Vector-based Integration of Local and Long-range Information in Visual Cortex

David C. Somers, Emanuel V. Todorov, Athanassios G. Siapas†, Mriganka Sur

Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

† Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

This publication can be retrieved by anonymous ftp to publications.ai.mit.edu.

Abstract

Integration of inputs by cortical neurons provides the basis for the complex information processing performed in the cerebral cortex. Here, we propose a new analytic framework for understanding integration within cortical neuronal receptive fields. Based on the synaptic organization of cortex [1, 2, 3, 4, 5, 6], we argue that neuronal integration is better understood in terms of local cortical circuitry than at the level of single neurons, and we present a method for constructing self-contained modules which capture (nonlinear) local circuit interactions. In this framework, receptive field elements naturally have dual (rather than the traditional unitary [7, 8, 9, 10, 11]) influence since they drive both excitatory and inhibitory cortical neurons. This vector-based analysis, in contrast to scalar approaches, greatly simplifies integration by permitting linear summation of inputs from both “classical” and “extraclassical” receptive field regions. We illustrate this by explaining two complex visual cortical phenomena, which are incompatible with scalar notions of neuronal integration.

Copyright © Massachusetts Institute of Technology, 1995.
Central to an understanding of cortical function is the question of how cortical neurons integrate inputs to produce outputs. In sensory cortex, the unit of integration is a neuron’s receptive field (RF)\[7, 9\]. Experiments in primary visual cortex (V1) demonstrate that the spatial extent of integration includes not only “classical” regions, where visual stimuli elicit responses (presumably through thalamocortical axons), but also “extraclical” regions, where stimuli largely modulate responses evoked by other stimuli (presumably via long-range intracortical or inter-areal axons)\[5, 12, 13, 14, 15\]. The traditional view of integration holds that each portion of a neuron’s receptive field in response to a given stimulus element has \textit{either} an excitatory or an inhibitory (i.e., a scalar) influence\[7, 8, 9, 10, 11\] (see fig 1a). 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 \[16, 17, 18, 19, 20\]. 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 (“supersaturate”)\[16, 22, 23, 24\] (see fig 2a). Second, adding a distal stimulus facilitates responses to a weak central stimulus, but suppresses responses to a strong stimulus\[17, 18, 19, 20\] (see fig 2b).

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 (see fig 1b): i) receptive field regions (via either thalamocortical or long-range intracortical axons) drive both excitatory and inhibitory cortical neurons \[1, 2, 6\]; ii) these neurons form dense, recurrent local connections \[1, 3, 4, 5\]; and iii) different portions of the receptive field provide converging inputs to a shared population of cortical neurons\[5, 13, 14, 15, 21\]. Based on this anatomy, we propose that: i) each RF region in response to a given stimulus 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 will demonstrate this approach by capturing the paradoxical local and long-range phenomena within a large-scale visual cortical model, and later we will 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 \[25, 26, 27\] or have oversimplified local cortical excitatory and inhibitory interactions in order to obtain closed-form (scalar) analysis\[28, 29, 30\]. Recently, we and others \[25, 26, 31\] have demonstrated that consideration of short-range recurrent cortical excitatory (and inhibitory) connections yields a concise account of a broad range of experimental data on orientation and direction selectivity. Here, we extend our model to incorporate long-range intracortical excitation (see fig 3 and [25] for model description).

Physiological responses to oriented grating stimuli of differing contrasts within the classical RF are captured by the model (see fig 4a). The responses shown here and below are for the excitatory subpopulation. Responses saturate at contrast levels below which thalamic responses saturate \[22\], can decline for high contrasts (supersaturation) \[16, 22, 23, 24\], and have firing rates well below maximal cellular firing rates \[32\]. Inhibitory neurons, on average, saturate at higher contrasts than do excitatory neurons (not shown). While preserving classical RF properties, our model also captures paradoxical
(a) The traditional view of fixed receptive field regions [7, 8, 9, 10, 11] implies an oversimplified view of cortical circuitry. For a given stimulus element, each portion of the RF is assumed to have either an excitatory or inhibitory influence. This view considers only feedforward cortical circuitry: afferent projections corresponding to excitatory RF regions (+) provide direct excitatory input to the neuron, while afferent projections from inhibitory subfields (−) excite local interneurons which inhibit the neuron. This receptive field concept provides a simple explanation of isolated stimulus effects but it neglects (local feedback) interactions among excitatory and inhibitory neurons and is ill-suited to address complex RF integration.

