February 9, 2000

GCNU TR 2000–001

Computational design and nonlinear dynamics of recurrent network models of the primary visual cortex

Zhaoping Li
Gatsby Unit

Abstract

The recurrent neural interaction in the primary visual cortex makes its outputs complex nonlinear functions of its inputs. This nonlinear transform serves the role of pre-attentive visual segmentation, i.e., the autonomous transformation from visual inputs to processed outputs that selectively emphasize certain features (e.g., pop-out features) for segmentation. Understanding the nonlinear dynamics of the neural circuit is a key to appreciating the cortical computational potential and tasks. However, the complex nonlinear dynamics of recurrent networks makes it extremely difficult to build a well-behaved and computationally functional model of the cortex merely by simulation trials. This paper describes an analytical study of the recurrent neural dynamics. We derive requirements on the neural architecture, components, and connection weights of a biologically plausible model of the cortex to achieve simultaneously different components of pre-attentive segmentation: region segmentation, figure-ground segregation, and contour enhancement. In addition, we analysis conditions for behaviors such as neural oscillations, illusory contours, and visual halluciations. Many of our analytical techniques can be applied to other recurrent networks with translation invariant neural and connection structures.

Computational design and nonlinear dynamics of recurrent network models of the primary visual cortex

Zhaoping Li
Gatsby Unit

1 Introduction

Recurrent neural dynamics is a basic computational substrate for cortical processing. The primary visual cortex is an example where recurrent dynamics is enabled by intra-cortical finite range lateral connections between neurons. The cortical input is the retinal image filtered through the cortical receptive fields (RFs) shaped like small edges or bars and retinotopically distributed in the visual space. The cortical outputs given the inputs are the cell activities or outputs under the recurrent interactions. In isolation, this cortical network can be seen as an autonomous system that transforms (nonlinearly) any cortical inputs to a different or processed output, namely the responses of the neurons to the input. Two immediate characteristics of this transform follow. 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 computation can doubtlessly be performed when V1 interacts dynamically with other visual areas, however, this is beyond the scope of this paper. Second, the recurrent dynamics enables a *global* scale computation to occur as, despite the local connectivities, the output of a V1 cell depends non-locally on its inputs in a way that it is hard to achieve in non-recurrent networks with only local connections.

Understanding the neural dynamics by computational modeling is essential to understanding the underlying signal transformations and thus the computational power. In this paper, we do not deal in depth with neural properties and behaviors (e.g., tuning properties of single units) that are not as directly concerned with global scale visual computation. Physiological and psychophysical data suggest that the pre-attentive computation in the primary cortex includes tasks 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 achieve these tasks, V1 functions as a saliency circuit to give higher responses to locations of higher saliencies in inputs, such as borders between texture regions, pop-out figures against backgrounds, and smooth contours (Li 1998, 1999a, 1999c). This saliency computation serves the purpose of pre-attentive segmentation. It is known to be a very difficult computation, especially considering that the same cortical circuit needs to achieve both contour enhancement and region or figure/ground segmentations, and that there is still no known general solution to segmentation after decades of research in machine and natural visual algorithms. There have been various models of the cortex which aim to

model particular components of the cortical computation, such as contour enhancement, i.e., relatively higher neural activities for cells receiving inputs arising from bars belonging to smooth contours (Grossberg and Mingolla 1985, Zucker, Dobbins, Iverson 1989, Yen and Finkel 1996). It is already very hard merely to successfully model the contour enhancement task in the cortex (Li 1998). Until recently, no models of the primary visual cortex are successful in capturing in a single model both the contour enhancement and texture segmentation components in the pre-attentive computation. In this paper, we describe the analytical study of the nonlinear recurrent dynamics to enable computational design of the recurrent network model to simultaneously achieve these components of the primary cortical computation.

Different neural circuits in different cortical areas share many common properties of neural connections, elementary operations, and the canonical microcircuit (Shepherd 1990). Some of our analytical techniques, e.g., the analysis of the cortical microcircuit, can be applied to other models of recurrent networks. It will be clear that much of our analysis relies on the fact that the primary visual cortex is a translation invariant system in the sense that the neural properties and connection patterns do not depend very much on the location of the neurons in the cortex. Many of our analytical techniques can be applied to other recurrent networks with translation invariant properties.

2 A minimal model of the primary visual cortex

In this section we introduce a model of the neural circuit in the cortex. Throughout the paper, when discussing general characteristics of the recurrent dynamics, we try to keep our analysis general. However, specific model details are needed to demonstrate particular analytical results or to illustrate some approximation and simplification techniques. For this, we use the model of V1 whose specifics and numerical parameters are available (Li 1998, 1999a), so that the readers can try out our simulations.

A minimal model of the cortex is the one which has enough components for necessary computations but not the excess details which are not essential for the computations concerned. 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 desired computation, but for which simplified versions fail.

We adopt a level of description for which a model neuron is described by its membrane potential x and its firing rate $g_x(x)$, which is a sigmoid-like function of x. In the model, visual cortical cells have their orientative selective receptive fields, which are arranged on a regular 2-dimensional array in image coordinates, indexed by $i = (m_i, n_i)$ where m_i and n_i are the horizontal and vertical coordinates. At each array point i, there is a group of cells, one each for K preferred orientations $\theta = k\pi/K$ for $k = 0, 1, ..., K - 1$ spanning 180°. Cell $i\theta$ has its RF located at i and preference for orientation θ. We further simplify matters by focusing only on the cells in layer 2-3 in the cortex, which are mainly responsible for the recurrent dynamics. Let the external visual inputs to cell $i\theta$ be $I_{i\theta}$. This is the result of processing the visual image through the RF of the cell. The responses $g_x(x_{i\theta})$ from the cells are the results of both the external input $I_{i\theta}$ and the recurrent interaction in the neural circuit.

The desired computation $\{I_{i\theta}\} \to \{g_x(x_{i\theta})\}$ is to give higher responses $g_x(x_{i\theta})$ to input bars $i\theta$ of higher perceptual saliency. For instance, if a visual image consists of a collection of bars $\{i\theta\}$ and all of them have the same input strength $\{I_{i\theta}\}$, the responses $\{g_x(x_{i\theta})\}$ to them can have different strength $g_x(x_{i\theta})$ for different bars — they should be relatively higher (compared to responses to other bars in the image) if the bars $i\theta$ are part of isolated smooth contours, or are at boundaries of texture regions, or are targets against backgrounds. For a well behaved model, this computation requires that, 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 output level to every bar should be the same.

2.1 A less-than-minimal recurrent model of V1

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

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

where $-\alpha_x x_{i\theta}$ model the decay in membrane potential, and I_o is the background input. This is a recurrent network with units $x_{i\theta}$, inputs $I_{i\theta}$, outputs $g_x(x_{i\theta})$, and recurrent connections $T_{i\theta,j\theta'}$ linking 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 modified by the network interaction, making $g_x(x_{i\theta})$ dependent on input $I_{j\theta'}$ for $(j\theta') \neq (i\theta)$. The computation is defined by the input-output transform $\{I_{i\theta}\} \rightarrow \{g_x(x_{i\theta})\}$, which, in turn, is determined by the neural connections. Translation invariance in the connections means that $T_{i\theta,j\theta'}$ depends only on the vector $i - j$ and relative angles of this vector to the orientations θ and θ'. For visual computations, mirror symmetry is expected, hence, the connections are symmetric, $T_{i\theta,j\theta'} = T_{j\theta',i\theta}$. One might also wish to distinguish interactions within a hypercolumn, i.e., $i = j$, from interactions between non-overlapping receptive fields, i.e., $i \neq j$, so $T_{i\theta,j\theta'}$ could be characterized by two subsets $T_{i\theta,i\theta'}$ and $T_{i\theta,j\theta'}$ for $j \neq i$.

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 in these models 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 1984) 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

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

Empirically, this convergence behavior to attractors still holds when the network is a modified or distorted (e.g., imperfect symmetry in neural connections) version of that in equation (1).

The energy landscape is sculpted by the intrinsic connections $T_{i\theta,j\theta'}$ while the external inputs $I_{i\theta}$ shift the locations of the local minima. The energy minimum state (the fixed point) $\bar{x}_{i\theta}$ where $\partial E/\partial g_x(x_{i\theta}) = 0$ for all $i\theta$ is (when $I_o = 0$)

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

Without recurrent interactions ($T = 0$), this minimum $\bar{x}_{i\theta} = I_{i\theta}/\alpha_x$ is exactly a scaled and faithful copy of the visual input $I_{i\theta}$. However, when the interaction T is strong enough, the recurrent interaction term $-\frac{1}{2} \sum_{i\theta,j\theta'} T_{i\theta,j\theta'} g_x(x_{i\theta}) g_x(x_{j\theta'})$ pulls the energy minimum state towards ones that are shaped by T and are 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 $\lambda_{\mathbf{T}} > \alpha_x$ (here g_x' is the gain in the sigmoid function $g_x(.)$). For instance, when the input $I_{i\theta}$ is homogeneous or translation invariant such that $I_{i\theta} = I_{j\theta}$ for all $i = j$, a strong interactions T could 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., $g_x(x_{i\theta}) \not\propto I_{i\theta}$, is desirable to a limited degree since this is how a saliency circuit produces differential outputs (saliencies) $g_x(x_{i\theta})$ to input bars of different saliencies even if they have the same input strength $I_{i\theta}$. However, this unfaithfulness should be controlled and driven

by the nature, in particular, the spatial configuration, of the input pattern $\{I_{i\theta}\}$ (e.g., the smooth contours or figures against a background) or the nature of deviations of input pattern $\{I_{i\theta}\}$ from homogeneity or homogeneous saliency. However, it is computational undesirable to have spontaneous or non-input-driven network behavior, such as inhomogeneous outputs to homogenous texture inputs, since this would lead to visual hallucinations (Ermentrout and Cowan 1979).

To illustrate, consider an example when the interactions T only link cells that prefer the same orientation, i.e., $T_{i\theta,j\theta'} = 0$ when $\theta \neq \theta'$. This is an idealization from experimental indications (Gilbert and Wiesel 1983, Rockland and Lund 1983) that cortical lateral interactions tend of link cells prefering similar orientations. In this idealization, the network can be seen as multiple subnetworks each for a particular orientation θ with no interaction between subnetworks. Let us take just one of the subnetworks, and for convenience we will drop the subindex for its angle θ. Consider a very simple interaction T_{ij} between i and j to be local and center-surround in shape, i.e., self excitation and nearest neighbor inhibition, such that
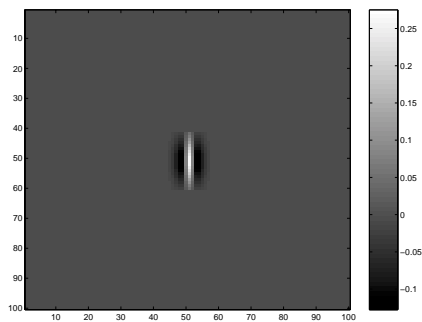
$$T_{ij} \propto \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } i \text{ and } j \text{ are nearest neighbors.} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

