Published in *Neural Computation* 13/8, p. 1749-1780. 2001.

Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex[1]

Zhaoping Li
17 Queen Square, GCNU, University College London
London WC1N 3AR, UK

Abstract

Recurrent interactions in the primary visual cortex makes its output a complex nonlinear transform of its input. This transform serves pre-attentive visual segmentation, i.e., autonomously processing visual inputs to give outputs that selectively emphasize certain features for segmentation. An analytical understanding of the nonlinear dynamics of the recurrent neural circuit is essential to harness its computational power. We derive requirements on the neural architecture, components, and connection weights of a biologically plausible model of the cortex such that region segmentation, figure-ground segregation, and contour enhancement can be achieved simultaneously. In addition, we analyze the conditions governing neural oscillations, illusory contours, and the absence of visual hallucinations. Many of our analytical techniques can be applied to other recurrent networks with translation invariant neural and connection structures.

1 Introduction

Recurrent neural dynamics is a basic computational substrate for cortical processing. In the primary visual cortex, this recurrent dynamics is instantiated by finite range, lateral, intra-cortical neural connections. The input to the cortex is the retinal image filtered through cortical receptive fields (RFs) shaped like small edges or bars and retinotopically distributed in visual space. The outputs of the cortex are the cell activities, which can be viewed as a complex nonlinear transform of the input under the recurrent interactions. Two characteristics of this transform follow immediately. First, if we focus on cases when top-down feedback from higher visual areas does not change during the course of the transform, the primary cortical computation is autonomous, suggesting that the computation concerned is pre-attentive in nature. In other words, we consider cases when feedback from higher visual areas is purely passive and its role is merely to set a background or operating point for V1 computation. This enables us to isolate the recurrent dynamics in V1 for thorough study. Of course, more extensive computations can doubtless be performed when V1 interacts dynamically with other visual areas; however, this

[1]A preliminary version of this paper was published as "Neural dynamics in a recurrent network model of primary visual cortex" in Proceedings of ICANN99.

is beyond the scope of the paper. The second characteristic is that the recurrent dynamics enables computations to occur at a *global* scale, despite the local connectivity. The output of a V1 cell depends non-locally on its inputs in a way that it is hard to achieve in feed-forward networks with only retinotopically organized connections.

Physiological and psychophysical data suggest that V1 implements preattentive computations such as contour enhancement, texture segmentation, and figure-ground segregation (Kapadia, Ito, Gilbert, and Westheimer1995, Gallant, Nothdurft, van Essen 1995, Knierim and van Essen 1992). To perform these tasks, V1 functions as a saliency circuit that gives higher responses to locations of higher saliency in inputs, such as the borders between texture regions, pop-out figures against backgrounds, and smooth contours (Li 1997, 1998, 1999a, 1999b). Such preattentive segmentation is known to be quite difficult, especially considering that the same cortical circuit needs to achieve both contour enhancement and region or figure/ground segmentation, and that there is still no known general solution to segmentation after decades of research in machine and natural visual algorithms. Various models of the cortex have addressed particular components of the cortical computation, such as contour enhancement, i.e., the relatively higher activities of cells receiving inputs arising from bars belonging to smooth contours (Grossberg and Mingolla 1985, Zucker, Dobbins, Iverson 1989, Yen and Finkel 1998). It is already very hard to model successfully contour enhancement in the cortex (Li 1998). Previous efforts (Grossberg and Mingolla 1985) have been made to capture in a single model both contour enhancement and texture segmentation. However, a fully functional and dynamically well-behaved model has only recently been proposed (Li 1997, 1999a).

Understanding the complex, recurrent, and nonlinear neural dynamics underlying the computation is essential to marshall its power. Li (1997, 1998, 1999a) introduced and described the structure and behavior of a recurrent model of V1 that simultaneously achieves the desired components of the computation. In this paper, we describe the mathematical analysis of the nonlinear dynamics that enabled the computational design of our model. We study issues such as network architecture, computational constraints, and dynamic stability that are directly relevant to the global scale computation. By contrast, single unit properties, such as orientation tuning, that are less relevant to the global scale computation will not be a focus. Some of our analytical techniques, e.g., the analysis of the cortical microcircuit and the stability study of the translation invariant networks, can be applied to study other cortical areas that share the common properties of neural elements, connections, and the canonical microcircuit (Shepherd 1990).

2 A minimal model of the primary visual cortex

A minimal model of the cortex is the one which has just enough components to execute the necessary computations without excess details. It is essentially a subjective matter as to what a minimal model is, since there is no recipe for a minimalist design. However, we present, as a candidate, a model that instantiates all the de-

sired computation, but for which simplified versions fail. Throughout the paper, we try to keep our analysis general in discussing characteristics of the recurrent dynamics. However, to illustrate or demonstrate particular analytical results, and approximation and simplification techniques, we often use a model of V1 whose specifics and numerical parameters are available (Li 1998, 1999a)[2], so that the readers can try out our simulations.

We model only layer 2-3 cells in the cortex, which are mainly responsible for the recurrent dynamics. A model neuron has membrane potential x and output or firing rate $g_x(x)$, which is a sigmoid-like function of x. Model cells have orientation selective RFs arranged on a regular 2-dimensional grid in image coordinates. At each grid point $i = (m_i, n_i)$, where m_i and n_i are the horizontal and vertical coordinates, there are K units, one each for preferred orientations $\theta = k\pi/K$ for $k = 0, 1, ..., K-1$ spanning 180^o. Unit $i\theta$ has its RF located at i and prefers orientation θ. It receives external visual inputs $I_{i\theta}$, which is the result of pre-processing the visual image through the RF. Its response $g_x(x_{i\theta})$ is the result of both $I_{i\theta}$ and the recurrent interactions. The image grid and the interactions are treated as translation invariant, allowing us to use many powerful analytical techniques. However, we should keep in mind that translation symmetry holds approximately only over a sufficiently small portion of the visual field, since our visual system has different resolutions at different eccentricities.

The desired computation $\{I_{i\theta}\} \rightarrow \{g_x(x_{i\theta})\}$ gives higher responses $g_x(x_{i\theta})$ to input bars $i\theta$ of higher perceptual saliency. For instance, even if two input bars $i\theta$ and $j\theta'$ have the same input contrast $I_{i\theta} = I_{j\theta'}$, the response $g_x(x_{i\theta})$ to $i\theta$ may be higher if $i\theta$ (but not $j\theta'$) is part of an isolated smooth contour, or is at the boundary of a texture region, or is a pop-out target against a background. Conversely, if the input bars are of the same saliency, e.g., when the input consists merely of bars of the same contrast from a homogeneous texture without any boundary, the the output level to every bar should be the same.

2.1 A less-than-minimal recurrent model of V1

A very simple recurrent model of the cortex can be described by equation:

$$\dot{x}_{i\theta} = -x_{i\theta} + \sum_{j\theta'} T_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta} + I_o \tag{1}$$

where $-x_{i\theta}$ models the decay in membrane potential, and I_o is the background input. The recurrent connections $T_{i\theta,j\theta'}$ link cells $i\theta$ and $j\theta'$. Visual input $I_{i\theta}$ persists after onset, and initializes the activity levels $g_x(x_{i\theta})$. The activities are then

[2]In Oct. 2000, a typo was discovered in the Appendix of the published version of Li (1998, 1999a) for the model parameter for $W_{i\theta,j\theta'}$. In Li (1998, 1999a), it was mistakenly written that "$W_{i\theta,j\theta'} = 0$" when "$d \geq 10$" or other conditions listed in Li (1998, 1999a) are satisfied. The correct model parameter, which have been used to produce all the published model results so far (including the ones in Li (1998, 1999a)), should be such that the condition "$d \geq 10$" printed in Li (1998, 1999a) be changed to condition "$d/\cos(\beta/4) \geq 10$". Here d and β are just as defined in Li (1998, 1999a). The typo should lead to quantitative changes in the model behavior from those published so far or those presented in this paper.

modified by the network interaction, making $g_x(x_{i\theta})$ dependent on input $I_{j\theta'}$ for $(j\theta') \neq (i\theta)$. Translation invariance in the connections means that $T_{i\theta,j\theta'}$ depends only on the vector $i - j$ and the relative angles of this vector to the orientations θ and θ'. Reflection symmetry means that $T_{i\theta,j\theta'} = T_{j\theta',i\theta}$.

Many previous models of the primary visual cortex (e.g., Grossberg and Mingolla 1985, Zucker, Dobbins, Iverson 1989, Braun, Niebur, Schuster, and Koch 1994) can be seen as more complex versions of the one described above. The added complexities include stronger nonlinearities, global normalization (e.g., by adding a global normalizing input to the background I_o), and shunting inhibition. However, they are all characterized by reciprocal or symmetric interactions between model units. It is well known (Hopfield 1982, Cohen and Grossberg 1983) that in a symmetric recurrent network as in equation (1), given any stationary input $I_{i\theta}$, the dynamic trajectory $x_{i\theta}(t)$ will converge in time t to a fixed point which is a local minimum (attractor) in an energy landscape

$$E(\{x_{i\theta}\}) = -\frac{1}{2}\sum_{i\theta,j\theta'} T_{i\theta,j\theta'} g_x(x_{i\theta}) g_x(x_{j\theta'}) - \sum_{i\theta} I_{i\theta} g_x(x_{i\theta}) + \sum_{i\theta} \int_0^{g_x(x_{i\theta})} g_x^{-1}(x)dx \quad (2)$$

Empirically, this convergence behavior to attractors still holds when the complexities above are added to the network.

The fixed point $\bar{x}_{i\theta}$ of the motion trajectory, or the minimum energy state where $\partial E/\partial g_x(x_{i\theta}) = 0$ for all $i\theta$, is (when $I_o = 0$)

$$\bar{x}_{i\theta} = I_{i\theta} + \sum_{j\theta'} T_{i\theta,j\theta'} g_x(\bar{x}_{j\theta'}) \quad (3)$$

Without recurrent interactions ($T = 0$), this minimum $\bar{x}_{i\theta} = I_{i\theta}$ is a faithful copy of the input $I_{i\theta}$. Sufficiently strong interactions T shape $\bar{x}_{i\theta}$ and make them unfaithful to the input. This happens when T is so strong that one of the eigenvalues $\lambda_{\mathbf{T}}$ of the matrix $\mathbf{T}$ with elements $\mathbf{T}_{i\theta,j\theta'} \equiv T_{i\theta,j\theta'} g'_x(\bar{x}_{j\theta'})$ satisfies $\lambda_{\mathbf{T}} > 1$ (here g'_x is the slope of $g_x(.)$). For instance, when the input $I_{i\theta}$ is translation invariant such that $I_{i\theta} = I_{j\theta}$ for all $i \neq j$, there is a translation invariant fixed point $\bar{x}_{i\theta} = \bar{x}_{j\theta}$ for all $i \neq j$. Strong interactions T could make this fixed point unstable and no longer a local minimum of the energy, and pull the state into an attractor in the direction of an eigenvector of $\mathbf{T}$ which is not translation invariant, i.e., $x_{i\theta} \neq x_{j\theta}$ for $i \neq j$. Computationally, the input unfaithfullness, i.e., $g_x(x_{i\theta})$ is not a function of $I_{i\theta}$ alone, is desirable to a limited degree since this is how a saliency circuit produces differential outputs $g_x(x_{i\theta})$ to input bars of same contrast $I_{i\theta}$ but different saliencies. However, this unfaithfulness should be driven by the nature of the input pattern $\{I_{i\theta}\}$ or its deviation from homogeneity (e.g., the smooth contours or figures against a background). Otherwise, visual hallucinations (Ermentrout and Cowan 1979) result when spontaneous or non-input-driven network behaviors — *spontaneous pattern formation or symmetry breaking* — happen. Note that if $\{x_{i\theta}\}$ is an attractor under homogeneous input $I_{i\theta} = I_{j\theta}$, so is a translated state $\{x'_{i\theta}\}$ such that $x'_{i\theta} = x_{i+a,\theta}$ for any translation a, since $\{x_{i\theta}\}$ and $\{x'_{i\theta}\}$ have the same energy value E. Hence, the absolute positions of the hallucinated patterns are random and shiftable. When the translation a is one dimensional, such a continuum of attractors has been called a "line attractor" (Zhang 1996). For two or more dimensional patterns, the continuum is a "surface attractor".

