designated aRF. The novel rigorous approach presented here applies not only to the Linsker-type Hebbian learning model but also to other related self-organization models about neural development (see Section 5 for further discussions).

1.1 Formulation Of The Linsker's Developmental Model

The Linsker's network is essentially a feed-forward neural network in which neurons are arranged in layers lettered from $\mathcal{A}$ to $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ etc., and synaptic strengths are modified according to the Hebbian rule until they arrive at boundary (see Fig. 1). In the Linsker's network, each cell in the present layer $\mathcal{M}$ receives synaptic inputs from a number of cells in the preceding layer $\mathcal{L}$. The density of these synaptic connections decreases monotonically with distance from the point underlying the

[1]Abbreviation: aRF, afferent receptive field.

$\mathcal{M}$-cell's position. More precisely, assume that $\xi_\tau(i)$ is the activity of the i-th neuron in layer $\mathcal{L} = \{1, \cdots, N_\mathcal{L}\}$ at time τ, where $1 \leq i \leq N_\mathcal{L}$, and $\eta_\tau(k)$ is the activity of the k-th neuron in the next layer $\mathcal{M} = \{1, \cdots, N_\mathcal{M}\}$ at time τ. Then, $\eta_\tau(k), 1 \leq k \leq N_\mathcal{M}$ is a linear weighted sum of the inputs $\xi_\tau(i)$'s:

$$\eta_\tau(k) = \sum_{i=1}^{N_\mathcal{L}} \xi_\tau(i)\omega_\tau(k,i)r(k,i) + a_1, \tag{1}$$

where $\omega_\tau(k,i)$ is the connection strength between the i-th $\mathcal{L}$-cell and the k-th $\mathcal{M}$-cell at time τ, $r(\cdot,\cdot)$ is a non-negative synaptic density function (SDF)[2] that satisfies

$$\sum_{i \in \mathcal{L}} r(k,i) = 1, \ \forall \ 1 \leq k \leq N_\mathcal{M},$$

and a_1 is a constant. Thus each $\mathcal{M}$-cell (e.g., the k-th $\mathcal{M}$-cell here) has a bundle of synaptic connections with a number of $\mathcal{L}$-cells through a *weight vector* or a *connection pattern* $\omega_\tau(k,\cdot) = \{\omega_\tau(k,i), i \in \mathcal{L}\}$, where i is numbered as shown in Fig. 1.

For the development process of $\omega_\tau(k,i)$, the Hebb-type learning rule is used:

$$\omega_{\tau+1}(k,i) = \omega_\tau(k,i) + a_2 + a_3(\eta_\tau(k) - a_4)(\xi_\tau(i) - a_5), \tag{2}$$

where $a_3 > 0$ is the step size of Hebbian learning, and a_2, a_4, and a_5 are all constants. Since the synaptic strengths change on a long time scale compared to the variation of random inputs, by averaging the Hebbian rule over the ensemble of input activities in layer $\mathcal{L}$, the dynamical equation for the development of the synaptic strength $\omega_\tau(k,i)$ between the k-th $\mathcal{M}$-cell and the i-th $\mathcal{L}$-cell at time τ is

$$\langle\omega_{\tau+1}(k,i)\rangle = \langle\omega_\tau(k,i)\rangle + k_1 + \sum_{j=1}^{N_\mathcal{L}}[Q^\mathcal{L}_{ij} + k_2]r(k,j)\langle\omega_\tau(k,j)\rangle, \tag{3}$$

where $\langle\cdot\rangle$ represents the average over the ensemble of activity patterns $\xi_\tau(i)$ in layer $\mathcal{L}$, and k_1, k_2 are system parameters that are particular combinations of the constants of the Hebbian rule:

$$\begin{aligned} k_1 &= a_2 + a_3(a_1 - a_4)[\langle\xi_\tau(i)\rangle - a_5], \\ k_2 &= a_3[\langle\xi_\tau(i)\rangle - a_5]\langle\xi_\tau(j)\rangle. \end{aligned}$$

Here $\langle\xi_\tau(i)\rangle$ is a constant [3] independent of i and τ, and the covariance matrix

$$\{Q^\mathcal{L}_{ij}\} = \{a_3\langle[\xi_\tau(i) - \langle\xi_\tau(i)\rangle][\xi_\tau(j) - \langle\xi_\tau(j)\rangle]\rangle, 1 \leq i,j \leq N_\mathcal{L}\} \tag{4}$$

of layer $\mathcal{L}$ describes the correlation of activities of the i-th and the j-th $\mathcal{L}$-cells. Actually, the covariance matrix of each layer is determined by SDFs $r(\cdot,\cdot)$'s of all preceding layers, and it is independent of τ as well.

By Perron-Frobenius theorem [37], we know that under certain conditions, $\langle\omega_\tau(k,i)\rangle$ will simply increase (or decrease) to positive (or negative) infinity or zero for the linear dynamics defined by

[2] Abbreviation: SDF, synaptic density function.

[3] In the Linsker's network, it is assumed that the input signal at the first layer ($\mathcal{A}$) is random, spontaneous, and spatially uncorrelated, i.e., $\{\xi^\mathcal{A}_\tau(i), i = 1, ..., N_\mathcal{A}\} \sim \mathcal{N}(0, \Sigma), \Sigma = \{\sigma_{ij}\}, \sigma_{ij} = 0$ for $i \neq j$ and $= 1$ for $i = j$, $i,j = 1, ..., N_\mathcal{A}$. The development process of connection strengths is run layer by layer, one layer at a time, so this assumption is readily verified.

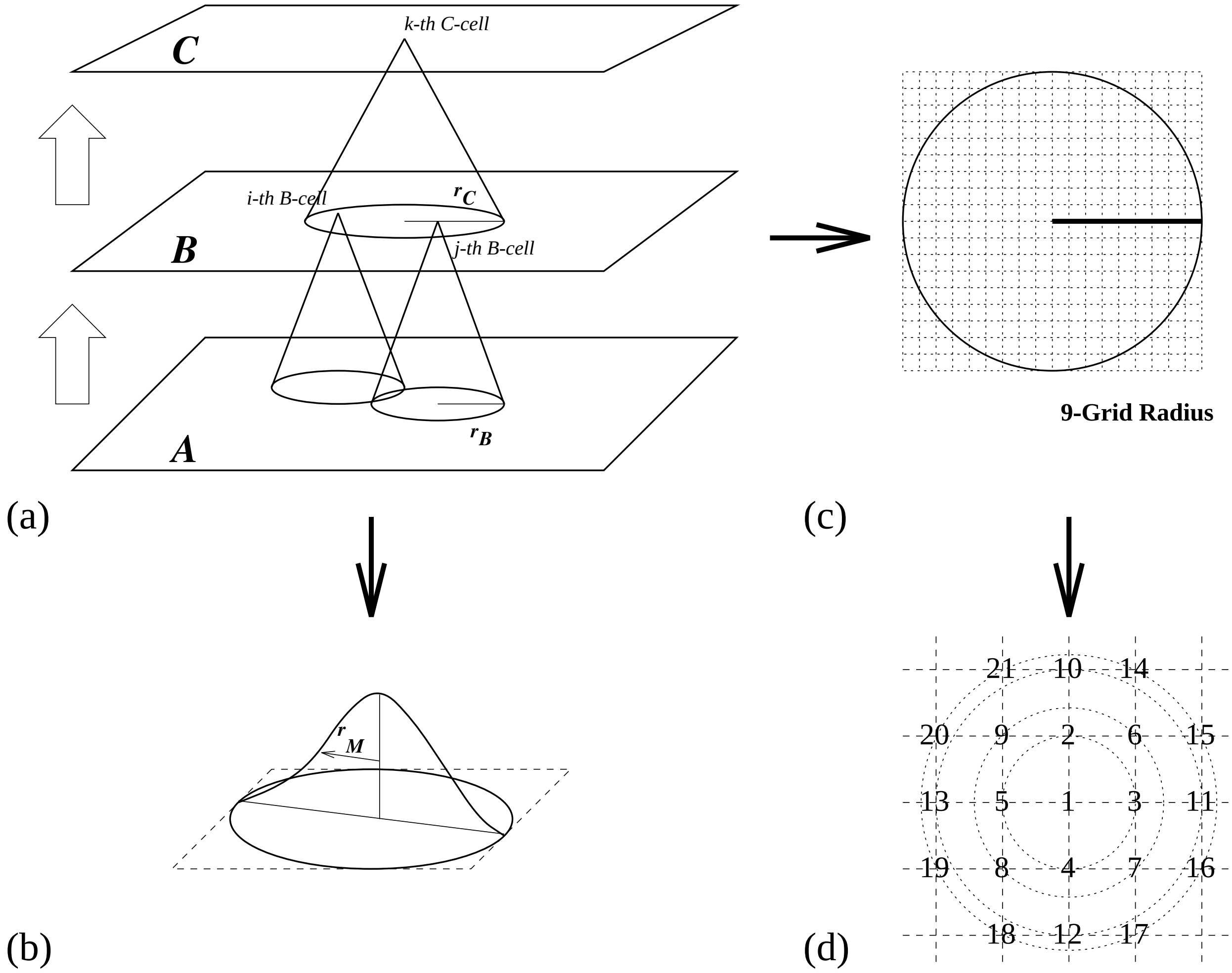


Figure 1: An illustration of the Linsker's network. **(a)** The first three layers of the Linsker's network. Each neuron in each layer receives synaptic inputs from a number of cells in the preceding layer. The density of synapses accepted by an $\mathcal{M}$-cell decreases monotonically with a range $r_{\mathcal{M}}$ from the site underlying the $\mathcal{M}$-cell's position. **(b)** The density of synapses received by an $\mathcal{M}$-cell has a range $r_{\mathcal{M}}$. For a Gaussian SDF $r^{\mathcal{LM}}(k,i) \sim \exp(-\frac{|k-i|^2}{r_{\mathcal{M}}^2})$, $i \in \mathcal{L}, k \in \mathcal{M}$, the range $r_{\mathcal{M}}$ is its standard deviation. **(c)** In our examples, each neuron is assumed to be on a grid site, and receives synaptic inputs from 253 sites within a circle with grid radius 9 in the preceding layer. **(d)** The grid sites are numbered in the order as shown. *This numbering scheme transforms a two-dimensional synaptic connection pattern into a one-dimensional weight vector.*

equation (3). One can show [11] that if equation (2) is regarded as a random perturbation of equation (3), then the underlying random variable $\omega_\tau(k,i)$ will also become unbounded. To avoid the unboundedness of $\langle\omega_\tau(k,i)\rangle$, upper and lower bounds for synaptic strengths in equation (3) are added in Linsker's simulation, i.e., $-\omega_{max} \leq \langle\omega_\tau(k,i)\rangle \leq \omega_{max}$, where $\omega_{max} > 0$. Therefore, we could rewrite the dynamical equation in (3) as follows:

$$\omega_{\tau+1}(i) = f\{\omega_\tau(i) + k_1 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij}^{\mathcal{L}} + k_2]r(j)\omega_\tau(j)\}, \tag{5}$$

where we have omitted the argument k which denotes the k-th $\mathcal{M}$-cell, and have denoted $\langle\omega_\tau(i)\rangle$ as $\omega_\tau(i)$ for the simplicity of notation. Here $f(\cdot)$ is a *limiter function* defined by

$$f(x) = \begin{cases} \omega_{max}, & \text{if } x > \omega_{max} \\ x, & \text{if } |x| \leq \omega_{max} \\ -\omega_{max}, & \text{if } x < -\omega_{max}. \end{cases}$$

Remark 1 **(a)** From the procedures of our proofs (see Theorem 1 and Corollary 1), one can verify that the results of this paper are still valid for the case where $f(\cdot)$ is defined as: $f(x) = n_{EM}$, if $x > n_{EM}$; $= x$, if $n_{EM} - 1 \leq x \leq n_{EM}$; $= n_{EM} - 1$, if $x < n_{EM} - 1$ for $0 < n_{EM} < 1$, which is the nonlinear function used in Linsker's simulations [20].

(b) The SDF is explicitly incorporated into the dynamics (5) which are equivalent to the Linsker's formulation. A rigorous explanation for this equivalence is given in Section 3.2 and Appendix A in [25]. □

1.2 Qualitative Analysis Of Nonlinear System And aRFs

In fact, the Linsker's model is a system of first-order nonlinear difference equations, taking the form

$$\omega_{\tau+1}(i) = f[\omega_\tau(i) + h_i(\omega_\tau, k_1, k_2)], \qquad i = 1, ..., N_{\mathcal{L}}, \tag{6}$$

where $h_i(\omega_\tau, k_1, k_2) = k_1 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij}^{\mathcal{L}} + k_2]r(j)\omega_\tau(j)$, and $\omega_\tau = \{\omega_\tau(j), j = 1, ..., N_{\mathcal{L}}\}$. The aRFs observed in Linsker's simulations are the saturated fixed point attractors (as defined later on) of this *parameterized* nonlinear system. As a first step, one is concerned mainly with the generic properties of the set of fixed points of the mapping (6), rather than the details of the transient process of the dynamics. Toward this end, we use the theory of *qualitative analysis*, and describe how the set of aRFs in Linsker-type networks can be completely characterized by performing a qualitative analysis of the system in (6). We next discuss several concepts related with fixed points in nonlinear systems.

The limiter function $f(\cdot)$ defines a hypercube $\Omega = [-\omega_{max}, \omega_{max}]^{N_{\mathcal{L}}}$ in the state space $\Re^{N_{\mathcal{L}}}$ and, therefore, the weight vector $\omega_\tau = \{\omega_\tau(i), i \in \mathcal{L}\} \in \Omega \subset \Re^{N_{\mathcal{L}}}$. The hypercube Ω is a closed convex set, i.e., a compact set on $\Re^{N_{\mathcal{L}}}$. Therefore, the equations (6) define a mapping from the compact set Ω into itself. On the other hand, it is well-known that a *fixed point* or an *equilibrium state* of equation (6) satisfies

$$\omega_\tau(i) = f[\omega_\tau(i) + h_i(\omega_\tau, k_1, k_2)]. \tag{7}$$

So we next show, using Brouwer's fixed point theorem [17], that the equation (7) has at least one solution.

Lemma 1 *There exists at least one solution of the equation*

$$\nu(i) = f[\nu(i) + h_i(\nu, k_1, k_2)],$$

where $\nu = \{\nu(j), j = 1, ..., N_{\mathcal{L}}\} \in \Omega$.

Proof: Note that $\Omega = [-\omega_{max}, \omega_{max}]^{N_{\mathcal{L}}} \subset \Re^{N_{\mathcal{L}}}$ is a closed convex set on $\Re^{N_{\mathcal{L}}}$. Define a mapping $F(\nu) : \Omega \to \Omega$ by

$$F(\nu) = (f(\nu(1) + h_1(\nu, k_1, k_2)),\ f(\nu(2) + h_2(\nu, k_1, k_2)),\ \cdots,\ f(\nu(N_{\mathcal{L}}) + h_{N_{\mathcal{L}}}(\nu, k_1, k_2)))$$

for $\nu \in \Omega$. From the definition of the limiter function f, we know F is a continuous mapping. Now, the proof of the lemma follows from the Brouwer's fixed point theorem [17], i.e., if F is a continuous mapping of a closed convex set Ω in itself, then there is at least one fixed point for the mapping F on Ω. $\square$

A fixed point is called a *stable fixed point* or a *fixed point attractor* of the system (6) if there exists a neighborhood of the fixed point in the state space such that if the system starts at any state in the neighborhood, then it will finally be attracted to the fixed point. A relevant concept is the *attractive basin* of a steady state. The basin of attraction of a fixed point is the set of all initial conditions in the state space that makes the trajectories of the dynamics (6) tend towards this fixed point attractor as $\tau \to \infty$. We denote the set of all extreme points or vertices of the hypercube Ω as $V(\Omega) = \{-\omega_{max}, \omega_{max}\}^{N_{\mathcal{L}}}$. A fixed point ω is called a *saturated fixed point* if $\omega \in V(\Omega)$ as well. And finally, an aRF is a *saturated fixed point attractor* of equation (6) [4].

Among the primary properties of a nonlinear system, the *coexistence* phenomenon makes it distinct from a linear system. It is well known that for a linear system, the entire state space is the attractive basin of the eigenvector with the largest eigenvalue. But as might be expected for a nonlinear dynamical system, usually, there *coexist* several attractors in equation (6) for the same set of the system parameters. That is, for a given set of the parameters, the state space comprises several attractive basins, each corresponding to an attractor (not necessarily a fixed point). The initial condition determines which attractor will be eventually reached. At the same time, a parameterized nonlinear system (e.g., the system in (6)) could have different groups of coexisting attractors for different sets of system parameters. That is, one could make the presence or absence of a specific state as a fixed point attractor by varying the set of the parameters. For a developmental model like the Linsker's network, it is expected that the different aRFs should emerge under different sets of parameters and should be relatively insensitive to the initial conditions. In other words, the dynamics should avoid the coexistence of several attractors in an appropriate way. In Linsker's work [20], this was achieved through the careful selection of parameters made possible by extensive simulations.

The task for theoretical work about this nonlinear system is to reveal its intrinsic properties in such a way that one could precisely predict the relationships between the different sets of system parameters and the corresponding fixed point attractors. In [25, 26], MacKay and Miller provided a linearization approach to this task by assuming that the principal features of the dynamics are

[4] In this paper, we mainly consider saturated fixed point attractors of equation (6), which are observed in Linsker's numerical simulations and studied theoretically in [20, 21, 22, 23, 24, 25, 26, 42]. Moreover, it is more likely that only this kind of steady state makes sense for modeling the biological development of sensory systems.

established before the weight boundaries are reached. Since within the hypercube Ω the dynamics are dominated by the linear system

$$\omega_{\tau+1}(i) = \omega_{\tau}(i) + h_i(\omega_{\tau}, k_1, k_2),$$

the short-time evolution of weight vectors can be fully characterized in terms of the properties of the eigenvectors and eigenvalues of the above linear system [25, 26]. With this approach, a number of informative results that explain Linsker's simulations (for example, the general spectrum of eigenvalues and the four principal parameter regimes) have been obtained. But as it has been already pointed out in [25, 26], this method of stability analysis is not suitable for the long-time evolution of equation (6): clearly, as the first largest component of ω_{τ} reaches saturation, the system is no longer linear. Moreover, as pointed out in [25, 26], the estimates of the division of parameter regimes obtained from a linear analysis may be in error. Therefore, we propose a totally different approach by turning back to the very beginning point of a qualitative analysis of the nonlinear system without making any approximation. It turns out to be a general framework for analyzing systems that can be described as in (6) (see Section 5.3), and provides a precise picture of the division of parameter regimes, and makes comprehensive predictions about the system behaviors.

1.3 Summary Of Our Approach

It is well known that in general it is intractable to completely characterize the behavior of a nonlinear dynamics. The difficulties lie in both determining the set of terminal attractors, as well as, in characterizing their basins of attraction in the state space. For some cases, however, it is relatively easy to derive a necessary and sufficient condition to check whether a given state is a fixed point of the dynamics. In terms of this condition, the *whole* parameter regime for the emergence of a fixed point of the dynamics may be obtained in the parameter space. If we are further able to prove the stability of the fixed point, then we can assert the occurrence of this fixed point attractor in that parameter regime. For the Linsker's network, fortunately, the above idea can be carried out because of the specific form of the nonlinear function $f(\cdot)$.

Basically, given a parameterized dynamical system, a rigorous analysis of its set of fixed points involves addressing the following issues:

1. For any given set of the parameters, what are the characteristics of the whole set of fixed points? What are the characteristics of the *stable* fixed points?

2. Given a fixed point, what is the regime in the parameter space that makes it a stable fixed point?

3. The preceding two issues imply that in general, for any given set of the parameters, several fixed points might coexist, i.e., in the parameter space, the regimes over which different fixed points are stable might overlap. Thus, a relevant question is:
   What is the complete and precise picture of the parameter space that defines the relationships among different fixed point attractors?

To address the above issues, the structure of the paper is as follows:

1. In Section 2, we present basic theorems for determining the whole set of saturated fixed point attractors of the system in (6) and prove their stability. These general theorems are suitable

for any difference equation of the type stated in (6) that uses a limiter function, and are not dependent on the structure of the second item, $h_i(\omega_\tau)$.

2. In Section 3, based on the basic results in Section 2, a necessary and sufficient condition is derived to determine whether a given aRF will emerge. We show that for any given set of SDFs (i.e., $r(\cdot)$'s), the stable parameter regime for a structured aRF is a band between two parallel lines in the parameter subspace of (k_1, k_2). These results lead to a general picture of principal parameter regimes and an important conclusion: Between any two adjacent layers in the Linsker's network, the existence of a parameter regime for the occurrence of an aRF is determined by the SDFs of all *preceding* layers and the layer under consideration, and *not* by the other two system parameters k_1, k_2. That is, the relative ranges of the SDFs determine the set of aRFs with stable regimes in the (k_1, k_2) subspace, and the choice of (k_1, k_2) determines the specific aRF (with average synaptic strength approximately equal to $-\frac{k_1}{k_2}$ (as observed in [20])) that would emerge. The conclusions of Sections 2 and 3 are general and can be applied to every layer of the Linsker's network.

3. In Section 4, we confine ourselves to synaptic connections in the first three layers of the Linsker's network. Some of the results we present include:

   (a) In Section 4.2, we first provide the precise explanations for the development of connections between layers $\mathcal{A}$ and $\mathcal{B}$, and show why it is impossible to robustly emerge any structured aRFs between layers $\mathcal{A}$ and $\mathcal{B}$ without regard to the initial condition.

   (b) In Section 4.3, we give analytic predictions about the influence of the SDF's ranges ($r_\mathcal{B}$ and $r_\mathcal{C}$) on the dynamics in the first three layers by changing $r_\mathcal{B}$ from the smallest extreme to the largest one with respect to $r_\mathcal{C}$. If the synaptic connections from layer $\mathcal{B}$ to $\mathcal{C}$ are fully feedforward connected with constant density and those from $\mathcal{A}$ to $\mathcal{B}$ are concentrated enough, then we show that any kind of connection pattern has a stable parameter regime and emerges under certain parameters. As $r_\mathcal{B}$ is changed from the smallest to the largest extreme, the development of synaptic connections between layers $\mathcal{B}$ and $\mathcal{C}$ will depend on each other stronger and stronger in the sense that most connections have the same sign as their neighbors in an aRF. So for the largest extreme of $r_\mathcal{B}$ (i.e., the weights from layer $\mathcal{A}$ to $\mathcal{B}$ are fully feedforward connected with constant density but there is no constraint on the SDF $r^{\mathcal{BC}}(\cdot)$), *no* structured aRFs (except for the all-excitatory and the all-inhibitory connection patterns) will emerge, although there exist correlation in input activities. The second extreme case indicates that the structured aRFs cannot emerge in a Linsker's network that is fully feedforward connected without localized SDFs.