For instance, in a Manhattan grid $i = (m_i, n_i)$, the nearest neighbors of i are $(m_i \pm 1, n_i)$ and $(m_i, n_i \pm 1)$. With strong enough T, the network under constant and uniform external input can settle into an "antiferromagnetic" state for which neighboring units x_i exhibit one of the two different activities, e.g., x_i is high when $m_i + n_i$ is even and x_i is low otherwise. This energy minimum state pattern $\{x_i\}$ is just a spatial array of replicas of the center-surround local interaction pattern T.
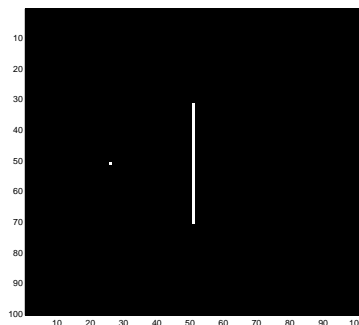
Consider another example, again in an idealization of a subnetwork of a single orientation, say vertical bars ($\theta = 0$). Now we consider the interaction T_{ij} that depends on the orientation of $i - j$ and is no longer rotationally invariant. The rotational invariance comes from both experimental indications and theoretical suggestions (e.g., Kapadia et al 1995, Polat and Sagi 1993, Field, Hayes, and Hess 1993) that i and j excite each other ($T_{ij} > 0$) when the two bars are co-aligned and inhibit each other when they are not co-aligned. (This also holds for cells tuned to similar, but not exactly the same, orientations, see Fig. (2B), and such interaction pattern has been called association field (Field et al 1993)). For vertical bars $\theta = 0$, T_{ij} takes the shape like that in Fig (1), $T_{ij} > 0$ between local and roughly vertically displaced i and j and $T_{ij} < 0$ between local and more horizontally displaced i and j. When such interactions are strong enough, the network is likely to settle into states consisting of stripy patterns (like finger-print or ocular dominance patterns) for which the orientation of the stripes is parallel to the associative interaction pattern T and the thicknesses of the stripes and the distances between neighboring stripes are approximately the spatial range or scale in T. Fig (1) shows a simple example when we look at an isolated subnet for $\theta = 0$, i.e., a network of units prefering vertical orientation. Although the system 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 the network state is uncompromising in such recurrent models. When the computation requires strong enough interaction T, for instance, to enhance contours and lines in inputs against background input noise using the associative field interaction pattern, the network is prone to "see" contours and lines (the stripes mentioned above) even when the input I suggests no lines to human eyes. This is called spontaneous pattern formation. To avoid this hallucination, the model has to weaken its capability to enhance contours. The computational capability of the network could be accessed by the maximum contour enhancement capability of the network without hallucination. Mathematical analysis (Li and Dayan 1999) shows that this symmetric network model of the cortex has a much reduced or insufficient computational capability than
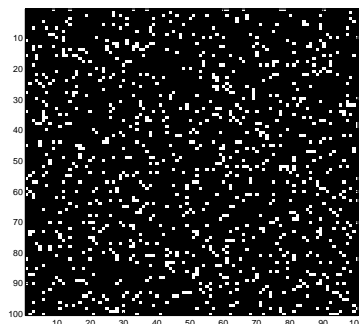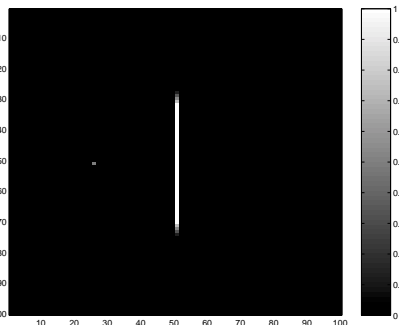
Figure 1: A submodel example describing a network of symmetrically connected cells tuned to vertical orientation ($\theta = 0$). Shown here are 5 gray scale images, each has a scale bar on the right. The network has 100x100 cells arranged in a 2-d array, with wrap around boundary condition. Each cell 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 to other cells. This pattern is local and translation invariant, it gives local colinear excitation between cells vertically displaced, but 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.

that of a more realistic model with an additional complexity which will be introduced below. In other words, although symmetric recurrent networks are useful for associative memory computations, 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 where greater faithfulness to visual input is required.

A property of this translation invariant recurrent network is that, in the continuous limit, different attractors (the energy minima) are connected. Given a translation invariant input $I_{i\theta}$ such that $I_{i\theta} = I_{j\theta}$ for all $i \neq j$, if state $\{x_{i\theta}\}$ is an attractor, so is a spatially translated state $\{x'_{i\theta}\}$ such that $x'_{i\theta} = x_{i+a,\theta}$ for any spatial translation a — one can easily varify that $\{x_{i\theta}\}$ and $\{x'_{i\theta}\}$ have the same energy value E. Hence, if the network hallucinates a stripe pattern or a spotted pattern (as in the "antiferromagnetic" case) under translation invariant input, the absolute (but not relative) positions of the spots and the stripes are random and can be shifted. When the translation happens over one dimension, such a continuum of attractors has been called "line attractor" (Zhang 1996). For two dimensional images, the continuum is a "surface attractor".

2.2 A minimal recurrent model with hidden units

The recurrent model above is apparently too impoverished to capture cortical computations. We need to add into the model the right neural elements or mechanisms to make the model computationally sufficient. 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 $\{I_{i\theta}\}$. Such attractor dynamics can not be removed by introducing ion channels or spiking neurons (rather than firing rate neurons), for instance, because the attractor behavior is largely dictated by the symmetric neural connections. For the same reason, mechanisms such as 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, also do not directly change the attractor dynamics. Attractor dynamics are untenable, however, in the face of the well established fact, called Dale's law, that a real neuron is either exclusively excitatory or exclusively inhibitory. Since it is impossible to have symmetric connections between excitatory and inhibitory neurons, a recurrent network with interactions between these two types of neurons will in general no longer support attractor dynamics. Thus we model the principal neurons $x_{i\theta}$ as exclusively excitatory pyramidal cells, and introduce inhibitory interneurons to mediate indirect, or disynaptic, inhibition between them, as in the real cortex (White 1989, Gilbert 1992, Rocklandand Lund 1983). The simplest step is to introduce one inhibitory interneuron $y_{i\theta}$ for each excitatory principal unit $x_{i\theta}$, and to connect each pair reciprocally. The principal units $x_{i\theta}$ still receive inputs $I_{i\theta}$ and send cortical outputs. The inhibitory units $y_{i\theta}$ are treated as hidden units in this recurrent network. The model is now described by equations:

$$\dot{x}_{i\theta} = -\alpha_x x_{i\theta} - g_y(y_{i,\theta}) + J_o g_x(x_{i\theta}) - \sum_{\Delta\theta \neq 0} \psi(\Delta\theta) g_y(y_{i,\theta+\Delta\theta})$$

$$+ \sum_{j \neq i, \theta'} J_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta} + I_o \tag{5}$$

$$\dot{y}_{i\theta} = -\alpha_y y_{i\theta} + g_x(x_{i\theta}) + \sum_{j \neq i, \theta'} W_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_c \tag{6}$$

where α_y and $g_y(y)$ model the inhibitory 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) the 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 interaction between $x_{i\theta}$ and $x_{j\theta'}$ (e.g., whether it is facilitation or inhibition) is determined by the combined action of $J_{i\theta,j\theta'}$ and $W_{i\theta,j\theta'}$. In stationary (but not in dynamic) states, the net effective connection between $x_{i\theta}$ and $x_{j\theta'}$ is, for example, $J_{i\theta,j\theta'} - W_{i\theta,j\theta'}/\alpha_y$ if $g_y(y) = y$. The term $\psi(\Delta\theta) \leq 1$ represents the local inhibitory connection at location i, and $J_o g_x(x_{i\theta})$ represents self excitation. Both $\psi(\Delta\theta)$ and J_o are merely explicit extensiations of the original interaction $T_{i\theta,i\theta'}$ between units within

a hypercolumn. Fig. (2C) schematically shows an example of the network. I_c and I_o are background inputs, including neural noise, feedback from higher centers, 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).

Because each cell has a finite orientation tuning width, 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. In V1, J and W tend to link cells preferring similar orientation $\theta \sim \theta'$ (Gilbert 1992, Rockland and Lund 1983). To implement the interaction resembling the association field, the J connections should be dominant between units preferring co-aligned bars ($\theta \sim \theta' \sim \angle(i-j)$), while the W connections should be dominant between units preferring non-aligned ones (Fig. (2B), Zucker et al 1989, Field et al 1993, Li 1998, 1999a). A simplest model of this interaction is to set $J > 0$ and $W = 0$ for mutual excitation and $J = 0$ and $W > 0$ for mutual inhibition, as in the model of Li (1998, 1999a). We will call such a connection pattern as in Fig. (2B) a bow-tie pattern due to the spatial arrangement of the J and W connections. However, having both $J_{i\theta,j\theta'} > 0$ and $W_{i\theta,j\theta'} > 0$ between two linked cells (as is the case in physiology, Hirsch and Gilbert 1991) and letting the ratio $J_{i\theta,j\theta'} : W_{i\theta,j\theta'}$ determine the overall sign of interaction gives extra computational flexibility.

A model with the extra complexity of Dale's law has been demonstrated to be able to capture the desired computations (Li 1999a). In this paper, we analyse recurrent dynamics in this minimal model of the cortex. We typically use the bow-tie connection (in Fig. (2B)) for examples.

3   Dynamic analysis

The model state can be characterized by $\{x_{i\theta}, y_{i\theta}\}$, or simply $\{x_{i\theta}\}$ if one ignores the hidden units $\{y_{i\theta}\}$. Interaction between excitatory and inhibitory cells makes the network states $\{x_{i\theta}(t)\}$ (as a function of time t) intrinsically oscillatory, and, given an input $\{I_{i\theta}\}$, the model does not guarantee attractor dynamics converging to a fixed point where $\dot{x}_{i\theta} = \dot{y}_{i\theta} = 0$. However, the model output after the initial transient following an input can be well defined by the temporal average $\{\bar{x}_{i\theta}\}$ of the network state $\{x_{i\theta}(t)\}$ if the state oscillates periodically in time or converges to a fixed point. A stationary state $\{x_{i\theta}(t \to \infty)\} = \{\bar{x}_{i\theta}\}$ is itself a fixed point, while an oscillatory state $\{x_{i\theta}(t \to \infty)\}$ oscillates around a fixed point which can often be approximated by the temporal average $\{\bar{x}_{i\theta}\}$. We will henceforth simply use the notation $\{\bar{x}_{i\theta}\}$ to denote both the fixed point and the temporal average. When the model is well behaved, there will be no spontaneous formation of network state $\{\bar{x}_{i\theta}\}$ that is not input driven. Then the computational performance of the model can be analyzed by studying the behavior of $\{g_x(\bar{x}_{i\theta})\}$, as we will do in sections 3.1-3.6. In section 3.7, we will study stability and dynamics around the fixed point $\{\bar{x}_{i\theta}\}$, to ensure that the model is well behaved, unlike the model with symmetrically connected units. This ensures that the behavior of $\{g_x(\bar{x}_{i\theta})\}$ we study in section 3.1-3.6 is a valid description of the model computation.

3.1   A single pair of neurons

Ignoring the recurrent interactions between different neural pairs, 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, for simplicity, we take $\alpha_x = \alpha_y = 1$ (as in the rest of the paper), omit indices $i\theta$, and denote $I = I_{i\theta} + I_o$. The fixed point $(\bar{x}, \bar{y})$ (where $\dot{x} = \dot{y} = 0$) under input (I, I_c) has the

A sampling location | One of the edge detectors

B Neural connection pattern.
Solid: J, Dashed: W

C Model Neural Elements

Edge outputs to higher visual areas

Inputs Ic to inhibitory cells

An interconnected neuron pair for edge segment i θ

Inhibitory interneurons

Excitatory neurons

Visual inputs, filtered through the receptive fields, to the excitatory cells.
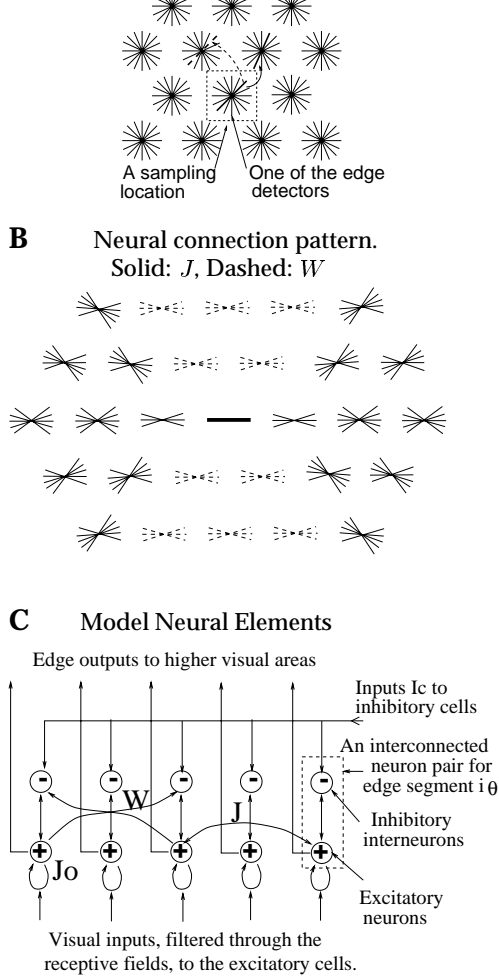
Figure 2: A schematic of the minimal model of the primary visual cortex. **A:** Visual inputs are sampled in a discrete grid by edge/bar detectors, modeling RFs for V1 layer 2-3 cells. Each grid point has K neuron pairs (see **C**), one per bar segment. All cells at a grid point share the same RF center, but are tuned to different orientations spanning 180°, 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.