To illustrate, consider an example when $T_{i\theta,j\theta'} \propto \delta_{\theta,\theta'}$ only links cells that prefer the same orientation, an idealization from observations (Gilbert and Wiesel 1983, Rockland and Lund 1983) that the lateral interactions tend to link cells preferring similar orientations. The network contains multiple, independent, subnetworks, one each for every θ. Take the $\theta = 0^o$ subnet, and for convenience drop the subindex θ. Consider a simple center-surround interaction, such that in a Manhattan grid,

$$T_{ij} \propto \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } (m_j, n_j) = (m_i \pm 1, n_i) \text{ or } (m_i, n_i \pm 1) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

With sufficiently strong T, the network under homogeneous input $I_i = I_j$ for all i, j can settle into an "antiferromagnetic" state in which neighboring units x_i exhibit one of the two different activities $x_{m_i,n_i} = x_{m_i+1,n_i+1} \neq x_{m_i+1,n} = x_{m_i,n_i+1}$. This pattern $\{x_i\}$ is just a spatial array of replicas of the center-surround interaction pattern T.

Fig (1) shows the behavior of a subnet for $\theta = 0$, the vertical bars, when the interaction T_{ij} depends on the orientation of $i - j$ and is *not* rotationally invariant. Here, $T_{ij} > 0$ between local and roughly vertically displaced i and j, and $T_{ij} < 0$ between local and more horizontally displaced i and j. Hence, two nearby bars i and j excite each other when they are co-aligned and inhibit each other otherwise, as suggested by experimental observations and theoretical studies (e.g., Kapadia et al 1995, Polat and Sagi 1993, Field, Hayes, and Hess 1993). Although the network enhances an input (vertical) line relative to the isolated (short) bar, it also hallucinates other vertical lines under noisy inputs.

The competition between internal interactions T and the external inputs I to shape $\{x_i\}$ is uncompromising in such recurrent networks. For analysis, take for simplicity T_{ij} such that $T_{ij} \neq 0$ only when i and j are in the same or nearest neighbor (vertical) columns. $T_0 \equiv \sum_{j,m_j=m_i} T_{ij} > 0$ is the total recurrent excitation within a vertical contour (column) for contour enhancement, and $T_1 \equiv -\sum_{j,m_j=m_i+1} T_{ij} = -\sum_{j,m_j=m_i-1} T_{ij} > 0$. is the total recurrent suppression between two nearby vertical columns, to suppress noise and background. Both T_0 and T_1 should be large for a large difference in saliency between a vertical contour and a homogeneous background input $I_i = I_j$ for all i, j. However, when $T_0 + 2T_1 > 1$ (for $g'_x = 1$), this network under homogeneous input spontaneously breaks symmetry to hallucinate saliency waves — two alternate saliencies for neighboring columns. (Mathematically, $T_0 + 2T_1$ is the eigenvalue of the matrix $\mathbf{T}$ with this saliency wave as the eigenvector.) Hence, contour enhancement makes the network prone to "see" contours even when there is none. The orientations and widths of the "ghost contours" match the interaction structure T. Avoiding the hallucination forces T_0 and/or T_1 to be smaller, and so harms the capability of the network to enhance contours relative to backgrounds (Li and Dayan 1999). In other words, although symmetric recurrent networks are useful for associative memory computations (Hopfield 1982), for which correcting significant input errors or filling-in extensively missing inputs is exactly what is needed, such an input distortion is too strong for early visual tasks that require greater faithfulness to visual input.

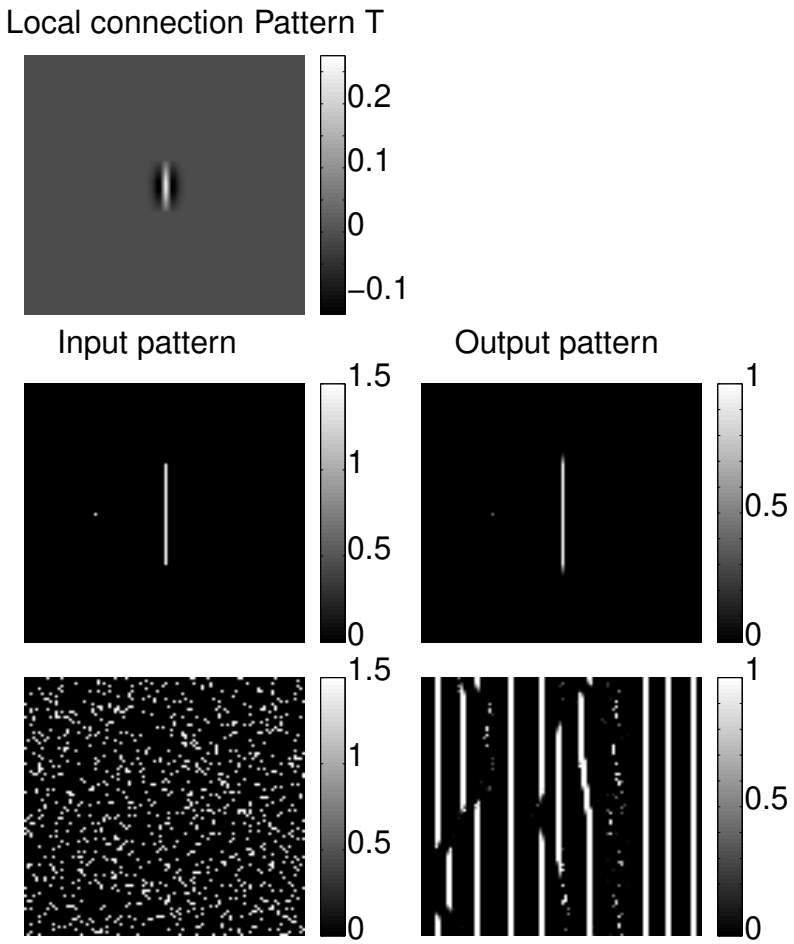

Figure 1: A reduced model consisting of symmetrically coupled cells tuned to vertical orientation ($\theta = 0$). Shown here are 5 gray scale images, each has a scale bar on the right. The network has 100x100 cells arranged in a 2-d array, with wrap around boundary condition. Each cell models a cortical cell tuned to vertical orientation, in a retinotopic manner. The sigmoid function $g_x(x)$ of the cells is $g_x(x) = 0$ when $x < 1$, $g_x(x) = x - 1$ when $1 \leq x < 2$, and $g_x(x) = 1$ when $x > 2$. The top image shows the connection pattern between the center cell and the other cells. This pattern is local and translation invariant, it gives local colinear excitation between cells vertically displaced, and local inhibition between cells horizontally displaced. Middle left: 2-d input pattern I, an input line and a noise spot. Middle right: 2-d output pattern $g_x(x)$ to the input at middle left — the line induces a response that is $\sim 100\%$ higher than the noise spot. Bottom left: 2-d input pattern I for noise input. Bottom right: 2-d output pattern $g_x(x)$ to the noisy input — hallucination of vertical streaks.

2.2 A minimal recurrent model with hidden units

The major weakness of the symmetrically connected model is the attractor dynamics which strongly attract the network state $\{x_{i\theta}\}$ away from the ones guided by the visual input $\{I_{i\theta}\}$. Since this attractor dynamics is largely dictated by the symmetry of the neural connections, it can not be removed by introducing ion channels or

spiking neurons (rather than firing rate neurons), for instance, nor by mechanisms like shunting inhibition, global activity normalization, and input gating (Grossberg and Mingolla 1985, Zucker et al 1989, Braun et al 1994), which are used by many models despite their questionable biological foundations. Attractor dynamics is untenable, however, in the face of the well established fact that a real neuron is either exclusively excitatory or exclusively inhibitory. It is obviously impossible to have symmetric connections between excitatory and inhibitory neurons. Mathematical analysis by Li and Dayan (1999) showed that asymmetric recurrent E-I networks with separate excitatory (E) and inhibitory (I) cells can indeed perform computations that symmetric ones cannot. Thus we model neurons $x_{i\theta}$ as exclusively excitatory pyramidal cells, and introduce one inhibitory interneuron (hidden units) $y_{i\theta}$ for each $x_{i\theta}$ to mediate indirect, or disynaptic, inhibition between $x_{i\theta}$'s, as in the real cortex (White 1989, Gilbert 1992, Rocklandand Lund 1983, also see Grossberg and Raizada 2000 for a much more fully elaborated model). The units $x_{i\theta}$ and $y_{i\theta}$ in such an E-I pair are reciprocally connected. Hence the dynamical equations become:

$$\begin{aligned} \dot{x}_{i\theta} &= -x_{i\theta} - g_y(y_{i,\theta}) + J_o g_x(x_{i\theta}) - \sum_{\Delta\theta \neq 0} \psi(\Delta\theta) g_y(y_{i,\theta+\Delta\theta}) \\ &+ \sum_{j \neq i, \theta'} J_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta} + I_o \qquad (5) \\ \dot{y}_{i\theta} &= -\alpha_y y_{i\theta} + g_x(x_{i\theta}) + \sum_{j \neq i, \theta'} W_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_c \qquad (6) \end{aligned}$$

where α_y and $g_y(y)$ model the interneuron $y_{i\theta}$, which inhibits its partner $x_{i\theta}$. The longer range connections $T_{i\theta,j\theta'}$ (between cells in different hypercolumns $i \neq j$) are now separated into two terms: (1) monosynaptic excitation $J_{i\theta,j\theta'} \geq 0$ between $x_{i\theta}$ and $x_{j\theta'}$ and (2) disynaptic inhibition $W_{i\theta,j\theta'} \geq 0$ between $x_{i\theta}$ and $x_{j\theta'}$ via the interneuron $y_{i\theta}$. Including both the monosynaptic and disynaptic pathways, the net effective connection between $x_{i\theta}$ and $x_{j\theta'}$ in stationary (but not in dynamic) states is, for example, $J_{i\theta,j\theta'} - W_{i\theta,j\theta'}/\alpha_y$ if $g_y(y) = y$, and it can be either facilitatory or inhibitory. Both $\psi(\Delta\theta)$ and J_o are explicit representations of the original interaction $T_{i\theta,i\theta'}$ between units within a hypercolumn. $\psi(\Delta\theta) \leq 1$ models local inhibition and $J_o g_x(x_{i\theta})$ models self excitation. Fig. (2C) schematically shows an example of the network. I_c and I_o are background inputs, including neural noise, feedback from higher areas, and inputs modeling the general and local normalization of activities (Li 1998) (which are omitted in the analysis in this paper, though are present in the simulations). An edge of input strength $\hat{I}_{i\beta}$ at i with orientation β in the input image contributes to $I_{i\theta}$ (for $\theta \approx \beta$) by an amount $\hat{I}_{i\beta}\phi(\theta - \beta)$, where $\phi(\theta - \beta)$ is the orientation tuning curve.

Lateral connections link cells preferring similar orientations. To implement net colinear facilitation and non-colinear flank inhibition (between similarly oriented bars), the excitatory J connections are dominant between units preferring coaligned bars ($\theta \sim \theta' \sim \angle(i-j)$), while the inhibitory W connections are dominant between units preferring non-aligned (but still similarly oriented) ones (Zucker et al 1989, Field et al 1993, Li 1998, 1999a). Such an interaction pattern has been called

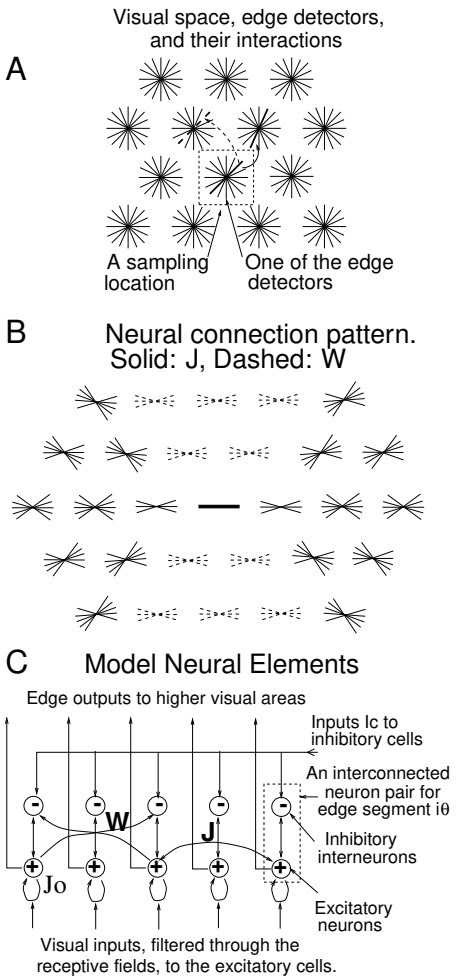

Figure 2: A schematic of the minimal model of the primary visual cortex. **A:** Visual inputs are sampled in a discrete grid by edge/bar detectors, modeling receptive fields (RFs) for V1 layer 2-3 cells. Each grid point has K neuron pairs (see **C**), one per bar segment. All cells at a grid point share the same RF center, but are tuned to different orientations spanning 180^o, thus modeling a hypercolumn. A bar segment in one hypercolumn can interact with another in a different hypercolumn via monosynaptic excitation J (the solid arrow from one thick bar to another), and/or disynaptic inhibition W (the dashed arrow to a thick dashed bar). See also **C**. **B**: A schematic of the neural connection pattern from the center (thick solid) bar to neighboring bars within a finite distance (a few RF sizes). J's contacts are shown by thin solid bars. W's are shown by thin dashed bars. All bars have the same connection pattern, suitably translated and rotated from this one. **C:** An input bar segment is associated with an interconnected pair of excitatory and inhibitory cells, each model cell models abstractly a local group of cells of the same type. The excitatory cell receives visual input and sends output $g_x(x_{i\theta})$ to higher centers. The inhibitory cell is an interneuron. The visual space has toroidal (wrap-around) boundary conditions.

the *association field* (Field et al 1993)). A simple model of this interaction is the *bi-phasic* pattern as in Fig. (2B): $J > 0$ and $W = 0$ for mutual excitation and $J = 0$ and

$W > 0$ for mutual inhibition (Li 1998, 1999a). Physiological evidence (Hirsch and Gilbert 1991) suggests that both $J_{i\theta,j\theta'} > 0$ and $W_{i\theta,j\theta'} > 0$ contribute to the links between a given pair of pyramidal cells $x_{i\theta}$ and $x_{j\theta'}$. This gives extra computational flexibility (e.g., contrast dependence of contextual influences, see section 3) by letting the ratio $J_{i\theta,j\theta'} : W_{i\theta,j\theta'}$ determine the overall sign of the interaction. For illustrative convenience, however, the simpler bi-phasic connection is sometimes used in this paper to demonstrate our analysis and is used for all the examples in the figures.

