A Local Circuit Integration Approach to Understanding Visual Cortical Receptive Fields

David C. Somers, Emanuel V. Todorov, Athanassios G. Siapas, and Mriganka Sur
Department of Brain and Cognitive Science, MIT, Cambridge, MA 02139
somers@ai.mit.edu, emo@ai.mit.edu, thanos@ai.mit.edu, msur@wccf.mit.edu

Abstract

The traditional concept of the receptive field (e.g., [4, 6]) holds that each portion of the receptive field (RF), in response to a stimulus element, has unitary (excitatory or inhibitory) influence on neuronal response. Here, we argue: i) receptive field components naturally have dual or vector (both excitatory and inhibitory) influence; ii) neuronal integration is better understood in terms of local cortical circuitry than single neurons. Using a large-scale model of primary visual cortex, we demonstrate that the net effect of a given stimulus element within either the classical or extraclassical RF can switch between excitatory and inhibitory as global stimulus conditions change. We analyze and explain these effects by constructing self-contained modules (via a novel technique) which capture local circuit interactions. These modules illustrate a new vector-based RF analysis which unifies notions of classical and extraclassical RF, treating long-range intracortical inputs on equal footing with thalamocortical inputs.

Introduction

Neuronal receptive fields in primary visual cortex (V1) have not only "classical" regions, where visual stimuli elicit responses (presumably through thalamocortical axons), but also have "extraclassical" regions, where stimuli largely modulate responses evoked by other stimuli (presumably via long-range intracortical or inter-areal axons) [3, 16]. The traditional view of integration holds that each portion of a neuron's receptive field in response to a given stimulus element has either an excitatory or an inhibitory (i.e., a scalar) influence [4, 6, 15]. Although this approach has substantial explanatory power, it cannot account for phenomena in which the net effect of a stimulus element in a given portion of the receptive field appears to switch between excitatory and inhibitory as global stimulus conditions change [8, 10, 17, 19]. Two such phenomena, involving local and long-range integration respectively, are paradigmatic. First, increasing the luminance contrast of an oriented visual stimulus causes responses in primary visual cortex to initially increase, but subsequently saturate and even decrease ("super-saturate") [1, 10, 11] (see data of [10] in fig 1a). Second, adding a distal stimulus facilitates responses to a weak central stimulus, but suppresses responses to a strong stimulus [8, 9, 19, 17] (see data of [8, 17] in fig 1b).

Our goal is to develop an expanded notion of the visual cortical receptive field which can explain stimulus-dependent responses such as these. Three basic features of cortical anatomy, which are overlooked by the traditional receptive field view, are central to the expanded view: i) receptive field regions (via either thalamocortical or long-range intracortical axons) drive both excitatory and inhibitory cortical neurons [13, 25]; ii) different portions of the receptive field provide converging inputs to a shared population of cortical neurons [3, 16]; and iii) these neurons form dense, recurrent local connections [3, 7, 25]. Based on this anatomy, we propose that: i) each RF region in response to a given stim-

ulus has both excitatory and inhibitory influences on neuronal responses which in general cannot be reduced to a scalar quantity but rather should be considered separately (i.e., RF input is a vector); ii) receptive field inputs are integrated by the local cortical circuitry; and iii) the net effect of a receptive field input depends both on the excitatory-inhibitory bias of the afferent inputs and on how other receptive field regions activate the local cortical circuitry. First we demonstrate this approach by capturing the paradoxical local and long-range phenomena within a large-scale visual cortical model, and later we present an analytic explanation. In contrast, prior computational investigations of local circuit influences either have captured anatomical details only in simulations with little formal analysis [21, 22] or have oversimplified local cortical excitatory and inhibitory interactions in order to obtain closed-form (scalar) analysis [5].