input-output $(I, I_c \rightarrow g_x(\bar{x}))$ gain

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

$\frac{\delta g_x(\bar{x})}{\delta I}$ increases with $g'_x(\bar{x})$ but decreases with $g'_y(\bar{y})$. We only consider cases for which $\delta g_x(\bar{x})/\delta I \geq 0$ (otherwise, the pyramidal cell would fire vigorously even with zero input I). The input-output function $I \rightarrow g_x(\bar{x})$ depends on I_c — Fig. (3 A,B,C) presents an example for which

8

both $g_x(x)$ and $g_y(y)$ are piece-wise linear functions. The behavior of threshold, input gain control, and saturation are apparent. When this neural pair receives contextual inputs from other pairs $j\theta' \neq i\theta$ via connections J and W, (I, I_c) effectively changes to $(I + \Delta I, I_c + \Delta I_c)$. According to equation (9), $\Delta g_x(\bar{x}) \approx (\delta g_x(\bar{x})/\delta I)(\Delta I - g'_y(\bar{y})\Delta I_c)$. The overall contextual effect is excitatory if $\Delta I/\Delta I_c > g'_y(\bar{y})$ and inhibitory otherwise. In our example when $g'_y(\bar{y})$ increases with I (or I_c), the contextual inputs could switch from effectively facilitation to suppression as external input level I increases (Fig. (3 D)). This is the model's account of the physiological-ly observed dependence of the contextual influences on input contrast I (Sengpiel, Baddeley, Freeman, Harrad, and Blakemore 1995), which has also been modelled by others (Stemmler, Usher, Niebur 1995, Somers, Todorov, Siapas, Toth, Kim, Sur 1998).
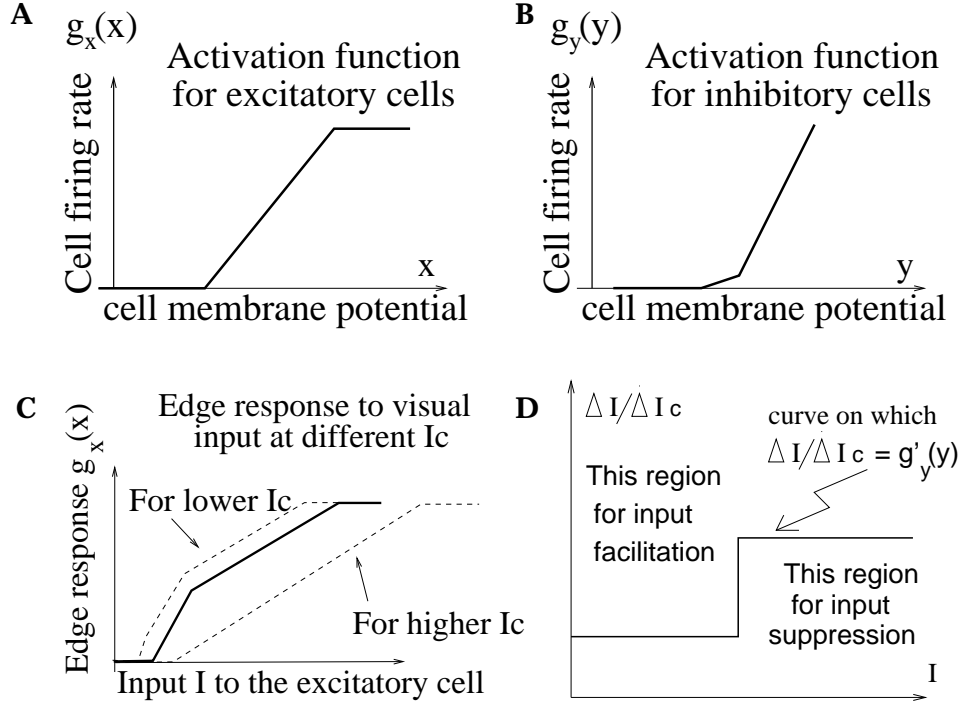

Figure 3: A,B: examples of $g_x(x)$ and $g_y(y)$ functions. C: Input-output function $I \to g_x(\bar{x})$ for an isolated neural pair without inter-pair neural interactions, under different levels of I_c. D: The overall effect of the external or contextual inputs $(\Delta I, \Delta I_c)$ on a neural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

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

Two pairs of neurons (i_1, θ_1) and (i_2, θ_2) with nearby but non-overlapping receptive fields can interact with each other via longer range connections $J_{i_1\theta_1,i_2\theta_2}$ and $W_{i_1\theta_1,i_2\theta_2}$. For simplicity, we use indices $a = 1, 2$ to denote the two pairs and their associated quantities, and denote the connection strength between the pair as $J_{12} = J_{21}$ and $W_{12} = W_{21}$. The equations of motion are

$$\dot{x}_a = -x_a - g_y(y_a) + I_a + I_o + J_o g_x(x_a) + J_{12} g_x(x_b)$$
$$\dot{y}_a = -y_a + g_x(x_a) + W_{12} g_x(x_b) + I_c$$

where $a, b = 1, 2$ and $a \neq b$. The contextual influence from input stimulus I_2 to neuron x_1 is monosynaptic excitation $J_{12} g_x(x_2)$ and disynaptic inhibition $W_{12} g_x(x_2)$ via inhibitory activi-

ties $g_y(y_1)$. The total effective connection from x_2 to x_1 directly and indirectly, according to the gain functions $\delta g_x(\bar{x})/\delta I$ and $\delta g_x(\bar{x})/\delta I_c$, is $J_{12} - g'_y(\bar{y}_1)W_{12}$.

In a simple case when $I \equiv I_1 = I_2$, we have the fixed point $\bar{x} \equiv \bar{x}_1 = \bar{x}_2$ and $\bar{y} \equiv \bar{y}_1 = \bar{y}_2$. The two bars can excite or inhibit each other depending on whether $J_{12} - g'_y(\bar{y})W_{12} > 0$. According to the bow-tie connection in Fig. (2B), mutual excitation is more likely when two bars are near parallel and aligned, while mutual inhibition is likely when the two bars are near parallel but not aligned. Again, high input strength I makes mutual inhibition more likely by raising $g'_y(\bar{y})$. Hence, the mutual interaction is again input contrast dependent. If the two input stimulus bars are near orthogonal to each other, J_{12} and $W_{12} \approx 0$, making for the least contextual interaction and contrast dependency.

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 $\bar{y}$. Consequently, $J_{12} - g'_y(\bar{y}_1)W_{12} < J_{21} - g'_y(\bar{y}_2)W_{21}$. In particular, it can happen that $J_{12} - g'_y(\bar{y}_1)W_{12} < 0$ and $J_{21} - g'_y(\bar{y}_2)W_{21} > 0$, i.e., x_1 effectively excites x_2 while x_2 effectively inhibits x_1. This means, two interacting pairs tend to have closer activity values x_1 and x_2 than two non-interacting pairs.

A simple application is the phenomenon called the tilt illusion, which is the slightly distorted perception of the orientation of a stimulus bar $\hat{I}_{i_1\beta_1}$ at position i_1 and orientation β_1 near a contextual bar at position i_2 and orientation β_2. Without loss of generality, we take $i_1 = (0,0)$ at the center of the visual field, and $\beta_1 = 0$ oriented vertically. As described above, the stimulus $\hat{I}_{i_1\beta_1}$ excites multiple pyramidal cells $(i_1, \theta_1 \approx \beta_1)$ with θ_1 close to but not necessarily equal to β_1. The perceived orientation for the stimulus (i_1, β_1) may be approximated by:

$$\bar{\theta} = \sum_{\theta_1} g_x(\bar{x}_{i_1,\theta_1})\theta_1 \Big/ \sum_{\theta_1} g_x(\bar{x}_{i_1,\theta_1}) \tag{10}$$

at i_1. Without any contextual bars, the response $g_x(\bar{x}_{i_1,\theta_1})$ is determined by the cell's orientation tuning curve $\phi(\theta_1 - \beta_1)$, which is an even function of $\theta_1 - \beta_1$. This means $\bar{\theta} = \beta_1$. The tilt illusion results when the contextual bar causes responses $g_x(\bar{x}_{i_1,\theta_1})$ to be asymmetrical about $\beta = 0$, most noticably $g_x(\bar{x}_{i_1,\delta}) \neq g_x(\bar{x}_{i_1,-\delta})$ for small angle δ, making $\bar{\theta} \neq \beta_1 = 0$.

The tilt illusion is most noticable when $\beta_2 \approx \beta_1$. Let us take $\delta > 0$ as a small clockwise rotation away from vertical. The contextual bar has a small orientation $\beta_2 \approx > 0$ and is located at one of the 3 nearby positions (Fig. 4):

(a) $i_2 =$ $(0, n_2 > 0)$ directly above i_1

(b) $i_2 =$ $(m_2 > 0, 0)$ to the right of i_1

(c) $i_2 =$ $(m_2 > 0, n_2 > 0)$ upper-right from i_1

In condition (a), (i_2, β_2) is roughly parallel and aligned with (i_1, β_1), making the contextual influences likely to be excitatory. However, (i_2, β_2) is more aligned with slighted left tilted $(i_1, \theta_1 = -\delta < 0)$ than with slightly right tilted $(i_1, \theta_1 = \delta > 0)$, i.e., a curve of smaller curvature is needed to connect (i_2, β_2) to $(i_1, -\delta)$ than to (i_1, δ). Hence, the contextual excitation J_{12} to the left tilted $(i_1, -\delta)$ is stronger than that to the right tilted (i_1, δ), making $g_x(\bar{x}_{i_1,-\delta}) > g_x(\bar{x}_{i_1,\delta})$ and thus the tilt illusion is a slighted left tilted perception $\bar{\theta} < 0$. In condition (b), (i_2, β_2) is roughly parallel to but not aligned with (i_1, β_1), making the contextual influences likely to be inhibitory. Since (i_2, β_2) is more parallel to (i_1, δ) than to $(i_1, -\delta)$, the disynaptic inhibitory connection W_{12} from the contextual bar is stronger to (i_1, δ) than to $(i_1, -\delta)$. This again causes $g_x(\bar{x}_{i_1,-\delta}) > g_x(\bar{x}_{i_1,\delta})$ and a slightly left tilted illusion $\bar{\theta} < 0$. Both (a) and (b) cause the perceived test bar to tilt slightly away (in orientation) from the contextual bar, as usually observed psychophysically. However in condition (c), when the contextual bar is placed at the upper-right from the test bar, the contextual facilitation is stronger to the slighted right tilted (i_1, δ) than the left tilted $(i_1, -\delta)$ because of the difference in the alignment. This causes a right tilted illusion. Although this illusion is not as usually documented in the literature, recent
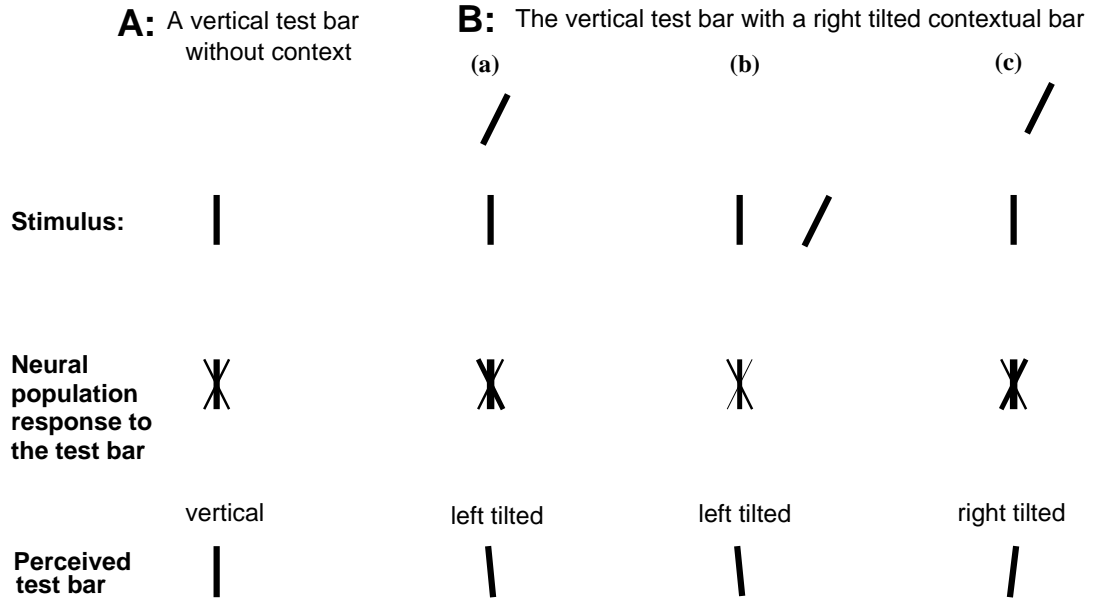
A: A vertical test bar without context

B: The vertical test bar with a right tilted contextual bar

(a)   (b)   (c)

Stimulus:

Neural population response to the test bar

vertical   left tilted   left tilted   right tilted

Perceived test bar

Figure 4: Schematic of tilt illusions — the contextual influences from a contextual bar (tilted β_2 degree from vertical) to a vertical test bar. A: the test bar without context, and B: with context. From top to bottom are the input stimulus, the population response to the test bar, and the perceived test bar. The neural population responses are from neurons with receptive fields centered on the test bar and preferring vertical ($\theta_1 = 0$) and near vertical ($\theta_1 = \delta > 0$ or $-\delta$, tilted to the right or left from vertical) orientations. For schematic visualization, the thicknesses of each bar (in the population response) is drawn to increase with the response levels of the corresponding neurons. The perceived orientation of the test bar is the response weighted average (center of gravity) of the preferred orientations of the population of responsive neurons. The contextual bar is (a) at top of, (b) on the right of, and (c) at upper-right of, the test bar.

psychophysical observations (Kapadia 1998) confirm this phenomena, as readers themselves can roughly test on the figure.

