Title: Detecting object boundaries in natural images requires 'incitatory' cell-cell interactions

Authors: Gabriel C. Mel[1*], Chaithanya A. Ramachandra[2], and Bartlett W. Mel[2,3]

Affiliations:

[1]Department of Computer Science, University of Southern California, Los Angeles, CA.

[2]Eyenuk, Inc, Los Angeles, CA, USA.

[3]Neuroscience Graduate Program, University of Southern California, Los Angeles, CA.

*Correspondence to: meldefon@stanford.edu

Abstract:

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 determine how a simple cell should influence a nearby boundary cell depending on its relative offset in receptive field position and orientation. We identified three basic types of cell-cell interactions: rising and falling interactions with a range of slopes and saturation rates, as well as non-monotonic (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 simple (1-layer) learning rule. We conclude that incitatory interconnections are a generally useful computing mechanism that the cortex may exploit to help solve difficult natural classification problems.

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, multi-purpose 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 longstanding hypothesis is that orientation-tuned V1 cells somehow participate in object boundary detection, a core process in biological vision (Biederman, 1987; Gilbert and Wiesel, 1990; Heydt and Peterhans, 1989; Hubel and Wiesel, 1962a; Kapadia et al., 1995a) 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 due to 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 non-boundaries.

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 (Sanguinetti et al., 2010; Sigman et al., 2001), and how these arrangements relate to human contour grouping performance (W. S. Geisler, Perry, Super, & Gallogly, 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 (Figure 1A), we began with a known cell type – orientation-tuned "simple cells" (as defined by Hubel & Wiesel, 1962, and typically modeled as divisively normalized oriented linear filters (Carandini and Heeger, 2012) – and asked how the outputs of a population of simple cells (SCs), whose receptive fields (RFs) densely cover an area of the visual field, should be combined to produce a "boundary cell" (BC) whose firing rate represents the probability that an object boundary is present within its RF (Figure 1B). 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" (see methods for a detailed discussion), simple cell-boundary cell interactions are captured by the log-likelihood ratios (LLRs) embedded in Bayes' rule (colored expressions in Figure 1C), which represent the evidence that a given simple cell provides about the presence of an object boundary within a neighboring BC's receptive field (Figure 1D). We found that SC-BC interactions were diverse, and in some cases involved compound excitatory and inhibitory 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 LLR functions for a population of 300 simple cells neighboring a "reference location" where a boundary might be detected. We find that the simple cell-boundary cell interactions implied by the LLRs 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 data set. Finally, we demonstrate that a simple (1-layer) supervised learning rule is capable of setting the parameters of the incitation circuit that are needed to produce boundary cells from conventional simple cells.

Materials and methods:

Image preprocessing

As in Ramachandra and Mel (2012), 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 20x20 pixel image patches for labelling. The "reference location" representing a hypothetical boundary cell's receptive field location was defined as the elongated, horizontal 2x4 pixel region at the center of the patch (dashed box, Figure 1A, B). The original color image patches were converted to single-channel (monochrome) intensity images $(0.29\,R + 0.59\,G + 0.11\,B)$. Simple cell-like oriented "filters" were created by rotating a 2x4 pixel

horizontal filter kernel (see f_1 in Figure 1B) in 15° increments from 0 to 165° (i.e. 12 orientations). Computing the filter coefficients for rotated filters required sampling the horizontal filter kernel in a pattern rotated off the original pixel-aligned grid; the sampling of interstitial pixel values was done using bilinear interpolation (https://en.wikipedia.org/wiki/Bilinear_interpolation). Filtering consisted of computing the dot product between the filter kernel and the underlying image pixels. The monochrome image patches were filtered at 12 orientations (as indicated above) at 25 positions on a 5x5 pixel lattice centered at the reference location. This resulted in 300 filter values representing the firing of 300 simple cells covering the central neighborhood of each patch at all orientations. (Given the symmetric form of the filter kernel, filter values for orientations from 180° to 345° were simply the negatives of the responses at the first 12 orientations). Note that the coverage of the 300 filters was not limited to the 5x5 block of pixels at the center of the image patch, since, depending on its center position and orientation, a 4x2 filter could, and in in most cases did, extend slightly into the surrounding region.

To minimize filter pairwise correlations, data collection was restricted to patches with a certain fixed "normalizer" value, calculated by summing the absolute values of a fixed subset of 100 simple cell responses surrounding the reference location (see normalizer_filter_subs.mat in supplementary materials). The normalizer bin used for all of the analysis reported below was 200±40. The selection of the 100 filters used for the normalizer was the result of an ad-hoc procedure whose goal was to jointly minimize (1) correlations between absolute filter scores in (normalized) image patches, and (2) the number of filter values that needed to be computed. To verify that the particular choice of filters used for the normalizer did not affect our main results, in a control experiment we renormalized the image patches based on a generic normalizer formed by the sum of the absolute values of all 300 filters

covering the 5x5 central image patch. The results derived from the renormalized patches were very similar, and supported all the same conclusions. The results presented below are therefore limited to the 100-filter version of the normalizer.

Data gathering

A horizontal 2x4 pixel rectangular box was drawn around the reference location and human labelers were asked to answer the question, “On a scale from 1 to 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, boundary segments had to be visible and unoccluded throughout the box. We restricted labelling to horizontal boundaries since pixel lattice discretization made it more difficult to judge oblique orientations, and because we expected filter statistics to be roughly orientation invariant. (This expectation was supported by subsequent tests showing that LLRs 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. Our informal observations, based in part on occasions when two labelers worked together, was that the agreement was very high. Rare ambiguous patches that could cause labeler disagreement were often conservatively labelled as 3 and consequently excluded from later 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.

Bayesian formalism

We assume that a boundary cell computes $p(\, yes \mid f_1\, f_2\, f_3 \cdots)$ (or some other monotonically related quantity). Using Bayes' rule we obtain

$$p(\, yes \mid f_1\, f_2\, f_3 \cdots) = \frac{p(\, f_1\, f_2\, f_3 \cdots \mid yes)\, p(yes)}{p(f_1\, f_2\, f_3 \cdots \mid yes) p(yes) + p(f_1\, f_2\, f_3 \cdots \mid no) p(no)}$$

Dividing through by the numerator and rearranging, we find

$$p(\, yes \mid f_1\, f_2\, f_3 \cdots) = \frac{1}{1 + \frac{p(no)}{p(yes)} exp\left(-\log \frac{p(f_1\, f_2\, f_3 \cdots \mid yes)}{p(f_1\, f_2\, f_3 \cdots \mid no)}\right)}$$

Assuming class condition independence (see below for a discussion of this assumption), the joint log likelihoods split into a sum of individual filter log likelihoods:

$$p(\, yes \mid f_1\, f_2\, f_3 \cdots) = \frac{1}{1 + \frac{p(no)}{p(yes)} exp\left(-\sum_i \log \frac{p(f_i \mid yes)}{p(f_i \mid no)}\right)}$$

In intuitive terms, this equation says that to compute boundary probability at a particular location, each neighboring simple cell's response should be passed through a log likelihood ratio (LLR) function, $\log \frac{p(f_i \mid yes)}{p(f_i \mid no)}$, the results should be summed, and the sum should be passed through a fixed sigmoidal nonlinearity, $\sigma(x) = \frac{1}{1+\frac{p(no)}{p(yes)} exp(-x)}$.

Extracting the LLRs

Histograms were collected of each of the 300 filter responses separately for "yes" patches (using 8 to 20 evenly spaced bins depending on the smoothness of the histogram) and "no" patches (using 50 evenly spaced bins). "Yes" histograms were binned more coarsely because our dataset had many fewer yes patches than no patches. The yes and no histograms for each filter were then normalized to

probability distributions. LLRs were computed as $\log \frac{\mathrm{p}(f|yes)}{\mathrm{p}(f|no)}$, where $p(f|yes)$ and $p(f|no)$ are the boundary and non-boundary pdfs evaluated at the filter response f, respectively. To control noise, for each filter, LLR analysis was restricted to a central set of filter values where $p(f|yes) > 0.005$ and $p(f|no) > 0.002$. (Different thresholds were used because smaller probabilities could be estimated more reliably for the much larger "no" patch set). Only data inside this region is plotted in Figs. 2-4 and Figure 6. The same procedure was repeated using different filter profiles (2x6, 2x8, 4x8, and 6x8 pixels) to generate the LLR curves shown in Figure 3.

Class-conditional independence (CCI)

To make the analysis tractable, we assumed class-conditional independence (CCI) between nearby filters, formally $p(f_1\, f_2 \cdots |\, yes) = p(f_1 \mid yes)p(\, f_2 |\, yes) \cdots$ and $p(f_1\, f_2 \cdots |\, no) = p(f_1 \mid no)p(\, f_2 |\, no) \cdots$. This assumption does not hold in general in natural images, so that the strict application of a naive Bayesian approach that assumes CCI among filters is not expected to perform very well. Indeed, in our experiments, if filters are chosen randomly the classifier performs poorly – often worse than using a single filter at the reference location. However, either of two simple strategies, both biologically plausible, can mitigate the problems arising from the violation of CCI. The first strategy is to explicitly select small subsets of cells from the neighborhood whose responses are mostly uncorrelated with each other; this was the approach taken in Ramachandra and Mel (2013). In that paper, in addition to selecting for decorrelation, we enforced two ad hoc constraints on the choice of filters, namely that (1) the filters should be individually informative about the presence of an object boundary, and (2) the resulting classifier should be sharply oriented tuned. The second strategy, and

the one that we pursue in this paper, is to use a learning rule to modify the synaptic weights in an incitatory circuit, without the need for any purpose-driven filter pre-selection (Figure 7A).

Modeling log likelihood ratios (LLRs) as differences of sigmoids

Each LLR can be thought of as a function of its filter value f. We fit each of the LLR functions by a difference of 2 sigmoids of the form $s(f_i) = \frac{A}{1+\exp[-g(f_i-t)]}$, where f_i was the filter's response (plotted on the x-axis in all LLR plots). For each LLR, an approximate amplitude A, gain g, and threshold t for the two sigmoids was chosen automatically, and then these 3 parameters were adjusted by hand so that the difference of the two sigmoids visually matched the LLR as closely as possible. We found visually-guided optimization better captured the essential shape structure of the LLR compared to conventional quantitative measures such as MSE. A similar fitting procedure was used for the three models in Figure 6 (model details shown in figure). The risk that human visually-guided optimization of curve shape would alter our conclusions was minimal since (1) human visually-guided optimization is based on a much more sophisticated shape-based metric than, say, MSE, and can therefore be reasonably considered as "ground truth"; (2) our conclusions do not depend on quantitative comparisons of fit quality for different models; and (3) the ability to precisely match individual LLR 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 drive down classification errors during learning (see Figure 7). For the surfaces in Figure 5E, and Figure 6, excitatory gain was computed by measuring the excitatory component's average slope between f = 0 and f = 10. Inhibitory gain in Figure 5 was computed in the same way. In Figure 6, each of the inhibition

198 families had only a single parameter; this parameter is what is plotted in the inhibitory gain surface
199 plots.
200

201 *Learning experiments*

202 For each patch, each filter value was passed through 8 different fixed sigmoid functions (using the
203 functional form given above). The 8 sigmoid functions were identical but for systematically increasing
204 thresholds (8 evenly spaced values from -6 to 35), simulating the responses of 8 simple cells with
205 identical receptive fields but slightly different output nonlinearities. The result was 300 filters x 8
206 nonlinear variants = 2,400 model simple cell responses per image patch. We then used logistic
207 regression to train a linear classifier to distinguish boundary from non-boundary patches using the
208 simple cell responses as inputs. A subset of the data (25,000 of the ~30,000 labeled patches) was used
209 for training. During training, data was balanced by duplicating boundary-containing patches such that
210 boundary and non-boundary exemplars were equal in number. Training was done using batch gradient
211 descent with a learning rate of $\eta = 0.1$, performed for 1000 iterations. The net effect of the 8 simple
212 cells sharing the same RF on the boundary probability was visualized in Figure 7B by systematically
213 increasing the underlying filter value (x axes) while holding all other filters constant, and plotting the
214 change in linear classifier score resulting from that filter's 8 nonlinear simple cells combined using their
215 learned weights (y axes). To facilitate comparison of the shape of each filters' learned net effect on the
216 boundary cell with that filter's explicit LLR, we scaled the colored interaction functions within each
217 plot. Each plot has one scaling factor that applies to all 5 colored curves in the plot. The inverse of this
218 scaling factor, which can be thought of as the weight that the classifier puts on the curves drawn in the
219 subplot, is shown by the grey bars.

Precision-Recall curves

Precision-Recall (PR) curves were generated for learned boundary cell classifiers, as well as for the naïve Bayes classifier (based on a sum of all filter LLRs; Figure 1BC) and other classifier variants (Figure 7). A classifier consisting of a single linear filter at the reference location provided the PR baseline (Figure 7C, blue curve). To generate a PR curve, a classifier was applied to each of the 5,000 labeled (untrained) test patches, and the patches were sorted by their 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 patches were called putative boundaries and below-threshold patches were called putative non-boundaries, and (1) "Precision" was calculated by asking what fraction of patches identified as putative boundaries contained true boundaries (according to their human assigned labels), and (2) "Recall" was calculated by asking what fraction of true boundaries were identified as putative boundaries. As the threshold increased, the P-R values swept out a curve in Precision-Recall space. Perfect performance would be 100% Precision and Recall simultaneously, corresponding to the top right corner of the PR graph.