Methods

Cortical circuitry under a 2.5mm by 5mm patch of primary visual cortex was represented by a model with 20,250 spiking cortical neurons and over 1.3 million cortical synapses. Neurons were organized into a 45 by 90 grid of "mini-columns" based on an orientation map obtained by optical recording of intrinsic signals of cat visual cortex (data from [23]). Each mini-column contains 4 excitatory and 1 inhibitory neurons modeled separately as "integrate-and-fire" neurons with realistic currents and experimentally-derived intracellular parameters [12] (see methods of [21] for equations and parameters). Intracortical connections provide short-range excitation (connection probabilities fall linearly from $\rho_{excit-excit} = 0.1$, $\rho_{excit-inhib} = 0.1$ at distance zero to $\rho = 0$ at $d = 150\mu$m), short-range inhibition (linear from $\rho_{inhib-excit} = 0.12$, $\rho_{inhib-inhib} = 0.06$ at $d = 0$ to $\rho = 0.5\rho_{peak}$ at $d = 500\mu$m; $\rho = 0$ elsewhere), and long-range excitation (linear with orientation difference ϕ between pre- and post-synaptic columns, from $\rho = 0.005$ at $\phi = 0°$ to $\rho = 0.001$ at $\phi = 90°$). Peak synaptic conductances, by source, onto excitatory cells are $g_{excit} = 7$nS, $g_{inhib} = 15$nS, $g_{lgn} = 3$nS, and $g_{long} = 1.2$nS and onto inhibitory cells are $g_{excit} = 1.5$nS, $g_{inhib} = 1.5$nS, $g_{lgn} = 1.5$nS, and $g_{long} = 1.2$nS. Cortical magnification is 1 mm/deg, cortical RF diameters are roughly 0.75°, and thalamocortical spikes are modeled as Poisson processes. Each thalamic neuron projects to cortical neurons over an area 0.6mm^2 and responds linearly with log stimulus contrast. Results are averaged over 20 networks constructed with these probability distributions.

Additional analysis was performed by constructing self-contained modules which capture local circuit properties of the large-scale models. Given a local neuronal population P whose mean firing rate $\mathbf{M} = \mathbf{F}(\mathbf{I}_d, \mathbf{I}_l)$ is a function $\mathbf{F}$ of the long-distance (intracortical and thalamocortical) inputs $\mathbf{I}_d$ and local (intracortical) inputs $\mathbf{I}_l$, we want to construct a closed system (module) whose response approximates $\mathbf{M}$ as a function of $\mathbf{I}_d$ only. All bold face quantities denote vectors with components corresponding to excitatory and inhibitory populations. Local inputs are defined as arriving from within a radius R, which is chosen to minimize approximation error. Module construction is only possible if $\mathbf{I}_l$ can be expressed as a function of $\mathbf{M}$ and $\mathbf{I}_d$. To that end we use a local homogeneity assumption $\mathbf{M} = \mathbf{I}_l$, i.e. neurons within R (not just P) have mean firing rates $\mathbf{M}$. Thus the module output $\mathbf{M}^*$ is the solution of $\mathbf{M} = \mathbf{F}(\mathbf{I}_d, \mathbf{M})$. This equation can be solved numerically if we model the response functions of integrate-and-fire neurons [18, 24]. Here, we compute $\mathbf{M}^*$ by simulating a module composed of excitatory and inhibitory neurons, in which neurons receive the same average number and strength of synapses as neurons in P receive from within the radius

