

# UCSF

## UC San Francisco Previously Published Works

### Title

A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex

### Permalink

<https://escholarship.org/uc/item/35m0c6s3>

### Journal

Neuron, 108(6)

### ISSN

0896-6273

### Authors

Keller, Andreas J  
Dipoppa, Mario  
Roth, Morgane M  
et al.

### Publication Date

2020-12-01

### DOI

10.1016/j.neuron.2020.11.013

Peer reviewed



Published in final edited form as:

*Neuron*. 2020 December 23; 108(6): 1181–1193.e8. doi:10.1016/j.neuron.2020.11.013.

## A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex

**Andreas J. Keller**<sup>1,3,8,\*</sup>, **Mario Dipoppa**<sup>4,8,\*</sup>, **Morgane M. Roth**<sup>1,3,8,\*</sup>, **Matthew S. Caudill**<sup>2,3,6,8</sup>, **Alessandro Ingrassia**<sup>4,7</sup>, **Kenneth D. Miller**<sup>4,5,9,\*</sup>, **Massimo Scanziani**<sup>1,2,3,9,10,\*</sup>

<sup>1</sup>Department of Physiology, University of California San Francisco, San Francisco, California 94158-0444, USA.

<sup>2</sup>Center for Neural Circuits and Behavior, Neurobiology Section and Department of Neuroscience, University of California San Diego, La Jolla, California 92093-0634, USA.

<sup>3</sup>Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California, USA.

<sup>4</sup>Center for Theoretical Neuroscience, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, New York 10027, USA.

<sup>5</sup>Dept. of Neuroscience, Swartz Program in Theoretical Neuroscience, Kavli Institute for Brain Science, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, New York, USA

<sup>6</sup>Present Address: Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA, and Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA.

<sup>7</sup>Present Address: Quantitative Life Sciences, The Abdus Salam International Centre for Theoretical Physics - ICTP, Trieste 34151, Italy

<sup>8</sup>These authors contributed equally

<sup>9</sup>These authors contributed equally

<sup>10</sup>Lead Contact

### SUMMARY

\*Correspondence: andreasjakob.keller@ucsf.edu (A.J.K.), md3681@columbia.edu (M.D.), morgane.roth@ucsf.edu (M.M.R.), kdm2103@columbia.edu (K.D.M.), massimo@ucsf.edu (M.S.).

**Author contributions.** M.S., A.J.K., and M.S.C. designed the experimental study. A.J.K. and M.M.R. conducted all experiments and experimental data analysis. M.S.C. performed preliminary experiments. M.D. and K.D.M. designed the model. M.D. and A.I. developed the training algorithm of the model. M.D. performed the numerical simulations and, with K.D.M., analyzed the model results. A.J.K., M.M.R., M.D., K.D.M., and M.S. wrote the manuscript.

**Declaration of interests.** The authors declare no competing interests.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Context guides perception by influencing stimulus saliency. Accordingly, in visual cortex, responses to a stimulus are modulated by context, the visual scene surrounding the stimulus. Responses are suppressed when stimulus and surround are similar but not when they differ. The underlying mechanisms remain unclear. Here we use optical recordings, manipulations, and computational modelling to show that disinhibitory circuits consisting of vasoactive intestinal peptide expressing (VIP) and somatostatin expressing (SOM) inhibitory neurons modulate responses in mouse visual cortex depending on similarity between stimulus and surround, primarily by modulating recurrent excitation. When stimulus and surround are similar, VIP neurons are inactive and activity of SOM neurons leads to suppression of excitatory neurons. However, when stimulus and surround differ, VIP neurons are active, inhibiting SOM neurons, which leads to relief of excitatory neurons from suppression. We have identified a canonical cortical disinhibitory circuit which contributes to contextual modulation and may regulate perceptual saliency.

## eTOC Blurp

Context provides meaning by influencing perception. In the visual world, context is the visual environment surrounding a visual scene. Here, Keller et al. report that a canonical disinhibitory circuit controls the response of mouse visual cortex to a visual stimulus depending on the context within which that stimulus is presented.

---

## INTRODUCTION

The perception of a sensory stimulus is markedly influenced by the context in which the stimulus is embedded. In the visual system, the context is the visual scene surrounding the stimulus. Through the influence of its surround, the same visual stimulus may be perceived as more or less salient, allowing it to pop out or merge with the rest of the visual scene (Figure 1A; Bergen and Julesz, 1983; Lamme, 1995; Treisman and Gelade, 1980). This aspect of sensory processing represents a fundamental computation to extract meaning from visual scenes.

