

# Object Boundary Detection in Natural Images May Depend on “Incitatory” Cell–Cell Interactions

Gabriel C. Mel,<sup>1</sup> Chaithanya A. Ramachandra,<sup>2</sup> and Bartlett W. Mel<sup>3</sup>

<sup>1</sup>Department of Computer Science, University of Southern California, Los Angeles, California 90089, <sup>2</sup>Eyenuk, Inc, Los Angeles, California 90089, and <sup>3</sup>Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089

Detecting object boundaries is crucial for recognition, but how the process unfolds in visual cortex remains unknown. To study the problem faced by a hypothetical boundary cell, and to predict how cortical circuitry could produce a boundary cell from a population of conventional “simple cells,” we labeled 30,000 natural image patches and used Bayes’ rule to help determine how a simple cell should influence a nearby boundary cell depending on its relative offset in receptive field position and orientation. We identified the following three basic types of cell–cell interactions: rising and falling interactions with a range of slopes and saturation rates, and nonmonotonic (bump-shaped) interactions with varying modes and amplitudes. Using simple models, we show that a ubiquitous cortical circuit motif consisting of direct excitation and indirect inhibition—a compound effect we call “incitation”—can produce the entire spectrum of simple cell–boundary cell interactions found in our dataset. Moreover, we show that the synaptic weights that parameterize an incitation circuit can be learned by a single-layer “delta” rule. We conclude that incitatory interconnections are a generally useful computing mechanism that the cortex may exploit to help solve difficult natural classification problems.

**Key words:** cortical circuits; natural image statistics; V1; vision; excitatory-inhibitory interactions

## Significance Statement

Simple cells in primary visual cortex (V1) respond to oriented edges and have long been supposed to detect object boundaries, yet the prevailing model of a simple cell—a divisively normalized linear filter—is a surprisingly poor natural boundary detector. To understand why, we analyzed image statistics on and off object boundaries, allowing us to characterize the neural-style computations needed to perform well at this difficult natural classification task. We show that a simple circuit motif known to exist in V1 is capable of extracting high-quality boundary probability signals from local populations of simple cells. Our findings suggest a new, more general way of conceptualizing cell–cell interconnections in the cortex.

## Introduction

The primary visual cortex (area V1) is a complex, poorly understood, multipurpose image processor optimized to extract information from natural scenes, which are themselves complex, poorly understood signals. Thus, understanding how V1 operates presents a challenging reverse engineering problem. A long-standing hypothesis is that orientation-tuned V1 cells somehow participate in object boundary detection, a core process in biological vision (Hubel and Wiesel, 1962; Biederman, 1987; von der Heydt and Peterhans, 1989; Gilbert and Wiesel, 1990; Kapadia et al., 1995) that is crucial for the functions of both

ventral and dorsal streams (Biederman, 1987; Hoffman, 2000; Rust and Dicarlo, 2010; Theys et al., 2015). However, little progress has been made in refining or testing this hypothesis, in part because of our lack of understanding of the structure of natural object boundaries, and, particularly, what a V1 cell needs to do to reliably distinguish boundaries from nonboundaries.

This uncertainty has made it difficult to form specific computational hypotheses as to how V1 circuits perform this behaviorally relevant classification task. Previous work has analyzed natural image statistics to determine how local boundary segments are arranged in images (Sigman et al., 2001; Sanguinetti et al., 2010), and how these arrangements relate to human contour grouping performance (Geisler et al., 2001). However, no study has yet attempted to deconstruct the natural boundary detection problem in detail, or to link the computations necessary for boundary detection to particular neural mechanisms.

With the goal to better understand the computations underlying object boundary detection in V1 (Fig. 1), we began with a question that could be used to sort natural image patches into boundary and non-boundary cases (Fig. 1A), and the

Received Oct. 5, 2018; revised Aug. 16, 2022; accepted Oct. 6, 2022.

Author contributions: G.C.M., C.A.R., and B.W.M. designed research; G.C.M., C.A.R., and B.W.M. performed research; G.C.M. and B.W.M. analyzed data; G.C.M. and B.W.M. wrote the paper.

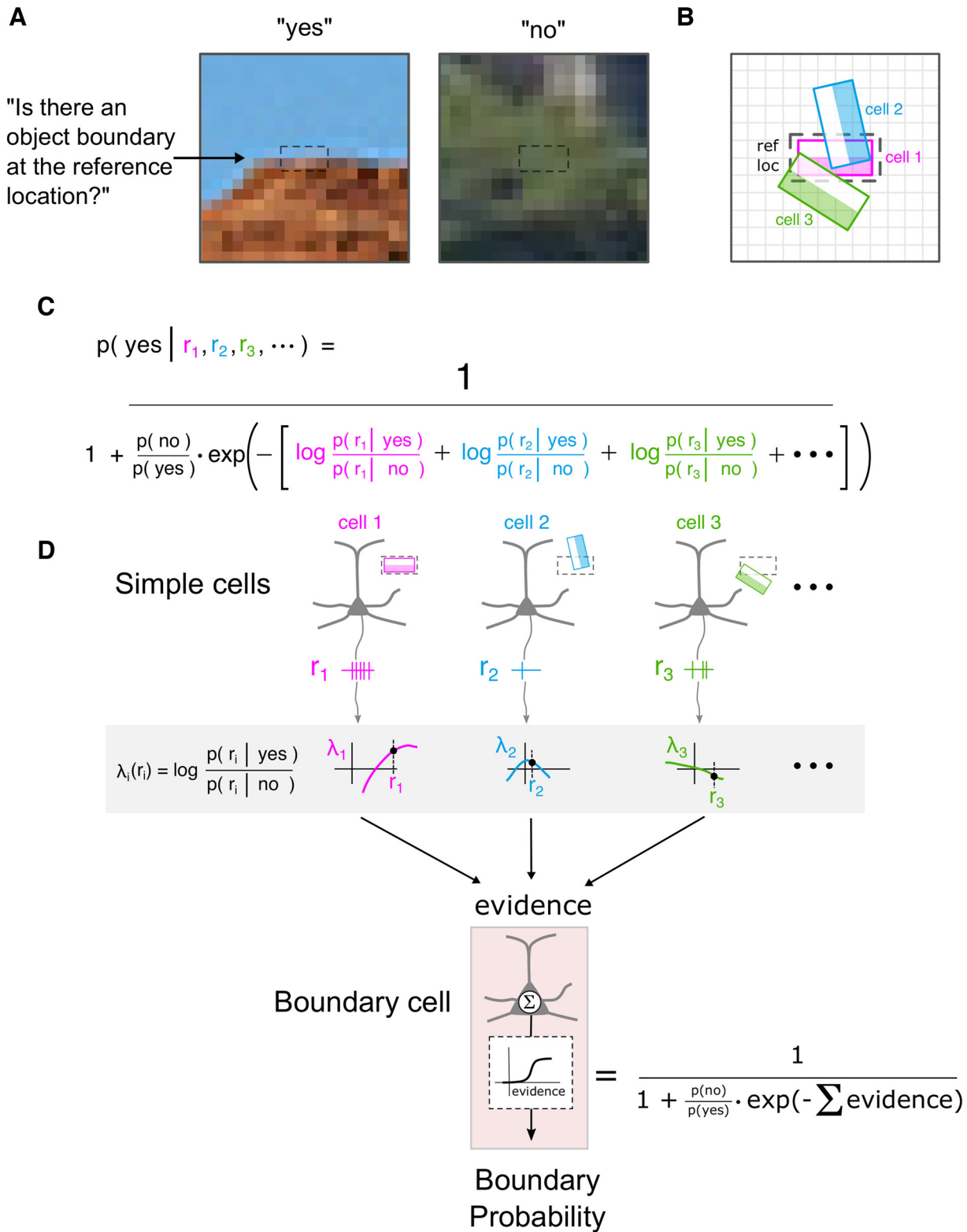
Funding for this work was provided by National Institutes of Health/National Eye Institute Grant EY-016093.

The authors declare no competing financial interests.

Correspondence should be addressed to Gabriel C. Mel at meldefon@gmail.com.

<https://doi.org/10.1523/JNEUROSCI.2581-18.2022>

Copyright © 2022 the authors



**Figure 1.** Calculating boundary probability from natural images using Bayes' rule. **A**, The boundary detection problem can be encapsulated by the question and answers shown; ~30,000 natural image patches were classified in this way. Dashed box indicates a reference location where a boundary might appear. Patches shown during labeling were  $20 \times 20$  pixels. **B**, Three (of 300) oriented cells with responses  $r_1$ ,  $r_2$ , and  $r_3$  are shown in the vicinity of the reference location. **C**, Under the assumption that simple cell responses are class-conditionally independent (see main text), Bayes' rule gives an expression for boundary probability in terms of individual cell LLRs (colored terms in denominator). **D**, Cell responses  $r_i$  are passed through their respective LLR functions  $\lambda_i()$ , and the results are summed and passed through a fixed sigmoidal "f-I curve" (frequency-current) to yield boundary probability.

assumption that a known cell type—orientation-tuned “simple cells” (SCs; as defined by Hubel and Wiesel, 1962), typically modeled as divisively normalized oriented linear filters (Carandini and Heeger, 2011)—covered the site of the putative boundary at a range of positions and orientations; a sample of 3 SC receptive fields (RFs) is shown in Fig. 1*B*. We then asked how the outputs of a population of SCs should be combined to produce a “boundary cell” (BC), whose firing rate represents the probability that an object boundary is present within its RF (Fig. 1*C,D*). When framed in this way, Bayes’ rule tells us what data to extract from natural images to obtain an answer to the question. In a previous study (Ramachandra and Mel, 2013), we noted that under the simplifying assumption of “class conditional independence” (CCI; for a detailed discussion, see Materials and Methods), simple cell–boundary cell interactions are captured by the log-likelihood ratio (LLR) functions embedded in Bayes’ rule (Fig. 1*C*, colored expressions), which represent the evidence that a given simple cell provides about the presence of an object boundary within the receptive field of a neighboring boundary cell (Fig. 1*D*). We found that SC–BC interactions were diverse, and in some cases involved compound excitatory (E) and inhibitory (I) effects. However, since only a small number of cells was analyzed in that study, we could not come to general conclusions about the types of cell–cell interactions needed to compute boundary probability, making it difficult to compare and contrast possible neural mechanisms.

In this study, we analyze a much larger dataset and compute the full set of simple cell–boundary cell interaction functions for a population of 300 odd-symmetric simple cells surrounding a “reference location” (RL) where a boundary might be detected. We find that the simple cell–boundary cell interactions suggested by the natural image LLR functions follow a predictable pattern that depends on the offset in position and orientation between simple cell and boundary cell receptive fields, and we show that a well known cortical circuit motif can implement the entire spectrum of SC–BC interactions found in our dataset. Finally, we demonstrate that a cortically inspired neural network can produce a boundary-detecting cell from simple cells with a single layer of excitatory synapses and a single inhibitory interneuron. Our findings suggest that cortical sensory computations, including the detection of natural object boundaries, may depend on a specific class of structured excitatory–inhibitory cell–cell interactions.

## Materials and Methods

**Image preprocessing.** As in the study by Ramachandra and Mel (2013), we used a modified version of the COREL database for boundary labeling in natural images. Several image categories, including sunsets and paintings were removed from the full COREL database since their boundary statistics differed markedly from that of typical natural images. Custom code was used to select ~30,000  $20 \times 20$  pixel image patches for labeling. The reference location representing the receptive field location of a hypothetical boundary cell was defined as the elongated, horizontal  $2 \times 4$  pixel region at the center of the patch (Fig. 1*A,B*, dashed box).

**Natural image data collection.** To collect ground-truth data relating to natural contour statistics, for each image patch to be labeled, a horizontal  $2 \times 4$  pixel rectangular box was drawn around a centered reference location and human labelers were asked to answer the question, “On a scale from 1–5, with 1 meaning ‘extremely unlikely’ and 5 meaning ‘extremely likely’—how likely is it that there is an object boundary passing horizontally through the reference box, end to end, without leaving the box?” To qualify as valid, boundary segments also had to be visible

and unoccluded within the box. We restricted labeling to horizontal boundaries (i.e., horizontal reference boxes) since pixel lattice discretization made it more difficult to judge oblique orientations, and because we expected cell response statistics in natural images to be approximately orientation invariant. (This expectation was supported by subsequent tests showing that LLR functions obtained for horizontal boundaries also led to high boundary detection performance on oblique boundaries.) Labeler responses were recorded, and patches with scores of 1 or 2 were classified as “no” patches, while patches with scores of 4 or 5 were classified as “yes” patches. Agreement between labelers was very high, based on informal observations when two labelers worked together. Rare ambiguous patches that could cause labeler disagreement were often given scores of 3, so these patches were excluded from our analyses. After labeling, the dataset was doubled by adding left–right flipped versions of each patch, and assigning the same label as the unflipped counterpart.

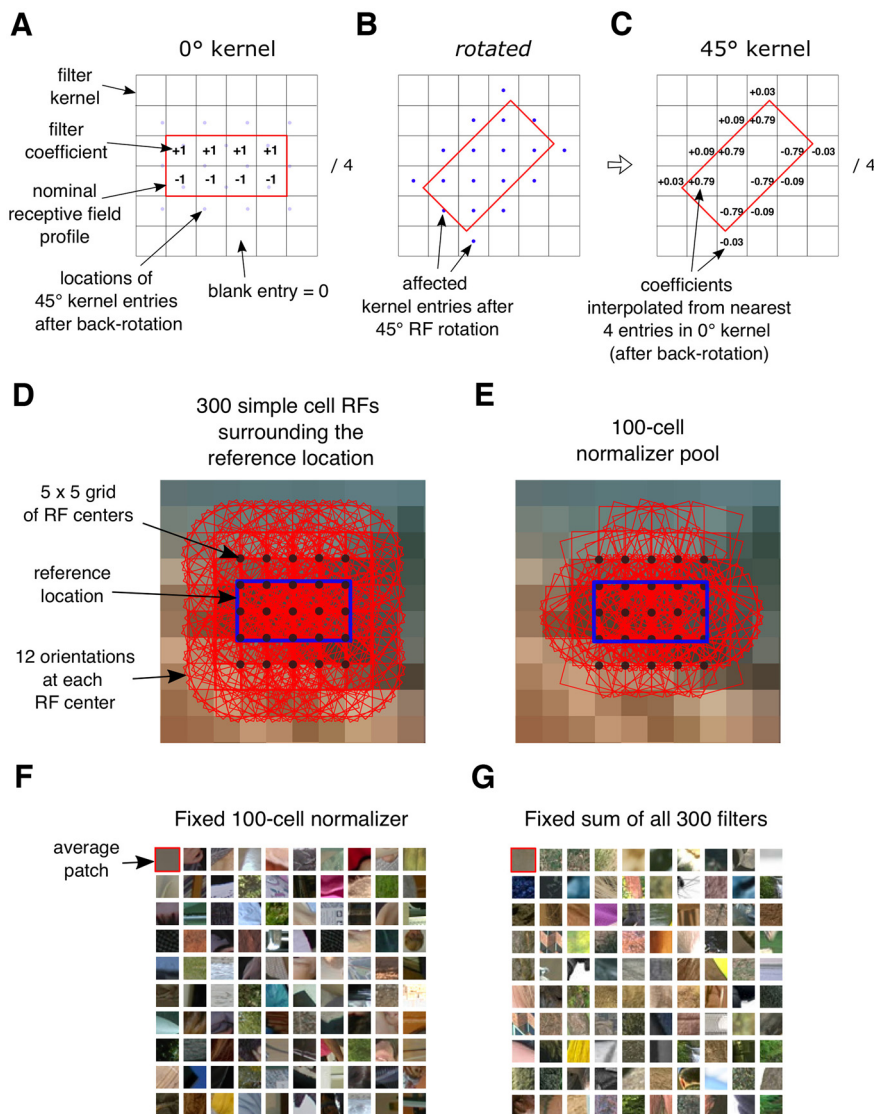
**Collecting virtual simple cell responses on and off natural boundaries.** Given a large set of image patches, some labeled as boundaries and others not, the next step was to collect virtual simple cell responses densely covering boundary versus nonboundary image patches so that their different statistics could be analyzed. Original color image patches were converted to single-channel (monochrome) intensity images ( $0.29R + 0.59G + 0.11B$ ). Simple cell-like-oriented linear “filters” were created by rotating a  $2 \times 4$  pixel horizontal filter kernel in  $15^\circ$  increments from  $0^\circ$  to  $165^\circ$  (i.e., 12 orientations). The filter kernel for the horizontal ( $0^\circ$ ) cell is shown in Figure 2*A*. Positive filter coefficients represent ON subregions, and negative coefficients represent OFF subregions of the receptive field of the virtual simple cell. Blank kernel entries represent zeros. Filter coefficients in rotated kernels were computed by rotating the horizontal kernel using “bilinear interpolation” ([https://en.wikipedia.org/wiki/Bilinear\\_interpolation](https://en.wikipedia.org/wiki/Bilinear_interpolation); Fig. 2*B,C*). In essence, any entry in an oriented kernel that was affected by the rotated horizontal pattern would be assigned a coefficient that was interpolated based on the nearest four coefficients in the original horizontal pattern after back-rotation. (The locations of the faint blue dots in Fig. 2*A* show the locations of the dark blue dots in Fig. 2*B* rotated back by  $45^\circ$  into the horizontal frame of references; the four coefficients closest to each faint blue dot were used for the interpolation.) A similar procedure was used to generate all other rotated filter kernels. Note that we modeled only “odd-symmetric” simple cells with a single ON and a single OFF subfield, as we found that even-symmetric cells (e.g., with a central ON region and two flanking OFF regions), which are also found in V1, are of limited use in classifying the type of “step” edges contained in our human-labeled natural image dataset.