Boundary cell stimulus responses

The idealized boundary image analogous to a spike-triggered average stimulus was computed by averaging all natural image patches weighted by their boundary cell response (Figure 9A). For Figure 9B,C, Grating stimuli were generated by sampling a sinusoidal grating wave on a 20x20 pixel grid. Frequency was chosen at 0.25 cycles/pixel because it led to relatively artifact-free stimuli and evoked

robust boundary cell responses. All stimuli were then scaled to have the fixed normalizer value used in the above LLR analysis – in keeping with the assumption that simple cell responses are divisively normalized. After creating each grating, the value (N) of the normalizer was computed on the grating patch, and the patch was divided by N/N_i, where $N_i = 2.0$ was the normalizer value used to collect the natural image patches. In this way, filters applied to the artificial grating patches could be run through the same set of LLRs as were collected from the natural image data set. Gratings were presented to the boundary cell at 15° steps in orientation. For Figure 9C responses were averaged over all phases of the grating at each orientation. Tuning curves in Figure 9D were obtained by presenting natural image stimuli with a fixed normalizer value. 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.

Results

To develop a more complete picture of the cell-cell interactions needed for natural boundary detection, in this study we collected and labeled 30,000 natural image patches, with scores ranging from 5, indicating high confidence that a boundary was present at a "reference location" (RL, indicated by a dashed box in Figure 1A), to 1, indicating high confidence that a boundary was *not* present at the RL. From these labeled patches, we histogrammed oriented linear filter values (representing simple cell responses) separately for "yes" (scores of 4-5) and "no" (scores of 1-2) cases (red and blue histograms in Figure 2A, respectively). From the responses of 300 neighboring simple cells at 12 orientations on a 5x5 pixel lattice centered on the RL, we computed the likelihoods $p(f_i \mid yes)$ and $p(f_i \mid no)$, meaning the probability of the ith filter having a particular response f_i when the patch does

263 (“yes”) or does not (“no”) contain a horizontal boundary. We show in Methods that, for a boundary cell
264 to compute the probability of a boundary, and contingent on the assumption that the different filter
265 responses are class conditionally independent (conditional on the patch being “yes” or “no”), the
266 boundary cell should have as its input the sum of the log likelihood ratios (LLRs), $\log \frac{p(f_i|yes)}{p(f_i|no)}$, of the
267 different simple cells. The boundary cell's output, representing the probability of a boundary, should
268 then be a certain sigmoidal function of this input (Figure 1C, D).
269
270 Accordingly, we computed the LLRs for all of the 300 simple cells. Examples of LLRs are shown in Figure
271 2B, and the full set is shown in Figure 2C grouped across 5 horizontal shifts at each orientation and
272 vertical position. The LLRs varied considerably with position and orientation relative to the RL, but
273 nonetheless conformed to a small number of qualitative shape prototypes (rising, falling, and bump-
274 shaped). When we generated LLRs for simple cells of different scales and shapes (2x6, 4x6, 4x8, and
275 6x8 pixel filters) we found a qualitatively similar pattern of results, indicating that the basic LLR shape
276 prototypes do not depend sensitively on the assumed sizes and shapes of SC receptive fields (Figure 3).
277
278 To gain insight into the forms of the LLRs we observed, we developed a simple mathematical model of
279 the process of LLR formation. If each filter’s yes and no distributions are approximated as gaussian with
280 a different mean and variance, the resulting LLR is quadratic. Since the no distribution is virtually
281 always wider than the yes distribution, i.e. has greater variance, the LLR will take the form of a
282 downward-pointing parabola (Figure 4A), qualitatively resembling the LLRs seen in Figure 2. The height
283 and width of the LLR is determined by the mean and variance of the yes and no distributions (Figure
284 4A, different colored curves). In addition to qualitatively capturing the range of observed LLR shapes,
285 this model has a simple interpretation in terms of natural image statistics: for each filter, responses to

286 non-boundaries vary widely about their mean, while responses to boundaries are clustered more
287 tightly around a characteristic value. The upside-down U-shaped LLRs simply reflect the fact that a
288 filter value provides the maximum positive evidence for a boundary when it is well centered within the
289 "yes" distribution.
290
291 To facilitate the interpretation of the LLRs as cell-cell interactions, we slightly reformatted the LLR
292 curves, in two ways. First, the LLRs were shifted vertically in order that they passed through the origin,
293 reflecting the idea that when a simple cell is not firing (corresponding to x=0 on the graph), its
294 influence on the boundary cell (y-value on the graph) should also be zero. This shift was justified given
295 that the LLRs would later be combined additively (Figure 1C, D), and thus the offsets across the entire
296 population of simple cells could be collapsed into a single net offset at the level of the boundary cell
297 (that would likely be small due to cancellation of positive and negative shifts. Second, simple cell firing
298 rates can only be positive, so the left half of each LLR function, corresponding to a negative simple cell
299 firing rate, was truncated. Information was not lost since the same or very similar function would be
300 covered by a different simple cell with the same RF but opposite contrast polarity. The right panel of
301 Figure 4A shows the combined effect of the shift and rectify operations. The full set of "shifted LLRs"
302 (hereafter sLLRs) obtained this way is shown in Figure 4B, with the plots corresponding to the
303 conceptual curves in A marked by asterisks.
304
305 Returning to the interpretation of sLLRs as simple cell-boundary cell interactions, for some simple cells
306 the sLLR increased monotonically from the origin, meaning that, as the simple cell's response increased
307 from zero, the evidence it provided to the boundary cell grew steadily more positive. This type of pure
308 "excitatory" SC-BC interaction was seen for simple cells that were most directly supportive of the

309 hypothesis that a boundary was present at the RL, such as the simple cell directly overlapping with the
310 RL (middle column in top row of Figure 4B). Referring to the model of Figure 4A, this was a case where
311 the downward-pointing "parabola" peaked far to the right of the origin, so that over the simple cell's
312 entire firing range, its effect on the boundary cell remained on the rising limb of the parabola (case 1 in
313 the right panel of Figure 4A). At even higher firing rates than are plotted in Figure 4B, the sLLR would
314 eventually reach its peak and turn back downward, but such high filter values were so rare in yes
315 patches in our natural image data set that the LLR curves could not be reliably estimated beyond the
316 range shown. Two other cases of pure excitatory sLLR are worth noting: the lower left and right
317 corners of Figure 4B. These cases apply to simple cells whose RFs are nearly "upside down" (i.e.
318 polarity reversed) versions of the reference filter profile, but shifted vertically either 2 pixels above or
319 below the reference "edge". The fact that these cells are monotonically supportive of the reference
320 hypothesis can be attributed to the existence of many 1-2 pixel wide light and dark horizontal bands in
321 our natural image data set.
322
323 For other simple cells, the sLLRs *decreased* monotonically from the origin, meaning that, as the simple
324 cell's response increased from zero, the evidence it provided to the boundary cell grew increasingly
325 more negative. This type of pure "inhibitory" SC-BC interaction was seen for simple cells whose firing
326 supported a hypothesis *incompatible* with the hypothesis that a boundary was present at the RL. The
327 clearest examples of such cells are those with RFs perpendicular to the RL (middle row, green LLR
328 curves. Referring again to the quadratic LLR model of Figure 4A, these monotonically decreasing sLLRs
329 arose from cases where the downward-pointing sLLR "parabola" was peaked at, or to the left of the
330 origin, so that over the entire positive response range of the simple cell, the sLLR fell continuously
331 along its descending limb (as in case 2 in Figure 4A).

For the majority of simple cells, however, the sLLR was bump-shaped, first rising and then falling as the simple cell's firing rate increased from zero. This type of "incitatory" SC-BC interaction was seen for cells that had sufficient overlap in position and orientation with the RL that they were most likely to be active at a middling level when a boundary was present at the RL. These cells provided increasing positive evidence for a boundary at the RL up to a moderate level of activity, but as their activity level increased further, they began to provide stronger evidence for a boundary *at their own RF location and orientation*, rather than at the RL.

A known circuit mechanism can produce the entire observed spectrum of sLLRs

Given that the input to a boundary cell should be the sum of the sLLRs associated with different simple cells, and that sLLRs can be either monotonic or nonmonotonic functions of the simple cell's response, we next asked what kind of neural interconnection circuit is capable of producing either monotonic or non-monotonic functions of the pre-synaptic cell's response. For monotonic excitatory and inhibitory interactions the answer is straightforward, but non-monotonic cell-cell interactions require a compound excitatory-inhibitory (E-I) interconnection scheme. One candidate mechanism is the ubiquitous circuit motif in which a cortical cell both directly excites and disynaptically inhibits other cells in its neighborhood (Buzsáki, 1984a; George et al., 2011; Isaacson and Scanziani, 2011a; Klyachko and Stevens, 2006a; McBain and Fisahn, 2001; Pfeffer et al., 2013a; Pouille and Scanziani, 2001a; Swadlow, 2002; Wehr and Zador, 2003) (Figure 5A, rightmost case). If the excitatory effect dominates at low firing rates and the inhibitory effect dominates at high firing rates, the neighbor cell's net effect on its target can be non-monotonic. When the circuit is simplified to consist of only the excitatory or

the inhibitory pathway, the cell-cell interaction reduces to conventional monotonic excitation or inhibition (Figure 5A, left and middle cases). To determine whether this circuit motif can produce the full range of cell-cell interactions contained in our data set, we assumed that both the direct excitatory and indirect inhibitory pathways exert a sigmoidal effect on the boundary cell, and therefore fit each LLR function with the difference of two sigmoid functions. Each of the sigmoids was allowed to vary in threshold, gain, and amplitude (Figure 5B). The fits are shown in Figure 5C, confirming that the range of cell-cell interactions needed to calculate boundary probability in natural images, including non-monotonic interactions, can be produced by a simple model of a circuit motif known to be present in V1 (see reference list just above). To determine whether the successful fitting of sLLRs depended on our particular choice of sigmoidal E and I basis functions, we repeated the fitting procedure using 3 alternative sets of E and I sigmoidal basis functions and obtained similar results (Figure 6). This indicates that the cell-cell interactions needed to detect object boundaries in natural images can be produced easily by this general type of compound E-I, or “incitatory” circuit.

We next looked for regularities in the progression of excitatory-inhibitory curve pairs used to fit the LLRs as a function of a neighbor cell's offset in position and orientation from the RL (Figure 5D). We observed the following patterns. First, as the neighbor's orientation offset from the RL increases and approaches 90 degrees (indicated by lightness changes within each plot), excitation becomes weaker, and inhibition becomes both stronger and lower in threshold, resembling cross-orientation suppression (a staple function of V1 (Bishop et al., 1973; DeAngelis et al., 1992; Geisler and Albrecht, 1992); though see Priebe and Ferster, 2006). Second, we observed a gradual weakening of both excitation and inhibition as a neighbor cell moves further from the RL in the direction perpendicular to the boundary orientation (different plot columns), reflecting the expected decline in informativeness

377 as a neighbor cell moves further from the boundary cell in question. To probe this effect further, we
378 characterized each excitatory and inhibitory curve by its gain parameter and plotted the gains
379 separately as a function of a neighbor's orientation difference and spatial offset relative to the RL
380 (Figure 5E). These surfaces confirm that, under this simple difference-of-sigmoids model, the strength
381 of the excitation and inhibition imparted to a boundary cell by neighboring simple cells varies
382 systematically with offset in RF position and orientation. The pattern is non-obvious, however, so if
383 seen experimentally, could be difficult to interpret without the benefit of a normative framework, such
384 as the one we have adopted here.
385

386 **Learning the parameters of the incitation circuit**

387 We showed that the incitatory interconnection circuit of Figure 5A is capable of producing the diverse
388 forms of simple cell-boundary cell interactions needed to compute boundary probability. However, the
389 circuit contains parameters that would need to be set, presumably during development, to allow each
390 simple cell to exert the appropriate effect on every surrounding boundary cell. We asked whether
391 these cell-cell interactions could be learned by a simple synaptic learning rule operating within the
392 slightly elaborated incitation circuit shown in Figure 7A. In particular, we assumed that each of the 300
393 oriented receptive fields surrounding a boundary cell is represented by a population of 8 simple cells,
394 all sharing the same oriented receptive field, but each having a different firing threshold, which was
395 intended to reflect natural variation in neuron size, morphology, firing dynamics, etc. The regularly
396 spaced threshold settings for the 8 cells are given in the Methods section. Each pre-synaptic simple
397 cell acted on the boundary cell through two adjustable weights, one excitatory weight directly onto the
398 boundary cell, and one excitatory weight onto the boundary cell's inhibitory "partner" cell, which

399 would contribute to disynaptic inhibition (Figure 7A). Three examples of filters (red, green, and yellow)
400 and their associated simple cell variants are depicted schematically in Figure 7A. Labeled image
401 patches containing boundaries and non-boundaries were presented to the 2,400 (=300x8) simple cells,
402 ground-truth labels from the natural image dataset were presented to the boundary cell (1 for
403 boundary, 0 for no boundary), and the excitatory synapses between the simple cells and the boundary
404 cell and its associated inhibitory neuron were adjusted using a single-layer supervised logistic
405 regression learning rule (identical to the perceptron learning rule, but with a smooth, sigmoidal neuron
406 activation function – Murphy, 2012). We then performed virtual neurophysiology to probe the net
407 effect of each underlying oriented filter on the boundary cell's response, induced by that filter's 8
408 simple cell variants each acting through its learned synaptic weights. These learned sLLR-like functions
409 again included monotonic rising and falling as well as non-monotonic bump-shaped functions (Figure
410 7B, colored curves). For some cells the learned SB-BC interaction functions corresponded closely to
411 the actual sLLRs (thin grey lines), most notably the cells centered on the RL at all different orientations
412 (middle column of Figure 7B). In other cases, one or two of the learned SC-BC interaction functions in
413 each group overlapped with the explicit LLR curves, having a similar parabolic shape, whereas the
414 other curves in the group were driven apart by the learning rule, covering a much wider range of
415 generally parabolic shapes.
416
417 Thus, the incitatory interconnection network depicted in Figure 7A can learn to produce the spectrum
418 of SC-BC interactions needed for boundary detection using a small number of neurons covering each
419 receptive field location.
420

421 In an interesting difference between the learned SC-BC interaction functions compared to the explicitly
422 calculated LLRs, we noted that the learned sLLR-like functions also included rightside up U-shaped
423 interactions (e.g. pink curves in lower right corner of Figure 7B). This and other more subtle differences
424 relative to the literal sLLRs can be attributed to the fact that the learning rule attempts to compensate
425 for statistical dependencies between input features, whereas the literal sLLRs shown in Figure 4B
426 reflect the simplifying assumption that each oriented filter contributes independently to the response
427 of a boundary cell. (The effect of conditional dependencies among filters is taken up again below, in
428 the text surrounding Figure 8). The overall similarity of the pattern of learned cell-cell interactions
429 compared to the literal sLLR's, however, validates the Bayesian-inspired reverse-engineering approach
430 to study cell-cell interactions in V1, in that Bayes rule allows us to convert intuitive labels assigned to
431 natural images into predictions as to how cortical neurons should influence each other to solve a
432 particular task (Figure 1).