Consistent with perceptual phenomena, neuronal responses to a visual stimulus are modulated by the visual scene surrounding the stimulus. This surround modulation occurs at several stages of the visual system including the retina (Alitto and Usrey, 2008; Chiao and Masland, 2003; Huang et al., 2019; McIlwain, 1964; Ölveczky et al., 2003; Solomon, 2006), the thalamus (Alitto and Usrey, 2008; Jones et al., 2012, 2015; Levick et al., 1972), and the visual cortex (Alexander and Van Leeuwen, 2010; Angelucci et al., 2017; Fitzpatrick, 2000; Kapadia et al., 2000; Knierim and van Essen, 1992; Rossi et al., 2001; Schnabel et al., 2018; Sillito et al., 1995), progressively increasing the complexity of the spatial features that are contextualized.

The classical feedforward receptive field (ffRF) of a neuron in primary visual cortex (V1) is the region in space in which a visual stimulus evokes a response (Hubel and Wiesel, 1962). The magnitude of this response can be modulated by stimulating the regions surrounding the ffRF. When a stimulus is large enough to cover both the ffRF and its surround, for example, the neuron's responses are generally suppressed. This phenomenon, called surround

Author Manuscript

suppression, is a well-established example of surround modulation (Blakemore and Tobin, 1972; Hubel and Wiesel, 1965; Kapadia et al., 1999; Knierim and van Essen, 1992; Nelson and Frost, 1978). It has been shown that anatomical substrates for surround suppression include feedback connections (Angelucci et al., 2017; Keller et al., 2020; Nurminen et al., 2018; Vangeneugden et al., 2019; Zhang et al., 2014), interlaminar connections (Bolz and Gilbert, 1986) and specific subtypes of inhibitory neurons (Adesnik et al., 2012; Haider et al., 2010), such as SOM inhibitory neurons. Indeed, the tuning properties of SOM inhibitory neurons (Adesnik et al., 2012; Dipoppa et al., 2018; Keller et al., 2020; Ma et al., 2010) and the fact that they connect to nearly all nearby excitatory neurons (Fino et al., 2013) make them ideal to contribute to surround suppression. Accordingly, functional elimination of SOM neurons partially relieves excitatory neurons from surround suppression (Adesnik et al., 2012).

Author Manuscript

However, not all combinations of stimuli in the fRF and surround generate suppression. Surround suppression occurs when the stimulus in the fRF and in the surround share similar features. For example, the response of a neuron to a grating stimulus of a given orientation in its fRF is suppressed when stimulating the surround with a grating of similar orientation. When the orientation of the grating in the surround differs from that in the fRF, the response of the neuron is much less or no longer suppressed (Coen-Cagli et al., 2015; Self et al., 2014; Sillito et al., 1995; Walker et al., 1999). Thus, the magnitude of surround suppression depends on the visual scene surrounding the stimulus in the fRF. The mechanism that regulates surround suppression depending on the similarity between the stimulus in the fRF and that in the surround remains elusive. We refer to this phenomenon as “contextual modulation”.

Author Manuscript

Author Manuscript

To investigate the mechanisms of contextual modulation, we presented visual stimuli with different surrounds to awake mice while imaging calcium responses in excitatory and inhibitory neurons of V1. We focused on the three major classes of inhibitory neurons, parvalbumin-expressing (PV), SOM and VIP neurons (Lee et al., 2010; Pfeffer et al., 2013; Tremblay et al., 2016). PV and SOM neurons are the two principal sources of inhibition of cortical excitatory neurons in mouse V1. In contrast, VIP neurons primarily provide inhibition to SOM neurons, thus representing a key component of cortical disinhibitory circuits (Jiang et al., 2015; Karnani et al., 2016; Millman et al., 2020; Pfeffer et al., 2013; Pi et al., 2013). We show that, as for excitatory neurons, the responses of VIP and PV neurons were suppressed by surrounds that shared similar features to the stimulus presented in the fRF but not when they differed. Strikingly, the responses of SOM neurons were modulated in a manner opposite to all other neuron types, being specifically suppressed by surrounding stimuli that differ from those in the fRF. This suggests that the suppression of SOM neurons to these stimuli may be mediated by VIP neurons. To determine whether the interaction between VIP and SOM neurons could account for the contextual modulation observed in excitatory neurons, we developed a circuit model respecting biological constraints, which we trained to reproduce our measurements. The modelling results suggest that the VIP-SOM disinhibitory circuit plays a key role in contextual modulation through its role in complex recurrent-circuit dynamics. Specifically, our model predicts that silencing VIP neurons reduces contextual modulation in excitatory neurons. Indeed, when VIP neurons in V1 were silenced optogenetically, surround suppression in excitatory neurons became less sensitive to

the stimulus features in the surround, which reduced contextual modulation. Thus, we show that a canonical cortical disinhibitory circuit contributes to the contextual modulation of excitatory neurons in V1.