The “response” value of a simple cell at a particular image location was computed as the dot product between the filter kernel of the cell and the underlying image intensity pixel values. A simple cell generated a positive response when its positive kernel coefficients mostly overlapped with bright image pixels and its negative coefficients mostly overlapped with dark image pixels. The largest positive responses occurred on light–dark boundaries of the preferred orientation of a cell. A negative response was treated as a positive response of a distinct simple cell with opposite contrast polarity (i.e., rotated by  $180^\circ$ ).

Given a labeled image patch, simple cell responses covering that patch were collected on a  $5 \times 5$  grid centered on the patch, at each of 12 orientations (Fig. 2*D*). This led to  $5 \times 5 \times 12 = 300$  simple cells responding to any given image patch. Simple cell response data were accumulated separately for yes and no labeled patches.

**Restriction to “normalized” image patches.** The prevailing model of a simple cell consists of a linear filter whose output is divisively normalized by activity in the surround of the cell (Carandini and Heeger, 2011). Normalizing the response of a simple cell involves (1) computing the prenormalized response of the cell (e.g., as we do above using an oriented linear filter); (2) calculating the sum  $N$  of certain other cells’ prenormalized responses in the surround of the cell and (3) inhibiting the simple cell’s response as an increasing function of  $N$ . Normalization is often called “divisive” as the surround activity term  $N$  generally appears





**Figure 2.** Simple cell “filter” construction, layout, and normalization pools. **A**, Diagram shows layout of positive and negative filter coefficients for computing simple cells responses. Positive coefficients represent the ON subregion of the RF, and negative coefficients represent the OFF subregion. **B**, Faint blue dots indicate off-grid positions (in  $x$ - $y$  space) of the 18 pixel locations holding coefficients for the 45° case, back-rotated to the horizontal orientation. **B**, Blue dots indicate all pixels that overlap with the red box, representing the original horizontal receptive field rotated by 45°. **C**, Each of these 18 pixels will contain a coefficient (C) computed from bilinear interpolation of the coefficients in the original horizontal layout (A). The coefficients “down the middle” of the 45° filter are zero, reflecting an equal contribution from the ON and OFF subregions of the RF. **D**, Layout of 300 filter locations (red) in  $5 \times 5$  grid surrounding the reference location (blue). **E**, Subset of 100 of the 300 simple cell RFs chosen as the normalizer pool in the following figures. See text for details. **F**, Examples of image patches drawn at random from the image database whose normalizer values (i.e., the sum of the response magnitudes of the 100 cells) fell into the same bin ( $200 \pm 40$ ); see text for details. Top left patch is the average of all. **G**, Same as **F**, but with patch selection based on a different normalizer: the sum of the response magnitudes of all 300 simple cells covering the region (D, RFs).

code (Liang et al., 2000; Wainwright and Simoncelli, 2000; Schwartz and Simoncelli, 2001; Zetsche and Rohrbein, 2001; Fine et al., 2003; Karklin and Lewicki, 2003, 2005; Zhou and Mel, 2008).

The importance of normalization with respect to our analysis is that neural activity in the nonclassical surround of a simple cell can powerfully influence its response, and, given that our study involves collecting and statistically analyzing simple cell responses in natural images, we made efforts to take account of the effects of normalization circuitry on simple cell responses. We selected a pool of 100 oriented cells from the 300 shown in Figure 2D using an *ad hoc* procedure designed to find a subset of the cells that would be minimally correlated in normalized patches. (Limiting the normalizing pool to a subset of the 300 cells turned out to be unnecessary, as we found that including all 300 cells in the normalizer pool led to functionally equivalent results, but we describe the selection procedure below since the 100-cell normalizer was what was actually used to generate our result figures. The reader not interested in the details of the normalization approach can skip to the next paragraph.) We started by defining a “basic normalizer” consisting of the single linear filter value at the reference location. We then culled out a large set of natural image patches whose basic normalizer values fell into a narrow range, that is, all the culled patches had approximately the same filter value at the reference location. The value used for the basic normalizer was  $10 \pm 2$ , but the particular value mattered little; what mattered was that the value was fixed across all image patches collected. We then incrementally “grew” a normalization pool as follows. The cell at the reference location was considered to be the first cell in the normalization pool ( $C_1$ ). A second cell, call it  $C_2$ , was added to the pool by choosing that cell (from the 299 remaining) for which the correlation of its absolute value to that of the other cells in the pool (which was only  $C_1$  in this case) was closest to zero. The absolute value was used because negative values were considered to be responses of distinct cells with opposite polarity. A third cell,  $C_3$ , was then chosen (from the 298 remaining) for which the correlation of its absolute value to that of  $C_1$  and  $C_2$

was, on average, closest to zero. This “greedy” (i.e. “choose the best each time”) procedure continued until 100 cells were chosen, leading to the particular subset of cell RFs shown in Figure 2E. Having chosen the 100-cell normalization pool, we could now cull out a second generation-normalized set of image patches from the image database by selecting patches for which the sum of the responses of all 100 cells in the normalization pool (converted to absolute values) fell into a fixed response range  $N = 200 \pm 40$ . The data in our results figures was generated from this set of normalized image patches. However, as mentioned above, random image patches normalized using all 300 filters surrounding the reference location (and  $N \sim 600 \pm 40$ ; the exact bin center was chosen to match the mean on our normalized labeled

in the denominator of the overall cell response expression. Response normalization plays a variety of roles in the brain (for review, see Carandini and Heeger, 2011), though in the visual system normalization is most often discussed as a means of counteracting the effect of multiplicative “nuisance” factors that modulate the responses of entire local population of neurons in a correlated fashion. For example, if the level of illumination varies across the visual field, a common occurrence in natural scenes, the spatial variation in light level tends to drive up the response rates of all neurons in brighter areas and tamp down the response rates of all neurons in darker areas. Normalization circuitry helps to cancel out these correlated, population-level neural response variations, leading to a greater degree of illumination invariance across the visual field, and a greater degree of statistical independence in the neural

patch database) were indistinguishable as a group (Fig. 2F,G) had indistinguishable averages (upper left patch in each image grid), and led to LLR functions for boundary detection that were very similar.

The restriction of our analysis to image patches within a narrow range of normalizer values was functionally similar to carrying out our analysis “under controlled lighting conditions.” The benefit of this restriction was that it allowed us to probe the relationships between simple cell responses and boundary probability without having to assume that we know the precise form of the function that normalizes neural responses against changes in lighting or other regional correlating factors. The cost of this restriction is that we obtain information only about the boundary probability computation at a specific normalization level. This would be a serious problem if the boundary probability computation changed significantly at different levels of normalization. To mitigate this risk, we verified that our analysis produced similar main results for two other value ranges of the 100-cell normalizer ( $N = 300 \pm 40$  and  $N = 400 \pm 40$ ). We accomplished this by assuming that the normalizing function divisively scales filter responses, which, given that the filters are linear, is equivalent to divisively scaling the image patches themselves. When we scaled down the image patches from the other two normalizer bins by factors of  $300/200 = 1.5$  and  $400/200 = 2$ , respectively, and thereafter processed all image patches equivalently, we obtained results that were functionally indistinguishable from those produced from the original “200” dataset (data not shown). In our main results figures, we include only the data derived from image patches in the original “200” normalizer bin.

**Bayesian formalism.** We assume that a boundary cell computes  $p(\text{yes}|r_1 r_2 r_3 \dots)$ , that is, the probability that a boundary is present at the reference location given a set of simple cell responses  $r_1 r_2 r_3$ , etc., whose receptive fields surround and cover the reference location (Fig. 1D). Using Bayes’ rule we obtain the following:

$$p(\text{yes}|r_1 r_2 r_3 \dots) = \frac{p(r_1 r_2 r_3 \dots | \text{yes}) p(\text{yes})}{p(r_1 r_2 r_3 \dots | \text{yes}) p(\text{yes}) + p(r_1 r_2 r_3 \dots | \text{no}) p(\text{no})}. \quad (1)$$

Dividing through by the numerator and rearranging, we find:

$$p(\text{yes}|r_1 r_2 r_3 \dots) = \frac{1}{1 + \frac{p(\text{no})}{p(\text{yes})} \exp\left(-\log \frac{p(r_1 r_2 r_3 \dots | \text{yes})}{p(r_1 r_2 r_3 \dots | \text{no})}\right)}. \quad (2)$$

The log term in the denominator of Equation 2 consists of a ratio of two “likelihoods.” The numerator and denominator of the likelihood ratio represent, respectively, the probability of measuring a particular combination of simple cell responses ( $r_1 r_2 r_3 \dots$ ) when a boundary is present at the reference location (yes), and the probability of measuring the same combination of filter responses when a boundary is not present at the reference location (no). When a simple cell response combination ( $r_1 r_2 r_3 \dots$ ) is measured in the vicinity of a reference location, and based on statistical record keeping we know that that response combination is more likely to be associated with a yes case than a no case, then the likelihood ratio is  $>1$ , meaning that there is positive evidence for a boundary. In this case, the logarithm term is  $>0$ , and the exponential term drops to  $<1$  because of the negative sign of the exponent. When the likelihood ratio is much  $>1$ , the exponential term approaches zero, and the denominator of the full expression approaches 1, indicating that a boundary at the reference location is highly probable. The other term in the denominator of Equation 2 is the prior “odds” of finding an object boundary at a randomly sampled location, which is the ratio of the prior probabilities of yes and no cases. Yes cases made up 2.4% of our default dataset (for  $N = 200$ ), meaning the odds of a boundary was  $2.4/97.6 = 2.45\%$ . The odds term functions as a sort of evidence threshold: the lower the prior odds of a boundary, the stronger the evidence must be to reach a 50% probability that a

boundary is present at the reference location. For prior odds of 2.45%, a likelihood ratio of 41 would be needed to reach a 50% boundary probability.

**Simplifying Bayes’ rule by assuming class conditional independence.** Using Bayes’ rule as a tool for interpreting cortical circuit interactions runs into the obstacle that the likelihood expressions contributing to the likelihood ratio involve high-dimensional probability densities, where the dimension corresponds to the number of simple cells that contribute to the boundary probability calculation (which could number in the 100s). Collecting high-dimensional probability density functions (pdfs) from natural images, and representing them mathematically, by rote tabulation, or neurally, is for all intents and purposes intractable. However, if we assume that neighboring simple cells display a certain kind of statistical independence (i.e., “class conditional” independence), this radically simplifies the computational problem by allowing the high-dimensional likelihoods in Bayes’ rule to be factored and re-expressed as a sum of one-dimensional LLR functions  $\lambda_i(r_i)$ . These one-dimensional LLR functions can be interpreted as “the effect that simple cell  $i$  has on a nearby boundary cell over its range of responses  $r_i$ .” Fortunately, the functions  $\lambda_i$  are simple in form (i.e., smooth and unimodal) and can be easily computed by a known cortical circuit motif. We therefore begin our analysis under the CCI assumption, as it provides a stepping stone to understanding the types of cell-cell interactions we should expect to find in a cortical circuit in which boundary probability is computed. We then take steps to bridge the gap between the predictions of the simplified analysis that assumes class conditional independence, and the types of cell-cell interactions that we should expect to find in the more realistic scenario where the simple cells contributing to a boundary probability computation do not satisfy the CCI assumption.

The assumption of class-conditional independence in our scenario implies that simple cell responses are statistically independent both for the set of image patches where a boundary is present at the reference location, and for the set of image patches where a boundary is not present at the reference location. When CCI holds, both the numerator and denominator terms in the likelihood ratio of Equation 2 can be factored into individual cell response terms, and converted to a sum of cell-specific LLR functions, as follows:

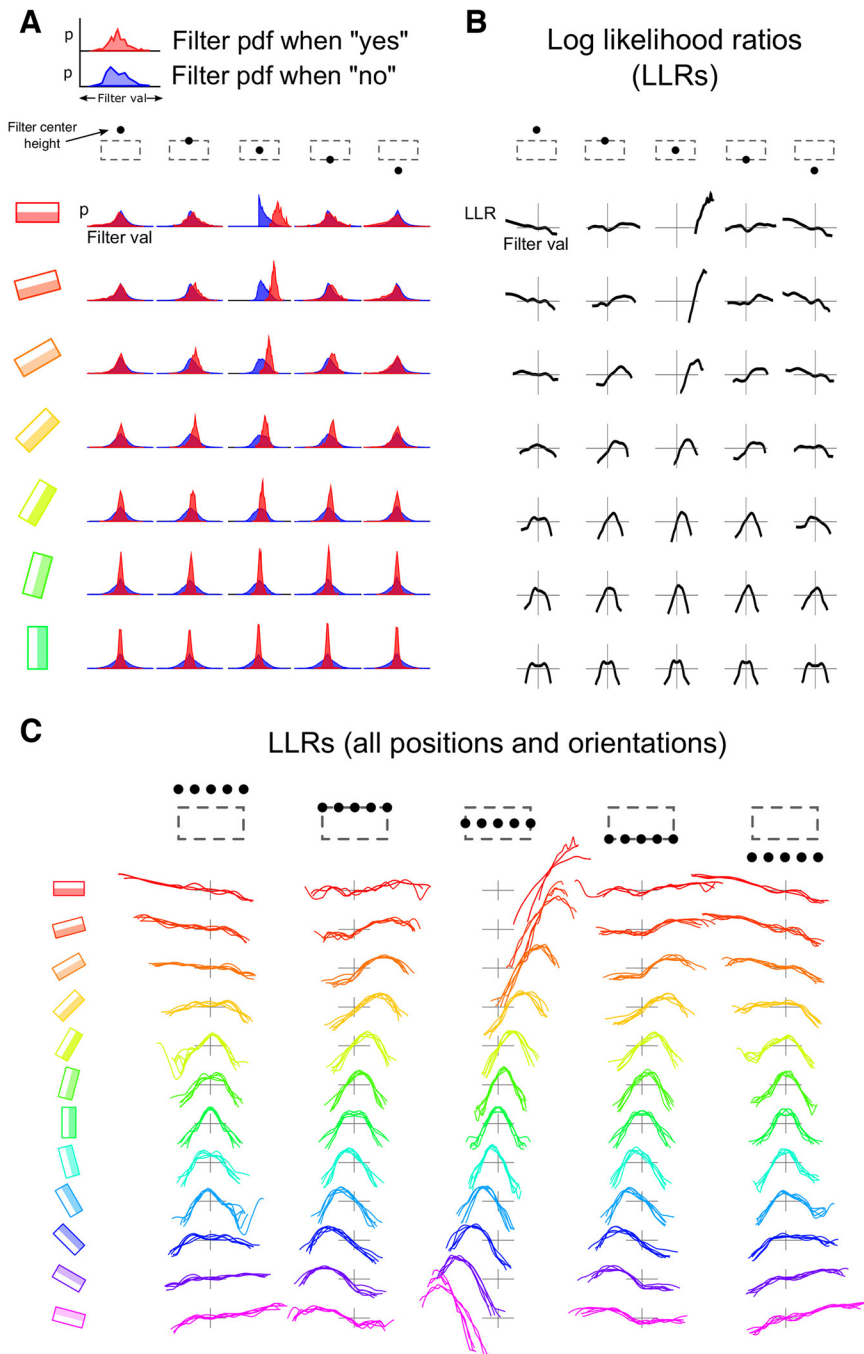
$$\begin{aligned} \log \frac{p(r_1 r_2 r_3 \dots | \text{yes})}{p(r_1 r_2 r_3 \dots | \text{no})} &= \\ &= \log \frac{p(r_1 | \text{yes}) \cdot p(r_2 | \text{yes}) \cdot p(r_3 | \text{yes}) \dots}{p(r_1 | \text{no}) \cdot p(r_2 | \text{no}) \cdot p(r_3 | \text{no}) \dots} \\ &= \log \frac{p(r_1 | \text{yes})}{p(r_1 | \text{no})} + \log \frac{p(r_2 | \text{yes})}{p(r_2 | \text{no})} + \log \frac{p(r_3 | \text{yes})}{p(r_3 | \text{no})} + \dots \\ &= \lambda_1(r_1) + \lambda_2(r_2) + \lambda_3(r_3) + \dots, \end{aligned}$$