3.3 A one dimensional array of identical bars

Consider an input consisting of identical bars oriented at $\theta = \theta_1$, lined up horizontally and separated by equal spaces as in Fig. (5A). The input $I_{i\theta}$ can be approximated as

$$
I_{i\theta} = \begin{cases} I_{array} & \text{for } i = (m_i, n_i = 0) \text{ on the horizontal axis and } \theta = \theta_1 \\ 0 & \text{otherwise} \end{cases} \tag{11}
$$

Here the approximation $I_{i\theta} = 0$ for $\theta \neq \theta_1$ is valid when each cell has a near zero orientation tuning width or when the input contrast is small. In the simplest case when bars $i\theta$ outside that array are silent $g_x(x_{i\theta}) = 0$ due to insufficient excitation, we ignore all bars beyond the array, treat the system as one dimensional, omit index θ, and let i denote the one dimensional location of the bar in the array. Then

$$
\dot{x}_i = -x_i - g_y(y_i) + I_{array} + I_o + J_o g_x(x_i) + \sum_{j \neq i} J_{ij} g_x(x_j) \tag{12}
$$

$$
\dot{y}_i = -y_i + I_c + g_x(x_i) + \sum_{j \neq i} W_{ij} g_x(x_j) \tag{13}
$$

For an infinitely long array, translation symmetry implies that all units have the same equilibrium point $\bar{x}_i = \bar{x}$ and $\bar{y}_i = \bar{y}$ and

$$\dot{\bar{x}} = 0 = -\bar{x} - g_y(\bar{y}) + I_{array} + I_o + (J_o + \sum_{i \neq j} J_{ij})g_x(\bar{x}) \tag{14}$$

$$\dot{\bar{y}} = 0 = -\bar{y} + I_c + (1 + \sum_{i \neq j} W_{ij})g_x(\bar{x}) \tag{15}$$

Comparing with equations (7) and (8), we see that this array is equivalent to a single neural pair with stronger self-excitation $J_o \to J_o + \sum_j J_{ij}$ and effective self-inhibition $g'_y(\bar{y}) \to g'_y(\bar{y})(1 + \sum_j W_{ij})$. The input gain is

$$\frac{\delta g_x(\bar{x})}{\delta I_{array}} = \frac{g'_x(\bar{x})}{1 + g'_x(\bar{x})(g'_y(\bar{y}) - J_o + g'_y(\bar{y})\sum_j W_{ij} - \sum_j J_{ij})} \tag{16}$$

Comparing with equation (9), we see that the response to bars in the array is higher than that to an isolated bar if the net extra excitatory connection

$$\mathcal{E} \equiv \sum_j J_{ij} \tag{17}$$

is stronger than the net extra inhibitory (effective) connection

$$\mathcal{I} \equiv g'_y(\bar{y})\sum_j W_{ij}. \tag{18}$$

The connections J_{ij} and W_{ij} depend on θ_1. When the bow-tie connections are used, and the bars are parallel to the array, making a straight line (Fig (5B)), we have $\mathcal{E} > \mathcal{I}$ giving an enhanced response. The total contour facilitation is

$$F_{contour} = (\mathcal{E} - \mathcal{I})g_x(\bar{x}) \tag{19}$$

When the bars are orthogonal to the array (Fig (5C)), $\mathcal{E} < \mathcal{I}$ and thus the responses are suppressed. In fact, our analysis can be extended to other translation invariant one dimensional arrays like in Fig (5D, E), for which the index i simply denotes a bar in a location along the array. In fact, the straight line in Fig (5B) can be seen as the limit of a circle in Fig (5D) when the radius goes to infinity. Let us again consider the bow-tie connections in which the J connections are assumed stronger between bars that are better aligned with each other. The excitatory strength $\mathcal{E}$ among the bars will be larger, and the inhibitory strength $\mathcal{I}$ will be smaller, for the straight line in Fig (5B) than the circle in Fig (5D). Consequently, contour enhancement is stronger for the straight line than the circle, and the enhancement decreases as the radius of the circle becomes smaller. Similarly, the pattern in Fig (5C) is a special case of the one in Fig (5E) when the radius goes to infinity.

The approximations used in equations (11 -15) can be quite good in some situations, but quite poor in others. This is demonstrated in the simulated examples in Fig. (6). Fig. (6G) shows the control case for which there is no contextual input to the stimulus bar. Contextual facilitation in Fig. (6A, B, E) is visualized by the thicker bars than those in Fig. (6G); while contextual suppression in Fig. (6C, D, F) is visualized by the thinner bars. In Fig. (6A), 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, since in population coding, cells preferring orientations near the stimulus bar are often excited also.) This is caused by direct visual input $I_{i\theta}$ for $\theta \neq \theta_1$ ($\theta \approx \theta_1$), as well as 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 stimulus array is compromised when the array is not
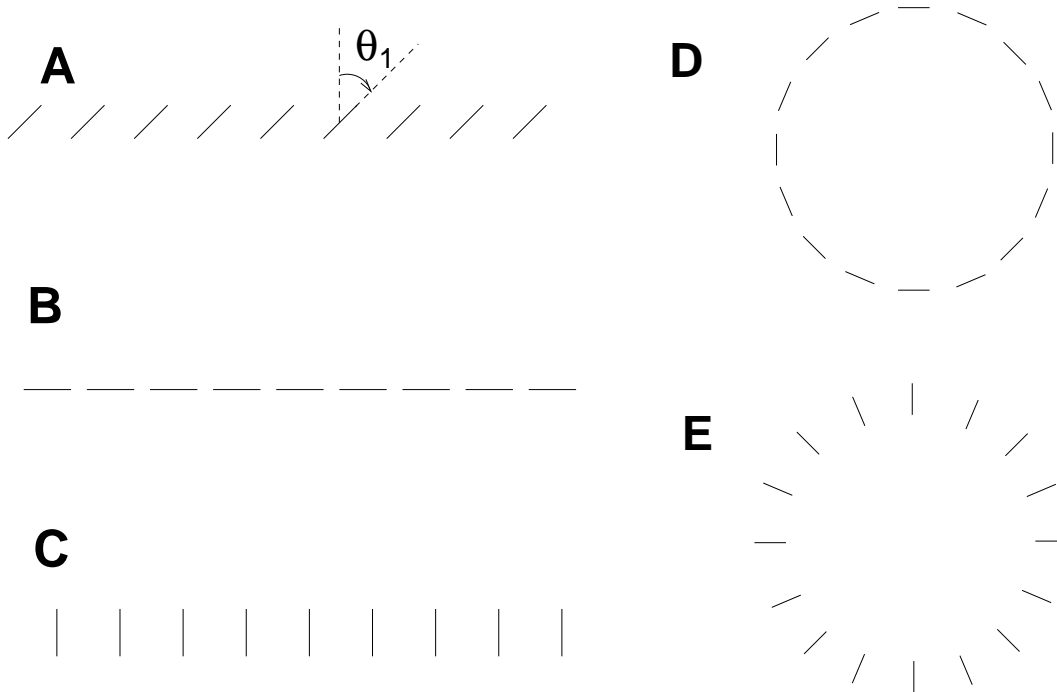
Figure 5: Examples of the one dimensional input stimulus mentioned in the text. **A**: horizontal array of identical bars oriented at angle θ_1. **B**: A special case of **A** when $\theta_1 = \pi/2$ and, in **C**, when $\theta_1 = 0$. **D**: an array of bars aligned into, or tangential to, a circle, 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 circumsference, the pattern in **C** is a special case when the radius is infinitely large.

infinitely long, e.g., Fig. (6B), or when the bars in the array are not equally spaced, e.g., Fig. (6E,F). In Fig. (6B), the bars at or near the left end of the line receive less or no contextual facilitation from their left, meaning that they are less enhanced than bars away from this end. In Fig. (6F), the more densely spaced bars receive more contextual suppression than others, giving weaker responses.

3.4 Two dimensional textures and texture boundaries

The analysis for 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. (7A)). The sums $\mathcal{E} = \sum_j J_{ij}$ and $\mathcal{I} = g'_y(\bar{y}) \sum_j W_{ij}$ are taken over locations j in the two dimensional grid.

It is observed physiologically that the cell's response to a bar is reduced when the bar is part of an extended texture (Knierim and van Essen 1992). This can be achieved when

$$\mathcal{E} < \mathcal{I}. \tag{20}$$

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 (7)). The texture can be seen as a horizontal array of vertical arrays of bars, e.g., a horizontal array of vertical contours when the bars are vertically oriented (Fig. (7B)). (The same texture can also be viewed as a vertical array of horizontal arrays of bars, or even an oblique array of arrays of bars. This will become useful for analysis in section 3.7.) We can

13

A: Infinitely long line

E: Uneven circular array

B: Half infinitely long line,
ending on the left

F: Uneven radiant array
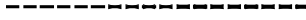
C: Infinitely long array
of oblique bars

D: Infinitely long horizontal
array of vertical bars
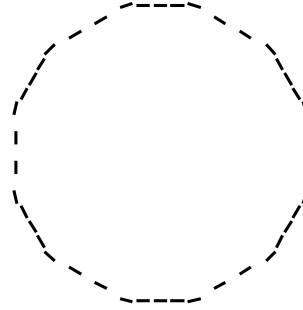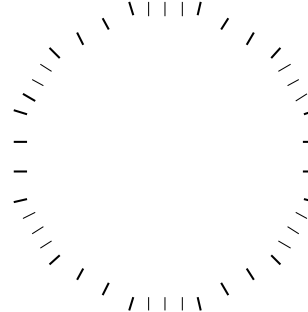
G: An isolated bar

Figure 6: 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 (suprethreshold) 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 weakers to bars that are more densely packed in **F**. In **A**, cells preferring neighboring orientations (near horizontal) at the line are also excited above threshold, unlike the approximated treatment in the text.

define the effective connections between the vertical arrays (Fig. (7DEF)) distanced a apart:

$$J_a' \equiv \sum_{j, m_j = m_i + a} J_{ij}, \qquad W_a' \equiv \sum_{j, m_j = m_i + a} W_{ij}. \qquad (21)$$

A A texture of bars oriented at θ_1

B A texture of vertical bars

C Two neighboring textures of bars.

Four example pairs of vertical arrays of bars.