## RESULTS

### Contextual modulation in excitatory neurons

To assess contextual modulation in V1, we used two-photon calcium imaging to record activity from layer 2/3 (L2/3) excitatory neurons in awake head-fixed mice. Contextual modulation was assessed by comparing the baseline-subtracted responses of individual neurons to small patches of drifting sinusoidal oriented gratings presented alone (“center stimulus”), or together with two different surrounds: An iso-oriented surround (“iso stimulus”; i.e. a grating in the surround whose orientation and phase matches that of the grating in the center), or a cross-oriented surround (“cross stimulus”; i.e. a grating in the surround whose orientation is orthogonal relative to that of the grating in the center; Figure 1B, top). The location of the center stimulus was centered on the fRFs of the neurons (see STAR Methods). The magnitude of the response of L2/3 excitatory neurons to center stimuli alone was larger than that to iso stimuli, consistent with iso stimuli generating surround suppression (Figure 1B, left). In contrast, the response to cross stimuli was similar to the response to center stimuli alone, consistent with the fact that cross stimuli generate less surround suppression than iso stimuli, or no surround suppression, as previously described (Self et al., 2014; Sillito et al., 1995; Walker et al., 1999). We computed the magnitude of surround suppression as the difference in response to center stimuli and response to iso or cross stimuli, normalized by the response to center stimuli. Accordingly, surround suppression in L2/3 excitatory neurons was larger for iso stimuli than for cross stimuli (Figure 1C, D). To compare the modulation by the iso surround to that of the cross surround, we defined a contextual modulation index (CMI) for each neuron as the difference divided by the sum of the responses to cross and iso stimuli (Figure 1E; see STAR Methods). The distribution of CMIs of excitatory neurons was skewed to positive values, indicating that their responses were stronger to the cross than to the iso stimulus. Since the distribution of CMIs was similar irrespective of whether or not the orientation of the center stimulus matched the neuron’s orientation preference (Figure S1A), our analysis includes neurons independently of their orientation preference. Moreover, behavioral states of the mice only had a minor effect on the contextual modulation of excitatory neurons (Figure S1C–F). Thus, data were pooled irrespective of behavioral state. Overall, excitatory neurons in L2/3 were strongly modulated by context, i.e. the strength of their responses depended on the stimulus features in the surround relative to those in the center.

To what extent is the contextual modulation of excitatory L2/3 neurons inherited from earlier stages of cortical processing? To answer this question, we measured the responses of excitatory neurons in layer 4 (L4), the main thalamic input layer, to center, iso and cross stimuli (Figure 1B, right). While L2/3 neurons, on average, were only suppressed by the iso stimulus, L4 neurons showed suppression in response to both iso and cross stimuli (Figure 1C). Thus, contextual modulation of L2/3 neurons is unlikely to be entirely inherited from L4 and may rely on local circuitry.

## Complementary contextual modulation in SOM and VIP neurons

What relieves L2/3 excitatory neurons from surround suppression when the stimulus in the surround differs from the stimulus in the center? Since surround suppression of L2/3 excitatory neurons relies, at least in part, on the activation of SOM inhibitory neurons (Adesnik et al., 2012), we compared the response of SOM neurons to iso and cross stimuli. We thus repeated the visual stimulation protocol used above while recording in SOM neurons (Figures 2A–C, S1B–F). Strikingly, the responses of SOM neurons to iso and cross stimuli were opposite to what we observed in excitatory neurons. While iso stimuli elicited strong responses in SOM neurons, as previously observed (Adesnik et al., 2012; Dipoppa et al., 2018; Keller et al., 2020), cross stimuli elicited smaller responses (Figure 2B). Accordingly, the distribution of their CMI was shifted towards negative values (Figure 2C). The smaller response of SOM neurons to cross than to iso stimuli was not a general characteristic of inhibitory neurons. PV neurons, the other large class of inhibitory neurons that targets excitatory neurons in mouse V1 (Pfeffer et al., 2013), showed larger responses to cross than to iso stimuli (Figures 2D, E, S1B, C). Therefore, the distribution of their CMI was shifted towards positive values (Figures 2F, S1D–F), similar to excitatory neurons and opposite to SOM neurons. Thus, SOM neurons are unique in the way they respond to cross and iso stimuli.

What prevents SOM neurons from responding to cross as much as to iso stimuli? SOM neurons receive excitatory input from L2/3 neurons. Given that L2/3 neurons strongly respond to cross stimuli, it is unlikely that the excitatory input to SOM neurons is reduced in response to cross stimuli. We thus reasoned that cross stimuli may generate inhibition onto SOM neurons. VIP inhibitory neurons are a class of cortical neurons that preferentially inhibits other inhibitory neurons, including SOM neurons (Pfeffer et al., 2013). If VIP neurons prevent SOM neurons from responding to cross but not to iso stimuli, they should be more excited by cross than by iso stimuli. To test this hypothesis, we repeated the visual stimulation protocol used above while recording in VIP neurons. Consistent with our prediction, VIP neurons responded more strongly to cross than to iso stimuli, as shown by their positively shifted CMI (Figures 2G–I, S1B–F).

Taken together, these results are consistent with a mechanism in which the response modulation by the visual stimulus surrounding the fRF of excitatory neurons is controlled by a disinhibitory circuit formed by the reciprocal inhibition of VIP and SOM neurons (Figure 2J).

