Unifying model for three forms of contextual modulation including feedback input from higher visual areas

Serena Di Santo[1], Mario Dipoppa[1], Andreas Keller[2], Morgane Roth[2], Massimo Scanziani[3], and Kenneth D. Miller[1]

[1]Center for Theoretical Neuroscience, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, NY 10027, USA.
[2]Institute of Molecular and Clinical Ophthalmology, Mittlere Strasse 91, CH-4031 Basel, Switzerland
[3]Department of Physiology, University of California, San Francisco, San Francisco, CA 94158-0444, USA

May 27, 2022

Abstract

Neural responses to a localized visual stimulus are modulated by the content of its surrounding. This phenomenon manifests in several forms of contextual modulation, including three interrelated properties of the visual cortex: surround suppression, inverse response and surround facilitation. We devise a unified biologically realistic circuit model accounting for all these phenomena and show that i) surround suppression in L2/3 is only partially due to the recruitment of lateral inhibition; ii) long-range feedback projections are necessary for inverse response and iii) the width of the response profile in the feedback layer determines inverse size tuning. The model predicts the modulations induced by silencing somatostatin-expressing cells or higher visual areas or changing the stimulus contrast. These predictions are consistent with the experimental observations when available and can be tested in existing setups otherwise. We then show the robustness of the identified mechanisms in a model with three interneuron subclasses, built to fit the classical responses and able to predict inverse size-tuning curves.

Keywords— contextual modulation, visual cortex, feedback, Higher Visual Areas, inhibitory subclasses

Highlights

- One model explains three different types of contextual modulation: (classical) surround suppression, (inverse) response to 'holes' in full field drifting gratings and cross orientation surround facilitation.
- Feedback, feedforward and lateral inhibitory inputs contribute to classical surround suppression in L2/3 of mouse V1 in different amounts.
- Observed responses to 'holes' in full field drifting gratings require long-range feedback projections.
- Surround modulation to the response to 'holes' in full field drifting gratings requires an increase in the characteristic length scale of the spatial pattern of activity in higher visual areas.
- The mechanisms uncovered by an analytically tractable model are also at work in a cell-type specific model that predicts response to a 'hole' stimulus.

Contents

1 Introduction **3**

2 Results **6**

2.1 A spatial schematic description preserving anatomical length-scales . 6

2.1.1 The widths of spatial response patterns grow much more slowly than the stimulus size 6

2.2 A minimal model for contextual modulation 7

2.3 Classical surround suppression in the minimal model 9

2.3.1 Feedforward, feedback and lateral SOM input shape classical surround suppression 10

2.3.2 SOM cells enhance classical surround suppression and generate more complex spatial profiles 11

2.4 Inverse response and size tuning properties of L2/3 are shaped by the feedback input 13

2.5 Inverse response requires wide enough feedback projections and feedback activity profiles scaling with stimulus size 15

2.6 A computational model trained to fit cell-type-specific recordings for classical stimuli predicts inverse responses 17

2.7 Surround facilitation can be explained by the inverse response to the orthogonal surround 19

3 Discussion **23**

1 Introduction

Objects in the visual scene rarely appear in isolation. Our visual perception of an object is influenced by the context in which it is presented. This phenomenon manifests in the visual system through the modulation of neural responses to a localized stimulus by the content of its visual surrounding. In the primary visual cortex (V1), a small stimulus elicits a response only if presented within a limited region of the visual scene, which is defined as the classical receptive field (CRF). When a stimulus within the CRF is presented conjointly with a stimulus in its surround (or extra-classical receptive field , eCRF), this surround can substantially influence the cell's response, an effect generally referred to as contextual modulation (see (Angelucci et al., 2017; Ziemba et al., 2018; Zipser et al., 1996) and references therein). The ability of the information at one location in the visual field to influence responses to stimuli at other retinotopic locations is likely to be critical to the parsing of objects and, more generally, the formation of a coherent perceptual representation of the visual scene (Field et al., 1993; Pak et al., 2020). We will define here three different forms of contextual modulation (Fig.1):

- *Surround suppression* (Allman et al., 1985; Angelucci et al., 2017; Hubel and Wiesel, 1965; Sceniak et al., 2001; Shushruth et al., 2009a; Spillmann et al., 2015): the response of a cell whose CRF is centered on a stimulus decreases when a similar stimulus is presented in the surround.
- *Inverse response:* when a large drifting grating is shown, and a 'hole' (a uniform patch of mean luminance) of varying sizes is imposed, centered on the CRF center (an "inverse stimulus"), neurons show responses and size tuning very much like their responses and size tuning to patches of drifting gratings (a "classical stimulus")(Keller et al., 2020b).
- *Surround facilitation*: the response of a cell whose receptive field is centered on a stimulus increases when an orthogonal stimulus is presented in the surround (Cavanaugh et al., 2002; Keller et al., 2020a; Paffen et al., 2005; Self et al., 2014; Xing and Heeger, 2001).

Modulation by a surround stimulus has been observed in a variety of species, cortical areas and sensory modalities, suggesting the relevance of contextual modulation as a principle of cortical computation (Cavanaugh et al., 2012, 2002; Chou et al., 2020; Falkner et al., 2010; Keller et al., 2020a; Petro et al., 2017; Schwartz and Simoncelli, 2001; Shushruth et al., 2009b; Sundberg et al., 2009; Tanaka et al., 1987; Tsui and Pack, 2011; Van den Bergh et al., 2010; Yu and Levi, 2000). Inverse responses have thus far only been observed in layers 2/3 (L2/3) of mouse V1 (Keller et al., 2020b) and are substantially reduced if higher visual areas (HVAs) are selectively suppressed through optogenetic techniques, suggesting that feedback connections play a key functional role.

The concept of CRF can be roughly understood based on simple anatomical observations considering connectivity structures that feed forward localized stimulation of the retina (Angelucci et al., 2002; Troyer et al., 1998). However, a

mechanistic understanding of the eCRF's effects is more complex and requires taking into account lateral (recurrent) excitatory and inhibitory inputs (Adesnik et al., 2012; Angelucci et al., 2017; Blakemore and Tobin, 1972; Keller et al., 2020a; Obeid and Miller, 2021b; Rubin et al., 2015) as well as feedback from HVAs (Angelucci et al., 2017; Nurminen and Angelucci, 2014). In particular, suppression of HVAs somewhat decreases surround suppression in V1 of primates (Nassi et al., 2014; Nurminen et al., 2018) and mice (Keller et al., 2020b; Vangeneugden et al., 2019), but largely eliminates inverse responses in mouse V1 (Keller et al., 2020b).

Because of its multifaceted nature, a uniform mechanistic understanding of V1 contextual modulation is still missing.

Mouse V1 provides an extraordinary paradigm to study circuit mechanisms of cortical computation across stimuli and brain states. Thanks to extensive experimental efforts targeting this brain area, an unprecedented quantity of simultaneous and cell-type-specific recordings is now available for analyses. Detailed anatomical measures such as the spatial range of connectivity among different cells have been established within reasonable error margins (Billeh et al., 2020; Campagnola et al., 2021; Rossi et al., 2020) (see Supplementary Section S6). The physiology of contextual modulation has been described under a variety of stimulus conditions.

A number of theoretical works, at different levels of abstraction, have been developed to account for various aspects of contextual modulation. Phenomenological models place it within a broad range of nonlinear response properties described as "divisive normalization" (*e.g.*, Carandini and Heeger, 2012; Carandini et al., 2005; Coen-Cagli et al., 2015). Circuit models have addressed various aspects of contextual modulation, though none have previously addressed inverse tuning. These models (*e.g.*, Dipoppa et al., 2018; Keller et al., 2020a; Li and Young, 2021; Mossing et al., 2021; Obeid and Miller, 2021b; Rubin et al., 2015; Schwabe et al., 2006, 2010; Shushruth et al., 2012b) differ as to whether they treat space discretely (*i.e.* two local circuits, one representing center and one surround) or continuously as a 2D plane of cells; as to which contextual modulation phenomena they address; as to what degree they incorporate constraints from biological measurements, *e.g.* of the anatomical spreads of connections; as to the degree to which they attempt to explain responses qualitatively vs. quantitatively; as to the number of cell types modeled; and as to the mechanisms proposed to underly observed responses.

Here we develop a unified model of the three contextual modulation phenomena described above in L2/3 of mouse V1. We consider a model of the four cortical cell types that have been prominently studied in recent years: excitatory or pyramidal (Pyr) and parvalbumin-expressing (PV), somatostatin-expressing (SOM) and vaso-active-intestinal-peptide-expressing (VIP) inhibitory interneurons. We study a model of continuous space, which allows us to model the cellular response patterns across space and to determine how the geometry of spatial connections contributes to contextual modulation. We incorporate experimental constraints on the spatial extents of connections (Billeh et al., 2020; Marques et al., 2018; Rossi et al., 2020) and data from Keller et al. (2020b) on the response profiles across space of each cell type to each stimulus type. We focus on qualitatively explaining response properties and understanding their underlying mechanisms, pinpointing the multiple

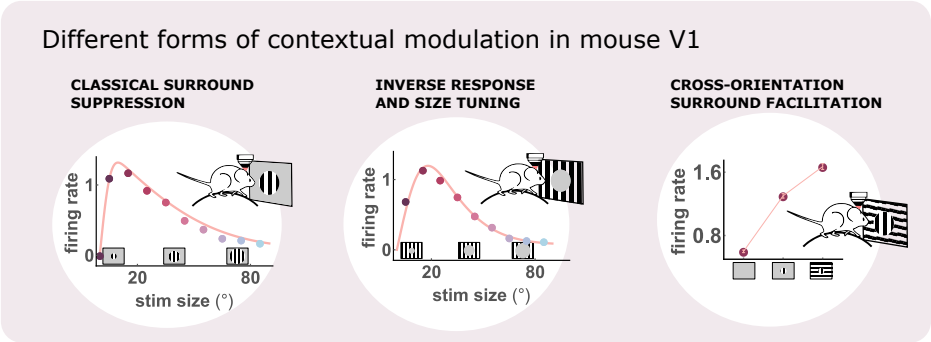

Figure 1: **Contextual modulation takes different forms in mouse V1 a**. Classical surround suppression. Response of Pyramidal cells as a function of stimulus size, also called size-tuning curve. The full circles are data points from the dataset in (Keller et al., 2020b) and the full line is a fit. The response decreases when the stimulus (a patch of drifting grating) size is increased beyond the CRF size. **b**. Inverse response and size tuning. Same plot as in a. Inverse response indicates the response of Pyramidal cells in L2/3 (but not L4) to a 'hole' or inverse stimulus in the CRF, i.e. a patch with zero contrast on a full field drifting grating. The responses also show tuning to the size of the hole similar to the classical size tuning. **c**. Cross-orientation surround facilitation. Response of Pyramidal cells to a patch of drifting grating are smaller than responses to a cross stimulus, i.e. a patch of drifting grating surrounded by a drifting grating at an orthogonal orientation. Data replotted with permission from (Keller et al., 2020a).

factors that contribute to contextual modulation and showing how these can be leveraged to generate testable predictions.

To isolate basic principles, we begin by studying an abstract, "minimal" non-linear model of a single recurrently-connected cell type that can be fully solved analytically. This allows us to break down the features of the inputs that contribute to direct and inverse size-dependent responses. Successively we build up to a full model of the four cell types to bridge the gap with data and test that the mechanisms at play are the same as the ones uncovered through the minimal model. We first apply this approach to understand both classical and inverse size tuning. We then show how the model can be seamlessly extended to include orientation tuning. This allows us to account for more complex stimuli that include patches of different orientations, and in particular to extend our model to address cross-orientation surround facilitation.

2 Results

2.1 A spatial schematic description preserving anatomical length-scales

As pointed out by Dipoppa et al. (2018) and Keller et al. (2020b), to understand surround modulation, it is crucial to take into account the input from neurons with CRFs distributed across visual space. This requires knowledge of two factors: i) the dependence of the mean effective strength of synaptic connections on the distance in visual space between the CRF centers of the pre- and post-synaptic cells and ii) the responses of cells to a stimulus as a function of the distance between the cell's CRF and the stimulus center. For the first, we assume that synaptic strengths decrease with distance between two neurons as a Gaussian function, with length scales taken from experimental measurements (Billeh et al., 2020; Keller et al., 2020b; Marques et al., 2018; Rossi et al., 2020) (see Supplementary Section S6). We take this approach, rather than trying to infer connections that lead to responses that fit the data (*e.g.*, Dipoppa et al., 2018), to more tightly constrain our models, and to focus more on general and qualitative rather than fine-tuned quantitative behaviors. We show that these qualitative results are robust when considering very large relative errors in the experimental estimates (see Supplementary Section S18 C). For the second, we use the dataset of Keller et al. (2020b) and a smoothing method introduced in Dipoppa et al. (2018) to reconstruct the firing-rate field (firing rate vs. CRF spatial position) of Pyr, PV, SOM and VIP cells in L2/3 of V1, as well as Pyr cells in L4 of V1 and in the higher visual area LM (Fig.2, see also Supplementary Section S1 A). We will also describe these firing-rate fields as Gaussian (or linear combination of Gaussian) functions centered on the stimulus center. Describing both spatial connectivity and firing rate fields with Gaussian functions of distance simplifies the model, reducing the number of parameters (Fig.2).