where  $\lambda_i(r_i)$  is the log-likelihood ratio function for the  $i$ th neighboring cell. In simple terms, the LLR function  $\lambda_i(r_i)$  captures how the evidence for a boundary at the reference location depends on the response of the  $i$ th neighboring simple cell. We can write the CCI version of Bayes’ rule as follows:

$$p(\text{yes}|r_1 r_2 r_3 \dots) = \frac{1}{1 + \frac{p(\text{no})}{p(\text{yes})} \cdot \exp\left(-\sum_i \lambda_i(r_i)\right)},$$

which reduces to the remarkably simple formula:

$$= \text{sigmoid}\left(\sum_i \lambda_i(r_i)\right).$$



**Figure 3.** Computing LLR functions from natural images. **A**, Histograms were constructed for simple cell responses from 30,000 labeled image patches potentially containing boundaries at the reference location (dashed box) separately for yes (red) and no (blue) cases. Yes (no) cases were those with confidence scores of 4 and 5 (1 and 2). A subset of simple cell response histograms is shown for 7 orientations and 5 vertical positions (centered horizontally). **B**, By dividing the yes and no distributions and taking natural logarithms, one obtains the LLR functions  $\lambda_i()$  for each cell, which vary as a function of the cell response  $r_i$ . **C**, The full set of 300 LLR functions reveals a regular pattern over orientation and location. Cases grouped within each subplot are for five horizontal shifts (indicated by black dots at top). Many LLR functions are nonmonotonic functions of the response of the cell.

In intuitive terms, this equation says that for a boundary cell at the reference location to compute boundary probability within its receptive field, the response of each neighboring simple cell  $r_i$  should be passed through a nonlinear function  $\lambda_i$  whose form depends on the neighbor's offset in position and orientation from the reference location; the results from all contributing simple cells should be summed; and the total should be passed through a final sigmoidal nonlinearity to produce the boundary cell response (Fig. 1D).

*Extracting the log-likelihood ratio functions  $\lambda_i(r_i)$  from natural images.* Histograms of the responses of the 300 simple cells surrounding the reference location were collected separately for yes and no image patches (total of 600 histograms). Histograms for no patches contained 50 evenly spaced bins. Yes histograms were binned more coarsely because our dataset had many fewer yes patches than no patches; between 8 and 20 evenly spaced bins were used to ensure smoothness of the response histograms for all cells. The yes and no histograms for each simple cell were converted to yes and no pdfs by dividing the sample count in each bin by the following two factors: (1) the total sample count in the respective histogram; and (2) the bin width. From these pdfs, LLR functions  $\lambda(r)$  were constructed for each simple cell as  $\lambda(r) = \log \frac{p(r|\text{yes})}{p(r|\text{no})}$ ,

where  $p(r|\text{yes})$  and  $p(r|\text{no})$  were the yes and no pdf bin values of the cell, respectively. The log function was the natural logarithm (base  $e$ ). To control noise levels, the  $\lambda$  function of each cell was considered valid only at nonextreme response levels for which the probability in both the yes and no pdfs exceeded minimum thresholds ( $p(r|\text{yes}) > 0.005$  and  $p(r|\text{no}) > 0.002$ ). Different thresholds were used because smaller probabilities could be estimated more reliably in the no histograms given the much larger set of no image patches. Only data inside valid response regions is plotted in Figures 3–5.

The same procedure was repeated using different simple cell profiles ( $2 \times 6$ ,  $2 \times 8$ ,  $4 \times 8$ , and  $6 \times 8$  pixels) to generate the LLR functions shown in Figure 4.

*Modeling log-likelihood ratio functions  $\lambda(r)$  as incitatory interactions.* As a step in the direction of a circuit-level model, we fit the measured LLR functions  $\lambda(r)$  for all 300 cells surrounding the reference location with an additive combination of an excitatory and inhibitory interaction function. This was accomplished in two steps. First, for each plot in Figure 6C, we combined the data from the 5 LLR functions (i.e., combining across the five vertical RF shifts) and used least-squares regression to fit a quadratic function  $ar^2 + br + c$  using the data from positive filter values only. Whenever the fitted quadratic had a negative mode ( $-\frac{b}{2a} < 0$ ), which meant the data used

to fit the quadratic would not have included the mode, or if the fit had a positive second derivative corresponding to an upward-pointing parabola ( $a > 0$ ), which implied that the fit must have suffered from insufficient data, we refit the quadratic model including LLR values for negative filter values (corresponding to the left sides of each plot; data not shown). We then decomposed the quadratic fits into excitatory and inhibitory components as follows: the excitatory component  $E(r)$  reproduces the rising portion of the quadratic, and then saturates at the maximum value, while the inhibitory component  $I(r)$  is fixed to 0 for all filter values less than the mode, and thereafter rises quadratically, giving

the boundary cell response (Fig. 1D).



rise to the falling portion of the quadratic.

$$\text{Thus } E(r) = \begin{cases} ar^2 + br, & r < -\frac{b}{2a} \\ -\frac{b^2}{4a}, & r \geq -\frac{b}{2a} \end{cases} \quad \text{and}$$

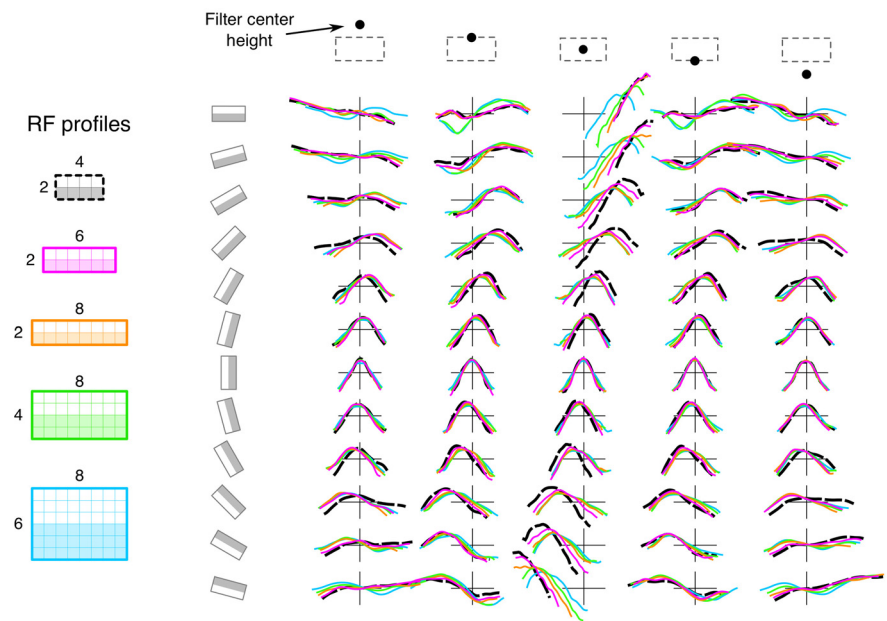
$$I(r) = \begin{cases} 0, & r < -\frac{b}{2a} \\ -a\left(r + \frac{b}{2a}\right)^2, & r \geq -\frac{b}{2a} \end{cases} \quad \text{and}$$

$E(r) - I(r)$  gives the quadratic fit obtained above (up to an additive constant, which is chosen so that all interactions are 0 when  $r = 0$ ). Note that the ability to precisely match individual LLR function shapes using a difference of two simple functions is mainly of didactic interest; the more practically significant question is whether a weighted sum of simple excitatory and inhibitory functions (which will in general involve more than two curves) can produce the LLR-like interactions needed to produce improved boundary detection performance (Fig. 7).

**Optimizing cell–cell interactions in a cortically inspired network using gradient descent learning.** In the data analysis portions of the article, the term “simple cell responses” referred to the value obtained by computing the dot product between the receptive field kernel and the underlying intensity image. This operation yielded the linear response value of the simple cell  $r_i$  for that receptive field location. In a step toward greater biological realism, for the network-level simulation of Figure 7, each of the 300 oriented receptive fields surrounding the reference location was represented by a population of eight model simple cells, all sharing the same receptive field and thus the same linear response value  $r$ , but having different output nonlinearities (reflecting natural variability within the neuronal population). The differentiation of responses within each eight-cell subpopulation was achieved by modifying the threshold in the sigmoidal output function applied to the linear response of each cell,  $y_i(r) = \frac{A}{1 + \exp[-g(r - t)]}$ . All cells shared the same amplitude and gain parameters ( $A = 1, g = 0.5$ ), but spanned a range of thresholds from  $t = -6$  to 35 in even steps.

The synaptic weights connecting each model simple cell to the boundary cell were obtained as follows. Each image patch created a pattern of activation across the 2400 model simple cells (300 linear RFs  $\times$  8 output threshold variants). We used logistic regression to train a linear classifier to distinguish boundary from nonboundary image patches using the 2400 model simple cells as inputs. A subset of the data (25,000 of the  $\sim$ 30,000 labeled patches) was used for training. During training, data were balanced by duplicating boundary-containing patches such that boundary and nonboundary exemplars were equal in number. Training was done using batch gradient descent with a learning rate of  $\eta = 0.1$ , performed for 1000 iterations. The net effect on the boundary cell of the eight model simple cells sharing the same RF location is visualized in Figure 7B. Each graph shows the weighted sum of the outputs of the eight simple cell variants as a function of  $r$ , the shared underlying linear response value of the cells. To facilitate comparison of these graphs with the explicit LLR function  $\lambda_i(r)$  for that RF location, we scaled the colored interaction functions within each plot. Each plot has one scaling factor that applies to all five colored curves in the plot. The inverse of the scaling factor, which can be thought of as the weight that the classifier puts on the curves contained in the subplot, is shown by gray bars.

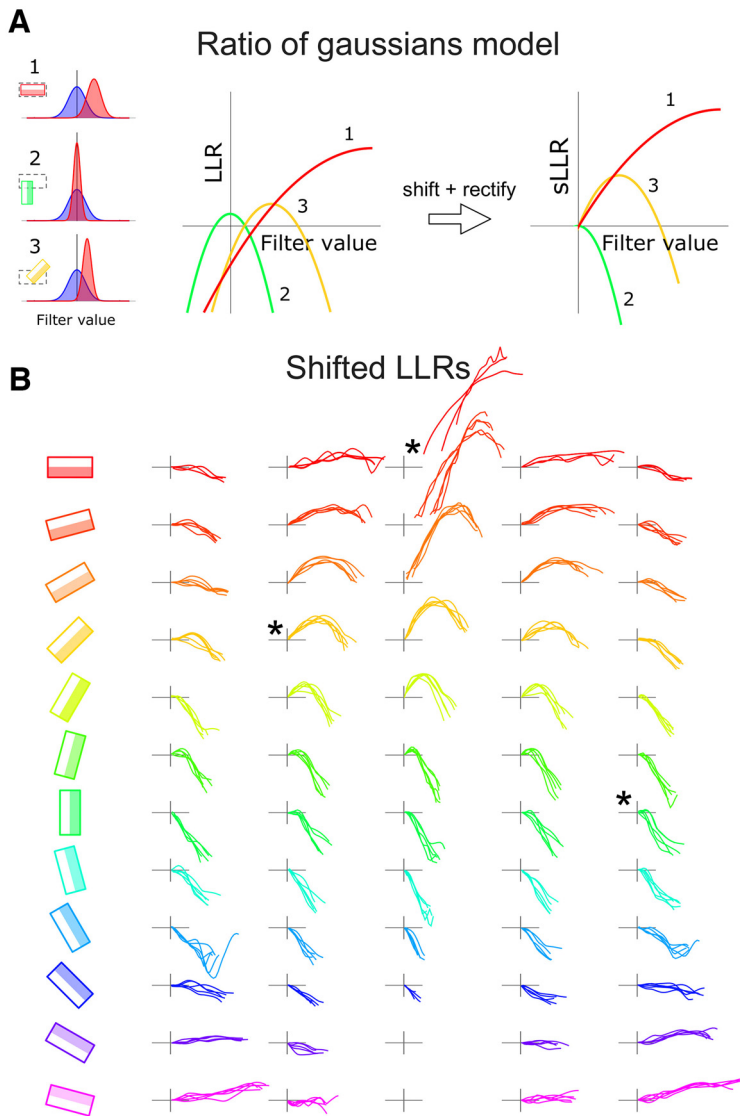
**Precision–recall curves measure boundary classification performance.** Precision–recall (PR) curves were generated for the boundary cell network with optimized weights, as well as for the naive Bayes’ classifier (based on a literal sum of all cell LLRs; Fig. 1C,D) and other classifier variants (see Fig. 9). The term “classifier” refers to any cell or network or



**Figure 4.** The basic pattern of LLR forms is conserved across different filter spatial profiles. LLRs were generated for each of the filter profiles shown on the left ( $2 \times 6$ ,  $2 \times 8$ ,  $4 \times 8$ , and  $6 \times 8$  pixels). The overall spectrum of LLR shapes remains similar for the different cases.

mathematical formula that produces a value in response to an image patch, where larger values are meant to signify higher boundary probability. A classifier consisting of the single simple cell at the reference location provided the PR baseline (see Fig. 9 blue curve). To generate a PR curve, a classifier was applied to each of the 5000 labeled but untrained “test” image patches, and the patches were sorted by the classifier output. A threshold was set at the lowest classifier output obtained over the entire test set and was systematically increased until the highest output in the test set was reached. For every possible threshold, above-threshold image patches were called putative boundaries and below-threshold patches were called putative nonboundaries. “Precision” was calculated by asking what fraction of patches identified as putative boundaries were true boundaries (according to the human ground truth labels), and “recall” was calculated by asking what fraction of true boundaries were identified as putative boundaries. As the threshold increased, the precision and recall values swept out a curve in PR space. Perfect performance would mean 100% precision and recall simultaneously, corresponding to the top right corner of the PR graph. Precision–recall curves are an alternative to receiver operating characteristic curves and are preferable in domains where the classes are very unbalanced in terms of prior probability. This is the case in our study: boundary images made up only 2.4% of the overall dataset, versus 97.6% for nonboundary cases. See this excellent article for a discussion of this issue (<https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>).