## A circuit model reproduces contextual modulation across neuron types

To test our intuition that the VIP-SOM disinhibitory circuit contributes to contextual modulation in L2/3 excitatory neurons, we developed a circuit model in which the model ‘units’ had supralinear input-output functions, consistent with experimental results (Adesnik, 2017; Priebe and Ferster, 2008; Priebe et al., 2004). Each unit of the circuit represented the average activity of a given neuron type (i.e. L2/3 excitatory, VIP, SOM, and PV neurons and L4 excitatory neurons), integrated in a ‘subnetwork’ with the other unit types (Figure 3A). Four such subnetworks were each assigned to one of two spatial locations (each considered the ‘surround’ of the other) and one of two preferred orientations (that were orthogonal to

each other; Figure 3B). For the units sharing the same spatial location (both within and across subnetworks), we allowed all connections except those known to be weak (Adesnik et al., 2012; Karnani et al., 2016; Pfeffer et al., 2013; Tremblay et al., 2016), similar to a previous model (Lee et al., 2017). Subnetworks across spatial locations were connected only through L2/3 excitatory projections. We will mostly focus on the subnetwork whose location and preferred orientation match the center stimulus, which we will refer to as the “centered and preferred-orientation subnetwork”.

We trained the model to reproduce our measurements. To this end, we let the network evolve over time (see STAR Methods). Thereby, we obtained the fixed-point responses, that are the final, time-invariant responses to which the network converges for a given stimulus. These fixed-point responses were then matched to the experimental responses by optimizing synaptic strengths. To determine the optimal synaptic strengths, we used a two-step procedure. We first generated many candidate solutions by performing non-negative regression (non-negative least squares), similarly to a previous study (Dipoppa et al., 2018), but on many sets of pseudo data obtained by randomly perturbing the experimental data. We then used the best solutions as initial conditions for a gradient-based optimization in a recurrent neural network (RNN; backpropagation through time with convolutional connections; Sporer et al., 2017; see STAR Methods). The top 115 solutions with the closest fits to the experimental data were used for further analysis (Figures 3C, S2A; see STAR Methods).

One feature of these solutions was their strong recurrent excitatory connections within a subnetwork (Figures 3D, S2B), consistent with previous experimental observations (Cossell et al., 2015; Hofer et al., 2011; Ko et al., 2011; Peron et al., 2020). Due to these strong excitatory connections, the fixed points of most solutions would be unstable if the inhibitory activities were frozen at their fixed-point levels, but the network is stabilized by feedback inhibition. That means that most of the networks that best fit the data are inhibition-stabilized networks (ISN; of the 115 top solutions, 72%, 98%, 88%, and 99% were ISNs at the fixed point for spontaneous, center, iso, and cross input, respectively), as has been found to underlie surround suppression (Adesnik, 2017; Ozeki et al., 2009) and spontaneous activity (Sanzeni et al., 2020). In addition, the combination of the supralinear input-output function and the inhibition stabilization implies that the circuit is a supralinear stabilized network, which can account for many nonlinear response properties of visual cortex through network recurrence (Ahmadian et al., 2013; Hennequin, 2018; Rubin et al., 2015).

### **The disinhibitory circuit is both necessary and sufficient for contextual modulation**

To determine the circuit elements critical to contextual modulation, we perturbed specific inputs, starting from the fixed-point responses to an iso stimulus, and observed the resulting changes in network responses (Figure 4).

To examine the transition of the network from iso to cross fixed-points, we simply switched the L4 inputs to their cross-response levels (Figure 4A). As expected, the whole network evolved from the iso to the cross fixed-point, in which excitatory, PV and VIP units' firing increased and, for most of the solutions, SOM units' firing decreased (Figure 4B). To understand the evolution of the excitatory units' firing during this transition, we looked at



the change in the inputs they received between the iso and cross fixed-points. The largest change in inputs to excitatory units was an increase in their recurrent excitatory inputs (Figure 4C). Surprisingly, the inhibition that the excitatory units received from the SOM units changed only modestly between the iso and cross fixed-points (Figure 4C). However, during the transition from the iso to the cross fixed points, the activity of SOM units was typically biphasic, first displaying a large change from their iso-response level but then returning closer to their initial response-level (Figure 4D). This suggests that SOM units could play an important role in contextual modulation, even though their contribution is largely hidden in the network dynamics. Thus, the disinhibitory pathway could play a key role in the facilitation of excitatory response from their iso- to cross-response levels, primarily by modulating recurrent excitation rather than direct suppression.

