Neural circuit models for computations in early visual cortex
Li Zhaoping
Department of Computer Science, University College London, UK

Abstract:
Some visual circuit models study the mechanisms of the mapping from visual inputs to responses in individual neurons. They model neural properties such as feature tuning and input sensitivity, and how these properties are modified by intra-cortical connections and contextual inputs. Other cortical models are more concerned with neural responses at a spatial population level, to identify the spatio-temporal visual inputs favored and extracted by the visual system, the computational goal of the input-to-response transform, and the cortical mechanisms responsible. The complexity of a cortical circuit model depends on the question asked. Modeling neural circuits of many interacting hypercolumns is a necessary challenge, which is providing insights to the cortical computation and linking physiology with global visual cognitive behavior.

Introduction
A full computational understanding of the visual cortex requires understanding the processes from visual inputs to neural responses and from neural responses the visual behavior. There are two broad classes of computational models for the early visual cortex. The first class focuses on modeling the mapping from inputs to neural responses. The interested input-to-response mapping include, for example, (1) visual (e.g., to orientation, spatial frequency) feature tuning of neural responses, (2) the mapping from the strength of input features to response levels, and (3) how the responses to the preferred stimuli (e.g., preferred orientation) are modulated by additional visual inputs at the same or surrounding locations. The second class of models is interested in the role of the cortex in achieving certain computational or cognitive tasks, and in how the cortical mechanisms enable these roles. These computational issues include, for example, (1) how neural circuits enable invariance of the orientation tuning width of a neuron to input contrast, (2) how intra-cortical interactions enable optimal estimations of input features (e.g., input orientation) from noisy neural responses, (3) whether the cortex serves to guide bottom-up attentional selection, to perform perceptual grouping, and/or to serve other roles.

Among the simplest cortical models are ones of two interacting neurons, an excitatory principal neuron connected with an inhibitory interneuron[1] (Fig. (1A)), often used to study gain control by the strength and form of the direct and contextual visual inputs. More complex are models with multiple interacting principal neurons tuned to different feature values in a single feature dimension such as orientation or spatial location[2–4, 4, 5] (Fig. (1B)), They can study feature tuning and response selectivity of input patterns (e.g., oriented inputs or particular spatial patterns) in the feature dimension concerned. However, among the most meaningful inputs for visual cognitive behavior are object contours and surface textures. Their description requires two feature dimensions, orientation and spatial location. This calls for more complex models of interacting neurons, each is tuned to space and orientation and prefers one of the oriented image elements, and the interaction between neurons depend on their preferred features in both dimensions (Fig. (1C)). More generally, one of the feature dimensions can also be color, spatial frequency, depth, or other dimensions. Understanding these complex circuits is necessary to investigate the computational role of the cortex[6, 7].
The representative circuit models

A circuit of an excitatory and an inhibitory (EI) neurons for input gain control

This models a local circuit of a principal excitatory (E) cell, modeling a local group of cortical pyramidal cell, connected with a model inhibitory(I) interneuron (also modeling a local interneuron group), in response to feedforward visual inputs from lateral geniculate nucleus (LGN), as well as lateral or feedback inputs via intra-cortical/recurrent or feedback connections. This is called a E-I circuit with an E-I pair. Let membrane potentials of the E and I be \( x \) and \( y \) respectively, with their neural responses \( g_x(x) \) and \( g_y(y) \) as nonlinear, sigmoid-like, functions of \( x \) and \( y \), the equations for the circuits are

\[
\begin{align*}
\dot{x} &= -\alpha_x x + J_{ee} g_x(x) - J_{ei} g_y(y) + I_e \\
\dot{y} &= -\alpha_y y + J_{ie} g_x(x) - J_{ii} g_y(y) + I_i
\end{align*}
\]

where \( J_{ee}, J_{ie}, J_{ei}, \) and \( J_{ii} \) model the synaptic connection strengths between the neurons, while \((I_e, I_i)\) are inputs converged from the thalamic inputs \( I_t \) and lateral or central inputs \( I_c \).

\[
\begin{align*}
I_e &= T_e I_t + C_e I_c \\
I_i &= T_i I_t + C_i I_c
\end{align*}
\]

where \( T_e, T_i, C_e \) and \( C_i \) are the synaptic weights from thalamic and central/lateral sources. The fixed point \((\bar{x}, \bar{y})\), where \( \dot{x} = \dot{y} = 0 \), typically approximates the temporal averages of neural activities \( x \) and \( y \) under static inputs \((I_t, I_c)\).