2.1.1 The widths of spatial response patterns grow much more slowly than the stimulus size

The rate fields elicited by classical stimuli of varying sizes estimated from the recordings in (Keller et al., 2020b) are reported in Fig.2(see also Supplementary Section S1 A). We plot the rate fields for a joint population of Pyr and PV cells, because their responses are qualitatively very similar and for better comparison with the model (see Section 2.2). These are fully specified by only two quantities: the amplitude and the width (σ parameter, or standard deviation) of the Gaussians. The amplitude represents the response of neurons that are centered on the stimulus and, when considered as a function of stimulus size, it is also called the size tuning curve. The Gaussian widths describe the geometry of the spatial patterns of activity, which, to the best of our knowledge, is systematically discussed here for the first time. Intriguingly, but consistent with plots in (Dipoppa et al., 2018), the growth of the width of the rate field with increasing stimulus size is much weaker

than expected if cells simply responded when their CRF is covered by a stimulus. In that case the rate field's width would grow at the same pace as the stimulus width (notice that for consistency with previous literature we define the stimulus size s as the diameter of the patch of drifting grating or the diameter of the hole for inverse stimuli, see also Supplementary Section S1 C). The assumption that rate field width follows stimulus width is at the core of existing theoretical models of surround suppression (Angelucci et al., 2017; Dipoppa et al., 2018; Li and Young, 2021; Obeid and Miller, 2021a; Rubin et al., 2015), in which larger stimuli recruit more wide-ranging lateral activation, and these activated lateral cells have a suppressive influence on the center. Our analysis of the dataset in (Keller et al., 2020b) reveals that the growth of the width of the rate field with the stimulus size is roughly linear, but has a slope much smaller than the stimulus radius (Fig.2). This feature is especially salient for L4, for which the width of the rate field is roughly constant (see also Supplementary Section S1 A). In what follows we refer to this behavior as the weak scaling of the rate fields with the stimulus size.

In what follows we first develop a model for ordinary surround suppression: tuning for the size of a drifting grating, with response suppressed for size increasing beyond an optimal size (Allman et al., 1985; Angelucci et al., 2017; Hubel and Wiesel, 1965; Sceniak et al., 2001; Shushruth et al., 2009a; Spillmann et al., 2015). We will then apply this model to understand inverse response: similar tuning for the size of a circular patch of zero contrast (gray color) within a large drifting grating (Keller et al., 2020b). Finally, we will combine these models to understand cross-orientation surround facilitation: facilitation of response by a drifting grating in the surround with orientation perpendicular to that of the drifting grating in the CRF center (Keller et al., 2020a).

2.2 A minimal model for contextual modulation

Here we develop a minimal model with one recurrent population disentangling the factors contributing to contextual modulation (Fig.3a). The aim of this model is to address the question: how do the properties of feedforward, feedback and lateral input currents contribute to contextual modulation in L2/3 of mouse V1? In this minimal model, there is only a single recurrently connected cell type. This cell type represents the combination of Pyr and PV cells, based on the observations that PV interneurons track the activity of Pyr cells in multiple contexts and over a wide dynamic range of inputs(Dipoppa et al., 2018; Keller et al., 2020a,b; Mossing et al., 2021; Palmigiano et al., 2020), thereby dynamically balancing (preventing an excess of) excitation and functioning as a stabilizer for the network (Adesnik, 2017; Ahmadian and Miller, 2021; Denève and Machens, 2016; Dipoppa et al., 2018; Isaacson and Scanziani, 2011; Kato et al., 2017; Okun and Lampl, 2008; Ozeki et al., 2009; Palmigiano et al., 2020; Renart et al., 2010; Sanzeni et al., 2020). In contrast, VIP and particularly SOM cell responses generally differ substantially from those of Pyr and PV cells(Dipoppa et al., 2018; Keller et al., 2020a,b; Mossing et al., 2021; Palmigiano et al., 2020). We will treat SOM cells, responding across space

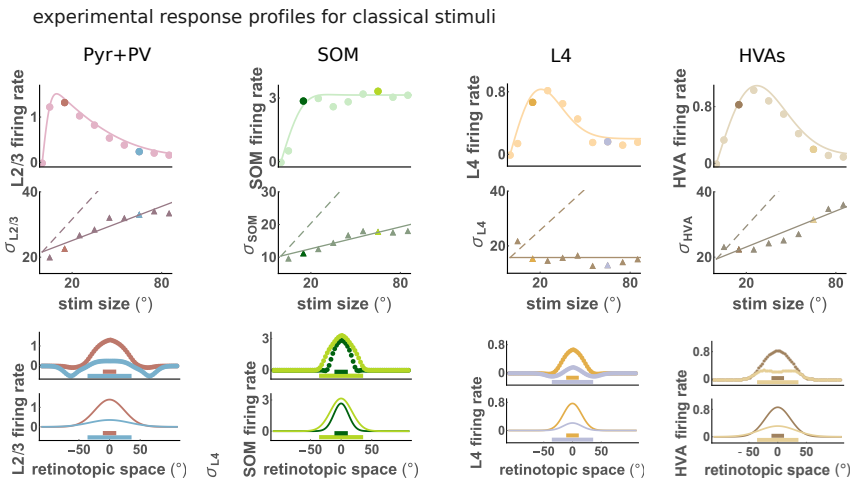

Figure 2: **The widths of spatial response patterns grow much more slowly than the stimulus size**. Pyr+Pv, SOM, L4 and HVAs spatial responses, calculated from the dataset (see Supplementary Section S1 A). Top panels: response as a function of the stimulus size (size tuning curve). Full circles are the data points, full line is a fit. Center panels: Full triangles are the widths of the best Gaussian fit to the rate field for each stimulus size, the full line is a linear fit and the dashed line represents the case in which the width of the rate field grows as the radius of the stimulus. Symbols of different colors with respect to the rest indicate the sizes for which we plot the firing rate field in the bottom panels (same color code as the top panels). Bottom panels: experimental rate field (top) and best Gaussian fit (bottom) for stimulus sizes $s = \{15°, 65°\}$. The bars represents the size of the corresponding stimuli (same color code).

to each stimulus as measured experimentally, as a static external input in this one-population model, and ignore VIP cells, which act primarily on SOM cells (Billeh et al., 2020; Campagnola et al., 2021; Pfeffer et al., 2013). We also take the measured responses across space of layer 4 (L4) and of the Lateromedial area (area LM), which represents HVAs, as external inputs to the model.

We assume neurons have a supralinear (rectified quadratic) input/output function, which describes the steady-state firing rate induced by a given net input to the neuron. Such an expansive input/output function is expected when firing activity is driven by input fluctuations rather than by the mean input (Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002), and has been shown to yield network behavior that reproduces a variety of nonlinear visual cortical behaviors that have often been summarized phenomenologically as "divisive normalization" (Carandini and Heeger, 2012), including surround suppression and its nonlinear dependence on stimulus contrast (Ahmadian and Miller, 2021; Ahmadian et al., 2013; Rubin et al., 2015). The use of these power-law input/output functions, combined with our assumptions that connectivity and firing-rate fields are described by Gaussian functions, allow us, building on previous theoretical work (Persi et al., 2011), to develop an approximate analytic solution to the one-population model, which yields deeper insight into the mechanisms driving model behavior.

A crucial aspect of the model is to include inputs from HVAs. One difficulty in the interpretation of the experimental results derives from the fact that cells in L2/3 are targeted by all of the HVAs (Billeh et al., 2020; Keller et al., 2020b; Siegle et al., 2021), but we have recordings of the V1 boutons from only one HVA, area LM (Keller et al., 2020b). Extended Data Fig.9 in Keller et al. (2020b) shows that optogenetic silencing of individual HVAs generates a quantitatively different reduction of classical versus inverse responses for each HVA considered. Therefore we take the minimal assumption that the HVAs considered all together have a rate field profile that is proportional to the one recorded in LM, but we allow the proportionality constant to be different in the classical and inverse stimulus condition for more flexibility. We infer these constants based on the amplitudes of the inverse vs. classical response in L2/3 (see Supplementary Section S2).

2.3 Classical surround suppression in the minimal model

In Fig.3 we report results of the minimal model for classical stimuli of increasing size. The inputs to L2/3 are given by the convolution of the rate fields of L4 and HVAs (and SOM when specified), which were derived from the data (Fig.2), with the corresponding Gaussian connectivity (Fig.3b). The recurrent rate fields resulting from the analytical solution and the simulation of the minimal model are in good agreement (Fig.3c), thus in what follows we rely on the analytical formulas to efficiently compute the results of the model (for a comparison of the model results with the experimental rate field of Pyr+PV cells see Fig.2 and Supplementary Section S4). The free parameters of the model are the feedforward, feedback and recurrent connection strengths. Since we consider the recurrent population to be a joint population of Pyr and PV cells, the recurrent connection strength W_0 can

be either positive or negative, depending on whether Pyr or PV dominates. In Supplementary Section S9 B we show that L2/3 responses are unstable if the net input is excessively dominated by L2/3 Pyr cells (i.e. the recurrent connection strength is positive and large). The feedforward and feedback connection strenghts were set to reproduce the amplitude of the response in L2/3. In Supplementary Section S9 C we show how the response of L2/3 changes as a function of the free parameters.

2.3.1 Feedforward, feedback and lateral SOM input shape classical surround suppression

A large body of experimental evidence shows that feedforward, horizontal, and feedback connections all contribute to classical surround suppression in V1 (see for instance (Angelucci et al., 2017) and references therein). The minimal model provides a scaffold to analyze the interplay of the different components that shape classical surround suppression. In fact, we can write analytically the form of the recurrent rate field and its derivative with respect to the stimulus size, a negative derivative indicating surround suppression. In Supplementary Section S7 we show that surround suppression in the recurrent layer can be produced by either or both of: 1) surround suppression in the input population and 2) broadening of the input rate field as stimulus size is increased. These two components induce surround suppression by two different mechanisms: in case 1) L2/3 responses for large stimuli are lower due to a smaller external input, that is, the suppression is inherited from the input (although it may be enhanced by L2/3 network mechanisms that increase response differences relative to input differences; and in case 2) they are lower because external excitation recruits more lateral activity, causing greater suppression of the center response. The fact that the scaling of the widths is weak for LM and absent for L4 suggests that the second mechanism contributes only marginally to surround suppression in L2/3.

Figure3d shows the effects of (counterfactually) eliminating surround suppression in L4 and HVAs (top and center row) as well as the effect of eliminating the growth of the input rate field from HVAs as stimulus size is increased (bottom row; note the width of the L4 rate field does not increase with stimulus size). As expected from the relatively weak increase in HVAs rate field width with stimulus size, this latter modification produces only a marginal change in surround suppression, predicting that surround suppression in L2/3 is mostly inherited from its inputs.

In Supplementary Section S9 A we compare the recurrent rate field obtained from a supralinear vs a linear transfer function and we show that the former supports larger surround suppression, conceivably because the nonlinearity increases the gap between responses to stronger stimuli and responses to weaker stimuli. In addition, in Supplementary Section S8 we analyze the contrast dependence of size tuning in the model. We show that a supralinear one-population model with Gaussian input whose width increases with stimulus size can generate the observed contrast-dependent size tuning, i.e. higher contrasts produce larger surround modulation and smaller preferred sizes, in line with the results in (Rubin et al., 2015). However, when the connectivity length scales are matched with the anatomically

relevant ones, those elements are not sufficient. In Supplementary Section S10 we reason that the decrease of the preferred size as contrast increases is i) partially inherited from the feedforward input, ii) possibly related to the scaling of sigma in the feedback input and iii) supported by large VIP/small SOM response for low contrasts (in agreement with (Mossing et al., 2021)).

2.3.2 SOM cells enhance classical surround suppression and generate more complex spatial profiles

In this section we investigate the role played by the lateral inhibitory input from SOM neurons in classical surround suppression.

The rate field of SOM neurons in L2/3 shows a lack of surround suppression (see Supplementary Section S1 D), in agreement with previous results (Adesnik et al., 2012). However, a minor amount of surround suppression has been reported in (Dipoppa et al., 2018) for SOM cells whose CRF distance from the stimulus center is very small. It is possible that the precision in the measure of CRF mapping in the dataset of (Keller et al., 2020b) may not suffice to resolve this small amount of surround suppression.

Inspired by experiments hyperpolarizing SOM cells in (Adesnik et al., 2012), we compare the size tuning curve of the model recurrent population in the presence or absence of input from SOM interneurons (see Fig. 3e, and Fig.4h in (Adesnik et al., 2012)). Since SOM cells prefer large stimulus sizes, they inhibit L2/3 the most for large stimuli, thus enhancing classical surround suppression of L2/3.

This effect is robust against changes in the free parameters, i.e. the connection strengths. However, when the strength of SOM projections onto the Pyr+PV population is large enough, the model Pyr+PV rate field shows a "ring" shape for large stimulus sizes, which is well captured by the analytic solution of the minimal model (see Fig.3e). This ring shape, appearing for large stimulus sizes, has been reported in Dipoppa et al. (Fig. S11, panels D1 and D2 of 2018) and is replotted for the joint population of Pyr and PV in Supplementary Section S4 A).

The absence of this feature in the dataset in (Keller et al., 2020b) (see Fig. 3d) could conceivably be explained if those animals had a smaller effective strength of SOM projections than in (Dipoppa et al., 2018).. When the "ring" shape appears, the neurons whose CRF is aligned with the stimulus center are more suppressed than the ones that are offset (only partially surrounded by the stimulus). From a geometric point of view, SOM input to Pyr+PV units is a relatively narrow Gaussian-shaped bump. For large stimuli its amplitude becomes larger with respect to the amplitude of all other input currents, generating a concave Pyr+PV rate field However, the input from HVAs can also contribute to this effect, since for large stimuli it also exhibits a ring profile (see Fig.2 and also Supplementary Section S1 D).