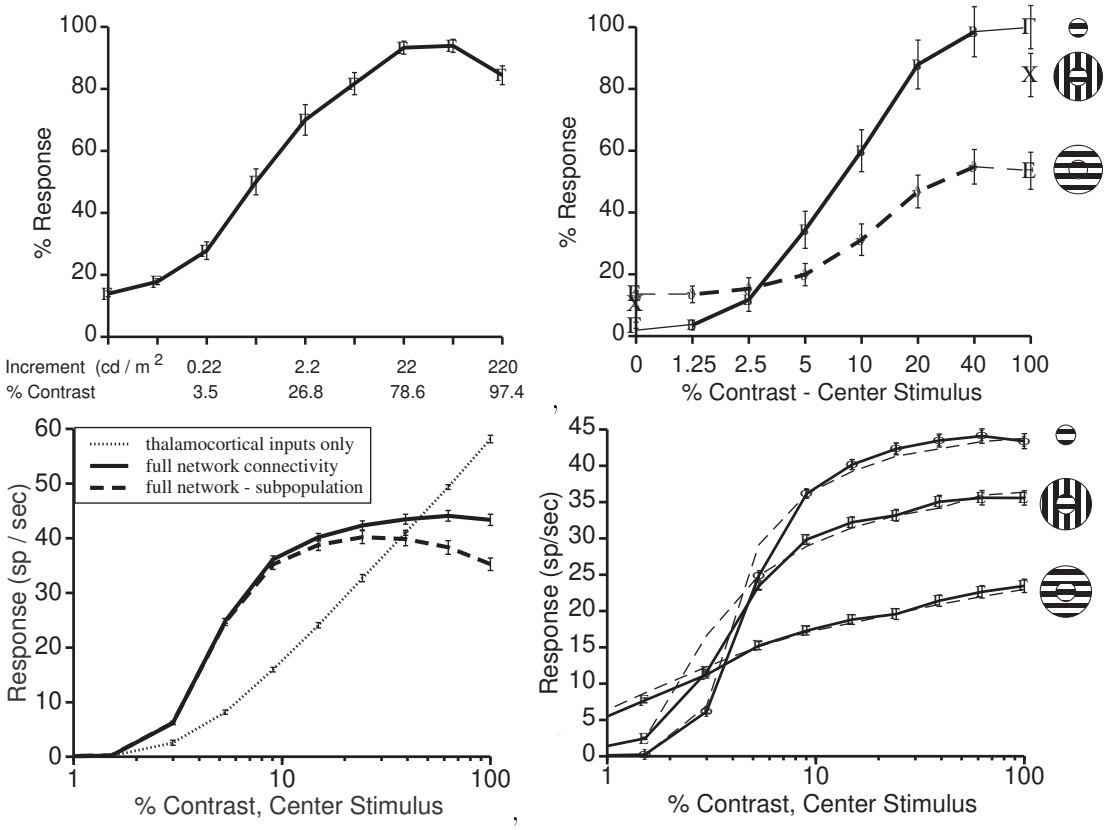

Figure 1: Experimental (a,b) and Simulation Results (c,d) for "super-saturating" contrast response functions (a,c) and surround facilitation/suppression of contrast responses (b,d). Solid and dashed lines in (d) are model and module responses, respectively.

R. The homogeneity assumption is equivalent to isolating P and compensating for the "cut" connections from R by adding extra connections within P. Inhibition is treated as purely local (long-distance inhibition can be addressed by doubling the system dimensions). The radius R that minimizes approximation error is a balance between two conflicting constraints: homogeneity of local firing, which favors smaller R, and inclusion of cortical inhibition, which favors bigger R.

Results

Physiological responses to oriented grating stimuli of differing contrasts within the classical RF are captured by the model (see fig 1c). The responses shown here and below are for the excitatory subpopulation. Responses saturate at contrast levels below which thalamic responses saturate [11], can decline for high contrasts (super-saturation) [1, 10, 11], and have firing rates well below maximal cellular firing rates [12]. Inhibitory neurons, on average, saturate at higher contrasts than do excitatory neurons (not shown). While preserving classical RF properties, our model also captures paradoxical extraclassical RF modulations [8, 9, 19, 17]. The modulatory influence of (fixed contrast) "surround" gratings on responses to optimal orientation "center" stimuli shifts from facilitatory to suppressive as center stimulus contrast increases (see fig 1d; see also [22]). These effects emerge from

the local intracortical interactions (as will be shown below) and do not require synaptic plasticity or complex cellular properties. Our model is the first to provide a unified account of these classical and extraclassical RF phenomena.