As we mentioned, in principle, an E-I recurrent model can perform computations that symmetric models cannot. In practice, this is not guaranteed and has to be ensured by designing the right model parameters, in particular, J and W, guided by an analytical understanding of the nonlinear dynamics.

3 Dynamic analysis

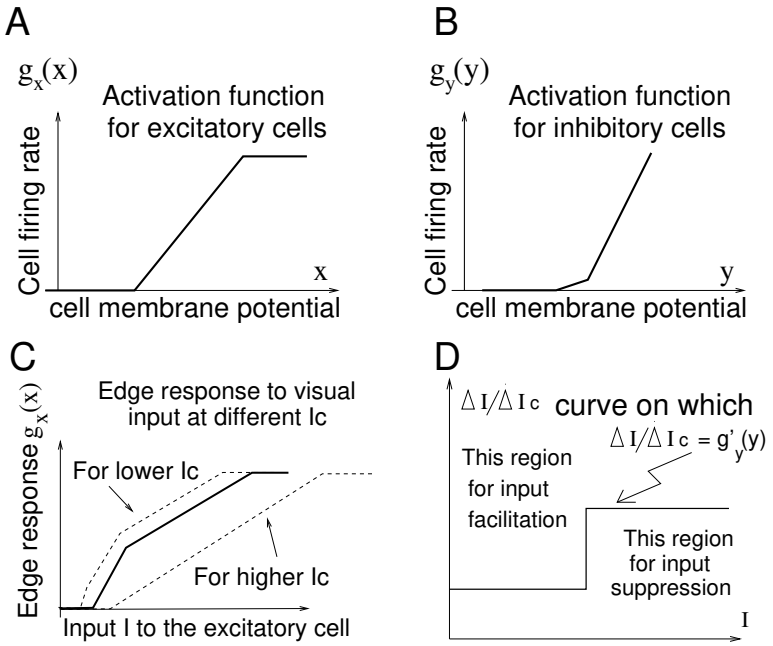

Figure 3: A,B: examples of $g_x(x)$ and $g_y(y)$ functions. C: Input-output function $I \to g_x(\bar{x})$ for an isolated neural pair without inter-pair neural interactions, under different levels of I_c. D: The overall effect of the external or contextual inputs $(\Delta I, \Delta I_c)$ on a neural pair is excitatory or inhibitory if $\Delta I/\Delta I_c$ is large or less than $g'_y(\bar{y})$, which depends on I.

The model state is characterized by $\{x_{i\theta}, y_{i\theta}\}$, or simply $\{x_{i\theta}\}$, omitting the hidden units $\{y_{i\theta}\}$. The interaction between excitatory and inhibitory cells makes $\{x_{i\theta}(t)\}$ intrinsically oscillatory in time. Given an input $\{I_{i\theta}\}$, the model does not guarantee convergence to a fixed point where $\dot{x}_{i\theta} = \dot{y}_{i\theta} = 0$. However, if $\{x_{i\theta}(t)\}$

converges to, or oscillates periodically around, a fixed point, after the transient following the onset of $\{I_{i\theta}\}$, the temporal average $\{\bar{x}_{i\theta}\}$ of $\{x_{i\theta}(t)\}$ can characterize the model output and approximate the fixed point. We henceforth use the notation $\{\bar{x}_{i\theta}\}$ to denote either the fixed point or the temporal average, and denote the computation as $I \rightarrow g_x(\bar{x}_{i\theta})$. Section 3.1-3.6 will analyze $I \rightarrow g_x(\bar{x}_{i\theta})$ and derive constraints on J and W in order to make $I \rightarrow g_x(\bar{x}_{i\theta})$ achieve the desired computations. Other investigators have also analyzed the fixed point behavior $I \rightarrow g_x(\bar{x}_{i\theta})$ in such E-I networks or the corresponding symmetric ones (Ben-Yishai et al 1995, Stemmler et al 1995, Somers et al 1998, Mundel et al 1997, Tsodyks et al 1997), mainly to model a local circuit of a hypercolumn (or a CA1 region) with simplified or no spatial organization beyond the hypercolumn. Our analysis emphasizes the spatial or geometrical organization of visual inputs in order to study global visual computations. Section 3.7 studies the stability and dynamics around $\{\bar{x}_{i\theta}\}$ and derives constraints on the model parameters coming from the need to avoid visual hallucination (Ermontrout and Cowan 1979) — the curse of symmetric networks.

3.1 A single pair of neurons

In isolation, a single pair $i\theta$ follows equations

$$\dot{x} = -x - g_y(y) + J_o g_x(x) + I \tag{7}$$
$$\dot{y} = -y + g_x(x) + I_c \tag{8}$$

where $\alpha_y = 1$ for simplicity (as in the rest of the paper), index $i\theta$ is omitted, and $I = I_{i\theta} + I_o$. The input-output $(I, I_c \rightarrow g_x(\bar{x}))$ gain at a fixed point $(\bar{x}, \bar{y})$ is

$$\frac{\delta g_x(\bar{x})}{\delta I} = \frac{g'_x(\bar{x})}{1 + g'_x(\bar{x}) g'_y(\bar{y}) - J_o g'_x(\bar{x})}, \qquad \frac{\delta g_x(\bar{x})}{\delta I_c} = -g'_y(\bar{y}) \frac{\delta g_x(\bar{x})}{\delta I} \tag{9}$$

When both $g_x(x)$ and $g_y(y)$ are piece-wise linear (Fig. (3A,B)) functions, so is the input-output relation $I \rightarrow g_x(\bar{x})$ (Fig. (3C)). The threshold, input gain control, and saturation in $I \rightarrow g_x(\bar{x})$ are apparent. The slope $\frac{\delta g_x(\bar{x})}{\delta I}$ is non-negative, otherwise, $I = 0$ gives non-zero output $x \neq 0$. It increases with $g'_x(\bar{x})$, decreases with $g'_y(\bar{y})$, and depends on I_c. Shifting (I, I_c) to $(I + \Delta I, I_c + \Delta I_c)$ changes $g_x(\bar{x})$ by $\Delta g_x(\bar{x}) \approx (\delta g_x(\bar{x})/\delta I)(\Delta I - g'_y(\bar{y})\Delta I_c)$, which is positive or negative depending on whether $\Delta I/\Delta I_c > g'_y(\bar{y})$. Hence, a more elaborate model could allow a fraction of the external visual input to go onto interneurons, as suggested by physiology (White 1989) and modeled by Grossberg and Raizada (2000), provided that $\Delta I/\Delta I_c > g'_y(\bar{y})$. Contextual inputs from other neuron pairs (via J and W) effectively give $(\Delta I, \Delta I_c)$. In our example when $g'_y(\bar{y})$ increases with I (or I_c), the contextual inputs can switch from being facilitatory to being suppressive as I increases (Fig. (3 D)). This input contrast dependence of the contextual influences has been observed physiologically (Sengpiel, Baddeley, Freeman, Harrad, and Blakemore 1995) and modelled by others (Stemmler, Usher, Niebur 1995, Somers, Todorov, Siapas, Toth, Kim, Sur 1998).

3.2 Two interacting pairs of neurons with non-overlapping receptive fields

Using indices $a = 1, 2$ to denote the two pairs and their associated quantities ($J_{12} = J_{21} = J$ and $W_{12} = W_{21} = W$),

$$\begin{aligned} \dot{x}_a &= -x_a - g_y(y_a) + J_o g_x(x_a) + J g_x(x_b) + I_a + I_o \\ \dot{y}_a &= -y_a + g_x(x_a) + W g_x(x_b) + I_c \end{aligned}$$

where $a, b = 1, 2$ and $a \neq b$. Including monosynaptic and disynaptic pathways, the net effective connection from x_2 to x_1, according to the gain functions $\delta g_x(\bar{x})/\delta I$ and $\delta g_x(\bar{x})/\delta I_c$, is $J - g'_y(\bar{y}_1)W$. When $I \equiv I_1 = I_2$ in the simplest case, $\bar{x} \equiv \bar{x}_1 = \bar{x}_2$ and $\bar{y} \equiv \bar{y}_1 = \bar{y}_2$. The two bars can excite or inhibit each other depending on whether $J - g'_y(\bar{y})W > 0$. This in turn depends on the input I through $g'_y(\bar{y})$. When $I_1 > I_2$, we have $(\bar{x}_1, \bar{y}_1) > (\bar{x}_2, \bar{y}_2)$. Usually, $g'_y(\bar{y})$ increases with $\bar{y}$, hence $J_{12} - g'_y(\bar{y}_1)W_{12} < J_{21} - g'_y(\bar{y}_2)W_{21}$. In particular, it can happen that $J_{12} - g'_y(\bar{y}_1)W_{12} < 0 < J_{21} - g'_y(\bar{y}_2)W_{21}$, i.e., x_1 excites x_2 which in turn inhibits x_1. This implies that two interacting pairs tend to have closer activity values x_1 and x_2 than two non-interacting pairs.

Even this very simple contextual influence can already account for some perceptual phenomena involving sparse visual inputs consisting only of single test and contextual bars. Examples include the altered detection threshold (Polat and Sagi 1993, Kapadia et al 1995) or perceived orientation (tilt illusion, Mundel et al 1997, Kapadia 1998) of a test bar when a contextual bar is present.