   (c) In Section 4.4 and Appendix B, we apply our general theorems and work out specific examples to determine the explicit regime of the family of parameters (including k_1, k_2, and the ranges of SDFs) for the occurrence of center-surround cells and other oriented aRFs. This enables us to obtain the complete and precise picture about the distribution and relationships of principal parameter regimes for the various aRFs. For example, we show that there exist different critical values (or thresholds) of the range of SDF for various aRFs above which those aRFs cannot be observed as terminal attractors, whatever the initial conditions are. Our results match Linsker's simulations (see for example, Fig. 10), give accurate explanations of the system behaviors, and provide a method to predict whether a specific connection pattern will emerge under a given set of parameters without running a dynamical simulation of the model.

4. In Section 5, we summarize the important role played by the localized synaptic arbor density in the development of aRFs, expound on the applicability of the general framework presented here to analyzing other kinds of learning and retrieval models, and discuss the Linsker's network from a biological perspective.

5. In Appendix A, we discuss the transform relationship between the difference equation (6) we have studied and its continuous time case. It turns out that all results for fixed points of the discrete time equation (6) are true for the corresponding continuous time version.

2 General Theorems About Fixed Points And Their Stability

The key observation in this approach is fairly direct, and is based on the special form of the nonlinear function $f(\cdot)$.

1. If a connection pattern ω is a solution of the fixed point equation (7) and $h_i(\omega, k_1, k_2) \neq 0$, then we have
$$| \omega(i) + h_i(\omega, k_1, k_2) | \geq \omega_{max}.$$
So $\omega(i)$ must have the same sign as $h_i(\omega, k_1, k_2)$, i.e.,
$$\omega(i) h_i(\omega, k_1, k_2) = \omega(i)\{k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j) \omega(j)\} > 0. \tag{8}$$
This gives a condition for checking whether a given connection pattern is a saturated fixed point of the equation (6).

2. Whenever for some i, if $h_i(\omega, k_1, k_2)$ becomes zero before $\omega(i)$ reaches saturation, then $\omega(i)$ will stay invariant at this iterating step. Clearly, this observation shows that there exists a possibility of getting an unsaturated fixed point. But this is a very special case that is of no interest in this developmental model (refer to Lemma 2 and Linsker's simulations).

Next we state one of the general theorems of this paper which makes the above ideas precise. Moreover, we assert that a connection pattern satisfying the above inequality (8) is an attractor of the equation (6). Before further development, we need the following lemma.

Lemma 2 *If*

1. $\lim_{\tau \to \infty} \omega_\tau$ *exists (denote* $\omega = \lim_{\tau \to \infty} \omega_\tau$*), and*

2. for $i = 1, \cdots, N_{\mathcal{L}}$, $h_i(\omega, k_1, k_2) = k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j) \omega(j) \neq 0$,

then $\exists\ \mathcal{T}$ *such that for all* $\tau > \mathcal{T}$
$$\omega_\tau = \omega \quad \textit{and} \quad \omega \in V(\Omega).$$

Proof: Since $h_i(\omega, k_1, k_2) = k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j) \omega(j) \neq 0$, for $i = 1, \cdots, N_{\mathcal{L}}$, we observe that there exists a neighborhood $\delta_i(\omega)$ of ω such that for all $y \in \delta_i(\omega)$,
$$h_i(y, k_1, k_2) = k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j) y(j)$$
is definitely positive or negative. Denote $\delta(\omega) = \bigcap_{i=1}^{N_{\mathcal{L}}} \delta_i(\omega)$, then we have
$$a = \min_i \inf_{y \in \delta(\omega)} |h_i(y, k_1, k_2)| > 0.$$

On the other hand, the existence of the limit $\lim_{\tau\to\infty}\omega_\tau$ implies that $\exists\ \mathcal{T}'$ with the property that, as $\tau > \mathcal{T}'$

$$\omega_\tau \in \delta(\omega).$$

For a fixed i, without loss of generality, we assume that

$$h_i(y, k_1, k_2) > 0.$$

Hence, for all $\tau > \mathcal{T}' + [\frac{2\omega_{max}}{a}] + 1 \stackrel{\text{def}}{=} \mathcal{T}$, we obtain

$$\omega_\tau(i) = \lim_{\tau\to\infty}\omega_\tau(i), \qquad \text{and} \ \ \omega_\tau(i) = \omega_{max}.$$

Therefore, we assert that $\omega = \omega_\tau$ and $\omega \in \{-\omega_{max}, \omega_{max}\}^{N_\mathcal{L}} = V(\Omega)$. □

Remark 2 The set of all fixed points of the equation (6) can be divided into two subsets: (1) the set of all saturated fixed points, and (2) the set of unsaturated fixed points satisfying the following condition

$$\exists i \in \mathcal{L} \quad \text{such that} \quad h_i(\omega, k_1, k_2) = k_1 + \sum_{j=1}^{N_\mathcal{L}}[Q_{ij}^\mathcal{L} + k_2]r(j)\omega(j) = 0. \tag{9}$$

Therefore, those unsaturated fixed points in equation (6) are only some special cases, compared with the general case $h_i(\omega, k_1, k_2) \neq 0$. Moreover, notice that for the set of all saturated fixed points, the set of parameters (k_1, k_2) satisfying the above condition (9) is a union of finitely many lines, therefore it is a set with measure zero. Without loss of generality, in this paper we only consider the saturated fixed point attractors which are observed in Linsker's simulations. □

With the above lemma in hand we can claim the following basic theorem.

Theorem 1 *The whole set of saturated fixed points (sFP) of the dynamics in equation (6) is given by*

$$\Omega_{sFP} = \{\omega \mid \omega(i)h_i(\omega, k_1, k_2) > 0, \quad 1 \leq i \leq N_\mathcal{L}\}, \quad \textit{and} \ \ \omega \in V(\Omega),$$

where $h_i(\omega, k_1, k_2) = k_1 + \sum_{j=1}^{N_\mathcal{L}}[Q_{ij}^\mathcal{L} + k_2]r(j)\omega(j)$.

Proof: In terms of Lemma 2, we see that for all $\tau > \mathcal{T}$,

$$\omega_\tau(i) = f(\omega_\tau(i))$$

and $\omega_\tau(i) \in \{-\omega_{max}, \omega_{max}\}$. So because of the special form of the nonlinear function $f(\cdot)$, The whole set of saturated fixed points of the dynamics in equation (6) should satisfy

$$\Omega_{sFP} \subseteq \{\omega \mid \omega(i)h_i(\omega, k_1, k_2) > 0, \quad 1 \leq i \leq N_\mathcal{L}\}.$$

The proof of the relation

$$\{\omega | \omega(i)h_i(\omega, k_1, k_2) > 0, \quad 1 \leq i \leq N_\mathcal{L}\} \ \subseteq \Omega_{sFP}$$

is trivial. Then the characteristics of the whole set of saturated fixed points of equation (6) is

$$\Omega_{sFP} = \{\omega \mid \omega(i)h_i(\omega, k_1, k_2) > 0, \quad 1 \le i \le N_{\mathcal{L}}\}.$$

□

Without loss of generality, we assume from now on that $\omega_{max} = 1$, i.e., we consider the equation (6) with the weight state space $\Omega = [-1, 1]^{N_{\mathcal{L}}}$ and the set of vertices $V(\Omega) = \{-1, 1\}^{N_{\mathcal{L}}}$.

For an $\omega \in \Omega_{sFP}$, let $B(\omega)$ be the attractive basin of ω defined by

$$B(\omega) = \{\zeta \mid \zeta \in \Omega = [-1, 1]^{N_{\mathcal{L}}}, \omega_{\tau=0} = \zeta, \text{ and } \lim_{\tau \to \infty} \omega_\tau = \omega\}.$$

Hence, if ω_τ starts at a state ζ in $B(\omega)$ when $\tau = 0$, then ω_τ will be attracted to ω. It is natural to ask whether every $\omega \in \Omega_{sFP}$ is stable, in other words, whether there exists a neighborhood $o(\omega)$ (an open ball with center at ω) of $\omega \in \Omega_{sFP}$ such that $o(\omega) \subset B(\omega)$.

Theorem 2 *$\forall \omega \in \Omega_{sFP}$, there exists a nonempty neighborhood $o(\omega)$ of ω satisfying $o(\omega) \subset B(\omega)$.*

Proof: In fact, combining Lemma 2 and Theorem 1, we assert that if we take $o(\omega) = \delta(\omega)$, defined in the procedure of the proof of Lemma 2 , then the result follows. □

Theorem 2 claims that as long as the dynamical system (6) starts at a state in the vicinity of an $\omega \in \Omega_{sFP}$, then it will finally stop at ω within finite time. The characteristics of the whole set of saturated fixed point attractors (denoted as Ω_{sFPA}) are exactly the same as that of Ω_{sFP} stated in Theorem 1, i.e.,

$$\Omega_{sFPA} = \Omega_{sFP} = \{\omega \mid \omega(i)h_i(\omega, k_1, k_2) > 0, \quad 1 \le i \le N_{\mathcal{L}}\}, \quad \text{and} \quad \omega \in V(\Omega), \tag{10}$$

where $h_i(\omega, k_1, k_2) = k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j)\omega(j)$.

The upper and lower bounds in the definition of the limiter function need not be symmetric, and the limiter function is often defined as in (11) and (13). The following corollary shows that Theorems 1 and 2 can be easily modified for such definitions of the limiter function.

Corollary 1 *If*

$$f(x) = \begin{cases} \omega_{max}, & \text{if } x > \omega_{max} \\ x, & \text{if } 0 \le x \le \omega_{max} \\ 0, & \text{if } x < 0\ , \end{cases} \tag{11}$$

for $\omega_{max} > 0$, then

$$\Omega_{sFPA} = \{\omega \mid h_i(\omega) > 0 \ \forall \omega(i) = \omega_{max},\ h_i(\omega) < 0 \ \forall \omega(i) = 0, \quad 1 \le i \le N\}\ , \tag{12}$$

and $\omega \in \{0, \omega_{max}\}^N$.

If

$$f(x) = \begin{cases} 0, & \text{if } x > 0 \\ x, & \text{if } \omega_{min} \le x \le 0 \\ \omega_{min}, & \text{if } x < \omega_{min}\ , \end{cases} \tag{13}$$

for $\omega_{min} < 0$, then

$$\Omega_{sFPA} = \{\omega \mid h_i(\omega) > 0 \ \forall \omega(i) = 0,\ h_i(\omega) < 0 \ \forall \omega(i) = \omega_{min}, \quad 1 \le i \le N\}\ , \tag{14}$$

and $\omega \in \{\omega_{min}, 0\}^N$.

3 The Criterion For The Division Of Parameter Regimes For The Occurrence Of Attractors

We next derive an explicit necessary and sufficient condition for the emergence of various aRFs, i.e., we derive conditions to determine whether a given ω belongs to Ω_{sFPA}. Using this condition, we can obtain, without referring to the equation (6) itself, the *whole* parameter regime of (k_1, k_2) in which a given connection pattern will be an attractor of the equation (6). Our examples in Section 4.4 and Appendix B fully explain this point.

3.1 The Necessary And Sufficient Condition For The Emergence Of aRFs

Definition 1 For any $\omega \in V(\Omega)$, define $J^+(\omega) = \{i \mid \omega(i) = 1\}$ as the index set of cells at layer $\mathcal{L}$ with excitatory weight for a connection pattern ω between layers $\mathcal{M}$ and $\mathcal{L}$, and $J^-(\omega) = \{i \mid \omega(i) = -1\}$ as the index set of $\mathcal{L}$-cells with inhibitory weight for ω. (see Fig. 11) $\square$

Note from the property of saturated fixed point attractors (see Theorem 1) that a connection pattern ω is an attractor of the equation (6) if and only if we have for all $i = 1, ..., N_{\mathcal{L}}$,

$$\begin{aligned} &\omega(i)\{k_1 + \sum_j [Q^{\mathcal{L}}_{ij} + k_2] r(j)\omega(j)\} = \\ &\omega(i)\{k_1 + \sum_{j\in J^+(\omega)} [Q^{\mathcal{L}}_{ij} + k_2] r(j)\omega(j) + \sum_{j\in J^-(\omega)} [Q^{\mathcal{L}}_{ij} + k_2] r(j)\omega(j)\} > 0. \end{aligned} \tag{15}$$

For all $i \in J^+(\omega)$, the above inequality reduces to (by the definition of $J^+(\omega)$ and $J^-(\omega)$)

$$k_1 + \sum_{j\in J^+(\omega)} [Q^{\mathcal{L}}_{ij} + k_2] r(j) - \sum_{j\in J^-(\omega)} [Q^{\mathcal{L}}_{ij} + k_2] r(j) > 0. \tag{16}$$

Equivalently,

$$k_1 + k_2[\sum_{j\in J^+(\omega)} r(j) - \sum_{j\in J^-(\omega)} r(j)] > \sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j) - \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j). \tag{17}$$

The above inequality is satisfied for all i in $J^+(\omega)$, and the left hand side is independent of i. Taking maximum over the set $i \in J^+(\omega)$ on both sides of this inequality, we thus obtain

$$k_1 + k_2[\sum_{j\in J^+(\omega)} r(j) - \sum_{j\in J^-(\omega)} r(j)] > \max_{i\in J^+(\omega)} [\sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j) - \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j)]. \tag{18}$$

On the other hand, for $i \in J^-(\omega)$, we can similarly deduce that

$$k_1 + k_2[\sum_{j\in J^+(\omega)} r(j) - \sum_{j\in J^-(\omega)} r(j)] < \min_{i\in J^-(\omega)} [\sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j) - \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j)]. \tag{19}$$

Definition 2 We define the *slope function*:

$$c(\omega) \stackrel{\text{def}}{=} \sum_{j\in J^+(\omega)} r(j) - \sum_{j\in J^-(\omega)} r(j)$$

which is the difference of sums of the SDF $r(\cdot)$ over $J^+(\omega)$ and $J^-(\omega)$, and is also the *average synaptic strength* of the connection pattern ω ;
and two k_1-*intercept functions*:

$$d_1(\omega) \stackrel{\text{def}}{=} \begin{cases} \max_{i\in J^+(\omega)} (\sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j) - \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j)), & \text{if } J^+(\omega) \neq \emptyset \\ -\infty, & \text{if } J^+(\omega) = \emptyset \end{cases}$$

and

$$d_2(\omega) \stackrel{\text{def}}{=} \begin{cases} \min_{i\in J^-(\omega)} (\sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j) - \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j)), & \text{if } J^-(\omega) \neq \emptyset \\ \infty, & \text{if } J^-(\omega) = \emptyset. \end{cases}$$

□

The definition of the slope function $c(\omega)$ implies that it only depends on the SDF $r(\cdot)$ between the two successive layers under consideration and does not relate to SDFs $r(\cdot)$'s of the other preceding layers. Two k_1-intercept functions $d_1(\omega)$ and $d_2(\omega)$ embody the dependence of equation (6) on the covariance matrix $Q^{\mathcal{L}}$ of the preceding layer, and the SDF $r(\cdot)$ of the layer under consideration (i.e., layer $\mathcal{M}$). Therefore, these two k_1-intercept functions are determined by the SDFs of all preceding layers and the present layer. The reason to call them as the slope function and the k_1-intercept functions is made clear in the following theorem, which gives the new rigorous criterion for the division of stable parameter regimes to ensure the emergence of various aRFs.

Theorem 3 *For every layer of the Linsker's network, a connection pattern ω is a saturated fixed point attractor of equation (6) if and only if*

$$d_2(\omega) > k_1 + c(\omega)k_2 > d_1(\omega). \tag{20}$$

Hence, if $d_2(\omega) < d_1(\omega)$, then the aRF ω is not an attractor of equation (6).

Proof: As derived in equations $(15) - (19)$, a connection pattern ω is an attractor of the equation (6) iff for $i \in J^+(\omega)$:

$$k_1 + k_2 \sum_{j\in J^+(\omega)} r(j) + \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j) > k_2 \sum_{j\in J^-(\omega)} r(j) + \sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j),$$

and for $i \in J^-(\omega)$

$$k_1 + k_2 \sum_{j\in J^+(\omega)} r(j) + \sum_{j\in J^+(\omega)} Q^{\mathcal{L}}_{ij} r(j) < k_2 \sum_{j\in J^-(\omega)} r(j) + \sum_{j\in J^-(\omega)} Q^{\mathcal{L}}_{ij} r(j).$$

Now, the result of the theorem follows directly from the definitions of $c(\omega)$, $d_1(\omega)$, and $d_2(\omega)$. □

The above theorem shows that for any given set of SDFs, the parameter regime of (k_1, k_2) to ensure that a structured aRF ω is an attractor of equation (6) is a band between two parallel lines $k_1 + c(\omega)k_2 > d_1(\omega)$ and $k_1 + c(\omega)k_2 < d_2(\omega)$ (see Regime E and Regime F in Fig. 2). Note that if $d_1(\omega) > d_2(\omega)$, then there is no regime of (k_1, k_2) for the occurrence of that aRF ω as an attractor of equation (6). Therefore, between any two given layers in the Linsker's network, the existence of a structured aRF ω as an attractor of equation (6) is determined by k_1-intercept functions $d_1(\cdot)$ and $d_2(\cdot)$, and therefore by the SDFs $r(\cdot)$'s of all preceding layers and the present layer.

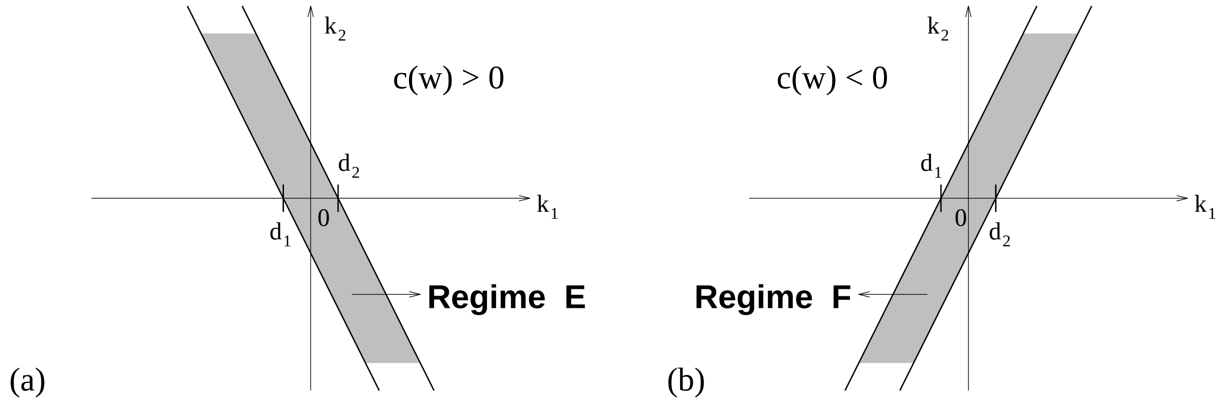


Figure 2: The parameter regime of (k_1, k_2), in which ω is a saturated fixed point attractor of equation (6). The regimes with grey texture is determined by $d_1(\omega) < k_1 + c(\omega)k_2 < d_2(\omega)$. **(a)** The case (Regime E) with a positive slope function (i.e., with a positive average synaptic strength). For example, the ON-center aRF with a large central excitatory region or the OFF-center aRF with a small inhibitory central region. **(b)** The case (Regime F) with a negative slope function (i.e., with a negative average synaptic strength). For example, the OFF-center aRF with a large central inhibitory region or the ON-center aRF with a small excitatory central region.

Remark 3 Since $r(\cdot)$ is normalized, the width of the band $[d_2(\omega) - d_1(\omega)]$ is narrow usually for an individual connection pattern ω. Hence, we have the following approximate relationship between two main system parameters k_1, k_2 and the *average synaptic strength* g of a stable connection pattern

$$\frac{-k_1}{k_2} \approx c(\omega) \stackrel{\text{def}}{=} g,$$

a relationship observed empirically by Linsker's simulations. Therefore, our analysis provides an explanation of why $\frac{-k_1}{k_2}$ is a good approximation of the mature g value. Moreover, it also shows that the difference between g and $\frac{-k_1}{k_2}$ originates from the SDFs $r(\cdot)$'s. □

Unlike for any other aRF, there always exists a stable parameter regime for the all-excitatory and the all-inhibitory connection patterns. We denote the k_1-intercept function $d_1(\omega)$ for the all-excitatory aRF as $d_1(+)$ and $d_2(\omega)$ for the all-inhibitory aRF as $d_2(-)$ respectively. From the above theorem and the definition of k_1-intercept functions, the all-excitatory aRF is an attractor of equation (6) when $k_1 + k_2 > d_1(+)$, and so is the all-inhibitory aRF when $k_1 - k_2 < d_2(-)$ (see Regime A and Regime B in Fig. 3). Thus the parameter plane of (k_1, k_2) is divided into four regimes by these two criteria, in which the regime determined by $d_1(+) < k_1 + k_2$ and $k_1 - k_2 < d_2(-)$ is the coexistence regime of all-excitatory and all-inhibitory connection patterns (see Regime C in Fig. 3). Although the absolute values of $d_1(+)$ and $d_2(-)$ may change from layer to layer, the division of the parameter regimes shown in Fig. 3 remains invariant. As an example of evaluating $d_1(+)$ and $d_2(-)$, we show the case for the first three layers in Section 4.

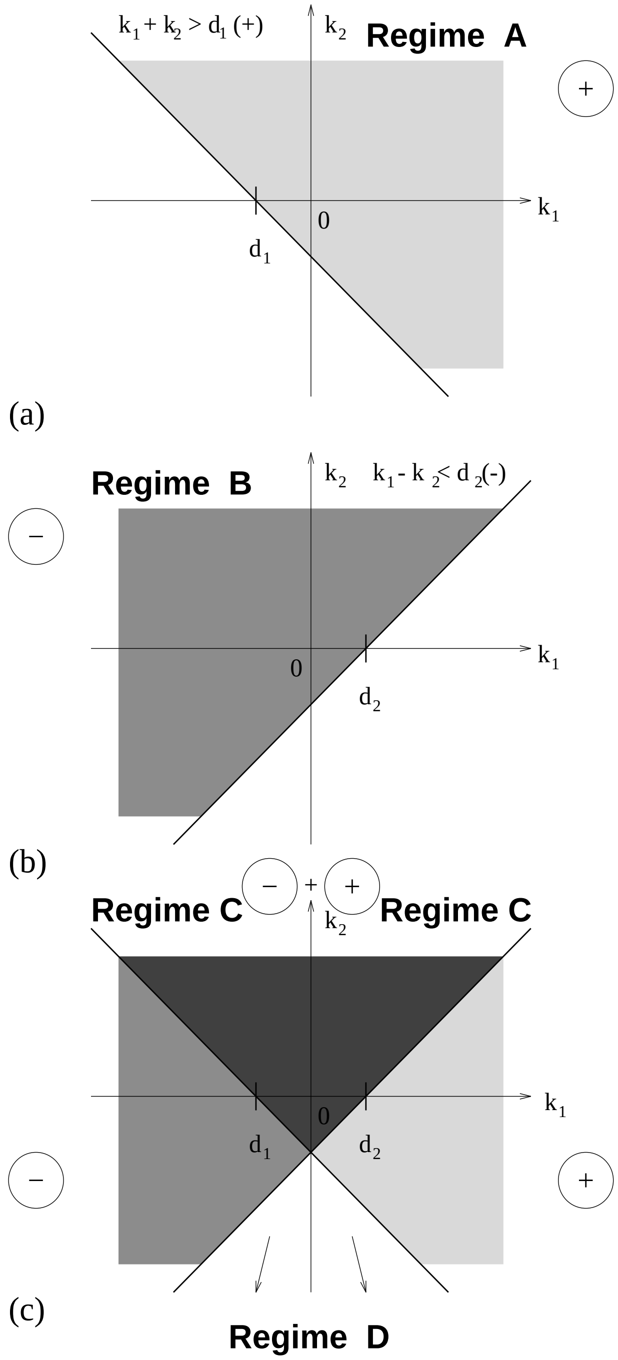


Figure 3: Parameter regimes in the (k_1, k_2) subspace for the emergence of all-excitatory (light grey texture regime A) and all-inhibitory (middle grey regime B) connection patterns. The dark grey regime (Regime C) is the coexistence regime for both all-excitatory and all-inhibitory connection patterns. The fourth regime without any texture (i.e., Regime D) is where neither the all-excitatory nor the all-inhibitory connection patterns are stable. Linsker's simulation results on the emergence of structured aRFs are obtained in Regime D.