Is the disinhibitory pathway necessary to drive the transition from iso- to cross-response levels? To answer this question, we switched the L4 inputs from their iso- to their cross-response levels as above but kept the input from VIP to SOM units frozen to their iso-response level (Figure 4E). Interestingly, this simple manipulation prevented the change in the activity and in inputs of all the units (Figure 4F, G), despite some minor transient fluctuations (Figure 4H). The disinhibitory pathway is therefore necessary to drive the transition from iso- to cross-response levels. But is the disinhibitory pathway also sufficient? To address this question, we kept the L4 inputs at their iso-response level but switched and froze the input from VIP to SOM units to their cross-response level (Figure 4I). The manipulation of this single input between these two inhibitory units influenced all the units in the network (Figure 4J, K). Interestingly, it reproduced the changes in activity observed when changing the L4 inputs to the cross-responses level as well as the change in inputs to excitatory units, but with a larger initial transient decrease in SOM-unit activity (Figure 4L). The robustness of this effect was further tested using a small drive onto the VIP unit instead of the large perturbation of changing its input to the SOM unit from its iso- to its cross-response level. Adding a tiny excitatory input to the VIP unit (Figure S3 vs Figure 4I–L) led to qualitatively similar results. Thus, activation of the disinhibitory pathway is both necessary and sufficient to produce the transition from iso- to cross-response levels.

During the transition from iso- to cross-response levels, the most prominent difference in inputs to excitatory units is the recurrent excitation. We therefore asked if this connection is critical to contextual modulation. Freezing the L2/3 excitatory to excitatory input to the iso-response level, while switching L4 input to the cross level, led the excitatory activities to return to response levels similar to their iso-response level, thus demonstrating the necessity of this connection (Figure S4A). However, keeping the L4 inputs at the iso-response level while switching and freezing L2/3 excitatory to excitatory inputs to the cross-response level was typically only sufficient for excitatory units to settle near their cross-response levels, but not for the rest of the network (Figure S4B). We next asked if any other inputs were sufficient to promote the transition from iso- to cross-response levels (Figure S4C–K). While one other input could move excitatory units in the right direction (Figure S4G), excitatory-to VIP-unit input was the only input besides the VIP- to SOM-unit input that, if frozen to its cross level, was sufficient to produce excitatory cross responses and move all firing rates in the same direction as the transition from iso- to cross-response levels (Figure S4E). This



again emphasizes the role of the disinhibitory circuit and in particular of the VIP unit in contextual modulation.

### **The model predicts that silencing VIP neurons reduces contextual modulation**

If disinhibition from VIP units plays a critical role in contextual modulation, then silencing VIP units should greatly reduce contextual modulation. To test this, we set the activity of VIP units to zero (Figure 5A). Silencing VIP units caused a larger absolute decrease in responses of excitatory units to cross than to iso stimuli for almost all solutions (Figures 5B). While PV units were affected similarly to excitatory units, SOM units showed the opposite changes (Figure 5B). In principle, a stronger absolute reduction in responses of excitatory units to cross than to iso stimuli is consistent with two possibilities. VIP units could simply regulate the overall gain in the network, that is, having the same relative impact on the responses of excitatory units to cross and iso stimuli. Alternatively, they could differentially regulate the responses of excitatory units depending on the stimulus. To distinguish between these two possibilities, we compared the CMI of the different units under control conditions with their CMI during the silencing of VIP units. Consistent with VIP units differentially regulating the responses to iso and cross stimuli, their silencing consistently decreased the CMI of excitatory units in the 115 solutions that most accurately fit the experimental data, indicating that cross responses decreased proportionately more than iso responses (Figure 5C). While PV units showed a decrease similar to excitatory units, the CMI of SOM units increased upon silencing of VIP units (Figure 5C).

The decrease in CMI upon silencing of VIP units was a robust feature of those solutions that most accurately replicated our experimental data (Figure S5). Thus, an important test of the solutions that best fit the data is whether silencing of VIP neurons in mouse V1 decreases the CMI of excitatory neurons.

### **Silencing of VIP neurons reduces contextual modulation**

Does the functional elimination of VIP neurons preferentially decrease the response of excitatory neurons to cross stimuli compared to iso stimuli as predicted by the model? Since excitatory neurons are already almost maximally suppressed by iso stimuli, we reduced the contrast of all stimuli to 50%. This reduced the suppression of excitatory neurons by iso stimuli (suppression with iso stimuli; 100% contrast:  $0.85 \pm 0.02$ ; 50% contrast:  $0.58 \pm 0.08$ ; mean  $\pm$  SEM; paired two-sided sign-rank;  $p < 10^{-10}$ ; 641 neurons in 6 mice), consistent with previous observations (Kapadia et al., 1999), and allowed us to better compare the impact of silencing VIP neurons on the response to iso and cross stimuli. We optogenetically suppressed VIP neurons while recording their activity and the activity of putative excitatory neurons (Figure 6A). To determine the efficiency of optogenetic silencing of VIP neurons, we recorded their responses to center, iso and cross stimuli with and without photo-activation of an inhibitory opsin (see STAR Methods). Photo-activation reduced both baseline activity as well as stimulus evoked responses of VIP neurons (Figure S6A–D). Furthermore, in agreement with a previous study (Attinger et al., 2017), silencing VIP neurons had a suppressive effect on the baseline activity of putative excitatory neurons, confirming the disinhibitory impact of VIP neurons (Figure S6E, H). Consistent with the predictions of our model, silencing VIP neurons reduced the responses of putative excitatory