433 **Comparing boundary detection performance of four models**

434 As an additional performance-based approach to evaluating the trained incitation circuit as a model of
435 boundary detection in V1, we compared the Precision-Recall curves of 5 different boundary detectors:
436 (1) the "null hypothesis", consisting of a single conventional simple cell centered at the RL (Figure 7C,
437 blue curve); (2) an unweighted sum of 300 literal LLRs (orange curve); this is essentially a direct
438 application of Bayes rule under the assumption of class-conditional independence as shown in Figure
439 1D, without any learning (see Ramachandra and Mel 2013); (3) a *weighted* sum of the 300 neighboring
440 literal sLLR's (green curve); again corresponding to the model of Figure 1D, but augmented with
441 weights optimized by logistic regression); the learned weights in this case help compensate for the CCI
442 violations among the filters, explaining the significantly improved PR performance; and (4) the learned

443 neuromorphic classifier shown in Figure 7A, with 8 simple cell variants per oriented RF (red curve; total
444 number of simple cell variants is $8 \times 300 = 2400$). We note that learning is possible even though all
445 modifiable weights are constrained to be positive, so that the model does not require either that
446 synaptic weights change sign, or that inhibitory weights are modifiable, both of which are more
447 difficult to justify biologically.
448
449 These results lead us to 4 conclusions: (1) the superior performance of all 3 multi-input classifier
450 variants compared to a single conventional simple cell reinforces the point that individual simple cells
451 are poor-performing natural boundary detectors, that can be significantly improved upon using
452 neurally plausible local circuit computations; (2) the superior performance of the 2 classifier variants
453 that depend on populations of simple cells (300, green curve, or 2400, red curve) with optimized
454 weights, compared to a classifier with unweighted inputs (Figure 7C, orange curve), points to the value
455 of learning the circuit parameters from natural image statistics; (3) the similar performance of the
456 learned neuromorphic classifier compared to a weighted sum of literal sLLRs reinforces the close
457 connection between Bayes rule and the circuit of Figure 7A, and shows that (4) the incitation circuit of
458 Figure 7A can extract critical information needed for boundary detection from relatively few neurons.
459

460 In light of these comparisons, we conclude that the requirements for developing a cortical circuit that
461 significantly improves boundary detection performance compared to a lone simple cell are modest,
462 including mainly (1) a compound E-I circuit motif known to exist in V1; (2) natural variation in firing
463 thresholds across the population of simple cells; and (3) a synaptic learning rule to adjust a single layer
464 of excitatory weights, (Possible sources of “supervision” during learning are taken up in the
465 Discussion). In the cortex, the modifiable parameters could be the strengths of synaptic contacts on

466 the dendrites of different boundary cells and on interneurons, as shown here, or varying dendritic
467 locations of those synaptic contacts[22–25], or a choice among interneurons having varying gains and
468 thresholds (Druckmann, Hill, Schürmann, Markram, & Segev, 2013; Markram et al., 2004).

469 **Probing the relationship between the incitation circuit and Bayes rule?**

470 To better understand the relationship between the Bayes' rule classifier shown in Figure 1 and the
471 incitation circuit shown in Figure 7, we ran simple learning experiments with pairs of simple cells
472 overlapping to varying degrees. For each simple cell pair, we fit the parameters of the incitation circuit
473 either separately (Figure 8A, left) or jointly (Figure 8A, right). We tested pairs of filters ranging from
474 very dependent (Figure 8B, middle columns) to nearly independent (Figure 8B; outer columns). Scatter
475 plots of joint filter responses to boundary (red) and non-boundary (black) patches are shown below
476 each pair. When the SC-BC interaction functions were learned separately, they were nearly identical to
477 the literal LLRs (Figure 8B, first row of blue and orange curves; solid curves show learned interactions,
478 dashed curves show LLRs). On the other hand, when the SC-BC interactions were learned jointly, for SC
479 pairs with heavily overlapping RFs, which led to a breakdown of the CCI assumption, the learned
480 interactions differed significantly from the pure LLRs (Figure 8B, middle columns). Consistent with
481 these observations, we show analytically in Appendix 1 that an incitation circuit like the one shown in
482 Figure 7 will learn LLRs if the input features are CCI. Consequently any observable differences between
483 the learned incitation functions and the LLRs can be attributed to a lack of exact class-conditional
484 independence.

485

486 **Experimentally distinguishing boundary cells from conventional simple cells**

487 Having shown that V1 circuitry is capable of producing boundary cells from simple cells using only a
488 single layer of modifiable excitatory weights, we next considered the question as to how BCs could be
489 detected experimentally, and distinguished from conventional simple cells (or the simple cell-like
490 subunits of complex cells – Hubel & Wiesel, 1962b; Movshon, Thompson, & Tolhurst, 1978; Ohzawa,
491 DeAngelis, & Freeman, 1997).
492
493 To determine how BCs would respond to various stimuli, stimulus patches were scaled to have the
494 same fixed value of the normalizer used in earlier figures, once again reflecting a simple form of
495 divisive normalization (see Methods section *Boundary cell stimulus responses*). We first constructed a
496 canonical stimulus for a boundary cell akin to a spike triggered average by averaging all image patches
497 weighted by their evoked boundary cell response. As expected, the STA stimulus appears as a localized,
498 polarized, oriented boundary segment reminiscent of a simple cell's receptive field (Figure 9A). We
499 then presented drifting sine wave gratings covering a boundary cell's "classical receptive field", leading
500 to the unremarkable phase response and orientation tuning curves shown in Figure 9B and C. Next, we
501 used labeled natural edges with the same normalizer value to explore the effect of increasing center
502 contrast on orientation tuning curve width. (This was not a perfectly controlled experiment because
503 variations in center contrast at a fixed normalizer value would have led to *anti*-variations in surround
504 contrast, but given the filter value at the RF center was only one of 100 filters of many orientations
505 used to compute the normalizer value, this effect was likely small). Subject to this limitation, as shown
506 in Figure 9B, the boundary cell's tuning width is essentially constant across a roughly 2-fold change in
507 center contrast – the limit of analysis allowed by our labeled database (average tuning curve has full
508 width at half height for high contrast stimuli 43.6°; for low contrast stimuli, 39.2°).
509

510 Thus, for oriented edges and gratings presented within the CRF, boundary cells behave similarly to
511 conventional simple cells in that they (1) have phase-dependent responses; (2) are orientation tuned;
512 and (3) have tuning curves whose widths are roughly contrast invariant (Alitto and Usrey, 2004). It is
513 therefore possible that boundary cells exist and have been classified as conventional simple cells in
514 previous experiments using simplified stimuli. Among the multiple types of V1 cells that have been
515 previously described, boundary cells share most in common with double opponent cells, which are
516 orientation tuned, have mostly odd-symmetric receptive field profiles as would be expected for
517 boundary detecting cells (Ringach, 2002), and respond to boundaries whether defined by luminance or
518 color (Johnson et al., 2008).
519
520 In future neurophysiological studies, an efficient means of dissociating conventional simple cells, which
521 respond to oriented contrast independent of boundary probability, from putative boundary cells,
522 which respond to boundary probability independent of oriented contrast, would be to use natural
523 image stimuli drawn from the four corners of the oriented contrast – boundary probability space
524 (Figure 10A). Image patches with low oriented contrast and low boundary probability scores (purple
525 dots) tend to contain flat, unstructured image regions; patches with low contrast and high probability
526 (green dots) tend to contain well-structured, faint edges; patches with high contrast but low
527 probability (blue dots) tend to contain contrasty noise or misaligned edges; and regions with high
528 contrast *and* high probability (red dots) are typically well-structured, strong edges (Figure 10B). This
529 factorial stimulus set would make it possible to identify pure simple cells, pure boundary cells, as well
530 as cells of intermediate type.
531

532 **DISCUSSION**

In the 60 years since Hubel and Wiesel first discovered orientation-tuned simple cells in V1, it has been generally assumed that these cells contribute in some way to the detection of object boundaries (Angelucci et al., 2002; Field et al., 1993; Grosof et al., 1993; Kapadia et al., 1995b, 2000a; Polat et al., 1998; Sceniak et al., 1999). 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; Krizhevsky et al., 2012; Lades et al., 1993; Lecun et al., 1998; Mel, 1997; Riesenhuber and Poggio, 1999). Surprisingly, however, the quantitative relationship between simple cell responses, typically modeled as divisively normalized linear filters (Carandini and Heeger, 2012), 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 (see also Arbelaez et al., 2011): as shown in Figure 7C (blue curve), if we use a simple cell's response 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 will be false positives (corresponding to a Precision score of ~35%). The reason a simple cell is such an unreliable edge detector is that true object boundaries are rare (note the overwhelming majority of points in Figure 10A are piled in the lower half of the plot), and when they do occur, they are very often of low contrast. Much more common are high contrast non-edge structures (e.g. textures) that contain sufficient oriented energy to strongly drive simple oriented filters.

555 The poor boundary detection performance of a lone simple cell leads to the conjecture that V1 also
556 contains "smarter" cells that compute boundary probability by combining the responses of multiple
557 simple cells covering a local neighborhood. In a previous study, we suggested that the appropriate
558 strategy for constructing a boundary cell from a local population of simple cells was to (1) select a small
559 set of simple cells (e.g. 6 cells) that were both individually informative and class-conditionally
560 independent (see Methods for discussion of the CCI assumption); (2) evaluate the log-likelihood ratios
561 for each of the participating simple cells, which would be the optimal functional interconnections
562 between each simple cell and the boundary cell (according to Bayes rule); and (3) sum the LLRs and
563 apply a fixed sigmoidal nonlinearity to compute boundary probability (Ramachandra and Mel, 2013)
564 (Figure 1B). The present study extends that previous work in eight ways: (1) we collected and analyzed
565 individual LLRs for *all* of the simple cells at all orientations covering a 5x5 pixel neighborhood in the
566 vicinity of a boundary cell's RF (300 cells total); (2) we show that the idealized functional
567 interconnections between simple cells and boundary cells depend systematically on the relative
568 positions and orientations of the simple cell and boundary cell RFs (Figure 2) – but are relatively
569 insensitive to the scale or aspect ratio of the simple cell receptive fields (Figure 3); (3) we developed a
570 simple analytical model (i.e. gaussian likelihoods->quadratic LLRs) that shows how the three seemingly
571 different types of SC-BC interaction functions – rising, falling, and bump-shaped functions – represent
572 different ranges of the same underlying (quadratic) function class (Figure 4); (4) we show that a mixed
573 excitatory-inhibitory, or "incitatory", circuit motif that is known to exist in V1 is capable of producing
574 the entire spectrum of natural image-derived SC-BC interaction functions (Figures 5,6); (5) we show
575 that the parameters of a boundary-detecting incitation circuit can be learned by adjusting a single layer
576 of excitatory weights (Figure 7A); (7) we show that a learned incitation circuit can improve the
577 precision of boundary detection in the high-recall range by 43% to 121% compared to a conventional

simple cell model (Figure 7C); and (8) 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 due to class-conditional dependencies among simple cells whose receptive fields overlap heavily with each other. This could be helpful in interpreting the results of future neurophysiological experiments in V1.

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; Bell and Sejnowski, 1995; Laughlin, 1989; 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 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 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. and Sigman et al. 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.

Non-monotonic 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", nor can they be 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 (Figure 7B). Should we find the idea surprising that nearby cells in the cortex act on each other non-monotonically?