3.2 The General Principal Parameter Regimes

TYPE	PARAMETER REGIME	ATTRACTOR
Regime A (Fig. 3(a))	$k_1 + k_2 > d_1(+)$ (approximately $\frac{-k_1}{k_2} < 1$)	All-excitatory connection pattern
Regime B (Fig. 3(b))	$k_1 - k_2 < d_2(-)$ (approximately $\frac{-k_1}{k_2} > -1$)	All-inhibitory connection pattern
Regime C=A∩B (Fig. 3(c))	$k_1 + k_2 > d_1(+)$ and $k_1 - k_2 < d_2(-)$	All-excitatory and all-inhibitory aRFs coexist
Regime D=$(\mathrm{A} \cup \mathrm{B})^c$ (Fig. 3(c))	$k_1 + k_2 < d_1(+)$ and $k_1 - k_2 > d_2(-)$	The structured aRFs may have separate parameter regimes
Regime E (Fig. 2(a))	$d_2(\omega) > k_1 + c(\omega)k_2 > d_1(\omega)$ where $c(\omega) > 0$	Any aRF in which the excitatory connections constitute the majority (e.g., the ON-center cell with large excitatory center radius r_{core} or the OFF-center cell with small inhibitory r_{core})
Regime F (Fig. 2(b))	$d_2(\omega) > k_1 + c(\omega)k_2 > d_1(\omega)$ where $c(\omega) < 0$	Any aRF in which the inhibitory connections constitute the majority (e.g., the ON-center cell with small excitatory r_{core} or the OFF-center cell with large inhibitory r_{core})
Regime G = E∩F∩A∩B (Fig. 4)	$d_2(\omega^1) > k_1 + c(\omega^1)k_2 > d_1(\omega^1)$ ...... $d_2(\omega^\mu) > k_1 + c(\omega^\mu)k_2 > d_1(\omega^\mu)$	A small coexistence regime of many aRFs around the origin point of the parameter plane of (k_1, k_2)
Regime H = E or F with $c(\omega) \approx 0$ (Fig. 4)	$d_2(\omega) > k_1 + c(\omega)k_2 > d_1(\omega)$ where $c(\omega) \approx 0$	many aRFs with $c(\omega) \approx 0$ coexist

Table 1: The General Principal Parameter Regimes

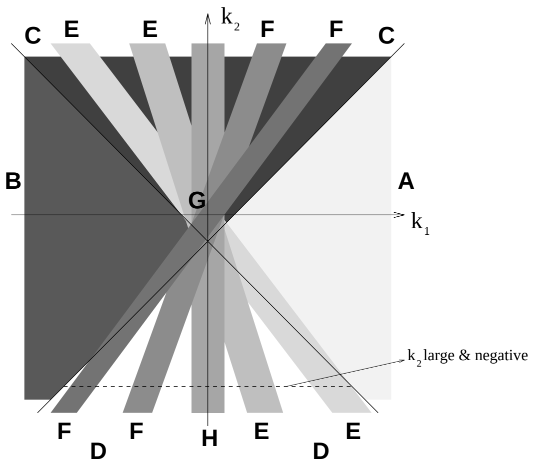


Figure 4: The general principal parameter regimes for every layer in the Linsker's network (also see Table 1). Notice that there exists a coexistence regime (Regime G) of many connection patterns around the origin point of the parameter plane of (k_1, k_2). Regime H covering the axis k_2 is slightly complicated as well, since any kind of structured aRF with average synaptic strength near zero may have a stable parameter regime in the sense as shown in Appendix B. As an example here, the bands correspond to ON-center aRFs with different r_{core} when $r_{\mathcal{C}} = 10$ and $r_{\mathcal{B}} = 1$. For other kinds of aRFs, the picture of principal parameter regimes is similar (refer to Theorem 4, i.e., if the SDF from $\mathcal{B}$ to $\mathcal{C}$ is flat but the SDF from $\mathcal{A}$ to $\mathcal{B}$ is concentrated enough, then any kind of aRF has a stable parameter regime).

In general, one would like to ask the following questions concerning the relationships between the different parameter regimes and the corresponding aRFs:

1. What is the complete and precise picture of the parameter regimes for various aRFs?
2. What does the relative stability of various aRFs mean in the context of critical values of parameters?
3. Based on the answers to the above two issues, what is the appropriate way to ensure the emergence of a specially designated aRF?

To address the above questions, we summerize the principal parameter regimes from our general theorems applicable to all layers in Table 1 and Figure 4, and make the following observations:

1. Regime A, B, and C (i.e., when $k_1 + k_2 > d_1(+)$ and/or $k_1 - k_2 < d_2(-)$) are dominated by the all-excitatory and the all-inhibitory aRFs. There exists a small regime (i.e., Regime G in Fig. 4) around the origin point in the parameter subspace of (k_1, k_2), in which many attractors coexist. Therefore, it is difficult to develop a structured aRF in Regimes A, B, C, and G, even though there exist the possibility of emerging it in Regime C (i.e., in $\mathrm{C} \cap \mathrm{E}$ or $\mathrm{C} \cap \mathrm{F}$) if one carefully chooses the initial condition. However, in a developmental model the evolution process of weights should be relatively insensitive to the initial conditions. So the only parameter regime in which the structured aRFs may have separate parameter regimes is Regime D (see Fig. 3(c) and Fig. 4), i.e., when $k_1 + k_2 < d_1(+)$ and $k_1 - k_2 > d_2(-)$.
2. Within Regime D, if one further chooses k_2 to be large and negative (e.g., $k_2 = -3$ as taken in Linsker's simulations), then various Regime E's and Regime F's corresponding to different aRFs will be removed from Regime A, B, C, and G, where the all-excitatory and all-inhibitory aRFs are dominant or many kinds of attractors coexist. Moreover, once k_2 is chosen to be large and negative, the Regime E's and Regime F's with different slope values will be separate from each other in Regime D (see Fig. 4). Thus, in Regime D one has the potential for emerging a designated aRF if the corresponding parameter regime (E or F) no longer coexists with others.
3. A number of different aRFs, however, may have the same average synaptic strength (i.e., the same slope value $c(\omega)$), and hence, their stable regimes (type E or F) may still completely overlap with each other in the (k_1, k_2) subspace. That is, if there exist two aRFs, ω^1 and ω^2, such that $c(\omega^1) = c(\omega^2)$ and they both have stable parameter regimes (i.e., $d_2(\omega^1) > d_1(\omega^1)$ and $d_2(\omega^2) > d_1(\omega^2)$), then their stable regimes might overlap, and irrespective of the choice of k_1 and k_2, they both might coexist (see Fig. 5). The only way to avoid such coexistence, is to ensure that only one of these aRFs will have a stable regime in the (k_1, k_2) subspace. According to Theorem 3, the existence of stable parameter regime for any aRF ω, is determined by k_1-intercept functions $d_1(\cdot)$ and $d_2(\cdot)$, and therefore by the relative ranges of SDFs $r(\cdot)$'s of all layers under consideration. For instance, by changing the relative ranges of SDFs, one can reach a parameter regime in the parameter space (k_1, k_2, the ranges of SDFs) in which $d_2(\omega^1) > d_1(\omega^1)$ and $d_2(\omega^2) < d_1(\omega^2)$, i.e., the connection pattern ω^1 is still an attractor but ω^2 becomes unstable (see Fig. 5). That is, one can further make the presence or absence of a specific aRF by adjusting the relative ranges of SDFs, and finally make the corresponding Regime E or F (with a certain slope value) become unique in Regime D with

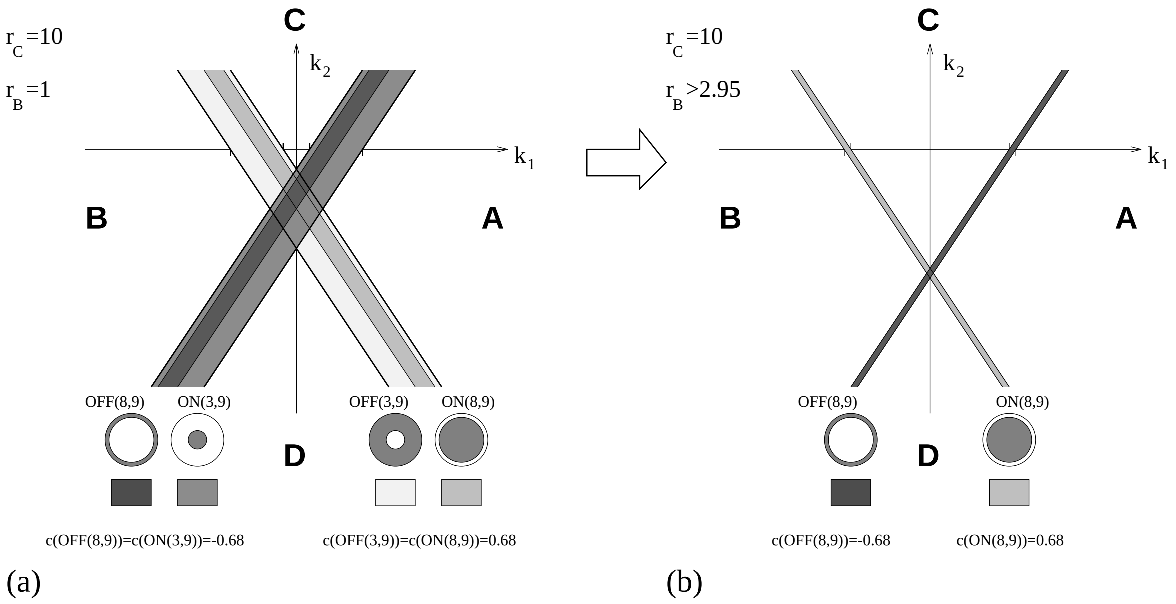


Figure 5: **(a)** An example that shows the completely overlaping stable regimes for the different kinds of aRFs with the same slope values. **(b)** Only way to avoid such coexistence, is to adjust the relative ranges of SDFs, and finally make Regime E or F for one of these aRFs (with a certain slope value) become unique in Regime D with large and negative k_2. (As an example here, we use the data from our Examples 1 and 2 in Appendix B.)

large and negative k_2. Our general theorems, thus, provide a methodology to avoid the coexistence of many potential aRFs and to ensure the emergence of a specially designated aRF; moreover, as it will be shown later, the parameter choices predicted by our theory match those reported in Linsker's simulations.

In order to further address the above questions we raised, we mainly focus on the weight evolution in the first three layers of the Linsker's network in the next section. We shall apply our general theorems to analytically predict the influence of the SDF's ranges $r_{\mathcal{B}}, r_{\mathcal{C}}$ on the dynamics, and work out specific examples to determine the explicit regime of the family of parameters (including k_1, k_2, and the ranges of SDFs) for the occurrence of center-surround and certain oriented aRFs between layers $\mathcal{B}$ and $\mathcal{C}$.

4 The aRFs In The First Three Layers

4.1 Description Of The First Three Layers Of The Linsker's Network

A given SDF $r^{\mathcal{LM}}(i,l), i \in \mathcal{M}, l \in \mathcal{L}$ will be said to have a *range* $r_{\mathcal{M}}$ if $r^{\mathcal{LM}}(i,l)$ is 'sufficient small' for $||i-l|| \geq r_{\mathcal{M}}$. For instance, if we assume the SDF is Gaussian, i.e., $r^{\mathcal{LM}}(i,l) \sim \exp(-\frac{|i-l|^2}{r_{\mathcal{M}}^2})$, $i \in \mathcal{M}, l \in \mathcal{L}$, then the standard deviation $r_{\mathcal{M}}$ of this distribution is its range (see Fig. 1). Denote the SDF from layer $\mathcal{A}$ to $\mathcal{B}$ and the SDF from $\mathcal{B}$ to $\mathcal{C}$ as $r^{\mathcal{AB}}(\cdot,\cdot)$ with range $r_{\mathcal{B}}$ and $r^{\mathcal{BC}}(\cdot)$ with range $r_{\mathcal{C}}$ respectively. A slight restriction on the SDF is imposed for our consideration. $r^{\mathcal{LM}}(i,l)$ is a function dependent only on the distance between i and l, expressed as $||i-l||$, and is nonincreasing with respect to $||i-l||$. Obviously, the Gaussian function, which is the SDF used in [20, 21, 22, 25, 26], is a special case of the above general restriction.

As in Linsker's simulations, we assume that the random input at first layer $\mathcal{A}$ has an independent normal distribution with mean 0 and variance 1 like in [20]. That is, $Q^{\mathcal{A}}_{ij} = 1$ if $i = j$, and $Q^{\mathcal{A}}_{ij} = 0$ if $i \neq j$. Hence, for the weight development between layers $\mathcal{A}$ and $\mathcal{B}$, the equation (6) for a $\mathcal{B}$-cell is:

$$\omega^{\mathcal{AB}}_{\tau+1}(i) = f(\omega^{\mathcal{AB}}_{\tau}(i) + k_1 + r^{\mathcal{AB}}(i)\omega^{\mathcal{AB}}_{\tau}(i) + k_2 c(\omega)) \ , \tag{21}$$

where $i \in \mathcal{A}$ [5].

Under the condition that the saturated connection strengths from layer $\mathcal{A}$ to $\mathcal{B}$ are all-excitatory like in [20], the covariance matrix of layer $\mathcal{B}$ is

$$Q^{\mathcal{B}}_{ij} = \sum_{l=1}^{N_{\mathcal{A}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l)$$

where $i, j \in \mathcal{B}$. Therefore, for the development process of aRF of a $\mathcal{C}$-cell, the equation (6) in Section 1.1 becomes

$$\omega^{\mathcal{BC}}_{\tau+1}(i) = f(\omega^{\mathcal{BC}}_{\tau}(i) + k_1 + \sum_{j=1}^{N_{\mathcal{B}}} [Q^{\mathcal{B}}_{ij} + k_2] r^{\mathcal{BC}}(j)\omega^{\mathcal{BC}}_{\tau}(j)). \tag{22}$$

4.2 Development Of Connections Between Layers $\mathcal{A}$ And $\mathcal{B}$

Applying Theorem 3, one can verify that for (21): (1) every structured aRF ω has a corresponding stable parameter regime that satisfies:

$$d_2(\omega) = \min_{i \in J^-(\omega)} r^{\mathcal{AB}}(i) > k_1 + c(\omega)k_2 > -\min_{i \in J^+(\omega)} r^{\mathcal{AB}}(i) = d_1(\omega) \ ; \tag{23}$$

(2) the stable parameter regime for the all-excitatory pattern satisfies: $k_1 + k_2 > -\min_{i \in \mathcal{A}} r^{\mathcal{AB}}(i) \geq -1/N_{\mathcal{A}}$; and (3) the stable parameter regime for the all-inhibitory pattern satisfies: $k_1 - k_2 < \min_{i \in \mathcal{A}} r^{\mathcal{AB}}(i) \leq 1/N_{\mathcal{A}}$. If the SDF, $r^{\mathcal{AB}}(\cdot)$, is positive (e.g., Gaussian with range $r_{\mathcal{B}}$ like in [20]), then $d_2(\omega) > d_1(\omega)$, and hence, there always exists a stable parameter regime for every structured aRF ω. Therefore, the existence of a stable parameter regime for any aRF is *independent of the third parameter* $r_{\mathcal{B}}$. These analytic predictions for the first two layers exactly match the four parameter regimes reported in [20] as described in the following:

[5] From now on we let $a_3 = 1$ without loss of generality.

1. In Linsker's simulations, all mature $\mathcal{B}$-cells are *all-excitatory* when $k_2 < 0$ and $-\dfrac{k_1}{k_2} > 1$, or $k_2 > 0$ and $-\dfrac{k_1}{k_2} < -1$ (i.e., when $k_1 + k_2 > 0$ and $k_1 - k_2 > 0$). Our results show that if the system (6) is in Regime $\mathrm{A} \cap \mathrm{B}^c$(see Fig. 4 and Table 1), i.e., when $k_1 + k_2 > -1/N_{\mathcal{A}}$ and $k_1 - k_2 > 1/N_{\mathcal{A}}$, then the all-excitatory aRF is the only attractor. Since $N_{\mathcal{A}} \gg 1$, Regime $\mathrm{A} \cap \mathrm{B}^c$ approximately comprises the region determined by $k_1 + k_2 > 0$ and $k_1 - k_2 > 0$, and thus matches the regime observed in Linsker's simulations.

2. Symmetrically, in Linsker's simulations, all mature $\mathcal{B}$-cells are *all-inhibitory* when $k_2 < 0$ and $-\dfrac{k_1}{k_2} < -1$, or $k_2 > 0$ and $-\dfrac{k_1}{k_2} > 1$ (i.e., when $k_1 - k_2 < 0$ and $k_1 + k_2 < 0$). Our results show that if the system (6) is in Regime $\mathrm{B} \cap \mathrm{A}^c$, i.e., when $k_1 - k_2 < 1/N_{\mathcal{A}}$ and $k_1 + k_2 < -1/N_{\mathcal{A}}$, then the all-excitatory aRF is the only attractor.

3. In Linsker's simulations, when $k_2 > 0$ and $-1 < -\dfrac{k_1}{k_2} < 1$, and if the initial value of average synaptic strength $g = c(\omega)$ for a given $\mathcal{B}$-cell is less than (or greater than) $-\dfrac{k_1}{k_2}$, then the cell will become all-inhibitory (or all-excitatory). This is because, in our approach the system (6) is in Regime C (i.e., $k_1 + k_2 > -1/N_{\mathcal{A}}$ and $k_1 - k_2 < 1/N_{\mathcal{A}}$), where the all-excitatory and the all-inhibitory aRFs are coexisting and dominant, and the outcome is dependent on the initial condition.

4. In Linsker's simulations, when $k_2 < 0$ and $-1 < -\dfrac{k_1}{k_2} < 1$, the mature pattern will have $g = c(\omega) = -\dfrac{k_1}{k_2}$, but its spatial structure is totally determined by the initial condition. Our results show that the system (6) is in Regime D (i.e., $k_1 + k_2 < -1/N_{\mathcal{A}}$ and $k_1 - k_2 > 1/N_{\mathcal{A}}$), where neither the all-excitatory nor the all-inhibitory connection patterns are stable. However, since *every* kind of structured aRF has a stable parameter regime as shown in (23), the stable regimes for different aRFs will overlap with each other, and there is no way to avoid this kind of coexistence. Thus, the weight development in Regime D is extremely sensitive to the initial condition.

Therefore, it is impossible to robustly emerge any structured aRF between layers $\mathcal{A}$ and $\mathcal{B}$ without regard to the initial condition. But, as described next, the situation will be fundamentally changed in the deeper layers of the Linsker's network: the system will incorporate more parameters (i.e., more ranges of SDFs), so that one can indeed make the presence or absence of a specially designated aRF by adjusting the relative ranges of the SDFs, and finally make the corresponding Regime E or F (with a certain slope value) become unique in Regime D (with large and negative k_2).

4.3 Analytic Studies Of SDFs' Influences In The First Three Layers

The relative ranges of two SDFs $r^{\mathcal{AB}}(\cdot,\cdot)$ and $r^{\mathcal{BC}}(\cdot)$ will play a key role in the evolution of various structured aRFs from $\mathcal{B}$ to $\mathcal{C}$. For example, in Linsker's simulations it is required that the range of $r^{\mathcal{BC}}(\cdot)$ be larger than the range of $r^{\mathcal{AB}}(\cdot,\cdot)$ for certain structured aRFs to emerge. In this section, we first explore an extreme case of this constraint on the relative ranges (Theorem 4): $r^{\mathcal{BC}}(\cdot)$ is a constant (hence, the range of $r^{\mathcal{BC}}(\cdot)$ is equal to its largest extreme) and $r^{\mathcal{AB}}(i,\cdot)$ almost concentrates on the $\mathcal{A}$-cell underlying the i-th site in layer $\mathcal{B}$ (hence, the range of $r^{\mathcal{AB}}(\cdot,\cdot)$ is small). That is, the synaptic connections from layer $\mathcal{B}$ to $\mathcal{C}$ are fully feedforward connected with constant density, and those from $\mathcal{A}$ to $\mathcal{B}$ are concentrated enough (obviously including the extreme one-to-one projection). In Theorem 5, we justify the necessity of the constraints in Linsker's simulations by exploring the other extreme case, where connections from layer $\mathcal{A}$ to $\mathcal{B}$ are fully feedforward connected with constant density.