In conclusion, our model predicts that classical surround suppression in L2/3 is partially inherited from surround suppression of its feedforward (L4) and feedback (HVAs) inputs (consistently with the experimental literature (Angelucci et al.,

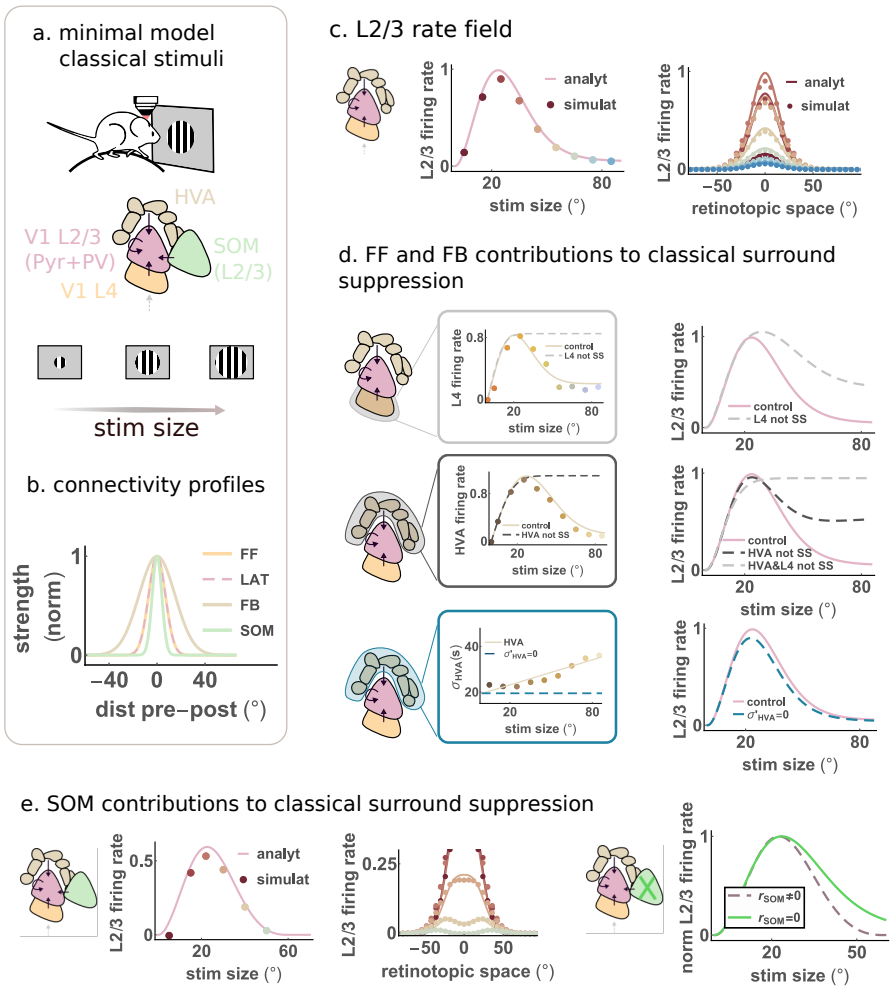

Figure 3: **Feedforward, feedback and lateral SOM inputs shape classical surround suppression.** **a**. From top to bottom, sketches for: experimental setup with classical stimuli (Keller et al., 2020b); minimal model with one recurrent cell type in L2/3 of V1 receiving feedforward input from L4, feedback input from HVAs and lateral input from SOM cells; classical size tuning stimuli. **b**. Spatial dependence of the strength of the connections on the CRF distance between pre- and post-synaptic neurons, for modeled feedforward, lateral effective recursive connections of the joint population, feedback and lateral connections from SOM cells.

Figure 3: **c**. Size-tuning curve (left) and rate field (right) for the recurrent layer of the model. Here and in what follows the color code of the full circles indicates stimulus size. From hereafter, in plots that show results from the model, full circles represent simulations and full lines represent analytical results. **d**. Counterfactual modifications of the inputs. The left panels define the counterfactual modification and the same color and dashing in the right panels shows its effect on the recurrent population, together with the control condition (light red). From top to bottom we eliminate in each case: surround suppression in L4, surround suppression in HVAs or the growth of the rate field of HVAs. The middle panel also shows the case in which we eliminate surround suppression both in HVAs and in L4. **e**: Left: Same as c but now adding input from SOM neurons in L2/3. Compare with experimental profiles reported in Supplementary Section S4 A. Right : Size tuning curve of the recurrent layer with (dashed) and without (green; same as in panel c) input from SOM neurons. SOM neurons increase surround suppression, consistently with (Adesnik et al., 2012) and in particular Fig.4d ibid.

2017; Vangeneugden et al., 2019)), marginally enhanced by lateral (PV) inhibition progressively recruited as the stimulus size increases (Obeid and Miller, 2021a; Rubin et al., 2015), enhanced by SOM lateral input (Adesnik et al., 2012) and partially supported by the supralinear input/output function, which derives from the noisy nature of the neural dynamics (Miller and Troyer, 2002).

2.4 Inverse response and size tuning properties of L2/3 are shaped by the feedback input

In the previous Sections we developed the minimal model, with most parameters set from experimental data and the connection strengths set in a range that matches classical surround suppression. Here we use the same model, with no adjustment of parameters save changing the amplitude of the HVAs response to inverse stimuli (as described in Section S2), to reproduce and understand the response patterns for inverse stimuli as a function of stimulus size, i.e. the diameter of the 'hole' (Fig.5a). In the data, the spatial representation of the inverse stimulus in the feedforward and feedback layers differs significantly from its representation in L2/3 (Figure4, see also Supplementary Section S1 C and S4). In particular the firing rate of cells in L4 and HVAs whose CRF is aligned with the stimulus center is very low, which lead the authors of (Keller et al., 2020b) to conclude that the inverse tuning properties of L2/3 are not directly inherited from its inputs.

To leverage our analytical framework, we parameterize the input rate fields with a difference of Gaussian functions whose parameters vary continuously with stimulus size (see Supplementary Sections S1 C, S1 D). In doing so, we trade-off

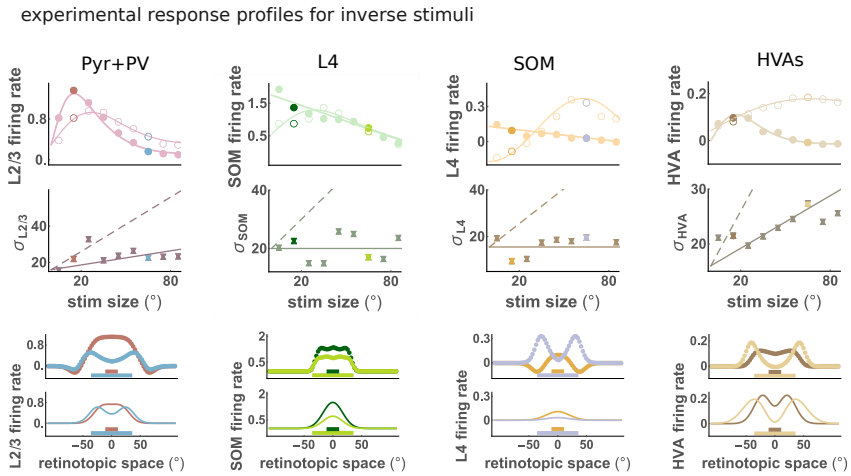

Figure 4: **Inverse response and size tuning properties of V1 and HVAs**.Same as Fig.2 but for inverse stimuli. Left panel of each cell class: response of the centered and offset cells as a function of the stimulus size (size tuning curve). Full circles are the data points for centered cells, empty circles are the data points for cells with an offset of 30°, full lines (lighter color) are fits to the aligned and offset size tuning curves. Full up triangles and down triangles are the widths of the best difference-of-Gaussian-fit to the rate field. The difference in the widths is always small (the triangles are almost overlapping). The full line (darker color) is a linear fit ans the dashed line represents the case in which the width of the rate field grows as the radius of the stimulus. Right: experimental rate field (top) and best difference-of-Gaussians fit (bottom) for two stimulus sizes highlighted in the left panel with symbols of the same color code. The bars represents the size of the corresponding inverse stimuli (same color code).

some details of the spatial profiles of the responses, to gain interpretability. For instance the growth of the profile of the input rate fields with stimulus size is encoded in one parameter (since the difference of the widths of the two Gaussian is kept constant and fixed by the recordings see Supplementary Section S1 C and Fig.4). Similarly to the classical stimulus condition, in HVAs this quantity scales weakly with the stimulus size (see Supplementary Sections S1 D). In Supplementary Section S3 we show that for intermediate sizes HVAs cells respond maximally when their receptive field is close to the edge of the hole.

Fig.5b shows that the minimal model reproduces the spatial profiles of inverse response as a function of size: centered cells show high response to small inverse stimuli, while for large inverse stimuli offset cells show a large response and centered cells respond less. This gives rise to inverse size tuning. This result leads to two observations: i) the spatial features of the input populations' response are sufficient for the emergence of inverse response and inverse size tuning, ii) the parameterization chosen captures the relevant features of the inputs: the ring profile (difference of Gaussians) and its growth in width as a function of the stimulus size.

Next we investigate the role that SOM neurons play in inverse size tuning. Hyperpolarizing SOM neurons in the model increases inverse surround suppression – the opposite of the effect seen in the classical case (Fig.5c, and see Fig.3d for comparison). This can be understood by looking at the response profiles of SOM cells (see Supplementary Section S1 D): for large inverse stimulus sizes, their response is lower, thus they inhibit Pyr and PV cells less. This is a testable prediction of the minimal model and could be falsified by an experimental setup analogous to the one employed in (Adesnik et al., 2012). Inverse response is substantially reduced when input from HVAs is reduced by a factor 0.65 (Fig.5d), consistent with experimental results on optogenetic silencing of HVAs (see Fig.5c in (Keller et al., 2020b)). Finally, the model shows an increase of inverse responses with contrast (Fig. 5e) similar to that found experimentally Keller et al. (Extended Data Fig. 3h of 2020b). In the model, the degree of suppression of inverse response with increasing stimulus size may either strengthen or stay the same with increasing contrast, depending on precise parameters (not shown).

2.5 Inverse response requires wide enough feedback projections and feedback activity profiles scaling with stimulus size

Here we demonstrate how our analytically tractable model guides us in gaining mechanistic understanding of the inverse response properties. For a fixed stimulus size ($s = 15°$, Fig.6a) we investigate the role played by the width of the projections from HVAs to L2/3, $\sigma_{L23-HVA}$, by varying its value around the experimental estimate (Supplementary Section S6, see also Fig 6b). The rate field of the inputs is kept constant. In the dataset, the L2/3 rate field in response to small inverse stimuli is peaked at its center, *i.e.* the cells aligned with the hole have the strongest response. Fig.6c highlights that in this simplified setup, a necessary condition for this feature to appear is that the span of the feedback projections from HVAs is

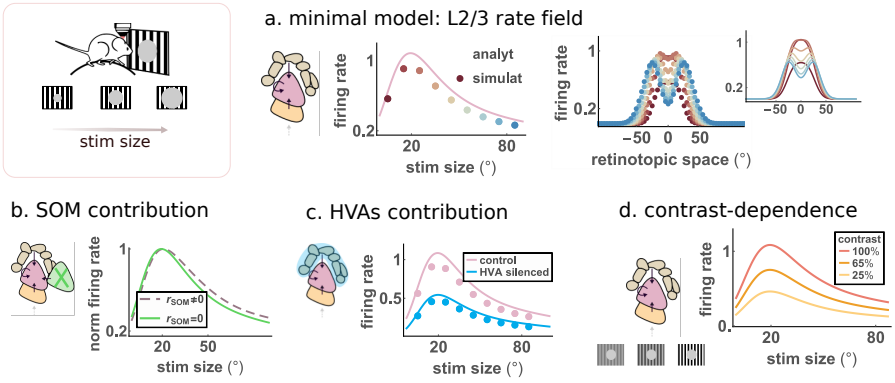

Figure 5: **Inverse response and size tuning properties of L2/3 are shaped by its feedback input**. **a**. Size-tuning curve (left) and rate fields (right) for the recurrent layer of the model. The size of the inverse stimulus is the diameter of the hole. For rate fields, results for the simulations (dots) and the analytics (curves)are represented separately for better readability. The model recovers inverse surround suppression as well as the spatial profile of responses. **b**. Same as Fig.3d for inverse stimuli. SOM neurons decrease inverse surround suppression. **c**. Size tuning curve in the control case (light red) and for silencing of HVAs (the rate field of HVAs is reduced by a factor 0.65). **d**. Size tuning curves for varying contrast (both L4 and HVAs rate fields are reduced by a factor $c = \log_2 2C_0$. C_0 values shown in the legend).

large enough, i.e. the CRF of HVAs and the one of their targets in L2/3 is offset enough, as quantified in the following paragraph.

To quantify the range of values of $\sigma_{L23-HVA}$ compatible with the inverse response, we compute the second derivative of the recurrent rate field in the center as a function of $\sigma_{L23-HVA}$ (Fig.6c left). A negative second derivative denotes a maximum in the center, characteristic of inverse response observed at small stimulus sizes. Fig.6c shows that $\sigma_{L23-HVA}$ needs to fall within the range $\{12°, 35°\}$. The experimental estimate in (Marques et al., 2018) is indicated by the shaded area in Fig.6c. Fig.6d shows that the same results hold in a model of L2/3 with 4 recurrent cell types (see Section 2.6).

The feedback input current from HVAs is given by a convolution between the Gaussian connectivity profile and the rate field of HVAs, which is concave in an interval around the center, *i.e.* it has a local minimum at the center. The result of this convolution is convex in the center (*i.e.* is peaked in the center) when the Gaussian is large enough with respect to that interval (see Supplementary Section S11 B).

With a similar argument we can rationalize inverse size tuning. The fact that the width of the HVAs rate field grows with stimulus size (*i.e.*the diameter of the hole) (Supplementary Section S1 C), means that the interval of concavity of HVAs becomes larger for large stimulus sizes. With a fixed $\sigma_{L23-HVA}$, large enough stimuli generate a concave feedback input current, eliciting a concave recurrent rate field. This suppression of response in the center gives rise to inverse surround suppression, even as responses of cells that are offset from the center grow with increasing inverse stimulus size.

We can test this mechanistic understanding by (counterfactually) eliminating the scaling of the width of the HVAs rate field (Fig.6e). If our explanation is right, this should prevent inverse surround suppression. Indeed, inverse response at the stimulus center keeps increasing as a function of stimulus size when the width scaling is removed (Fig.6f). In Supplementary Section S11 we prove this mathematically for the parameterization chosen. Note that this is in contrast with the classical stimulus condition, where surround suppression is only slightly decreased removing the scaling of the width (see Fig.3f). This difference in surround modulation suggests that despite the similarity of the size tuning curves in the classical and inverse case, their origin is substantially different.