neurons to cross stimuli significantly more than those to iso stimuli (Figure 6B, C). Importantly, as in our model, silencing VIP neurons also reduced the CMI of excitatory neurons, indicating that VIP neurons regulate V1 activity in a context dependent manner (Figures 6D, E; similar results obtained with different opsins, S7A-D; also true for 100% contrast, Figure S6E-G; similar results obtained under different behavioral states, Figure S8A-F). During this manipulation, excitatory neurons with both a positive and negative CMI shifted their CMI towards zero (Figure 6E), implying that their responses were less dependent on the specific features of the surround.

To determine whether the perturbation of the activity of VIP neurons affects the activity of SOM neurons, we repeated our silencing protocol, however, this time, while recording from SOM neurons (Figure 6F). Upon VIP-neurons silencing, SOM neurons were significantly less suppressed by cross stimuli than by iso stimuli (Figure 6G, H). Moreover, SOM neurons with a negative CMI, which dominated the overall sample of SOM neurons (Figure 2C), shifted their CMI towards zero, while the ones with positive CMIs did not change on average (Figures 6I, J; similar results obtained with different opsins, S7E-H; similar results obtained under different behavioral states, Figure S8G-I). Thus, the preferential suppression of SOM neurons by cross stimuli relies, at least in part, on the preferential activation of VIP neurons by these stimuli.

Taken together, based on optogenetic perturbations and computational modelling, these results demonstrate that the VIP-SOM disinhibitory circuit plays a key role in the contextual modulation of excitatory neurons.

## DISCUSSION

This study provides a mechanism for contextual modulation in V1 and reveals a disinhibitory circuit as a key mediator. Using imaging, optogenetic manipulations, and computational modeling, we find that the relationship between VIP and SOM inhibitory neurons contributes to the response profiles of L2/3 excitatory neurons in V1. When a uniform full-field stimulus is presented, VIP neurons are silent, while SOM neurons dominate the network and inhibit excitatory neurons. With a discontinuity in orientation between center and surround, VIP neurons are excited, inhibiting SOM neurons and effectively relieving excitatory neurons from the inhibition of SOM neurons. In addition, modelling suggests that the disinhibitory circuit is both necessary and sufficient for contextual modulation but exerts its function through its role in complex circuit dynamics rather than through simple feedforward disinhibition.

### Local circuits

The connectivity motifs between inhibitory neurons has been previously described (Jiang et al., 2015; Karnani et al., 2016; Pfeiffer et al., 2013; Pi et al., 2013): SOM neurons inhibit all other classes of neurons in L2/3 while VIP neurons preferentially inhibit SOM neurons. In addition, SOM neurons receive excitatory input from L2/3 neurons distributed over a relatively large retinotopic space (Adesnik et al., 2012). We therefore cannot exclude a contribution of these excitatory inputs to the contextual modulation of SOM neurons. Regardless, our results indicate that when SOM neurons prevail over VIP neurons,

excitatory neurons are suppressed, i.e. surround suppressed. Conversely, when VIP neurons prevail over SOM neurons, excitatory neurons are relieved from suppression.

In the present study the activity of SOM neurons is, on average, similar for center and iso stimuli. This may seem to contradict the assertion that SOM neurons mediate surround suppression of excitatory neurons by preferentially responding to iso stimuli. However, the similar response magnitude of SOM neurons to center and iso stimuli likely reflects the fact that the center stimuli presented here were already large enough to trigger, on average, the maximal response in SOM neurons. Moreover, while center stimuli trigger responses in all inhibitory neurons, i.e. VIP, PV and SOM neurons, iso stimuli trigger response in SOM neurons only. Thus, in response to iso stimuli, the principal source of inhibition onto excitatory neurons are SOM neurons, explaining their preferential role in surround suppression. Finally, surround suppression, though triggered by a transient increase in inhibition received by suppressed neurons, involves a steady-state decrease in the inhibition they receive. However, this decrease in inhibition is outweighed by a decrease in the excitation the neurons receive (Adesnik, 2017; Ozeki et al., 2009). This reflects the “paradoxical” response of inhibitory neurons in an inhibition-stabilized network (Ozeki et al., 2009; Tsodyks et al., 1997). As a result, SOM neurons can also show reduced firing at the fixed point of the network, when excitatory neurons are suppressed. Thus, the similar response magnitude of SOM neurons to center and iso stimuli is not inconsistent with the role of SOM neurons in surround suppression.