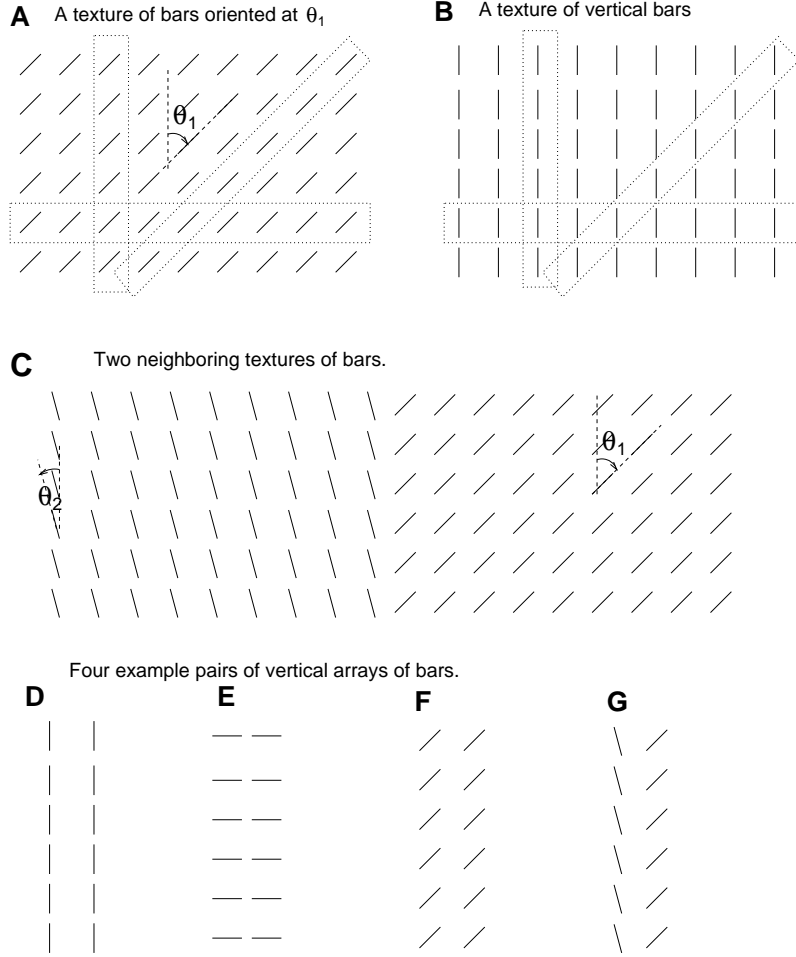**D**   **E**   **F**   **G**

Figure 7: Examples of the two dimensional textures and their interactions. **A:** texture made of bars oriented at $\theta_1 = 0$ 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.

Then

$$\mathcal{E} = \sum_j J_{ij} = \sum_a J'_a, \qquad \mathcal{I} = g'_y(\bar{y}) \sum_j W_{ij} = g'_y(\bar{y}) \sum_a W'_a \qquad (22)$$

The effective connection within a single vertical array is J'_0 and W'_0. It is possible to have the connection structure J and W such that $J'_0 > g'_y(\bar{y})W'_0$ and $\mathcal{E} = \sum_a J'_a < \mathcal{I} = g'_y(\bar{y}) \sum_a W'_a$ can be simultaneously satisfied for some orientation θ_1, as long as there is sufficient excitation within a vertical array and sufficient inhibition between vertical arrays. In particular, when the vertical array is a long straight line ($\theta_1 = 0$), one can have contour enhancement (i.e., $J'_0 > g'_y(\bar{y})W'_0$) when that line is isolated, but overall suppression (i.e., $\mathcal{E} = \sum_a J'_a < \mathcal{I} =$

15

$g'_y(\bar{y}) \sum_a W'_a$) when that line is embedded within a texture of lines. Hence contour enhancement and texture suppression can be achieved within a single neural circuit, i.e., a single set of J and W connections.

Computationally, contextual suppression within a texture means that the boundaries of a texture region can give rise to relatively higher responses, thereby serving segmentation. The contextual suppression on a bar within a texture of bars of orientation θ_1 is

$$C_{whole-texture}^{\theta_1} \equiv \sum_a (g'_y(\bar{y}_{\theta_1})W_a'^{\theta_1} - J_a'^{\theta_1})g_x(\bar{x}_{\theta_1}) = (\mathcal{I} - \mathcal{E})g_x(\bar{x}_{\theta_1}) > 0 \tag{23}$$

Here $\bar{x}_{\theta_1}$ denotes the (translation invariant) fixed point for all texture bars, index θ_1 denotes the dependences of parameters and variables on θ_1. Consider the bars on the vertical axis $i = (m_i = 0, n_i)$. Removing the texture bars on the left $i = (m_i < 0, n_i)$ removes the contextual suppression from the left half of the texture area. Now the activity level $\bar{x}_{i\theta_1}$ depends on the horizontal location m_i of the bars, or their distance from the texture boundary. As the distance $m_i \to \infty$, $\bar{x}_{i\theta_1}$ approaches $\bar{x}_{\theta_1}$. The contextual suppression on the bars on the vertical axis is

$$C_{half-texture}^{\theta_1} \equiv \sum_{m_j \geq 0} (g'_y(\bar{y}_{\theta_1})W_{m_j}'^{\theta_1} - J_{m_j}'^{\theta_1})g_x(\bar{x}_{j\theta_1}) \tag{24}$$

$$\approx \sum_{m_j \geq 0} (g'_y(\bar{y}_{\theta_1})W_{m_j}'^{\theta_1} - J_{m_j}'^{\theta_1})g_x(\bar{x}_{\theta_1}) \tag{25}$$

where we approximate $\bar{x}_{i\theta_1} \approx \bar{x}_{\theta_1}$. Since $C_{half-texture}^{\theta_1} < C_{whole-texture}^{\theta_1}$, the reduced suppression leads to relatively higher response of the bars on the vertical axis, thus marking the boundary for visual segmentation purposes.

Consider two neighboring textures of bars oriented at θ_1 for $i = (m_i \geq 0, n_i)$ and θ_2 for $i = (m_i < 0, n_i)$ (Fig. (7C)). The connections between arrays in different textures (Fig. (7G)) are

$$J_a'^{\theta_1\theta_2} \equiv \sum_{j, m_j = m_i + a} J_{i\theta_1 j\theta_2} \qquad W_a'^{\theta_1\theta_2} \equiv \sum_{j, m_j = m_i + a} W_{i\theta_1 j\theta_2} \tag{26}$$

By symmetry $J_a'^{\theta_1\theta_2} = J_{-a}'^{\theta_1\theta_2} = J_a'^{\theta_2\theta_1}$, and similarly for $W_a'^{\theta_1\theta_2}$. Further, when $\theta_1 = \theta_2$, $J_a'^{\theta_1\theta_2} = J_a'^{\theta_1}$ and $W_a'^{\theta_1\theta_2} = W_a'^{\theta_1}$. The contextual suppression on the border bar at $i = (m_i = 0, n_i)$ is contributed by both textures, and is thus

$$C_{2-half-textures}^{\theta_1,\theta_2} \equiv C_{half-texture}^{\theta_1} + C_{neighbor-half-texture}^{\theta_1,\theta_2}$$
$$\text{where}$$
$$C_{neighbor-half-texture}^{\theta_1,\theta_2} \equiv \sum_{m_j < 0} (g'_y(\bar{y}_{i\theta_1})W_{m_j}'^{\theta_1\theta_2} - J_{m_j}'^{\theta_1\theta_2})g_x(\bar{x}_{j\theta_2})$$
$$\approx \sum_{m_j < 0} (g'_y(\bar{y}_{\theta_1})W_{m_j}'^{\theta_1\theta_2} - J_{m_j}'^{\theta_1\theta_2})g_x(\bar{x}_{\theta_2})$$

Here $C_{neighbor-half-texture}^{\theta_1,\theta_2}$ is the contextual suppression from neighboring texture to the border texture bars. Again, we use approximation $\bar{x}_{j\theta_2} \approx \bar{x}_{\theta_2}$. Note that 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).

The reduction in contextual suppresion on the border bars, or, border enhancement, is then

$$\delta C \equiv C_{whole-texture}^{\theta_1} - C_{2-half-texture}^{\theta_1,\theta_2} \tag{27}$$

$$\approx C_{neighbor-half-texture}^{\theta_1,\theta_2=\theta_1} - C_{neighbor-half-texture}^{\theta_1,\theta_2} \tag{28}$$

$$\approx \sum_{a<0} (g'_y(\bar{y}_{\theta_1})W_a'^{\theta_1} - J_a'^{\theta_1})g_x(\bar{x}_{\theta_1}) - \sum_{a<0} (g'_y(\bar{y}_{\theta_1})W_a'^{\theta_1\theta_2} - J_a'^{\theta_1\theta_2})g_x(\bar{x}_{\theta_2}) \tag{29}$$

which is responsible for relatively higher responses at the texture border than texture regions away from the border, serving pre-attentive texture segmentation.

If J and W tend to link cells preferring similar orientations as in the bow-tie pattern, $J_a'^{\theta_1\theta_2}$ and $W_a'^{\theta_1\theta_2}$ decrease with increasing $|\theta_1 - \theta_2|$, and vanish for $|\theta_1 - \theta_2| = \pi/2$. We further note that $\bar{x}_{\theta_1} \approx \bar{x}_{\theta_2}$ when $\theta_1 \approx \theta_2$. Then,

$$\delta C \approx \begin{cases} 0 & \text{for } \theta_1 \approx \theta_2 \\ \sum_{a<0}(g_y'(\bar{y}_{\theta_1})W_a'^{\theta_1} - J_a'^{\theta_1})g_x(\bar{x}_{\theta_1}) > 0 & \text{for } \theta_1 \perp \theta_2 \\ \text{roughly increases} & \text{as } |\theta_1 - \theta_2| \text{ increases} \end{cases} \tag{30}$$

Thus the texture border gives relatively higher responses when the orientation contrast between the textures is high, this highlight diminishes as the orientation contrast approaches 0, see Fig. (8). Furthermore, even at a given contrast $|\theta_1 - \theta_2|$, the border enhancement δC still depends on θ_1. For instance, when $|\theta_1 - \theta_2| = \pi/2$ and the bow-tie connections are used, the enhancement δC for border bars parallel to the border $\theta_1 = 0$ is higher than that for border bars perpendicular to the border $\theta_1 = \pi/2$. This is partly because the strength of contextual suppression $g_y'(\bar{y}_{\theta_1})W_a'^{\theta_1} - J_a'^{\theta_1}$ between parallel contours ($\theta_1 = 0$ and $a \neq 0$) (Fig. (7D)) is much stronger than the strength of suppression between two vertical arrays of horizontal bars (Fig. (7E)). Thus we predict a tuning of the strength of the border highlight to the relative orientation θ_1 between the border and the bars (Li 1999b).

Clearly, the approximation ($\bar{x}_{i\theta_1} \approx \bar{x}_{\theta_1}$ for $m_i \geq 0$ and $\bar{x}_{i\theta_2} \approx \bar{x}_{\theta_2}$ for $m_i < 0$), which is used to arrive at equation (30), breaks down at the border. This breakdown is more severe at stronger or more salient borders such as the one in Fig. 8C, (which is particularly strong due to the colinear excitation along the vertical contour at the border). This accentuates the tuning of the border highlight to the relative angle between the border and the texture bars.

The mechanism underlying the border highlight is texture suppression or iso-orientation suppression within a texture region. The stronger the suppression, the stronger the border highlight. By equation (23), this suppression strength is contrast dependent through $g_y'(\bar{y})$. Since $g_y'(\bar{y})$ usually increases with increasing $\bar{y}$, the border highlight or texture segmentation is more effective at higher contrast. The connection weights in the model can be designed such that at very low input contrast, this suppression diminishes or even becomes facilitation. There could be computational reasons for such a change: facilitation certainly helps texture detection, which at low input contrast could be more important than texture segmentation. Psychophysically, texture segregation does require an input contrast that is well above the texture detection threshold (Nothdurft 1994).