This circuit models the following phenomena observed physiologically[1]. The response \( g_x(\bar{x}) \) vs. visual input \( I_t \) exhibit threshold and saturation and even super-saturation behavior. The contextual input outside the classical receptive field of the E-I pair (with substantial \((C_e, C_i)\) when they are parallel to the optimal inputs within the receptive field) enhances \( g_x(\bar{x}) \) at low input \( I_t[8, 9] \) and suppresses \( g_x(\bar{x}) \) at high \( I_t[1, 9–11] \). These effects are mainly caused by the inhibition from the I cell to the E cell, and the inhibition is stronger when the I cell is activated and more sensitive to inputs, as modelled explicitly below (see Fig. (2A)).

The sensitivity of the response \( g_x(\bar{x}) \) to thalamic input \( I_t \) can be shown as(see[12] for a proof in simplified case)

\[
\frac{\delta g_x(\bar{x})}{\delta I_t} = \frac{g_x'(\bar{x})}{(\alpha_x - J_{ee} g_x'(\bar{x}))(\alpha_y + J_{ii} g_y'(\bar{y})) + J_{ei} J_{ie} g_x'(\bar{x}) g_y'(\bar{y})}
\]

It depends on \((\bar{x}, \bar{y})\) through the slopes \( g_x'(\bar{x}) \) and \( g_y'(\bar{y}) \) of \( g_x(\bar{x}) \) and \( g_y(\bar{y}) \), increasing with E cell’s gain \( g_x'(\bar{x}) \) but decreasing with I cell’s gain \( g_y'(\bar{y}) \). It is highest with weak thalamic and lateral/feedback inputs \( I_t \) and \( I_c \) such that the I cell is not activated and \( g_y'(\bar{y}) = 0 \), giving input sensitivity

\[
\frac{\delta g_x(\bar{x})}{\delta I_t} = \frac{g_x'(\bar{x}) T_e}{\alpha_x - J_{ee} g_x'(\bar{x})}
\]

in which the negative feedback from the I cell is non-existent. Increasing thalamic input \( I_t \) raises \( y_t \) either directly via \( T_t I_t \) or indirectly via E to I excitation, making \( g_y'(\bar{y}) > 0 \) and thereby reducing the input sensitivity \( \frac{\delta g_x(\bar{x})}{\delta I_t} \). When the I cell is activated sufficiently, E cell’s response to \( I_t \) saturate or super-saturate, i.e., \( \delta g_x(\bar{x})/\delta I_t \leq 0 \), when \( (\alpha_y + J_{ii} g_y'(\bar{y})) T_i \leq J_{ei} g_y'(\bar{y}) T_i \), even before the E cell saturates by itself when \( g_y'(\bar{y}) = 0 \).

Meanwhile, contextual visual inputs outside the receptive field of the E-I pair typically cause lateral input \( (\delta I_c, \delta I_t) = (C_e I_c, C_i I_c) \) to the E-I pair, more so when they are parallel to the inputs.
within the receptive fields[13–16]. According to equation (7), the change in $g_x(x)$ caused by the lateral inputs $(\delta I_e, \delta I_l) = (C_e I_e, C_l I_l)$ is

$$
\delta g_x(x) = \frac{g'_x(x) I_e}{(\alpha_x - J_x g'_y(y)) + (\alpha_y + J_y g'_y(y))} \frac{(\alpha_y + J_y g'_y(y)) C_e - J_x g'_y(y) C_l}{(\alpha_x - J_x g'_y(y)) + (\alpha_y + J_y g'_y(y))} g'_y(y)
$$

Hence, $I_e$ is facilitative when

$$(\alpha_y + J_y g'_y(y)) C_e > J_x g'_y(y) C_l$$

and suppressive otherwise. Hence, the contextual influence depends on the I cell’s state $y$. Under weak feedforward input $I_f$ when the I cell is weak and $g'_y(y) \approx 0$, contextual influences are facilitatory. Stronger input $I_f$ activates the I cell and increases $g'_y(y)$, and can switch the contextual influences to suppressive, as physiologically observed. When the contextual inputs are perpendicular to the (optimal) input within the receptive field of the EI pair, $(C_e, C_l)$ are much smaller and thus reducing the lateral inputs $(\delta I_e, \delta I_l) = (C_e I_e, C_l I_l)$, making the E response $g_x(x)$ vs. $I_f$ more similar to the case without the contextual inputs.