We understand the integration of classical and extraclassical RF influences by analyzing local circuitry as a unit. Neuronal responses in the model depend not only on thalamocortical and long-range intracortical inputs [3, 16], but also on recurrent local inputs. We simplify analysis by isolating nonlinear local interactions within a closed system (module) which receives only long-distance (thalamocortical and long-range intracortical) inputs and generates approximately the same mean responses as a local neuronal population embedded in the model. This task is non-trivial, because intracortical connections form a continuum. Simply isolating a small group of cells (together with the connections among them) will remove many local connections from across the group boundary, and thus lead to inaccurate responses. The module we construct preserves the distribution of cellular properties and interactions within the local population, and compensates for the missing local connections by making extra connections within the isolated group (see methods). This module will produce correct responses whenever mean firing rates are locally homogeneous. Note that the method can easily incorporate multiple distinct neuronal subpopulations (e.g. cell types and/or layers), and multiple sources of long-distance input (e.g. feedback projections). This technique differs from "mean-field" approximations (e.g., [20]) in that analysis is local and does not require oversimplification of cellular and network properties.

We construct a module consisting of two interacting homogeneous populations, excitatory and inhibitory neurons (see methods). Afferent inputs to the module excite both neuronal populations and thus must be treated as two-dimensional vectors; this contrasts with standard single neuron RF analyses in which inputs are scalars [4, 5, 6, 15]. Thalamocortical and long-range intracortical inputs are combined linearly (summed) for each subpopulation. Since these two input sources activate excitatory and inhibitory neurons in different proportions, the corresponding input vectors have different angles; vector magnitudes vary directly with stimulus strength. Module responses are a function of the summed input vectors, and mean firing rates of the module's excitatory neurons are completely characterized by the surface plotted in figure 2a. Increasing the contrast of the classical RF stimulus (in the absence of extraclassical stimulation) scales inputs to both cell populations, defining a straight line in the input plane (bottom plane of 2a). Presentation of a fixed surround stimulus activates long-range intracortical inputs; the effect of these inputs can be understood as a simple translation of the contrast input line via vector addition (surround stimulus effects mediated by feedback projections from area V2 can be treated similarly). Contrast response functions (CRFs) predicted by the module are obtained by projecting the resulting input line onto the surface. These predicted CRFs are also shown as the dashed lines in figure 1d. Note that they closely approximate the CRFs generated by the model for all tested stimulus conditions (see fig 1c,d) as well as experimental CRFs (see fig 1a,b). Thus, the paradoxical classical and extraclassical RF integration phenomena are captured by local circuit interactions alone.

Local interactions are described by module response surface shapes. The surface shape shown in figure 2a is characteristic of a large class of recurrently connected excitatory-inhibitory circuits and can be thought of as providing generalized gain control. Note that integrate-and-fire neurons have approximately threshold-linear feedforward responses (fig 1c), and thus the module output (fig 2a) is a smoothed version of an underlying piecewise-