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 non-monotonic interactions between cells. For example, in the "inhibition-stabilized network" model (Jadi and

Sejnowski, 2014; Ozeki et al., 2009), which accounts for a number of V1 cell response properties, "non-binary" 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 (Adesnik et al., 2012; Cavanaugh et al., 2002; Schwabe et al., 2010). 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 upon. For example, K.D. Miller's models of map development include short range excitation and medium-range inhibition (Miller, 1994); Angelucci and Bressler's models include near and far suppressive surrounds (Angelucci and Bressloff, 2006); 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 (e.g. Bosking, Zhang, Schofield, & Fitzpatrick, 1997; Field et al., 1993; W. S. Geisler et al., 2001; Kapadia, Westheimer, & Gilbert, 2000b; Li, 1999; Sigman et al., 2001). In all of these cases, one neuron's effect 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 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

645 modulation can switch from facilitating to suppressive with increasing center contrast (Ichida et al.,
646 2007; Nauhaus et al., 2009; Polat et al., 1998; Schwabe et al., 2006; Somers et al., 1998); length-tuned
647 neurons respond best to an oriented stimulus up to a certain length, but are then progressively
648 inhibited as the stimulus grows longer (Anderson et al., 2001); and non-monotonic modulatory
649 interactions between a neuron's classical and extra-classical receptive fields have been reported (Polat
650 et al., 1998). These data, though unaccompanied by normative explanations, do support the idea that
651 the sign and magnitude one neuron's effect on another can depend not only on the relative position
652 and orientation of their receptive fields (in the case of vision), but also on their relative activity levels.
653
654 Our paper represents a fleshing out of this type of effect, and is to our knowledge the first normative
655 theory, parameterized by natural images, that specifies how intracolumnar cell-cell interactions may
656 help solve a specific, biologically-relevant *classification* problem. By analyzing natural image data on
657 and off object boundaries, we showed that the local cell-cell interactions needed to solve this
658 classification problem are not capturable by scalar weights, but are in general nonlinear functions that
659 depend on "all of the above" – relative location, relative orientation, and relative activity levels of the
660 sending and receiving cells. And while such connections can only in special cases be described by
661 scalar weights, we showed that they are easily produced by a compound E-I circuit motif (see Figure 5)
662 that is known to exist in the cortex (Buzsáki, 1984; Isaacson & Scanziani, 2011; Klyachko & Stevens,
663 2006; McBain & Fisahn, 2001; Pfeffer et al., 2013; Pouille & Scanziani, 2001; Swadlow, 2002; Wehr &
664 Zador, 2003) . Further, we showed that the synaptic weights that control the net effect of an
665 "incitation" motif are easily learned. Future experiments will be needed to establish whether trainable
666 incitation circuits are actually used to help solve the difficult natural classification problems faced by
667 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 an explicit Bayesian classifier. We found this to be true (Figure 7C), 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 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 MT and IT in the monkey (Marcar et al., 1995, 2000; Sary et al., 1995; Zeki et al., 2003); and in multiple areas of the human visual cortex (Larsson et al., 2010; Mysore et al., 2006; Zeki et al., 2003). 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 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 underlyingly a feedforward computation whose core operation is a sum of LLR terms (Figure 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, omitting local or long-range feedback connections that are known to exist in the neocortex (Angelucci et al., 2017), and furthermore omits 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).

Abbreviations

SC: simple cell
BC: boundary cell
RF: receptive field
RL: reference location
LLR: log-likelihood ratio
sLLR: shifted log-likelihood ratio
CCI: class conditional independence
PR: precision-recall

References

Adesnik, Hillel, William Bruns, Hiroki Taniguchi, Z Josh Huang, and Massimo Scanziani. (2012). "A Neural Circuit for Spatial Summation in Visual Cortex." *Nature* 490 (7419): 226–31. https://doi.org/10.1038/nature11526.

Alitto, Henry J., and W. Martin Usrey. (2004). "Influence of Contrast on Orientation and Temporal Frequency Tuning in Ferret Primary Visual Cortex." *Journal of Neurophysiology* 91 (6): 2797–2808. https://doi.org/10.1152/jn.00943.2003.

Anderson, Jeffrey S., Ilan Lampl, Deda C. Gillespie, and David Ferster. (2001). "Membrane Potential and Conductance Changes Underlying Length Tuning of Cells in Cat Primary Visual Cortex." *Journal of Neuroscience* 21 (6): 2104–12.

Angelucci, Alessandra, and Paul C. Bressloff. (2006). "Contribution of Feedforward, Lateral and Feedback Connections to the Classical Receptive Field Center and Extra-Classical Receptive Field Surround of Primate V1 Neurons." In *Progress in Brain Research*, 154:93–120. Elsevier.

Angelucci, Alessandra, Jonathan B. Levitt, Emma J. S. Walton, Jean-Michel Hupé, Jean Bullier, and Jennifer S. Lund. (2002). "Circuits for Local and Global Signal Integration in Primary Visual Cortex." *The Journal of Neuroscience* 22 (19): 8633–46.

Arbelaez, P., M. Maire, C. Fowlkes, and J. Malik. (2011). "Contour Detection and Hierarchical Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (5): 898–916. https://doi.org/10.1109/TPAMI.2010.161.

Barlow, H. B. (1981). "The Ferrier Lecture, 1980: Critical Limiting Factors in the Design of the Eye and Visual Cortex." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 212 (1186): 1–34. https://doi.org/10.1098/rspb.1981.0022.

Behabadi, Bardia F, Alon Polsky, Monika Jadi, Jackie Schiller, and Bartlett W Mel. (2012). "Location-Dependent Excitatory Synaptic Interactions in Pyramidal Neuron Dendrites." *PLoS Computational Biology* 8 (7): e1002599. https://doi.org/10.1371/journal.pcbi.1002599.

Bell, A J, and T J Sejnowski. (1995). "An Information-Maximization Approach to Blind Separation and Blind Deconvolution." *Neural Computation* 7 (6): 1129–59.

Biederman, Irving. (1987). "Recognition-by-Components: A Theory of Human Image Understanding." *Psychological Review* 94 (2): 115–147.

Bishop, P. O., J. S. Coombs, and G. H. Henry. (1973). "Receptive Fields of Simple Cells in the Cat Striate Cortex." *The Journal of Physiology* 231 (1): 31–60. https://doi.org/10.1113/jphysiol.1973.sp010218.

Bosking, William H., Ying Zhang, Brett Schofield, and David Fitzpatrick. (1997). "Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex." *Journal of Neuroscience* 17 (6): 2112–27.

Buzsáki, G. (1984). "Feed-Forward Inhibition in the Hippocampal Formation." *Progress in Neurobiology* 22 (2): 131–53.

Carandini, Matteo, and David J. Heeger. (2012). "Normalization as a Canonical Neural Computation." *Nature Reviews Neuroscience* 13 (1): 51–62. https://doi.org/10.1038/nrn3136.

Cavanaugh, James R., Wyeth Bair, and J. Anthony Movshon. (2002). "Selectivity and Spatial Distribution of Signals from the Receptive Field Surround in Macaque V1 Neurons." *Journal of Neurophysiology* 88 (5): 2547–56. https://doi.org/10.1152/jn.00693.2001.

DeAngelis, G. C., J. G. Robson, I. Ohzawa, and R. D. Freeman. (1992). "Organization of Suppression in Receptive Fields of Neurons in Cat Visual Cortex." *Journal of Neurophysiology* 68 (1): 144–63.

Druckmann, Shaul, Sean Hill, Felix Schürmann, Henry Markram, and Idan Segev. (2013). "A Hierarchical Structure of Cortical Interneuron Electrical Diversity Revealed by Automated Statistical Analysis." *Cerebral Cortex* 23 (12): 2994–3006. https://doi.org/10.1093/cercor/bhs290.

Field, David J., Anthony Hayes, and Robert F. Hess. (1993). "Contour Integration by the Human Visual System: Evidence for a Local 'association Field.'" *Vision Research* 33 (2): 173–93. https://doi.org/10.1016/0042-6989(93)90156-Q.

Fukushima, K., S. Miyake, and T. Ito. (1983). "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition." *IEEE Transactions on Systems, Man and Cybernetics* SMC-13 (5): 826–34. https://doi.org/10.1109/TSMC.1983.6313076.

Geisler, W. S., J. S. Perry, B. J. Super, and D. P. Gallogly. (2001). "Edge Co-Occurrence in Natural Images Predicts Contour Grouping Performance." *Vision Research* 41 (6): 711–24.

Geisler, Wilson S., and Duane G. Albrecht. (1992). "Cortical Neurons: Isolation of Contrast Gain Control." *Vision Research* 32 (8): 1409–10. https://doi.org/10.1016/0042-6989(92)90196-P.

George, Andrew A., Ariel M. Lyons-Warren, Xiaofeng Ma, and Bruce A. Carlson. (2011). "A Diversity of Synaptic Filters Are Created by Temporal Summation of Excitation and Inhibition." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 31 (41): 14721–34. https://doi.org/10.1523/JNEUROSCI.1424-11.2011.

Gilbert, Charles D., and Torsten N. Wiesel. (1990). "The Influence of Contextual Stimuli on the Orientation Selectivity of Cells in Primary Visual Cortex of the Cat." *Vision Research*, Optics Physiology and Vision A Festschrift Honoring Professor Gerald Westheimer on His 65th Birthday, 30 (11): 1689–1701. https://doi.org/10.1016/0042-6989(90)90153-C.

Grosof, D H, R M Shapley, and M J Hawken. (1993). "Macaque V1 Neurons Can Signal 'Illusory' Contours." *Nature* 365 (6446): 550–52. https://doi.org/10.1038/365550a0.

Heydt, R. von der, and E. Peterhans. (1989). “Mechanisms of Contour Perception in Monkey Visual Cortex. I. Lines of Pattern Discontinuity.” *The Journal of Neuroscience* 9 (5): 1731–48.

Hoffman, Donald D. (2000). *Visual Intelligence: How We Create What We See*. WW Norton & Company. https://books.google.com/books?hl=en&lr=&id=_2iBn_G3E7MC&oi=fnd&pg=PR9&dq=david+hoffman+visual+intelligence&ots=abym9UY5d5&sig=2qNW0k6I-4bd5iPzlrRGWWOwbto.

Hubel, D H, and T N Wiesel. (1962a). “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex.” *The Journal of Physiology* 160 (January): 106–54.

1962b. “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex.” *The Journal of Physiology* 160 (January): 106–54.

Ichida, Jennifer M., Lars Schwabe, Paul C. Bressloff, and Alessandra Angelucci. (2007). “Response Facilitation From the ‘Suppressive’ Receptive Field Surround of Macaque V1 Neurons.” *Journal of Neurophysiology* 98 (4): 2168–81. https://doi.org/10.1152/jn.00298.2007.

Isaacson, Jeffry S, and Massimo Scanziani. (2011). “How Inhibition Shapes Cortical Activity.” *Neuron* 72 (2): 231–43. https://doi.org/10.1016/j.neuron.2011.09.027.

Jadi, M. P., and T. J. Sejnowski. (2014). “Regulating Cortical Oscillations in an Inhibition-Stabilized Network.” *Proceedings of the IEEE* 102 (5): 830–42. https://doi.org/10.1109/JPROC.2014.2313113.

Jadi, Monika, Alon Polsky, Jackie Schiller, and Bartlett W Mel. (2012). “Location-Dependent Effects of Inhibition on Local Spiking in Pyramidal Neuron Dendrites.” *PLoS Computational Biology* 8 (6): e1002550. https://doi.org/10.1371/journal.pcbi.1002550.

Johnson, Elizabeth N., Michael J. Hawken, and Robert Shapley. (2008). “The Orientation Selectivity of Color-Responsive Neurons in Macaque V1.” *Journal of Neuroscience* 28 (32): 8096–8106. https://doi.org/10.1523/JNEUROSCI.1404-08.2008.

Kapadia, Mitesh K., Minami Ito, Charles D. Gilbert, and Gerald Westheimer. (1995a). "Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys." *Neuron* 15 (4): 843–56. https://doi.org/10.1016/0896-6273(95)90175-2. 1995b. "Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys." *Neuron* 15 (4): 843–56. https://doi.org/10.1016/0896-6273(95)90175-2.

Kapadia, Mitesh K., Gerald Westheimer, and Charles D. Gilbert. (2000a). "Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception." *Journal of Neurophysiology* 84 (4): 2048–62. 2000b. "Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception." *Journal of Neurophysiology* 84 (4): 2048–62.

Klyachko, Vitaly A., and Charles F. Stevens. (2006). "Excitatory and Feed-Forward Inhibitory Hippocampal Synapses Work Synergistically as an Adaptive Filter of Natural Spike Trains." *PLOS Biol* 4 (7): e207. https://doi.org/10.1371/journal.pbio.0040207.

Koch, C, T Poggio, and V Torre. (1983). "Nonlinear Interactions in a Dendritic Tree: Localization, Timing, and Role in Information Processing." *Proceedings of the National Academy of Sciences of the United States of America* 80 (9): 2799–2802.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Lades, M., J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. (1993). "Distortion Invariant Object Recognition in the Dynamic Link Architecture." *IEEE Transactions on Computers* 42 (3): 300–311. https://doi.org/10.1109/12.210173.

Laughlin, S. B. (1989). "The Role of Sensory Adaptation in the Retina." *Journal of Experimental Biology* 146 (1): 39–62.

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. (1998). "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324. https://doi.org/10.1109/5.726791.

Li, Z. (1999). "Visual Segmentation by Contextual Influences via Intra-Cortical Interactions in the Primary Visual Cortex." *Network (Bristol, England)* 10 (2): 187–212.

Major, Guy, Alon Polsky, Winfried Denk, Jackie Schiller, and David W. Tank. (2008). "Spatiotemporally Graded NMDA Spike/Plateau Potentials in Basal Dendrites of Neocortical Pyramidal Neurons." *Journal of Neurophysiology* 99 (5): 2584–2601. https://doi.org/10.1152/jn.00011.2008.

Markram, Henry, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. (2004). "Interneurons of the Neocortical Inhibitory System." *Nature Reviews Neuroscience* 5 (10): 793–807. https://doi.org/10.1038/nrn1519.

McBain, C. J., and A. Fisahn. (2001). "Interneurons Unbound." *Nature Reviews. Neuroscience* 2 (1): 11–23. https://doi.org/10.1038/35049047.

Mel, B W. (1997). "SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally Inspired Approach to Visual Object Recognition." *Neural Computation* 9 (4): 777–804.

Miller, K. D. (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." *Journal of Neuroscience* 14: 409–409.

Movshon, J. A., I. D. Thompson, and D. J. Tolhurst. (1978). "Spatial and Temporal Contrast Sensitivity of Neurones in Areas 17 and 18 of the Cat's Visual Cortex." *The Journal of Physiology* 283 (October): 101–20.

Murphy, Kevin P. "Machine Learning." The MIT Press, 2012. https://mitpress.mit.edu/books/machine-learning-1.