Our model predicts that classical surround suppression in L2/3 is mostly inherited from the aligned cells of its inputs and enhanced by both PV inhibitory input and the SOM-VIP subcircuit. On the other hand, inverse size tuning is mediated by offset cells in HVAs. As the stimulus becomes larger, peak responses in HVAs move to cells at larger offsets from the center (see Supplementary Section S3). These cells provide less and less excitatory input to the center as the stimulus radius grows beyond the span of the feedback projections.

The arguments presented here provide a purely geometrical explanation of inverse response and inverse size tuning, illustrating the importance of considering a spatial framework with realistic anatomical and physiological length scales to understand contextual modulation phenomena.

2.6 A computational model trained to fit cell-type-specific recordings for classical stimuli predicts inverse responses

The operating regime of cortical computations emerges from the interplay of Pyramidal cells with multiple interneuron types, mainly PV, SOM and VIP cells. In this Section we show that our results hold in a network model that includes all four cell types. We call this the 'full model' (see Fig.7a).

Since the synaptic coupling weights are not easily measured in the lab (*e.g.*, see varying results in Billeh et al., 2020; Karnani et al., 2016b; Pfeffer et al., 2013), we infer them (as in Dipoppa et al., 2018; Keller et al., 2020a; Mossing et al., 2021; Palmigiano et al., 2020), together with threshold offsets specific to each cell type, by constraining the model to reproduce the recurrent rate fields recorded in the classical stimulus condition (see Supplementary Section 3). In order to avoid a proliferation of parameters: i) we set to zero the connections that have been shown experimentally to be very small (Billeh et al., 2020; Campagnola et al., 2021;

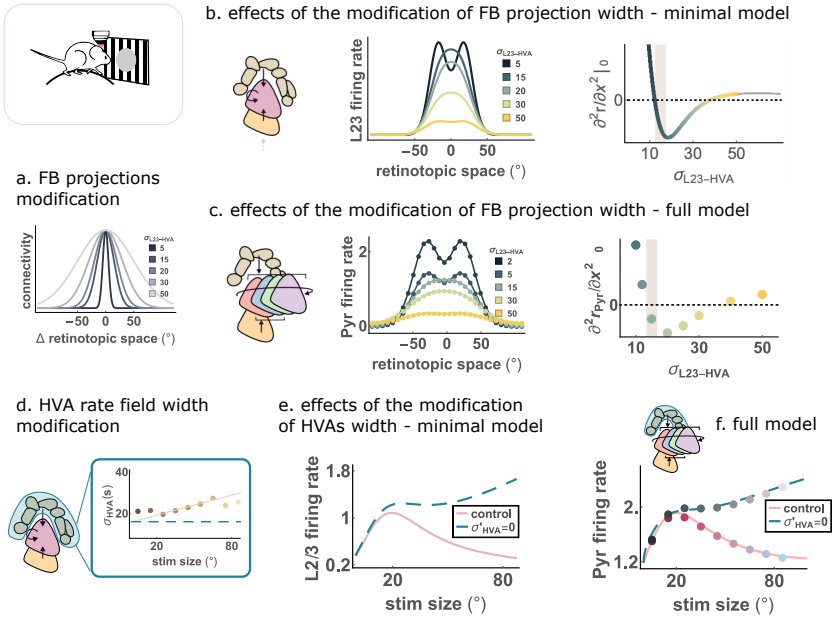

Figure 6: **Inverse response requires wide enough feedback projections and feedback activity profiles that scale with stimulus size** **a**. Counterfactual modification of the width of projections from HVAs to L2/3, $\sigma_{L23-HVA}$. **b**. inverse response depends on feedback projection's width. Rate field of the recurrent layer for different values of $\sigma_{L23-HVA}$ in different colors. Right: second derivative evaluated at the origin of the rate field, plotted as a function of $\sigma_{L23-HVA}$. The color code indicates the value of $\sigma_{L23-HVA}$ as in the left panel, while the shaded area corresponds to the experimental estimate in (Marques et al., 2018). Inverse response, seen as a larger response in the center of the hole than in the surround, is recovered in the region where the second derivative is negative. **c**. Same plots as **b**, here obtained by modifying $\sigma_{L23-HVA}$ in the full model, where 4 recurrent cell types are explicitly considered (see Sec.2.6). **d**. Counterfactual modification in which the width of the rate field of HVAs does not change with changing stimulus size (blue dashed line), together with the widths taken from the data (full circles indicate data, the full light brown line is a linear fit and represents the widths used in Fig. 5.). **e**. Inverse size tuning depends on HVAs' rate field width. The counterfactual modifications in **d** prevent inverse surround suppression. Size tuning curve for unmodified (modified) input is the light red full (blue dashed) line. **f**. Same as panel **e**, but for the full model as in **c** (see Sec.2.6).

Fu et al., 2014; Garcia-Junco-Clemente et al., 2017; Jiang et al., 2015; Karnani et al., 2016a,b; Pfeffer et al., 2013; Pi et al., 2013) as in (Dipoppa et al., 2018; Keller et al., 2020a) and ii) we fix the connection widths based on anatomical measurements (Billeh et al., 2020; Karnani et al., 2016b; Rossi et al., 2020) (see Fig.7b and Supplementary Section S6).

We develop a semianalytical Non-Negative-Least-Squares approach to fit the free parameters of the model (see Supplementary Section 3). Since all the inputs that SOM cells receive in the classical stimulus condition are surround suppressed, yet SOM cells are not, it is particularly hard to recover SOM size tuning properties. Possible solutions to this problem are to consider divisive inhibition from SOM to VIP (Dipoppa et al., 2018) or additional inputs that grow with stimulus size(Mossing et al., 2021). Here for simplicity we adopt the latter solution (see Supplementary Section S16). This additional input might arise from asymmetries or feature-dependence in lateral or feedback connectivity profiles (Iacaruso et al., 2017; Marques et al., 2018; Samonds et al., 2017). We fit the model to reproduce the classical response profiles and we compare it with the inverse response profiles. We find a collection of parameter sets that fit the data well (for a model comparison see Supplementary Section S17). In Fig. 7b we plot one of the inferred parameter sets, and the resulting match of experimental and model classical size tuning curves (Fig.7c). With the inferred parameters, and given the external inputs from HVAs and L4, the model qualitatively reproduces the inverse response and size tuning of all cell types in L2/3 (Fig.7d). This confirms that the inputs (in particular the input from HVAs) play a fundamental role in shaping the inverse patterns of activity, as suggested in (Keller et al., 2020b) and clarified by the minimal model. Moreover, we show that the results presented here are robust against reasonable changes in the connection width, accounting for errors in the experimental estimates considered (see Supplementary Section S18 C).

The full model can be leveraged to test the robustness of the mechanisms uncovered through the minimal model. In particular Fig.6d and g confirm that inverse response and size tuning in the full model require: i)broad enough feedback connections, ii) growth of the width of the HVAs rate field. Therefore the principles discussed and analysed in Section 2.5 extend to this more biologically realistic framework.

2.7 Surround facilitation can be explained by the inverse response to the orthogonal surround

When a stimulus is presented together with an orthogonal surround ("cross" stimulus), the response of L2/3 – but not L4 – Pyr cells centered on the stimulus is facilitated (Keller et al., 2020a; Shushruth et al., 2012a; Sillito et al., 1995). We propose that this effect may be the counterpart of the inverse response. More specifically we propose that the response of aligned Pyr cells increases due to the inverse response to the orthogonal surround. To test this hypothesis, we extend the minimal model to include orientation preference (see Fig.8a and Supplementary Section S14). We set the decay of the connectivity in orientation space by experimental

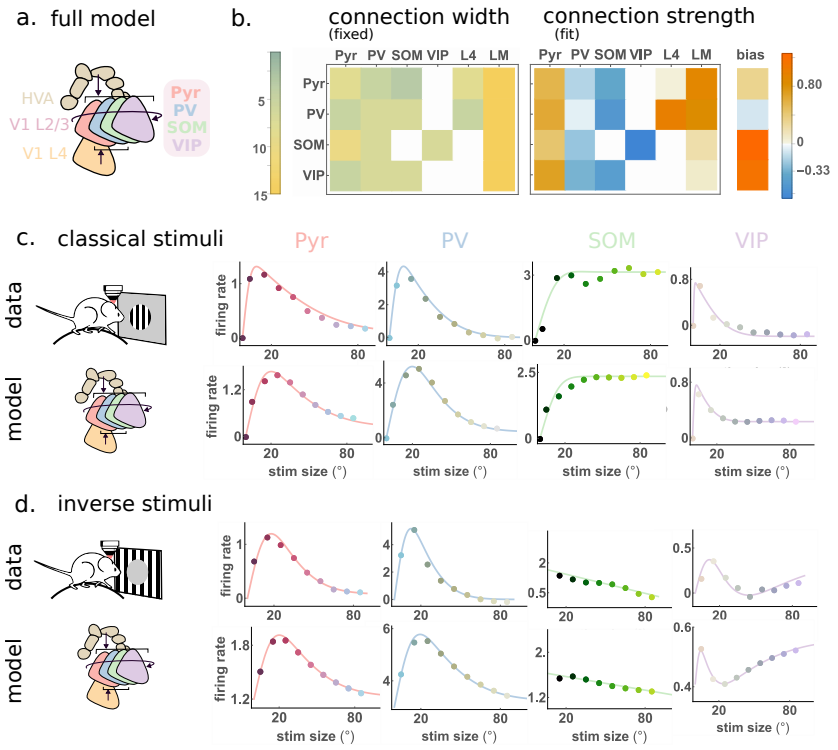

Figure 7: **A computational model trained to fit cell-type-specific recordings for classical stimuli predicts inverse responses a**. Sketch of the full model, with 4 recurrent cell types in L2/3 of V1 (Pyr, PV, SOM, VIP), plus feedforward input from L4 (targeting only Pyr and PV) and feedback input from HVAs (targeting all cell types). **b**. Left: a matrix of connection widths estimated from the literature (see Supplementary Section S6 and references therein). Right: the set of connection weights and biases, obtained by the non-negative least squares semi-analytic procedure (Supplementary Section 3). **c**. Classical size tuning curves. Comparison between experimental tuning curves (full circles top row) and steady states of the simulations of the full model (full circles bottom row) with free parameters as in b. The full lines are fits to the data/simulations. **d**. Same as c for inverse size tuning stimuli. The model with parameters inferred to fit the responses to classical stimuli generalizes to reproduce the inverse size tuning curves for all 4 cell types in L2/3.

estimates (Ko et al., 2011, 2013; Rossi et al., 2020) (see Supplementary Section S6).

We assume that the rate fields of the input populations for the cross stimulus can be approximated as the sum of the rate fields for a direct stimulus of one orientation in the center plus an inverse stimulus of the orthogonal orientation (Fig.8b left panel). In general the response fields are not additive. Nevertheless, for cross stimuli the classical and inverse stimuli are largely encoded by complementary sets of cells, i.e. cells with orthogonal orientation preferences. Our assumption could be tested experimentally by recording the spatial pattern of responses to the three stimuli – center-only, inverse, and cross).

The extended minimal model shows that the input from HVAs cells to L2/3 Pyr+PV cells centered on the stimulus are larger for the cross stimulus than for classical stimuli (Fig. 8b, right panel). Thus L2/3 cells centered on the stimulus and tuned to its orientation are surround facilitated (Fig. 8c). In addition, the model predicts that Pyr and PV cells in L2/3 centered on the stimulus and tuned to the orientation orthogonal to it should show significant response in the cross stimulus condition (Fig.8c).

Moreover if our hypothesis is correct, then optogenetically silencing HVAs during cross stimulus presentation should reduce surround facilitation. Likewise, the model predicts that hyperpolarizing SOM cells during cross stimulus presentation should also reduce surround facilitation. These points are illustrated in (Fig.8d,e). In Fig.8d, we show the firing rate of centered cells normalized to the only center stimulus condition within each experimental setup (i.e. control, HVAs silenced, SOM hyperpolarized). Surround facilitation corresponds to peak responses larger than 1. For instance, when HVAs are silenced, L2/3 responses to both classical and cross stimuli decrease. Fig.8d shows that their ratio also decreases, implying that HVAs support cross-orientation surround facilitation.

To better quantify surround facilitation we define the Surround Facilitation Index as $SFI = (cross - center)/(cross + center)$. In Fig.8e we show the behavior of the SFI as a function of the relative weight of excitation and inhibition in the joint population of Pyr+PV, controlled by the amplitude of the recurrent connections W_0. The model shows that systems with larger W_0, i.e. systems with more recurrent excitation, have larger SFI. This suggests that, when optogenetically silencing a fraction of PV neurons, surround facilitation should increase. Fig.8e also shows that SFI drops when HVAs or SOM inputs are reduced. Finally, in Supplementary Section S15 we show that a lower SFI following SOM hyperpolarization is implied by the lower ratio of inverse to classical response for SOM compared to Pyr and PV cells.

To summarize, both HVAs and lateral SOM inputs contribute to cross-orientation surround facilitation, while lateral PV input tends to reduce it. These are all testable predictions of the model.