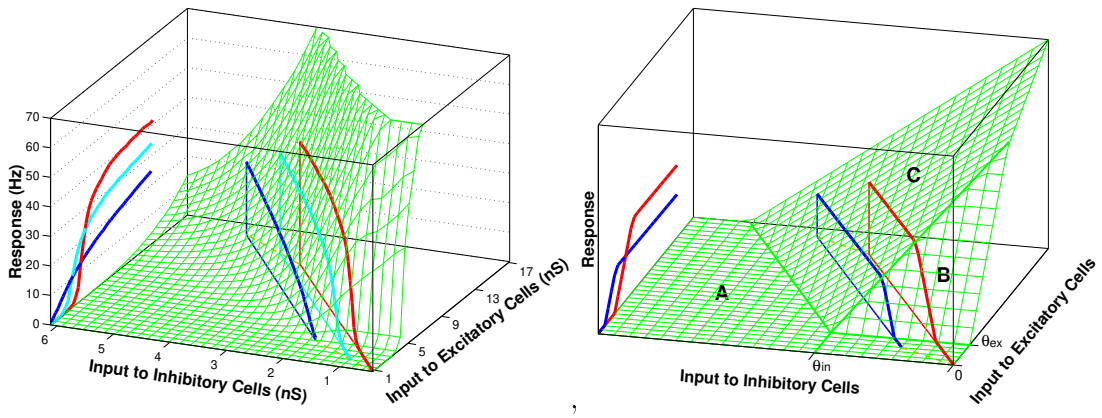

Figure 2: Responses of self-contained module (a) and a linear approximation of module (b). Axes represent total excitatory input to the two module populations, in units of average synaptic conductance. Total long-distance input converging on the center of the full model is plotted in the input plane for all stimulus conditions (medium - center only; light - orthogonal surround; dark - iso-orientation surround). Surround stimulation provides a vector input that translates the thalamic input line (which represents the set of vectors for all center contrasts). Module response curves are obtained by surface projection (also shown on backplane and in dashed lines of fig 1d).

linear surface. This underlying surface can be obtained from a simplified module, composed of interconnected threshold-linear neurons - a typical example is shown in fig 2b. Assume excitatory and inhibitory neurons have thresholds $\theta_{\mathrm{ex}}, \theta_{\mathrm{in}}$, and gains $K_{\mathrm{ex}}, K_{\mathrm{in}}$; total afferent inputs to the two populations are $I_{ex} = M_t T_{ex} + M_h H_{ex}$, $I_{in} = M_t T_{in} + M_h H_{in}$, where M_t, H_t are thalamic and long-range horizontal inputs, and $T_{ex}, T_{in}, H_{ex}, H_{in}$ are the corresponding synaptic efficacies. The synaptic weights among excitatory (e) and inhibitory (i) cells in the module are $W_{ee}, W_{ei}, W_{ie}, W_{ii}$. Then the mean firing rates in the module satisfy the following piecewise-linear system of equations: $M_{ex} = K_{ex}(I_{ex} + W_{ee}M_{ex} - W_{ie}M_{in} - \theta_{ex})$, $M_{in} = K_{in}(I_{in} + W_{ei}M_{ex} - W_{ii}M_{in} - \theta_{in})$. The response surface in fig 2b is $M_{ex}(I_{ex}, I_{in})$, as obtained from the above system. The surface has three planar regions, corresponding to (A) no excitatory firing, (B) recurrent self-excitation with no inhibition, and (C) balanced (competing) excitatory and inhibitory firing. Response saturation occurs when the contrast input line crosses region (B) and is parallel to the contours in region (C), i.e. $\theta_{in}/\theta_{ex} > T_{in}/T_{ex} = (W_{ii} + 1/K_{in})/W_{ie}$ (shown with red curve). Super-saturation results from increasing the slope of the contrast input line, so that $T_{in}/T_{ex} > (W_{ii} + 1/K_{in})/W_{ie}$. The surround facilitation/suppression effect (compare blue curve to red curve) is obtained when the translation vector resulting from surround stimulation has a bigger slope than the contrast input line, i.e. $H_{in}/H_{ex} > T_{in}/T_{ex}$. This corresponds to the physiological prediction that long-range intracortical inputs are less biased towards excitatory (vs. inhibitory) neurons than are thalamocortical inputs.

Conclusions

Modularity has long been proposed as a means of resolving the complexity of cortical function [14, 2]. Here we have constructed modules (corresponding to dense local cortical

circuitry) which are quasi-autonomous: their response properties, as studied in isolation, are preserved in the larger system. Our modular analysis illustrates an expanded concept of the cortical receptive field: each portion of the RF has a dual excitatory-inhibitory influence whose net effect on a neuron depends on how other RF components activate the recurrent local cortical circuitry. This vector-based RF integration fully encompasses the traditional (scalar) view as a special case. Furthermore, this approach unifies notions of classical and extraclassical RFs by showing how long-range inputs can be considered on equal footing with thalamocortical inputs and how the effects of both can be analyzed together. Based on this analysis we predict that for different types of stimulation (involving, for example, luminance, orientation, or motion contrast), the influence of extraclassical stimulation shifts from facilitatory to suppressive as center RF drive increases. Since the properties of neurons and connections in visual cortex exploited here are common to other cortical areas, vector-based integration appears well-suited to other cortex as well.