Nauhaus, Ian, Laura Busse, Matteo Carandini, and Dario L. Ringach. (2009). "Stimulus Contrast Modulates Functional Connectivity in Visual Cortex." *Nature Neuroscience* 12 (1): 70–76. https://doi.org/10.1038/nn.2232.

Ohzawa, I, G C DeAngelis, and R D Freeman. (1997). "Encoding of Binocular Disparity by Complex Cells in the Cat's Visual Cortex." *Journal of Neurophysiology* 77 (6): 2879–2909.

Olshausen, Bruno A., and David J. Field. (1996). "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images." *Nature* 381 (6583): 607. https://doi.org/10.1038/381607a0.

Ozeki, Hirofumi, Ian M. Finn, Evan S. Schaffer, Kenneth D. Miller, and David Ferster. (2009). "Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression." *Neuron* 62 (4): 578–92. https://doi.org/10.1016/j.neuron.2009.03.028.

Pfeffer, Carsten K., Mingshan Xue, Miao He, Z. Josh Huang, and Massimo Scanziani. (2013). "Inhibition of Inhibition in Visual Cortex: The Logic of Connections between Molecularly Distinct Interneurons." *Nature Neuroscience* 16 (8): 1068–76. https://doi.org/10.1038/nn.3446.

Polat, Uri, Keiko Mizobe, Mark W. Pettet, Takuji Kasamatsu, and Anthony M. Norcia. (1998). "Collinear Stimuli Regulate Visual Responses Depending on Cell's Contrast Threshold." *Nature* 391 (6667): 580–84. https://doi.org/10.1038/35372.

Pouille, F, and M Scanziani. (2001). “Enforcement of Temporal Fidelity in Pyramidal Cells by Somatic Feed-Forward Inhibition.” *Science (New York, N.Y.)* 293 (5532): 1159–63. https://doi.org/10.1126/science.1060342.

Priebe, Nicholas J., and David Ferster. (2006). “Mechanisms Underlying Cross-Orientation Suppression in Cat Visual Cortex.” *Nature Neuroscience* 9 (4): 552–61. https://doi.org/10.1038/nn1660.

Ramachandra, Chaithanya A, and Bartlett W Mel. (2013). “Computing Local Edge Probability in Natural Scenes from a Population of Oriented Simple Cells.” *Journal of Vision* 13 (14). https://doi.org/10.1167/13.14.19.

Riesenhuber, M, and T Poggio. (1999). “Hierarchical Models of Object Recognition in Cortex.” *Nature Neuroscience* 2 (11): 1019–25. https://doi.org/10.1038/14819.

Ringach, Dario L. (2002). “Spatial Structure and Symmetry of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex.” *Journal of Neurophysiology* 88 (1): 455–63. https://doi.org/10.1152/jn.2002.88.1.455.

Rust, Nicole C, and James J Dicarlo. (2010). “Selectivity and Tolerance (‘invariance’) Both Increase as Visual Information Propagates from Cortical Area V4 to IT.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (39): 12978–95. https://doi.org/10.1523/JNEUROSCI.0179-10.2010.

Sceniak, Michael P., Dario L. Ringach, Michael J. Hawken, and Robert Shapley. (1999). “Contrast’s Effect on Spatial Summation by Macaque V1 Neurons.” *Nature Neuroscience* 2 (8): 733–39. https://doi.org/10.1038/11197.

Schwabe, Lars, Jennifer M. Ichida, S. Shushruth, Pradeep Mangapathy, and Alessandra Angelucci. (2010). “Contrast-Dependence of Surround Suppression in Macaque V1: Experimental Testing

of a Recurrent Network Model." *NeuroImage*, Computational Models of the Brain, 52 (3): 777–92. https://doi.org/10.1016/j.neuroimage.2010.01.032.

Schwabe, Lars, Klaus Obermayer, Alessandra Angelucci, and Paul C. Bressloff. (2006). "The Role of Feedback in Shaping the Extra-Classical Receptive Field of Cortical Neurons: A Recurrent Network Model." *Journal of Neuroscience* 26 (36): 9117–29. https://doi.org/10.1523/JNEUROSCI.1253-06.2006.

Schwartz, Odelia, and Eero P. Simoncelli. (2001). "Natural Signal Statistics and Sensory Gain Control." *Nature Neuroscience* 4 (8): 819. https://doi.org/10.1038/90526.

Sigman, Mariano, Guillermo A. Cecchi, Charles D. Gilbert, and Marcelo O. Magnasco. (2001). "On a Common Circle: Natural Scenes and Gestalt Rules." *Proceedings of the National Academy of Sciences* 98 (4): 1935–40. https://doi.org/10.1073/pnas.98.4.1935.

Somers, D. C., E. V. Todorov, A. G. Siapas, L. J. Toth, D. S. Kim, and M. Sur. (1998). "A Local Circuit Approach to Understanding Integration of Long-Range Inputs in Primary Visual Cortex." *Cerebral Cortex* 8 (3): 204–17. https://doi.org/10.1093/cercor/8.3.204.

Suga, N., and T. Manabe. (1982). "Neural Basis of Amplitude-Spectrum Representation in Auditory Cortex of the Mustached Bat." *Journal of Neurophysiology* 47 (2): 225–55.

Swadlow, Harvey A. (2002). "Thalamocortical Control of Feed–forward Inhibition in Awake Somatosensory 'barrel' Cortex." *Philosophical Transactions of the Royal Society B: Biological Sciences* 357 (1428): 1717–27. https://doi.org/10.1098/rstb.2002.1156.

Theys, Tom, Maria C. Romero, Johannes van Loon, and Peter Janssen. (2015). "Shape Representations in the Primate Dorsal Visual Stream." *Frontiers in Computational Neuroscience* 9. https://doi.org/10.3389/fncom.2015.00043.

Wehr, Michael, and Anthony M. Zador. (2003). “Balanced Inhibition Underlies Tuning and Sharpens Spike Timing in Auditory Cortex.” *Nature* 426 (6965): 442–46. https://doi.org/10.1038/nature02116.

Acknowledgments

Funding for this work was provided by NIH/NEI (EY016093).

Author contributions

BM and CR conceived of the original project. GM and BM collected the natural image data, performed the analysis, developed the models, and wrote the paper.

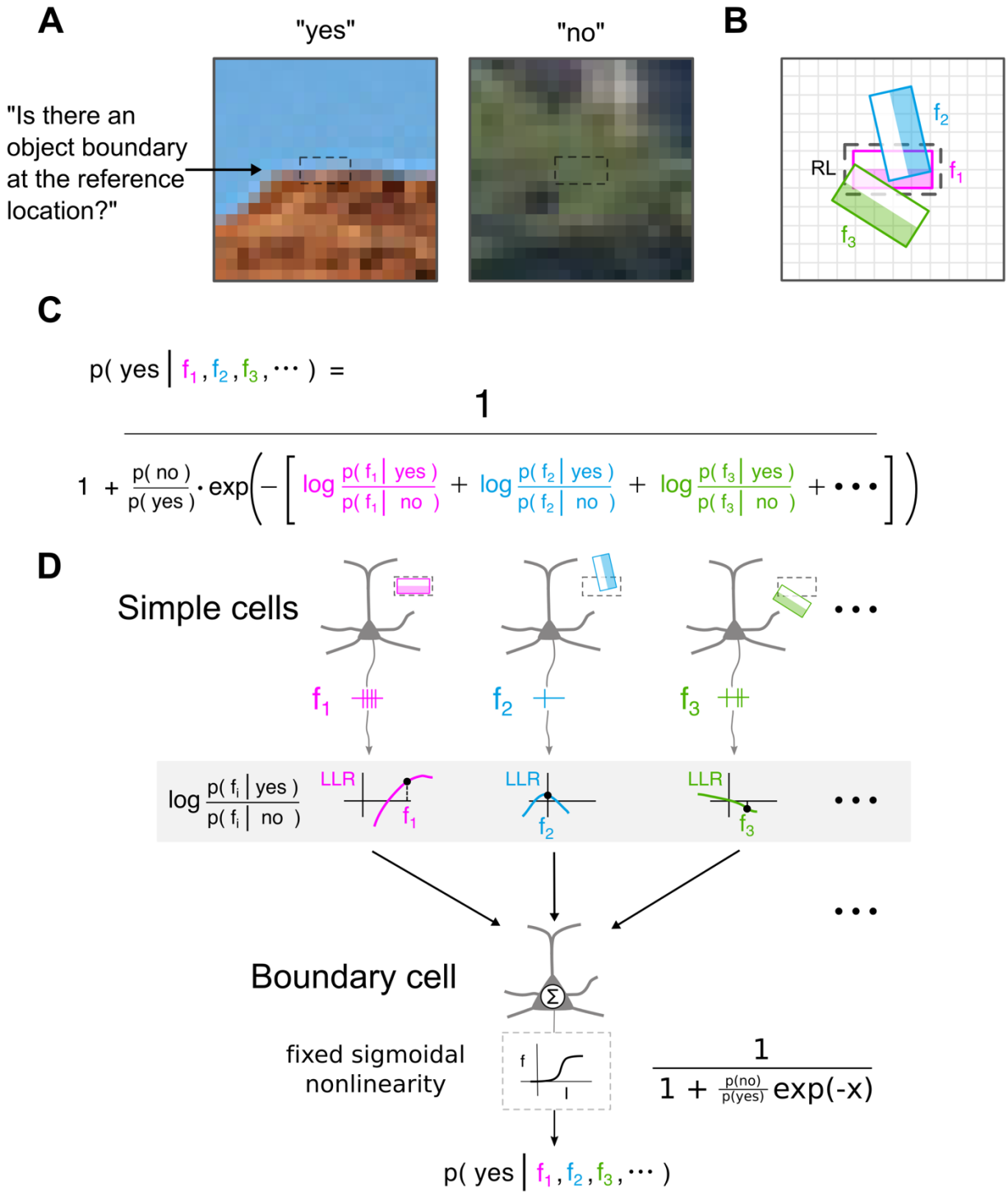
A
"yes"
"no"
"Is there an object boundary at the reference location?"
B
RL
f2
f1
f3
C
p(yes | f1, f2, f3, ···) =
1
1 + p(no) / p(yes) · exp(− [log p(f1 | yes) / p(f1 | no) + log p(f2 | yes) / p(f2 | no) + log p(f3 | yes) / p(f3 | no) + ···])
D
Simple cells
f1
f2
f3
log p(fi | yes) / p(fi | no)
LLR
Boundary cell
fixed sigmoidal nonlinearity
f
I
1 / (1 + p(no) / p(yes) exp(-x))
p(yes | f1, f2, f3, ···)

Figure 1

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 labelling were 20x20 pixels. (**B**) 3 (of many) oriented linear filters with responses f_1, f_2, f_3 are shown in the vicinity of the RL. Filters kernels consisted of two rows of values as follows 0.25 * {+1, +1, +1, +1; -1, -1, -1, -1}, or the rotated equivalent (using bilinear interpolation). (**C**) Under the assumption that filters are class-conditionally independent (see Methods), Bayes' rule gives an expression for boundary probability in terms of individual filter log-likelihood ratios (LLRs) (colored terms in denominator). (**D**) Measured filter values are passed through their respective LLR functions, and the results are summed and passed through a fixed sigmoidal "f-I curve" to yield boundary probability.

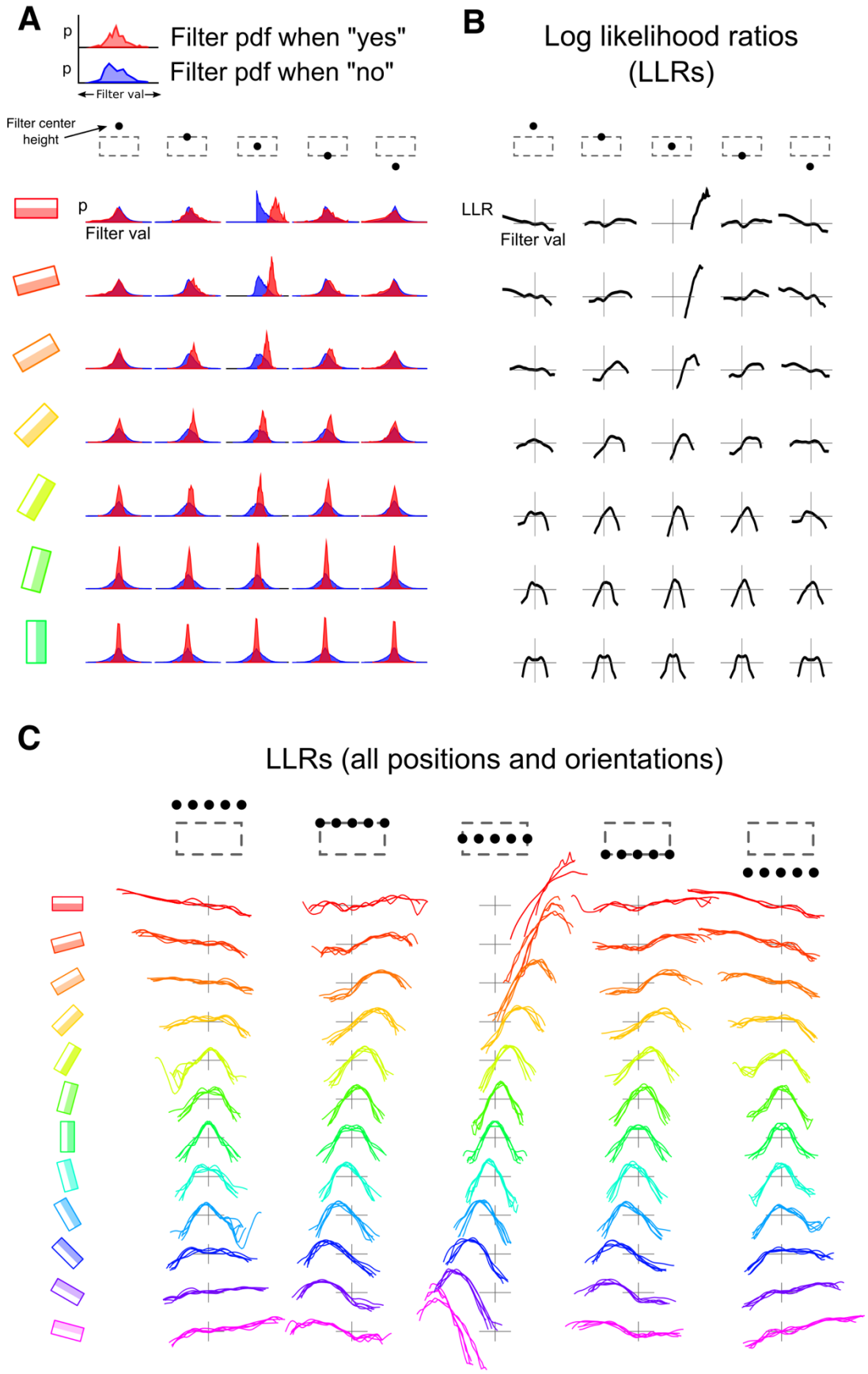
A
p
Filter pdf when "yes"
p
Filter pdf when "no"
Filter val
Filter center height
p
Filter val
B
Log likelihood ratios (LLRs)
LLR
Filter val
C
LLRs (all positions and orientations)