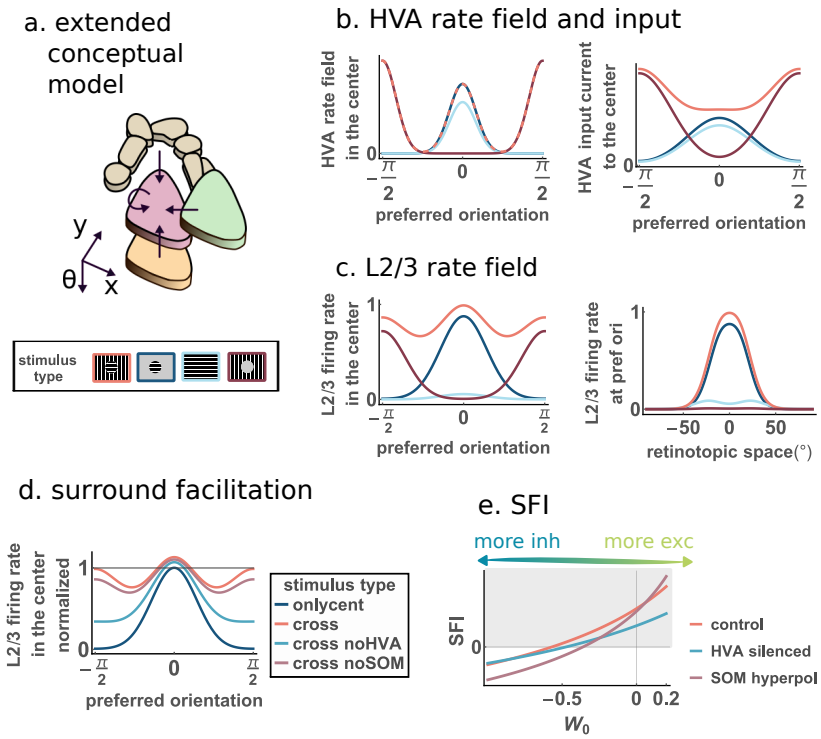

Figure 8: **Surround facilitation is generated by inverse response to the orthogonal surround. a**. Sketch of the extended minimal model, accounting for orientation preference as an additional dimension for the rate fields. Bottom: Legend of the stimuli analysed: cross (red), classical/only center (blue), iso-oriented surround (light blue), inverse (dark red). **b** Left: HVAs rate fields of cells whose CRF is centred on the stimulus center and with all possible orientation selectivities (x axis). We assume that the rate field of HVAs for the cross stimulus is given by the sum of a classical stimulus at orientation $\theta = 0$ plus an inverse stimulus at $\theta = \pi$. Right: input current from HVAs, given by the convolution of the rate fields with the connectivity profile. **c**. The response of centered L2/3 cells preferring orientation $\theta = 0$ is larger for the cross stimulus than for the classical stimulus alone (surround facilitation). Left: Rate field of L2/3 neurons at the center of the stimulus, as a function of their preferred orientation. Right: rate field of L2/3 cells that prefer orientation $\theta = 0$, as a function of their retinotopic location. **d**. Rate fields of L2/3 as in **c**, but normalized to the amplitude of the response in the only-center condition for the given experimental manipulation (control, no HVAs, no SOM).

Figure 8: The responses > 1 (< 1) represent the presence (absence) of surround facilitation. Surround facilitation decreases when HVAs are silenced or SOM cells are hyperpolarized. **e**. Dependence of surround facilitation index (SFI) on the relative weight of excitation and inhibition in the recurrent layer, specified by the amplitude of the recurrent connections W_0. Consistently with panel **d**, SFI decreases both when HVAs are silenced and when SOM cells are hyperpolarized.

3 Discussion

We propose a unifying model accounting for three types of contextual modulation and connecting two previously unrelated phenomena: inverse response and cross-orientation surround facilitation. Our model suggests that classical surround suppression in L2/3 is partially inherited from surround suppression of its feedforward and feedback inputs (Angelucci et al., 2017; Vangeneugden et al., 2019), marginally enhanced by lateral inhibition progressively recruited as the stimulus size increases (Obeid and Miller, 2021a; Rubin et al., 2015), enhanced by SOM lateral input (Adesnik et al., 2012) and partially supported by the supralinear input/output function (Miller and Troyer, 2002). In addition we address contrast dependence of classical surround suppression (Mossing et al., 2021), as well as the change in convexity of the spatial response profile of Pyr and PV cells to large classical stimuli (Dipoppa et al., 2018).

We establish a quantitative relation between inverse response and the range of feedback projections. In particular we show that if the feedback projections were too short-ranged, or too long-ranged, there would be no inverse response in L2/3. Moreover, despite the similarity of inverse size tuning curves and classical ones, we characterize inverse surround suppression as a purely geometrical effect, that would not appear if the HVAs rate field profile did not get wider with increasing stimulus size. On the contrary classical size tuning is only marginally enhanced by the same component. Additionally, we propose that cross-orientation surround facilitation is due to the inverse response to the surround.

Since we intend to gain robust mechanistic insight, we include in the picture as few elements as possible, to easily disentangle the effect of each one of them. Moreover we keep the parameters space as small dimensional as possible (as little as 3-dimensional in the case of the minimal model). Instead of fine tuning the model to a particular data set, we aim to describe results at a coarser level, at which they are more robust and consistent across different data sets.

Despite the very limited number of parameters, our minimal approach is highly informative. In addition to the three overall phenomena of classical and inverse size tuning and surround facilitation, it reproduces stabilization of the circuit by PV cells, the modulations of the classical and inverse response when silencing HVAs, and the effects of changing stimulus contrast and hyperpolarizing SOM cells. At the same time, it generates insights that translate into several experimentally testable

predictions: hyperpolarization of SOM cells decreases inverse surround modulation; silencing HVAs and/or hyperpolarizing L2/3 SOM cells and/or silencing L2/3 PV cells reduces surround facilitation; and the spatial activity patterns evoked by a center stimulus of one orientation and an inverse stimulus of the orthogonal orientation should add up roughly linearly, to give the spatial pattern evoked by a cross-oriented center/surround stimulus.

Keller et al. (Keller et al., 2020b) highlight the importance of input from offset neurons in HVAs for the inverse response to take place. One modelling approach to describe spatial effects is to divide space into two bins, representing the activity of the center and surround subnetworks (Angelucci et al., 2017; Keller et al., 2020a; Mossing et al., 2021). Here, we adopt a continuous description of retinotopic space, as in (Dipoppa et al., 2018) to preserve the information about length scales of the spatial patterns of activity. In addition, we constrain the range of the projections between different populations based on anatomical results (Billeh et al., 2020; Campagnola et al., 2021; Karnani et al., 2016b; Pfeffer et al., 2013; Rossi et al., 2020).

The spatial profile of the response of a population is implemented through the the concept of firing rate field. To simplify the experimental dataset we standardize the rate fields to have Gaussian shapes. This schematic description of space leads to the key observation that the length scales of the recorded spatial profiles grow much more slowly than the stimulus size for the dataset in (Keller et al., 2020b). This holds true independently of the preprocessing procedure used to define the rate fields (see Supplementary Section S1 A FigureS2 and S3) and it is consistent with the recordings in (Dipoppa et al., 2018). Pinpointing this counterintuitive feature of neural responses is one of our major findings. This feature is at odds with the assumptions taken in previous modelling works (Obeid and Miller, 2021a; Rubin et al., 2015), based on the intuition that cells in V1 respond when their CRF is covered by a stimulus.

Our schematic description of space opens the way to an analytic treatment of the problem (Persi et al., 2011), which allows us to determine robust conditions for qualitative response properties, such as: what are the properties of the inputs that give rise to surround suppression and that determine the preferred size of the response, possible causes for the decrease of the preferred size when the stimulus contrast is increased, the conditions that allow inverse response rate fields that peak at the center and the conditions that make center responses to large holes smaller than center responses to small holes.

We treated HVAs as a static input, even though they are part of a coupled system, receiving input from V1 and projecting back to V1. In the data set considered, L2/3 responses to small-size classical and inverse stimuli are very similar, while LM responses differ substantially. This implies that LM must receive a significant part of its input from other sources, that we are not modelling (Blot et al., 2021; Harris et al., 2019; Tohmi et al., 2014).

While the minimal model neglects the full recurrent microcircuitry of L2/3, it serves as a blueprint for isolating mechanistic principles of cortical computation, that apply to a fully recurrent 4-cell-type circuit. We fit the parameters of the more

biologically plausible model to reproduce the size-dependent response profiles for classical stimuli. The model is able to predict the inverse size tuning curves, given the feedforward and feedback inputs.

Context shapes the readout not only at a neural level, but also at a perceptual level. Living beings are active interpreters of the visual scene, as opposed to passive decoders (Agrillo et al., 2015; Agrochao et al., 2020; Bååth et al., 2014; Bach and Poloschek, 2006; Carbon, 2014; Eagleman, 2001; Endler et al., 2010; Gori et al., 2014; Keemink et al., 2018; Luo et al., 2019; Mély et al., 2018; Pak et al., 2019, 2020; Santacà et al., 2019). The study of the information that higher processing stages project back onto the canvas of visual cortex opens a window to understand the 'illusions' that create our perceptual experiences.

We hope that the simplicity, analytical tractability, and spatial nature of the model presented here, together with the fact that it can be easily extended to account for feature spaces, will be useful for further investigations of visual cortical circuitry and responses.

References

H. Adesnik. Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. Neuron, 95:1147–1159, 2017.

H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani. A neural circuit for spatial summation in visual cortex. Nature, 490(7419):226–231, 2012.

C. Agrillo, S. Gori, and M. J. Beran. Do rhesus monkeys (macaca mulatta) perceive illusory motion? Animal cognition, 18(4):895–910, 2015.

M. Agrochao, R. Tanaka, E. Salazar-Gatzimas, and D. A. Clark. Mechanism for analogous illusory motion perception in flies and humans. Proceedings of the National Academy of Sciences, 117(37):23044–23053, 2020.

Y. Ahmadian and K. D. Miller. What is the dynamical regime of cerebral cortex? Neuron, 2021.

Y. Ahmadian, D. B. Rubin, and K. D. Miller. Analysis of the stabilized supralinear network. Neural Computation, 25:1994–2037, 2013.

J. Allman, F. Miezin, and E. McGuinness. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. Annual Review of Neuroscience, 8(1):407–430, 1985. doi: 10.1146/annurev.ne.08.030185.002203. URL `https://doi.org/10.1146/annurev.ne.08.030185.002203`. PMID: 3885829.

A. Angelucci, J. B. Levitt, and J. S. Lund. Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area v1. Progress in brain research, 136:373–388, 2002.

A. Angelucci, M. Bijanzadeh, L. Nurminen, F. Federer, S. Merlin, and P. C. Bressloff. Circuits and mechanisms for surround modulation in visual cortex. Annual review of neuroscience, 40:425–451, 2017.

R. Bååth, T. Seno, and A. Kitaoka. Cats and illusory motion. Psychology, 5(9): 1131–1134, 2014.

M. Bach and C. M. Poloschek. Optical illusions. Adv Clin Neurosci Rehabil, 6(2): 20–21, 2006.

Y. N. Billeh, B. Cai, S. L. Gratiy, K. Dai, R. Iyer, N. W. Gouwens, R. Abbasi-Asl, X. Jia, J. H. Siegle, S. R. Olsen, et al. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. Neuron, 106(3):388–403, 2020.

C. Blakemore and E. A. Tobin. Lateral inhibition between orientation detectors in the cat's visual cortex. Experimental brain research, 15(4):439–440, 1972.

A. Blot, M. M. Roth, I. Gasler, M. Javadzadeh, F. Imhof, and S. B. Hofer. Visual intracortical and transthalamic pathways carry distinct information to cortical areas. Neuron, 109(12):1996–2008.e6, 2021. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2021.04.017. URL `https://www.sciencedirect.com/science/article/pii/S089662732100283X`.

L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli, et al. Connectivity and synaptic physiology in the mouse and human neocortex. bioRxiv, 2021.

M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. Nat. Rev. Neurosci., 13:51–62, 2012.

M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? Journal of Neuroscience, 25(46):10577–10597, 2005.

C. Carbon. Understanding human perception by human-made illusions. Frontiers in Human Neuroscience, 8, 2014.

J. Cavanaugh, W. M. Joiner, and R. H. Wurtz. Suppressive surrounds of receptive fields in monkey frontal eye field. J. Neurosci., 32:12284–12293, 2012.

J. R. Cavanaugh, W. Bair, and J. A. Movshon. Selectivity and spatial distribution of signals from the receptive field surround in macaque v1 neurons. Journal of neurophysiology, 2002.

X.-l. Chou, Q. Fang, L. Yan, W. Zhong, B. Peng, H. Li, J. Wei, H. W. Tao, and L. I. Zhang. Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression. Elife, 9:e54157, 2020.

R. Coen-Cagli, A. Kohn, and O. Schwartz. Flexible gating of contextual influences in natural vision. Nat Neurosci, 18:1648–1655, 2015.

S. Denève and C. K. Machens. Efficient codes and balanced networks. Nature Neuroscience, 19(3):375–382, 2016. ISSN 1546-1726. doi: 10.1038/nn.4243. URL https://doi.org/10.1038/nn.4243.

M. Dipoppa, A. Ranson, M. Krumin, M. Pachitariu, M. Carandini, and K. D. Harris. Vision and locomotion shape the interactions between neuron types in mouse visual cortex. Neuron, 98(3):602–615, 2018.

D. M. Eagleman. Visual illusions and neurobiology. Nature Reviews Neuroscience, 2(12):920–926, 2001. ISSN 1471-0048. doi: 10.1038/35104092. URL https://doi.org/10.1038/35104092.

J. A. Endler, L. C. Endler, and N. R. Doerr. Great bowerbirds create theaters with forced perspective when seen by their audience. Current Biology, 20(18): 1679–1684, 2010.

A. L. Falkner, B. S. Krishna, and M. E. Goldberg. Surround suppression sharpens the priority map in the lateral intraparietal area. J. Neurosci., 30:12787–12797, 2010.

D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". Vision Research, 33(2): 173–193, 1993. ISSN 0042-6989. doi: https://doi.org/10.1016/0042-6989(93)90156-Q. URL https://www.sciencedirect.com/science/article/pii/004269899390156Q.

Y. Fu, J. M. Tucciarone, J. S. Espinosa, N. Sheng, D. P. Darcy, R. A. Nicoll, Z. J. Huang, and M. P. Stryker. A cortical circuit for gain control by behavioral state. Cell, 156(6):1139–1152, 2014.

P. Garcia-Junco-Clemente, T. Ikrar, E. Tring, X. Xu, D. L. Ringach, and J. T. Trachtenberg. An inhibitory pull–push circuit in frontal cortex. Nature neuroscience, 20(3):389–392, 2017.

S. Gori, C. Agrillo, M. Dadda, and A. Bisazza. Do fish perceive illusory motion? Scientific reports, 4(1):1–6, 2014.

D. Hansel and C. van Vreeswijk. How noise contributes to contrast invariance of orientation tuning in cat visual cortex. J. Neurosci., 22:5118–5128, 2002.