3.3 A one dimensional array of identical bars

An infinitely long, horizontal array of evenly spaced, identical, bars gives an input pattern approximated as

$$I_{i\theta} = \begin{cases} I_{array} & \text{for } i = (m_i, n_i = 0) \text{ on the horizontal axis and } \theta = \theta_1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The approximation $I_{i\theta} = 0$ for $\theta \neq \theta_1$ is good for small orientation tuning width and low input contrast. When bars $i\theta$ outside that array are silent $g_x(x_{i\theta}) = 0$ due to insufficient excitation, we omit them and treat the remaining system as one dimensional. Omitting index θ and using i to denote bars according to their one dimensional location, we get

$$\dot{x}_i = -x_i - g_y(y_i) + J_o g_x(x_i) + \sum_{j \neq i} J_{ij} g_x(x_j) + I_{array} + I_o \tag{11}$$

$$\dot{y}_i = -y_i + g_x(x_i) + \sum_{j \neq i} W_{ij} g_x(x_j) + I_c \tag{12}$$

Translation symmetry implies that all units have the same equilibrium point $(\bar{x}_i, \bar{y}_i) = (\bar{x}, \bar{y})$, and

$$\dot{\bar{x}} = 0 = -\bar{x} - g_y(\bar{y}) + (J_o + \sum_{i \neq j} J_{ij}) g_x(\bar{x}) + I_{array} + I_o \tag{13}$$

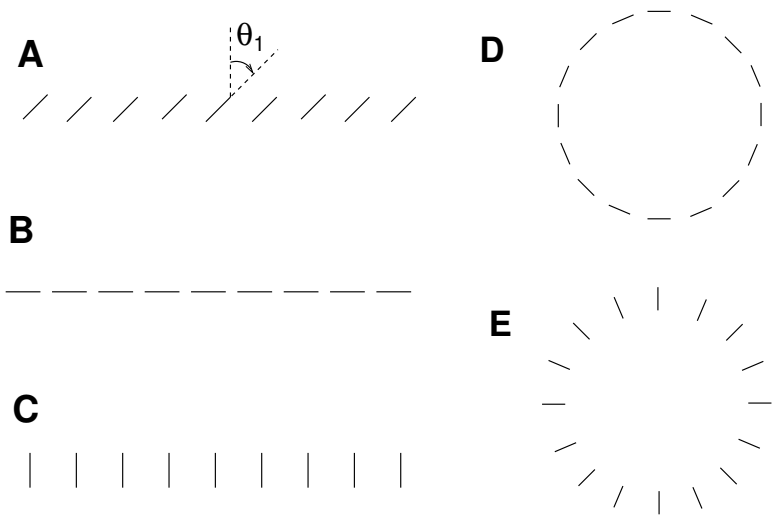

Figure 4: Examples of the one dimensional input stimuli mentioned in the text. **A**: horizontal array of identical bars oriented at angle θ_1. **B**: A special case of **A** when $\theta_1 = \pi/2$ and, in **C**, when $\theta_1 = 0$. **D**: an array of bars aligned into, or tangential to, a circle, the pattern in **B** is a special case of this circle when the radius is infinitely large. **E**: same as D except that the bars are perpendicular to the circle circumference, the pattern in **C** is a special case when the radius is infinitely large.

$$\dot{\bar{y}} \;=\; 0 = -\bar{y} + (1 + \sum_{i \neq j} W_{ij}) g_x(\bar{x}) + I_c \tag{14}$$

This array is then equivalent to a single neural pair (cf. equations (7) and (8)) with the substitution $J_o \rightarrow J_o + \sum_j J_{ij}$ and $g'(\bar{y}) \rightarrow g'_y(\bar{y})(1 + \sum_j W_{ij})$. The response to bars in the array is thus higher than that to an isolated bar if the net extra excitatory connection

$$\mathcal{E} \equiv \sum_j J_{ij} \tag{15}$$

is stronger than the net extra inhibitory (effective) connection

$$\mathcal{I} \equiv g'_y(\bar{y}) \sum_j W_{ij}. \tag{16}$$

The input-output relationship $I \rightarrow g_x(\bar{x})$ is qualitatively the same as that of a single bar, with a quantitative change in the gain

$$\frac{\delta g_x(\bar{x})}{\delta I} = \frac{g'_x(\bar{x})}{1 + g'_x(\bar{x})(g'_y(\bar{y}) - (\mathcal{E} - \mathcal{I})) - J_o g'_x(\bar{x})}. \tag{17}$$

$\mathcal{E}$ and $\mathcal{I}$ depend on θ_1. Consider the case of association field connections. When the bars are parallel to the array, making a straight line (Fig (4B)), $\mathcal{E} > \mathcal{I}$. The condition for contour enhancement is

$$\text{Contour facilitation} F_{\text{contour}} \equiv (\mathcal{E} - \mathcal{I}) g_x(\bar{x}) > 0 \quad \text{and is sufficiently strong.} \tag{18}$$

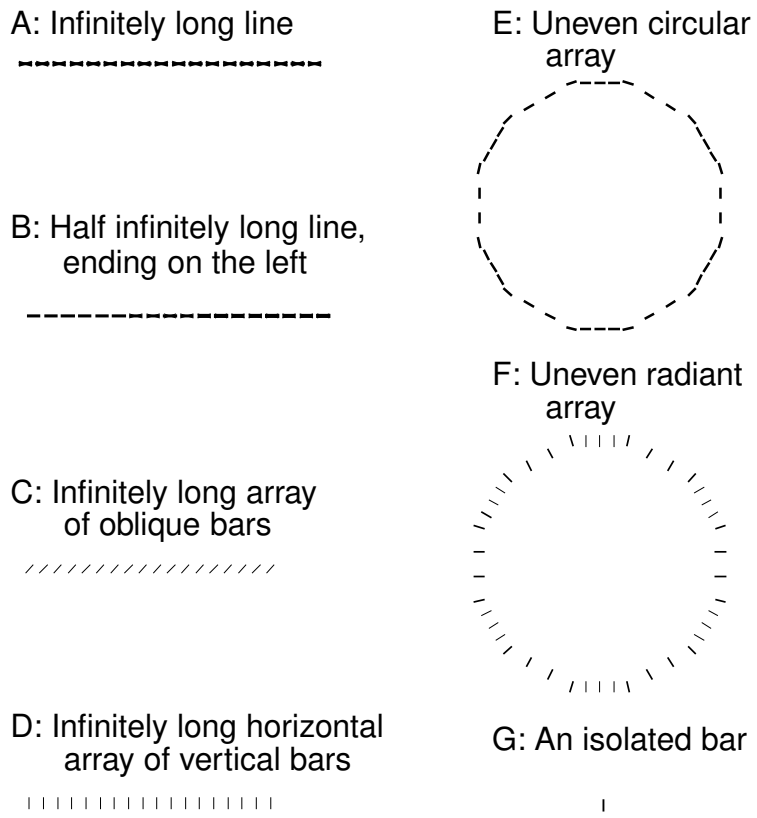

Figure 5: Simulated outputs from a cortical model to corresponding visual input patterns of 1 dimensional arrays of bars. The model transforms input $I_{i\theta}$ to cell output $g_x(x_{i\theta})$. The thicknesses of the bars $i\theta$ are proportional to temporally averaged model outputs $g_x(x_{i\theta})$. The corresponding (suprathreshold) input $\hat{I}_{i\theta} = 1.5$ is of low/intermediate contrast and is the same for all 7 examples and all visible bars. Different outputs $g_x(x_{i\theta})$ for different examples or for different bars in each example are caused by contextual interactions. Overall contextual facilitations cause higher outputs in **A, B, E** than that of an isolated bar in **G**, while overall contextual suppressions cause lower outputs in **C, D, F** (compare the different thicknesses of the bars). Note the deviations from the idealized approximations in the text. Uneven spacing between the bars (**F, G**) or an end of a line (at the left end of **B**) cause deviations from the translation invariance of responses. Note that the responses taper off near the line end in **B**, and that the responses are noticably weaker to bars that are more densely packed in **F**. In **A, B**, cells preferring neighboring orientations (near horizontal) at the line are also excited above threshold, unlike the approximated treatment in the text.

When the bars are orthogonal to the array (Fig (4C)), $\mathcal{E} < \mathcal{I}$ and the responses are suppressed. This analysis extends to other translation invariant one dimensional arrays like those in Fig (4D, E), for which the index i simply denotes a bar at a location along the array (Li 1998). The straight line in Fig (4B) is in fact the limit of a circle in Fig (4D) when the radius goes to infinity. Similarly, the pattern in Fig (4C) is a special case of the one in Fig (4E).

The qualities of the approximations in equations (10 -14) depend on the in-

put, as shown in Fig. (5). Contextual facilitation in Fig. (5A, B, E) and contextual suppression in Fig. (5C, D, F) are visualized by the thicker and thinner bars, respectively, than the isolated bar in Fig. (5G). In Fig. (5A), cells whose RFs are centered on the line but not oriented exactly horizontally are also excited above threshold, unlike our approximation $g_x(x_{i\theta}) = 0$ for non-horizontal θ. (This should not cause perceptual problems, though, given population coding.) This is caused by direct visual input $I_{i\theta}$ for $\theta \neq \theta_1$ ($\theta \approx \theta_1$) *and* the colinear facilitation from other bars in the line. The approximation of translation invariance $\bar{x}_i = \bar{x}_j$ for all bars in the array is compromised when the array has an end, e.g., Fig. (5B), or when the bars are unevenly spaced, e.g., Fig. (5E,F). In Fig. (5B), the bars at or near the left end of the line are less enhanced since they receive less or no contextual facilitation from their left. In Fig. (5F), the more densely spaced bars receive more contextual suppression than others.

3.4 Two dimensional textures and texture boundaries

The analysis of the one dimensional array also applies to an infinitely large two dimensional texture of uniform input $I_{i\theta_1} = I_{texture}$ when $i = (m_i, n_i)$ sit on a regularly spaced grid (Fig. (6A)). The sums $\mathcal{E} = \sum_j J_{ij}$ and $\mathcal{I} = g'_y(\bar{y}) \sum_j W_{ij}$ are taken over all j in that grid.

Physiologically the response to a bar is reduced when the bar is part of a texture (Knierim and van Essen 1992). This can be achieved when $\mathcal{E} < \mathcal{I}$. Consider, for example, the case when $i = (m_i, n_i)$ form a Manhattan grid with integer values of m_i and n_i (Fig (6)). The texture can be seen as a horizontal array of vertical arrays of bars, e.g., a horizontal array of vertical contours in Fig. (6B). The effective connections between the vertical arrays (Fig. (6DEF)) distance a apart are:

$$J'_a \equiv \sum_{j, m_j = m_i + a} J_{ij}, \qquad\qquad W'_a \equiv \sum_{j, m_j = m_i + a} W_{ij}. \tag{19}$$

Then $\mathcal{E} = \sum_a J'_a$ and $\mathcal{I} = g'_y(\bar{y}) \sum_a W'_a$. The effective connection within a single vertical array is J'_0 and W'_0. One has to design J and W such that contour enhancement and texture suppression can occur using the same neural circuit (V1). That is, when the vertical array is a long straight line ($\theta_1 = 0$), contour enhancement (i.e., $J'_0 > g'_y(\bar{y}) W'_0$) occurs when the line is isolated, but overall suppression (i.e., $\mathcal{E} = \sum_a J'_a < \mathcal{I} = g'_y(\bar{y}) \sum_a W'_a$) occurs when that line is embedded within a texture of lines (Fig. (6B)), as long as there is sufficient excitation within a line and sufficient inhibition between the lines.

Computationally, contextual suppression within a texture means that the boundaries of a texture region induce relatively higher responses, thereby marking the boundaries for segmentation. The contextual suppression of a bar within a texture is

$$C^{\theta_1}_{\text{whole}-\text{texture}} \equiv \sum_a (g'_y(\bar{y}_{\theta_1}) W'^{\theta_1}_a - J'^{\theta_1}_a) g_x(\bar{x}_{\theta_1}) = (\mathcal{I} - \mathcal{E}) g_x(\bar{x}_{\theta_1}) > 0 \tag{20}$$

where $\bar{x}_{\theta_1}$ denotes the (translation invariant) fixed point for all texture bars. Consider the bars on the vertical axis $i = (m_i = 0, n_i)$. Removing the texture bars on the

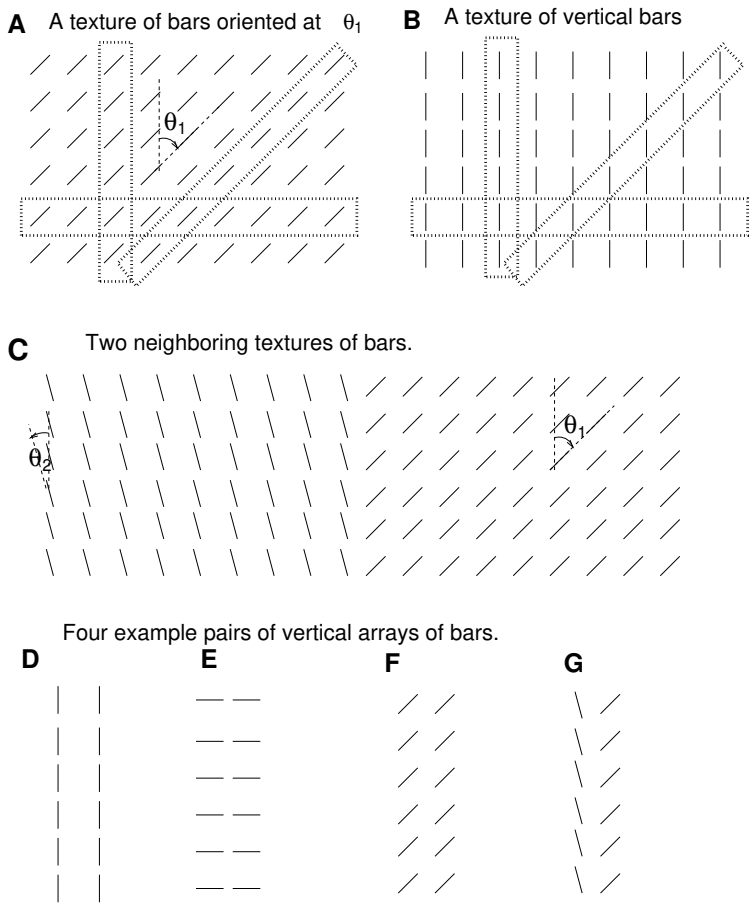

Figure 6: Examples of the two dimensional textures and their interactions. **A:** texture made of bars oriented at θ_1 and sitting on a Manhattan grid. This can be seen as a horizontal array of vertical array of bars. **B:** a special case of **A** when $\theta_1 = 0$. This is a horizontal array of vertical lines. Each texture can also be seen as a vertical array of horizontal arrays of bars, or an oblique array of oblique arrays of bars. Each vertical, horizontal, or oblique array can be viewed as a single entity, shown as examples in the dotted boxes. **C:** Two nearby textures and the boundary between them. **D, E, F**: examples of nearby and identical vertical arrays. **G:** two nearby but different vertical arrays. When each vertical array is seen as an entity, one can calculate effective connections J' and W' (defined in the text) between these vertical arrays.