Figure 2

Figure 2. Computing LLRs from natural images. (**A**) Filter responses from 30,000 labeled image patches potentially containing boundaries at the RL (dashed box) were separately histogrammed 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 filter histograms is shown for 7 orientations and 5 vertical positions (centered horizontally). (**B**) By dividing the yes and no distributions and taking logs, one obtains the LLRs. (**C**) Full set of 300 LLRs reveals a regular pattern over orientation and location. Cases grouped within each subplot are for 5 horizontal shifts (indicated by black dots at top). Many LLRs are non-monotonic functions of the filter values.

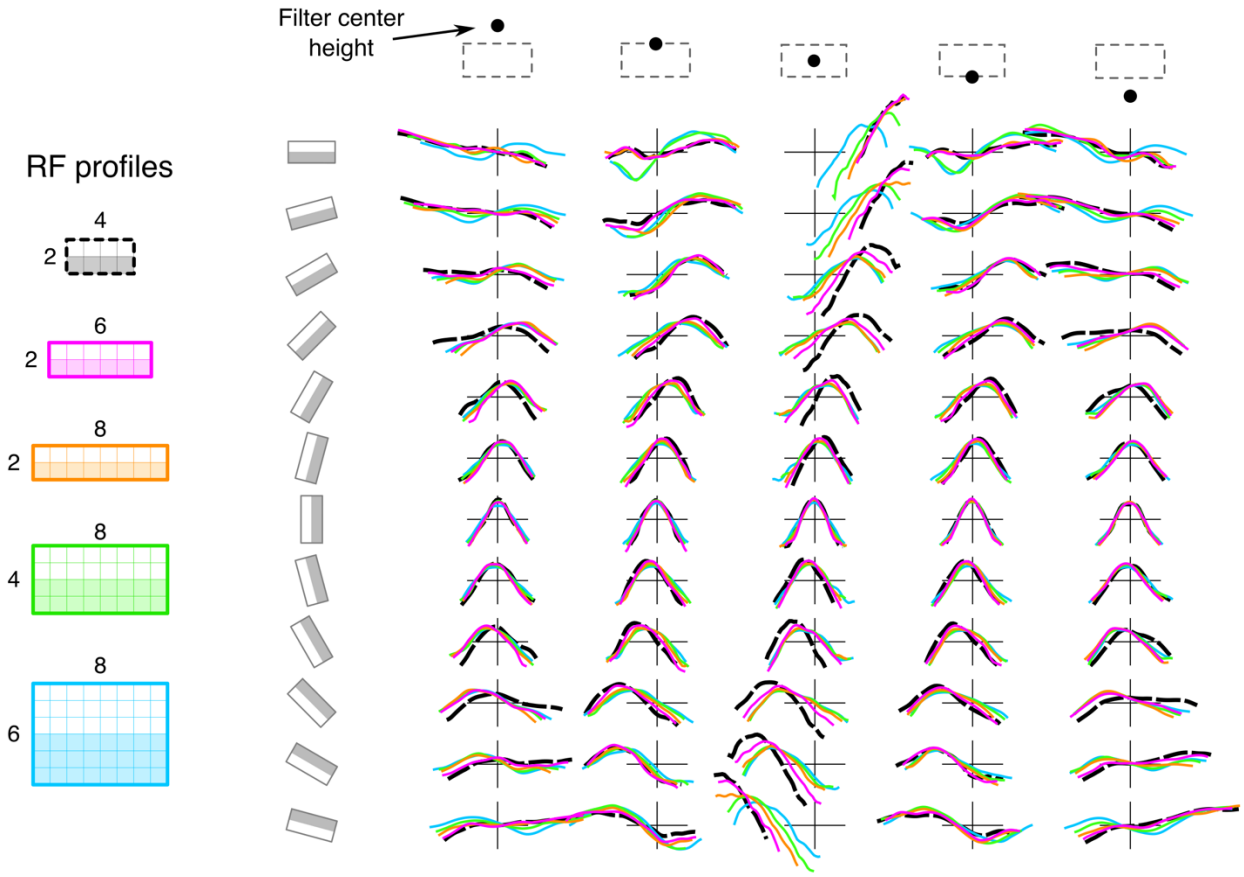

Figure 3

Figure 3. The basic pattern of LLRs forms is conserved across different filter spatial profiles. LLRs were generated for each of the filter profiles shown on the left (2x6, 2x8, 4x8, and 6x8 pixels). The overall spectrum of LLR shapes remains similar for the different cases.

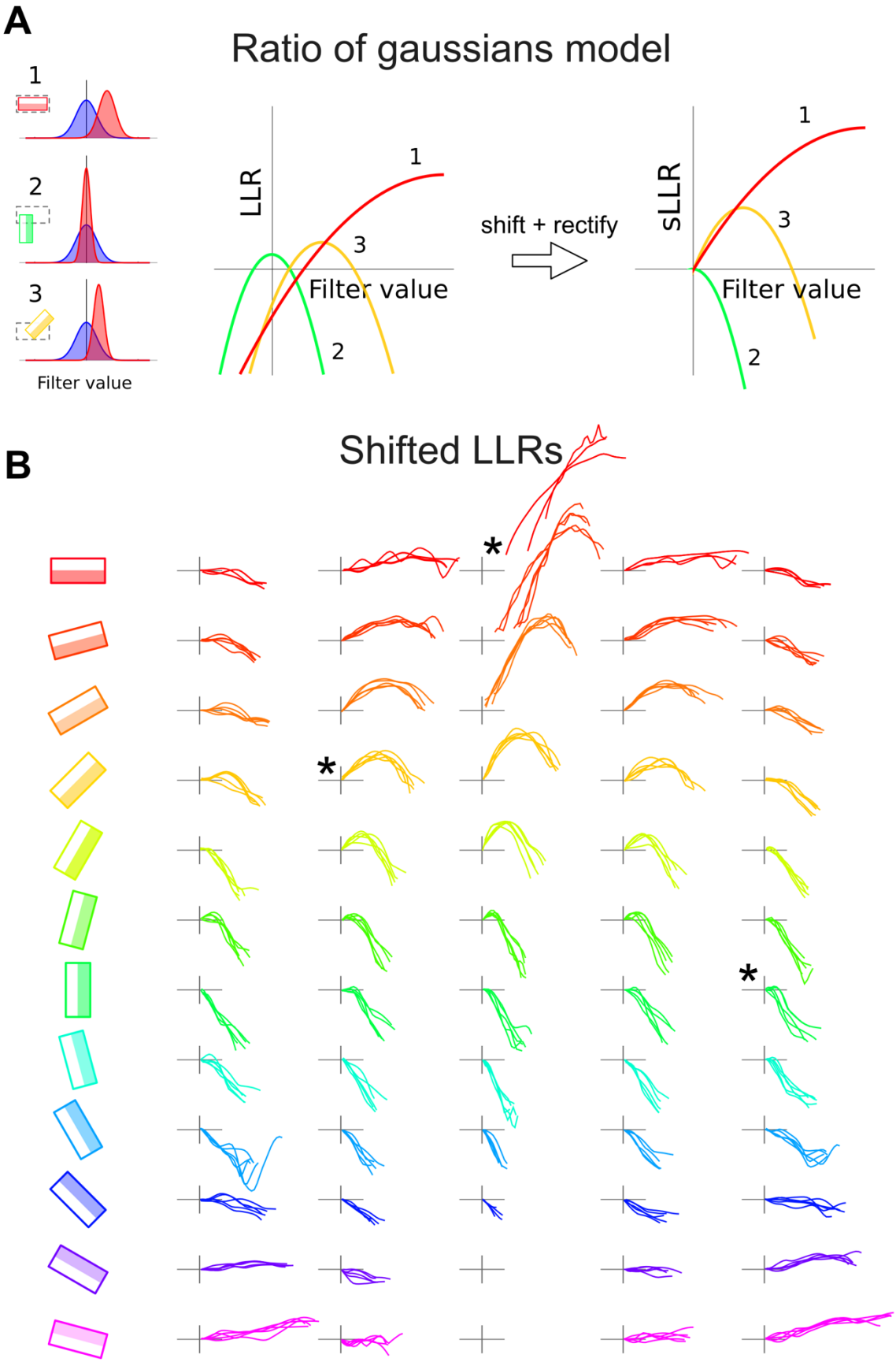
A
Ratio of gaussians model
1
2
3
Filter value
LLR
1
3
2
Filter value
shift + rectify
sLLR
1
3
2
Filter value
B
Shifted LLRs
*
*
*

Figure 4

Figure 4. Interpreting the LLRs as cell-cell interaction functions. (**A**) Modelling the "yes" and "no" distributions as gaussians (left panel) leads to parabolic LLRs (middle panel). In order to interpret the LLRs as cell-cell interactions functions, we perform two additional processing steps: (1) When a simple cell is inactive, it should not influence the boundary cell; this is accomplished by shifting the LLR to have zero output (y=0) when the input is zero (x=0). (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; sLLR stands for "shifted LLR". (**B**) The full set of LLR interactions processed in this way. Many of them are non-monotonic, indicating that that simple cell should have a non-monotonic effect on the boundary cell. The plots corresponding to the three LLRs modelled in (**A**) are marked with asterisks.

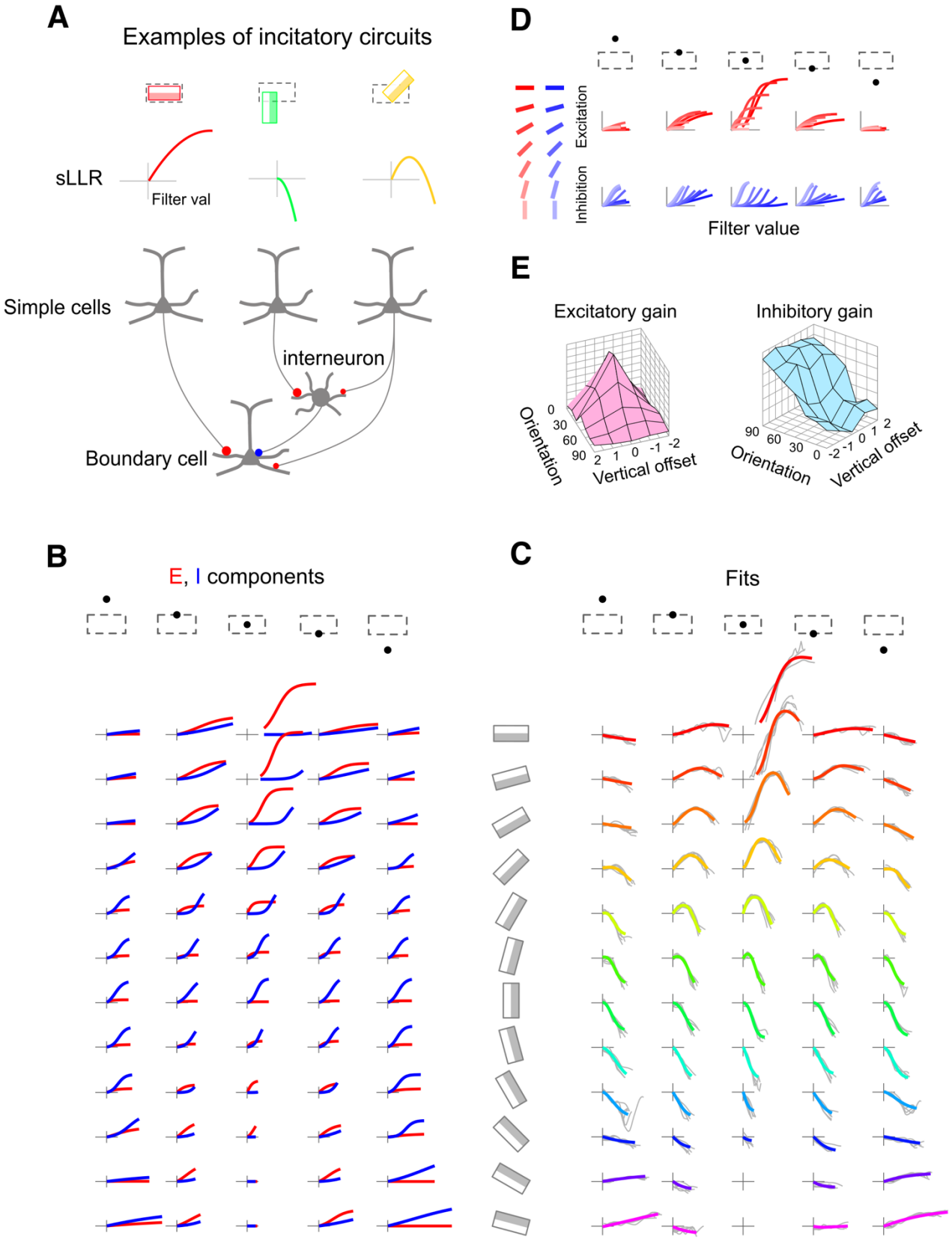
A
Examples of incitatory circuits
sLLR
Filter val
Simple cells
interneuron
Boundary cell
D
Excitation
Inhibition
Filter value
E
Excitatory gain
Inhibitory gain
Orientation
Vertical offset
B
E, I components
C
Fits

Figure 5

Figure 5. Fitting simple cell–boundary cell interactions (LLRs) with a difference of sigmoids representing separate E and I 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) sigmoidal curves were optimized by manipulating their thresholds, slopes and asymptotes so that their difference fit the corresponding LLR shown in (**C**). (**C**) LLR fits are shown in color, on top of the 5-curve groups from Figure 2C shown in light grey. (**D**) E and I sigmoids from b are collected across orientations within each subplot, showing smooth progressions of sigmoid parameters. (**E**) Plots show gains for the E and I interaction components. For groups of simple cells horizontally centered at the RL, excitation delivered to the boundary cell becomes weaker and inhition grows stronger as the neighbor's orientation deviates from the reference orientation.