**Boundary cell stimulus responses.** The idealized boundary image, analogous to a spike-triggered average (STA) stimulus, was computed by averaging all natural image patches weighted by their boundary cell response (see Fig. 10A). Sinusoidal grating stimuli were generated on a  $20 \times 20$  pixel grid at 24 orientations in  $15^\circ$  steps, and at 60 phase shifts per cycle (see Fig. 10B,C). The spatial frequency was chosen to be 0.25 cycles/pixel because it led to relatively artifact-free stimuli at  $20 \times 20$  pixel resolution, and evoked robust boundary cell responses. For consistency with earlier results, the contrast of each grating image was adjusted to have the same normalizer value (200) as the natural image patches used in the LLR analysis. This was done by generating the grating patch at 100% contrast, computing the normalizer value  $N$  on the generated patch using the 100-cell normalization pool (see section above on normalization for details), and scaling down the grating patch by the factor  $N/200$ . This procedure reduced the worry that boundary cell responses to the artificial grating stimuli would be distorted by floor or ceiling effects.



**Figure 5.** Interpreting the LLRs as cell–cell interaction functions. **A**, Modeling the yes and no distributions as gaussians (left) leads to parabolic LLRs (middle). To interpret the LLRs as cell–cell interactions functions, we perform the following two additional processing steps: (1) when a simple cell is inactive, it should not influence the boundary cell, and this is accomplished by shifting the LLR to have zero output ( $y = 0$ ) when the input is zero ( $x = 0$ ); and (2) simple cells cannot have negative firing rates, and so the left halves of the LLRs, corresponding to negative simple cell firing rates, are discarded (these cases are handled by an opponent SC whose RF is identical but with the ON and OFF subfields reversed). This produces the curves in the right panel. **B**, The full set of LLR interactions processed in this way. Many of them are nonmonotonic, indicating that that simple cell should have a nonmonotonic effect on the boundary cell. The plots corresponding to the three LLRs modeled in **A** are marked with asterisks.

Simple cell responses to the grating stimuli were presented to the network of Figure 7 just as was done with natural image patches (see above). Responses were averaged over all phases of the grating at each orientation (see Fig. 10C). Tuning curves (see Fig. 10D) were obtained by presenting natural image stimuli from the  $N = 200$  set of normalized image patches. Red and blue curves are for images with 90th and 10th percentile contrast at the reference location, respectively. These percentiles varied in their contrast by approximately a factor of 2. Contrast was defined as the linear filter response at the reference location divided by the average intensity over the  $2 \times 4$  pixel region of support of the reference filter.

## Results

To gain insight into the cell–cell interactions needed for natural boundary detection, we collected and labeled 30,000 natural image

patches, with scores ranging from 5, indicating high confidence that a boundary was present at an RL (Fig. 1A, dashed box), down to 1, indicating high confidence that a boundary was not present at the RL. From these labeled patches, we constructed histograms of oriented linear filter values (representing simple cell responses) separately for yes (scores of 4–5) and no (scores of 1–2) cases (Fig. 3A, red and blue histograms, respectively). From the responses of 300 neighboring simple cells at 12 orientations on a  $5 \times 5$  pixel lattice centered on the RL, we computed the likelihoods  $p(r_i|\text{yes})$  and  $p(r_i|\text{no})$ , meaning the probability of the  $i^{\text{th}}$  simple cell having a particular response  $r_i$  when the patch either does (yes) or does not (no) contain a horizontal boundary. We show in the Materials and Methods section that, for a boundary cell to compute the probability of a boundary, and contingent on the assumption that the different filter responses are class conditionally independent, the boundary cell should have as its input the sum of the LLR

functions,  $\lambda_i(r_i) = \log \frac{p(r_i|\text{yes})}{p(r_i|\text{no})}$  of all contributing simple cells. Each function  $\lambda_i(r_i)$  represents the evidence—positive or negative—that simple cell  $i$ , when responding at level  $r_i$ , provides about the existence of a boundary at the reference location. This evidence should be summed by the boundary cell, and then passed through a sigmoidal output function to yield an actual boundary probability (Fig. 1C,D).

Accordingly, we computed the LLR functions for all of the 300 simple cells surrounding the reference location. Examples of LLR functions are shown in Figure 3B, and the full set is shown in Figure 3C grouped across five horizontal shifts at each orientation and vertical position. The LLR functions varied considerably with position and orientation relative to the reference location, but nonetheless conformed to a small number of qualitative shape prototypes (rising, falling, and bump shaped). When we generated LLR functions for simple cell receptive fields of different sizes and aspect ratios ( $2 \times 6$ ,  $4 \times 6$ ,  $4 \times 8$ , and  $6 \times 8$  pixel RF profiles), we found a qualitatively similar pattern of results, indicating that the basic shapes of the LLR functions do not depend sensitively on the RF profiles of the simple cell receptive fields (Fig. 4).

The LLR functions  $\lambda_i(r_i)$  are important functions to understand since they encode the way the response of a simple cell,  $r_i$ , individually affects boundary probability at a neighboring receptive field location (positively, negatively, or a combination of both). To gain insight into the forms of the  $\lambda_i(r_i)$  functions that we collected, we developed a simple mathematical model of the process of LLR function formation. If yes and no distributions of each cell are approximated as Gaussian (i.e., unimodal bell-shaped curves) with different means and variances, the resulting LLR functions will be quadratic in form, that is, they will have parabolic shapes. Since the no response distribution is virtually always more variable than the yes distribution, the LLR functions take the form of downward-pointing parabolas (Fig. 5A),



qualitatively resembling the LLR functions seen in Figure 3. The particular height and width of each LLR function is determined by the means and variances of the yes and no distributions for that cell (Fig. 5A, different colored curves). In addition to qualitatively capturing the range of observed LLR function shapes, this model has a simple interpretation in terms of natural image statistics: in image patches that do not contain boundaries at the reference location (which is to say most image patches), the responses of simple cells in the neighborhood tend to vary widely. On the other hand, image patches that do contain boundaries at the reference location are more constrained, and the responses of nearby simple cells tend to be clustered more tightly around characteristic values,  $r_i^*$ . The bump-shaped LLR functions, peaked at  $r_i^*$ , simply reflect the fact that a simple cell provides the strongest positive evidence for a boundary when its response is more typical of yes image patches than no patches.

To facilitate the interpretation of the  $\lambda$  functions as cell–cell interactions, we slightly reformatted them, in two ways. First, the  $\lambda$  functions were shifted vertically in order that they passed through the origin, reflecting the idea that when a simple cell is not firing (corresponding to  $r_i = 0$  on the graph), its influence on the boundary cell ( $y$ -value on the graph) should also be zero. This shift was justified given that the outputs of these functions would later be combined additively (Fig. 1C,D), and thus the vertical offsets across the entire population of simple cells could be collapsed into a single net offset at the level of the boundary cell (that would likely be small because of the cancellation of positive and negative shifts). Second, simple cell firing rates can only be positive, so the left half of each LLR function, corresponding to a negative simple cell firing rate, was “rectified” (i.e., set to zero). Information was not lost since the same or very similar function would be covered by a different simple cell with the same RF but opposite contrast polarity. The right panel of Figure 5A shows the combined effect of the shift and rectify operations. The full set of shifted LLR (sLLR) functions  $\lambda_i^s(r_i)$  obtained this way is shown in Figure 5B, with the plots corresponding to the conceptual curves in Figure 5A marked by asterisks.

Returning to the interpretation of shifted LLR functions as simple cell–boundary cell interactions, for some simple cells  $\lambda^s$  increased monotonically from the origin, meaning that, as the simple cell response increased from zero, the evidence it provided to the boundary cell grew steadily more positive. This type of monotonic positive SC–BC interaction was seen for simple cells that were most directly supportive of the hypothesis that a boundary was present at the reference location, such as the simple cell directly overlapping with the RL (Fig. 5B, middle column, top row). Referring to the model of Figure 5A, this was a case where the downward-pointing “parabola” peaked far to the right of the origin, so that over the entire observed firing range of the simple cell, its effect on the boundary cell remained on the rising limb of the parabola (Fig. 5A, case 1, right panel). At even higher firing rates than are plotted in Figure 5B, the  $\lambda^s$  function would eventually reach its peak and turn back downward, but such high filter values were so rare in yes patches in our natural image dataset that the LLR curves could not be reliably estimated beyond the range shown. Two other cases of pure positive  $\lambda^s$  functions are worth noting (Fig. 5B, bottom left and right corners). These cases apply to simple cells whose RFs are nearly “upside down” (i.e., polarity reversed) versions of the reference filter kernel, but shifted vertically 2 pixels either above or below the reference location. The fact that these cells are monotonically supportive of the reference hypothesis can be attributed to the existence of

many 1- to 2-pixel-wide light and dark horizontal bands in our natural image dataset.

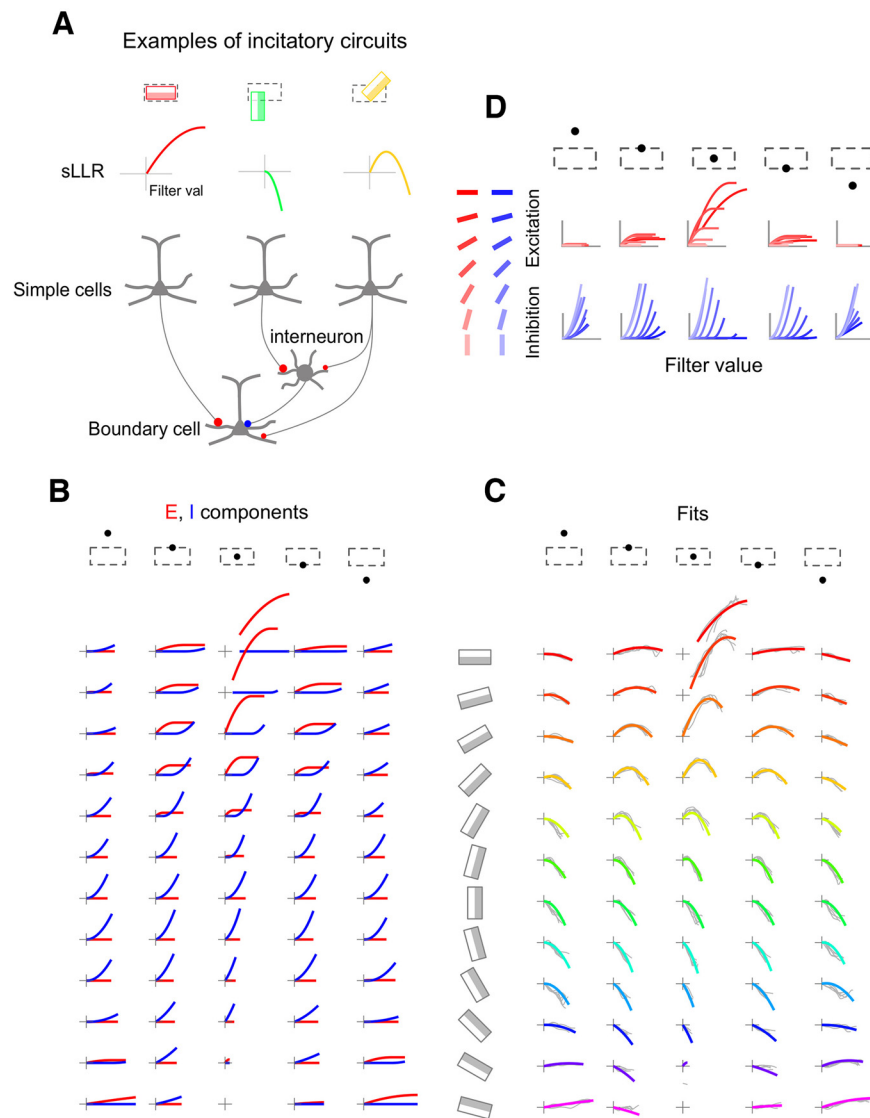
For other simple cells, the  $\lambda^s$  functions decreased monotonically from the origin, meaning that, as the firing rate of the simple cell increased from zero, the evidence it provided to the boundary cell grew increasingly more negative. This type of monotonic negative cell–cell interaction was seen for simple cells whose firing supported a hypothesis incompatible with the hypothesis that a boundary was present at the RL. The clearest examples of such cells are those with RFs perpendicular to the RL (Fig. 5B, middle row, green LLR curves). Referring again to the quadratic LLR model of Figure 5A, these monotonically decreasing  $\lambda^s$  functions arose from cases where the downward-pointing LLR parabola was peaked at, or to the left of the origin, so that over the entire response range of the simple cell (to the right of the origin), the  $\lambda^s$  function fell continuously along its descending limb (Fig. 5A, case 2).

For the majority of simple cells, however, the  $\lambda^s$  functions were bump shaped, first rising and then falling as the firing rate of the simple cell increased from zero. This type of nonmonotonic cell–cell interaction was seen for simple cells whose receptive fields had some degree of overlap in position and orientation with the reference location, so that when a boundary was present at the reference location, their characteristic response levels  $r_i^*$  fell in the “middling” range. As a result, these cells provided increasing positive evidence for a boundary at the reference location for response levels up to  $r_i^*$ , but at even higher response levels, they began to signal stronger evidence for a boundary at their own RF location and orientation rather than the reference location. This led to a progressive weakening of the evidence signal sent to the reference boundary cell at response levels above  $r_i^*$ .

### A known circuit mechanism can produce the entire observed spectrum of $\lambda^s$ functions

Given that the input to a boundary cell can be approximated as the sum of the  $\lambda^s$  functions associated with different simple cells (Fig. 1D), and that the shifted LLR functions for different simple cells can be either monotonic or nonmonotonic functions of the simple cell response, we next asked what kind of neural interconnection circuit is capable of producing the types of monotonic and nonmonotonic interaction functions that we observed. One candidate mechanism is the ubiquitous circuit motif in which a cortical cell both directly excites and disinaptically inhibits other cells in its neighborhood (Buzsáki, 1984; McBain and Fisahn, 2001; Pouille and Scanziani, 2001; Swadlow, 2002; Wehr and Zador, 2003; Klyachko and Stevens, 2006; George et al., 2011; Isaacson and Scanziani, 2011; Pfeffer et al., 2013). For monotonic positive or negative SC–BC interactions, the effect could in principle be mediated by pure excitatory or inhibitory connections, respectively (through an interneuron in the case of a pure inhibitory connection; Fig. 6A, left and middle cases). Nonmonotonic cell–cell interactions, however, would seem to require a compound E–I interconnection scheme (Fig. 6A, rightmost case), wherein the excitatory effect dominates at low firing rates and the inhibitory effect dominates at high firing rates.