left $i = (m_i < 0, n_i)$ removes the contextual suppression from them, and so gives them higher responses. This highlights the texture boundary $m_i = 0$. Now the activity $\bar{x}_{i\theta_1}$ depends on m_i, i.e., the distance of the bars from the texture boundary. As $m_i \to \infty$, $\bar{x}_{i\theta_1} \to \bar{x}_{\theta_1}$. The contextual suppression of the bars on the boundary, $m_i = 0$, is

$$C^{\theta_1}_{\rm half-texture} \equiv \sum_{m_j \geq 0} (g'_y(\bar{y}_{i\theta_1}) W'^{\theta_1}_{m_j} - J'^{\theta_1}_{m_j}) g_x(\bar{x}_{j\theta_1}) \tag{21}$$

$$\approx \sum_{a \geq 0} (g'_y(\bar{y}_{\theta_1}) W'^{\theta_1}_{a} - J'^{\theta_1}_{a}) g_x(\bar{x}_{\theta_1}) < C^{\theta_1}_{\rm whole-texture}, \tag{22}$$

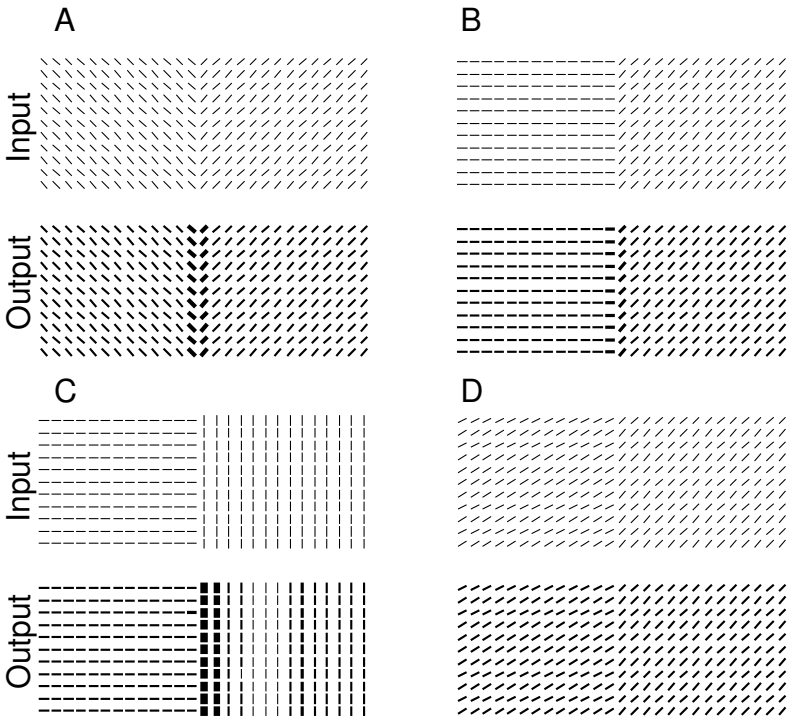

Figure 7: Simulated examples of texture boundary highlights between different pairs of textures, defined by bar orientations. In each example, we show the input image $I_{i\theta}$ above the output image $g_x(x_{i\theta})$ averaged in time. Each image shows a small region out of an extended input area. **A:** $\theta_1 = 45^o$, $\theta_2 = -45^o$. **B:** $\theta_1 = 45^o$, $\theta_2 = 90^o$. **C:** $\theta_1 = 0^o$, $\theta_2 = 90^o$. **D:** $\theta_1 = 45^o$, $\theta_2 = 60^o$. The texture border is vertical in the middle of each stimulus pattern. Note how border highlights increase with increasing orientation contrast $\theta_1 - \theta_2$. The orientation contrast of 15^o in **D** is difficult to detect by the model or humans. The orientation contrast $\theta_1 - \theta_2 = 90^o$ for both **A** and **C**. Note how the responses to the boundary bars decrease with increasing orientation differences between the bars and the boundary.

where we approximate $(\bar{x}_{j\theta_1}, \bar{y}_{j\theta_1}) \approx (\bar{x}_{\theta_1}, \bar{y}_{\theta_1})$ for all $m_j \geq 0$.

The boundary highlight persists when a neighboring, different, texture of bars oriented at θ_2 for $i = (m_i < 0, n_i)$ is present (Fig. (6C)). To analyze this, define connections between arrays in different textures (Fig. (6G)) as

$$J'^{\theta_1\theta_2}_a \equiv \sum_{j, m_j = m_i + a} J_{i\theta_1 j\theta_2} \qquad W'^{\theta_1\theta_2}_a \equiv \sum_{j, m_j = m_i + a} W_{i\theta_1 j\theta_2} \tag{23}$$

When $\theta_1 = \theta_2$, $J'^{\theta_1\theta_2}_a = J'^{\theta_1}_a$ and $W'^{\theta_1\theta_2}_a = W'^{\theta_1}_a$. The contextual suppression from the neighboring texture (θ_2) on the texture boundary ($m_i = 0$) is $C^{\theta_1,\theta_2}_{\rm neighbor-half-texture} \equiv \sum_{m_j<0}(g'_y(\bar{y}_{i\theta_1})W'^{\theta_1\theta_2}_{m_j} - J'^{\theta_1\theta_2}_{m_j})g_x(\bar{x}_{j\theta_2})$. For the association field connection, $J_{i\theta_1,j\theta_2}$ and $W_{i\theta_1,j\theta_2}$ tend to link similarly oriented bars $\theta_1 \sim \theta_2$, we have $C^{\theta_1,\theta_2}_{\rm neighbor-half-texture}$ minimum or zero when $\theta_1 \perp \theta_2$ and increasing with decreasing $|\theta_1 - \theta_2|$. Hence, the boundary highlight is expected to increase with the orientation contrast $|\theta_1 - \theta_2|$. The net contextual suppression on the border, contributed by both textures, is $C^{\theta_1,\theta_2}_{\rm 2-half-textures} \equiv C^{\theta_1}_{\rm half-texture} + C^{\theta_1,\theta_2}_{\rm neighbor-half-texture}$. Hence, the border enhancement, or the reduction of contextual suppression at the border relative to regions further inside the texture is

$$\delta C \equiv C^{\theta_1}_{\rm whole-texture} - C^{\theta_1,\theta_2}_{\rm 2-half-texture} \tag{24}$$

$$\approx C^{\theta_1,\theta_2=\theta_1}_{\text{neighbor-half-texture}} - C^{\theta_1,\theta_2}_{\text{neighbor-half-texture}} \quad (25)$$

$$\approx \sum_{a<0}(g'_y(\bar{y}_{\theta_1})W'^{\theta_1}_a - J'^{\theta_1}_a)g_x(\bar{x}_{\theta_1}) - \sum_{a<0}(g'_y(\bar{y}_{\theta_1})W'^{\theta_1\theta_2}_a - J'^{\theta_1\theta_2}_a)g_x(\bar{x}_{\theta_2}) \quad (26)$$

Again, we made the approximation $\bar{x}_{j\theta_2} \approx \bar{x}_{\theta_2}$ for $m_j < 0$. Usually $\bar{x}_{\theta_2} \neq \bar{x}_{\theta_1}$ since the fixed point should depend on the relative orientation between the bars and the arrays (i.e., the axes). Assuming $J'^{\theta_1\theta_2}_a \approx 0$ and $W'^{\theta_1\theta_2}_a \approx 0$ when $|\theta_1 - \theta_2| = \pi/2$, and noting that $\bar{x}_{\theta_1} \approx \bar{x}_{\theta_2}$ when $\theta_1 \approx \theta_2$,

$$\delta C \approx \begin{cases} 0 & \text{for } \theta_1 \approx \theta_2 \\ \sum_{a<0}(g'_y(\bar{y}_{\theta_1})W'^{\theta_1}_a - J'^{\theta_1}_a)g_x(\bar{x}_{\theta_1}) > 0 & \text{for } \theta_1 \perp \theta_2 \\ \text{roughly increases} & \text{as } |\theta_1 - \theta_2| \text{ increases} \end{cases} \quad (27)$$

Thus the border highlight diminishes as the orientation contrast approaches 0, see Fig. (7). Furthermore, even at a given contrast $|\theta_1 - \theta_2|$, the border enhancement δC depends on θ_1. For instance, with $|\theta_1 - \theta_2| = \pi/2$ and the association field connections, the enhancement δC for border bars parallel to the border $\theta_1 = 0$ (which form a contour) is higher than that for border bars perpendicular to the border $\theta_1 = \pi/2$. This is because both the suppression $g'_y(\bar{y}_{\theta_1})W'^{\theta_1}_a - J'^{\theta_1}_a$ between parallel contours ($\theta_1 = 0$ and $a \neq 0$) and the facilitation $J'^{\theta_1}_0 - g'_y(\bar{y}_{\theta_1})W'^{\theta_1}_0$ within a contour (Fig. (6D)) are much stronger than their counterparts for the vertical arrays of horizontal bars (Fig. (6E)). Thus the strength of the border highlight is predicted to be tuned to the relative orientation θ_1 between the border and the bars (Li 2000). This explains the asymmetry in the outputs of Fig. (7C), the highlight of the vertical border is much stronger for the vertical than the horizontal texture bars.

Clearly, the approximations $\bar{x}_{i\theta_1} \approx \bar{x}_{\theta_1}$ for $m_i \geq 0$ and $\bar{x}_{i\theta_2} \approx \bar{x}_{\theta_2}$ for $m_i < 0$), which are used to arrive at equation (27), break down at the border, especially at more salient borders like that in Fig. 7C. This accentuates the tuning of the border highlight to θ_1.

Iso-orientation suppression underlies the border highlight, and by equation (20), its strength $\mathcal{I} - \mathcal{E}$ depends on contrast through $g'_y(\bar{y})$. Since $g'_y(\bar{y})$ usually increases with increasing $\bar{y}$, the highlight is stronger at higher contrast. Psychophysically, texture segmentation does require an input contrast well above the texture detection threshold (Nothdurft 1994). It is easy to tune the connection weights in the model quantitatively such that iso-orientation suppression holds at all input contrasts, or holds only at sufficient input contrast and becomes iso-orientation facilitation at very low contrast as in Li (1998, 1999a). Computationally, facilitation certainly helps texture detection, which at low input contrast could be more important than segmentation. On this note, contour facilitation ($F_{\text{contour}} > 0$) holds at all contrasts (Li 1998) using the bi-phasic connection, since no W connections link the contour segments. Non-bi-phasic connections should be employed to model diminished contour enhancement at high contrast (Sceniak et al 1999).