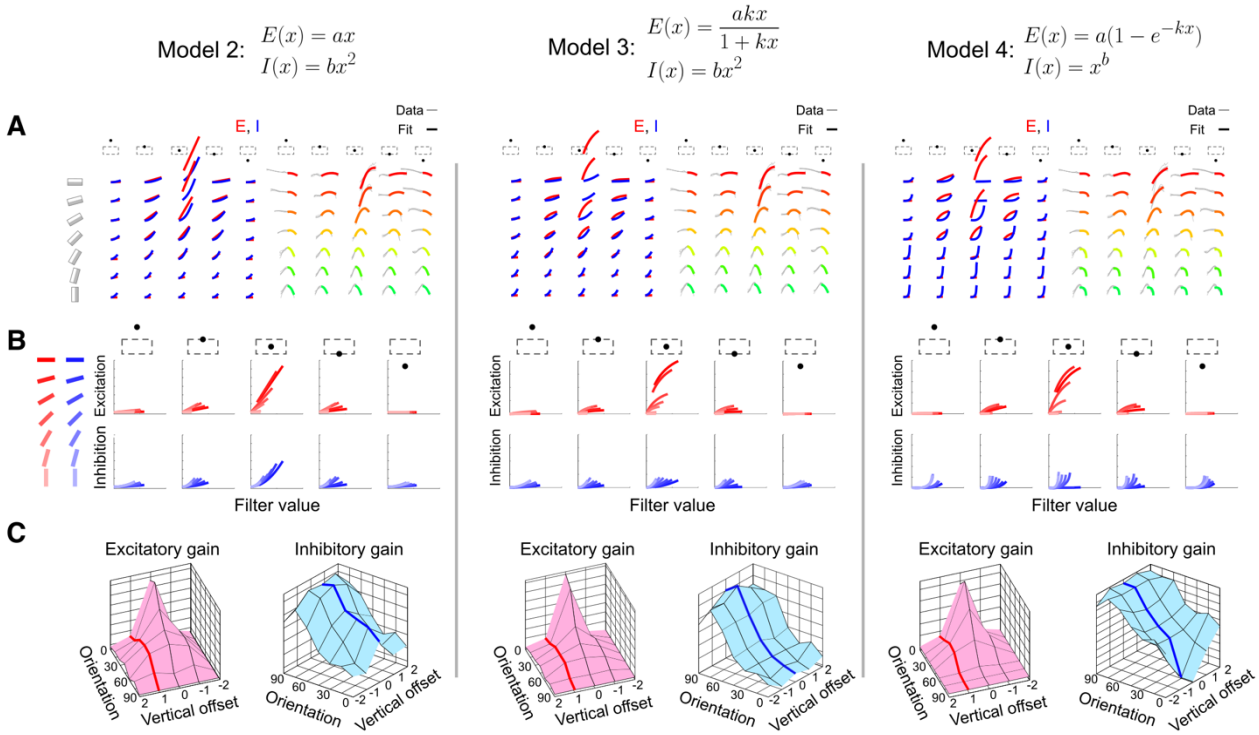
Model 2: $E(x) = ax$ $I(x) = bx^2$
Model 3: $E(x) = \frac{akx}{1+kx}$ $I(x) = bx^2$
Model 4: $E(x) = a(1 - e^{-kx})$ $I(x) = x^b$
A
B
C
E, I
Data
Fit
Excitation
Inhibition
Filter value
Excitatory gain
Inhibitory gain
Orientation
Vertical offset

Figure 6

Figure 6. Circuit-level predictions depend only weakly on the choice of parameters representing the excitatory and inhibitory component curves. *Related to Figure 5.* Three roughly similarly performing models are shown. (**A**) Excitatory (red) and inhibitory (blue) curve components (left) and resulting LLR fit (right) are shown for each model. Fit quality is comparable across all three models, and the original model shown in Figure 5C. (**B**) Despite having different E-I curve shapes, all three models show the same basic trends in the progression of excitation and inhibition as a function of orientation and vertical offset from the RL. (**C**) Summarizing each E and I curve with a single gain parameter shows a similar pattern for the three models.

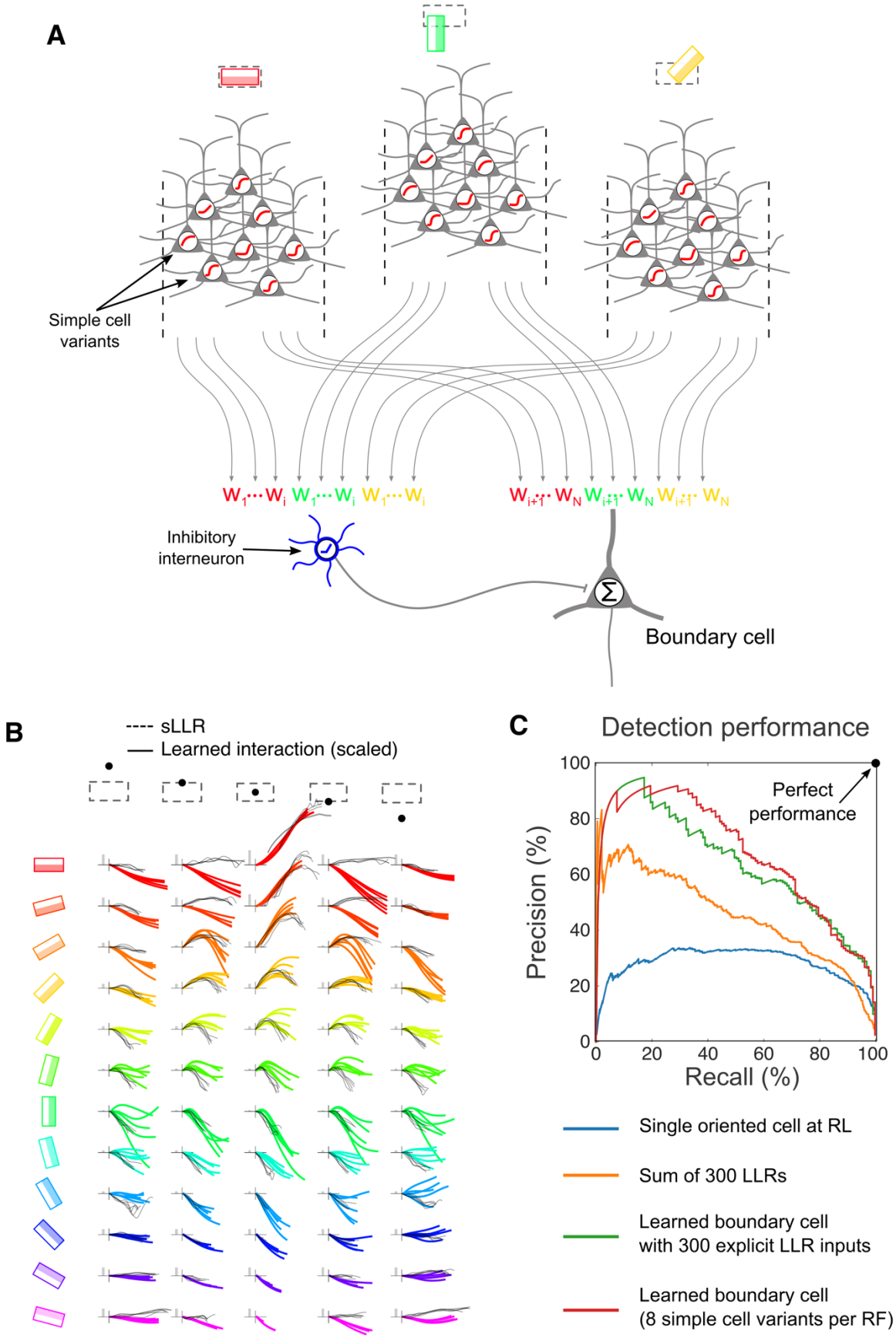
A
Simple cell variants
Inhibitory interneuron
Boundary cell
B
sLLR
Learned interaction (scaled)
C
Detection performance
Precision (%)
Recall (%)
Perfect performance
Single oriented cell at RL
Sum of 300 LLRs
Learned boundary cell with 300 explicit LLR inputs
Learned boundary cell (8 simple cell variants per RF)

Figure 7

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 i/o nonlinearity. Nonlinearities were sigmoids, $y = \frac{1}{1+e^{-g(\sum_i w_i x_i - t)}}$, 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 ith simple cell, η 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 η) to a learning rule which constrains all weights to be positive. (**B**) To determine the net effect of each filter on the boundary cell (for comparison to 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 8 associated simple cells was plotted (colored curves). Black dashed curves are averaged LLRs from Figure 2C. The gray bar in each plot represents the weight that the BC puts on that group of 5 colored curves (**C**) Precision-recall curves (on held out data) for the learned boundary cell (red) and weighted sum of LLRs (essentially the explicit Bayesian approach illustrated in Figure 1B and C, but with adjustable weights) (green) are very similar, indicating that the learned neural circuit behaves in accordance with the theoretical prediction. A pure sum of 300 filter LLRs is shown in orange. The lower blue PR curve shows that by comparison, a single oriented simple cell at the RL is a poor detector of natural object boundaries.

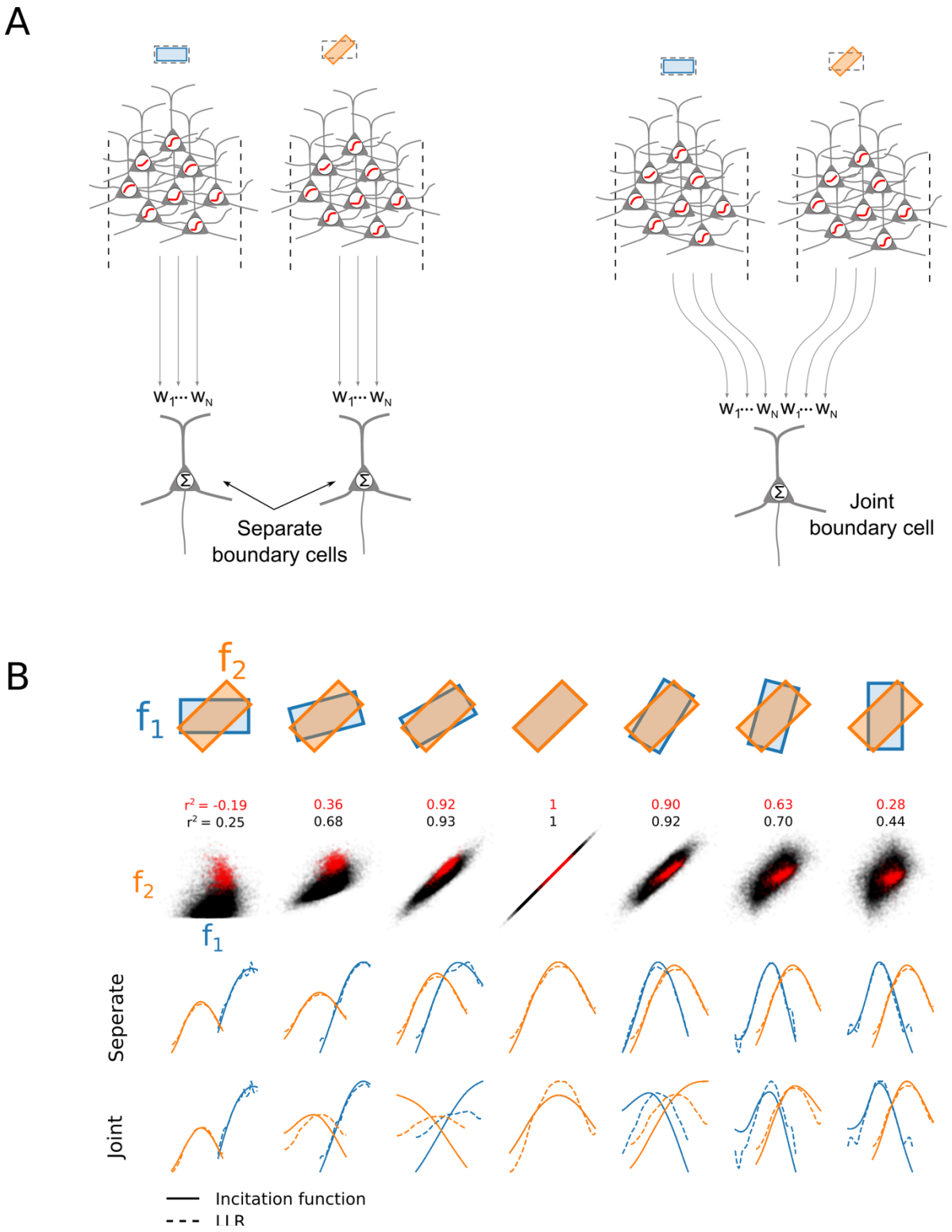
A
$w_1 \cdots w_N$
$w_1 \cdots w_N$
Separate
boundary cells
$w_1 \cdots w_N$ $w_1 \cdots w_N$
Joint
boundary cell
B
f_2
f_1
r^2 = -0.19
r^2 = 0.25
0.36
0.68
0.92
0.93
1
1
0.90
0.92
0.63
0.70
0.28
0.44
f_2
f_1
Seperate
Joint
Incitation function
LLR