Theorem 4 *If* $r^{\mathcal{BC}}(i) = \frac{1}{N_{\mathcal{B}}}$, *and* $\sum_{l=1}^{N_{\mathcal{A}}}[r^{\mathcal{AB}}(i,l)]^2 > \frac{1}{2}$, *for* $i = 1, \cdots, N_{\mathcal{B}}$, *then*

$$\{\omega \mid k_1 \cdot N_{\mathcal{B}} + k_2(|J^+(\omega)| - |J^-(\omega)|) = 0, \quad \omega \in V(\Omega) = \{-1,1\}^{N_{\mathcal{B}}}\} \subset \Omega_{sFPA},$$

where $|\cdot|$ *represents the total number of elements in a set.*

Proof: By Theorem 3 of Section 3, we only need to check that

$$d_2(\omega) > 0 > d_1(\omega),$$

i.e.,

$$\sum_{j\in J^+(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) > \sum_{j\in J^-(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j), \ \forall i \in J^+(\omega),$$

and

$$\sum_{j\in J^+(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) < \sum_{j\in J^-(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j), \ \forall i \in J^-(\omega)$$

hold true for any $\omega \in V(\Omega)$.
Observe that for each $i \in J^+(\omega)$,

$$Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) = r^{\mathcal{BC}}(j) \sum_{l=1}^{N_{\mathcal{A}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l) > 0, \quad \text{and } r^{\mathcal{BC}}(i) = \frac{1}{N_{\mathcal{B}}}, \forall i,j \in \mathcal{B}.$$

Hence, we can show that for each $i \in J^+(\omega)$,

$$\sum_{j\in J^+(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) > Q^{\mathcal{B}}_{ii} r^{\mathcal{BC}}(j) = \frac{1}{N_{\mathcal{B}}} \sum_{l=1}^{N_{\mathcal{A}}} [r^{\mathcal{AB}}(i,l)]^2.$$

On the other hand,

$$\begin{aligned}
\sum_{j\in J^-(\omega)} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) &\leq \sum_{j\neq i} Q^{\mathcal{B}}_{ij} r^{\mathcal{BC}}(j) \\
&\leq \sum_{j\neq i} \sum_{l=1}^{N_{\mathcal{A}}} \frac{1}{N_{\mathcal{B}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l) \\
&\leq \sum_{l=1}^{N_{\mathcal{A}}} \frac{1}{N_{\mathcal{B}}} r^{\mathcal{AB}}(i,l)(1 - r^{\mathcal{AB}}(i,l)) \\
&= \frac{1}{N_{\mathcal{B}}}(1 - \sum_{l=1}^{N_{\mathcal{A}}} [r^{\mathcal{AB}}(i,l)]^2),
\end{aligned}$$

where the third inequality is because of the symmetry of $r^{\mathcal{AB}}(\cdot,\cdot)$ and $\sum_{j\neq i} r^{\mathcal{AB}}(j,l) = 1 - r^{\mathcal{AB}}(i,l)$.

Since we assume $\sum_{l=1}^{N_{\mathcal{A}}}[r^{\mathcal{AB}}(i,l)]^2 > \frac{1}{2}$, for $i = 1,\cdots,N_{\mathcal{B}}$, we have the desired result:

$$\sum_{j\in J^+(\omega)} Q_{ij}^{\mathcal{B}} r^{\mathcal{BC}}(j) > \sum_{j\in J^-(\omega)} Q_{ij}^{\mathcal{B}} r^{\mathcal{BC}}(j),\ \forall i \in J^+(\omega).$$

The case for $i \in J^-(\omega)$ can be treated similarly. □

Theorem 4 shows that any $\omega \in V(\Omega)$ satisfying

$$\frac{|J^+(\omega)| - |J^-(\omega)|}{N_{\mathcal{B}}} = -\frac{k_1}{k_2} \tag{24}$$

is an attractor of equation (22). Namely, any extreme point in $V(\Omega) = \{-1,1\}^{N_{\mathcal{B}}}$ satisfying (24) is an attractor of the weight-development dynamics from $\mathcal{B}$ to $\mathcal{C}$, under the condition in which the synaptic connections from $\mathcal{A}$ to $\mathcal{B}$ are concentrated enough, and those from $\mathcal{B}$ to $\mathcal{C}$ are fully feedforward connected with constant density. In particular, for the grid system used in our examples (see Appendix B), if we choose the system parameters k_1 and k_2 such that $|\frac{k_1}{k_2}|N_{\mathcal{B}}$ is an integer less than $N_{\mathcal{B}}$, then any kind of connection pattern (with the average synaptic strength $-\frac{k_1}{k_2}$) will be an attractor of equation (22). In fact, since $\sum_{l=1}^{N_{\mathcal{A}}}[r^{\mathcal{AB}}(i,l)]^2 > \frac{1}{2}$, it follows that $Q_{ij}^{\mathcal{B}} \approx 1$, if $i = j$ and $Q_{ij}^{\mathcal{B}} \approx 0$ if $i \neq j$, i.e., each connection strength between layers $\mathcal{B}$ and $\mathcal{C}$ is developed independently. As a result of this, any pattern of connection strengths may emerge. This extreme case is corroborated by every example in Section 4.4 and Appendix B, and is similar with the situation between layers $\mathcal{A}$ and $\mathcal{B}$.

As the condition of this theorem fails, we expect the connection strengths to become more and more organized in the sense that most connections will have the same sign as their neighbors in an aRF. Next, we turn to another extreme case where the synaptic connections from layer $\mathcal{A}$ to $\mathcal{B}$ are fully feedforward connected, i.e., $r^{\mathcal{AB}}(\cdot,\cdot)$ is a constant.

Theorem 5 *Assume that*

$$r^{\mathcal{AB}}(j,i) = \frac{1}{N_{\mathcal{A}}},\ i \in \mathcal{A},\ j \in \mathcal{B},$$

then all saturated fixed points are not *stable except the all-excitatory and all-inhibitory connection patterns.*

Proof: Note that under the assumption of this theorem, we have

$$Q_{ij}^{\mathcal{B}} = \sum_{l=1}^{N_{\mathcal{A}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l) = \frac{1}{N_{\mathcal{A}}}.$$

Now equation (22) becomes

$$\omega_{\tau+1}(i) = f(\omega_\tau(i) + k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2] r^{\mathcal{BC}}(j)\omega_\tau(j)).$$

Notice that for any $\omega \in V(\Omega)$, the second term $k_1 + \sum_{j=1}^{N_\mathcal{B}}[\frac{1}{N_\mathcal{A}} + k_2] r^{\mathcal{BC}}(j)\omega_\tau(j)$ on the right hand side of the above equation is independent of i, and is either a positive or a negative constant. By Theorems 1 and 2, a saturated fixed point attractor ω should satisfy that for every i, $\omega(i)$ has the same sign as the second term which is independent of i. Therefore, we arrive at the final conclusion. □

Theorem 5 shows that only the all-excitatory and the all-inhibitory aRFs will be the stable saturated fixed points if the weights from layer $\mathcal{A}$ to $\mathcal{B}$ are fully feedforward connected with constant density but there is no constraint on the SDF $r^{\mathcal{BC}}(\cdot)$. However, it is still possible that unsaturated fixed points and other kinds of attractors (for example, limit cycle or other more complicated cases) might occur within certain parameter regimes. So a relevant question is: How large are the attractive basins for both all-excitatory and all-inhibitory aRFs under the assumptions of Theorem 5? To confirm this point, we should consider the initial conditions and study the parameter space and the weight state space together. As we shall show in the following corollary, the all-excitatory and the all-inhibitory aRFs are the only terminal attractors in the upper half of the parameter subspace of (k_1, k_2).

Corollary 2 *Assume that*

$$r^{\mathcal{AB}}(j,i) = \frac{1}{N_\mathcal{A}},\ i \in \mathcal{A},\ j \in \mathcal{B},\ \textit{and}\ k_2 \geq -\frac{1}{N_\mathcal{A}},$$

then we have

$$\lim_{\tau\to\infty} \omega_\tau(j) = +1,\ \forall j \in \mathcal{B}$$

if the initial condition ω_0 satisfies

$$k_1 + \sum_{j=1}^{N_\mathcal{B}}[\frac{1}{N_\mathcal{A}} + k_2] r^{\mathcal{BC}}(j)\omega_0(j) > 0;$$

and

$$\lim_{\tau\to\infty} \omega_\tau(j) = -1,\ \forall j \in \mathcal{B}$$

if the initial condition ω_0 satisfies

$$k_1 + \sum_{j=1}^{N_\mathcal{B}}[\frac{1}{N_\mathcal{A}} + k_2] r^{\mathcal{BC}}(j)\omega_0(j) < 0.$$

Proof: We only consider the case of

$$k_1 + \sum_{j=1}^{N_\mathcal{B}}[\frac{1}{N_\mathcal{A}} + k_2] r^{\mathcal{BC}}(j)\omega_0(j) > 0.$$

Following the same argument as in Theorem 5, the equation (22) becomes

$$\omega_{\tau+1}(i) = f(\omega_\tau(i) + k_1 + \sum_{j=1}^{N_\mathcal{B}}[\frac{1}{N_\mathcal{A}} + k_2] r^{\mathcal{BC}}(j)\omega_\tau(j)),$$

and the second term $k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2]r^{\mathcal{BC}}(j)\omega_\tau(j)$ on the right hand side of the above equation is independent of i. Because of the nondecreasing property of the function $f(\cdot)$, we see that

$$k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2]r^{\mathcal{BC}}(j)\omega_\tau(j) > 0$$

implies

$$\omega_{\tau+1}(i) \geq \omega_\tau(i), \forall i \in \mathcal{B}.$$

From this we assert that

$$k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2]r^{\mathcal{BC}}(j)\omega_{\tau+1}(j) \geq k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2]r^{\mathcal{BC}}(j)\omega_\tau(j)$$

because of the nonnegativity of $\frac{1}{N_{\mathcal{A}}} + k_2$. If we denote

$$\delta = k_1 + \sum_{j=1}^{N_{\mathcal{B}}}[\frac{1}{N_{\mathcal{A}}} + k_2]r^{\mathcal{BC}}(j)\omega_0(j) > 0,$$

then we have

$$\omega_{\tau+1}(i) \geq f(\omega_0(i) + (\tau+1)\delta).$$

Therefore, we arrive at the final conclusion. $\square$

Remark 4 In Theorems 4 and 5, we considered the influence of the SDF's ranges $r_{\mathcal{B}}, r_{\mathcal{C}}$ on the dynamics by changing their values from the smallest extreme to the largest one. When $r_{\mathcal{B}}$ is small but $r_{\mathcal{C}}$ is in the largest extreme (Theorem 4), we proved that any kind of connection pattern (even including the oriented cells or any other circularly asymmetric patterns) has a stable parameter regime and emerge under a certain set of parameters, because each synaptic connection within an aRF is developed independently. As $r_{\mathcal{B}}$ is changed from the smallest to the largest extreme, the development of synaptic connections between layers $\mathcal{B}$ and $\mathcal{C}$ will depend on each other stronger and stronger in the sense that most connections have the same sign as their neighbors in an aRF. So for the largest extreme of $r_{\mathcal{B}}$ (Theorem 5), any structured aRF will never emerge at all, although there exist correlations in input activities in layer $\mathcal{B}$. $\square$

Up to this point, we have already shown that the dynamical behaviors of the Linsker's network are determined by the complicated relationships among its system parameters k_1, k_2, and the ranges $r_{\mathcal{B}}$, $r_{\mathcal{C}}$, etc. While we have studied the properties of only saturated fixed point attractors of equation (6), the following corollary shows that the system can display limit cycles in certain parameter regimes. Thus, the intricate relationships among the parameters can lead to nontrivial system behaviors.

Corollary 3 *If we choose the parameters k_2 large and negative enough such that*

$$2 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j) < k_1 < -2 - \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j),$$

then once the dynamics converge to one of the all-excitatory and the all-inhibitory weight vectors, it will stay at a two-state limit cycle. This limit cycle is composed of the all-excitatory and the all-inhibitory weight states.

Proof: Because of the special form of $f(\cdot)$, if the weight vector ω is all-excitatory at the time τ, and

$$1 + k_1 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j) < -1,$$

then the weight vector ω will be all-inhibitory at time $\tau + 1$; if at the same time, k_1 and k_2 also satisfy

$$-1 + k_1 - \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j) > 1,$$

then ω will return to the all-excitatory state at time $\tau + 2$. Therefore, if we chose the parameters to fulfill the condition

$$2 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j) < k_1 < -2 - \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij} + k_2]r(j),$$

then the system at least has a two-state limit cycle. It is obvious that at this time both all-excitatory and all-inhibitory connection patterns no longer belong to the set of fixed points, since

$$\omega(i)\{k_1 + \sum_{j=1}^{N_{\mathcal{L}}}[Q_{ij}^{\mathcal{L}} + k_2]r(j)\omega(j)\} < 0, \quad \forall i \in \mathcal{B}.$$

□

In our following examples, we shall consider the general case for $r^{\mathcal{AB}}(\cdot,\cdot)$ and $r^{\mathcal{BC}}(\cdot)$, and investigate the effect of the set of parameters (including k_1, k_2, and the ranges of SDFs) on the emergence of various structured aRFs from layer $\mathcal{B}$ to layer $\mathcal{C}$.

4.4 Examples Of Structured aRFs Between Layers $\mathcal{B}$ And $\mathcal{C}$

As an application of our above analyses, we present several examples of the explicit parameter regimes of $(k_1, k_2, r_{\mathcal{B}}, r_{\mathcal{C}})$ for the formation of various structured aRFs, including ON-center, OFF-center, oriented, and bi-lobed cells in Appendix B. Without loss of generality, we assume that the connection strengths from layer $\mathcal{A}$ to $\mathcal{B}$ are all-excitatory like in Linsker's simulations. One has no difficulty to evaluate the k_1-intercept functions $d_1(+)$ and $d_2(-)$ for the all-excitatory and the all-inhibitory aRFs between layers $\mathcal{B}$ and $\mathcal{C}$ as follows:

$$d_1(+) = -\min_{1 \le i \le N_{\mathcal{B}}} \sum_{j=1}^{N_{\mathcal{B}}} r^{\mathcal{BC}}(j) \sum_{l=1}^{N_{\mathcal{A}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l) \ ,$$

and

$$d_2(-) = \min_{1 \le i \le N_{\mathcal{B}}} \sum_{j=1}^{N_{\mathcal{B}}} r^{\mathcal{BC}}(j) \sum_{l=1}^{N_{\mathcal{A}}} r^{\mathcal{AB}}(i,l) r^{\mathcal{AB}}(j,l) \ .$$

The aim of the demonstration here is to further answer those questions about the parameter space we raised in Section 3.2. We summarize the examples in Appendix B as follows:

1. For a fixed value of $r_{\mathcal{C}}$ ($= 2.5$, 10, and 100), the SDF's range $r_{\mathcal{B}}$ as the third system parameter has various *critical values* for different attractors. That is, several of the fixed points are no longer stable if $r_{\mathcal{B}}$ exceeds the corresponding critical values (see Figs.13,14,16, & 18).

2. For circularly symmetric ON-center cells, those receptive fields with large ON-center core (which have positive or small negative slope value $c(\omega) \approx -\frac{k_1}{k_2}$) always have a stable parameter regime. But for those ON-center cells with large negative slope value $c(\omega)$, their stable parameter regimes decrease in size with $c(\omega)$ (see Fig. 6).

   Similarly, circularly symmetric OFF-center cells with large OFF-center core (which have negative or small positive slope value $c(\omega)$) will be more stable than those with large positive average weights (see Fig. 7).

   But for non-circularly-symmetric patterns, only those attractors with small slope value $|c|$ might always have a stable parameter regime (see Fig. 8 & 9).

3. The results about the relative stability among different aRFs give us a good reason why we usually obtain ON-center or OFF-center receptive fields between layers $\mathcal{B}$ and $\mathcal{C}$ much more easily than other patterns. That is, the ON-center patterns are almost the only aRFs in Regime D∩E, and the OFF-center patterns are almost the only aRFs in Regime D∩F *if $r_{\mathcal{B}}$ is large enough to exceed all the critical values* for other kinds of aRFs (e.g., in Appendix B, $r_{\mathcal{B}} > 5.0724$ with respect to $r_{\mathcal{C}} = 10$) (see Fig. 6-9).

4. The situation around the axis k_2 is slightly complicated, since a number of aRFs with an average synaptic strength near zero might always have a stable regime covering the axis k_2 (see Regime H in Fig. 4). But it is important to note that the bi-lobed aRF has a broader stable regime than other kinds of aRFs, i.e., the bi-lobed aRF can be observed with higher possibility in the simulation of the weight evolution in equation (6), provided $c(\omega) \approx -\frac{k_1}{k_2}$ is chosen to be near zero and k_2 large and negative.

5. Our theorems and Example 3 in Appendix B indicate that there exists the possibility of emerging oriented aRFs in the first three layers of the Linsker's network. Actually, Theorem 4 shows that any kind of aRF could be an attractor of equation (22) for any two adjacent layers of the Linsker's network. The occurrence of a specially designated aRF (e.g., the oriented cells) in numerical simulations of the Linsker's network is determined by the appropriate choice of the parameters and the initial condition. As observed in our examples, the stable parameter regimes for oriented aRFs always overlap with other major aRFs (e.g., ON-center cells and OFF-center cells). That is, there is no parameter regime that could be made unique for the oriented cells in the first three layers of the Linsker's network (in the sense that we clarified in Section 3.2). Therefore, the emergence of an oriented aRF depends crucially on the choice of the initial condition, and cannot be easily observed if initial conditions are chosen at random. But for non-circularly-symmetric patterns (e.g., certain oriented aRFs), it is still possible to find a unique stable parameter regime in deeper layers of the Linsker's network (with more parameters in one's hand) as shown in Linsker's simulations.

6. **Linsker's simulations of $\mathcal{B}$-to-$\mathcal{C}$ aRFs and our explanations:**
   In conclusion, the complete and precise picture of the division of parameter regimes presented

here accurately match Linsker's simulation results of the first three layers reported in [20]. That is, given k_2 large and negative, Linsker passes through a series of regimes as k_1 is decreased from large and positive to large and negative as we have shown with the dashed line in Fig. 10.

(a) In Linsker's simulations, the *all-excitatory* pattern is the only attractor when $-\frac{k_1}{k_2} > 1$[6]. Our results show that under this set of parameters, the system (6) is in Regime $\mathrm{A} \cap \mathrm{B}^c$, where the all-excitatory aRF is the only attractor.

(b) When $0.3 < -\frac{k_1}{k_2} \approx g < 1$, the *ON-center* cells are the only aRFs that are observed in Linsker's simulations. Our results show that the system (6) is in Regime D∩E with $c(\omega) > 0.3$ within which the ON-center cells could be the only aRFs if $r_{\mathcal{B}}$ is chosen to be large enough (with respect to $r_{\mathcal{C}}$) to exceed its critical values for other aRFs.

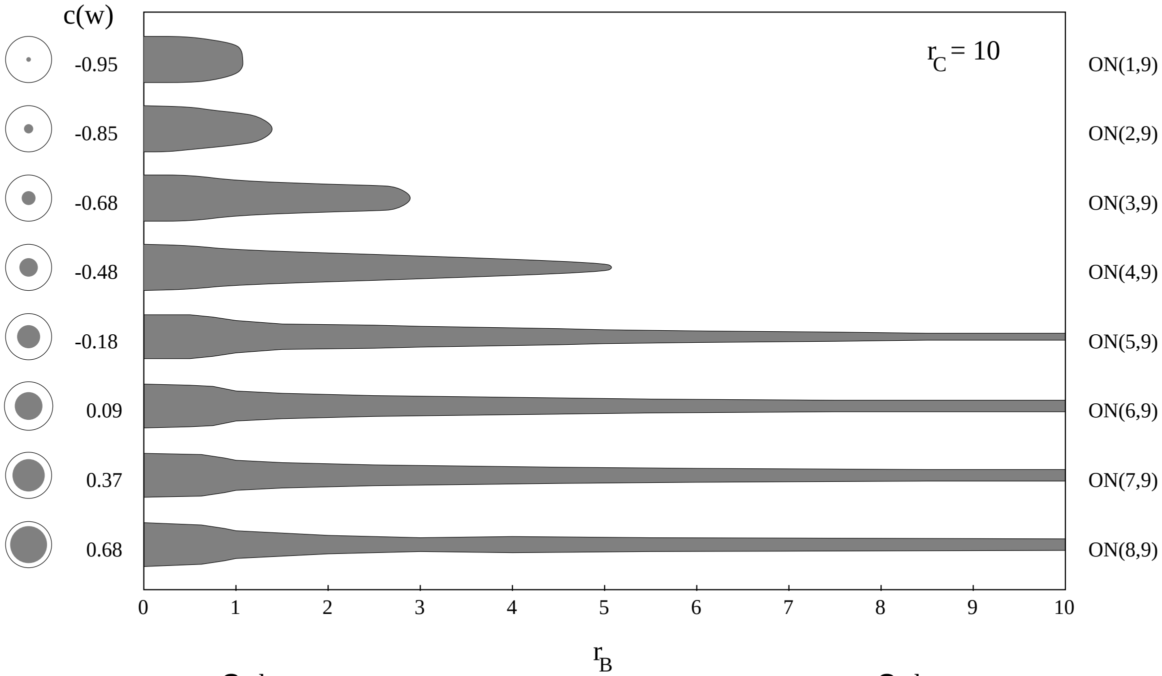
(c) Various *bi-lobed* cells are the outcome of his simulations when $-0.2 < -\frac{k_1}{k_2} \approx g < 0.2$. In our analysis, for the above range of $-\frac{k_1}{k_2}$, the system (6) is either in Regime H with $c(\omega) \approx 0$ or in Regime D∩E∩F with $-0.2 < c(\omega) < 0.2$. Our results show that in Regime H, various aRFs coexist but the bi-lobed cell has a broader stable regime. In Regime D∩E∩F with $-0.2 < c(\omega) < 0.2$, however, the ON-center and OFF-center cells coexist as attractors of the system (6), and could become dominant. That is, bi-lobbed and oriented aRFs could be made unstable if $r_{\mathcal{B}}$ is chosen to be large enough (with respect to $r_{\mathcal{C}}$) to exceed the corresponding critical values. In general, the outcome from this regime can be sensitive to the initial condition.

(d) Symmetrically, the *OFF-center* cells are the only aRFs when $-1 < -\frac{k_1}{k_2} \approx g < -0.3$. This is because the system (6) is in Regime D∩E with $c(\omega) < -0.3$ within which the OFF-center cells could be the only aRFs if $r_{\mathcal{B}}$ is chosen to be large enough (with respect to $r_{\mathcal{C}}$) to exceed its critical values for other kinds of aRFs.

(e) Finally, the *all-inhibitory* pattern is the only attractor when $-\frac{k_1}{k_2} < -1$. Under this set of parameters, the system (6) is in Regime $\mathrm{B} \cap \mathrm{A}^c$.

In conclusion, we have shown that the existence of a stable parameter regime for an aRF is determined by the SDFs of all preceding layers and the present layer. Thus, one could make the presence or absence of a designated aRF by choosing the relative ranges of the SDFs, and a methodology for achieving this is outlined in Sections 3.2 and 4.4. That is, for a given set of SDFs, one can determine a corresponding set of aRFs that have stable regimes in the (k_1, k_2) subspace, and any connection pattern not in this set cannot emerge under *any* choice of (k_1, k_2). The role of (k_1, k_2) is to put the system in Regime D (i.e., away from the stable regimes of the all-excitatory or all-inhibitory patterns by choosing k_2 large and negative), and then to choose a desired aRF (by choosing $-\frac{k_1}{k_2}$ approximately equal to the average synaptic strength).

[6] Linsker's simulations used $n_{EM} = 0.5$ (refer to Remark 1(a) in Section 1.1); hence, our average synaptic strength $g = 2g_{Linsker}$.