J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, A. Feiner, D. Feng, N. Gaudreault, C. R. Gerfen, N. Graddis, P. A. Groblewski, A. M. Henry, A. Ho, R. Howard, J. E. Knox, L. Kuan, X. Kuang, J. Lecoq, P. Lesnar, Y. Li, J. Luviano,

S. McConoughey, M. T. Mortrud, M. Naeemi, L. Ng, S. W. Oh, B. Ouellette, E. Shen, S. A. Sorensen, W. Wakeman, Q. Wang, Y. Wang, A. Williford, J. W. Phillips, A. R. Jones, C. Koch, and H. Zeng. Hierarchical organization of cortical and thalamic connectivity. Nature, 575(7781):195–202, 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1716-z. URL https://doi.org/10.1038/s41586-019-1716-z.

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of neurophysiology, 28(2): 229–289, 1965.

M. F. Iacaruso, I. T. Gasler, and S. B. Hofer. Synaptic organization of visual space in primary visual cortex. Nature, 547(7664):449–452, 2017.

J. S. Isaacson and M. Scanziani. How inhibition shapes cortical activity. Neuron, 72(2):231–243, 2011.

X. Jiang, S. Shen, C. R. Cadwell, P. Berens, F. Sinz, A. S. Ecker, S. Patel, and A. S. Tolias. Principles of connectivity among morphologically defined cell types in adult neocortex. Science, 350(6264), 2015.

M. M. Karnani, J. Jackson, I. Ayzenshtat, A. H. Sichani, K. Manoocheri, S. Kim, and R. Yuste. Opening holes in the blanket of inhibition: localized lateral disinhibition by vip interneurons. Journal of neuroscience, 36(12):3471–3480, 2016a.

M. M. Karnani, J. Jackson, I. Ayzenshtat, J. Tucciarone, K. Manoocheri, W. G. Snider, and R. Yuste. Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. Neuron, 90(1):86–100, 2016b.

H. K. Kato, S. K. Asinof, and J. S. Isaacson. Network-Level Control of Frequency Tuning in Auditory Cortex. Neuron, 95(2):412–423, 2017.

S. W. Keemink, C. Boucsein, and M. C. W. van Rossum. Effects of v1 surround modulation tuning on visual saliency and the tilt illusion. Journal of Neurophysiology, 120(3):942–952, 2018. doi: 10.1152/jn.00864.2017. URL https://doi.org/10.1152/jn.00864.2017. PMID: 29847234.

A. J. Keller, M. Dipoppa, M. M. Roth, M. S. Caudill, A. Ingrosso, K. D. Miller, and M. Scanziani. A disinhibitory circuit for contextual modulation in primary visual cortex. Neuron, 108(6):1181–1193, 2020a.

A. J. Keller, M. M. Roth, and M. Scanziani. Feedback generates a second receptive field in neurons of the visual cortex. Nature, 582(7813):545–549, 2020b.

H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. Nature, 473(7345):87–91, 2011.

H. Ko, L. Cossell, C. Baragli, J. Antolik, C. Clopath, S. B. Hofer, and T. D. Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. Nature, 496 (7443):96–100, 2013.

Y. Li and L.-S. Young. Unraveling the mechanisms of surround suppression in early visual processing. PLoS computational biology, 17(4):e1008916, 2021.

J. Luo, K. He, I. M. Andolina, X. Li, J. Yin, Z. Chen, Y. Gu, and W. Wang. Going with the flow: the neural mechanisms underlying illusions of complex-flow motion. Journal of Neuroscience, 39(14):2664–2685, 2019.

T. Marques, J. Nguyen, G. Fioreze, and L. Petreanu. The functional organization of cortical feedback inputs to primary visual cortex. Nature neuroscience, 21(5): 757–764, 2018.

K. D. Miller and T. W. Troyer. Neural noise can explain expansive, power-law nonlinearities in neural response functions. Journal of neurophysiology, 87(2): 653–659, 2002.

D. P. Mossing, J. Veit, A. Palmigiano, K. D. Miller, and H. Adesnik. Antagonistic inhibitory subnetworks control cooperation and competition across cortical space. bioRxiv, 2021. doi: 10.1101/2021.03.31.437953. URL `https://www.biorxiv.org/content/early/2021/03/31/2021.03.31.437953`.

D. Mély, D. Linsley, and T. Serre. Complementary surrounds explain diverse contextual phenomena across visual modalities. Psychological Review, 125, 09 2018. doi: 10.1037/rev0000109.

J. J. Nassi, C. Gomez-Laberge, G. Kreiman, and R. T. Born. Corticocortical feedback increases the spatial extent of normalization. Front Syst Neurosci, 8: 105, 2014.

L. Nurminen and A. Angelucci. Multiple components of surround modulation in primary visual cortex: multiple neural circuits with multiple functions? Vision research, 104:47–56, 2014.

L. Nurminen, S. Merlin, M. Bijanzadeh, F. Federer, and A. Angelucci. Top-down feedback controls spatial summation and response amplitude in primate visual cortex. Nat Commun, 9:2281, 2018.

D. Obeid and K. D. Miller. Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. bioRxiv, 2021a. doi: 10.1101/2020.12.30.424892. URL `https://www.biorxiv.org/content/early/2021/01/02/2020.12.30.424892`.

D. Obeid and K. D. Miller. Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. bioRxiv, https://doi.org/10.1101/2020.12.30.424892, 2021b.

M. Okun and I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nature neuroscience, 11(5):535–537, 2008.

H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. Neuron, 62(4):578–592, 2009.

C. L. Paffen, M. J. van der Smagt, S. F. Te Pas, and F. A. Verstraten. Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. Journal of Vision, 5(6):8–8, 2005.

A. Pak, E. Ryu, C. Li, and A. Chubykin. Top-down feedback controls the cortical representation of illusory contours in mouse primary visual cortex. The Journal of Neuroscience, 40:1998–19, 12 2019. doi: 10.1523/JNEUROSCI.1998-19.2019.

A. Pak, E. Ryu, C. Li, and A. A. Chubykin. Top-down feedback controls the cortical representation of illusory contours in mouse primary visual cortex. Journal of Neuroscience, 40(3):648–660, 2020. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1998-19.2019. URL `https://www.jneurosci.org/content/40/3/648`.

A. Palmigiano, F. Fumarola, D. P. Mossing, N. Kraynyukova, H. Adesnik, and K. D. Miller. Structure and variability of optogenetic responses identify the operating regime of cortex. bioRxiv, 2020. doi: 10.1101/2020.11.11.378729. URL `https://www.biorxiv.org/content/early/2020/11/11/2020.11.11.378729`.

E. Persi, D. Hansel, L. Nowak, P. Barone, and C. van Vreeswijk. Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. PLoS computational biology, 7(2):e1001078, 2011.

L. Petro, A. Paton, and L. Muckli. Contextual modulation of primary visual cortex by auditory signals. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1714):20160104, 2017.

C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nature neuroscience, 16(8):1068–1076, 2013.

H.-J. Pi, B. Hangya, D. Kvitsiani, J. I. Sanders, Z. J. Huang, and A. Kepecs. Cortical interneurons that specialize in disinhibitory control. Nature, 503(7477):521–524, 2013.

A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris. The asynchronous state in cortical circuits. science, 327(5965):587–590, 2010.

L. F. Rossi, K. D. Harris, and M. Carandini. Spatial connectivity matches direction selectivity in visual cortex. Nature, 588(7839):648–652, 2020.

D. B. Rubin, S. D. Van Hooser, and K. D. Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron, 85(2):402–417, 2015.

J. M. Samonds, B. D. Feese, T. S. Lee, and S. J. Kuhlman. Nonuniform surround suppression of visual responses in mouse v1. Journal of neurophysiology, 118(6): 3282–3292, 2017.

M. Santacà, M. E. Miletto Petrazzini, C. Agrillo, and A. Wilkinson. Can reptiles perceive visual illusions? delboeuf illusion in red-footed tortoise (chelonoidis carbonaria) and bearded dragon (pogona vitticeps). Journal of Comparative Psychology, 133(4):419, 2019.

A. Sanzeni, B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed. Inhibition stabilization is a widespread property of cortical networks. Elife, 9: e54875, 2020.

M. P. Sceniak, M. J. Hawken, and R. Shapley. Visual spatial characterization of macaque v1 neurons. Journal of neurophysiology, 85(5):1873–1887, 2001.

L. Schwabe, K. Obermayer, A. Angelucci, and P. Bressloff. The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. J. Neurosci., 26:9117–9129, 2006.

L. Schwabe, J. M. Ichida, S. Shushruth, P. Mangapathy, and A. Angelucci. Contrast-dependence of surround suppression in Macaque V1: experimental testing of a recurrent network model. Neuroimage, 52:777–792, 2010.

O. Schwartz and E. P. Simoncelli. Natural signal statistics and sensory gain control. Nature Neuroscience, 4:819–825, 2001.

M. W. Self, J. A. Lorteije, J. Vangeneugden, E. H. van Beest, M. E. Grigore, C. N. Levelt, J. A. Heimel, and P. R. Roelfsema. Orientation-tuned surround suppression in mouse visual cortex. Journal of Neuroscience, 34(28):9290–9304, 2014.

S. Shushruth, J. M. Ichida, J. B. Levitt, and A. Angelucci. Comparison of spatial summation properties of neurons in macaque v1 and v2. Journal of neurophysiology, 102(4):2069–2083, 2009a.

S. Shushruth, J. M. Ichida, J. B. Levitt, and A. Angelucci. Comparison of spatial summation properties of neurons in macaque V1 and V2. J. Neurophysiol., 102: 2069–2083, 2009b.

S. Shushruth, P. Mangapathy, J. M. Ichida, P. C. Bressloff, L. Schwabe, and A. Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. Journal of Neuroscience, 32 (1):308–321, 2012a.

S. Shushruth, P. Mangapathy, J. M. Ichida, P. C. Bressloff, L. Schwabe, and A. Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. J. Neurosci., 32:308–321, 2012b.

J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, P. A. Groblewski, R. Ahmed, A. Arkhipov, A. Bernard, Y. N. Billeh, D. Brown, M. A. Buice, N. Cain, S. Caldejon, L. Casal, A. Cho, M. Chvilicek, T. C. Cox, K. Dai, D. J. Denman, S. E. J. de Vries, R. Dietzman, L. Esposito, C. Farrell, D. Feng, J. Galbraith, M. Garrett, E. C. Gelfand, N. Hancock, J. A. Harris, R. Howard, B. Hu, R. Hytnen, R. Iyer, E. Jessett, K. Johnson, I. Kato, J. Kiggins, S. Lambert, J. Lecoq, P. Ledochowitsch, J. H. Lee, A. Leon, Y. Li, E. Liang, F. Long, K. Mace, J. Melchior, D. Millman, T. Mollenkopf, C. Nayan, L. Ng, K. Ngo, T. Nguyen, P. R. Nicovich, K. North, G. K. Ocker, D. Ollerenshaw, M. Oliver, M. Pachitariu, J. Perkins, M. Reding, D. Reid, M. Robertson, K. Ronellenfitch, S. Seid, C. Slaughterbeck, M. Stoecklin, D. Sullivan, B. Sutton, J. Swapp, C. Thompson, K. Turner, W. Wakeman, J. D. Whitesell, D. Williams, A. Williford, R. Young, H. Zeng, S. Naylor, J. W. Phillips, R. C. Reid, S. Mihalas, S. R. Olsen, and C. Koch. Survey of spiking in the mouse visual system reveals functional hierarchy. Nature, 592(7852):86–92, 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-03171-x. URL https://doi.org/10.1038/s41586-020-03171-x.

A. M. Sillito, K. L. Grieve, H. E. Jones, J. Cudeiro, and J. Davls. Visual cortical mechanisms detecting focal orientation discontinuities. Nature, 378(6556):492–496, 1995.

L. Spillmann, B. Dresp-Langley, and C. Tseng. Beyond the classic receptive field: The effect of contextual stimuli. Journal of Vision, 15:1–22, 2015.

K. A. Sundberg, J. F. Mitchell, and J. H. Reynolds. Spatial attention modulates center-surround interactions in macaque visual area v4. Neuron, 61:952–963, 2009.

K. Tanaka, I. Ohzawa, A. S. Ramoa, and R. D. Freeman. Receptive field properties of cells in area 19 of the cat. Exp. Brain Res., 65:549–558, 1987.

M. Tohmi, R. Meguro, H. Tsukano, R. Hishida, and K. Shibuki. The extrageniculate visual pathway generates distinct response properties in the higher visual areas of mice. Current Biology, 24:587–597, 2014.

R. Tremblay, S. Lee, and B. Rudy. Gabaergic interneurons in the neocortex: from cellular properties to circuits. Neuron, 91(2):260–292, 2016.

T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller. Contrast-invariant orientation tuning in cat visual cortex: Feedforward tuning and correlation-based intracortical connectivity. J. Neurosci., 18:5908–5927, 1998.

J. M. Tsui and C. C. Pack. Contrast sensitivity of MT receptive field centers and surrounds. J. Neurophysiol., 106:1888–1900, 2011.

G. Van den Bergh, B. Zhang, L. Arckens, and Y. M. Chino. Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. J. Comp. Neurol., 518:2051–2070, 2010.

J. Vangeneugden, E. H. van Beest, M. X. Cohen, J. A. Lorteije, S. Mukherjee, L. Kirchberger, J. S. Montijn, P. Thamizharasu, D. Camillo, C. N. Levelt, et al. Activity in lateral visual areas contributes to surround suppression in awake mouse v1. Current Biology, 29(24):4268–4275, 2019.

J. Xing and D. J. Heeger. Measurement and modeling of center-surround suppression and enhancement. Vision research, 41(5):571–583, 2001.

C. Yu and D. M. Levi. Surround modulation in human vision unmasked by masking experiments. nature neuroscience, 3(7):724–728, 2000.