This circuit with its nonlinear interactions between the E and I neurons has also been used to understand other phenomena such as context enabled learning[17] and faster responses to inputs while having positive feedback amplifications[18]

A recurrent network, e.g., a model hypercolumn, of neurons for a single visual feature dimension

To model feature tuning or selectivity to input patterns, multiple principal cells, each tuned to different features at least in a single feature dimension, are needed. For example, one could model how a hypercolumn of interacting neurons $i$, preferring orientations $\theta_i$ spanning the whole orientation range, responds to a pattern of input orientations $I = (I_1, I_2, ..., I_i, ..., I_N)$. An $I_i = A + C \cdot \cos(\theta_i - \theta)$ for $|\theta_i - \theta| \leq \pi/2$ (with periodic boundary condition identifying $\theta = 0$ and $\theta = \pi$) and $I_i = 0$ otherwise models an input oriented at $\theta$. This hypercolumn could be modelled as

$$
\dot{x}_i = -x_i + J_{ij} g(x_j) - I_i
$$

where $J_{ij}$ models the synaptic connection from neuron $x_j$ to $x_i$. Alternatively, the feature dimension can be spatial locations, so that $I_i$ models feedforward input at location $i$ and the circuit models cortical responses $g(x) = (g(x_1), g(x_2), ...)$ to spatial input patterns $I$. Due to translation invariance, $J_{ij}$ depends only on the difference $|i - j|$ or $|\theta_i - \theta_j|$, and $J_{ij} = J_{ji}$. Here, for simplicity, a neuron can be excitatory to some post-synaptic cells and inhibitory to others in many such models. These models aim to understand how recurrent interactions modify or enhance the initial responses $g(x)$ to the feedforward input $I$ which are the outcome of convoluting the raw visual inputs with the feedforward receptive fields. The hypothesized goals of the recurrent processing include those to sharpen the orientation tunings of the neurons, to made the responses to a noisy input pattern less noisy, to amplify responses to weak input patterns or even to maintain responses in the absence of inputs[2–5, 19].

Networks with symmetric connections between any two neurons are known as Hopfield networks[20]. Their states under static inputs always converge to stable (static) states. In the network of equation (9), this converged state $x$ satisfies

$$
-x + J g(x) - I = 0
$$

where $J$ is a matrix with elements $J_{ij}$ and $(J g(x)_i) = \sum_j J_{ij} g(x_j)$. Strong enough $J$ can give non-zero state $\bar{x}$ under zero input $I = 0$. For example, if $J_{ij} = 1.1$ for $|i - j| \leq 1$ and $J_{ij} = -1$ otherwise,
and if $g(x)$ is a step function with $g(x) = 1$ for $x \geq 1$ and $g(x) = 0$ otherwise, then a stable activity pattern is the unimodal $g(\bar{x}) = (..., 0, 0, 1, 1, 0, 0, ...)$. If a transient input $I$ brings the network near this state, the network is likely to settle and remain in it as if it is a memory. For example, if the feature space is orientation, and an initial noisy input pattern $I$ is biased towards orientation $\theta$, the response $g(x)$ can behave as if the input is noise free and oriented near $\theta$. Largely shaped by recurrent connections $J$, $x$ and the orientation tuning width of neurons are less sensitive to the contrast and anisotropy of the input $I$ than the they are when $J = 0$, and the neural tuning width could also be made narrower than that in the feedforward input $I_i$.[4, 5] (see Fig. (2B)). These properties can also help to optimally estimate the visual input orientation $\theta$ from a noisy $I$ by locating the peak position $\theta_i$ in the population activity pattern $g(x)$.[19] However, insensitivity of response patterns $g(x)$ to the actual input pattern $I$ makes the network unsuitable for faithfully representing input features. For example, the network could hallucinate one oriented input by an unimodel response pattern $g(x)$ while the actual input is non-oriented or have a bimodal pattern caused by two input orientations. However, using an E-I network orientation tuned E-I pairs[21], amplification of oriented inputs without tendencies to hallucinate input patterns can be achieved.