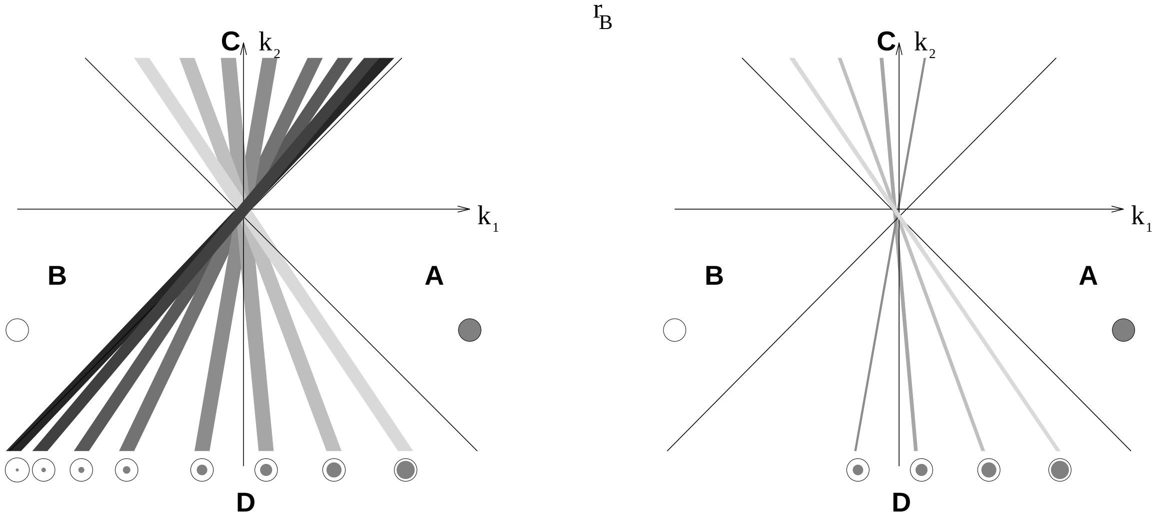


Figure 6: The critical values of $r_{\mathcal{B}}$ as a function of $c(\omega)$ with fixed $r_{\mathcal{C}} = 10$, where ω's are ON-center aRFs. We use a grid system in our examples, and assume that each cell in layer $\mathcal{C}$ receives synaptic inputs from 253 sites in layer $\mathcal{B}$, where 253 is the total number of sites inside the circle with grid radius 9 (see Fig. 1); see Appendix B for further details. In the upper figure, each horizontal band corresponds to a kind of ON-center cell denoted by the radius of its ON-center core as $ON(r_{core}, 9)$. Each circle of the left-vertical-axis label, represents the spatial pattern of an ON-center aRF from layer $\mathcal{B}$ to layer $\mathcal{C}$ (see Fig. 11). The vertical width of each band (as the function of $r_{\mathcal{B}}$) is equal to the difference of $d_2(\omega) - d_1(\omega)$ at that $r_{\mathcal{B}}$, and exists only if $d_2(\omega) > d_1(\omega)$ (refer to Theorem 3). The absolute value of the width shown here is proportional to $log(\alpha(d_2(\omega) - d_1(\omega)))$, where α is a positive constant. As $r_{\mathcal{B}}$ increases, the patterns, $ON(1,9), ON(2,9), ON(3,9)$, and $ON(4,9)$, become unstable. That is, as $r_{\mathcal{B}}$ exceeds the corresponding critical values (1.0674, 1.4634, 2.9485, and 5.0724 in the same order), the patterns no longer have a stable regime in the (k_1, k_2) subspace. But the rest of the ON-center cells (i.e., $ON(5,9), ON(6,9), ON(7,9)$, and $ON(8,9)$) always have a stable parameter regime. The two lower figures show the parameter subspace (k_1, k_2) when $r_{\mathcal{B}} < 1.0674$ and $r_{\mathcal{B}} > 5.0724$, respectively.

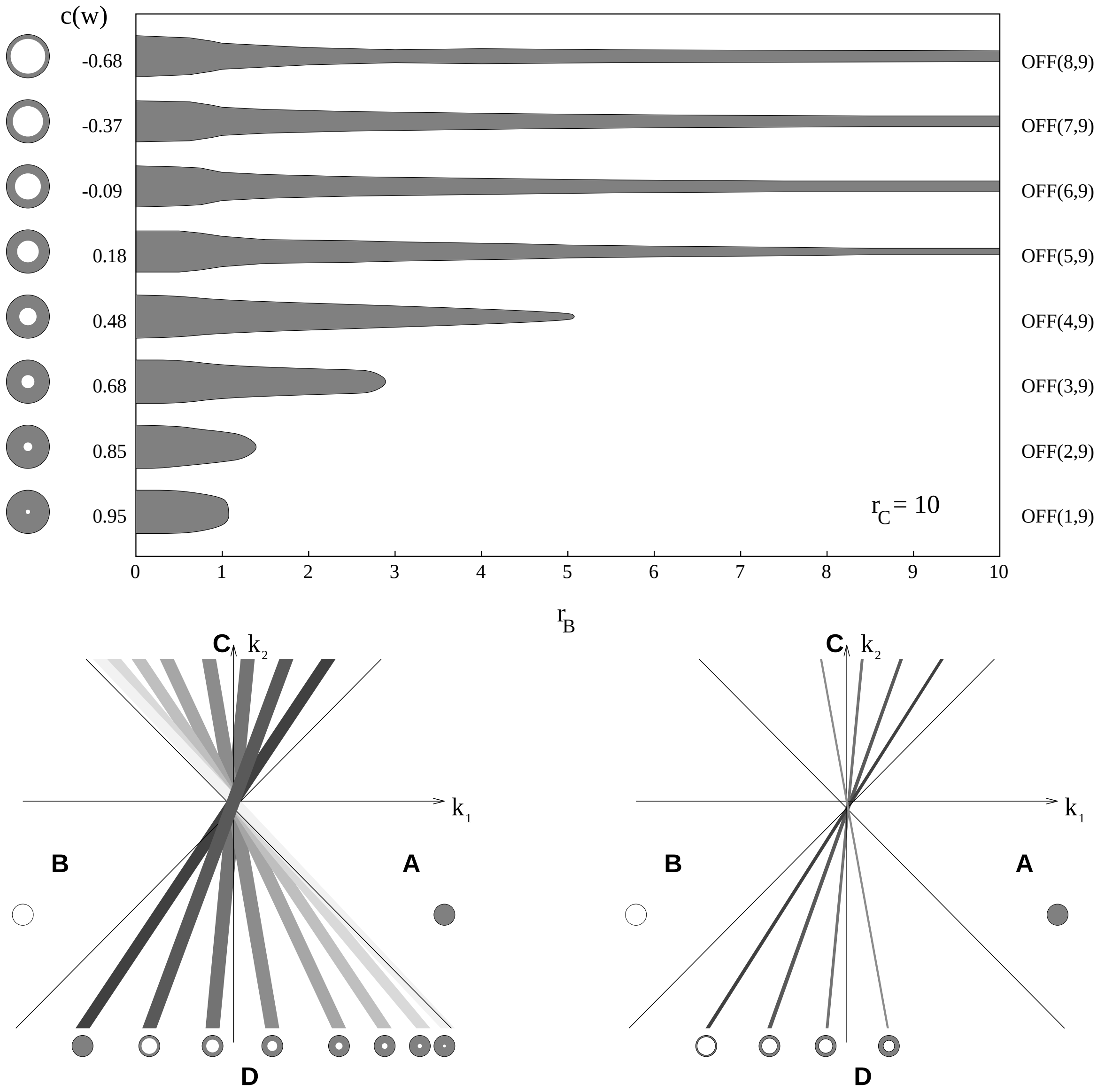


Figure 7: The critical values of $r_{\mathcal{B}}$ as a function of $c(\omega)$ with fixed $r_{\mathcal{C}} = 10$, where ω's are OFF-center aRFs (under the same setup as in Fig. 6). In the upper figure, each horizontal band corresponds to a kind of OFF-center cell denoted by the radius of its OFF-center core as $OFF(r_{core}, 9)$. Each circle of the left-vertical-axis label, represents the spatial pattern of an OFF-center aRF from layer $\mathcal{B}$ to layer $\mathcal{C}$ (see Fig. 11). As $r_{\mathcal{B}}$ increases, there will be no stable parameter regimes for $OFF(1,9), OFF(2,9), OFF(3,9)$, and $OFF(4,9)$, i.e., $r_{\mathcal{B}}$ exceeds the corresponding critical values (1.0674, 1.4634, 2.9485, and 5.0724 in the same order). But the rest of the OFF-center cells (i.e., $OFF(5,9), OFF(6,9), OFF(7,9)$, and $OFF(8,9)$) always have a stable parameter regime. The two lower figures show the parameter subspace (k_1, k_2) when $r_{\mathcal{B}} < 1.0674$ and $r_{\mathcal{B}} > 5.0724$, respectively.

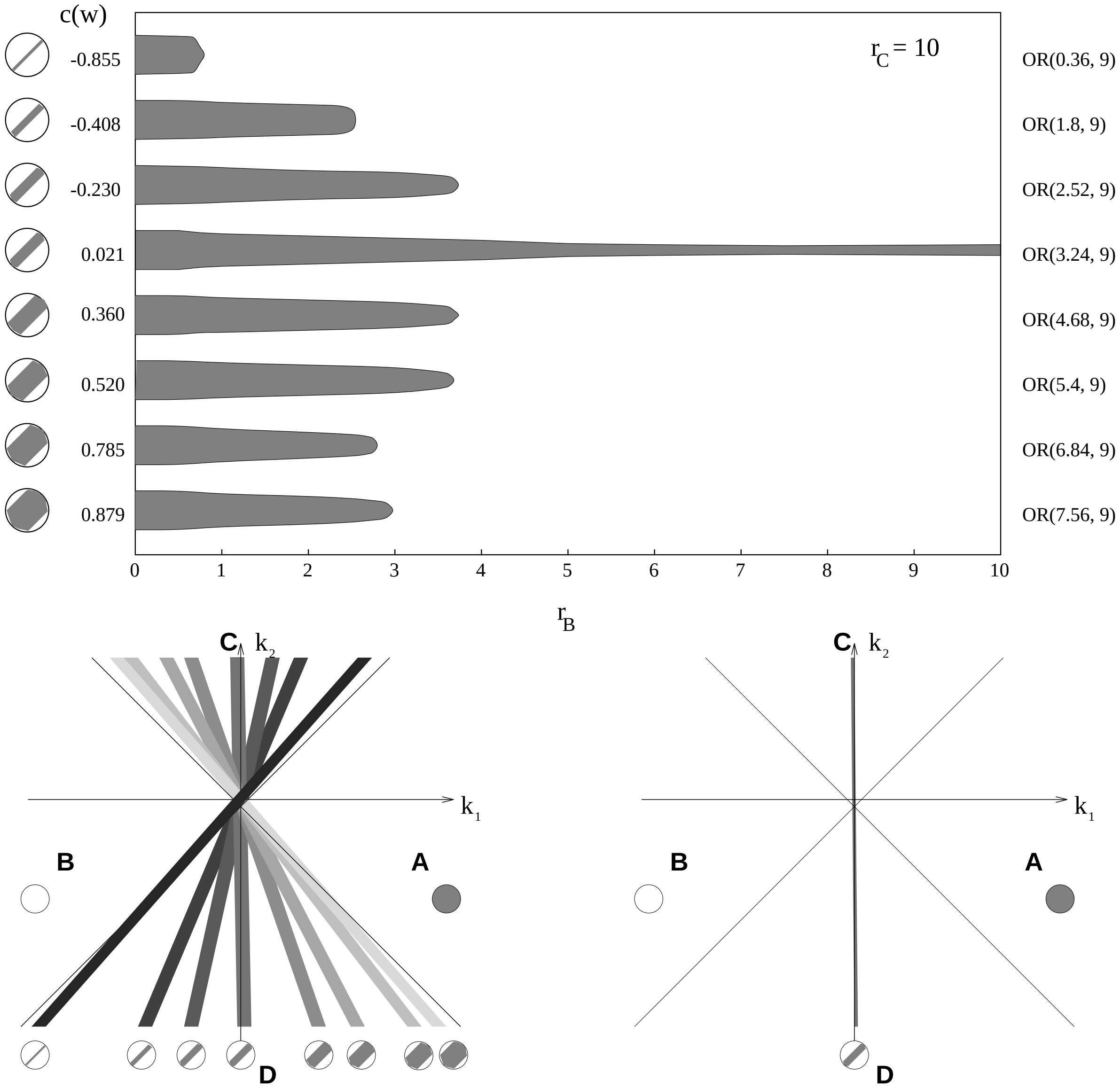


Figure 8: The critical values of $r_{\mathcal{B}}$ as a function of $c(\omega)$ with fixed $r_{\mathcal{C}} = 10$, where ω's are oriented aRFs (under the same setup as in Fig. 6). In the upper figure, each horizontal band corresponds to a kind of oriented cell denoted by the radius r_{width} as $OR(r_{width}, 9)$ (see Fig. 11). As $r_{\mathcal{B}}$ increases, there will be no stable parameter regimes for all of oriented cells except for $OR(3.24, 9)$ with an average synaptic strength near zero. The two lower figures show the parameter subspace (k_1, k_2) when $r_{\mathcal{B}} < 0.8073$ and $r_{\mathcal{B}} > 3.7741$, respectively.

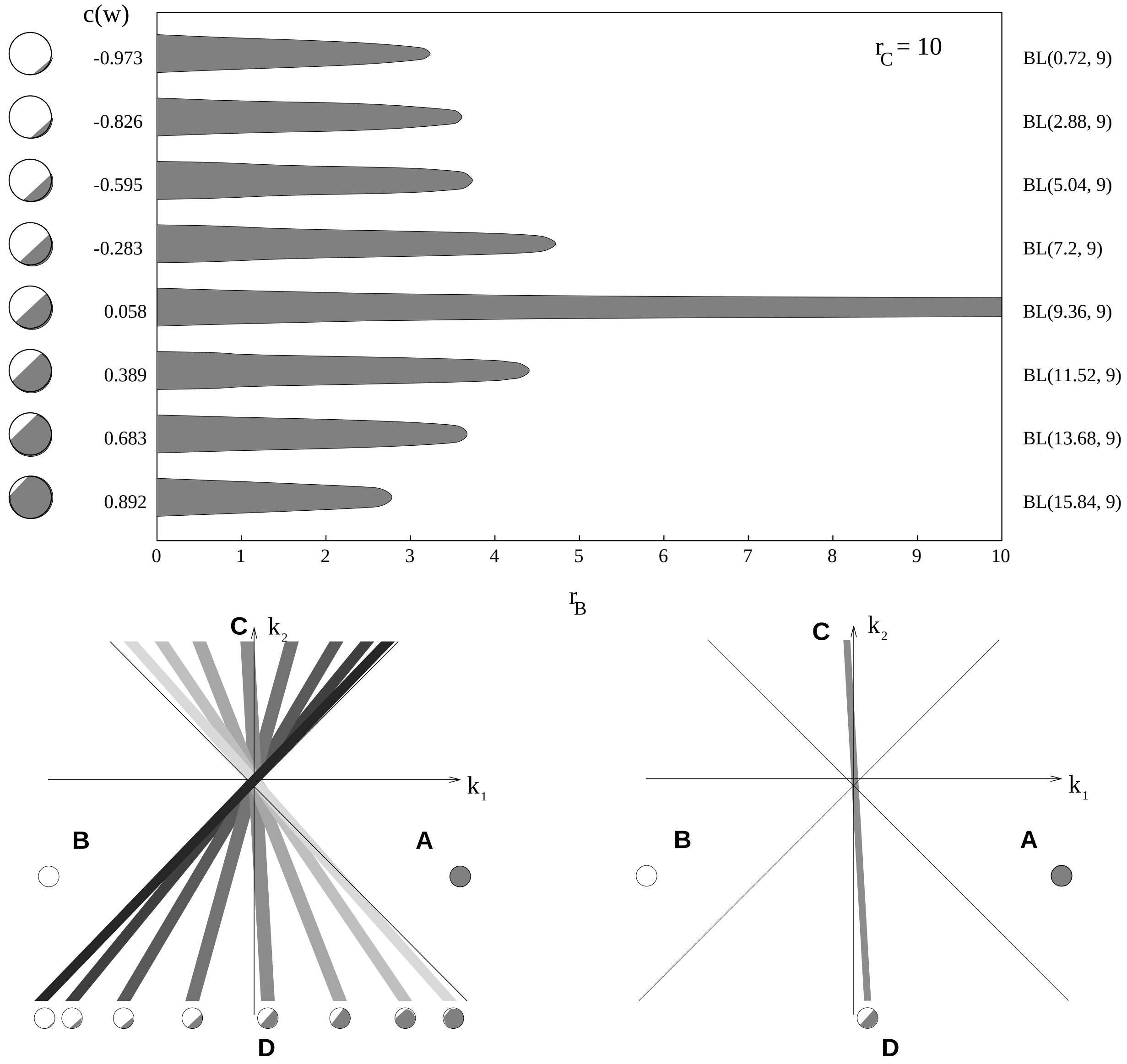


Figure 9: The critical values of $r_{\mathcal{B}}$ as a function of $c(\omega)$ with fixed $r_{\mathcal{C}} = 10$, where ω's are bi-lobed aRFs (under the same setup as in Fig. 6). In the upper figure, each horizontal band corresponds to a kind of bi-lobed cell denoted by the radius r_{BL} as $BL(r_{BL}, 9)$ (see Fig. 11). As $r_{\mathcal{B}}$ increases, there will be no stable parameter regimes for all of bi-lobed cells except for $BL(9.36, 9)$ with an average synaptic strength near zero. The two lower figures show the parameter subspace (k_1, k_2) when $r_{\mathcal{B}} < 2.8100$ and $r_{\mathcal{B}} > 4.7886$, respectively.

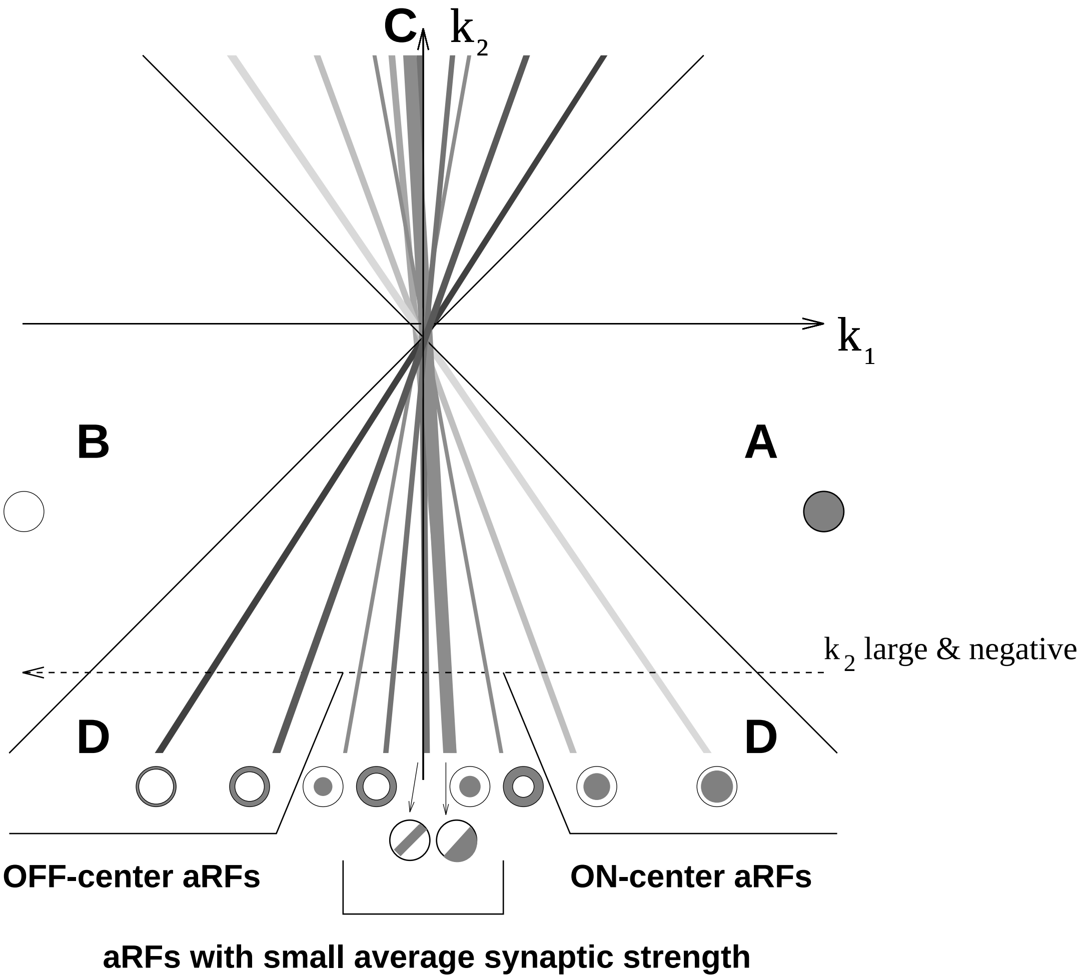


Figure 10: An example of the appropriate way to avoid the coexistence of $\mathcal{B}$-to-$\mathcal{C}$ aRFs by choosing a feasible set of parameters. (1) Only within Regime D, various Regime E's and Regime F's corresponding to various aRFs will be removed from Regime A, B, C, and G where the all-excitatory and all-inhibitory aRFs are dominant or many kinds of attractors coexist. (2) Only when k_2 is chosen to be large and negative, these Regime E's and Regime F's with different slope values will be separate from each other in Regime D. (3) Only when the ranges of SDFs are chosen appropriately (e.g., in our examples, the range $r_{\mathcal{B}}$ of the SDF from layer $\mathcal{A}$ to $\mathcal{B}$ is chosen to exceed its critical values for various aRFs, and to ensure ON-center aRFs or OFF-center aRFs are only aRFs when $0.3 < -\dfrac{k_1}{k_2} \approx g < 1$ and $-1 < -\dfrac{k_1}{k_2} \approx g < -0.3$ respectively), a Regime E or F for an aRF with a certain slope value may become unique in Regime D with large and negative k_2. (In our examples shown here, $r_{\mathcal{B}} > 5.0724$ with respect to $r_{\mathcal{C}} = 10$. The dashed line shows the procedure in which $-k_1/k_2$ is decreased from larger than $+1$ to less than -1 while k_2 is chosen to be large and negative.

5 Concluding Remarks

5.1 Synaptic Arbor Density Function

The crucial role played by the SDFs was recognized in Linsker's simulations, but a theoretical explanation has not been provided in previous investigations. Our analysis, for the first time, provides justification for the choice of the relative ranges of the SDFs in Linsker's simulations, and in addition provides novel theoretical predictions on the behavior of the Linsker's network. For example, Theorem 5 shows that if the all-excitatory connections from layer $\mathcal{A}$ to $\mathcal{B}$ are fully connected (i.e., the synaptic arbor density is constant), then no structured aRF will emerge between layers $\mathcal{B}$ and $\mathcal{C}$ (although there exists correlation in afferent activities of layer $\mathcal{B}$). In general, a necessary condition for the emergence of a structured aRF from layer $\mathcal{L}$ to $\mathcal{M}$ is that the covariance matrix $Q^{\mathcal{L}}$ of layer $\mathcal{L}$ in (4) should include *localized correlations*. The localized correlations in $Q^{\mathcal{L}}$ originate from the localized arbor density of synapses of cells in some layer preceding layer $\mathcal{L}$. In the extreme case studied in Theorem 5, full connectedness with constant density from $\mathcal{A}$ to $\mathcal{B}$ makes the afferent activities in layer $\mathcal{B}$ fully correlated, and irrespective of the SDF from $\mathcal{B}$ to $\mathcal{C}$, no structured aRF between $\mathcal{B}$ and $\mathcal{C}$ can occur. Therefore, the localized arborization of synapses between adjacent layers plays a crucial role in the emergence of structured aRFs.