The fact that our networks are ISNs likely underlies some of their intriguing behaviors. When an input is given only to VIP units, all the excitatory units increase their fixed-point firing rates. In this case, because the network is an ISN, a weighted sum of the inhibition received by each of the excitatory units will show biphasic behavior, initially decreasing due to inhibition of SOM units, but ultimately being increased in the new fixed point due to recurrent excitation of SOM and PV units (Litwin-Kumar et al., 2016; Rubin et al., 2015; see Supplemental Discussion). The biphasic behavior seen in the SOM unit of the centered and preferred-orientation subnetwork after VIP-unit activation (Figures 4I–L, S3) likely reflects its role in this larger ISN dynamics. In the transition from iso- to cross-response levels, this unit shows similar biphasic behavior in many solutions. While this case is more complex (see supplemental discussion), this could reflect similar dynamics.

The biphasic response of SOM units upon VIP-unit stimulation is a robust prediction of our model, true of 114 out of the 115 most accurate solutions (Figures 4I–L, S3). The model also suggests that SOM units may show a similar biphasic behavior in the transition from iso- to cross-response levels. Judging from previous observations of inhibitory ISN transients (Ozeki et al., 2009; Sanzeni et al., 2020), testing these predictions could require time resolution of 5–10 ms, and so is outside the range of the calcium imaging used here.

In summary, this four-unit-type circuit is dynamically complex, involving many interacting and competing positive and negative feedback loops. In particular, the disinhibitory circuit can serve as a positive feedback pathway by which excitatory firing can be amplified, e.g. via the loop excitatory-VIP-SOM-excitatory, as well as a negative feedback pathway, e.g. via

the loop excitatory-SOM-excitatory. All these loops must simultaneously reach equilibrium, making for complex dynamics.

### Feedback drive

In addition to feedforward and local recurrent inputs, feedback inputs may also contribute to contextual modulation. Excitatory L2/3 neurons are most active during the presentation of cross compared to other stimuli. However, their main feedforward inputs, L4 neurons, are not as active in response to cross stimuli as they are in response to center-alone stimuli (Figure 1B, C). Is the strong response of L2/3 excitatory neurons to cross stimuli only the result of a decreased inhibition by SOM neurons, or also supported by an additional source of excitation? We have recently shown that excitatory neurons in L2/3, but not L4, are strongly driven by feedback projections from higher visual areas when visual stimuli are placed in the surround of their classical feedforward receptive field (Keller et al., 2020). This drive generates the feedback receptive field of L2/3 neurons and might contribute to the excitation of L2/3 neurons to stimuli that also cover the surround. Moreover, this feedback drive might account for the fact that some L2/3 neurons respond more strongly to cross than to center stimuli.

VIP but not SOM neurons also have a feedback receptive field (Keller et al., 2020). Since most VIP neurons are strongly suppressed by iso stimuli, in fact more strongly than excitatory neurons (Figure 2G, H), their feedback receptive field may selectively contribute to the responses of VIP neurons to cross stimuli. This feedback drive may help VIP neurons to dominate the VIP-SOM circuit for cross stimuli, a possibility that will be addressed in the future.

### Behavioral states and population heterogeneity

VIP neurons have been shown to receive modulatory inputs mediating locomotion signals (Arroyo et al., 2014; Fu et al., 2014). Furthermore, other inhibitory subtypes are also modulated by locomotion (Dipoppa et al., 2018; Pakan et al., 2016). By acting on all elements of the local circuit in V1, locomotion provides a general gain control on neuronal responses (Heintz et al., 2020). Consistently, in our experiments, locomotion had only a minor impact on the distribution of contextual modulation indexes (Figure S1C, D). Moreover, arousal modulates the VIP-SOM disinhibitory circuit independently of locomotion (Vinck et al., 2015) and can be estimated based on the pupil size, in resting mice. Interestingly, contextual modulation of excitatory but not of inhibitory neurons changed with pupil size (Figure S1E) suggesting that the saliency of the cross stimulus can be further increased by the arousal of the mouse.

While both locomotion and arousal had minor effects on the contextual modulation, they did not explain the heterogeneity of the contextual modulation observed within all individual neuronal subtypes, the distribution remained wide even when controlling for behavioral state. This heterogeneity of contextual modulation has been reported in other animal models, such as the macaque visual system (Henry et al., 2020) and may reflect the functional diversity of neurons in V1 and their roles in computing contextual modulation.

## Contextual modulation and predictive processing

Contextual modulation represents a fundamental computation to extract meaning from visual scenes. It could support many perceptual phenomena, such as pop-out effects, figure-ground segregation, detection of borders, and object detection (Angelucci et al., 2017; Bergen and Julesz, 1983; Henry et al., 2020; Jones et al., 2001; Kapadia et al., 2000; Knierim and van Essen, 1992; Lamme, 1995; Rossi et al., 2001; Schnabel et al., 2018; Seriès et al., 2003; Treisman and Gelade, 1980). Furthermore, the dichotomy between surround suppression and cross-orientation facilitation is consistent with a predictive processing framework (Bastos et al., 2012; Keller and Mrsic-Flogel, 2018), that is, a framework in which the features of a stimulus at a given location can be used to estimate the features of a stimulus at an adjacent location (Rao and Ballard, 1999). Based on natural statistics of the visual environment, the spatial features in a small patch of visual world are likely to be similar to the spatial features in the adjacent patches. If the stimuli in the surround provide a correct estimate of the stimulus in the center, the response of the neuron can be suppressed, i.e. surround suppression, as there is less need to transmit a signal that is accurately predicted. On the other hand, when the center and the surround differ, the stimuli in the surround provide an incorrect estimate of the stimulus in the center and the signal of the neuron will not be suppressed but passed along and even enhanced, i.e. cross-orientation facilitation. Predictive processing provides one compelling framework for contextual modulation of visual responses in cortical circuits, and the VIP-SOM disinhibitory circuit may provide a key route by which predictive signals alter local processing.