(b) Consideration of cortical circuitry suggests a receptive field concept that is more difficult to analyze but better suited to address RF integration. Afferent projections (including thalamocortical and long-range intracortical connections) from a RF region contact both excitatory and inhibitory neurons [1, 2, 6] and thus have dual influences (+/−). Because local cortical neurons densely interconnect [1, 3, 4, 5], RF regions not only directly drive a cortical neuron but also have substantial indirect influence via a shared pool of local excitatory and inhibitory neurons. Thus, the net effect of a stimulus element in a given portion of the RF depends both on the excitatory-inhibitory bias of its afferent projections and on how other RF segments activate the local cortical circuitry.

Figure 1: Receptive Field Concepts and Cortical Circuitry.
(a) Experimental evidence for “supersaturation.” Data obtained by single unit recording of neurons in cat V1 (see fig. 2c of [16]) are replotted to show declining responses at high contrast levels. Mean values (with std. err.) of contrast-response curves of 29 cortical cells (12 complex, 17 simple). Response curves for each cell were normalized to its strongest response, which was defined as 100%. The abscissa shows the luminance increment of a moving bar relative to background (3cd/m²).

(b) Experimental evidence for surround facilitation/suppression. Data from single unit recording of cat V1 cells (kindly provided by F. Sengpiel [20]) shows that the presence of a high contrast surround stimulus can facilitate responses to a low contrast center stimulus, yet suppress responses to a high contrast center stimulus. Mean responses (N=10 cells; with std. err.) obtained for varying contrast levels of an (optimally-sized and optimally oriented) center grating, with no surround (solid line) and with a high contrast, iso-orientation surround (dashed line). Response curves for each cell were normalized for its strongest response to “center only” stimulation. Points at 0% and 100% contrast (open symbols; center only = □) replot data from monkey V1 (fig. 10 of [17]) showing same effect using both iso-orientation (○) and cross-orientation (△) surround stimuli (bar stimuli were used in the latter experiments).

Figure 2: Experimental Evidence.
Cortical circuitry under a 2.5mm by 5mm patch of primary visual cortex is represented by a model with 20,250 spiking cortical neurons and over 1.3 million cortical synapses. Neurons are 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 [45]). 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 [32] (see methods of [25] for equations and parameters). Color map represents orientation preference (shown on the scale at bottom), while surface amplitude represents the net (Σ excitatory - Σ inhibitory) strength of intracortical connections from each mini-column to the cells of the central (yellow) orientation mini-column. In the figure, long-range connections are scaled upward by a factor of 10 relative to short-range connections in order to aid visibility. Intracortical connections provide short-range excitation (connection probabilities fall linearly from \( \rho_{\text{excit-excit}} = 0.1 \), \( \rho_{\text{excit-inhib}} = 0.1 \) at distance zero to \( \rho = 0 \) at \( d = 150\mu m \)), short-range inhibition (linear from \( \rho_{\text{inhib-excit}} = 0.12 \), \( \rho_{\text{inhib-inhib}} = 0.06 \) at \( d = 0 \) to \( \rho = 0.5\rho_{\text{peak}} \) at \( d = 500\mu m \); \( \rho = 0 \) elsewhere), and long-range excitation (linear with orientation difference, from \( \rho = 0.005 \) at \( \phi = 0^\circ \) to \( \rho = 0.001 \) at \( \phi = 90^\circ \)). Peak synaptic conductances, by source, onto excitatory cells are \( g_{\text{excit}} = 7nS \), \( g_{\text{inhib}} = 15nS \), \( g_{\text{long}} = 3nS \), and \( g_{\text{long}} = 1.2nS \) and onto inhibitory cells are \( g_{\text{excit}} = 1.5nS \), \( g_{\text{inhib}} = 1.5nS \), \( g_{\text{long}} = 1.5nS \), and \( g_{\text{long}} = 1.2nS \). 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 0.6mm² and responds linearly with log stimulus contrast. Results are averaged over 20 networks constructed with these probability distributions.