To determine whether this circuit motif can in principle produce the full range of cell–cell interactions contained in our dataset, we assumed that the sLLRs representing simple cell–boundary cell interactions were parabolic in shape, as discussed above, and that each parabolic fit to an sLLR could be expressed as a sum of two simple monotonic functions  $E(r)$  and  $I(r)$  representing, respectively, the direct excitatory and indirect inhibitory effect of the simple cell on the boundary



**Figure 6.** Fitting simple cell–boundary cell interactions (LLRs) with a difference of excitatory and inhibitory effects. **A**, Each of the three sLLRs shown can be parametrized by an incitatory circuit. The circuits implementing the red and green sLLRs involve pure excitation and inhibition, special cases of incitation, while that of the orange sLLR involves a nontrivial combination of both excitation and inhibition. **B**,  $E$  (red) and  $I$  (blue) curves were optimized so that their difference fit the corresponding LLR shown in **C** (for details, see Materials and Methods). **C**, LLR fits are shown in color, on top of the five-curve groups from Figure 3C shown in light gray. **D**,  $E$  and  $I$  interactions from **B** are collected across orientations within each subplot, showing smooth progressions of parameters.

cell (for definitions of  $E(r)$  and  $I(r)$ , see Materials and Methods). The  $E$  and  $I$  interaction functions found through this procedure are shown in Figure 6B, and the parabolic fits to the  $\lambda^s$  functions, corresponding to the sum of the corresponding  $E$  and  $I$  curves, are shown in Figure 6C. The generally good fits to the measured sLLRs confirmed that the spectrum of cell–cell interactions needed to calculate boundary probability in natural images, including nonmonotonic interactions, can in principle be produced by the “incitatory” circuit motif shown in Figure 6A. To verify that good fits to the spectrum of  $\lambda^s$  functions did not depend on our particular choice of  $E$  and  $I$  functions (which in this case were constrained to sum to a downward-pointing parabola), we repeated the fitting procedure using several alternative sets of monotonic  $E$  and  $I$  functions, including sigmoids, lines, power functions, and various saturating functions. We obtained good qualitative fits in each case (data

not shown), and, based on this, concluded that the basic types of cell–cell interactions needed to detect object boundaries in natural images fall easily within the scope of the simple incitatory circuit shown in Figure 6A, which is known to be present in V1.

Returning to the parabolic fits of the sLLRs, we next looked for regularities in the progression of excitatory–inhibitory curve pairs used to fit the  $\lambda^s$  functions depending on the offset of a neighbor cell in position and orientation from the reference location (Fig. 6D). We observed the following patterns. First, for a neighboring cell centered at the RL (Fig. 6D, middle column) as the receptive field of the neighbor rotates out of alignment with the RL (indicated by a lightening of the curves within each subplot), excitation becomes weaker, and inhibition becomes both stronger and lower in threshold. This pattern amounts to a form of cross-orientation suppression, a staple function of V1 (Bishop et al., 1973; DeAngelis et al., 1992; Geisler and Albrecht, 1992; but see Priebe and Ferster, 2006). A second, more subtle pattern can also be seen: for simple cells whose RFs are slightly vertically shifted relative to the RL of the boundary cell (Fig. 5C, second and fourth columns), the most peaked  $\lambda^s$  functions are seen for simple cells whose RFs are slightly misoriented relative to the boundary cell RL (i.e., by  $\pm 30$ – $45^\circ$ ; Fig. 5C, third and fourth rows). The steeper sides of these peaky  $\lambda^s$  functions, combined with the left shift in their peaks, led our least-squares fitting procedure to assign stronger excitatory weights to these cells compared with simple cells perpendicularly offset but parallel to the boundary RL (Fig. 5C, first row). The effect can be seen in Figure 6D, where the boldest red curve corresponding to the preferred orientation of the boundary cell is not the curve with maximum amplitude. We found that this same pattern held for fits to the  $\lambda^s$  functions using the several

other combinations of monotonic  $E$  and  $I$  functions that we tested (data not shown), but the effect should be taken as a weak prediction of our model, given it was relatively mild and dependent on numerous parameters. A more general, stronger conclusion supported by Figure 6D is that the strength of the excitation and inhibition received by a boundary cell from neighboring simple cells will vary systematically with the offsets of simple cells in RF position and orientation relative to the boundary cell RF, and in a way that can be predicted within a natural image-based normative framework such as the one we have adopted here.

#### Optimizing the parameters for a boundary-detecting incitation network

In pursuit of our goal to understand how cells in V1 detect natural object boundaries, our approach thus far has been to frame

boundary detection as a Bayesian classification problem, where the inputs are simple cells (with the simplifying assumption that all simple cells are CCI; Fig. 1), collect ground truth data from human-labeled natural images (Figs. 3, 4), and calculate what the simple cell–boundary cell interactions should look like in situations where the CCI assumption holds true. The approach has led to two main findings. First, simple cell–boundary cell interaction functions  $\lambda_i^s$  are in general nonlinear and often nonmonotonic (Fig. 5), and so cannot be captured by a direct connection from an SC to a BC through a conventional positive or negative weight. On the other hand, we showed that the entire family of rising, falling, and bump-shaped SC–BC interaction functions can be captured by a two-stage “incitation” circuit motif that is known to exist in V1 (Fig. 6A).

The precise forms of the SC–BC interaction functions that we might expect to find in V1 remain in question, however: the  $\lambda_i^s$  functions shown in Figures 5B and 6C are only the optimal cell–cell interaction functions if all of the simple cells impinging on the boundary cell show CCI. To reiterate, this rather severe condition means that no simple cell provides any information about the activity level of any other simple cell either on or off object boundaries (for details, see Materials and Methods and Appendix). The condition is, of course, trivially true for a single simple cell, and in that case, the functional connection of the simple cell to the boundary cell should be exactly its LLR. The assumption may also hold true for a small population of minimally overlapping simple cells (see Ramachandra and Mel, 2013), but it is definitely not true for 300 simple cells with densely overlapping receptive fields, which is the scenario we consider here. Generally speaking, if the number of simple cells exceeds the number of underlying stimulus dimensions, which is at most 64 in our case (i.e., the number of pixels covered by the RF profiles of 300 simple cells), then the representation is “overcomplete,” and the simple cells are necessarily correlated. For such a correlated cell population, the likelihood terms in Equation 2 can no longer be factored, and the boundary probability can no longer be expressed as a sum of cell-specific LLR terms  $\lambda_i(r_i)$ .

This leads to the following interesting question: instead of imposing parameters on the incitation circuit to represent cell-specific  $\lambda_i^s$  functions, as we did in Figure 6, suppose the synaptic weights of an incitation circuit are trained to maximize boundary detection performance. What would the individual SC–BC interaction functions look like then? Would they be similar to their cell-specific  $\lambda_i^s$  functions? Or would the SC–BC interaction functions be significantly altered, presumably because of correlations within the simple cell population?

To answer this question, we set up a slightly augmented incitation network (Fig. 7A) whose modifiable parameters included (1) the excitatory weights connecting each simple cell to the boundary cell and (2) the excitatory weights connecting each simple cell to the inhibitory interneuron. In addition, to make it possible for the network to learn nonlinear SC–BC interaction functions by modifying only a single layer of excitatory weights, each simple cell was replicated eight times to form a small population of closely related cells, all sharing the same oriented receptive field, but each having a different firing threshold. The threshold variability can be seen as arising from, for example, natural variation in neuron size, morphology, and firing dynamics. The functional purpose of this scheme is that the activation level of each simple cell (before the threshold is applied) is effectively being recoded through a set of fixed nonlinear

basis functions, which facilitates learning. The regularly spaced threshold settings used for the groups of eight cells are given in the Materials and Methods section. Each presynaptic simple cell acted on the boundary cell through two adjustable weights, one excitatory weight directly onto the boundary cell, and one excitatory weight onto the inhibitory “partner” cell of the boundary cell, which would contribute to disinhibitory inhibition of the boundary cell (Fig. 7A). The inhibitory neuron was modeled as a linear cell whose firing rate was a weighted sum of its synaptic inputs. Three examples of oriented receptive fields (red, green, and yellow) and their associated simple cell variants are depicted schematically in Figure 7A.

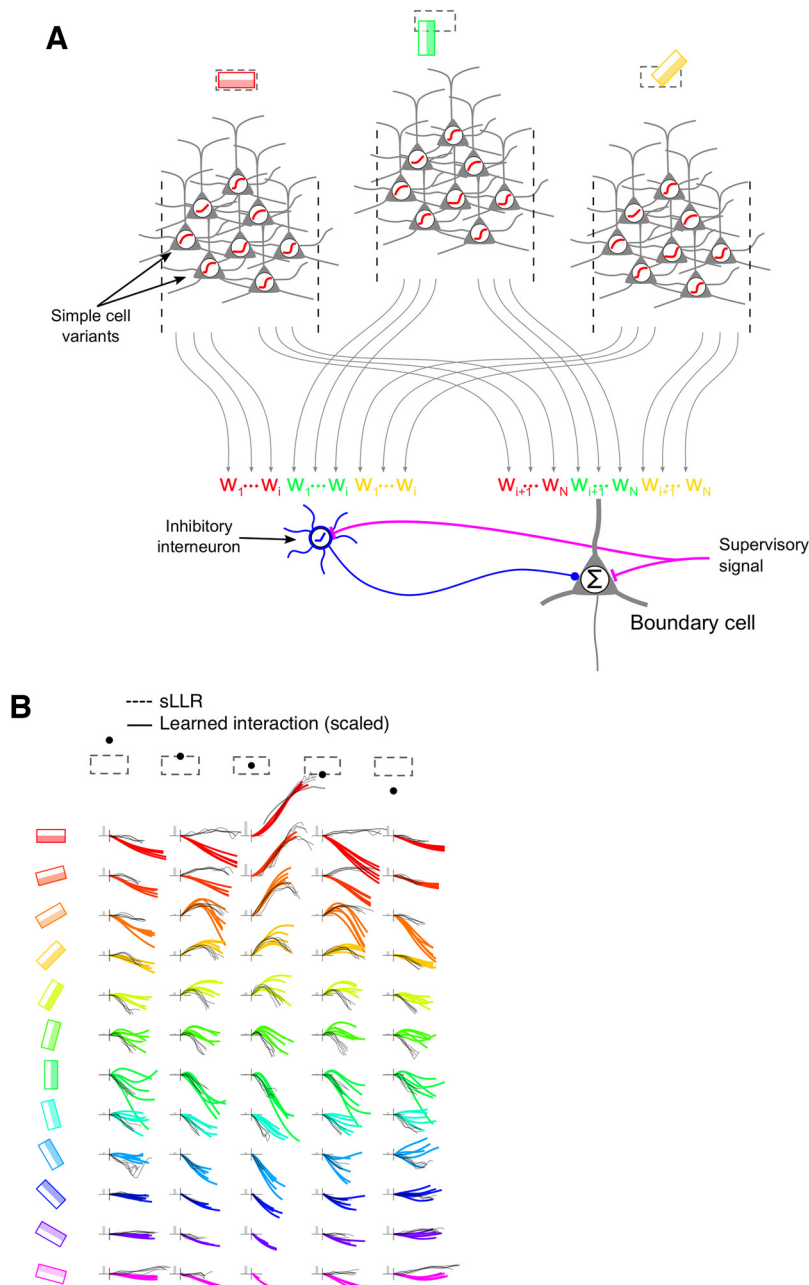
Training occurred as follows. Labeled image patches containing boundaries and nonboundaries (with equalized probability) were presented to the 2400 (= 300 × 8) simple cells; ground-truth labels from the natural image dataset were presented to the boundary cell (1 for boundary, 0 for no boundary); and the excitatory synapses between the simple cells and the boundary cell and its associated inhibitory neuron were adjusted using a supervised logistic regression learning rule (Murphy, 2012). We then performed virtual neurophysiology to probe the net effect of each oriented receptive field on the boundary cell response induced by the eight simple cell variants (i.e., basis functions) sharing that RF.

The learned interaction functions again included monotonic rising and falling as well as nonmonotonic bump-shaped functions (Fig. 7B, colored curves). For some cells, the learned SC–BC interaction functions corresponded closely to their respective sLLRs (thin gray lines), most notably the cells centered on the RL at all different orientations (Fig. 7B, middle column). For this category of cells, the simplified Bayesian formulation seems to explain their role in the boundary detection computation, but no strong inference can be made along these lines given that we could neither predict, nor retroactively account for, which cells fell into this category. In other cases, one or two of the learned SC–BC interaction functions in each group of five overlapped heavily with their corresponding sLLR curves, whereas the other curves in the group were driven apart by the learning rule to cover a much wider spread (vertically or horizontally or both) than the original set of sLLRs. In still other cases, the learned interaction functions were nearly “opposite” to their corresponding sLLR functions (Fig. 7B, red curves in columns 2 and 4 of the second row). In these cases, also, we were unable to explain why the learned interaction functions of the cells deviated from the predictions of the simplified Bayesian model. In the hopes of at least confirming that the deviations were caused by violations in the CCI assumption, we conducted a simple experiment, described next, in which correlations between pairs of simple cell inputs to a boundary cell were systematically manipulated.

### Probing the relationship between the incitation circuit and Bayes’ rule

To probe the role of correlations between simple cells in shaping SC–BC interaction functions in a boundary-detecting circuit, we ran a simple experiment in which a boundary cell received input from just two simple cells whose RFs overlapped to varying degrees. For each simple cell pair, we fit the parameters of the incitation circuit either separately (Fig. 8A, left) or jointly (Fig. 8A, right). We tested pairs of filters ranging from very dependent (Fig. 8B, middle columns) to nearly independent (Fig. 8B, outer columns). Scatter plots of joint filter responses to boundary (red)





**Figure 7.** Simple cell–boundary cell interactions can be learned by a biologically plausible synaptic plasticity rule. **A**, Each oriented filter was represented by a population of 8 simple cells, each with a different fixed input/output nonlinearity. Nonlinearities were sigmoids,  $y = \frac{1}{1 + e^{-x}}$ , with threshold  $t$  set at 8 evenly spaced values between  $-6$  and

35. The learning rule used to adjust the weights from each simple cell onto the inhibitory and boundary cell was  $\Delta w_i = \pm \eta (t - y) x_i$ , where  $t$  is the “training signal” (1 for boundary, 0 for no boundary),  $y$  is the response of the boundary cell,  $x_i$  is the response of the  $i$ th simple cell,  $\eta$  is the learning rate, and the positive (negative) sign was used for the boundary (inhibitory) cell. In the context of our model, this learning rule is mathematically equivalent (up to a transient initial difference in the learning rate parameter  $\eta$ ) to a learning rule that constrains all weights to be positive. **B**, To determine the net effect of each filter on the boundary cell (for comparison with the LLRs), the underlying linear filter value was increased from 0 to 1 while holding all other inputs constant, and the weighted sum of the eight associated simple cells was plotted (colored curves). Black curves are sLLRs from Figure 5B. The gray bar in each plot represents the weight that the BC puts on that group of five colored curves.