C. M. Ziemba, J. Freeman, E. P. Simoncelli, and J. A. Movshon. Contextual modulation of sensitivity to naturalistic image structure in macaque v2. Journal of neurophysiology, 120(2):409–420, 2018.

K. Zipser, V. A. Lamme, and P. H. Schiller. Contextual modulation in primary visual cortex. Journal of Neuroscience, 16(22):7376–7389, 1996.

Acknowledgments

We acknowledge funding sources NIH U01NS108683, NIH R01EY029999, NIH U19NS107613, NSF 1707398, Gatsby Charitable Foundation GAT3708.

Authors Contributions

Conceptualization, K.D.M., S.D.S and M.D.; Methodology, S.D.S.; Software, S.D.S.; Formal Analysis, S.D.S; Investigation, A.K.,M.R.,M.S. and M.D.; Resources, A.K.,M.R.,M.S. and M.D; Writing-Original Draft, S.D.S.; Writing-Review and Editing, K.D.M., S.D.S. and M.D.; Visualization, S.D.S.; Supervision, K.D.M.; Funding Acquisition,K.D.M., M.S.

Declaration of interests

The authors declare no competing interests.

STAR methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and code should be directed to and will be fulfilled by the lead contact, Serena Di Santo (serenadisanto14@google.com)

Materials availability

This study did not generate new unique reagents.

Data and code availability

All original code has been deposited at [] and is publicly available as of the date of publication. DOIs are listed in the key resources table.

METHOD DETAILS

The minimal model

We consider a system with one recurrent population defined on a two-dimensional retinotopic space of coordinates $\{\mathbf{x}\}$, whose rate field is $r(\mathbf{x})$ and which receives an isotropic Gaussian input $I(\mathbf{x}) = I_0 G(\mathbf{x}, \sigma)$ (where $G(\mathbf{x}, \sigma) = \frac{1}{2\pi\sigma} e^{-\frac{\mathbf{x}^2}{2\sigma}}$). Let the recurrent connections be $W(\mathbf{x}, \mathbf{y}) = W_0 G(\mathbf{x} - \mathbf{y}, \sigma_r)$, and let the response function be $r(\mathbf{x}) = [u(\mathbf{x})]_+^2$, where we defined the total input current $u(\mathbf{x})$. The SSN equations for the system read:

$$r(\mathbf{x}) = \left[I(\mathbf{x}) + \int d\mathbf{y} W(\mathbf{x}, \mathbf{y}) r(\mathbf{y}) \right]^2 \tag{1}$$

Where we omit the rectification sign, based on the observation that $r(\mathbf{x}) > 0$ wherever the total input current $u(\mathbf{x}) > 0$ is positive. Analogously we can write the SSN equations for the total input current as:

$$u(\mathbf{x}) = I(\mathbf{x}) + \int d\mathbf{y} W(\mathbf{x}, \mathbf{y}) u^2(\mathbf{y}) \tag{2}$$

To find $u(\mathbf{x})$ we take the ansatz:

$$u(\mathbf{x}) = \alpha(\mathbf{x}) G(\mathbf{x}, \sigma_u) \tag{3}$$

where $\alpha(\mathbf{x})$ is a slowly varying function of its argument, so that we can take it out of the integral in Eq. 2. More in general, if we consider:

$$\mathcal{I}(x) = \int_{-\infty}^{\infty} dy \alpha(y) G(y, \sigma) G(x - y, \sigma_r) \tag{4}$$

where we assume that $\alpha(y)$ is a slowly varying function of its argument x and in particular it varies slowly with respect to $G(x, \sigma + \sigma_r)$. Then:

$$\mathcal{I}(x) \simeq \alpha(\bar{y}) \int_{-\infty}^{\infty} dy G(y, \sigma) G(x - y, \sigma_r) \tag{5}$$

where $\bar{y}$ is the value that maximizes $G(y, \sigma)G(x - y, \sigma_r)$, i.e. $\bar{y} = \frac{\sigma}{\sigma_r + \sigma} x$ but since $\alpha(y)$ is slowly varying we will take the further approximation $\alpha(\bar{y}) \simeq \alpha(x)$, that corresponds to just carrying $\alpha(y)$ outside of the integral.

If we note that:

$$G^2(\mathbf{x}, \sigma_u) = \frac{1}{4\pi\sigma_u} G\left(\mathbf{x}, \frac{\sigma_u}{2}\right) \tag{6}$$

We can then plug the ansatz into Eq. 2 and solve the convolution (remembering that the convolution of two normalized Gaussians is a normalized Gaussian with variance equal to the sum of the variances):

$$\alpha(\mathbf{x}) G(\mathbf{x}, \sigma_u) = I_0 G(\mathbf{x}, \sigma) + \frac{W_0}{4\pi\sigma_u} \alpha^2(\mathbf{x}) G(\mathbf{x}, \sigma_{ru}) \tag{7}$$

where we defined $\sigma_{ru} = \sigma_r + \frac{\sigma_u}{2}$. We now solve the equation above for $\alpha(\mathbf{x})$ and get:

$$\alpha(\mathbf{x}) = \frac{2\pi\sigma_u}{W_0 G(\sigma_{ru})} \left(G(\sigma_u) \pm \sqrt{G^2(\sigma_u) - \frac{I_0 W_0 G(\sigma) G(\sigma_{ru})}{\pi\sigma_u}} \right) \tag{8}$$

where for better readability we are omitting the spatial dependence of the function $G(\cdot)$. If σ_u is such that:

$$2\sigma_u^{-1} = \sigma^{-1} + \sigma_{ru}^{-1} \tag{9}$$

then we have:

$$\frac{G(\sigma)G(\sigma_{ru})}{G^2(\sigma_u)} = \frac{\sigma_u^2}{\sigma\sigma_{ru}} \tag{10}$$

and we can write:

$$\alpha(\mathbf{x}) \simeq \frac{2\pi\sigma_u}{W_0 G(\sigma_{ru})} G(\sigma_u) \left(1 \pm \sqrt{1 - \frac{I_0 W_0 \sigma_u}{\pi\sigma\sigma_{ru}}} \right) \tag{11}$$

Note that the the equality would hold strictly only when Eq.9 is satisfied, but since we want to choose σ and σ_r at the same time, Eq.9 may hold only approximately. Finally, using Eq.10 again:

$$\begin{aligned} u(\mathbf{x}) &\simeq \frac{2\pi\sigma\sigma_{ru}}{W_0\sigma_u} \left(1 \pm \sqrt{1 - \frac{I_0 W_0 \sigma_u}{\pi\sigma\sigma_{ru}}} \right) G(\sigma) \\ &= \frac{\pi(\sigma + \sigma_{ru})}{W_0} \left(1 \pm \sqrt{1 - \frac{2 I_0 W_0}{\pi(\sigma + \sigma_{ru})}} \right) G(\sigma) \end{aligned} \tag{12}$$

Note that only one of the two solutions is actually acceptable, depending on whether the recurrent population described is excitatory or inhibitory, in order to ensure $u(\mathbf{x}) > 0$. Indeed the solution with + cannot hold for negative W_0 and for positive W_0 instead the solution requires that $2I_0W_0 < \pi(\sigma_r+\sigma_u)$, which means that for very large recurrent excitatory input the system diverges (see Supplementary Section S9 B).

Analogously, if the external input can be described by a sum of Gaussians:

$$I(\mathbf{x}) = \sum_{k=1}^{m} I_k G(\mathbf{x}, \sigma_k) \tag{13}$$

we can obtain

$$u(\mathbf{x}) = \frac{\pi(\sigma_1 + \sigma_{ru})}{W_0} \left(1 - \sqrt{1 - \frac{2I_1W_0}{\pi(\sigma_1 + \sigma_{ru})} - 4\sigma_1^2 W_0 \sum_{k=1}^{m} \frac{I_k G(\sigma_{1k})}{(\sigma_1 - \sigma_k)\sigma_{ru}}} \right) G(\sigma_1) \tag{14}$$

and we defined

$$\begin{aligned} \sigma_u &= \frac{2\sigma_1\sigma_{ru}}{\sigma_1 + \sigma_{ru}} \\ \sigma_{1k} &= \frac{\sigma_1\sigma_k}{\sigma_1 - \sigma_k} \end{aligned} \tag{15}$$

The free parameters in the minimal model are the strengths of the effective synaptic connections.

The full model

In the full model the recurrent layer is composed by Pyramidal neurons (Pyr) as well as the 3 main interneuron classes Parvalbumin (PV), Somatostatin (SOM) and Vaso-intestinal peptide (VIP), that account for about $80 - 90\%$ of the neurons in the Visual Cortex(Tremblay et al., 2016). In addition, as in the minimal model, we consider external feedforward input from L4 and external feedback input from HVAs. In order to keep the number of parameters as small as possible, we consider rectified quadratic transfer functions for all celltypes and we fix the projection span of each connection based on anatomical observations (see Supplementary Section S6). Since the firing rates are not 0, we assume that the total input current to each population is positive. Then we can write the following equation for the rate field of population A (A=1..4, since the description of the origin of L4 and HVAs size tuning profiles is beyond the scope of this work):

$$\sqrt{r_A(\mathbf{x})} = \sum_B \int W_{AB}(\mathbf{x} - \mathbf{y}) r_B(\mathbf{y}) d\mathbf{y} + T_A \tag{16}$$

where $B = \{E, P, S, V, L, H\}$ are the populations Pyr, PV, SOM, VIP, L4, and HVAs and $\mathbf{x}$ denotes 2 dimensional flat space. The term T_A is a bias that we

consider constant (across space and stimulus size), which can be understood as a cell-type specific firing threshold.

As in the minimal model, we consider Gaussian assumptions and ansatz:

$$\begin{aligned} W_{AB}(\mathbf{x}-\mathbf{y}) &= \frac{\bar{W}_{AB}}{2\pi\sigma_{AB}} e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2\sigma_{AB}}} \\ r_A(\mathbf{x}) &= \bar{r}_A e^{-\frac{\mathbf{x}^2}{2\sigma_A}} + b_A \end{aligned} \tag{17}$$

Where b is the baseline, i.e. the firing rate in absence of inputs, uniform across space. Here we aim at solving for the $\bar{W}_{AB}$ given the $r_A(\mathbf{x})$ (see Supplementary Section S1 A) and given the values of σ_{AB}, fixed according to anatomical estimates (see Supplementary Section S6). In order to regress the connection weights, we set up a non negative least squares (NNLS) approach similarly to (Dipoppa et al., 2018; Keller et al., 2020a). Let us consider first the NNLS procedure when we constrain only on one stimulus size. Then the loss function reads[1]:

$$\begin{aligned} S_A &= \int_{-\infty}^{\infty} d\mathbf{x} \left[\sqrt{r_A(\mathbf{x})} - \sum_B \int W_{AB}(\mathbf{x}-\mathbf{y}) r_B(\mathbf{y}) d\mathbf{y} - T_A \right]^2 \\ &= \int_{-\infty}^{\infty} d\mathbf{x} \left[\sqrt{\bar{r}_A e^{-\frac{\mathbf{x}^2}{2\sigma_A}} + b_A} - \sum_B \bar{W}_{AB} \left(r_B \frac{\sigma_B}{\sigma_{AB+}} e^{-\frac{x^2}{2\sigma_{AB+}}} + b_B \right) - T_A \right]^2 \end{aligned} \tag{19}$$

where we defined $\sigma_{AB+} = \sigma_{AB} + \sigma_B$. It is easy to verify that

$$\sqrt{r_A e^{-\frac{\mathbf{x}^2}{2\sigma_A}} + b_A} \simeq q_A e^{-\frac{\mathbf{x}^2}{4\sigma_A}} + \sqrt{b_A}, \tag{20}$$

where we defined $q_A = \sqrt{r_A + b_A} - \sqrt{b_A}$, is a very good approximation. The NNLS equations consist in minimizing the loss function (which codifies the error between the data and the model) with respect to the parameters. Thus when we minimize

[1] we make repeated use of the useful integrals:

$$\begin{aligned} \int e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2\sigma_1}} e^{-\frac{\mathbf{y}^2}{2\sigma_2}} d\mathbf{y} &= \frac{2\pi\sigma_1\sigma_2}{\sigma_1+\sigma_2} e^{-\frac{\mathbf{x}^2}{2(\sigma_1+\sigma_2)}} \\ \int e^{-\frac{\mathbf{x}^2}{2\sigma_1}} e^{-\frac{\mathbf{x}^2}{2\sigma_2}} d\mathbf{x} &= \frac{2\pi\sigma_1\sigma_2}{\sigma_1+\sigma_2} \end{aligned} \tag{18}$$

with respect to $\bar{W}_{AB}$ we obtain:

$$-2\int_{-\infty}^{\infty} d\mathbf{x}\left[q_A e^{-\frac{\mathbf{x}^2}{4\sigma_A}} - \sum_C \bar{W}_{AC}\left(\bar{r}_C \frac{\sigma_C}{\sigma_{AC+}} e^{-\frac{\mathbf{x}^2}{2\sigma_{AC+}}} + b_C\right) + \sqrt{b_A} - T_A\right]\left(\bar{r}_B \frac{\sigma_B}{\sigma_{AB+}} e^{-\frac{\mathbf{x}^2}{2\sigma_{AB+}}} + b_B\right) = 0$$

$$\frac{\bar{r}_B \sigma_B 2\sigma_A q_A}{\sigma_{AB+} + 2\sigma_A} - \bar{r}_B \sigma_B \sum_C \bar{W}_{AC} \frac{\bar{r}_C \sigma_C}{\sigma_{AB+} + \sigma_{AC+}} +$$

$$-\bar{r}_B \sigma_B \sum_C \bar{W}_{AC} b_C + \bar{r}_B \sigma_B (\sqrt{b_A} - T_A) + b_B q_A 2\sigma_A - b_B \sum_C \bar{W}_{AC} \bar{r}_C \sigma_C +$$

$$-\int_{-\infty}^{\infty} d\mathbf{x} \frac{b_B}{2\pi}\left(\sum_C W_{AC} b_C - (\sqrt{b_A} - T_A)\right) = 0 \tag{21}$$