3.5 Translation invariance and pop-out

In the examples above, orientation contrasts are highlighted because they mark boundaries between textures composed of bars of single orientations. However,

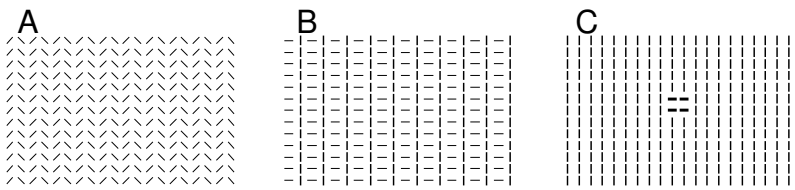

Figure 8: Model responses to homogeneous (**A, B**) and inhomogeneous (**C**) input images, each composed of bars of equal input contrasts. **A:** A homogeneous (despite of the orientation contrast) texture of bars of different orientations, a uniform output saliency results. **B:** Another homogeneous texture, vertical bars are more salient, however the whole texture has a translation invariant saliency distribution. **C:** The small figure pops out from the background because it is where translation invariance is broken in inputs, and the whole figure is its own boundary.

if orientation contrasts are homogeneous within the texture itself, they will not pop out. Fig. (8A) shows an example for which the texture is made of alternating columns of bars at $\theta_1 = 45^o$ (even a) and $\theta_2 = 135^o$ (odd a). The contextual suppression of a bar oriented at θ_1 is:

$$C_{\text{complex-texture}} = \sum_{\text{even } a} (g_y'(\bar{y}_{\theta_1}) W_a'^{\theta_1} - J_a'^{\theta_1}) g_x(\bar{x}_{\theta_1}) + \sum_{\text{odd } a} (g_y'(\bar{y}_{\theta_1}) W_a'^{\theta_1\theta_2} - J_a'^{\theta_1\theta_2}) g_x(\bar{x}_{\theta_2}) \tag{28}$$

Thus no bar oriented at θ_1 is less suppressed, or more salient, than other bars oriented at θ_1. Note that since $C_{\text{complex-texture}} \neq C^{\theta_1}_{\text{whole-texture}}$, the value of $\bar{x}_{\theta_1}$ is not the same as it would be in a simple texture of bars of a single orientation θ_1. This applies similarly to $\bar{x}_{\theta_2}$. For general θ_1 and θ_2, $\bar{x}_{\theta_1} \neq \bar{x}_{\theta_2}$. In Fig. (8A), reflection symmetry leads to $\bar{x}_{\theta_1} = \bar{x}_{\theta_2}$ or uniform saliency within the whole texture. In Fig. (8B), bars oriented at $\theta_1 = 0^o$ induced higher responses than those oriented at $\theta_2 = 90^o$. Nevertheless, looking at this texture which is defined by both the vertical and horizontal bars and their spatial arrangement, no local patch of the texture is more salient than any other patch. This translation invariance in saliency is simply the result of the network preserving the translation invariance in the input (texture), as long as the translation symmetry is not spontaneously broken (see section 3.7 for analysis).

A boundary between textures is one place where input is not translation invariant, and is highlighted by the cortical interactions. A special case of this is when one small texture patch is embedded in a large and different texture. The small texture is small enough that the whole texture is its own boundary, and thus pops out from the background (Fig. (8C)). In general, orientation contrasts do not correspond to texture boundaries and thus do not necessarily pop out. Through contextual influences, the highlight at a texture border can alter responses to nearby locations up to a distance comparable to the lateral connection lengths. Hence, the response to a texture region is not homogeneous unless this region is far enough away from the border. This is evident at the right side of the border in Fig. (7C). These effects are not to be confused with spontaneous symmetry breaking since they are generated by the input border and are local. See Li (2000) for more details about these effects and their physiological counterparts.

3.6 Filling-in and leaking-out

Small fragments of a contour or homogeneous texture can be missing in inputs due to input noise or to the visual scene itself. Filling-in is the phenomenon that the missing input fragments are not noticed, see Pessoa Thompson, and Noe (1998) for an extensive discussion. It could be caused by one of the following two possible mechanisms. The first is that, although the cells for the missing fragment do not receive direct visual inputs, they are excited enough by the contextual influences to fire as if there were direct visual inputs. (This is how (e.g.,) Grossberg and Mingolla (1985) model illusory contours.) The second possibility is that, even though the cells for the missing fragment do not fire, the regions near, but not at, the missing fragments are not salient or conspicuous enough to attract visual attention strongly. In the latter case, the missing fragments are only noticable by attentive visual scrutiny/search. It is not clear from physiology (Kapadia et al 1995) which mechanism is involved.

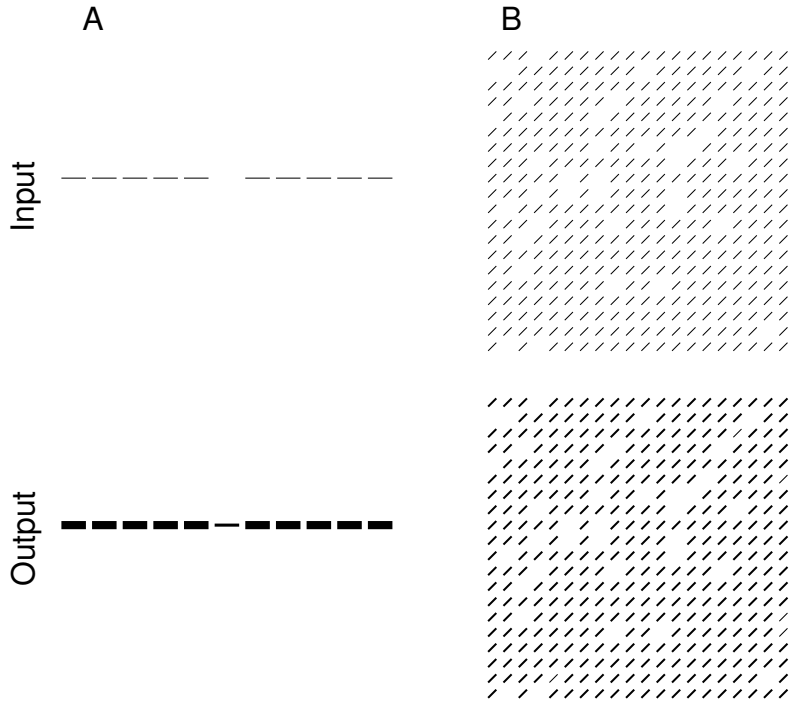

Figure 9: Examples of filling-in — model outputs from inputs composed of bars of equal contrasts in each example. **A:** A line with a gap, the response to the gap is non-zero, **B:** A texture with missing bars, the responses to bars near the missing bars are not significantly higher than the responses to other texture bars.

Consider a single bar segment $i = (m_i = 0, n_i = 0)$ missing in a smooth contour, say, a horizontal line like Fig. (9A), filling-in could be achieved by either of the two possible mechanisms. To excite the cell i to firing threshold T_x (such that $g_x(x_i > T_x) > 0$), contextual facilitation $\sum_j (J_{ij} - W_{ij} g'_y(\bar{y}_i)) g_x(\bar{x}_j)$ should be strong enough, or approximately

$$F_{\text{contour}} + I_o = (\mathcal{E} - \mathcal{I}) g_x(\bar{x}) + I_o > T_x \tag{29}$$

where I_o is the background input, F_{contour} and the effective net connections $\mathcal{E}$ and $\mathcal{I}$ are as defined in equations (15 - 18), and we approximate for all contour bars $(\bar{x}_j, \bar{y}_j)$ by $(\bar{x}, \bar{y})$, the translation invariant activity in a complete contour. If segments within a smooth contour facilitate each other's firing, then a missing fragment i reduces the saliencies of the neighboring contour segments $j \approx i$. The missing segment and its vicinity are thus not easily noticed, even if the cell i for the missing segment does not fire.

The cell $i = (m_i = 0; n_i = 0)$ should not be excited enough to fire if the left half $j = (m_j < 0, n_j = 0)$ of the horizontal contour are removed. Otherwise the contour extends beyond its end or grows in length — leaking out. To prevent leaking-out

$$F_{\text{contour}}/2 + I_o < T_x \tag{30}$$

since the contour facilitation to i is approximately $F_{\text{contour}}/2$, half of that F_{contour} in an infinitely long contour. The inequality (30) is satisfied for the line end in Fig. (5B), and should hold at any contour saliency $g_x(\bar{x})$. Not leaking out also means that large gaps in lines can not be filled in. To prevent leaking-out at $i = (m_i = 0, n_i = 1)$, the side of an infinitely long (e.g.,) horizontal contour on the horizontal axis in Fig. (4B) (thus to prevent the contour getting thicker), we require $\sum_{j \in \text{contour}} (J_{ij} - g'_y(\bar{y}_i) W_{ij}) g_x(\bar{x}) < T_x - I_o$ for $i \notin$ contour. This condition is satisfied in Fig. (5A).

Filling-in in a texture with missing fragments i (texture filling-in) is only feasible by the second mechanism — to avoid conspicuousness near i — since i can not be excited to fire if contextual input within a texture is suppressive. If i is not missing, its neighbor $k \approx i$ receives contextual suppression $(\mathcal{I} - \mathcal{E}) g_x(\bar{x}) \equiv \sum_{j \in \text{texture}} (g'_y(\bar{y}) W_{kj} - J_{kj}) g_x(\bar{x})$. A missing i makes its neighbor k more salient by the removal of its contribution $(W_{ki} g'_y(\bar{y}) - J_{ki}) g_x(\bar{x})$ to the suppression. This contribution should be a negligible fraction of the total suppression to ensure that the neighbors are not too conspicuous, i.e.,

$$g'_y(\bar{y}) W_{ki} - J_{ki} \ll (\mathcal{I} - \mathcal{E}) \equiv \sum_{j \in \text{texture}} (g'_y(\bar{y}) W_{kj} - J_{kj}). \tag{31}$$

This is expected when the lateral connections are extensive enough to reach sufficiently large contextual areas, i.e., when $W_{ki} \ll \sum_j W_{kj}$ and $J_{ki} \ll \sum_j J_{kj}$. Leaking-out is not expected outside a texture border when the contextual input from the texture is suppressive.

Note that filling-in by exciting the cells for a gap in a contour (equation (29)) works against preventing leaking-out (equation (30)) from contour ends. It is not difficult to build a model that achieves active filling-in. However, preventing the model from leaking-out and unwarranted illusory contours implies a small range of choices for the connection strengths J and W.

3.7 Hallucination prevention, and neural oscillations

To ensure that the model performs the desired computation analyzed in section 3.1-3.6, the mean or the fixed points $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ in these analysis should correspond

to the actual model behavior. (We use bold-faced character to represent vectors or matrices.) Section 2 showed that this is difficult to achieve in the symmetric networks, as the fixed points $(\bar{\mathbf{X}})$ for the desired computation are likely to be unstable, i.e., they are saddle points or local maximums in the energy function. In that case, the actual model output deviates drastically from the desired fixed point $(\bar{\mathbf{X}})$ or visual input, and, in particular, visual hallucinations occur. In the corresponding E-I network, the asymmetric connections between E and I give the network a tendency to oscillation around the fixed point. This oscillation enables our model to avoid the motion of $(\mathbf{X}, \mathbf{Y})$ towards hallucination (Li and Dayan 1999), making it possible to correspond the desired fixed points $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ with the (temporally averaged) model behavior. However, this correspondence is not guaranteed, and is in fact difficult to achieve without guided model design. It requires stability conditions on $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ to be satisfied, which constrains J and W, in addition to the conditions placed on J and W in section 3.1-3.6 for desired contour integration and texture segmentation (the inequalities (18), (20), (29), (30), and (31)). This section derives these stability constraints and their implications. Specifically, we derive the condition to prevent visual hallucinations or spontaneous formations of spatially inhomogeneous outputs given translation invariant visual inputs. Ermontrout and Cowan (1979) have analyzed the stability conditions to obtain hallucinations in a simplified model of V1 (see Bressloff et al (2000) for a more recent analysis), and studied the forms and dynamics of the hallucinations. Li (1998) analyzed the stability constraints to prevent hallucination under contour inputs. Here we generalize the analysis in Li (1998) to other homogeneous inputs, and in addition analyze the nonlinear dynamics of (non-hallucinating) homogeneous oscillations around $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$. We omit the analysis of the emergent hallucinations since the hallucinations are prevented by the model.