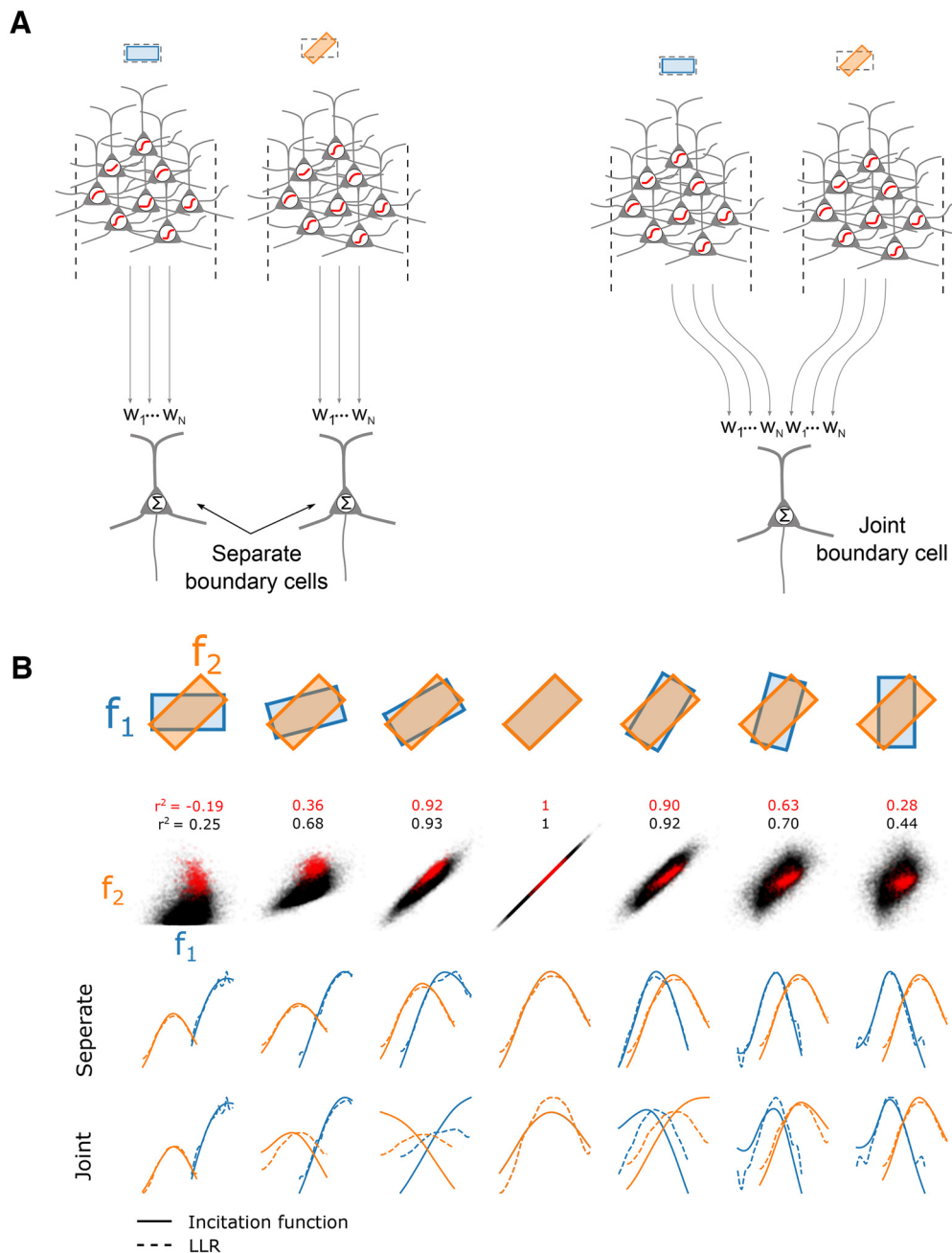
and nonboundary (black) patches are shown below each pair. When the SC–BC interaction functions were learned separately, they were nearly identical to the literal LLRs [Fig. 8B, first row of blue and orange curves (solid curves show learned interactions, dashed curves show LLRs)]. On the other hand, when the SC–

BC interactions were learned jointly, for SC pairs with heavily overlapping RFs, which led to a breakdown of the CCI assumption, the learned interactions differed significantly from the pure LLRs (Fig. 8B, middle columns). Consistent with these observations, we show analytically in the Appendix that an incitation circuit like the one shown in Figure 7A will learn LLRs if the input features have CCI. Consequently, any observable differences between the learned incitation functions and the LLRs must be attributable to a breakdown of class-conditional independence.

### Comparing boundary detection performance of four models

The optimization of synaptic weights in the incitation circuit of Figure 9A opens up an additional avenue for validation (or refutation) of our overarching boundary cell hypothesis. Our main premise is that a boundary cell in V1 should be able to significantly improve its boundary detection performance (compared with a single simple cell at the reference location) if it can access, through the local cortical circuit, a large and diverse set of simple cells covering the neighborhood. A major additional claim is that the incitation circuit motif, which is known to exist in V1, is well suited to deliver such a performance improvement. To directly test these claims, we compared the precision–recall performance of the trained incitation network (Fig. 9, red curve) to three other boundary detectors: (1) the “null hypothesis,” consisting of a single conventional simple cell centered at the reference location (Fig. 9, blue curve); (2) an unweighted sum of 300 literal LLRs (orange curve), which is essentially a direct implementation of Bayes’ rule under the CCI assumption (Fig. 1D); and (3) a weighted sum of the same 300 literal LLRs (Fig. 9, green curve). This hybrid model honors the basic structure of the Bayesian classifier of Figure 1D, but allows weights on each of the LLR inputs to help compensate for CCI violations in the simple cell population.

The results shown in Figure 9 support the following three conclusions: (1) the superior performance of all three multi-input classifier variants compared with a single conventional simple cell reinforces the point that individual simple cells are seriously underpowered as natural boundary detectors (see Ramachandra and Mel, 2013); (2) the superior performance of the two classifier variants with learned synaptic weights (Fig. 9, red and green curves) compared with the simplified Bayesian classifier that receives unweighted LLR inputs (Fig. 9, orange curve) attests to the



**Figure 8.** The incitation circuit learns literal LLRs when the filters are class-conditionally independent. **A**, We selected several pairs of filters and fitted either 2 separate incitation circuits, 1 for each filter (left), or 1 circuit with both inputs (right). **B**, Top, Filter pairs ranged from very different (left and right) to very similar or identical (middle) filters. Middle, Scatter plots of joint filter responses for boundary (red) and nonboundary (black) image patches. Bottom, When filters were fit separately, the learned incitation functions (solid curves) were nearly identical to the LLR curves of the filters (dashed). When the filters were fit jointly, pairs with very similar filters no longer learned LLR functions because of a breakdown of CCI.

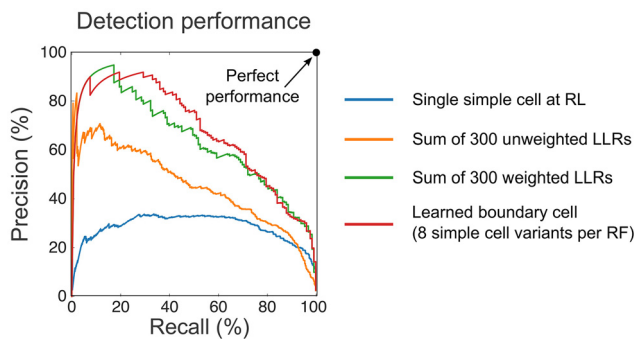
importance of a learning rule that is sensitive to natural image statistics and can help compensate for unwanted input correlations; and (3) the very similar performance of the optimized incitation network (Fig. 9, red curve) and the weighted sum of LLRs (green curve), especially in the high recall range, is again suggestive of a nontrivial connection between the simplified Bayesian classifier of Figure 1A and the behavior of the learned boundary detecting incitation circuit of Figure 7A.

We conclude by noting that the requirements for developing a cortical circuit that produces significantly improved boundary detection performance compared with a conventional simple cell are relatively modest, including (1) a compound E–I circuit

motif that we have dubbed an incitation circuit, which is known to exist in V1; (2) variability in firing thresholds across the population of simple cells; and (3) a supervised “delta” rule capable of setting the strengths (and/or dendritic locations) of the excitatory contacts from simple cells onto boundary cells and their associated interneurons. Possible sources of the supervisory signal are taken up in the Discussion.

## Discussion

In the 60 years since Hubel and Wiesel (1962) first discovered orientation-tuned simple cells in V1, it has been generally



**Figure 9.** Precision–recall curves for 4 boundary detecting classifiers. See Materials and Methods for definitions of precision and recall. Perfect performance is at top right-hand corner. A single oriented simple cell at the reference location performs poorly ( $\sim 20\%$  precision at 90% recall; blue curve). A classifier consisting of a sum of 300 unweighted simple cell LLRs is shown in orange. This case corresponds to the simplified Bayesian classifier depicted in Figure 1, C and D. Poor performance was expected given violations of the CCI assumption in the simple cell population. The green case is similar, but with weights on each of the 300 LLRs trained by logistic regression. Performance is dramatically improved, indicating that the learning rule helps to compensate for correlations in the simple cell population. The red curve shows performance of the trained incitation network of Figure 7A. Precision exceeds that of a single simple cell by  $\sim 1.5\times$  at 90% recall, and  $2.3\times$  at 50% recall.

assumed that these cells contribute in some way to the detection of object boundaries (Field et al., 1993; Grosf et al., 1993; Kapadia et al., 1995, 2000; Polat et al., 1998; Sceniak et al., 1999; Angelucci et al., 2002). Consistent with this idea, virtually every modern object recognition system, whether designed by hand or trained from natural image data, includes simple cell-like filtering in its early stages of processing (Fukushima et al., 1983; Lades et al., 1993; Mel, 1997; Lecun et al., 1998; Riesenhuber and Poggio, 1999; Krizhevsky et al., 2012). Surprisingly, however, the quantitative relationship between simple cell responses, typically modeled as divisively normalized linear filters (Carandini and Heeger, 2011), and object boundary probability in natural images has been little explored (though see Ramachandra and Mel, 2013), making it difficult to know whether or how V1 circuits contribute to this behaviorally relevant natural computation. It is important to emphasize that a simple cell on its own is a poor detector of natural object boundaries within its receptive field (but see Arbeláez et al., 2011): as shown in Figure 9 (blue curve), if we use the response of a simple cell as an indicator of the presence of an object boundary within its RF, even when the threshold for detection is raised to such a high value that half of all true boundaries are rejected (corresponding to a recall score of 50%), almost two-thirds of the “detected” edges at that threshold are false positives (corresponding to a Precision score of  $\sim 35\%$ ). The reason a simple cell is such an unreliable edge detector is that true object boundaries are rare (see Fig. 11A, where the overwhelming majority of points are piled in the bottom half of the plot), and when they do occur, they are very often of low contrast. Much more common are high-contrast nonedge structures (e.g., textures) that contain sufficient oriented energy to strongly drive simple oriented filters.

The poor boundary detection performance of a lone simple cell leads to the conjecture that V1 also contains “smarter” cells that compute boundary probability by combining the responses of multiple simple cells covering a local neighborhood. In a previous study, we suggested that the appropriate strategy for constructing a boundary cell from a local population of simple cells is as follows: (1) select a small set of simple cells (e.g., six cells) that are both individually informative and class-conditionally

independent (for discussion of the CCI assumption, see Materials and Methods); (2) evaluate the log-likelihood ratios for each of the participating simple cells, which tells us the optimal functional interconnections between each simple cell and the boundary cell (according to Bayes’ rule); and (3) sum the LLRs and apply a fixed sigmoidal nonlinearity to compute boundary probability (Ramachandra and Mel, 2013; Fig. 1C,D). The present study extends that previous work in eight ways: (1) we collected and analyzed individual LLRs for all of the even-symmetric simple cells at all orientations covering a  $5 \times 5$  pixel neighborhood in the vicinity of the RF of a boundary cell (300 cells total); (2) we show that the idealized functional interconnections between simple cells and boundary cells depend systematically on the relative positions and orientations of the simple cell and boundary cell RFs (Fig. 3), but are relatively insensitive to the scale or aspect ratio of the simple cell receptive fields (Fig. 4); (3) we developed a simple analytical model (i.e., Gaussian likelihoods, leading to quadratic LLRs) that shows how the three seemingly different types of SC–BC interaction functions—rising, falling, and bump-shaped functions—represent different ranges of the same underlying (quadratic) function class (Fig. 5); (4) we show that a mixed excitatory–inhibitory, or incitatory, circuit motif that is known to exist in V1 is capable of producing the entire spectrum of natural image-derived SC–BC interaction functions (Fig. 6); (5) we show that the parameters of a boundary-detecting incitation circuit can be learned by adjusting a single layer of excitatory weights (Fig. 7A); (6) we show that a learned incitation circuit can improve the precision of boundary detection in the high-recall range by 43% to 121% compared with a conventional simple cell model (Fig. 9); and (7) by “reading out” the weights of the learned incitation circuit, we show that the simple cell–boundary cell interaction functions that we would expect to find in the visual cortex are not likely to be verbatim LLRs, but rather, perturbed versions because of class-conditional dependencies among simple cells whose receptive fields overlap heavily with each other (Figs. 7B, 8). This could be helpful in interpreting the results of future neurophysiological experiments in V1.

## Experimental predictions

### *Distinguishing boundary cells from conventional simple cells*

Having shown that cortical circuitry is capable in principle of producing boundary cells from simple cells using only a single layer of modifiable excitatory weights, it is important to ask how BCs could be detected experimentally and distinguished from conventional simple cells (or the simple cell-like subunits of complex cells; Hubel and Wiesel, 1962; Movshon et al., 1978; Ohzawa et al., 1997).

To determine how BCs would respond to various stimuli, stimulus patches were scaled to have the same fixed value of the normalizer used in earlier figures, once again reflecting a simple form of divisive normalization (see Materials and Methods subsection Boundary cell stimulus responses). We first constructed a canonical stimulus for a boundary cell akin to a spike-triggered average by averaging all image patches weighted by their evoked boundary cell response. As expected, the STA stimulus appears as a localized, polarized, oriented boundary segment reminiscent of the receptive field of a simple cell (Fig. 10A). We then presented drifting sine wave gratings covering the “classical receptive field” of a boundary cell, leading to the unremarkable phase response and orientation tuning curves shown in Figure 10, B and C. Next, we used labeled natural edges with the same normalizer value to explore the effect of increasing center contrast



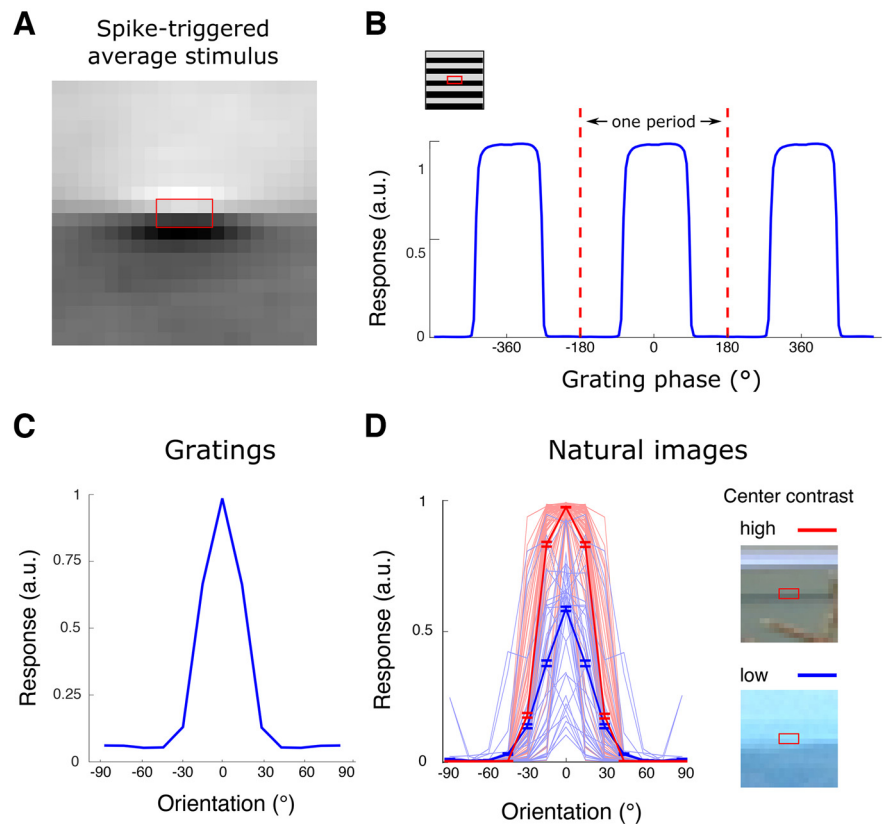
on the orientation tuning curve width. (This was not a perfectly controlled experiment because variations in center contrast at a fixed normalizer value would have led to antivariations in surround contrast, but given the filter value at the RF center was only 1 of 100 filters of many orientations used to compute the normalizer value, this effect was likely small.) Subject to this limitation, as shown in Figure 10D, the tuning width of the boundary cell is essentially constant across an approximately twofold change in center contrast, the limit of analysis allowed by our labeled database (average tuning curve has full-width at half-height for high-contrast stimuli, 43.6°; average tuning curve has full-width at half-height for low-contrast stimuli, 39.2°).

Thus, for oriented edges and gratings presented within the classical RF, boundary cells behave similarly to conventional simple cells in that they (1) have phase-dependent responses, (2) are orientation tuned, and (3) have tuning curves whose widths are approximately contrast invariant (Alitto and Usrey, 2004). It is therefore possible that boundary cells exist and have been classified as conventional simple cells in previous experiments using simplified stimuli. Among the multiple types of V1 cells that have been described previously, boundary cells share most in common with double opponent cells, which are orientation tuned, have mostly odd-symmetric receptive field profiles, as would be expected for boundary detecting cells (Ringach, 2002), and respond to boundaries whether defined by luminance or color (Johnson et al., 2008).