3.5 Translation invariance and pop-out

In the examples above, orientation contrast in inputs have relatively higher saliency because they mark boundaries between textures which are composed of bars of single orientations. However, if orientation contrast is homogeneous through the texture itself, then particular locations with particular orientation contrast will not have higher saliencies than their surroundings and hence will not pop out or attract visual attention. Fig. (9A) shows an example for which the texture is made of alternate columns of bars at $\theta_1 = 45^o$ (odd a) and $\theta_2 = 135^o$ (even a). The contextual suppression on a bar oriented at θ_1 is:

$$C_{complex-texture} = \sum_{\text{odd } a} (g_y'(\bar{y}_{\theta_1})W_a'^{\theta_1} - J_a'^{\theta_1})g_x(\bar{x}_{\theta_1}) + \sum_{\text{even } a} (g_y'(\bar{y}_{\theta_1})W_a'^{\theta_1\theta_2} - J_a'^{\theta_1\theta_2})g_x(\bar{x}_{\theta_2}) \tag{31}$$

Thus no bar oriented at θ_1 is less suppressed, or more salient, than other bars oriented at θ_1. Note that since $C_{complex-texture} \neq C_{whole-texture}^{\theta_1}$, 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 $\bar{x}_{\theta_2}$. In general, $\bar{x}_{\theta_1} \neq \bar{x}_{\theta_2}$. In the particular case of Fig. (9A), the bars oriented at θ_2 receive the same amount of
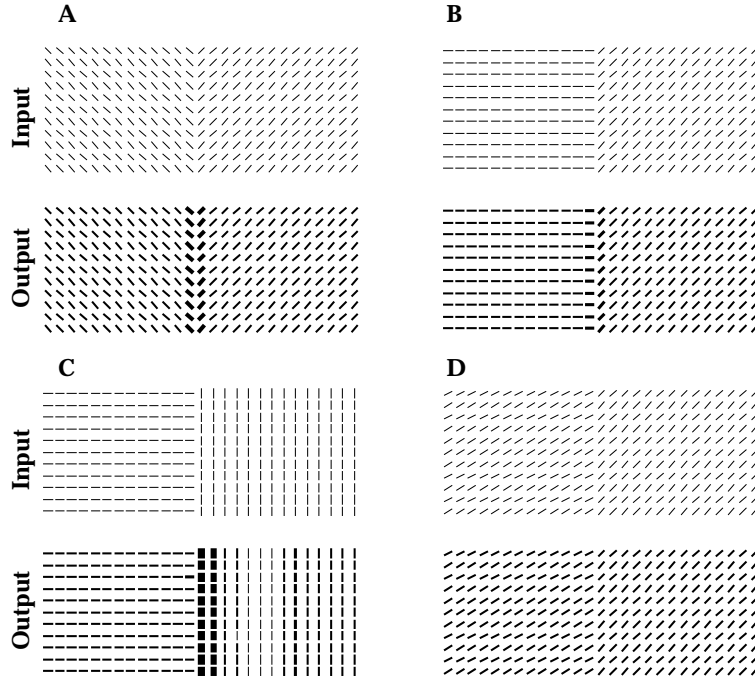
Figure 8: Simulated examples of texture boundary highlights between different pairs of textures, defined by bar orientations. In each example is the input image $I_{i\theta}$ above the output image $g_x(x_{i\theta})$ averaged in time. Each image plots a small region out of a much more extended input 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° 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.

contextual suppression as the bars oriented at θ_1. Hence, the whole texture appears uniform in saliency. Fig. (9B) shows an example of another homogeneous texture where the bars oriented at $\theta_1 = 0^o$ induced higher responses than bars oriented at $\theta_2 = 90^o$. However, when one looks at the 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 another patch. In other words, although the texture does not have uniform saliency, saliency as a function of spatial location is translation invariant. This translation invariance in saliency is induced by the translation invariance in the input (texture). If the neural circuit, with its translation invariant cortical interactions J and W, does not spontaneously break the translation symmetry (see section 3.7 for analysis), the input translation invariance is preserved at the output.

A boundary between textures is one place where translation invariance in the inputs breaks down. It induces higher outputs through 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 texture Fig. (9C) . Orientation contrast coincides with the texture boundary or translation invariance breaking only between two simple textures of bars of single orientations. In general, orientation contrast does not necessarily pop out.
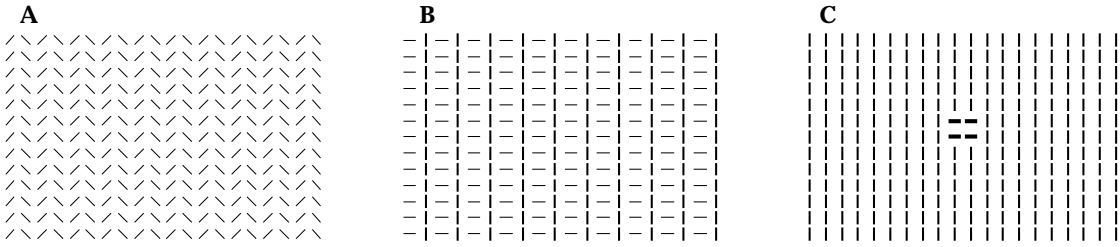
Figure 9: Model responses to homogeneous (**A, B**) and non-homogeneous (**C**) input images, each composed of bars of equal input contrasts. **A:** A homogeneous (despite of the orientation contrast) texture of bars of different orientations, a uniform output saliency results. **B:** Another homogeneous 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.

3.6 Filling-in and leaking-out

Sometimes small fragments of a smooth contour or homogeneous texture are missing in inputs due to input noise or to the visual scene itself. Filling-in is the phenomenon of that the missing input fragments are not noticed. Filling-in could be caused by at least two possible mechanisms. The first possibility is that, although the cells for the missing fragment do not receive direct visual inputs, contextual inputs from stimuli near the missing fragment provide enough excitation to these cells to make them fire as if there were direct visual inputs. (This is the way that some models cope with illusory contours, e.g., Grossberg and Mingolla 1985.) 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 do not have high enough saliencies, i.e., are not conspicuous enough, to strongly attract visual attention. In this case, the missing fragments are only noticable by attentive visual scrutiny/search.

For a single bar segment missing in a smooth contour, filling-in could be achieved by either of the two possible mechanisms. Let the missing bar be at $i = (m_i = 0, n_i = 0)$ in a contour like Fig. (5B), we ignore the spatial inhomogeneity and make the approximation $\bar{x}_j \approx \bar{x}$ where $\bar{x}_j$ is the activities for the non-missing segments and $\bar{x}$ is that in a complete contour (without missing segments). The contextual excitation to the missing segment is then $F_{contour}$ as in equation (19). The filling-in condition under which the cell x_i is excited beyond firing threshold is

$$F_{contour} + I_o = (\mathcal{E} - \mathcal{I})g_x(\bar{x}) + I_o > T_x \tag{32}$$

where T_o is the cell firing threshold such that $g_x(x_i) > 0$ when $x_i > T_o$, I_o is the background input to the excitatory cell, and the effective net connections $\mathcal{E}$, $\mathcal{I}$, and contour facilitation $F_{contour}$ are as defined in equations (17 - 19). Assuming that background input I_o alone does not excite the cell enough to fire, this filling-in condition can be satisfied with large enough contour facilitation strength $\mathcal{E} - \mathcal{I}$ and strong enough contour saliency $g_x(\bar{x})$ (which in turn is caused by strong enough input contrast). Assuming that segments within a smooth contour normally facilitate each other's firing, then a missing fragment i should usually lead to a reduction of contextual facilitation to the neighboring contour bars j. In this case, the bars near the missing segment can not be more salient or conspicuous than other contour bars far from the missing fragment, which is thus not easily noticed.

Contour enhancement should not be so strong that bar segments beyond the end of a long, but finite length, contour are excited towards nonzero outputs. This would make the output contour longer than the input contour or make it grow in length — leaking out. Consider the case of a horizontal, half infinitely long, contour with segments $j = (m_j > 0, n_j = 0)$

19

ending at $i = (m_i = 0, n_i = 0)$, we again make the approximation $\bar{x}_j \approx \bar{x}$ that cell activities in the contour are roughly that in an infinitely long contour. The contextual excitation on the segment $i = (m_i = 0, n_i = 0)$ outside the end of the contour is then $F_{contour}/2$, i.e., half the contour facilitation in a whole contour, since there is no contextual inputs from the left side of $i = (m_i = 0, n_i = 0)$. The condition to prevent leaking-out is then

$$F_{contour}/2 + I_o < T_x \tag{33}$$

This condition is satisfied for the line end in Fig. (6B), and should be satisfied for any contour saliency strength $g_x(\bar{x})$. Not leaking out also means that large gaps in lines can not be filled in. To prevent leaking-out at a location i, e.g., at $i = (m_i = 0, n_i = 1)$, at the side of an infinitely long (e.g.,) horizontal contour on the horizontal axis in Fig. (5B) (thus to prevent the contour getting thicker), we require $\sum_{j \in \text{contour}} (J_{ij} - g_y'(\bar{y})W_{ij})g_x(\bar{x}) < T_x - I_o$. for $i \notin$ contour. This condition is satisfied in Fig. (6A).

If the contextual inputs within a texture are suppressive, the cell for a missing fragment i in a texture can never be excited to fire. Further, the neighbors $k \approx i$ of the missing fragment i will experience a weaker contextual suppression than elements far away because of the absent inhibitory contribution from the missing fragment. The filling-in can thus be achieved only by the second possible mechanism outlined above, and this only when the saliencies of the neighbors near the missing fragment are no more than a small fraction higher than those for other texture elements (Fig 10B). Hence, the suppression strength $W_{ki}g_y'(\bar{y}) - J_{ki}$ on the neighboring bars k from the missing bar $i \approx k$ should be negiligible compared to the total suppression strength from the whole texture, i.e.,

$$g_y'(\bar{y}_k)W_{ki} - J_{ki} \ll (\mathcal{I} - \mathcal{E}) \tag{34}$$

where $\mathcal{I} \equiv \sum_{j \in \text{texture}} g_y'(\bar{y}_k)W_{kj}$ and $\mathcal{E} \equiv \sum_{j \in \text{texture}} J_{kj}$. This condition is expected to hold when the lateral connections are extensive enough to reach large enough 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 because the contextual input from the texture is suppressive.

It is apparent that the conditions that (a) small gaps in lines should be filled in (equation (32)), and (b) there should be no leaking-out at the ends of lines (equation (33)), work against each other. It is not difficult to build a model that achieves active filling-in. However, preventing the model from allowing leaking out and thus creating more illusory contours than is warranted by human vision implies that there is only a small range of choices for the connection strengths J and W.

3.7 Oscillations and hallucinations

The computational analysis in section 3.1-3.6 has been limited to looking only at the fixed point $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ which is a good representation of the input I. (Here we use bold-faced character to represent vectors or matrices.) Our analysis applies only if the fixed point is stable or when the fixed point is roughly the temporal average of the state variables, i.e., $\int dt\mathbf{X}(t)/\int dt \approx \bar{\mathbf{X}}$ and $\int dt\mathbf{Y}(t)/\int dt \approx \bar{\mathbf{Y}}$. When the fixed point is unstable, two possible situations can arise. First, the system variables $(\mathbf{X}, \mathbf{Y})$ could oscillate (Gray and Singer 1989, Eckhorn et al 1988) around the fixed point $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ such that the fixed point still roughly describes the temporal mean activities. Second, the system could approach a stable state (another fixed point) $(\mathbf{X}, \mathbf{Y}) \neq (\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ or even a dynamic state such that the temporal mean $\int dt\mathbf{X}(t)/\int dt$ is no longer close to the input driven fixed point $\bar{\mathbf{X}}$ which is the one we have been studying for model computation. In particular, assuming that the fixed point $\bar{\mathbf{X}}$ faithfully represents the visual input I, a network state whose temporal average state $\int dt\mathbf{X}(t)/\int dt$ deviates strongly from $\bar{\mathbf{X}}$ would imply that the system hallucinates patterns not present or usually perceived
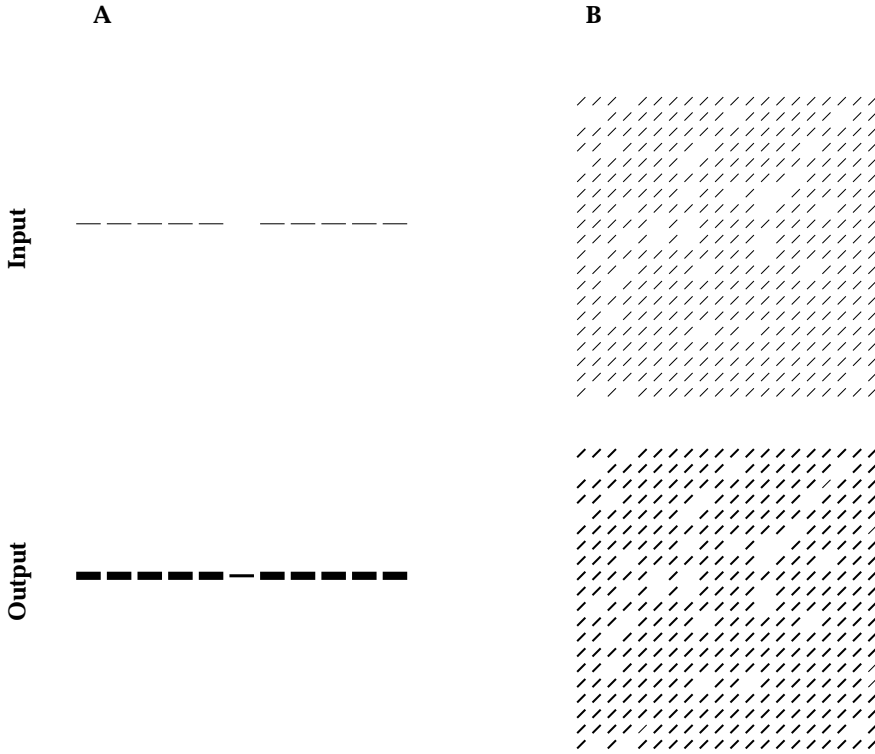
Figure 10: Examples of filling-in — model outputs from inputs composed of bars of equal contrasts in each example. **A:** A line with a gap, the response to the gap is non-zero, **B:** A texture with missing bars, the responses to bars near the missing bars are not significantly higher than the responses to other texture bars.