Figure 3: Visual Cortical Model
(a) Mean cortical contrast responses of central (yellow) orientation domain (36 excitatory neurons × 20 networks). Circular grating stimuli (diameter 0.75°) were presented for 300 msecs. Responses saturate as contrast increases (solid line). This property of cortical neurons has long been known [16, 22, 23, 24] (see fig. 2a) and likely results from cortical interactions since cortical saturation generally occurs at lower contrasts than for LGN cells and because saturation response levels are much lower than the responses of cortical neurons in vitro to strong injected currents [32]. When intracortical connections are silenced in simulations (leaving only thalamocortical inputs active) responses of the same neurons increase linearly without saturation (dotted line). Note switch from cortical amplification to attenuation as contrast increases. For a subpopulation of model neurons (N = 18 × 20), responses decline at high contrasts (dashed line). Such “supersaturation” has been described for V1 neurons [22, 23, 16, 24]; the decline in response beyond a maximal level implies that cortical contrast effects are more complex than simply saturation or normalization [29]. These saturation properties were also obtained for non-preferred stimulus orientations, both experimentally [42] and in the model (not shown).

(b) Simulation of “surround” stimulus effects. Annular “surround” grating stimuli (100% contrast; inner diameter 0.75°) facilitate responses to low contrast “center” stimuli, but suppress responses to high contrast “center” stimuli. Recent experiments [20] (see fig. 2b) have demonstrated that contrast-response functions for “surround” and “no surround” stimuli cross at a moderate contrast level with “surround” responses relatively stronger for low contrast centers and “no surround” responses dominating for high contrast center stimuli. Other researchers [17, 18, 19] have also observed that an oriented surround stimulus can facilitate responses to an empty center, yet suppress responses to a high-contrast, optimal orientation center stimulus. Both effects are strongest for iso-orientation surround stimuli [17, 18]. In addition, Sillito et al. [18] have reported that (high contrast) non-optimal orientation center stimuli, which fail to elicit responses alone, can produce vigorous responses when a surround stimulus of the preferred orientation (for the center field) is also presented. All such effects were observed in our simulations (non-preferred center results not shown), and these simulation results are closely approximated by the responses of a self-contained module (dashed lines) to the same long-range input (see also fig 6).

Figure 4: Simulations

5
extraclassical RF modulations [17, 18, 19, 20]. The modulatory influence of “surround” gratings on responses to optimal orientation “center” stimuli shifts from facilitatory to suppressive as center stimulus contrast increases (see fig 4b). 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 [5, 13] (which are directly related to classical and extraclassical stimulation [14], respectively), 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 fig 5 for a formal explanation). 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., [31]) 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 fig 5). 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 [7, 8, 9, 10, 11, 29, 30, 33]. Thalamocortical and long-range intracortical inputs activate excitatory and inhibitory neurons in different proportions and thus 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 6. 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. 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 [34] can be treated similarly). Contrast response functions (CRFs) predicted by the module are obtained by projecting the resulting input line onto the surface, and they closely approximate the CRFs generated by the model for all tested stimulus conditions (see fig 4b) as well as experimental CRFs (see fig 2a,b). Thus, the paradoxical classical and extraclassical RF integration phenomena are captured by local circuit interactions.

Local interactions are described by module response surface shapes. The surface shape shown in figure 6 is characteristic of a large class of recurrently connected excitatory-inhibitory circuits and can be thought of as providing generalized gain control: sigmoid-shaped response curves are generated for a wide-range of stimulus conditions. Analytical conditions for extraclassical facilitation/suppression effects are derived in figure 7 and correspond to the physiological prediction that long-range intracortical
Given a local neuronal population $P$ whose mean firing rate $M = F(I_d, I_l)$ is a function $F$ of the long-distance (intracortical and thalamocortical) inputs $I_d$ and local (intracortical) inputs $I_l$, we want to construct a closed system (module) whose response approximates $M$ as a function of $I_d$ only. All bold face quantities denote vectors with components corresponding to excitatory (triangles) and inhibitory (circles) cells; local inputs are defined as arriving from within a radius $R$, which is chosen to minimize approximation error. Module construction is only possible if $I_l$ can be expressed as a function of $M$ and $I_d$. To that end we use a local homogeneity assumption $M = I_l$, i.e. neurons within $R$ (not just $P$) have mean firing rates $M$. Thus the module output $M^*$ is the solution of $M = F(I_d, M)$. This equation can be solved numerically if we model the response functions of integrate-and-fire neurons [43, 44]. Here, we compute $M^*$ 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 $R$. 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 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 $R$.