In future neurophysiological studies, an efficient means of dissociating conventional simple cells, which respond to oriented contrast independent of boundary probability, from putative boundary cells, which respond to boundary probability independent of oriented contrast, would be to use natural image stimuli drawn from the four corners of the oriented contrast–boundary probability space (Fig. 11A). Image patches with low oriented contrast and low boundary probability scores (purple dots) tend to contain flat, unstructured image regions; patches with low contrast and high probability (Fig. 11A, green dots) tend to contain well structured, faint edges; patches with high contrast but low probability (Fig. 11A, blue dots) tend to contain contrasty noise or misaligned edges; and regions with high contrast and high probability (red dots) typically contain well structured, strong edges (Fig. 11B). This factorial stimulus set would make it possible to identify pure simple cells, pure boundary cells, as well as cells of intermediate type.

#### Diverse inputs to the dendrites of boundary cells?

One of our main findings is that a cell in visual cortex whose job is to detect object boundaries can improve its detection performance if it collects input from many simple cells in its vicinity with a diversity of receptive field positions and orientations.

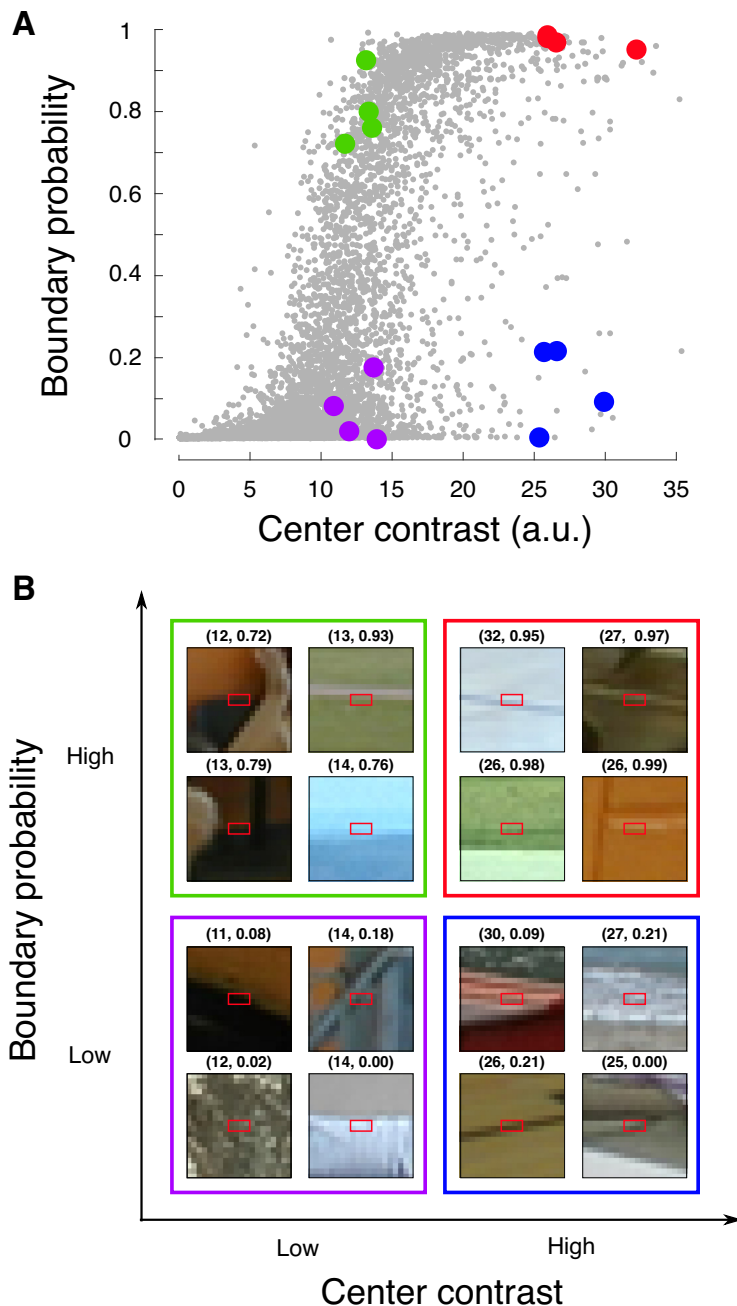


**Figure 10.** Boundary cell responses to parametric and natural stimuli resemble simple cell responses. To compute BC responses, the weighted sum of LLRs model (Fig. 9, green PR curve) was used. **A**, Spike-triggered average stimulus computed by averaging natural image stimuli weighted by their evoked boundary cell response. **B**, Response of a boundary cell to a grating presented at different phases. The boundary cell is like a simple cell in that it is sensitive to polarity, responding to only half of all phases. **C**, Orientation tuning curve to the same grating. At each orientation, responses were averaged over all phases of the grating. The resulting tuning curve is similar to those obtained for simple cells in V1. **D**, Patches with fixed surround contrast (normalizer value) and varying center contrast were selected and presented at 15° increments to the boundary cell. For a fixed surround contrast, center contrast increases cell response without increasing tuning width, a hallmark of contrast invariant orientation tuning found in V1 simple cells [full-width at half-height for high-contrast stimuli (red curve), 43.6°; full-width at half-height for low-contrast stimuli (blue curve), 39.2°].

Consistent with this, several recent two-photon calcium imaging studies have mapped the receptive field properties of individual dendritic spines on V1 neurons in mice, ferrets, and monkeys, and have shown that the inputs to a single V1 cell (and often a single dendrite) are quite variable in terms of their receptive field properties, covering a much wider range of preferred orientations and RF positions than might be expected given the more sharply tuned response preferences of a target cell (Jia et al., 2010; Wilson et al., 2016; Iacaruso et al., 2017; Scholl et al., 2017; Ju et al., 2020). These findings do not prove that many or most cells in V1 are boundary cells, only that most V1 cells appear to receive the requisite diversity of inputs from neighboring cells. What remains to be shown is that the excitatory and inhibitory inputs from surrounding cells are properly weighted and balanced by the local incitation circuit, so as to maximize boundary detection performance. The conceptual experiment described next could help to establish whether the local cortical circuit actually functions in this way.

#### A predictable spectrum of SC–BC interactions?

A key feature of the boundary cell hypothesis is that SC–BC interaction functions take on predictable forms, depending primarily on the offsets in position and orientation between the SC and BC receptive fields (Fig. 7B). It may be possible to empirically measure those interaction functions by applying stimuli that distinguish



**Figure 11.** Distinguishing linear filter responses from boundary probability responses. To determine whether a given cell is computing linear contrast or boundary probability, it is necessary to use a stimulus set that dissociates these two measures. Generally speaking, what is needed are stimuli whose linear filter and boundary probability scores are “well spread” throughout linear filter–boundary probability space. **A**, Plotting the two scores for all labeled patches shows that they are highly correlated, and that randomly selected patches are likely to lie at the bottom left and top right corners of this space, where linear contrast and boundary probability are either both low or high together. Therefore, if only these stimuli were presented to the cell, it would be difficult to know whether high cell responses were being driven by linear contrast or boundary probability. It would therefore be preferable to present stimuli that are well spread over the space of the two scores (colored dots) so that cell responses to each variable can be assessed separately. **B**, Examples of these stimuli are shown. They include low-contrast nonedges (purple cases), high-contrast nonedges (blue cases), low-contrast edges (green cases), and high-contrast edges (red cases).

simple cells from boundary cells (Fig. 11) while imaging V1 neurons in awake animals (Tang et al., 2018; Ju et al., 2020). The approach would require the ability (1) to identify pairs of nearby simple and boundary cells and to characterize their RFs, and (2) to transiently activate or inactivate identified cells optogenetically. Then, while presenting carefully curated natural image patches to

the boundary cell, which also overlap with the RF of the simple cell, the activity level of the simple cell could be perturbed, and the response changes in the boundary cell measured. The direction and magnitude of the change in the activity of the boundary cell could be compared with the prediction of a trained incitation network (Fig. 7B). For a particular simple cell, certain image patches will drive the cell to a level on the rising slope of its SC–BC interaction function, so that a boost in the cell activity (through optogenetic stimulation) should in turn boost the boundary cell activity—and conversely for suppression of the simple cell activity. For other image patches, the simple cell will be firing at or beyond the mode of its SC–BC interaction function with respect to a particular boundary cell, so that a boost in the activity of that simple cell will lead to a suppression of the activity of that boundary cell (and conversely if the activity of the simple cell is optogenetically suppressed).

It is worth noting that we cannot assume that the SC–BC interaction functions measured in this way will look exactly like those produced by a particular trained incitation network, since even if the overall idea holds true, the interaction functions depend on the complete set of simple cells providing input to a particular boundary cell, which cannot be known. Nonetheless, by repeating these types of response manipulations for a large number of SC–BC pairs, we can hope to find a basic correspondence between predicted and measured SC–BC interactions, in the sense that the measured interactions should include pure increasing cases, pure decreasing cases, and nonmonotonic cases, with a systematic dependence on the spatial and orientation offsets of the RFs of simple and boundary cells (Figs. 6, 7).

It is also worth noting that boundary cells need not reside in, or only in, V1. Nothing precludes that cells that signal boundary probability, rather than boundary contrast, may be found in higher visual areas.

### Relationship to previous work on natural image statistics

A number of previous studies have attempted to explain receptive field properties of cells in the retina, LGN, and primary visual cortex in terms of natural image statistics and principles such as efficient coding, sparse coding, and independent components analysis (Barlow, 1981; Laughlin, 1989; Bell and Sejnowski, 1995; Olshausen and Field, 1996; Schwartz and Simoncelli, 2001; Zhu and Rozell, 2013). These studies have been mainly concerned with neural representation, where the goal is fast/accurate information transmission through a noisy channel, and eventually faithful image reconstruction. In contrast, our work is primarily concerned with neural computation, where the goal is to transform

the goal is fast/accurate information transmission through a noisy channel, and eventually faithful image reconstruction. In contrast, our work is primarily concerned with neural computation, where the goal is to transform

the image into a more abstract shape representation that is more directly useful for guiding behavior.

From a different perspective and with a different goal, Geisler et al. (2001) collected co-occurrence statistics of pre-detected local boundary elements in natural scenes, with the aim to predict human contour grouping performance. Their measurements on natural images included the probability of finding a second boundary element in the vicinity of a first boundary element, depending on the relative offsets in the position and orientation of the two elements, or whether two spatially offset boundary elements were more likely to belong to the same or different object. Sigman et al. (2001) also studied co-occurrence statistics of pre-detected boundary elements, coming to the conclusion that boundary elements in natural scenes tend to lie on common circles. The goal to characterize the spatial distribution of pre-detected boundary elements in natural scenes in both of these studies contrasts with our focus here on the detection problem, that is, the problem of discriminating object boundaries from non-boundaries based on populations of simple cell responses collected from a local neighborhood in an image. Furthermore, all of the grouping statistics collected by Geisler et al. (2001) and Sigman et al. (2001) were represented as scalar values linking pairs of locations/orientations. In contrast, our natural image analysis produces functions linking pairs of locations/orientations, which capture how a given simple cell should influence a nearby boundary cell as a part of a boundary detection computation. Also unlike these previous studies, we use our data to constrain and to benchmark cortical circuit models.

### Nonmonotonic cell–cell interactions have been previously reported

One of our findings is that among the different types of local cell–cell interactions needed for object boundary detection in natural images, many cannot be described as “excitatory” or “inhibitory,” or represented by positive or negative synaptic weights, but are instead U-shaped functions wherein cell 1 might excite cell 2 at low firing rates, reach its peak excitatory effect at intermediate firing rates, and inhibit cell 2 at high firing rates. U-shaped functions of the opposite polarity can also occur (Fig. 7B). Should we find the idea surprising that nearby cells in the cortex act on each other nonmonotonically?

From one perspective, one might argue that whenever there are excitatory and inhibitory cells wired together in a circuit motif, perhaps we should be surprised if we did not find nonmonotonic interactions between cells. For example, in the “inhibition-stabilized network” model (Ozeki et al., 2009; Jädi and Sejnowski, 2014), which accounts for a number of V1 cell response properties, “nonbinary” interactions between cells would almost certainly be expected to occur. Nevertheless, there has been a historical tendency to think about cell–cell interactions in the cortex as being of a defined polarity, represented by a positive or negative scalar value, and often subject to simple geometric rules. The notion of “surround suppression,” for example, reflects both of these tendencies (Cavanaugh et al., 2002; Schwabe et al., 2010; Adesnik et al., 2012). Even as the geometric constraints governing cell–cell interactions become more intricate, such as where interconnection strength and polarity depend on distance or relative orientation, the simplification that cell–cell interactions have a defined polarity is often still relied on. For example, the models of Miller (1994) map the development of short-range excitation and medium-range inhibition; the models of Angelucci and Bressloff (2006) include near and far suppressive surrounds; and several studies support the idea that cortical cells affect each other laterally

through bowtie-shaped “extension fields” consisting of patterned arrays of positive and negative coefficients (Field et al., 1993; Bosking et al., 1997; Li, 1999; Kapadia et al., 2000; Geisler et al., 2001; Sigman et al., 2001). In all of these cases, the effect of one neuron on another neuron is described in terms of its scalar connection “strength.”

Not all functional interconnections that have been described in the cortex fit such descriptions, however. Examples of activity level-dependent interactions have been reported, where the strength and even the polarity of the connection between cells depends on the activity levels of the sending and/or receiving cells. For example, the responses of amplitude-tuned neurons in the auditory cortex grow stronger as the sound pressure level increases up to an optimal intensity level, and then are progressively inhibited as the sound grows louder (Suga and Manabe, 1982); in V1, surround modulation can switch from facilitating to suppressive with increasing center contrast (Polat et al., 1998; Somers et al., 1998; Schwabe et al., 2006; Ichida et al., 2007; Nauhaus et al., 2009); length-tuned neurons respond best to an oriented stimulus up to a certain length, but are then progressively inhibited as the stimulus grows longer (Anderson et al., 2001); and nonmonotonic modulatory interactions between the classical and extraclassical receptive fields of a neuron have been reported (Polat et al., 1998). These data, though unaccompanied by normative explanations, do support the idea that the sign and magnitude of the effect of one neuron on another can depend not only on the relative position and orientation of their receptive fields (in the case of vision), but also on their relative activity levels.

Our article represents a fleshing out of this type of effect and is to our knowledge the first normative theory, parameterized by natural images, that specifies how intracolumnar cell–cell interactions may help solve a specific, biologically relevant classification problem. By analyzing natural image data on and off object boundaries, we showed that the local cell–cell interactions needed to solve this classification problem are not capturable by scalar weights, but are in general nonlinear functions that depend on “all of the above”—relative location, relative orientation, and relative activity levels of the sending and receiving cells. We further showed that the SC–BC functional connections needed for boundary detection are easily produced by a compound E–I circuit motif (Fig. 6) that is known to exist in the cortex (Buzsáki, 1984; McBain and Fisahn, 2001; Pouille and Scanziani, 2001; Swadlow, 2002; Wehr and Zador, 2003; Klyachko and Stevens, 2006; Isaacson and Scanziani, 2011; Pfeffer et al., 2013). Finally, we showed that the synaptic weights that control the net effect of an incitation motif are easily learned. Future experiments will be needed to establish whether trainable incitation circuits are actually used to help solve the difficult natural classification problems faced by neurons in V1 and other areas of the cortex.