5.2 The Linsker's Network And The BSB Model

The dynamical behavior of the Linsker's network is much more complicated than the retrieval dynamics in the Brain-State-in-a-Box (BSB) model. In the BSB model, the weight matrix is diagonal-dominant and/or symmetric, which ensures that the extreme points on its state hypercube are the only stable states [12, 13, 16]. However, for the Linsker's network, since there are no such restrictions on the property of the second item $h_i(\omega_\tau, k_1, k_2)$ on the right hand side of equation (6), there indeed exist unsaturated fixed point attractors in equation (6) (see Lemma 2). There exists, even, a two-state limit cycle in difference equation (6) for any two adjacent layers, due to the complicated relationship among system parameters (see Corollary 3 in Section 4.3). One can easily verify that the dynamics of the original BSB model [1] are a special case of equation (5) if one replaces the weight variable ω with the state variable, the covariance matrix with the weight matrix, and let $r(j) = 1$, and $k_1 = k_2 = 0$; hence, the general results presented in Sections 2 and 3 can be applied to characterize the set of fixed points of this model. An extended BSB model with $k_1 \neq 0$, $k_2 \neq 0$, and a neighboring function $r(\cdot)$ can also be studied using our framework, i.e., the effect of the system parameters on the set of fixed points and the related retrieval dynamics can be precisely captured.

5.3 Dynamics With Limiter Function

Theorems 1 and 2 provide the general framework for the description of the fixed point attractors for any difference equation of the type stated in (6) that uses a limiter function. It is important to notice that there are no restrictions imposed on the second item $h_i(\omega_\tau)$ on the right hand side of equation (6), and on the covariance matrix Q in (4). Depending on the structure of the second item, $h_i(\omega_\tau)$, one can derive explicit conditions for determining the stable parameter regimes for any given fixed point. For example, such a result for the Linsker's network is expressed as a necessary and sufficient condition in Theorem 3. The system parameters in the Linsker's model are k_1, k_2 and the ranges of SDFs. k_1 and k_2 are used as the two major parameters, and the effect of the ranges of SDFs is captured in the two k_1-intercept functions, $d_1(\omega)$ and $d_2(\omega)$. One can then discuss the presence or absence of a desired fixed point ω by calculating $d_1(\omega)$ and $d_2(\omega)$ from different choices of the ranges of SDFs. Since the functions in the necessary and sufficient condition are computable (like our slope and k_1-intercept functions), one is always able to check whether a designated fixed point is stable for a specific set of parameters. Therefore, our theoretical approach and examples are applicable to analyzing other unsupervised self-organization systems for connection evolution [5].

5.4 Intralayer Interaction And Biological Discussion

The Linsker's developmental model studied in this paper represents an important step towards understanding the self-organization mechanism of primary visual system. This model (or any other extant model), clearly, does not incorporate all the factors known to play crucial roles in the biological system, and we briefly discuss the well-known factors and the extent to which they are represented in different models.

In all vertebrate visual systems, the dendritic field and axonal field for most feedforward cells are always far less than that of intralayer interaction. So it is the intralayer interaction that mainly causes correlation in afferent activity. It is well recognized that the afferent correlations are key to the emergence of structured aRFs (e.g., refer to the definitions of $d_1(\omega)$ and $d_2(\omega)$ offered in this paper) and orientation dominance maps. Hence, it is necessary to incorporate the role of lateral intralayer interaction in generating various ordered structures in visual system. We think that in the Linsker's network the particular choices of the relative ranges of SDFs (i.e., progressively increasing ranges) play in part the role of intralayer interaction, by introducing localized afferent correlations.

In fact, recent investigations suggest that the intralayer interconnections play an important role in generating center-surround aRFs in mammalian retina, and the underlying mechanism does not seem to involve self-organization. It has been observed that at birth the main types of ON- and OFF-center ganglion cells in cat have emerged, but they are not yet connected to their preceding retinal synaptic network because the photoreceptors are still proliferating and the bipolar axons are just descending towards the inner plexiform layer where ganglion cells arborize. The ontogenesis of retinal wiring seems to follow a genetically determined mechanisms and not to be very susceptible to modification by later visual experience [39]. In fact, the basic center-surround structure is established as early as at the receptor terminal before the first feedforward synapse. In the cone receptor-horizontal cell circuit, center-surround aRFs develop with the aid of the horizontal cells with broad uniform fields.

On the other hand, the development process of primary visual system in regions other than the retina, appears to become more dependent on self-organization mechanisms discussed in the Linsker's network. At early prenatal stages, the spontaneous random firing of young ganglion cells is crucial to shape the development of the geniculate cells [18]. Genetic code mainly defines the timing course of synaptic plasticity, and the Hebbian synaptic modification rule forms the basis of the organization among elements at neuronal level and gives rise to the emergence of more ordered structure [10, 36, 40]. Biologically, the spontaneous firing of ganglion cells, the correlated patterns of activity in the afferents, the lateral intralayer interaction, the feedforward synaptic arbor density function, and parallel ON-, OFF- arrays are all thought to contribute to the symmetry-breaking mechanism by which oriented aRFs and orientation dominance maps can arise. A considerable number of analyses and simulations [3, 4, 8, 9, 19, 28, 29, 30, 31, 32, 33, 38, 41] on the self-organization in visual system have recently expanded Linsker's ideas by introducing more biologically-motivated factors, and in particular, by including lateral intralayer interactions. Our approach can be adapted to explicitly analyze the role played by the lateral intralayer interaction, in the same way that we deal with the interlayer connection density (i.e., SDFs) [34].

Furthermore, in the context of neuroscience, the theoretical results for the networks we have studied not only explain the dynamical mechanisms underlying the early development of primary cortex, but also suggest and support a new understanding of the development process [35]. The various aRFs emerge not because some special functions of the brain (e.g., the edge detection and the orientation detection) demand them, but as a consequence of the microcosmic rule of synaptic formation and modification. Instead of using a single feature detection neuron, the network works as a whole system whose computational function has been clarified as maximization of information preservation [22].

References

[1] Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, **84**, 413-451.

[2] Cohen, M. A., Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-13**, 815-826.

[3] Erwin, E., Obermayer, K., & Schulten, K. (1992). Self-organization maps: stationary states, metastability and convergence rate. *Biol. Cybern.*, **67**, 35-45.

[4] Erwin, E., Obermayer, K., & Schulten, K. (1992). Self-organization maps: ordering, convergence properties and energy functions. *Biol. Cybern.*, **67**, 47-55.

[5] Erwin, E., Obermayer, K., & Schulten, K. (1994). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. Technical Report UIUC-BI-TB-94-08, The Beckman Institute, University of Illinois.

[6] Feng, J., & Pan, H. (1993). Analysis of Linsker-type Hebbian learning: Rigorous results. *Proc. 1993 IEEE Int. Conf. on Neural Networks - San Francisco Vol. III* (pp. 1516-1521). Piscataway, NJ: IEEE.

[7] Feng, J., Pan, H., & Roychowdhury, V.P. (1995). A rigorous analysis of Linsker-type Hebbian learning. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7*. Cambridge MA: MIT Press.

[8] Foldiak, P. (1989). Adaptive network for optimal linear feature extraction. *Proc. 1989 Int. Joint Conf. on Neural Networks - Washington D.C. Vol. I* (pp. 401-405). Piscataway, NJ: IEEE.

[9] Foldiak, P. (1990). Forming sparse representation by local anti-Hebbian learning. *Biol. Cybern.*, **64**, 165-170.

[10] Fregnac, Y., & Imbert, M. (1984). Development of neuronal selectivity in primary visual cortex of cat. *Physiol. Rev.*, **64**(1), 325-434.

[11] Freidlin, M. I., & Wentzell, A. D. (1984). *Random Perturbations of Dynamical System.* Berlin: Springer-Verlag.

[12] Golden, R. M. (1986). The "Brain-State-in-a-Box" neural model is a gradient descent algorithm. *Journal of Mathematical Psychology*, **30**, 73-80.

[13] Greeenberg, H. J. (1988). Equilibria of the Brain-State-in-a-Box (BSB) neural model. *Neural Networks*, **1**, 323-324.

[14] Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures, *Neural Networks*, **1**, 17-61.

[15] Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.*, **81**, 3088-3092.

[16] Hui, S., & Zak, S. H. (1992). Dynamical analysis of the Brain-State-in-a-Box (BSB) neural models. *IEEE Transactions on Neural Networks*, **3**(1), 86-94.

[17] Hurewicz, W., & Wallman, H. (1948). *Dimension Theory.* Princeton: Princeton University Press.

[18] Kalil, R.E. (1989). Synaptic formation in the developing brain. *Scientific American*, **261**(6), 38-45.

[19] Leen, T.K. (1991). Dynamics of learning in linear feature-discovery networks. *Network*, **2**, 85-105.

[20] Linsker, R. (1986). From basic network principle to neural architecture (series). *Proc. Natl. Acad. Sci. USA*, **83**, 7508-7512, 8390-8394, 8779-8783.

[21] Linsker, R. (1988a). Development of feature-analyzing cells and their columnar organization in a layered self-adaptive network. In R. Cotterill (Ed.), *Computer Simulation in Brain Science* (pp. 416-431). Cambridge: Cambridge University Press.

[22] Linsker, R. (1988b). Self-organization in a perceptual network. *Computer*, **21**(3), 105-117.

[23] Linsker, R. (1989). An application of the principle of maximum information preservation to linear systems. In D. Touretzky (Ed.), *Advances in Neural Information Processing Systems I* (pp. 186-194). San Mateo, CA: Morgan Kaufmann.

[24] Linsker, R. (1990). Perceptual neural organization: Some approaches based on network models and information theory. *Annu. Rev. Neurosci.*, **13**, 257-281.

[25] MacKay, D., & Miller, K. (1990a). Analysis of Linsker's application of Hebbian rules to linear networks. *Network*, **1**, 257-297.

[26] MacKay, D., & Miller, K. (1990b). Analysis of Linsker's simulations of Hebbian rules. *Neural Computation*, **2**, 173-187.

[27] von der Malsberg, C. (1995). Network self-organization in the ontogenesis of the mammalian visual system. In S.F. Zornetzer, J. Davis, & C. Lau (Eds.), *An Introduction to Neural and Electronic Networks, Second Edition.* New York: Academic Press.

[28] Miller, K.D. (1990a). Correlation-based mechanisms of neural development. In M.A. Gluck, & D.E. Rumelhart (Eds.), *Neuroscience and Connectionist Theory* (pp. 267-353). Hillsdale, NJ: Lawrence Erlbaum.

[29] Miller, K.D. (1990b). Derivation of linear Hebbian equations from a nonlinear Hebbian model of synaptic plasticity. *Neural Computation*, **2**, 321-333.

[30] 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. Neuroscience*, **14**(1), 409-441.

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

[32] Miller,K.D., & Stryker,M.P. (1989). The development of ocular dominance columns: Mechanisms and models. In S.J. Hanson, & C.R. Olson (Eds.), *Connectionist Modeling and Brain Function: The Developing Interfaces* (pp. 255-350). Cambridge, MA: MIT Press.

[33] Obermayer, K., Blasdel, G.G., & Schulten, K. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Phys. Rev. A*, **45** (10), 7568-7589.

[34] Pan, H., Roychowdhury, V.P., & Feng, J. (1995). On correlation-based cortical map models: Results of a nonlinear analysis. In preparation.

[35] Purves, D., Riddle, D.R., & LaMantia, A.-S. (1992). Iterated patterns of brain circuitry (or how the cortex gets its spots). *Trends in Neuroscience*, **15**(10), 362-368.

[36] Rauschecker, J.P. (1991). Mechanisms of visual plasticity: Hebb synapses, NMDA receptors, and beyond. *Physiol. Rev.*, **71**(2), 587-615.

[37] Seneta, E. (1973). *Non-Negative Matrices.* London: George Allen & Unwin.

[38] Sirosh, J., & Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biol. Cybern.*, **71**, 66-78.

[39] Sterling, P. (1990). Retina. In G.M. Shepherd (Ed.), *The Synaptic Organizations of the Brain* (pp. 170-213). New York: Oxford University Press.

[40] Sur, M., Pallas, S.L., & Roe, A.W. (1990). Cross-modal plasticity in cortical development: Differentiation and specification of sensory neocortex. *Trends Neurosci.*, **13**(6), 227-233.

[41] Tanaka, S. (1990). Theory of self-organization of cortical maps: Mathematical framework. *Neural Networks*, **3**, 625-640.

[42] Tang, D.S. (1989). Information-theoretic solutions to early visual information processing: Analytic results. *Phys. Rev. A*, **40**, 6626-6635.

Appendix A: On the Continuous Version of the Linsker's Model

It is well known that the Cohen-Grossberg model specified by the following equations [2, 14]:

$$\frac{dx_t(i)}{dt} = a_i(x_t(i))[b_i(x_t(i)) - \sum_{k=1}^{N} c_{ik} d_k(x_t(k))], \qquad i = 1, ..., N, \tag{25}$$

is a gradient system if the matrix $C = \{c_{ij},\ 1 \le i, j \le N\}$ is symmetric, the amplification function $a_i(\cdot)$ and the self-signal function $b_i(\cdot)$ are continuous, $a_i(\cdot) \ge 0$, and the other-signal function $d_i(x_t(i))$ is differentiable and monotone nondecreasing. The competitive network (25) can be written in the form

$$\frac{dx_t(i)}{dt} = -A(x_t) \ \ \text{grad}\ V(x_t),$$

where the matrix $A(x_t) = \{A_{ij}(x_t),\ 1 \le i, j \le N\}$ is chosen to satisfy

$$A_{ij}(x_t) = \frac{a_i(x_t(i))\delta_{ij}}{d_i'(x_t(i))} \qquad ('= \frac{d}{dt}),$$

and the function

$$V(x_t) = -\sum_{i=1}^{N} \int_0^{x_t(i)} b_i(s) d_i'(s) ds + \frac{1}{2} \sum_{j,k=1}^{N} c_{jk} d_j(x_t(j)) d_k(x_t(k)) \tag{26}$$

is a global Liapunov function for (25) since

$$\frac{dV(x_t)}{dt} = -\sum_{i=1}^{N} a_i(x_t(i)) d_i'(x_t(i))[b_i(x_t(i)) - \sum_{k=1}^{N} c_{ik} d_k(x_t(k))]^2 \le 0,$$

where $\delta_{ij} = 1$ if $i = j$; and $= 0$ if $i \ne j$.

In fact, the equation (6) represented in discrete time can be described in continuous time as

$$\frac{d\omega_t(i)}{dt} = -\omega_t(i) + f\{k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2 + \frac{\delta_{ij}}{r(i)}] r(j)\omega_t(j)\}, \tag{27}$$

in the sense that both versions have the same set of fixed point attractors. The existence and uniqueness of the solution for the above equation (27) is ensured since $f(\cdot)$ satisfies the Lipschitz condition. Similar to the type of transformation from the Brain-State-in-a-Box model to the Cohen-Grossberg form (25), let $B_{ij} = Q_{ij}^{\mathcal{L}} + k_2 + \frac{\delta_{ij}}{r(i)}$, and $h(x) = f(x + k_1)$, then the equation (27) becomes

$$\frac{d\omega_t(i)}{dt} = -\omega_t(i) + h\{\sum_{j=1}^{N_{\mathcal{L}}} B_{ij} r(j)\omega_t(j)\}.$$

Let $y_t(i) = \sum_{j=1}^{N_{\mathcal{L}}} B_{ij} r(j)\omega_t(j)$, by taking the summation $\sum_{j=1}^{N_{\mathcal{L}}} B_{ij} r(j)$ on both sides of the above equation, we have

$$\frac{dy_t(i)}{dt} = -y_t(i) + \sum_{j=1}^{N_{\mathcal{L}}} B_{ij} r(j) h(y_t(j)).$$

If we choose

$$a_i(y_t(i)) = 1,$$

$$b_i(y_t(i)) = -y_t(i),$$

$$c_{ij} = -B_{ij},$$

$$d_j(y_t(j)) = r(j)h(y_t(j)),$$

then the continuous time version of the Linsker's network (27) is transformed into the Cohen-Grossberg model (25), and there exists a Liapunov function for it because

$$a_i(y_t(i)) > 0,$$

$$c_{ij} = -B_{ij} = -B_{ji} = c_{ji},$$

so $C = \{c_{ij},\ 1 \leq i, j \leq N_{\mathcal{L}}\}$ is symmetric, and

$$d_j'(y_t(j)) = r(j)(D^+ + D_-)[h(y_t(j))] \geq 0,$$

where D^+ and D_- are the Dini derivatives and $r(\cdot)$ is non-negative according to its definition. Therefore there is no limit cycle for the continuous time case.

It is noticed that the relationship between the stable states of the discrete time and continuous time versions of the Linsker's model is not exactly the same as that in Hopfield model [15] because of the specific form of $f(\cdot)$. We have shown (see Corollary 3 in Section 4.3) that there exist a limit cycle in (6) within a certain parameter regime, so that the discrete time version of the Linsker's network is not an exact gradient system at all. Meanwhile, for the continuous time case, we have that the limit $\omega = \lim\limits_{\tau \to \infty} \omega_\tau$ does exist. Hence we only need to assume that

$$\omega \in V(\Omega) = \{-\omega_{max}, \omega_{max}\}^{N_{\mathcal{L}}},$$

and

$$\forall \omega \in V(\Omega), \qquad k_1 + \sum_{j=1}^{N_{\mathcal{L}}} [Q_{ij}^{\mathcal{L}} + k_2] r(j) \omega(j) \neq 0, \qquad \text{for } i = 1, \cdots, N_{\mathcal{L}},$$

then it will be easily confirmed that all conclusions from Theorem 1 to 5 for the discrete time case do hold for the continuous time case.

Appendix B: Examples of Structured aRFs between Layers $\mathcal{B}$ and $\mathcal{C}$ of the Linsker's Network

We use a grid system in our examples. We assume that each cell in layer $\mathcal{C}$ receives synaptic inputs from 253 sites in layer $\mathcal{B}$, where 253 is the total number of sites inside the circle with grid radius 9 (Fig. 1). We also assume that the SDFs $r^{\mathcal{AB}}(\cdot,\cdot)$ and $r^{\mathcal{BC}}(\cdot)$ are Gaussian functions with ranges $r_{\mathcal{B}}$,$r_{\mathcal{C}}$. In the following, we explore the stable parameter regimes for different classes of aRFs (see Fig. 11). For any aRF, ω, the existence of a stable parameter regime is determined by the two k_1-intercept functions, $d_1(\omega, r_{\mathcal{B}}, r_{\mathcal{C}})$ and $d_2(\omega, r_{\mathcal{B}}, r_{\mathcal{C}})$. In our calculations, we fix $r_{\mathcal{C}}$ to three different values, and for each $r_{\mathcal{C}}$, we calculate $d_1(\omega, r_{\mathcal{B}})$ and $d_2(\omega, r_{\mathcal{B}})$. An aRF is a stable attractor if and only if $d_2(\omega, r_{\mathcal{B}}) > d_1(\omega, r_{\mathcal{B}})$. We shall call the turning point of $r_{\mathcal{B}}$ that turns $d_2(\omega, r_{\mathcal{B}}) - d_1(\omega, r_{\mathcal{B}})$ from positive to negative as the *critical value* of $r_{\mathcal{B}}$, and denote it as $r_{\mathcal{B}}^{critical}$.

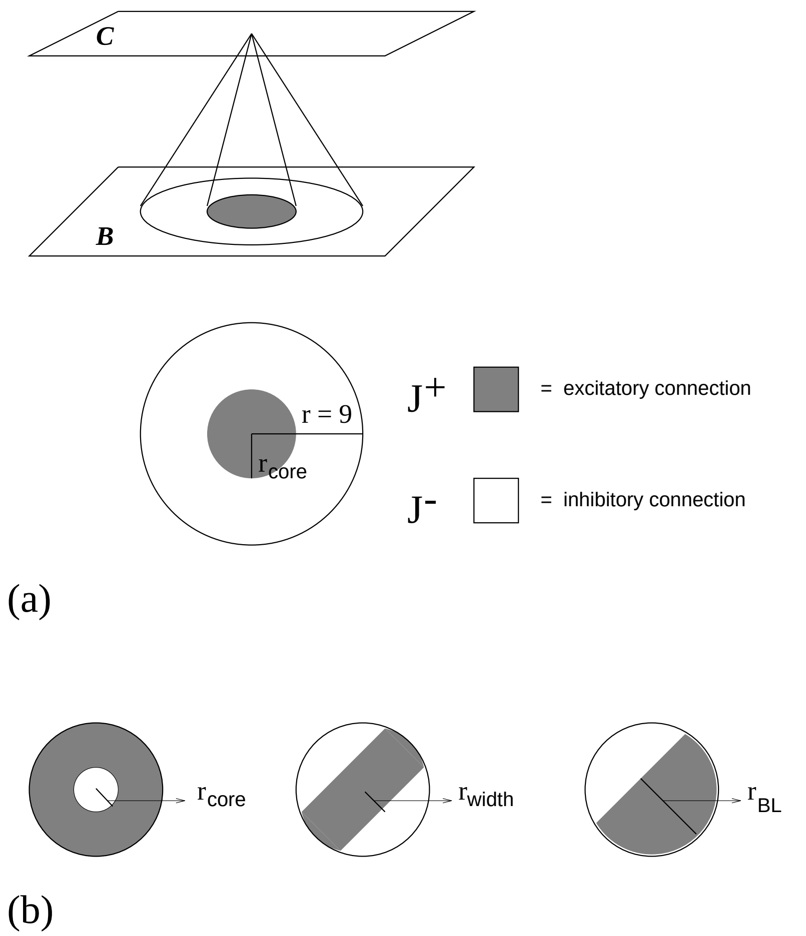


Figure 11: **(a)** An ON-center aRF between layers $\mathcal{B}$ and $\mathcal{C}$. The synaptic strengths between a $\mathcal{C}$-cell and a number of $\mathcal{B}$-cells within the circle with radius r_{core} are excitatory, and those outside the circle are all inhibitory. **(b)** Other kinds of aRFs (from left to right): OFF-center cell with radius r_{core}, oriented cell with radius r_{width},and bi-lobed cell with radius r_{BL}.

Example 1: *ON-center cells* $ON(r_{core}, 9)$.

Denote the radius of the excitatory central region of an ON-center aRF as r_{core} and denote an ON-center aRF simply as $ON(r_{core}, 9)$ (see Fig. 11).