A circuit of neurons tuned to both space and orientation (or other) features

To study spatial patterns of input features such as orientation, models should include visual features in at least two dimensions: space, indexed by $i$, and the interested feature dimension such as orientations, color, motion, depth, or other features, indexed by $\theta$. When $\theta$ denotes orientation, visual patterns like object contours and surface textures can be represented as spatial configurations of oriented elements ($i, \theta$). One such model[12, 22] is a spatial array of interacting hypercolumns $i$, each composed of E-I pairs ($i, \theta$) tuned to different features $\theta$:

$$x_{i\theta} = -\alpha_x x_{i\theta} - g_y(y_{i\theta}) - \sum_{\Delta \theta \neq 0} \psi(\Delta \theta) g_y(y_{i,\theta+\Delta \theta}) + J_o g_x(x_{i\theta}) + \sum_{j \neq i, \theta'} J_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta} + I_o$$

(11)

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

(12)

where $I_{i\theta}, x_{i\theta}, y_{i\theta}, g_x(x_{i\theta}), g_y(y_{i\theta})$ are analogous to the corresponding variables in equations (1-2), $J_o$ is the self-excitation factor of E neurons, $I_o$ and $I_c$ model background inputs, including noise, feature unspecific surround suppressions, and central feedbacks, $\psi(\Delta \theta)$ models connection strength for intra-hypercolumn interactions, $J_{i\theta,j\theta'}$ and $W_{i\theta,j\theta'}$ are synaptic connection strength from the E neurons to other E and I neurons respectively. Suitable neural interactions make response patterns $g(x) = (..., g(x_{i\theta}), ..., g(x_{j\theta'}), ...)$. enhance some visual inputs $I = (..., I_{\theta_1}, ..., I_{\theta_\nu}, ...)$ relative to others. For example, when the dominant interaction between colinearly aligned elements ($i, \theta$) is mutual excitation via $J_{i\theta,j\theta'}$ and that between near-parallel and non-aligned elements is mutual disinaptic inhibition via $W_{i\theta,j\theta'}$, responses to smooth contours, to orientation pop-outs (e.g., a vertical bar among horizontal bars), or to texture borders can be enhanced relative to those to homogeneous textures (see Fig. (2C)). It can be shown[7, 21] that, in order to achieve sensitive amplifications of conspicuous input elements relative to other inputs without sensory hallucinations, it is necessary to model the cortical circuit as an E-I network whose oscillatory tendency[23] have non-intuitive computational benefits, rather than a simplified Hopfield-like network[6, 24] (like in the recurrent circuit described by equation (9) with symmetric connection strengths between neurons. Harnessing this complex E-I circuit model for well-behaved responses to meaningful visual
inputs was instrumental in developing a theory that V1, via its intra-cortical interactions, creates a bottom-up saliency map to guide attention to conspicuous locations[25–27]. This theory subsequently provided non-trivial predictions, such as the non-intuitive one that visual attention can be captured by perceptually non-distinctive eye-of-origin singletons (e.g., a vertical bar presented to the left eye among many identical vertical bars to the right eye), which are subsequently confirmed experimentally[28–30].

Other E-I circuit models of comparable complexity have been used to demonstrate that V2’s neural tunings to surface border ownership could in principle emerge from intra-cortical interactions[31], and that a single network for stereo matching in random dot stereograms could also enhance responses to depth singletons and borders of depth surfaces as observed in V2 cells[32] while a Hopfield-like network for stereo matching could not[33]. There are other complex cortical circuit models involving space and other visual features, with added complexity including interactions between multiple visual areas or cortical laminar layers[6, 34]. Proposed functions of these model circuits include visual segmentation, boundary completion, and feature filling-in. However, an insufficient understanding of these complex models has so far made it difficult to make precise the proposed functions of the circuit.