(by humans) in the input, i.e., spontaneous pattern formation. This is exactly what happens in the model of symmetric neural connections in equation (1) and Fig. (1), where the fixed point $\{\bar{x}_{i\theta}\}$ faithful to the input is unstable because it is not an energy minimum, and the state variable $\{x_{i\theta}\}$ slides to an energy minimum state which is very far from the input.

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

$$\begin{pmatrix} \dot{\mathbf{X}} \\ \dot{\mathbf{Y}} \end{pmatrix} = \begin{pmatrix} -1 + \mathbf{J} & -\mathbf{G}'_{\mathbf{y}} \\ \mathbf{G}'_{\mathbf{x}} + \mathbf{W} & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{35}$$

where $\mathbf{J}$, $\mathbf{W}$, $\mathbf{G}'_{\mathbf{x}}$, and $\mathbf{G}'_{\mathbf{y}}$ are matrices with $\mathbf{J}_{i\theta j\theta'} = J_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$, $\mathbf{J}_{i\theta,i\theta} = J_o g'_x(\bar{x}_{i\theta})$, $\mathbf{W}_{i\theta j\theta'} = W_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$ 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 $\psi(0) = 1$. Let the matrix $\mathbf{M} \equiv \begin{pmatrix} -1 + \mathbf{J} & -\mathbf{G}'_{\mathbf{y}} \\ \mathbf{G}'_{\mathbf{x}} + \mathbf{W} & -1 \end{pmatrix}$ have eigenvalues γ_k and

eigenvectors $\mathbf{Z}^k \equiv \begin{pmatrix} \mathbf{X}^k \\ \mathbf{Y}^k \end{pmatrix}$, for $k = 1, 2, ...$, the small deviation follows trajectory $\mathbf{Z}(t) \approx$

$\sum_k c_k \mathbf{Z}^k e^{\gamma_k t}$, where c_k is determined by the initial deviation $\mathbf{Z}(t = 0)$. Hence, the fixed point is stable only when all eigenvalues of $\mathbf{M}$ have negative real parts. We can eliminate (hidden) variable $\mathbf{Y}$ to obtain:

$$\ddot{\mathbf{X}} + (2 - \mathbf{J})\dot{\mathbf{X}} + (\mathbf{G}'_{\mathbf{y}}(\mathbf{G}'_{\mathbf{x}} + \mathbf{W}) + 1 - \mathbf{J})\mathbf{X} = 0 \tag{36}$$

This enables us to focus on the output variables $\mathbf{X}$.

Analysis is possible in the cases of particular interest for which the input is an infinitely long array or an infinitely large texture of identical bars. In this case, there is a translation invariant fixed point $\bar{\mathbf{X}}$. When it is unstable, deviations $\mathbf{X}$ from it can be spatially or translationally variant, leading to spontaneous pattern formation and perceptual hallucination (Ermentrout and Cowan 1979) of spatially non-uniform saliencies. The locations of higher and hallucinated saliencies in a texture compete and confuse with highlights at texture borders, and this is undesirable if the output of the computation of texture segmentation uses higher saliencies at the border. We analyze the translation invariant fixed points by ignoring bars outside the array or texture and omit index θ, as we did in subsections (3.3) and (3.4). Now $\mathbf{X}$ and $\mathbf{Y}$ in equation (35) only include components x_i, y_i for the contour or texture bars i (omitting θ), and $\mathbf{J}_{ij}$ and $\mathbf{W}_{ij}$ only connect these bars to each other. Again, translation symmetry means

$$\mathbf{G}'_{\mathbf{y}ij} = \delta_{ij}g'_y(\bar{y}), \qquad\qquad \mathbf{G}'_{\mathbf{x}ij} = \delta_{ij}g'_x(\bar{x})$$
$$(\mathbf{G}'_{\mathbf{y}}\mathbf{G}'_{\mathbf{x}})_{ij} = g'_x(\bar{x})g'_y(\bar{y})\delta_{ij}, \qquad (\mathbf{G}'_{\mathbf{y}}\mathbf{W})_{ij} = g'_y(\bar{y})\mathbf{W}_{ij},$$
$$\mathbf{J}_{ij} = \mathbf{J}_{i+a,j+a} \qquad \mathbf{W}_{ij} = \mathbf{W}_{i+a,j+a} \quad \text{for any } a$$

Hence, $\mathbf{J}$ and $\mathbf{W}$ are Toplitz matrices. They commute with each other and thus share the same eigenvectors. One can easily check that, for the kth eigenvector, for $k = 1, 2, ..., N$ (where N is the size of the system):

$$\text{The shared eigenvector} \qquad \mathbf{X}^k_j = e^{\mathrm{i}f_k j} \qquad \text{where i} = \sqrt{-1}, \tag{37}$$

$$\text{The eigenvalue of } \mathbf{J} \qquad \lambda^k_J = \sum_j \mathbf{J}_{aj}e^{\mathrm{i}f_k(a-j)} \tag{38}$$

$$\text{The eigenvalue of } \mathbf{W} \qquad \lambda^k_W = \sum_j \mathbf{W}_{aj}e^{\mathrm{i}f_k(a-j)} \tag{39}$$

Here $\mathbf{X}^k_j$ is the jth component in the eigenvector $\mathbf{X}^k$, which are Fourier waves of spatial frequency f_k such that $e^{\mathrm{i}f_k N} = 1$. λ^k_J and λ^k_W are Fourier transforms (spectrum) of the row vectors in $\mathbf{J}$ and $\mathbf{W}$. Eq. (36) has solutions $\mathbf{X} = \sum_k c^k \mathbf{X}^k e^{\gamma^k t}$, where

$$\gamma^k \equiv -1 + \lambda^k_J/2 \pm \mathrm{i}\sqrt{g'_y(g'_x + \lambda^k_W) - (\lambda^k_J)^2/4} \tag{40}$$

and c^k, the amplitude of mode $\mathbf{X}^k$, depends on initial conditions $\mathbf{X}(t = 0)$. If $Re(\gamma^k)$, the real part of γ^k, is negative for all k, the fixed point $\bar{\mathbf{X}}$ is stable. Otherwise, the mode with the largest $Re(\gamma^k)$, let it be $k = 1$, will dominate the deviation $\mathbf{X}(t) \approx c_1\mathbf{X}^1 e^{\gamma^1 t}$ from the fixed point $\bar{\mathbf{X}}$. This deviation oscillates in time in a stable limit cycle when $g'_y(g'_x + \lambda^1_W) - (\lambda^1_J)^2/4 > 0$ or when there is no other fixed point for the system trajectory to approach. If $\mathbf{X}^1$ is the zero frequency $f_1 = 0$ Fourier wave, then $\mathbf{X}^1_i = \mathbf{X}^1_j$ for $i \neq j$, this means, the deviation from the fixed point is also translation invariant, and the neural oscillations will be synchronized between the bar elements. In this case, the unstable fixed point does not lead to the hallucination of spatial patterns that are non-homogeneous or not translation invariant. A model without hallucination should be such that either the fixed points under homogeneous inputs are stable, or the deviations from the fixed point does not break the translation symmetry. Thus, under any translation invariant input

$$Re(\gamma^k) < 0 \qquad \text{for all } k, \qquad \text{or} \qquad Re(\gamma^1)_{f_1=0} > Re(\gamma^k)_{f_k \neq 0} \tag{41}$$

One notes that both $\mathbf{J}$ and $\mathbf{W}$ are symmetric matrices with only positive elements. Thus, both λ^k_J and λ^k_W are real, and they both achieve largest values $Max(\lambda^k_J) = \sum_j \mathbf{J}_{aj}$ and $Max(\lambda^k_W) = \sum_j \mathbf{W}_{aj}$ for the zero frequency $f_k = 0$ wave or the translation invariant mode. However, it is the value of $Re(\gamma^k)$ that determines the dominant mode $k = 1$, and this may have $f_1 \neq 0$.

For illustrative purposes, we take the example of smooth contour inputs in Fig. (5B), and assume no suppression between contour elements $\mathbf{W}_{ij} = 0$. Hence, $\lambda_W^k = 0$, and $\gamma^k = -1 + \lambda_J^k/2 \pm i\sqrt{g_y' g_x' - (\lambda_J^k)^2/4}$ with the corresponding eigenvectors $\mathbf{X}_j^k = e^{i f_k j}$. The mode $k = 1$ with the highest $Re(\gamma^k)$ is the one with largest λ_J^k which occurs for $f_1 = 0$. In this case, $\mathbf{X}_j^1 = \mathbf{X}_i^1$, and $\lambda_J^1 = \sum_j \mathbf{J}_{ij}$. Thus the contour input does not induce spatial symmetry breaking although sufficient contour enhancement $\sum_{j \neq i} \mathbf{J}_{ij}$ increases λ_J^1 and can push $Re(\gamma^1)$ positive to generate synchronized oscillations among the contour segments (see Fig. (11B), Li 1998). This conclusion does not depend on the detailed form of $\mathbf{J}_{ij}$ as long as $\mathbf{W}_{ij} = 0$.

On the other hand, under one-dimensional non-contour inputs such as Fig. (5C,E), the interaction and dynamics are quite different. If one assumes that $\mathbf{J}_{ij} = \mathbf{J}_{ii}\delta_{ij}$. Then, $\lambda_J^k = \mathbf{J}_{ii}$, and $\gamma^k = -1 + \mathbf{J}_{ii}/2 \pm i\sqrt{g_y'(g_x' + \lambda_W^k) - \mathbf{J}_{ii}^2/4}$ for all k. Since $Re(\gamma^1) < -1 + \mathbf{J}_{ii} = -1 + J_o g_x'(\bar{x})$, the system is stable when there is insufficient self-excitation $J_o < 1/g_x'(\bar{x})$. This condition is always satisfied in a network of well behaved individual neurons for the following reason. An isolated principal unit x follows equation $\dot{x} = -x + J_x g_x(x) + I$. One can easily check that, under zero input $I = 0$, too much self-excitation, i.e., $J_o > 1/g_x'(x)$, would lead to unreasonable non-zero steady state output $g(\bar{x})$ which is a solution of the equation $-\bar{x} + J_x g_x(\bar{x}) = 0$. Hence input patterns in which there are no excitatory J links between units for the input elements, (as we are assuming for patterns in Fig. (5C,E)), induce well behaved model outputs regardless of the connections W.

The situation is much more complex for a texture input like that in (Fig (7A)). The eigenvectors or modes are now 2-d waves $\mathbf{X}_j^k \propto e^{i f_k j}$. Here $f_k = (f_x(k), f_y(k))$ is a two component (horizontal and vertical) wave vector perpendicular to the peaks and troughs of the waves. Likewise, locations of the texture bars are also described by a two component vector $j = (m_j, n_j)$. In the example when $f_k = (f_x(k), 0)$ is in the horizontal direction, we have

$$\lambda_J^k = \sum_j \mathbf{J}_{aj} e^{i f_k (a-j)} \tag{42}$$

$$= \sum_{m_j, n_j} \mathbf{J}_{aj} e^{i f_x(k)(m_a - m_j)} \tag{43}$$

$$= \sum_{m_j} e^{i f_x(k)(m_a - m_j)} \sum_{n_j} \mathbf{J}_{aj} \tag{44}$$

$$= g'(\bar{x}) \sum_b J_b' e^{i f_x(k) b} \tag{45}$$

$$\text{Similarly} \quad \lambda_W^k = g'(\bar{x}) \sum_b W_b' e^{i f_x(k) b} \tag{46}$$

where J_b' and W_b' are the effective connections between two texture columns as defined in equation (21). This array of columns is then similar to the one dimensional array of bars above. However, the column-to-column connections J' and W' are stronger than the bar-to-bar counterparts. More importantly, the connection structure J' and W' or the the spectrum (λ_J^k, λ_W^k) depend on the orientation θ_1 of the texture bars. Unlike the situations for one dimension arrays of bars, the simplification $J' = 0$ or $W' = 0$ between different columns is no longer reasonable for some θ_1.