Fixing $r_{\mathcal{C}} = 2.5$, 10, and 100 respectively, we determine the parameter regimes of $(k_1, k_2, r_{\mathcal{B}})$, in which ON-center aRFs are developed (Fig. 12(a)-(h) only shows the case when $r_{\mathcal{C}} = 10$).

(a) $r_{core} = 1$. In this case, we have 5 excitatory sites and 248 inhibitory sites with the slope $c(ON(1,9)) = -0.551$ when $r_{\mathcal{C}} = 2.5$; $c(ON(1,9)) = -0.948$ when $r_{\mathcal{C}} = 10$; $c(ON(1,9)) = -0.960$ when $r_{\mathcal{C}} = 100$. In Figure 12(a), the parameter regime of (k_1, k_2) is given for the occurrence of center-surround cell $ON(1,9)$ with different $r_{\mathcal{B}}$ but fixed $r_{\mathcal{C}} = 10$. It shows the two k_1-intercept functions $d_1(ON(1,9), r_{\mathcal{B}})$ and $d_2(ON(1,9), r_{\mathcal{B}})$ as the functions of $r_{\mathcal{B}}$. Note that as $r_{\mathcal{B}} \geq 1.0674$, $d_1(ON(1,9)) \geq d_2(ON(1,9))$, hence $(k_1, k_2) = \phi$, the empty set; namely, the connection pattern $ON(1,9)$ will no longer be an attractor of equation (6). Thus, for $ON(1,9)$, $r_{\mathcal{B}}^{critical} = 1.0674$.

(b)-(d) $r_{core} = 2, 3, 4$ in Figure 12(b)-(d). In these cases, we have $c(ON(2,9)) = -0.854$, $c(ON(3,9)) = -0.681$, and $c(ON(4,9)) = -0.478$ with $r_{\mathcal{C}} = 10$ respectively. Note that as $r_{\mathcal{B}} \geq 1.4634$ for $ON(2,9)$, $r_{\mathcal{B}} \geq 2.9485$ for $ON(3,9)$, and $r_{\mathcal{B}} \geq 5.0724$ for $ON(4,9)$, $(k_1, k_2) = \phi$, namely, those aRFs will no longer be attractors of equation (6).

(e) $r_{core} = 5$. In this case, we have 81 excitatory sites and 172 inhibitory sites with $c(ON(5,9)) = -0.179$ with $r_{\mathcal{C}} = 10$. The difference of Figure 12(e) from the above cases is that a stable regime of (k_1, k_2) always exists (i.e., there is no critical value for $r_{\mathcal{B}}$).

(f)-(h) $r_{core} = 6, 7, 8$ in Figure 12(f)-(h). In these cases, we have $c(ON(6,9)) = 0.092, c(ON(7,9)) = 0.366, c(ON(8,9)) = 0.685$ with $r_{\mathcal{C}} = 10$ respectively. These aRFs will always be attractors like when $r_{core} = 5$.

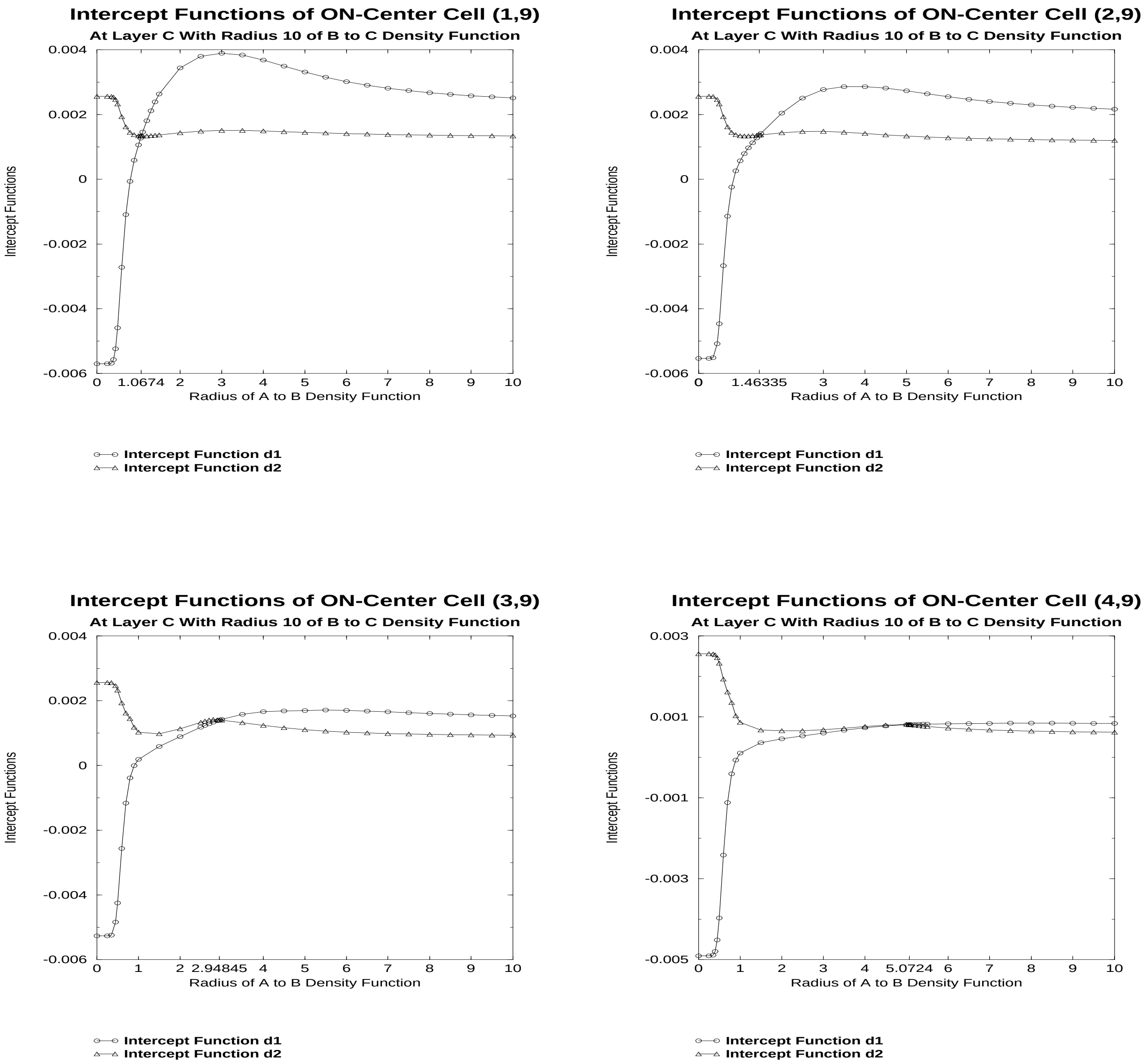


Figure 12: Fixing $r_{\mathcal{C}} = 10$, the parameter regime of $(k_1, k_2, r_{\mathcal{B}})$ is determined by the two k_1-intercept functions $d_1(ON(r_{core}, 9))$ and $d_2(ON(r_{core}, 9))$ (which are functions only of $r_{\mathcal{B}}$ in each plot). Note the critical values of $r_{\mathcal{B}}$ in (a)-(d), where $d_2(ON(r_{core}, 9)) - d_1(ON(r_{core}, 9))$ turns from positive to negative. When $r_{\mathcal{B}}$ is larger than the critical value for an aRF, then the corresponding aRF will no longer be an attractor of equation (6). **(a)** $ON(1, 9)$. **(b)** $ON(2, 9)$. **(c)** $ON(3, 9)$. **(d)** $ON(4, 9)$.

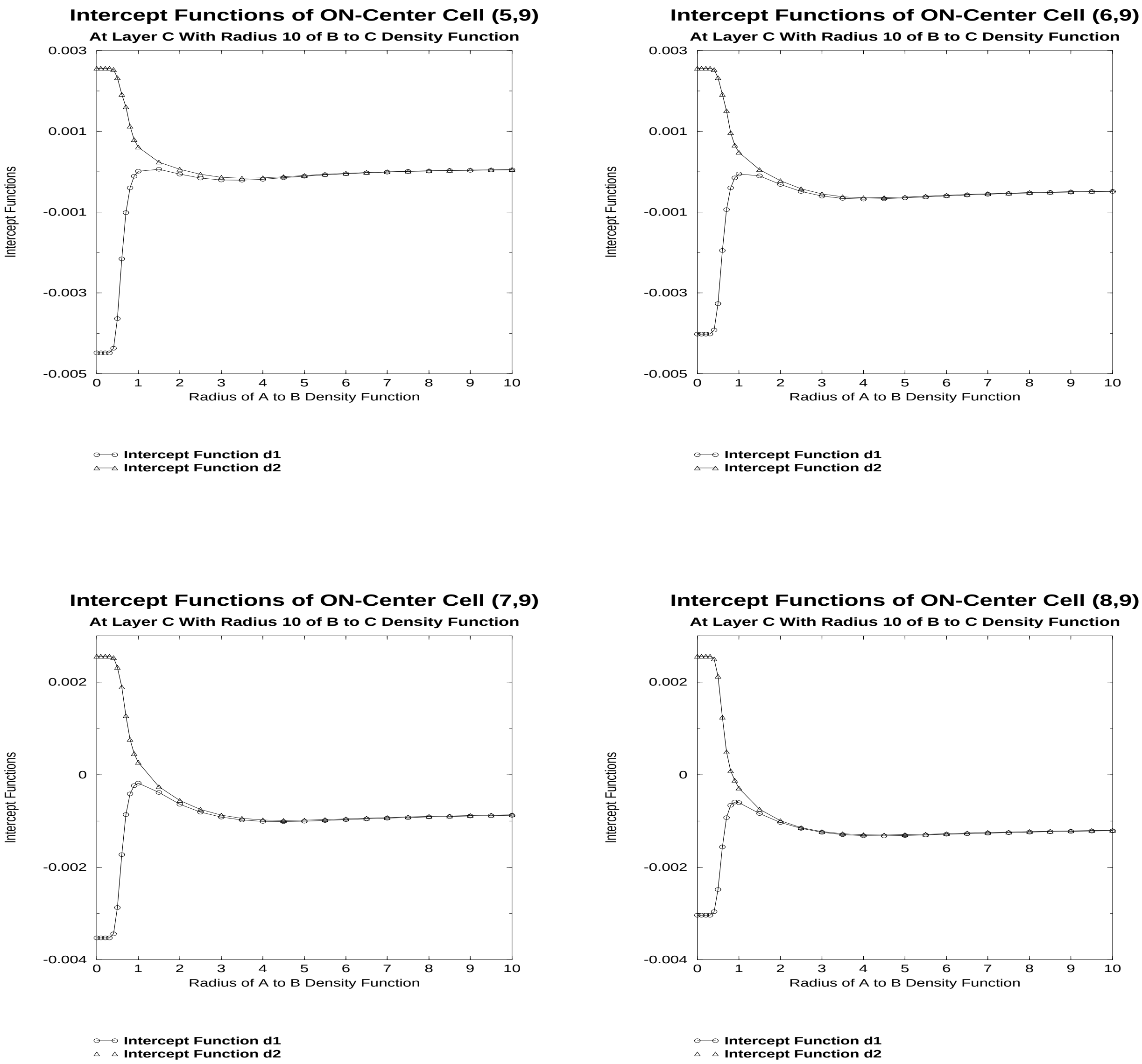


Figure 12 continued. **(e)** $ON(5,9)$. **(f)** $ON(6,9)$. **(g)** $ON(7,9)$. **(h)** $ON(8,9)$.

We summarize the above results about $ON(r_{core}, 9)$ in Table 2, and represent the critical values of $r_{\mathcal{B}}$ as a function of $c(ON(r_{core}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 in Figure 13.

Table 2: $ON(r_{core}, 9)$

			$r_{\mathcal{C}} = 2.5$		$r_{\mathcal{C}} = 10$		$r_{\mathcal{C}} = 100$	
r_{core}	$\oplus$ Sites	$\ominus$ Sites	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$
1	5	248	-0.551	1.0867	-0.948	1.0674	-0.960	1.3584
2	13	240	-0.040	–	-0.854	1.4634	-0.897	1.9335
3	29	224	0.536	–	-0.681	2.9485	-0.770	2.9442
4	49	204	0.834	–	-0.478	5.0724	-0.611	4.1820
5	81	172	0.966	–	-0.179	–	-0.358	9.1725
6	113	140	0.994	–	0.092	–	-0.105	–
7	149	104	0.999	–	0.366	–	0.180	–
8	197	56	1.000	–	0.685	–	0.559	–

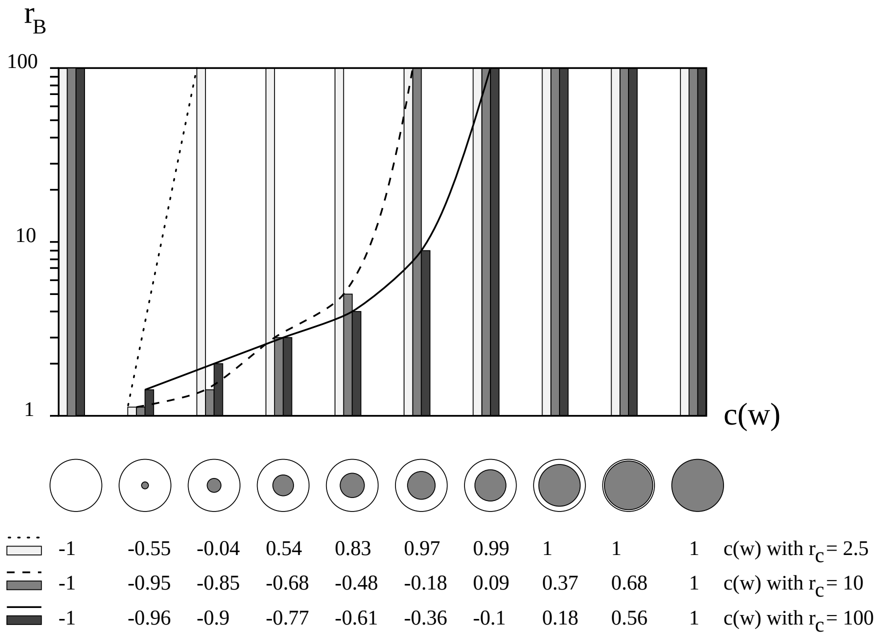


Figure 13: The critical values of $r_{\mathcal{B}}$ as a function of $c(ON(r_{core}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 respectively.

Summary of Example 1: We have shown that various aRFs as attractors have different relative stability. For a fixed $r_{\mathcal{C}}$ (the range of the SDF between layers $\mathcal{B}$ and $\mathcal{C}$), the range $r_{\mathcal{B}}$ of the SDF between layers $\mathcal{A}$ and $\mathcal{B}$ (as the third system parameter), has various critical values for different attractors. For circularly symmetric ON-center cells, those aRFs with large ON-center core (which have positive or small negative slope value $c(\omega) \approx -\frac{k_1}{k_2}$) always have a stable parameter regime. But for those ON-center cells with large negative slope value $c(\omega)$, their stable parameter regimes decrease in size with $c(\omega)$.

Similarly, circularly symmetric OFF-center cells with large OFF-center core (which have negative or small positive slope value $c(\omega)$) will be more stable than those with large positive average of weights as we shall learn from the next example. □

Example 2: *OFF-center cells* $OFF(r_{core}, 9)$.

We summarize the results about $OFF(r_{core}, 9)$ in Table 3, and represent the critical values of $r_{\mathcal{B}}$ as a function of $c(OFF(r_{core}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$,10, and 100 in Figure 14.

Table 3: $OFF(r_{core}, 9)$

r_{core}	⊖ Sites	⊕ Sites	$r_{\mathcal{C}} = 2.5$		$r_{\mathcal{C}} = 10$		$r_{\mathcal{C}} = 100$	
			$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$
1	5	248	0.551	1.0867	0.948	1.0674	0.960	1.3584
2	13	240	0.040	–	0.854	1.4634	0.897	1.9335
3	29	224	-0.536	–	0.681	2.9485	0.770	2.9442
4	49	204	-0.834	–	0.478	5.0724	0.611	4.1820
5	81	172	-0.966	–	0.179	–	0.358	9.1725
6	113	140	-0.994	–	-0.092	–	0.105	–
7	149	104	-0.999	–	-0.366	–	-0.180	–
8	197	56	-1.000	–	-0.685	–	-0.559	–

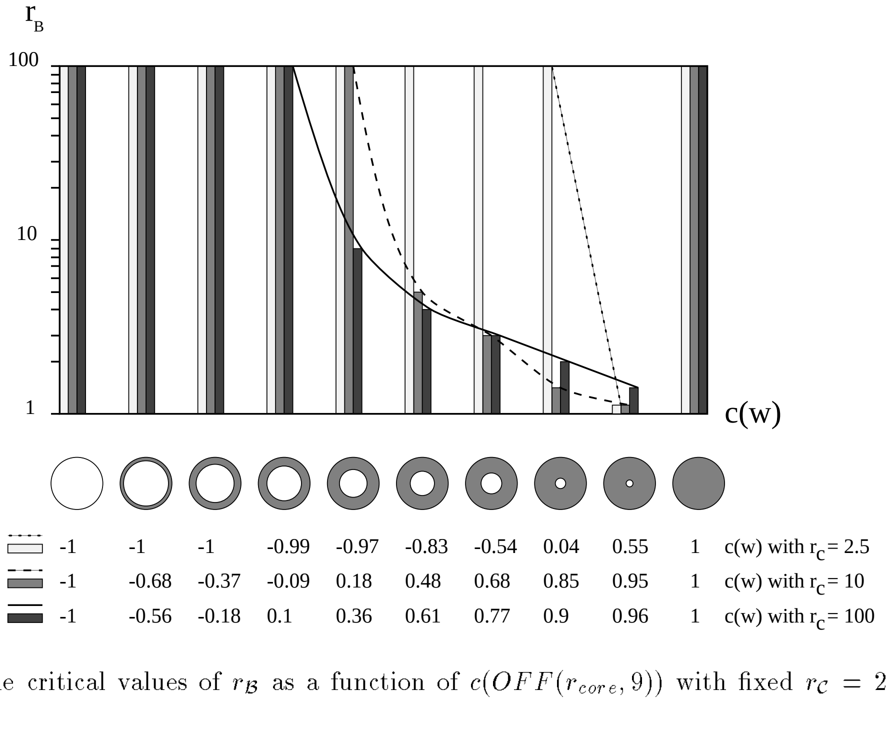


Figure 14: The critical values of $r_{\mathcal{B}}$ as a function of $c(OFF(r_{core}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 respectively.

Relationship between ON-center and OFF-center aRFs: For a given $r_{\mathcal{B}}$, $c(\omega)$ ranges from -1 to 1, when the number of excitatory connections in an aRF, $|J^{+}(\omega)|$, changes from 0 to $N_{\mathcal{B}}$ and the number of its inhibitory connections, $|J^{-}(\omega)|$, changes from $N_{\mathcal{B}}$ to 0. Recall that for every ON-center cell, ω, with excitatory connections inside the circle with radius r_{core} and inhibitory connections outside that circle, there exists a corresponding OFF-center cell, ω', with inhibitory connections inside the circle with the same radius r_{core} and excitatory connections outside that circle. Hence, from the definition of $c(\omega)$ (which is a measure of the average synaptic strength for the pattern), we observe that

$$c(\omega') = \sum_{j \in J^+(\omega')} r(j) - \sum_{j \in J^-(\omega')} r(j) = -c(\omega).$$

Similarly for d_1 and d_2, we have

$$d_1(\omega') = -d_2(\omega),$$

and

$$d_2(\omega') = -d_1(\omega).$$

Therefore, we have

$$d_2(\omega') - d_1(\omega') = d_2(\omega) - d_1(\omega).$$

That is, if there exists a stable parameter regime for ω, then so does for ω' (refer to Fig. 5). Thus, we only need to consider ON-center cells because of the symmetry between the slope and intercept functions of OFF-center and ON-center cells.

Now it is clear from our results (see Tables 2 and 3) that if $c(\omega) > 0$ (i.e., an ON-center connection pattern with a large central excitatory region or an OFF-center with a small inhibitory central region), then the ON-center cells will always have stable parameter regimes (i.e., $(k_1, k_2, r_{\mathcal{B}}, r_{\mathcal{C}})$), whereas the OFF-center cells can be made unstable. It is the same the other way round, i.e., if $c(\omega) < 0$, the OFF-center cell with large inhibitory r_{core} is less sensitive to parameters $(r_{\mathcal{B}}, r_{\mathcal{C}})$ than the ON-center cell with small excitatory r_{core}. This conclusion makes it clear why the parameter regime $(k_1 > 0, k_2 < 0)$ favors the occurrence of the ON-center cell with large r_{core}, although there also exists the possibility of the emergence of the OFF-center cell with small r_{core}. Similarly, the parameter regime $(k_1 < 0, k_2 < 0)$ prefers the OFF-center aRF with large r_{core} to the ON-center cell with small r_{core}. □

Example 3: *Oriented cells* $OR(r_{width}, 9)$.

Denote a half of the width of the excitatory central strip in an oriented aRF as r_{width} and denote this aRF simply as $OR(r_{width}, 9)$ (Fig. 11).

Fixing $r_{\mathcal{C}} = 2.5$, 10, and 100 respectively, we determine the parameter regimes of $(k_1, k_2, r_{\mathcal{B}})$ by varying r_{width} from 0.36 to 7.56. $d_1(OR(r_{width}, 9), r_{\mathcal{B}})$ and $d_2(OR(r_{width}, 9), r_{\mathcal{B}})$ and the critical values of $r_{\mathcal{B}}$ are shown in Figure 15 for the case when $r_{\mathcal{C}} = 10$.