References

[1] Bonds, A.B. *Visual Neurosci.* **6**, 239-255 (1991).
[2] Douglas, R.J., Martin, K.A.C., Whitteridge, D. *Neural Comp* **1**, 480-488 (1989).
[3] Gilbert, C.D. & Wiesel, T.N. *J. Neurosci.* **3**, 1116-1133 (1983).
[4] Hartline, H.K. *Am. J. Physiol.* **130**, 700-711 (1940).
[5] Heeger, D.J. *Visual Neurosci.* **70**, 181-197 (1992).
[6] Hubel, D.H. & Wiesel, T.N. *J. Neurophysiol.* **148**, 574-591 (1959).
[7] Kisvarday, Z.F., Martin, K.A.C., Freund, T.F., Magloczky, Z.F., Whitteridge, D., and Somogyi, D. *Exp. Brain Res.* **64**, 541-552 (1986).
[8] Knierim, J.J. & Van Essen, D.C. *J. Neurophysiol.* **67**, 961-980 (1992).
[9] Levitt, J.B. & Lund J.S. *Soc. Neurosci. Abstr.* **20**, 428 (1994).
[10] Li, C.Y. & Creutzfeldt, O.D. *Pflugers Arch.* **401**, 304-314 (1984).
[11] Maffei, L. & Fiorentini, A. *Vision Res.* **13**, 1255-1267 (1973).
[12] McCormick, D.A., Connors, B.W., Lighthall, J.W. & Prince, D.A. *J. Neurophysiol.* **54**, 782-806 (1985).
[13] McGuire, B.A., Gilbert, C.D., Rivlin, P.K. & Wiesel, T.N. *J. Comp. Neurol.* **305**, 370-392 (1991).
[14] Mountcastle, V.B. in *The Mindful Brain* (eds Edelman, G.M. & Mountcastle, V.B.) 7-50 (MIT Press, Cambridge, MA, 1978).
[15] Jones, J.P. & Palmer, L.A. *J Neurophysiol.* **58**, 1187-1211 (1987).
[16] Rockland, K.S. & Lund, J.S. *Science* **215**, 1532-1534 (1982).
[17] Sengpiel, F., Baddeley, R.J, Freeman, T.C.B., Harrad, R., & Blakemore, C. *Soc. Neurosci. Abstr.* **21**, 1649 (1995).
[18] Siapas, A.G., Todorov, E.V, & Somers, D.C. *Soc. Neurosci. Abstr.* **21**, 1651 (1995).
[19] Sillito, A.M., Grieve, K.L., Jones, H.E., Cudeiro, J., & Davis, J. *Nature*, **378**, 492 (1995).
[20] Ben-Yishai, R., Lev Bar-Or, R. & Sompolinsky, H. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3844-3848 (1995).
[21] Somers, D.C., Nelson, S.B. & Sur, M. *J. Neurosci.* **15**, 5448-5465 (1995).
[22] Stemmler, M., Usher, M. & Niebur, E *Science* **269**, 1877-1880, (1995).
[23] Toth, L.J., Rao, S.C., Kim, D.-S., Somers, D., and Sur, M. *Proc. Natl. Acad. Sci. USA*, **93**, 9869–9874 (1996).
[24] Tuckwell, H.G. *Stochastic Processes in the Neurosciences* (Soc. for Indust. & Appl. Math., Philadelphia, PA, 1989).
[25] White, E.L. *Cortical Circuits* 46-82 (Birkhauser, Boston, 1989).