Note that the second derivatives are always negative, thus the problem is always convex. Also note that these are 6 equations for each of the $\bar{W}_{AB}$ with $B = \{E, P, S, V, L, H\}$, i.e. one equation for each of the connection weights that enter into the self-consistent equation for the firing rate of population A. Since the space is continuous and infinite, to ensure convergence of the loss function, we have to impose:

$$\sqrt{b_A} - T_A = \sum_C \bar{W}_{AC} b_C \tag{22}$$

Conceptually this means that if we knew the $\bar{W}_{AB}$, then the knowledge of the baseline firings (and the transfer functions) would determine the firing thresholds. Moreover when we take the first derivative of the loss function with respect to T_A and we take into account Eq22 we obtain:

$$q_A 2\sigma_A - \sum_C \bar{W}_{AC} \bar{r}_C \sigma_C = 0 \tag{23}$$

Finally using both Eq22 and Eq23 into Eq21 we get

$$\sum_C (\bar{W}_{AC} \bar{r}_C \sigma_C)\left(\frac{1}{\sigma_{AB+} + 2\sigma_A} - \frac{1}{\sigma_{AC+} + \sigma_{AB+}}\right) = 0 \tag{24}$$

We can solve this linear equations to find $\bar{W}_{AC}$ and then find T_A from Eq.22. If we solve the linear system in Eq 24 without any further caution, we find an error function that is strictly 0, but values of the parameters that are not compatible with the excitatory or inhibitory nature of the different populations. Therefore we devise an algorithm to solve this system of linear equations with constraints on the positivity of each element W_{AB}. We start from a random initial set of W_{AB} that satisfies our positivity constraints and solve one equation at a time. We update the initial parameter set if two conditions are verified: i) the solution still satisfies the positivity constraints and ii) the loss function is actually smaller than the initial

value. This simple algorithm allows to find a family of solutions, compatible with the results from the simulation of the Eq.16 where the $\bar{W}_{AB}$ and the T_A are fixed and the $\bar{r}_A$ and b_A are the steady states reached by the system. The advantages with respect to the minimal model are that we are not assuming stereotyped Gaussian profiles for the recurrent populations and that the size tuning curves of all recurrent cell types are self-generated, once the parameters are optimized. A limitation of this approach is that the additional inputs that we are explicitly ignoring (e.g. other layers of V1) will be compensated for by the recurrent populations, thus conditioning the meaning of the effective synaptic coupling strengths.

Next, we repeat the procedure when constraining on all stimulus sizes in the classical stimulus condition. The loss function in this case reads:

$$S_A = \int ds \int d\mathbf{x} \left[\sqrt{r_A(\mathbf{x}, s)} - \sum_{B=\{E,P,S,V,L,H\}} \int W_{AB}(\mathbf{x}-\mathbf{y}) r_B(\mathbf{y}, s) d\mathbf{y} - T_A \right]^2 \tag{25}$$

where again $r_A(\mathbf{x}, s) = \bar{r}_A(s) e^{-\frac{\mathbf{x}^2}{2\sigma_A}} + b_A$, and we can follow all the same steps as above, until the very last one, where we cannot simplify due to the $\int ds$. We obtain:

$$\begin{aligned} \sqrt{b_A} - T_A - \sum_C \bar{W}_{AC} b_C = 0 \\ 2 \int ds \frac{\sigma_A(s)\sigma_B(s)}{2\sigma_A(s) + \sigma_{AB+}(s)} q_A(s) \bar{r}_B(s) - \sum_C \bar{W}_{AC} \int ds \frac{\sigma_B(s)\sigma_C(s)}{\sigma_{AB+}(s) + \sigma_{AC+}(s)} \bar{r}_C(s) \bar{r}_B(s) = 0 \end{aligned} \tag{26}$$

Which, as before, we solve numerically for $\bar{W}_{AC}$ with sign constraints.

Note that constraining on inverse size tuning curves would not introduce any conceptual complication (just more terms of the same type to calculate) but here we decide to not infer the model parameters to fit direct and inverse response, but instead we fit only the direct response and then benchmark against the ability to generate inverse size-dependent responses.

SOM cells aligned with the stimulus center respond strongly to large classical stimuli, suggesting that they receive a large excitatory current for such stimuli. Nevertheless, both L2/3 Pyramidal neurons and HVAs neurons are very surround suppressed, thus providing a much smaller excitatory current for large stimuli than for small stimuli. This suggests that SOM cells receive extra excitatory input, that increases with stimulus size. Since, to the best of our knowledge, empirical observations on the origin of such external current are still missing, in order to recover the size tuning curve and spatial profiles of SOM neurons we make a minimal assumption that they receive an extra excitatory current whose amplitude grows linearly with stimulus size. We then infer the effective strength of this interaction

by adding one external field X in the NNLS equations:

$$S_A = \int ds \int d\mathbf{x} \left[\sqrt{r_A(\mathbf{x}, s)} - \sum_{B=\{E,P,S,V,L,H,X\}} \int W_{AB}(\mathbf{x}-\mathbf{y}) r_B(\mathbf{y}, s) d\mathbf{y} - T_A \right]^2 \tag{27}$$

with

$$\begin{aligned} r_X(\mathbf{y}, s) &= \sqrt{s} e^{-\frac{\mathbf{x}^2}{2\sigma_X^2}} \\ W_{AX}(\mathbf{x}-\mathbf{y}) &= \delta_{AS} \frac{\bar{W}_{AX}}{2\pi\sigma_{AX}} e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2\sigma_{AX}}} \end{aligned} \tag{28}$$

The choice of σ_{37} and σ_7 is arbitrary and is made here following arguments of simplicity and analogy with other excitatory input currents ($\sigma_{37} = 8$, $\sigma_7 = 20$). We note that a different choice of these values would only weakly affect the geometry of the SOM rate field.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data preprocessing

The distance between the center of the stimulus and the center of the CRF of cells is estimated in (Keller et al., 2020b) through CRF mapping experiments, where a stimulus is presented at many positions spanning the area of the visual field accessible for recording and the response of thousands of neurons are recorded through 2-photon calcium imaging (Keller et al., 2020b). Since the distance between contiguous presentations of the stimulus is 5°, we take this value to be the error on the measure of the CRF location. For each cell we then take a Gaussian in its spatial location with $\sigma_{err} = 5°$ and with amplitude equal to the recorded (trial averaged) response of that cell, i.e. $r_i e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2v_{err}}}$ with $v_{err} = \sigma_{err}^2$. We then sum the Gaussian functions for all recorded cells and divide by the sum of unitary Gaussians, to account for the non-uniform distribution of data-points. The rate field of population A is thus given by:

$$\tilde{r}_A(\mathbf{x}, s) = \frac{\sum_{i=1}^{N_A} r_i(s) e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2v_{err}}}}{\sum_{i=1}^{N_A} e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2v_{err}}}} \tag{29}$$

where N_A is the number of recorded cells of type A and s indicates the stimulus size. In Supplementary Section S1 A we show an example of the smoothening that this procedure (first presented in (Dipoppa et al., 2018)) generates. Rate fields are plotted in 1 spatial dimension, but they are in fact 2-dimensional and isotropic in retinotopic space (see e.g. Supplementary Section 3). The retinotopic space is considered continuous, representing the limit of very large number of neurons.

Functional form of the rate fields

In order to be able to understand the system analytically we take the approximation:

$$\tilde{r}_A^{(c)}(\mathbf{x}, s) \simeq r_A^{(c)}(\mathbf{x}, s) = p_A^c(s) e^{-\frac{\mathbf{x}^2}{2\bar{v}_A^d(s)}} \tag{30}$$

where $A = E, P, S, V, L, M$ indicate respectively populations of Pyramidal, PV, SOM, VIP, Layer 4, LM cells and the superscript $^{(c)}$ stands for classical. The prefactor $p_A^c(s)$ represents the size tuning curve and the spatial dependence is purely Gaussian. Operatively, we fit the experimental rate fields for each stimulus condition with a Gaussian function. We report the goodness of fits applied to both preprocessing procedures in Supplementary Section S1 C.

For inverse rate fields the response profiles are more complex, therefore we schematize them through a difference of Gaussian functions:

$$\tilde{r}_A^{(i)}(\mathbf{x}, s) \simeq r_A^{(i)}(\mathbf{x}, s) = p_A^i(s) e^{-\frac{\mathbf{x}^2}{2\bar{v}_{pA}^i(s)}} - q_A^i(s) e^{-\frac{\mathbf{x}^2}{2\bar{v}_{qA}^i(s)}}, \tag{31}$$

with $p_A^i(s), q_A^i(s) > 0$.

Note that in the inverse stimulus condition we define scaling as the growth of the width of the outer Gaussian of the fit when stimulus size is increased. We do not distinguish between outer and inner scaling because they appear to be very similar. The difference between outer and inner scaling is mostly very small (less that 1 degree) and:

$$\frac{\bar{v}_{pA}^i - \bar{v}_{qA}^i}{\bar{v}_{pA}^i} < 0.05\%. \tag{32}$$

Moreover its distribution is very narrow.

In order to have a full functional description of the rate fields as a function of the classical stimulus size we need to define the size tuning curve $p_A^c(s)$. This will allow to give an analytical account of surround suppression by analyzing the interval where $\frac{\partial p_A^d(s)}{\partial s} < 0$ and preferred size (or CRF size) as $s^* : \frac{\partial p_A^d(s)}{\partial s}|_{s=s^*} = 0$ (see Supplementary Section S7). We use the parametrization:

$$\begin{aligned} p_A^d(s) &= \beta_1^A \frac{s}{S_{\beta 1}^A} - \beta_2^A \frac{s}{S_{\beta 2}^A} \\ \bar{v}_A^d(s) &= k_{d1}^A + k_{d2}^A s \end{aligned} \tag{33}$$

where the first equation encodes the size tuning curve and $\gamma_A = \beta_2^A / \beta_1^A$ modulates the amount of surround suppression, $S_{\beta 1}^A$ modulates the steepness of the increase of size tuning for small sizes and $^A_{\beta 2}$ modulates the steepness of the decrease of size tuning for large sizes. Moreover $k_{d2}^A > 0$ indicates the growth of the rate field width with stimulus size. The size tuning curves described by the functional form in 33, as well as the data and the fitted samples of the rate fields are shown in Supplementary Section S1 D.

For the inverse stimuli we need to specify the functions $p_A^i(s)$ and $q_A^i(s)$ for the populations that we are going to use as inputs for the minimal model, i.e.

$A = S, L, M$. We devise such functional forms and show a comparison with the experimental rate fields.

For SOM population we use the parametrization:

$$\begin{aligned} p_S^i(s) &= \alpha_1^S \frac{s}{S_{\alpha 1}^S} - \alpha_2^S \frac{s}{S_{\alpha 2}^S} \\ q_S^i(s) &= \rho_1^S \frac{s}{S_{\rho 1}^S} - \rho_2^S \frac{s}{S_{\rho 2}^S} \end{aligned} \tag{34}$$

The evidence presented in (Keller et al., 2020b) suggest that HVAs highly influence the inverse size-dependent response profiles, and on the contrary, the input from L4 is very weak in the inverse stimulus condition, due to classical surround suppression, thus while we focus on reproducing accurately the spatial nuances of LM, but we allow less accuracy for L4, in order to keep the model simple and the number of parameters as small as possible (for a more nuanced parametrization of the inputs see Supplementary Section S12). Thus for L4 cells we use:

$$p_L^i(s) = q_L^i(s) = \alpha_1^L \frac{s}{S_{\alpha 1}^L} - \alpha_2^L \frac{s}{S_{\alpha 2}^L} \tag{35}$$

that constrains the center to have zero response, consistently with the very low response of aligned cells in L4 shown in (Keller et al., 2020b). Finally for LM we use:

$$\begin{aligned} p_M^i(s) &= \alpha_1^M + \alpha_2^M s \\ q_M^i(s) &= p_M^i(s) - \rho_1^M \left(\frac{s}{S_{\rho 1}^M} - \frac{s}{S_{\rho 2}^M} \right) \end{aligned} \tag{36}$$

Note that the size tuning curve of the aligned cells can then be described as $p_A^i(s) - q_A^i(s)$. The size tuning curve of the offset cells is here defined as: $p_A^i(s) e^{-\frac{\tilde{\mathbf{x}}^2}{2\bar{v}_{pA}^i(s)}} - q_A^i(s) e^{-\frac{\tilde{\mathbf{x}}^2}{2\bar{v}_{qA}^i(s)}}$ with $\tilde{\mathbf{x}} = (30°, 30°)$. Aligned and offset size tuning curves pre and post parametrization are shown in Supplementary section S1 D.

Finally the growth of the inner and outer scaling is given by:

$$\begin{aligned} \bar{v}_{pA}^i(s) &= k_{i1}^A + k_{i2}^A s \\ \bar{v}_{qA}^i(s) &= k_{i3}^A + k_{i2}^A s \end{aligned} \tag{37}$$

The recorded rate field of LM shows a much lower amplitude in the inverse stimulus condition versus the classical one (as evaluated for example computing the integral over space of the rate field). If the input from HVAs consisted solely of LM contribution, we would not be able to recover an inverse response in L2/3 with firing rates as high as or even higher than the classical case, reported in (Keller et al., 2020b). But we know that V1 receives input from all HVAs. Moreover Extended Data Fig.9 in (Keller et al., 2020b) shows that optogenetic silencing of individual HVAs generates a quantitatively different reduction of classical versus

inverse responses for each HVAs considered. Thus, to quantitatively recover the relative amplitude of the rate field in L2/3, we take the minimal assumption that HVAs considered all together have the same rate field profiles of LM, but with an amplitude that is inferred according to the amplitude of the inverse response of L2/3 cells. In formulas:

$$r_H^{(i)}(\mathbf{x}, s) = \eta r_M^{(i)}(\mathbf{x}, s) \tag{38}$$

and from here on we will consider $A = \{E, P, S, V, L, H\}$, where H is a population representing the conjunction of all HVAs (see Supplementary Section S2).