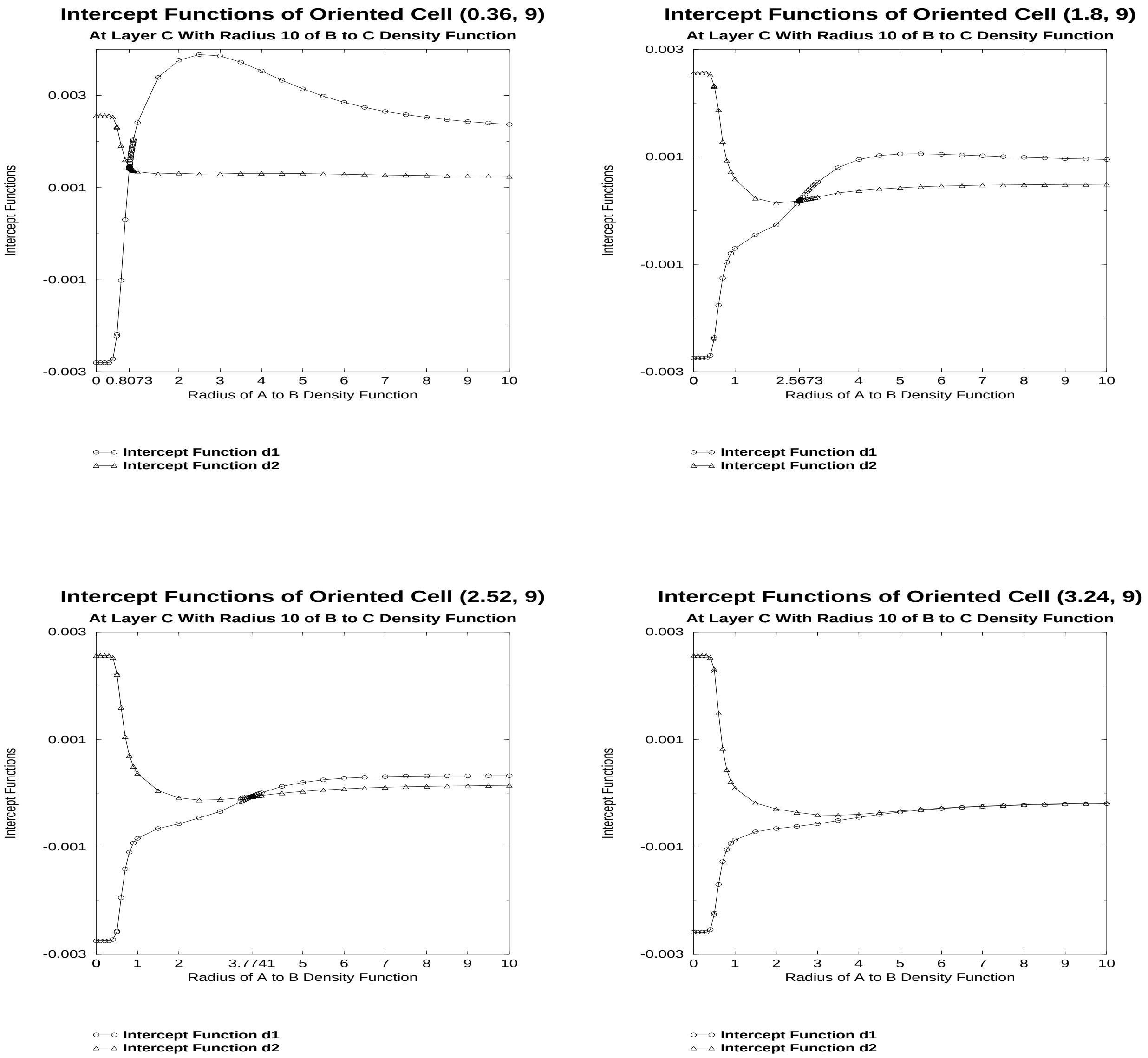


Figure 15: Fixing $r_{\mathcal{C}} = 10$, the parameter regime of $(k_1, k_2, r_{\mathcal{B}})$ is represented by two k_1-intercept functions $d_1(OR(r_{width}, 9))$ and $d_2(OR(r_{width}, 9))$ as the functions of $r_{\mathcal{B}}$. For every case, except in (d), there exists a critical value of $r_{\mathcal{B}}$ that turns $d_2(OR(r_{width}, 9)) - d_1(OR(r_{width}, 9))$ from positive to negative. When $r_{\mathcal{B}}$ is larger than its critical value, the aRF will no longer be an attractor of equation (6). **(a)** $OR(0.36, 9)$. **(b)** $OR(1.8, 9)$. **(c)** $OR(2.52, 9)$. **(d)** $OR(3.24, 9)$.

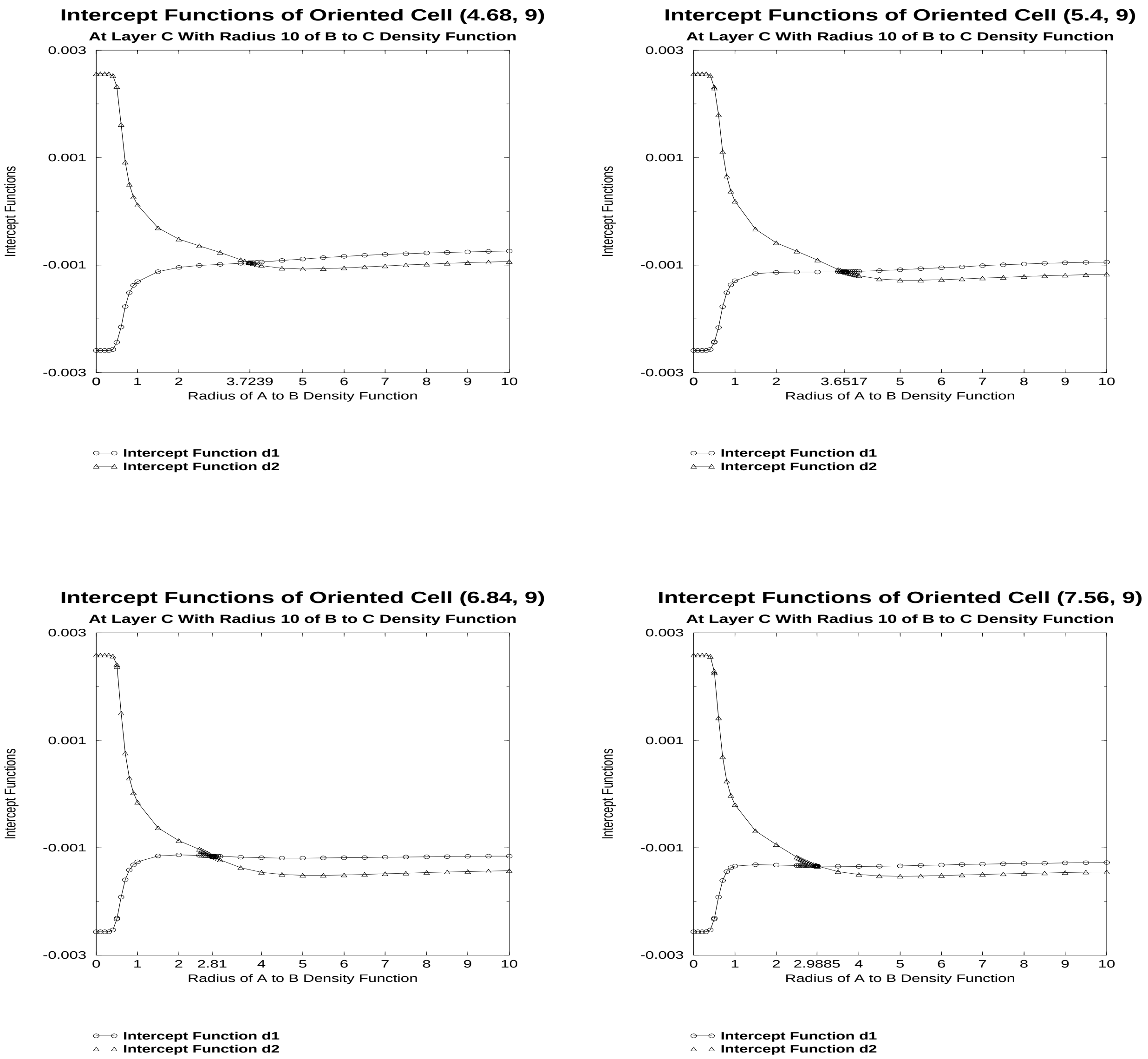


Figure 15 continued. **(e)** $OR(4.68, 9)$. **(f)** $OR(5.4, 9)$. **(g)** $OR(6.84, 9)$. **(h)** $OR(7.56, 9)$.

We summarize the above results about $OR(r_{width}, 9)$ in Table 4, and represent the critical values of $r_{\mathcal{B}}$ as a function of $c(OR(r_{width}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 in Figure 16. Notice that only the attractor with about zero slope value $|c|$ might always have a stable parameter regime. □

Table 4: $OR(r_{width}, 9)$

	$r_{\mathcal{C}} = 2.5$		$r_{\mathcal{C}} = 10$		$r_{\mathcal{C}} = 100$	
r_{width}	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$
0.36	-0.639	0.6050	-0.855	0.8073	-0.875	1.0238
1.80	0.417	0.7661	-0.408	2.5673	-0.479	2.4193
2.52	0.647	0.6738	-0.230	3.7741	-0.318	2.8310
3.24	0.859	0.6152	0.021	—	-0.074	3.2468
4.68	0.979	0.5813	0.360	3.7239	0.255	5.7863
5.40	0.991	0.5701	0.520	3.6517	0.417	5.2447
6.84	0.996	0.5641	0.785	2.8100	0.714	3.9169
7.56	0.996	0.5640	0.879	2.9885	0.832	4.1468

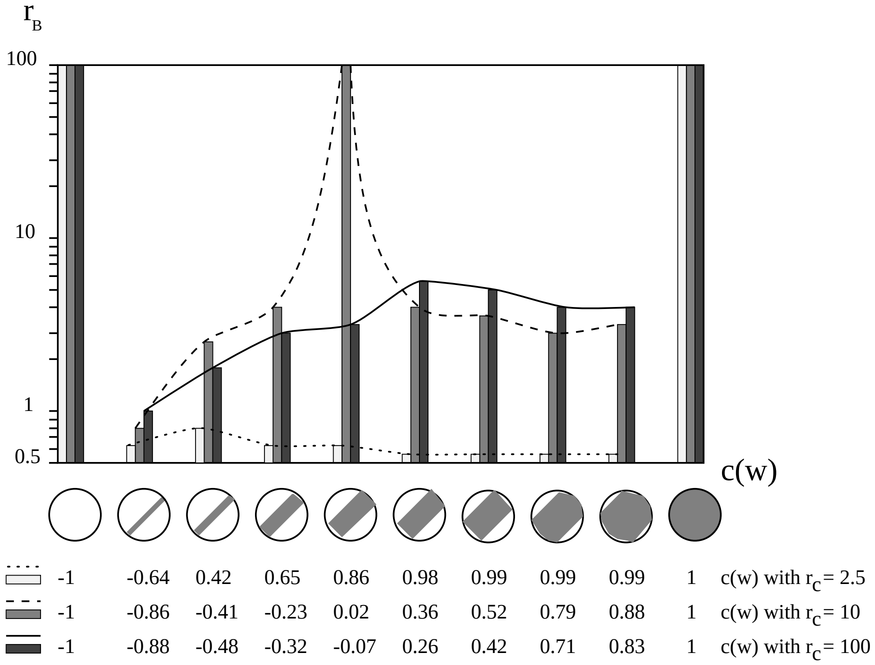


Figure 16: The critical values of $r_{\mathcal{B}}$ as a function of $c(OR(r_{width}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 respectively.

Example 4: *Bi-lobed cells* $BL(r_{BL}, 9)$.

Denote the maximum width of the excitatory area in a bi-lobed aRF as r_{BL} and denote this aRF as $BL(r_{BL}, 9)$ (Fig. 11).

Fixing $r_{\mathcal{C}} = 2.5$, 10, and 100, we determine the parameter regime of $(k_1, k_2, r_{\mathcal{B}})$ by varying r_{BL} from 0.72 to 15.84 similar with other examples. Figure 17 shows the case when $r_{\mathcal{C}} = 10$.

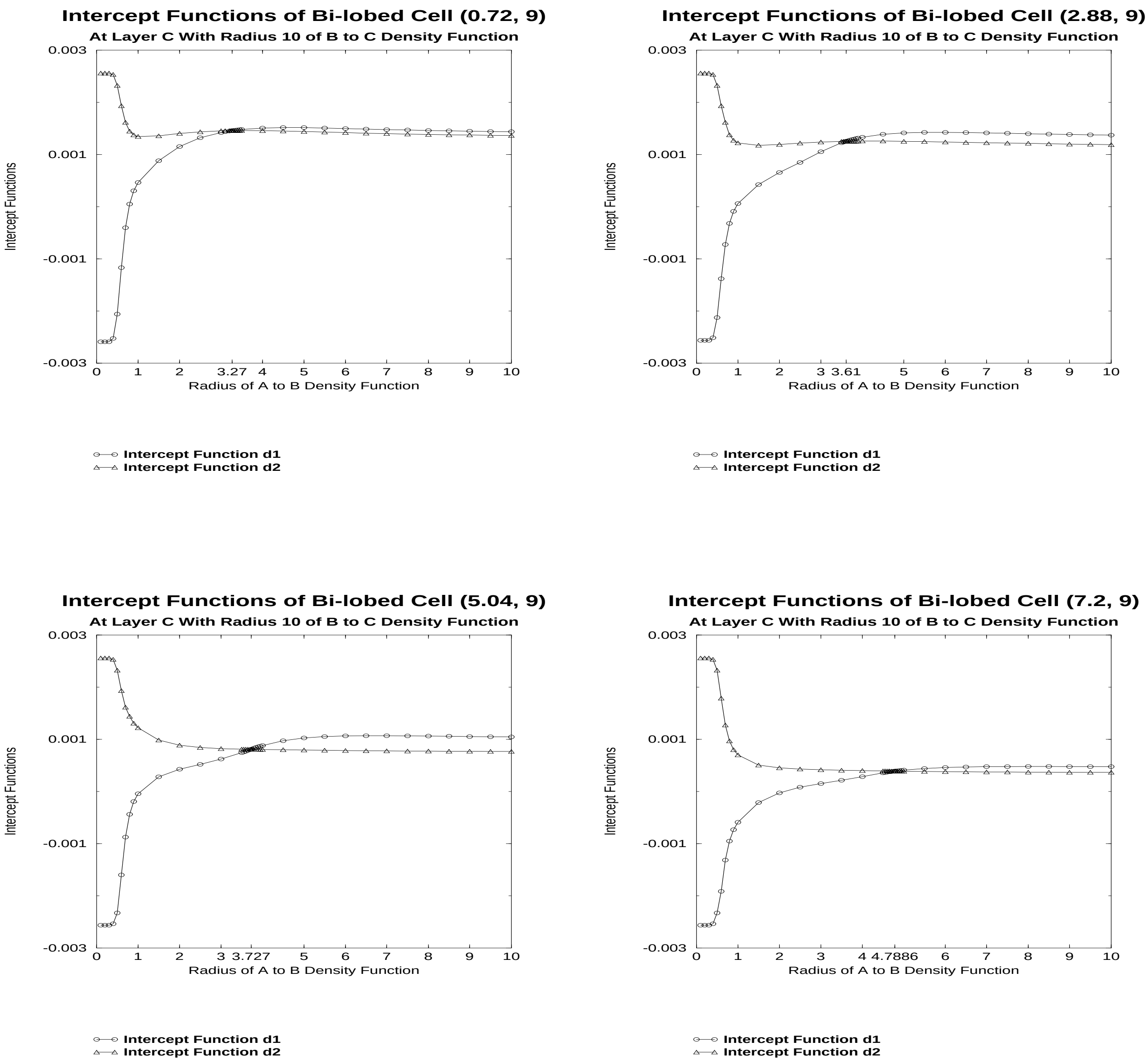


Figure 17: Fixing $r_{\mathcal{C}} = 10$, the parameter regime of $(k_1, k_2, r_{\mathcal{B}})$ is represented by two k_1-intercept functions $d_1(BL(r_{BL}, 9))$ and $d_2(BL(r_{BL}, 9))$ as the functions of $r_{\mathcal{B}}$. For every case, except in (e), there exist a critical value of $r_{\mathcal{B}}$ that turns $d_2(BL(r_{BL}, 9)) - d_1(BL(r_{BL}, 9))$ from positive to negative. When $r_{\mathcal{B}}$ is larger than its critical value, the aRF will no longer be an attractor of equation (6). **(a)** $BL(0.72, 9)$. **(b)** $BL(2.88, 9)$. **(c)** $BL(5.04, 9)$. **(d)** $BL(7.2, 9)$.

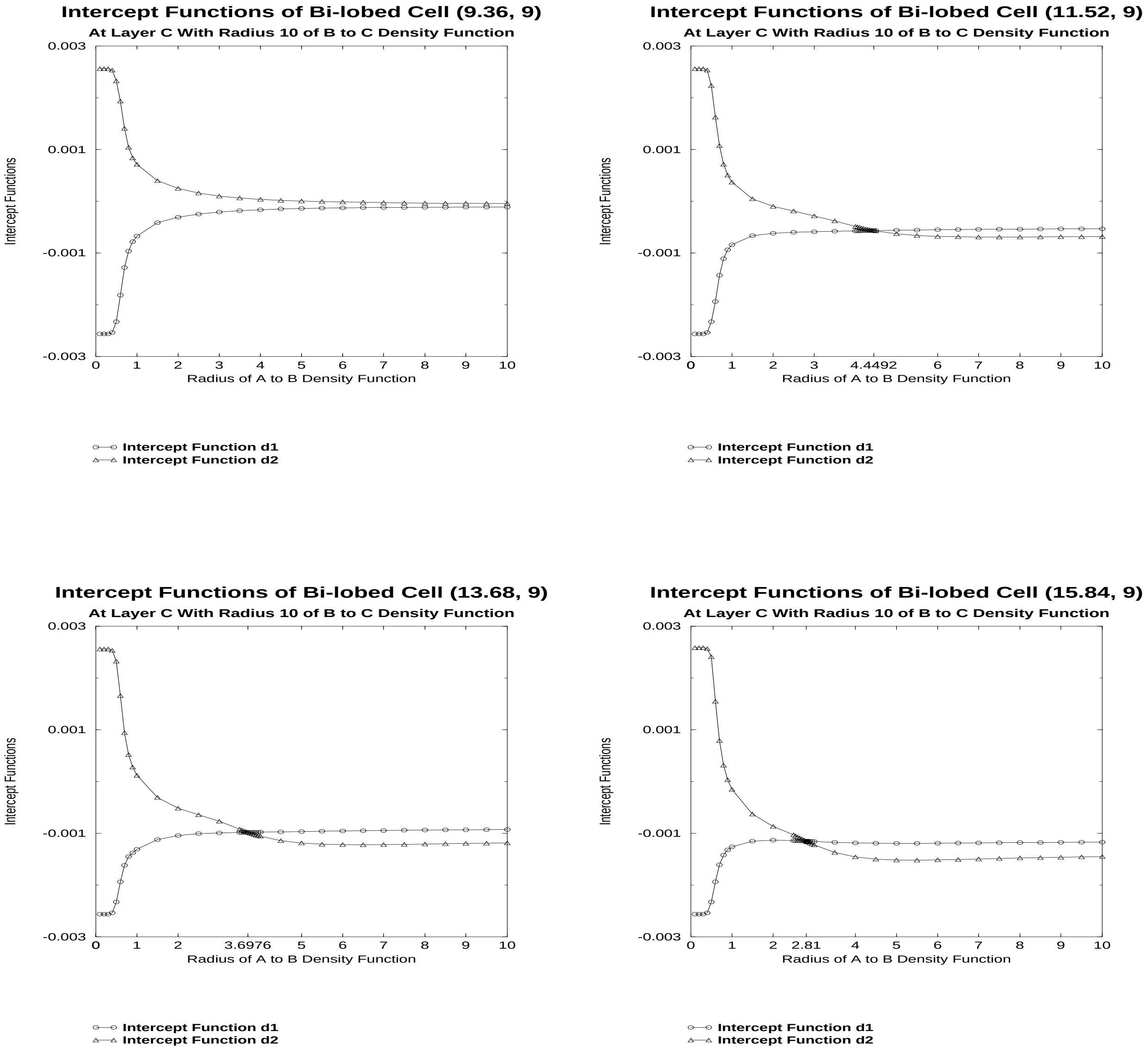


Figure 17 continued. **(e)** $BL(9.36, 9)$. **(f)** $BL(11.52, 9)$. **(g)** $BL(13.68, 9)$. **(h)** $BL(15.84, 9)$.

We summarize our results about $BL(r_{BL}, 9)$ in Table 5, and represent the critical values of $r_{\mathcal{B}}$ as a function of $c(BL(r_{BL}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$, 10, and 100 in Figure 18.

Table 5: $BL(r_{BL}, 9)$

	$r_{\mathcal{C}} = 2.5$		$r_{\mathcal{C}} = 10$		$r_{\mathcal{C}} = 100$	
r_{BL}	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$	$c(\omega)$	$r_{\mathcal{B}}^{critical}$
0.72	-1	0.5837	-0.973	3.2700	-0.961	4.5950
2.88	-0.999	0.5633	-0.826	3.6100	-0.779	5.0250
5.04	-0.973	0.5979	-0.595	3.7270	-0.542	5.4507
7.20	-0.686	0.7609	-0.283	4.7886	-0.249	7.8840
9.36	0.160	1.7739	0.058	–	0.051	–
11.52	0.831	0.6592	0.389	4.4492	0.344	6.8453
13.68	0.991	0.5790	0.683	3.6976	0.629	5.3186
15.84	1	0.5641	0.892	2.8100	0.858	3.9165

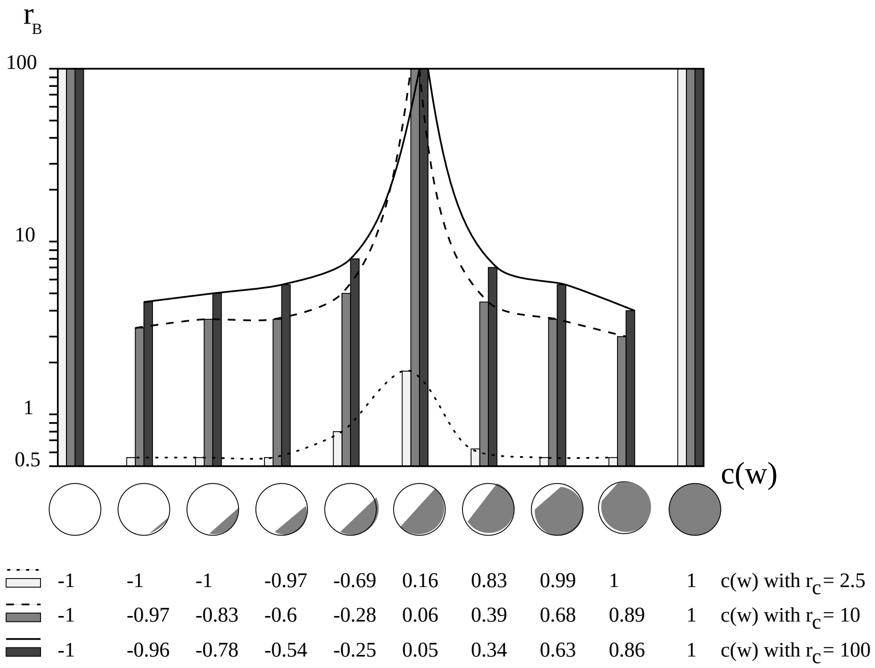


Figure 18: The critical values of $r_{\mathcal{B}}$ as a function of $c(BL(r_{BL}, 9))$ with fixed $r_{\mathcal{C}} = 2.5$,10, and 100 respectively.

From Examples 3 and 4, one can infer that for certain non-circularly-symmetric patterns, only those attractors with very small slope value $|c|$ might always have a stable parameter regime.

These results about relative stability give us a good reason why we usually obtain ON-center or OFF-center aRFs between layers $\mathcal{B}$ and $\mathcal{C}$ much more easily than other patterns, and the developmental dynamics further sustain the ON-center cells with positive $c(\omega)$ and the OFF-center cells with negative $c(\omega)$. $\square$