To analyze stability, we study how small deviations $(\mathbf{X} - \bar{\mathbf{X}}, \mathbf{Y} - \bar{\mathbf{Y}})$ from the fixed point evolve. Change variables $(\mathbf{X} - \bar{\mathbf{X}}, \mathbf{Y} - \bar{\mathbf{Y}}) \rightarrow (\mathbf{X}, \mathbf{Y})$. For small deviation $\mathbf{X}, \mathbf{Y}$, a Taylor expansion on equations (5) and (6) gives the linear approximation:

$$\begin{pmatrix} \dot{\mathbf{X}} \\ \dot{\mathbf{Y}} \end{pmatrix} = \begin{pmatrix} -1 + \mathbf{J} & -\mathbf{G}'_{\mathbf{y}} \\ \mathbf{G}'_{\mathbf{x}} + \mathbf{W} & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{32}$$

where $\mathbf{J}$, $\mathbf{W}$, $\mathbf{G}'_{\mathbf{x}}$, and $\mathbf{G}'_{\mathbf{y}}$ are matrices with $\mathbf{J}_{i\theta j\theta'} = J_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$ for $i \neq j$, $\mathbf{J}_{i\theta,i\theta} = J_o g'_x(\bar{x}_{i\theta})$, $\mathbf{W}_{i\theta j\theta'} = W_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$ for $i \neq j$, $\mathbf{W}_{i\theta,i\theta'} = 0$, $\mathbf{G}'_{\mathbf{x}i\theta j\theta'} = \delta_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$. and $\mathbf{G}'_{\mathbf{y}_{i\theta j\theta'}} = \delta_{ij}\psi(\theta - \theta') g'_y(\bar{y}_{j\theta'})$ where $\psi(0) = 1$. To focus on the output $\mathbf{X}$, eliminate (hidden) variable $\mathbf{Y}$:

$$\ddot{\mathbf{X}} + (\mathbf{2} - \mathbf{J})\dot{\mathbf{X}} + (\mathbf{G}'_{\mathbf{y}}(\mathbf{G}'_{\mathbf{x}} + \mathbf{W}) + \mathbf{1} - \mathbf{J})\mathbf{X} = 0 \tag{33}$$

Consider inputs of our interest which are bars arranged in a translation invariant fashion along a one or two dimensional array. For simplicity and approximation, we again omit bars outside the array and their associated quantities in equation (33), and omit the index θ, like we did in section 3.3 and 3.4. Translation symmetry implies $(\bar{x}_i, \bar{y}_i) = (\bar{x}, \bar{y})$, $\mathbf{G}'_{\mathbf{y}_{ij}} = \delta_{ij} g'_y(\bar{y})$, $\mathbf{G}'_{\mathbf{x}ij} = \delta_{ij} g'_x(\bar{x})$, $(\mathbf{G}'_{\mathbf{y}}\mathbf{G}'_{\mathbf{x}})_{ij} = g'_x(\bar{x}) g'_y(\bar{y})\delta_{ij}$, and $(\mathbf{G}'_{\mathbf{y}}\mathbf{W})_{ij} = g'_y(\bar{y})\mathbf{W}_{ij}$. Furthermore, $\mathbf{J}_{ij} = \mathbf{J}_{i+a,j+a} \equiv \mathbf{J}_{i-j}$ and $\mathbf{W}_{ij} = \mathbf{W}_{i+a,j+a} \equiv \mathbf{W}_{i-j}$ for any a. One can now go to the Fourier domain of the

spatial variables $\{\mathbf{X}_i\}$ and their associated quantities $\mathbf{J}, \mathbf{W}$ to obtain:

$$\ddot{\mathcal{X}}_k + (2 - J)\dot{\mathcal{X}}_k + (g_y'(\bar{y})(g_x'(\bar{x}) + \mathcal{W}_k) + 1 - \mathcal{J}_k)\mathcal{X}_k = 0 \tag{34}$$

where $\mathcal{X}_k, \mathcal{J}_k, \mathcal{W}_k$ are spatial Fourier transforms of $\mathbf{X}, \mathbf{J}, \mathbf{W}$ for frequency f_k such that $e^{\mathrm{i}f_k N} = 1$, where N is the size of the system. $\mathcal{X}_k$ is the amplitude of the spatial wave of frequency f_k in the deviation $\mathbf{X}$ from the fixed point, $\mathcal{J}_k = \sum_a \mathbf{J}_a e^{\mathrm{i}f_k a}$, and $\mathcal{W}_k = \sum_a \mathbf{W}_a e^{\mathrm{i}f_k a}$. $\mathcal{X}_k$ evolves as $\mathcal{X}_k(t) \propto e^{\gamma_k t}$ where

$$\gamma_k \equiv -1 + \mathcal{J}_k/2 \pm \mathrm{i}\sqrt{g_y'(g_x' + \mathcal{W}_k) - \mathcal{J}_k^2/4} \tag{35}$$

Denote $Re(\gamma_k)$ as the real part of γ_k, $Re(\gamma_k) < 0$ for all k makes any deviation $\mathbf{X}$ decay to zero, and hence no hallucination can occur. Otherwise, the mode with the largest $Re(\gamma_k)$, let it be $k = 1$, will dominate the deviation $\mathbf{X}(t)$. If this mode has zero spatial frequency $f_1 = 0$, then the dominant deviation is translation invariant and synchronized across space, and hence no spatially varying patterns can be hallucinated. Thus the conditions to prevent hallucinations are

$$Re(\gamma_k) < 0 \qquad \text{for all } k, \qquad \text{or} \qquad Re(\gamma_1)_{f_1=0} > Re(\gamma_k)_{f_k \neq 0} \tag{36}$$

When $Re(\gamma_1)_{f_1=0} > 0$, the fixed point is not stable, the divergent homogeneous deviation $\mathbf{X}$ is eventually confined by the threshold and saturation nonlinearity. It oscillates (synchronously) in time when $g_y'(g_x' + \mathcal{W}_1) - \mathcal{J}_1^2/4 > 0$ or when there is no other fixed point to which the system trajectory can approach. Denote this translation invariant oscillatory trajectory by $(x, y) = (x_i, y_i)$, which is the same for all i. Then,

$$\begin{aligned} \dot{x} &= -x - (g_y(y + \bar{y}) - g_y(\bar{y})) + \mathtt{J}_1(g_x(x + \bar{x}) - g_x(\bar{x})) \\ \dot{y} &= -y + (1 + \mathtt{W}_1)(g_x(x + \bar{x}) - g_x(\bar{x})) \end{aligned}$$

where $\mathtt{J}_1 = \mathcal{J}_1/g_x'(\bar{x})$ and $\mathtt{W}_1 = \mathcal{W}_1/g_x'(\bar{x})$. After converging to a finite oscillation amplitude, the trajectory $(x(t), y(t))$ is a closed curve in the (x, y) space. It oscillates with period T such that $(x(t + T), y(t + T)) = (x(t), y(t))$, and satisfies

$$\int_0^T dt[(1+\mathtt{W}_1)x(g_x(x+\bar{x})-g_x(\bar{x}))+y(g_y(y+\bar{y})-g_y(\bar{y}))] = \int_0^T dt\mathtt{J}_1(1+\mathtt{W}_1)(g_x(x+\bar{x})-g_x(\bar{x}))^2, \tag{37}$$

since over a cycle of the oscillation, the oscillation energy

$$\int_{\bar{x}}^{x+\bar{x}} (1 + \mathtt{W}_1)(g_x(s) - g_x(\bar{x}))ds + \int_{\bar{y}}^{y+\bar{y}} (g_y(s) - g_y(\bar{y}))ds, \tag{38}$$

(potential and kinetic energy, the two terms in the above expression) is dissipated and restored to a conservation, as the readers can verify. This is because the dissipation, on the left side of equation (37), is balanced by the self-excitation, on the right side of equation (37). At smaller oscillation amplitudes, the self-excitation dominates, as exemplified by the unstable fixed point; at larger amplitude, the

dissipation dominates because of the saturation and/or threshold behavior in self-excitation. Thus the oscillation trajectory converges to the balance of a periodic nonlinear oscillation.

Since $\mathbf{J}_a = \mathbf{J}_{-a} \geq 0$ and $\mathbf{W}_a = \mathbf{W}_{-a} \geq 0$, $\mathcal{J}_k$ and $\mathcal{W}_k$ are real and have maxima $Max(\mathcal{J}_k) = \sum_a \mathbf{J}_a$ and $Max(\mathcal{W}_k) = \sum_a \mathbf{W}_a$ for the zero frequency $f_k = 0$ mode. Many simple forms of $\mathbf{J}$ and $\mathbf{W}$ decay with f_k, for example, $\mathbf{J}_a \propto e^{-a^2/2}$ gives $\mathcal{J}_k \propto e^{-f_k^2/2}$. However, the dominant mode is determined by the value of $Re(\gamma_k)$, and may have $f_1 \neq 0$. In principle, given a model interaction J and W and a translation invariant input, whether it is arranged on a Manhattan grid or some other grid, $Re(\gamma_k)$ should be evaluated for all k to ensure appropriate behavior of the model or inequalities (36). In practice, the finite range of (J, W) and the discreteness and the (rotational) symmetry in the image grid implies that often only a finite, discrete, set of k needs to be examined.

Let us look at some examples using the bi-phasic connections. For 1-d contour input like that in Fig. (4B), $W_{ij} = 0$. Then $Re(\gamma_k) = Re(-1 + \mathcal{J}_k/2 \pm \mathrm{i}\sqrt{g'_y g'_x - \mathcal{J}_k^2/4})$ increases with $\mathcal{J}_k$, whose maximum occurs at the translation invariant mode $f_1 = 0$, and $\mathcal{J}_1 = \sum_j \mathbf{J}_{ij}$. Then no hallucination can happen, though synchronous oscillations can occur when enough excitatory connections J link the units involved. For 1-d non-contour inputs like Fig. (4C,E), $J_{ij} = 0$ for $i \neq j$, thus $\mathcal{J}_k = \mathbf{J}_{ii}$, $\gamma_k = -1 + \mathbf{J}_{ii}/2 \pm \mathrm{i}\sqrt{g'_y(g'_x + \mathcal{W}_k) - \mathbf{J}_{ii}^2/4}$. Hence $Re(\gamma_k) < -1 + \mathbf{J}_{ii} = -1 + J_o g'_x(\bar{x}) < 0$ for all k, since $-1 + J_o g'_x(\bar{x}) < 0$ is always satisfied (otherwise an isolated principal unit x, which follows equation $\dot{x} = -x + J_x g_x(x) + I$, is not well behaved). Hence there should be no hallucination or oscillation.

Under 2-dimensional texture inputs, frequency $f_k = (f_x(k), f_y(k))$ is a wave vector perpendicular to the peaks and troughs of the waves. When $f_k = (f_x(k), 0)$ is in the horizontal direction, $\mathcal{J}_k = g'(\bar{x}) \sum_a J'_a e^{\mathrm{i} f_x(k) a}$ and $\mathcal{W}_k = g'(\bar{x}) \sum_a W'_a e^{\mathrm{i} f_x(k) a}$, where J'_a and W'_a are the effective connections between two texture columns as defined in equation (19). Hence, the texture can be analyzed as a 1-dimensional array as above, substituting bar-to-bar connections (J, W) by column-to-column connections (J', W'). However, J' and W' are stronger, have a more complex Fourier spectrum $(\mathcal{J}_k, \mathcal{W}_k)$, and depend on the orientation θ_1 of the texture bars. Again use the bi-phasic connection for examples. When $\theta_1 = 90^o$ (horizontal bars), W'_b is weak between columns, i.e., $W'_b \approx \delta_{b0} W'_0$ and $\mathcal{W}_k \approx W'_0$. Then, $Re(\gamma^k)$ is largest when $\mathcal{J}_k$ is, at $f_x(k) = 0$ — a translation invariant mode. Hence, illusory saliency waves (peaks and troughs) perpendicular to the texture bars are unlikely. Consider however vertical texture bars for the horizontal wave vector $f_k = (f_x(k), 0)$. The bi-phasic connection gives nontrivial J'_b and W'_b between vertical columns, or non-trivial dependences of $\mathcal{J}_k$ and $\mathcal{W}_k$ on f_k. The dominant mode with the largest $Re(\gamma_k)$ is not guaranteed to be homogeneous, and J and W must be tuned in order to prevent hallucination.

Given a non-hallucinating system and under simple or translation invariant inputs, neural oscillations, if they occur, can only be synchronous and homogeneous (i.e., identical) among the units involved, i.e., $f_1 = 0$. Since $\gamma^1 = -1 + \mathcal{J}_1/2 \pm \mathrm{i}\sqrt{g'_y(g'_x + \mathcal{W}_1) - \mathcal{J}_1^2/4}$, and $\mathcal{J}_1 = \sum_j \mathbf{J}_{ij}$ for $f_1 = 0$, the tendency for oscillation increases with increasing excitatory-to-excitatory links J_{ij} between units