Let us consider the example when the texture bars are horizontal and parallel to f_k, i.e., $f_k = (f_x(k), 0)$, $\theta_1 = 90^o$, corresponding to the contour input discussed above. The inhibition $W_b' \neq 0$ between vertical columns is usually non-zero, unlike what one can assume between bars in a contour. However, with the bow-tie connections for which co-aligned bars are linked by J and non-aligned bars are linked by W, it is easy to see that J_b' is strong between different columns, while W_b' is strong within a column ($b = 0$) and weak between columns. To a good approximation, $W_b' \approx \delta_{b0} W_0'$, i.e., λ_W^k is roughly independent of k, and hence, $\gamma^k \approx -1 + \lambda_J^k/2 \pm i\sqrt{g_y'(g_x' + W_0') - (\lambda_J^k)^2/4}$. In this case, $Re(\gamma^k)$ is largest when λ_J^k is largest, and this

happens at $f_x(k) = 0$, which is a translation invariant mode. This means that, for a texture of identical bars, translation invariance is not likely to be broken in the direction parallel to the orientation of the bars, i.e., for an input texture of horizontal bars, the model using the bow-tie connections is unlikely to produce vertical columns of different saliencies.

However, if the texture bars are vertical (consider still the mode for $f_k = (f_x(k), 0)$), the bow-tie connections result in strong enough facilitation J'_b between different vertical columns. Further, the suppression W'_b between columns also depends non-trivially on b. Hence, both λ^k_J and λ^k_W depends on f_k non-trivially. Even though both λ^k_J and λ^k_W are largest for $f_k = 0$, by equation (40), the dominant mode with the largest $Re(\gamma)$ is quitely likely to be the one with $f_x(k) \neq 0$. The abundance of facilitation between the bars means that J'_b or λ^k_J can be strong enough to induce unstable modes. Thus, for texture inputs of identical bars, models using the bow-tie connections have a stronger tendency to hallucinate saliency columns parallel to the orientation of the bars.

A well behaved model should be designed such that the stability condition in equation (41) is satisfied considering all f_k whether it is parallel, perpendicular, or oblique, to the texture bars. Given f_k, it is easy to calculate λ^k_J, λ^k_W, and thus γ^k using equations (38) and (39). The spectrum λ^k_J and λ^k_W depend on the orientation of the bars and on the spatial arrangements of the texture bars (i.e., whether it is arranged in a Manhattan grid or some other fashion) via the summation in equations (38) and (39). Given a system that does not form patterns spontaneously from homogeneous inputs, oscillations can only happen when the dominant mode is the translation invariant one with $f_1 = 0$ and when $Re(\gamma^1) > 0$. Since $\gamma^1 = -1 + \lambda^1_J/2 \pm i\sqrt{g'_y(g'_x + \lambda^1_W) - (\lambda^1_J)^2/4}$, $Re(\gamma^1) > 0$ only when λ^1_J is large enough. For $f_1 = 0$, $\lambda^1_J = \sum_j \mathbf{J}_{ij}$. Consequently, λ^1_J is large enough and oscillations happen when the stimulus is such that the each stimulus bar receives more excitatory-to-excitatory connections J_{ij} from other bars in the visual input (Koenig and Schillen 1991). This may explain why neural oscillations are observed in some cases and not others in physiology. Under the bow-tie connections, a large texture input is more likely to induce neural oscillation than a long contour input, which is in turn more likely to induce oscillation than, say, a horizontal array of vertical bars or isolated individual bars, see Fig. (11). This prediction can be physiologically tested. It has indeed been observed that grating stimuli are more likely to induce oscillations than bar stimuli (Molotchnikoff, Shumikhina, and Moisan, 1996).

A model of the visual cortex should be designed such that it performs the desired computations such as contour enhancement and region segmentation. This computation places qualitative and quantitative requirements on the neural connections J and W through, e.g., equations (32) and (33), and conditions on minimal required contour enhancement $F_{contour}$ and texture suppresion $C_{whole-texture}$. These computational requirements on J and W have to be satisfied simultaneously with the stability requirements on J and W (equation (41)). Had we used a recurrent network with symmetric connections, as in equation (1), the stability conditions would involve the spectrum of eigenvalues of the matrix $\mathbf{T}$, where $\mathbf{T}_{i\theta j\theta'} = T_{i\theta j\theta'} g'_x(\bar{x}_{j\theta'})$. A stable fixed point should have $\lambda^{\mathbf{T}} < \alpha_x$ all the eigenvalues $\lambda^{\mathbf{T}}$ of this matrix, a condition that can be shown to be difficult to satisfy given the required visual computations (Li and Dayan 1999).

4   Summary and Discussion

We have studied the nonlinear neural dynamics in a recurrent model of primary visual cortex. An analytical understanding of the dynamics has been essential to reveal the computational potential of the recurrent model, a facet that is usually very difficult to understand using merely computer simulations. Our analysis has been used to build a model that successfully explains much physiological and psychophysical data (Li 1999a,b,c). The various simplifi-
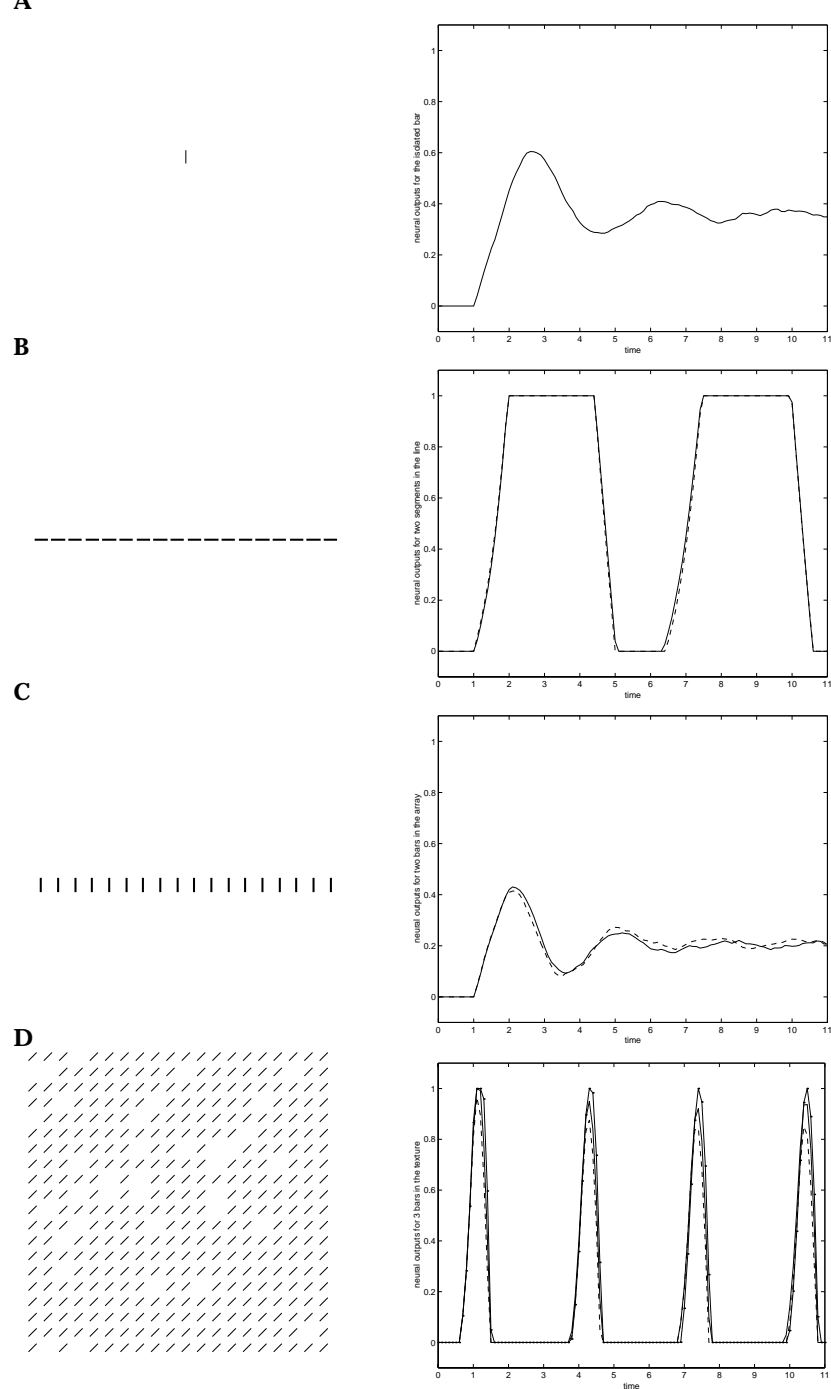
Figure 11: Different stimuli give different tendencies to oscillatory responses. Shown here are the visual input stimuli (all appear at time zero and stay on) and the time course of the neural activities. **A** an isolated bar and the neural response which stablizes after the initial oscillatory transients. **B** An input contour and the synchronized and sustained oscillatory neural responses from 2 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) from two non-neighboring neurons. **D:** An input texture (with some holes in it), and the sustained oscillatory responses from 3 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 bars next to the holes in the textures induce a little bit higher responses.

cations and approximations in the analysis can be justified by the fact that they have led to the design of computationally appropriate connections J and W in the model that give relative higher responses to smooth contour, pop-out targets, and boundaries between different

(simple or complex) regions, whilst not hallucinating spatial patterns not present in inputs (Li 1998, 1999a). The analysis techniques in this paper can be applied to other recurrent networks whose neural connections are translationally symmetric.

We presented in this paper an example of how nonlinear neural dynamics link computations with the model architecture, neural elements, and neural connections. Additional or different computational goals, including the ones which maybe performed by the primary visual cortex and which are not yet modelled by the model example studied here, should call for a more complex or different model design. For example, for contour inputs, a different computation could be (1) to fill-in the short gaps in a contour under low signal-to-noise inputs when the gaps are likely caused by input noise, but, (2) to highlight the locations near the gaps (to draw visual attention) without filling-in under high signal-to-noise inputs when the gaps are likely caused by the visual scene itself. Another example of desired computation is to prevent the spontaneous saliency differentiation even when the input is not homogeneous in the image plane but is generated from homogeneous flat texture surfaces slanted in depth, thus generalizing the translation invariance computation studied in section 3.5. This computation should require multiscale image representations and recurrent interactions between cells tuned to different scales. By studing the recurrent nonlinear dynamics and analysing the link between model structures and model computation, 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.

References

[1] Braun J., Niebur E., Schuster H. G., and Koch C. "Perceptual contour completion: A model based on local, anisotropic, fast-adapting interactions between oriented filters". (24th Annual Meeting of the Society for Neuroscience, Miami Beach, Florida, USA, November... Society for Neuroscience Abstracts, v.20, n.1-2, 1665 (1994).

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

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

[4] 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, 1993

[5] 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.

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

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

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

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

[10] 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.

[11] Hopfield, JJ (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science* **81**:3088-3092.

[12] 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.

[13] Kapadia, M.K., private communication. 1998.

[14] 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.

[15] Koenig P. and Schillen T.B. (1991) Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. Neural Computation 3:155-166

[16] 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, 1998.

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

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

[19] Li Zhaoping (1999b) Pre-attentive segmentation in the primary visual cortex. *Spatial Vision* Vol. 13, p.25-50.

[20] Li Zhaoping (1999c) "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

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

[22] Molotchnikoff S, Shumikhina S, Moisan LE "Stimulus-dependent oscillations in the cat visual cortex: differences between bar and grating stimuli." *Brain Res* 1996 Aug 26;731(1-2):91-100

[23] 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

[24] Polat U. and Sagi D. (1993) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vis. Res.* 33, p993-999.

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

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

[27] G. M. Shepherd *The synaptic organization of the brain* Second Ed. 1979, Third ed. 1990. Oxford University Press

[28] 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.

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

[30] White E. L. (1989) *Cortical circuits* (Birkhauser).

[31] Yen S-C. and Finkel L.H. "Salient contour extraction by temporal binding in a cortically-based network" *Natural Information Processing Systems* Denver. 1996

[32] 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.

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