### How could a properly parameterized incitation circuit develop?

A possible extension of this work would be to address the limitation that the incitation circuit we show in Figure 7A was trained by a supervised learning rule (logistic regression), but without our providing a biologically based account for the source of the supervision. The original purpose of the exercise was to test whether an incitation circuit with a single layer of modifiable excitatory weights is capable of performing object boundary detection at a level comparable to that of an explicit Bayesian classifier. We found this to be true (Fig. 9), suggesting that this particular Bayesian-inspired algorithm lies within the computational scope of



cortical tissue. The demonstration leaves open the question, however, as to where a supervisory signal might come from during visual development that alerts a boundary cell and its inhibitory partner when an object boundary is crossing through its receptive field. One possible source of supervision would be a population of neurons located within the same or different area that have access to different visual cues, such as cells sensitive to motion-defined boundaries. Such cells are found at many levels of the visual system, including the retinas of salamanders (Olveczky et al., 2003); V1, V2, V3, middle temporal cortex, and inferior temporal cortex in the monkey (Marcar et al., 1995, 2000; Sáry et al., 1995; Zeki et al., 2003); and in multiple areas of the human visual cortex (DuPont et al., 1997; Zeki et al., 2003; Mysore et al., 2006; Larsson et al., 2010). Topographic feedback projections from motion boundary-sensitive cells in these areas to V1 (or locally within V1) could help to instruct boundary cells in V1 so that they may perform well based purely on pictorial cues (i.e., when motion signals are unavailable).

### Limitations of the model

The boundary detection computation that we have studied was inspired by Bayes’ rule and is essentially a feedforward computation whose core operation is a sum of LLR terms (Fig. 1C). Our attempt to map this computation onto a simple, cortically plausible circuit is shown in Figure 7A, in which a layer of simple cells with varying output nonlinearities activates both (1) a “layer” of boundary cells (though only one BC is shown); and (2) a layer of inhibitory cells, one per BC (though only one inhibitory cell is shown—the one assigned to the one shown boundary cell). Each inhibitory cell, in turn, acts on its associated boundary cells through a fixed connection. Given that the circuit of Figure 7A is purely feedforward, ignoring (1) local or long-range feedback connections that are known to exist in the neocortex (Angelucci et al., 2017), (2) nonlinear dendritic integration effects that could also contribute to boundary detection (Jin et al., 2022), and (3) all dynamics at the synapse, cell, and circuit levels, it falls far short of a fully elaborated cortical circuit model. Rather, the circuit model of Figure 7A should be viewed as a demonstration that a known cortical circuit motif—the incitation motif—is capable of producing cells that superficially resemble simple cells, but are much better at detecting object boundaries in natural scenes than the standard simple cell model (Heeger, 1992). A worthy long-range goal would be to fold the boundary detection capability of a properly parameterized incitation circuit into a more comprehensive cortical circuit model that addresses a wider range of physiological phenomena (Ozeki et al., 2009; Zhu and Rozell, 2013).

### Appendix 1: logistic regression learns LLRs assuming CCI

We are interested in estimating the probability of some event  $y$ , in this case whether a patch contains a boundary, from input data  $\vec{x}$ , in this case the responses of several simple cells. Logistic regression builds a model of  $p(y|\vec{x})$  by computing (potentially nonlinear) features  $\Phi(\vec{x})$ , and assuming that the output probability is a sigmoid function of a linear combination of these features:

$$\hat{p}(y|\vec{x}) = \sigma(w^T \vec{x}) = \frac{1}{1 + e^{-w^T \Phi(\vec{x})}}.$$

The goal of learning is to pick weights  $w$  that minimize the expected cross-entropy between the true and model probabilities, as follows:

$$\begin{aligned} w^* &= \operatorname{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \hat{p}(y|\vec{x}), \\ &= \operatorname{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \frac{\hat{p}(y|\vec{x})}{p(y|\vec{x})} \\ &\quad - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln p(y|\vec{x}). \end{aligned}$$

The first term is the KL divergence between the true and model distributions, and the second term is a constant with respect to the weights, and can be ignored. The optimization is then:

$$w^* = \operatorname{argmin}_w D_{KL}(p(y, \vec{x}) || \hat{p}(y, \vec{x})).$$

This is minimized when the model distribution  $\hat{p}$  matches the true distribution  $p$ . To see that under the assumption of class conditional independence, learning the LLR functions achieves this minimum, observe the following:

$$\hat{p}(y, \vec{x}) = p(\vec{x}) \hat{p}(y|\vec{x}) = p(\vec{x}) \sigma(f(\vec{x})),$$

where  $f(\vec{x}) = w^T \Phi(\vec{x})$ . Further, class conditional independence implies the following:

$$\begin{aligned} p(y|\vec{x}) &= \frac{p(y)p(\vec{x}|y)}{p(y)p(\vec{x}|y) + p(\bar{y})p(\vec{x}|\bar{y})} = \frac{1}{1 + \frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}} \\ &= \sigma\left(-\log \frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}\right), \\ &= \sigma\left(-\log \frac{p(\bar{y})}{p(y)} + \sum_i \log \frac{p(x_i|y)}{p(x_i|\bar{y})}\right) \\ &= \sigma\left(-\log \frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)\right), \end{aligned}$$

so that the objective can be written as follows:

$$w^* = \operatorname{argmin}_w D_{KL} \left( p(\vec{x}) \sigma\left(-\log \frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)\right) || p(\vec{x}) \sigma(f(\vec{x})) \right).$$

One can see by inspection that the two distributions will be equal and the objective will be minimized to exactly 0 if and only if:

$$f(\vec{x}) = -\log \frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i),$$

that is, when the classifier simply combines the filter values by passing them through their LLR functions and summing the result. Furthermore, by convexity, all local minima achieve the global minimum cost which, assuming the features  $\Phi$  are diverse

enough to represent the LLR functions, is exactly 0. But we have just shown that 0 cost is possible only if  $f$  is of the form above. Thus, any learning rule for  $w$  that eventually converges to a local minimum – including gradient descent – must learn the LLR functions.

## References

- Adesnik H, Bruns W, Taniguchi H, Huang ZJ, Scanziani M (2012) A neural circuit for spatial summation in visual cortex. *Nature* 490:226–231.
- Alitto HJ, Usrey WM (2004) Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. *J Neurophysiol* 91:2797–2808.
- Anderson JS, Lampl I, Gillespie DC, Ferster D (2001) Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. *J Neurosci* 21:2104–2112.
- Angelucci A, Bressloff PC (2006) Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog Brain Res* 154:93–120.
- Angelucci A, Levitt JB, Walton EJS, Hupé J-M, Bullier J, Lund JS (2002) Circuits for local and global signal integration in primary visual cortex. *J Neurosci* 22:8633–8646.
- Angelucci A, Bijanzadeh M, Nurminen L, Federer F, Merlin S, Bressloff PC (2017) Circuits and mechanisms for surround modulation in visual cortex. *Annu Rev Neurosci* 40:425–451.
- Arbeláez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 33:898–916.
- Barlow HB (1981) The Ferrier Lecture, 1980: critical limiting factors in the design of the eye and visual cortex. *Proc R Soc Lond B Biol Sci* 212:1–34.
- Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7:1129–1159.
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147.
- Bishop PO, Coombs JS, Henry GH (1973) Receptive fields of simple cells in the cat striate cortex. *J Physiol* 231:31–60.
- Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci* 17:2112–2127.
- Buzsáki G (1984) Feed-forward inhibition in the hippocampal formation. *Prog Neurobiol* 22:131–153.
- Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62.
- Cavanaugh JR, Bair W, Movshon JA (2002) Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J Neurophysiol* 88:2547–2556.
- DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. *J Neurophysiol* 68:144–163.
- Dupont P, De Bruyn B, Vandenbergh R, Rosier AM, Michiels J, Marchal G, Mortelmans L, Orban GA (1997) The kinetic occipital region in human visual cortex. *Cereb Cortex* 7:283–292.
- Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local “association field”. *Vision Res* 33:173–193.
- Fine I, MacLeod DIA, Boynton GM (2003) Surface segmentation based on the luminance and color statistics of natural scenes. *J Opt Soc Am A Opt Image Sci Vis* 20:1283–1291.
- Fukushima K, Miyake S, Ito T (1983) Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Trans Syst Man Cybern SMC* 13:826–834.
- Geisler WS, Albrecht DG (1992) Cortical neurons: isolation of contrast gain control. *Vision Res* 32:1409–1410.
- Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res* 41:711–724.
- George AA, Lyons-Warren AM, Ma X, Carlson BA (2011) A diversity of synaptic filters are created by temporal summation of excitation and inhibition. *J Neurosci* 31:14721–14734.
- Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res* 30:1689–1701.
- Grosfeld DH, Shapley RM, Hawken MJ (1993) Macaque V1 neurons can signal “illusory” contours. *Nature* 365:550–552.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197.
- Heydt R, von der Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neurosci* 9:1731–1748.
- Hoffman DD (2000) Visual intelligence: how we create what we see. New York: WW Norton.
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol* 160:106–154.
- Iacaruso MF, Gasler IT, Hofer SB (2017) Synaptic organization of visual space in primary visual cortex. *Nature* 547:449–452.
- Ichida JM, Schwabe L, Bressloff PC, Angelucci A (2007) Response facilitation from the “suppressive” receptive field surround of macaque V1 neurons. *J Neurophysiol* 98:2168–2181.
- Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron* 72:231–243.
- Jadi MP, Sejnowski TJ (2014) Regulating cortical oscillations in an inhibition-stabilized network. *Proc IEEE* 102:830–842.
- Jia H, Rochefort NL, Chen X, Konnerth A (2010) Dendritic organization of sensory input to cortical neurons in vivo. *Nature* 464:1307–1312.
- Jin L, Behabadi BF, Jadi MP, Ramachandra CA, Mel BW (2022) Classical-contextual interactions in V1 may rely on dendritic computations. *Neuroscience* 489:234–250.
- Johnson EN, Hawken MJ, Shapley R (2008) The orientation selectivity of color-responsive neurons in macaque V1. *J Neurosci* 28:8096–8106.
- Ju N, Li Y, Liu F, Jiang H, Macknik SL, Martinez-Conde S, Tang S (2020) Spatiotemporal functional organization of excitatory synaptic inputs onto macaque V1 neurons. *Nat Commun* 11:697.
- Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–856.
- Kapadia MK, Westheimer G, Gilbert CD (2000) Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J Neurophysiol* 84:2048–2062.
- Karklin Y, Lewicki MS (2003) Learning higher-order structures in natural images. *Network* 14:483–499.
- Karklin Y, Lewicki MS (2005) A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Comput* 17:397–423.
- Klyachko VA, Stevens CF (2006) Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. *PLoS Biol* 4:e207.
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, Vol 25 (Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds), pp 1097–1105. Red Hook, NJ: Curran Associates.
- Lades M, Vorbruggen JC, Buhmann J, Lange J, Malsburg C, von der Wurtz RP, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans Comput* 42:300–311.
- Larsson J, Heeger DJ, Landy MS (2010) Orientation selectivity of motion-boundary responses in human visual cortex. *J Neurophysiol* 104:2940–2950.
- Laughlin SB (1989) The role of sensory adaptation in the retina. *J Exp Biol* 146:39–62.
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324.
- Li Z (1999) Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Network* 10:187–212.
- Liang Y, Simoncelli EP, Lei Z (2000) Color channels decorrelation by ICA transformation in the wavelet domain for color texture analysis and synthesis. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*. CVPR 2000 (Cat. No. PR00662), Vol 1, pp 606–611.
- Marcar VL, Xiao DK, Raiguel SE, Maes H, Orban GA (1995) Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *J Neurophysiol* 74:1258–1270.
- Marcar VL, Raiguel SE, Xiao D, Orban GA (2000) Processing of kinetically defined boundaries in areas V1 and V2 of the macaque monkey. *J Neurophysiol* 84:2786–2798.
- McBain CJ, Fisahn A (2001) Interneurons unbound. *Nat Rev Neurosci* 2:11–23.

- Mel BW (1997) SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput* 9:777–804.
- Miller KD (1994) A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *J Neurosci* 14:409–441.
- Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat’s visual cortex. *J Physiol* 283:101–120.
- Murphy KP (2012) *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT.
- Mysore SG, Vogels R, Raiguel SE, Orban GA (2006) Processing of kinetic boundaries in macaque V4. *J Neurophysiol* 95:1864–1880.
- Nauhaus I, Busse L, Carandini M, Ringach DL (2009) Stimulus contrast modulates functional connectivity in visual cortex. *Nat Neurosci* 12:70–76.
- Ohzawa I, DeAngelis GC, Freeman RD (1997) Encoding of binocular disparity by complex cells in the cat’s visual cortex. *J Neurophysiol* 77:2879–2909.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
- Olveczky BP, Baccus SA, Meister M (2003) Segregation of object and background motion in the retina. *Nature* 423:401–408.
- Ozeki H, Finn IM, Schaffer ES, Miller KD, Ferster D (2009) Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* 62:578–592.
- Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat Neurosci* 16:1068–1076.
- Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM (1998) Collinear stimuli regulate visual responses depending on cell’s contrast threshold. *Nature* 391:580–584.
- Pouille F, Scanziani M (2001) Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* 293:1159–1163.
- Priebe NJ, Ferster D (2006) Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat Neurosci* 9:552–561.
- Ramachandra CA, Mel BW (2013) Computing local edge probability in natural scenes from a population of oriented simple cells. *J Vis* 13(14):19, 1–22.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Ringach DL (2002) Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol* 88:455–463.
- Rust NC, Dicarlo JJ (2010) Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30:12978–12995.
- Sanguinetti G, Citti G, Sarti A (2010) A model of natural image edge co-occurrence in the rotation group. *J Vis* 10(14):37, 1–16.
- Sáry G, Vogels R, Kovács G, Orban GA (1995) Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. *J Neurophysiol* 73:1341–1354.
- Sceniak MP, Ringach DL, Hawken MJ, Shapley R (1999) Contrast’s effect on spatial summation by macaque V1 neurons. *Nat Neurosci* 2:733–739.
- Scholl B, Wilson DE, Fitzpatrick D (2017) Local order within global disorder: synaptic architecture of visual space. *Neuron* 96:1127–1138.e4.
- Schwabe L, Obermayer K, Angelucci A, Bressloff PC (2006) The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. *J Neurosci* 26:9117–9129.
- Schwabe L, Ichida JM, Shushruth S, Mangapathy P, Angelucci A (2010) Contrast-dependence of surround suppression in macaque V1: experimental testing of a recurrent network model. *Neuroimage* 52:777–792.
- Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci* 4:819–825.
- Sigman M, Cecchi GA, Gilbert CD, Magnasco MO (2001) On a common circle: natural scenes and Gestalt rules. *Proc Natl Acad Sci U S A* 98:1935–1940.
- 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* 8:204–217.
- Suga N, Manabe T (1982) Neural basis of amplitude-spectrum representation in auditory cortex of the mustached bat. *J Neurophysiol* 47:225–255.
- Swadlow HA (2002) Thalamocortical control of feed-forward inhibition in awake somatosensory “barrel” cortex. *Philos Trans R Soc Lond B Biol Sci* 357:1717–1727.
- Tang S, Zhang Y, Li Z, Li M, Liu F, Jiang H, Lee TS (2018) Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *eLife* 7:e33370.
- Theys T, Romero MC, van Loon J, Janssen P (2015) Shape representations in the primate dorsal visual stream. *Front Comput Neurosci* 9:43.
- Wainwright MJ, Simoncelli E (2000) Scale Mixtures of Gaussians and the Statistics of Natural Images. *Adv Neural Inf Process Syst* 12:855–861.
- Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426:442–446.
- Wilson DE, Whitney DE, Scholl B, Fitzpatrick D (2016) Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. *Nat Neurosci* 19:1003–1009.
- Zeki S, Perry RJ, Bartels A (2003) The processing of kinetic contours in the brain. *Cereb Cortex* 13:189–202.
- Zetsche C, Röhrbein F (2001) Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network* 12:331–350.
- Zhou C, Mel BW (2008) Cue combination and color edge detection in natural scenes. *J Vis* 8:4.1–25.
- Zhu M, Rozell CJ (2013) Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS Comput Biol* 9:e1003191.