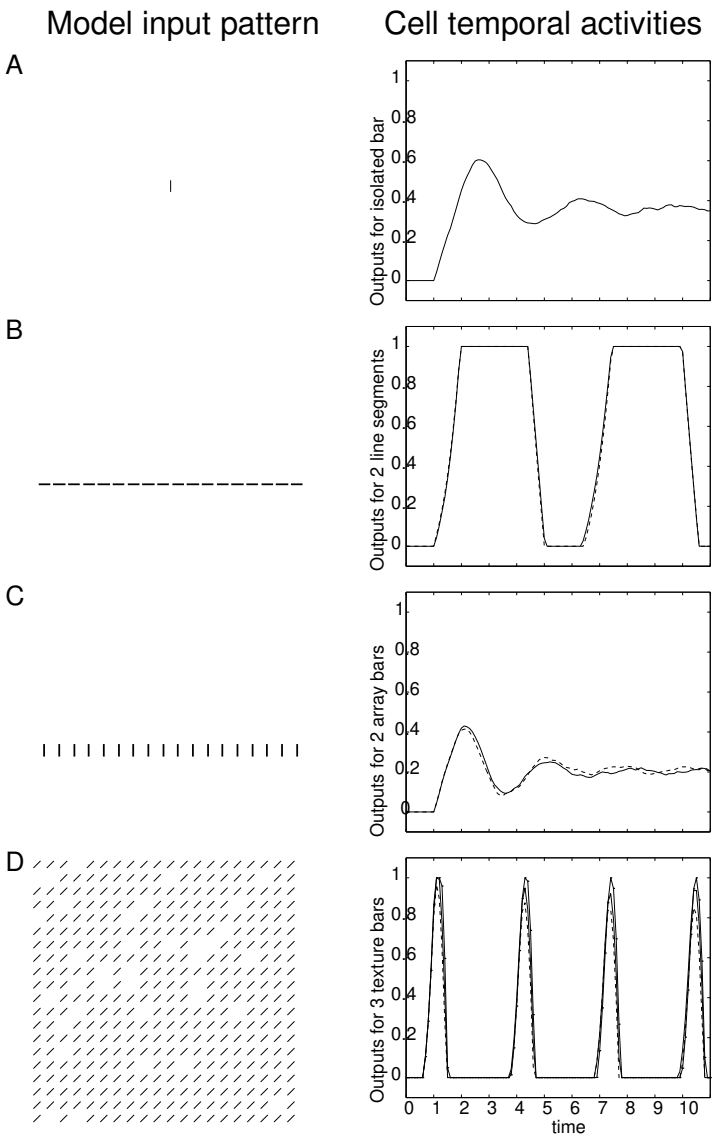

Figure 10: Different stimuli have different tendencies to cause oscillatory responses. The pictures on the left show visual stimuli (all appear at time zero and stay on), and the graphs on the right time course of the neural activities. **A** an isolated bar, the neural response to which stablizes after an initial oscillatory transient. **B** An input contour and the synchronized and sustained oscillatory neural responses of two non-neighboring neurons, all neurons corresponding to the contour segments respond similarly. **C:** A horizontal array of vertical bars, and the responses (decaying oscillations towards static values) of two non-neighboring neurons. **D:** An input texture (with some holes in it), and the sustained oscillatory responses of three neurons, whose spatial (horizontal, vertical) coordinates are (2, 2) (solid curve), (15, 2) (dotted curve), and (5, 9) (solid-dotted curve). The coordinate of the bottom left texture bar is (0, 0). Note that the responses to bars next to the holes in the textures are a little higher.

involved (Koenig and Schillen 1991). Hence, this tendency is likely to be higher for 2-d texture inputs than for 1-d array inputs, and lowest for a single, small, bar input. This may explain why neural oscillations are observed in some physiological experiments and not others. Under the bi-phasic connections, a long contour input is more likely to induce oscillation than an input of non-contour 1-d array, see Fig. (10). These predictions can be physiologically tested. Indeed, physiologically, grating stimuli are more likely to induce oscillations than bar stimuli (Molotchnikoff, Shumikhina, and Moisan, 1996).

4 Summary and Discussion

In this paper, we have argued that a recurrent model composed of interacting E-I pairs is a suitable minimal model of the primary visual cortex performing preattentive computation of contour integration and texture segmentation. We analyze the nonlinear input-output transform $I \rightarrow g_x(x)$ and the stability and temporal dynamics of the model. We derived design conditions on the intracortical connections such that (1) $I \rightarrow g_x(x)$ performs the desired computations, and (2) no hallucinations occur. Such an understanding has been essential to reveal the computational potential and limitations of the models, and led to a successful design (Li 1998, 1999a). The analysis techniques in this paper can be applied to other recurrent networks whose neural connections are translationally symmetric.

Note that the design conditions for a functional model can be satisfied by many quantitatively different models with qualitatively the same architecture. The model by Li (1998, 1999a) is one of them, and interested readers can find quantitative comparisons between the behavior of that model and experimental data. Although the behavior of Li's model agrees reasonably well with experimental data, there must be better and quantitatively different models. In particular, as discussed in this paper, non-bi-phasic connections (unlike those in Li's model) could be more computationally flexible, and thus account for additional experimental data. Emphasizing analytical tractability and a minimal design, this paper does not survey on other visual cortical models which have more elaborate structures and components. (See Li 1998, 1999a for such a survey.) One example is Grossberg and Raizada's recent model of V1 and V2, which evolved from earlier versions by Grossberg and Mingolla (1985), in which, in particular, the neuron units are not differentiated into excitatory and inhibitory ones.

In this paper, we have shown an example of how nonlinear neural dynamics link computations with the model architecture and neural connections. Additional or different computational goals, including the ones which maybe performed by the primary visual cortex and not yet modelled by our model example, might call for a more complex or different design. For example, our model lacks an end-stopping mechanism for V1 neurons. Such a mechanism could highlight the ends of, or gaps in, a contour, which in our model induce decreased responses (relative to the rest of the contour) due to reduced contour facilitation (Li 1998). Highlighting the line ends can be desirable, especially under high input contrasts when the gaps are clearly not due to input noise, and both the gaps and ends of con-

tours can be behaviorally very meaningful. Without end-stopping, our model is fundamentally limited in performing these computations. Our model also does not generate subjective contours like the ones that form the Kanizsa triangle or the Ehrenstein illusion (which could enable a perception of a circle whose contour connects the interior line ends of bars in Fig. (4E)). Evidence (von der Heydt et al 1984) suggests that area V2, rather than V1, is more likely to be responsible for these subjective contours, and this is addressed by models by Grossberg and colleagues (Grossberg and Mingolla 1985, Grossberg and Raizada 2000). Another desired computation is to generalize the notion of "translation invariance" to prevent the spontaneous saliency differentiation even when the input is not homogeneous in the image plane but is generated from a homogeneous flat texture surface slanted in depth. This will require multiscale image representations and recurrent interactions between cells tuned to different scales. By studying the recurrent nonlinear dynamics and analyzing the link between the structure and computation of a model, we hope to be able to better understand the computations in the primary visual cortex and in other visual or non-visual cortical areas where recurrent network dynamics play important roles.

Acknowledgement I am very grateful to Peter Dayan for conversations and discussions, and his very helpful comments on the drafts of this paper. Comments and suggestions from the anonymous reviewers of this paper are also very much appreciated. This work is supported in part by the Gatsby Foundation.

References

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) "Theory of orientation tuning in visual cortex. " *Proc Natl Acad Sci USA* 92(9):3844-8

Braun J., Niebur E., Schuster H. G., and Koch C. (1994) "Perceptual contour completion: A model based on local, anisotropic, fast-adapting interactions between oriented filters". Society for Neuroscience Abstracts, v.20, n.1-2, 1665.

Bressloff P.C., Cowan J. D., Golubitsky M., Thomas P.J., and Wiener M. C. (2000) "Geometric visual hallucinations, euclidean symmetry, and the functional architecture of striate cortex" Submitted to *Phil. Trans. Royal Soc. B.*

Cohen M A and Grossberg S. (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Systems., Man Cybern.* **13** 815-26.

Eckhorn R., Bauer R., Jordan W., Brosch M., Kruse W., Munk M., and Reitboeck H. J. (1988) " Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analysis in the cat." *Biol. Cybern.* 60:121-130.

Ermentrout G. B. and Cowan J.D. (1979) A mathematical theory of visual hallucination patterns *Biol. Cybern.* **34** 137-50.

Field D.J., Hayes A., and Hess R.F. (1993). "Contour integration by the human visual system: evidence for a local 'associat ion field'" *Vision Res.* 33(2): 173-93.

Gallant J.L., van Essen D. C. , and Nothdurft H. C. (1995) "Two-dimensional and three-dimensional texture processing in visual cortex of the macaque monkey" In *Early vision and beyond* eds. T. Papathomas, Chubb C, Gorea A., and Kowler E. (MIT press), pp 89-98.

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

Gilbert C.D. and Wiesel T.N. (1983) "Clustered intrinsic connections in cat visual cortex." *J Neurosci.* **3**(5), 1116-33.

Gray C.M. and Singer W. "Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex" (1989) *Proc. Natl. Acad. Sci. USA* 86: 1698-1702.

Grossberg S. and Mingolla E. (1985) "Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations" *Percept Psychophys.* **38** (2), 141-71.

Grossberg S, Raizada RD (2000) "Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex" *Vision Res.* 40(10-12):1413-32

Hirsch J. A. and Gilbert C. D. (1991) "Synaptic physiology of horizontal connections in the cat's visual cortex." *J. Neurosci.* **11**(6): 1800-9.

Hopfield, JJ (1982). Neural networks and systems with emergent selective computational abilities. *Proc. Natl. Acad. Sci. USA* **79** 2554-8.

Kapadia, M.K., Ito.M., , Gilbert, C. D.,, and Westheimer G. (1995) "Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys." *Neuron*. **15**(4), 843-56.

Kapadia, M.K., (1998) private communication.

Knierim J.J., and van Essen D.C. (1992) " Neuronal responses to static texture patterns ion area V1 of the alert macaque monkeys." *J. Neurophysiol.* **67**, 961-980.

Koenig P. and Schillen T.B. (1991) Stimulus-dependent assembly formation of oscil-

latory responses: I. Synchronization. Neural Computation 3:155-166

Li, Zhaoping (1997) "Primary cortical dynamics for visual grouping" in *Theoretical aspects of neural computation* Eds. Wong, K.Y.M, King, I, and D-Y Yeung, Springer-Verlag.

Li Zhaoping (1998) "A neural model of contour integration in the primary visual cortex" in *Neural Computation* 10, 903-940.

Li Zhaoping (1999) Visual segmentation by contextual influences via intracortical interactions in primary visual cortex. in *Network, Computation in neural systems* Vol 10, p.187-212.

Li Zhaoping (2000) Pre-attentive segmentation in the primary visual cortex. *Spatial Vision* Vol. 13, p.25-50.

Li Zhaoping (1999b) "Contextual influences in V1 as a basis for pop out and asymmetry in visual search " in *Proc. Natl. Acad. Sci. USA* 96(18) p.10530-10535

Li Zhaoping (2000) "Can V1 mechanisms account for figure-ground and medial axis effects?" in *Advances in Neural Information Processing Systems* 12 MIT Press, Cambridge, MA, Editors: SA Solla, TK Leen & K-R Muller P. 136-142.

Li, Zhaoping and Dayan P. (1999) "Computational differences between asymmetrical and symmetrical networks" *Network: Computation in Neural Systems* Vol. 10, 1, 59-77.

Molotchnikoff S, Shumikhina S, Moisan LE (1996) "Stimulus-dependent oscillations in the cat visual cortex: differences between bar and grating stimuli." *Brain Res* 731(1-2):91-100

Mundel T. Dimitrov A, and Cowan JD (1996) "Visual cortex circuitry and orientation tuning" in *Advances in Neural Information Processing Systems 9* p 887-93, MIT Press.

Nothdurft H. C. (1994) "Common properties of visual segmentation" in *Higher-order processing in the visual system* eds. Bock G. R., and Goode J. A. (Wiley & Sons), pp245-268

Pessoa L. Thompson E, Noe A (1998) "Finging out about filling-in: a guide to perceptual completion for visual science and the philosophy of perception." *Behav. Brain Sci.* 21(6) 723-48.

Polat U. and Sagi D. (1993) Lateral interactions between spatial channels: suppres-

sion and facilitation revealed by lateral masking experiments. *Vis. Res.* 33, p993-999.

Rockland K.S., and Lund J.S. (1983) " Intrinsic Laminar lattice connections in primate visual cortex" *J. Comp. Neurol.* **216**, 303-318

Sceniak MP, Ringach DL, Hawken MJ, Shapley R (1999) "Contrast's effect on spatial summation by macaque V1 neurons." *Nature Neurosci* 2(8):733-9

Sengpiel R., Baddeley R. Freeman T., Harrad R., and Blakemore C. (1995) *Soc. Neurosci. Abstr.* 21, 1629.

Shepherd G.M. (1990) *The synaptic organization of the brain* 3rd edition, Oxford University Press.

Somers DC, Todorov EV, Siapas AG, Toth LJ, Kim DS, Sur M (1998) A local circuit approach to understanding integration of long-range inputs in primary visual cortex. *Cereb. Cortex* 8(3) 204-17.

Stemmler, M., Usher, M. and Niebur, E. (1995). Lateral Interactions in Primary Visual Cortex: A Model Bridging Physiology and Psychophysics. *Science*, 269 : 1877-1880.

Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL (1997) "Paradoxical effects of external modulation of inhibitory interneurons" in *J Neurosci* 17(11):4382-8

von der Heydt R, Peterhans E, Baumgartner G (1984) "Illusory contours and cortical neuron responses. " *Science* 224(4654) 1260-2.

White E. L. (1989) *Cortical circuits* (Birkhauser).

Yen S-C. and Finkel L.H. (1998) "Extraction of perceptually salient contours by striate cortical networks" *Vision Res.* 38(5): 719-41

Zhang, K (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience* **16**:2112-2126.

Zucker S. W., Dobbins A. and Iverson L. (1989) "Two stages of curve detection suggest two styles of visual computation" *Neural Computation* 1, 68-81