Figure 8

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

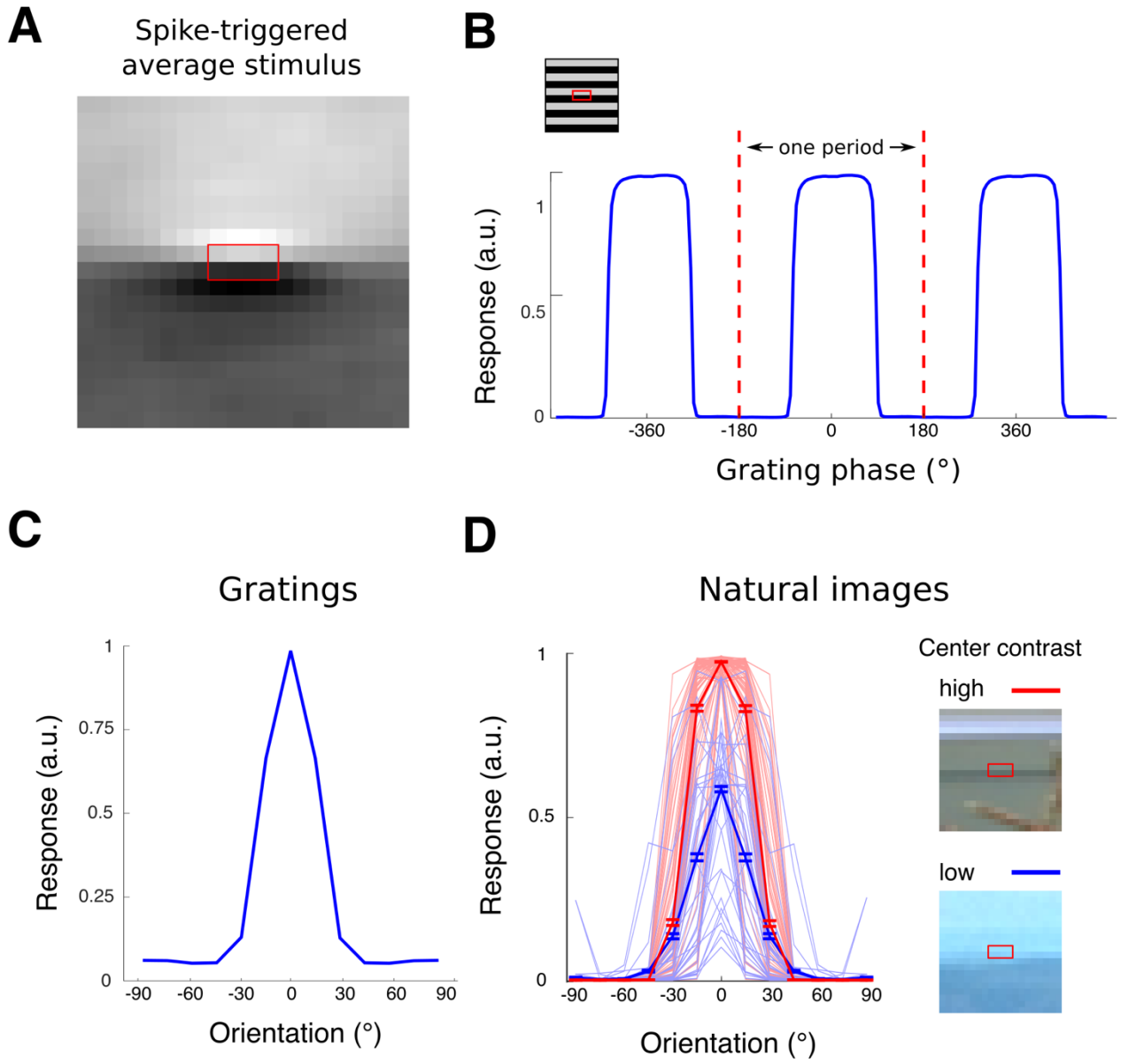
A
Spike-triggered average stimulus
B
one period
Response (a.u.)
1
0.5
0
-360
-180
0
180
360
Grating phase (°)
C
Gratings
Response (a.u.)
1
0.75
0.5
0.25
0
-90
-60
-30
0
30
60
90
Orientation (°)
D
Natural images
Response (a.u.)
1
0.5
0
-90
-60
-30
0
30
60
90
Orientation (°)
Center contrast
high
low

Figure 9

Figure 9. Boundary cell responses to parametric and natural stimuli resemble simple cell responses. To compute BC responses, the weighted sum of LLRs model (orange PR curve in Figure 7C) 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 phrases. The boundary cell is simple cell-like 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) is 43.6°, and for low contrast stimuli, 39.2°).

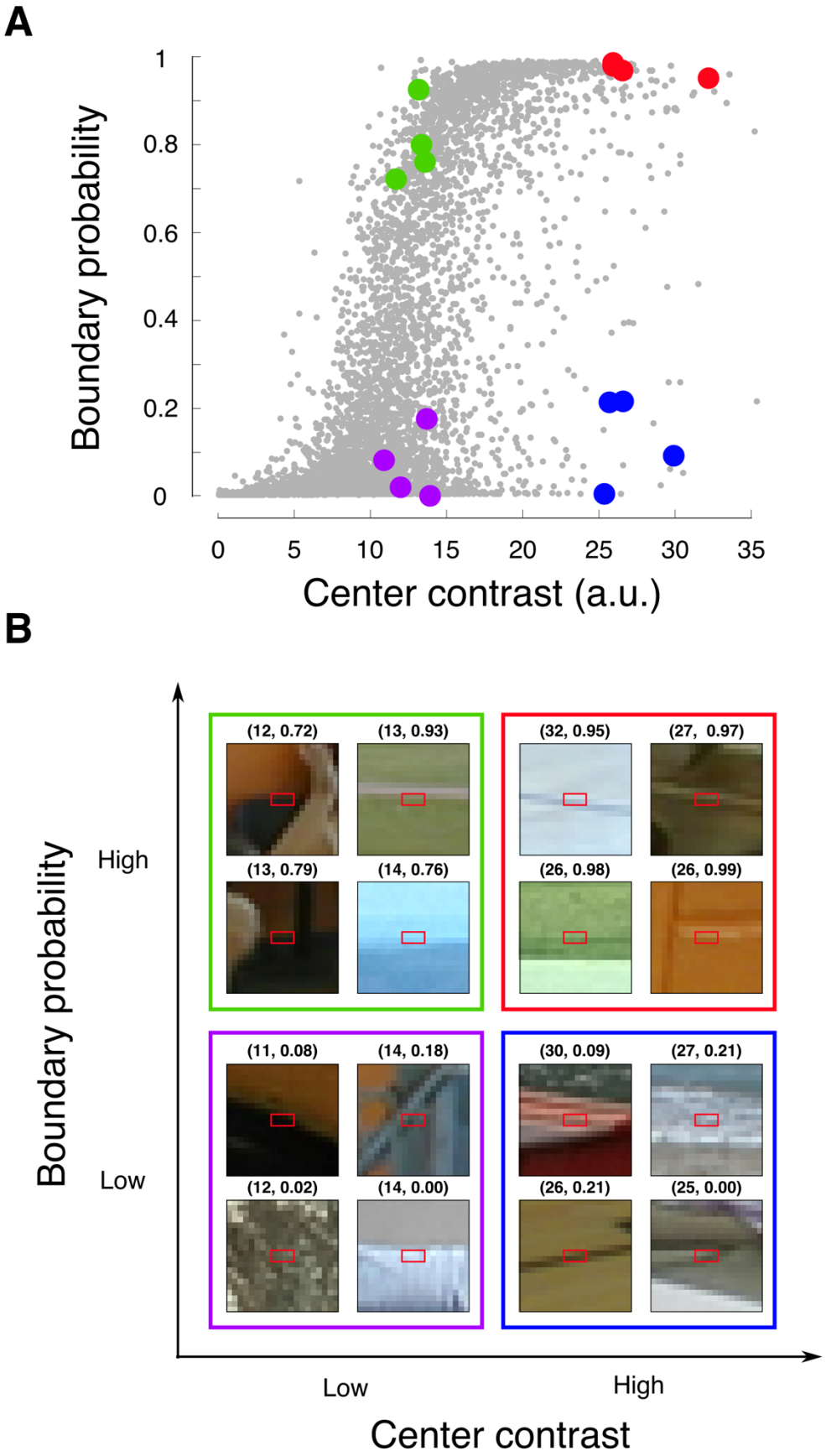
A
Boundary probability
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
Center contrast (a.u.)
B
Boundary probability
High
Low
(12, 0.72)
(13, 0.93)
(13, 0.79)
(14, 0.76)
(32, 0.95)
(27, 0.97)
(26, 0.98)
(26, 0.99)
(11, 0.08)
(14, 0.18)
(12, 0.02)
(14, 0.00)
(30, 0.09)
(27, 0.21)
(26, 0.21)
(25, 0.00)
Low
High
Center contrast

Figure 10

Figure 10. 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 which dissociates these two measures. Roughly 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 labelled patches shows that they are highly correlated, and that randomly selected patches are likely to lie at the lower left and upper 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 be better 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 non-edges (purple cases), high contrast non-edges (blue cases), low contrast edges (green cases), and high contrast edges (red cases).

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 features $\vec{x}$, in this case the responses of several simple cells. Logistic regression builds a model of $p(y|\vec{x})$ by assuming that the output probability is a sigmoid function of a linear combination of the features:

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

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

$$w^* = \text{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \hat{p}\,(y|\vec{x})$$

$$= \text{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \frac{\hat{p}(y|\vec{x})}{p(y|\vec{x})} - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln p\,(y|\vec{x})$$

The left term is the KL divergence between the true and model distributions, and the second term is constant with respect to the weights, and can be ignored. Our objective is then

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

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

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

Where $f(\vec{x}) = w^T\vec{x}$. Further, class conditional independence implies

$$p(y|\vec{x}) = \frac{p(y)p(\vec{x}|y)}{p(y)p(\vec{x}|y) + p(\bar{y})p(\vec{x}|\bar{y})} = \frac{1}{1+\frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}} = \sigma\left(-\log\frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}\right)$$

$$= \sigma\left(-\log\frac{p(\bar{y})}{p(y)} + \sum_i \log\frac{p(x_i|y)}{p(x_i|\bar{y})}\right) = \sigma\left(-\log\frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)\right)$$

so that the objective can be written

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

One can see by inspection that the two distributions will be equal and the objective will be minimized when

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

that is, when the classifier simply combines the filter values by passing them through their LLR functions and summing the result.

Now consider the problem of learning the optimal weights through gradient descent using a dataset of N input-output pairs. Call the ith input patch $x^{(i)}$ and the ith output label $y^{(i)}$. Label non-boundaries $y_i = -1$ and boundaries $y_i = 1$. The cost function, gradient, and Hessian can be written

$$J = \frac{1}{N}\sum_k \ln\left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)$$

$$\frac{\partial J}{\partial w_i} = \frac{1}{N}\sum_k \frac{-y^{(k)}x_i^{(k)}}{1 + e^{-y^{(k)}w^T x^{(k)}}}$$

$$\frac{\partial^2 J}{\partial w_i\, \partial w_j} = \frac{1}{N}\sum_k \frac{x_i^{(k)}x_j^{(k)}}{\left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)^2}$$

If we collect all of the input patches $x^{(k)}$ into the rows of the matrix X , the Hessian can be written simply as

$$\frac{\partial^2 J}{\partial w\, \partial w^T} = X^T D X$$

Where D is a positive diagonal matrix with entries $D_{kk} = \left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)^{-2}$. First, note that the Hessian is positive semidefinite since $v^T X^T D X v = \left|\left|\sqrt{D}Xv\right|\right|^2 \geq 0$ for all v. In fact, assuming that none of the K simple cells is expressible as a linear combination of the others (generically true if the input cells have distinct filter kernels) it is positive definite: X will have K linearly independent columns, implying that XX^T has K strictly positive eigenvalues. Because D is positive, $\sqrt{D}$ is also positive and has full rank, so $\sqrt{D}XX^T\sqrt{D}^T$ will also have K strictly positive eigenvalues. Each nonzero eigenvalue of $\sqrt{D}XX^T\sqrt{D}^T$ is also an eigenvalue of $X^T\sqrt{D}^T\sqrt{D}X = X^T D X$. Therefore, the Hessian $X^T D X$ has K strictly positive eigenvalues. But the Hessian has dimensions (K,K) , so it is positive definite. This implies the cost function J is strictly convex, and that gradient descent converges to the unique minimum which, given the argument above, is the LLR solution.