1 Concluding remarks

We have limited intuitions of the neural circuit models of more than two neurons. It is important to find a minimum model which has all the necessary complexity, but no more, to address the problem of interest. The three levels of model complexities in this paper, without using spiking neural models for example, demonstrate the complexity levels and problems that they can address or model. A complex model, such as the one of interacting hypercolumns or E-I pairs in equations (11-12) should also be first understood at the level of their simpler components, such as one of an E-I pair in equation (1-2) or a hypercolumn. It remains an exciting challenge to model neural circuits appropriately to discover and understand their computational roles.

Acknowledgement Work supported by the Gatsby Charitable Foundation.

References


A paper with clear analysis of how a model circuit for an E-I pair explains input gain control and contextual influences which switches from facilitatory to suppressive as direct input to the E-I papr increases


Perhaps the first model of spontaneous pattern formation by recurrent interactions, qualitative the same as the later models of visual hallucinations in space and orientation domain.


An earlier explicit model of visual hallucination in a recurrent cortical model.

Similar to Ben-Yishai et al. 1995, with more physiological considerations.


A well known ring model, a recurrent model of hypercolumn like the one in Fig. 2B.


A mathematical analysis on how the V1 saliency map model is designed.


First model of relative contour enhancement against backgrounds by model V1 neurons and their interactions. This is the model in equations (11-12) and its performance is demonstrated in Fig. 2.


The nonlinear interaction with an E-I pair is modelled to account for perceptual learning of contrast discrimination enabled by contextual inputs.


A clear analysis of how E-I interactions enable fast dynamic responses to amplified inputs.

Recurrent interactions in the hypercolumn model is used to estimate input orientation when there is only one input orientation.


Clear analysis of a model which is now known as the Hopfield model, which has synaptic connection strength between any two connected neurons and always converge to static stable states under static input. It can be used to model functions like memory recall and error correction. Many of the subsequent models of working memory and decision making are based on the Hopfield model.


Mathematical analysis why a E-I circuit model is necessary for the V1 saliency computation.


Same model as in Li 1998, first to show the same V1 model to enhance visual contours can also compute saliency in general, giving feature pop-out and texture segmentation.


One of the first analytical illustrations that an E-I pair is a neural oscillator, and that interactions between the E-I pairs lead to non-trivial computations.


Extension of the visual spatial hallucination model by Ermentrout and Cowan 1979 with the addition of orientation feature dimension.

[25] Li Z: Contextual influences in v1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci USA* 1999, **96**:10530–10535.

Showing that the capability of the V1 model in Li 1998, 1999.


A V2 model involving space and border-ownership feature dimensions, to demonstrate that V2 elements (with its neurons and intra-cortical interactions) are in principle sufficient for the emergence of neural tuning to border ownership.

A V2 model involving space and disparity feature dimensions, for stereo matching while having saliency highlights for disparity singletons and surface boundaries.


One of the first recurrent models for meaningful visual computation, in this case, a cooperative algorithm for stereo matching.

Figure 1: Three levels of complexity to model visual cortical circuits. 

**A:** an excitatory-inhibitory (EI) network of two neurons preferring a single visual feature, e.g., a particular orientation $\theta$ at a particular location $i$, in response to feedforward visual inputs, and contextual recurrent inputs or central feedback inputs. 

**B:** a network of neurons for a single feature dimension, e.g., a ring model of a cortical hypercolumn of neurons tuned to various orientations $\theta$ spanning $0 < \theta \leq 180^\circ$, or a model of neurons whose receptive fields $i$ collectively span the visual field. 

**C:** a network of neurons for processing two feature dimensions, space (indexed by $i$) and another dimension (indexed by $\theta$) such as orientation. Different neurons are tuned to different $(i, \theta)$. When $\theta$ is for orientation, the network processes spatial arrangements of orientation features $\{I_{i\theta}\}$, such as object contours and surface textures. Each computational issue is best addressed by a model of suitable complexity.
Figure 2: Example performances of models at the three levels of complexity in Fig. (1). A: an E-I pair model as in equations (1-2), with its sigmoid firing rate functions $g_x(x)$ and $g_y(y)$ and its input gain control under different contextual conditions. B: a hypercolumn circuit model of interacting neurons tuned to orientation, as in equation (9). Lateral connections help to smooth, amplify and sharpen oriented input patterns. C: a model of interactions between E-I pairs in various hypercolumns and tuned to different orientations, as in equations (11-12). The model gives relatively higher responses to conspicuous image locations.