Figure 5: Constructing self-contained modules from large-scale models.
Excitatory responses of a module (avg. 10 trials; 500 msecs) constructed for the center of the cortical model are shown by the green surface. 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 (red - center only; cyan - orthogonal surround; blue - iso-orientation surround). Module response curves are obtained by surface projection (see fig 4b for comparison with model results). Surround stimulation provides a vector input that translates (see dashed brown line in input plane) the thalamic input line (which represents the set of vectors for all center contrasts). Orthogonal surround stimulation activates fewer neurons that send long-distance projections to the central population and thus results in a smaller translation than iso-orientation surround stimulation. Note that other center and/or surround stimuli would produce different input vectors and thus would generate differing degrees of facilitation and suppression for the same module (see experiments of [17, 18, 20]).

Figure 6: Module response surface
Analysis of the requirements for saturation and facilitation/suppression effects. Note that integrate-and-fire neurons have approximately threshold-linear feedforward responses (fig 4a), and thus the module output (fig 6) is a smoothed version of an underlying piecewise-linear surface. This underlying surface can be obtained from a simplified module, composed of interconnected threshold-linear neurons - a typical example is shown in here. Assume excitatory and inhibitory neurons have thresholds \( \theta_{ex}, \theta_{in} \), and gains \( K_{ex}, K_{in} \); 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:

\[
M_{ex} = K_{ex} (I_{ex} + W_{ee} M_{ex} - W_{ie} M_{in} - \theta_{ex}), \quad M_{in} = K_{in} (I_{in} + W_{ei} M_{ex} - W_{ii} M_{in} - \theta_{in}).
\]

The response surface in the figure 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). Supersaturation 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. \( H_{in}/H_{ex} > T_{in}/T_{ex} \). Although this simplified module neglects driving force nonlinearities and response smoothing around threshold (see fig 4a, feedforward), its parameters can be extracted from the parameters of the large-scale model to provide a quantitative characterization of geometrical properties of the real response surface in fig 6. Note that analysis region B requires only relatively low inhibitory firing, not a lack of firing; this is made apparent if a piecewise-linear, rather than simply a threshold-linear model, is used.

Figure 7: Analysis
inputs are less biased towards excitatory (vs. inhibitory) neurons than are thalamocortical inputs. The response surface can be thought of as the collection of all possible contrast-response curves, and the contextual influence of surround stimulation then can be seen as “selecting” a new response curve. By comparison, scalar addition of thalamocortical and long-range inputs predicts surround effects along the same response curve, in which case contrast saturation (no decline) and long-range suppression are incompatible. Extracellular observations \([16, 24, 20, 35]\) of stimulus dependent shifts in contrast response functions (CRFs) which primarily shift the CRF downward (lower peak response) or to the right along the contrast axis (same peak response) have previously been interpreted in terms of divisive inhibition \([29, 30]\). Intracellular evidence, however, indicates that divisive or shunting inhibition plays little role in visual cortex \([36, 37]\). Our module surface (fig 6) illustrates how both effects can occur without divisive inhibition: i) additional drive to cortical inhibitory neurons will shift the contrast input line and yield a largely downward shift of the CRF; ii) inhibition of thalamic responses (or thalamocortical transmission) changes only the magnitude but not the angle of the input vector and thus the CRF will shift rightward.

Modularity has long been proposed as a means of resolving the complexity of cortical function \([38, 39, 40]\). 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; recent evidence from monkey V1 showing modulation of neuronal responses as a function of the orientation difference between center & surround stimuli \([18]\) are strongly consistent with our prediction. 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.

**Acknowledgments**

We wish to thank Sacha Nelson for thoughtful discussions about cortical gain control and long-range connections. We are grateful to Frank Sengpiel for kindly allowing us to display his experimental data. We also thank Peter Dayan, Ted Adelson, Michael Jordan, Tai Sing Lee, Louis Toth, Chris Moore, Bhavin Sheth, Gerry Sussman, and Josh Tenenbaum for their careful reading of this manuscript. A preliminary report of a portion of this work has appeared \([41]\). This work was funded in part by an NIMH postdoctoral fellowship to D.S., a Howard Hughes Predoctoral Fellowship to E.T., and an NEI grant to M.S.
References


[38] Szentagothai, J. Brain Res. 95, 475-496 (1975).