In conclusion, the VIP-SOM disinhibitory circuit is a canonical circuit that plays a key role in contextual modulation in primary visual cortex, and may perform similar functions across multiple cortical areas.

## STAR METHODS

### Resource Availability

**Lead Contact**—Requests for further information and resources should be directed to the lead contact, Massimo Scanziani (massimo@ucsf.edu).

**Materials Availability**—No new unique reagents or mouse lines were generated in this study.

**Data and Code Availability**—All data and analysis code related to the experimental dataset are available from the corresponding authors A.J.K. (andreasjakob.keller@ucsf.edu), M.M.R. (morgane.roth@ucsf.edu) and M.S. (massimo@ucsf.edu) on request.

All data and analysis code related to the modelling part are available from the corresponding authors M.D. (md3681@columbia.edu) and K.D.M. (kdm2103@columbia.edu) on request.

### Experimental Model and Subject Details

All experimental procedures were conducted in accordance with the regulation of the Institutional Animal Care and Use Committee of the University of California, San Francisco.

The mice were housed on a reverse light cycle (light/dark cycle: 12/12 hrs). At the start of the experiments, all mice were older than 2 months. Mice were of either sex and of the following genotype:

Gad2-IRES-cre (*GAD2<sup>tm2(cre)Zjh</sup>*; RRID: IMSR\_JAX:010802) × Ai14 (*Gt(ROSA)26Sor<sup>tm14(CAG-tdTomato)Hze</sup>*; RRID: IMSR\_JAX:007914) for imaging of layer 2/3 (L2/3) excitatory neurons (9 mice; Figures 1, 3, S1, S2); *Scnn1a-Tg3-cre (Tg(Scnn1a-cre)3Aibs/J*; RRID: IMSR\_JAX:009613) and *Scnn1a-Tg3-cre (Tg(Scnn1a-cre)3Aibs/J*; RRID: IMSR\_JAX:009613) × Ai148 (*Igs7<sup>tm148.1(tetO-GCaMP6f,CAG-tTA2)Hze</sup>*; RRID: IMSR\_JAX:030328) for imaging layer 4 (L4) excitatory neurons (4 mice and 1 mouse, respectively; Figures 1, 3, S2); *Sst-IRES-cre (Sst<sup>tm2.1(cre)Zjh</sup>*; RRID: IMSR\_JAX:028864) × Ai14 (*Gt(ROSA)26Sor<sup>tm14(CAG-tdTomato)Hze</sup>*; RRID: IMSR\_JAX:007914) for imaging of L2/3 somatostatin-expressing neurons (SOM; 13 mice; Figures 2, 3, S1, S2); *PV-cre (Pvalb<sup>tm1(cre)Arbr</sup>*; RRID: IMSR\_JAX:017320) × Ai14 (*Gt(ROSA)26Sortm14(CAG-tdTomato)Hze*; RRID: IMSR\_JAX:007914) for imaging of L2/3 parvalbumin-expressing inhibitory neurons (PV; 10 mice; Figures 2, 3, S1, S2); *VIP-IRES-cre (Vip<sup>tm1(cre)Zjh</sup>*; RRID: IMSR\_JAX:010908) × Ai14 (*Gt(ROSA)26Sor<sup>tm14(CAG-tdTomato)Hze</sup>*; RRID: IMSR\_JAX:007914) for imaging of L2/3 vasoactive-intestinal-peptide-expressing inhibitory neurons (VIP; 7 mice; Figures 2, 3, S1, S2); *VIP-IRES-cre (Vip<sup>tm1(cre)Zjh</sup>*; RRID: IMSR\_JAX:010908) for optogenetic manipulation of VIP neurons and imaging putative excitatory and VIP neurons (8 mice; Figures 6, S6–S8); and *VIP-IRES-cre (Vip<sup>tm1(cre)Zjh</sup>*; RRID: IMSR\_JAX:010908) × *Sst-IRES-Flp (Sst<sup>tm3.1(flpo)Zjh</sup>*; RRID: IMSR\_JAX:028579) for optogenetic manipulation of VIP neurons and imaging SOM neurons (8 mice; Figures 6, S7, S8).

## Method Details

**Surgery.**—Mice were anesthetized with 2% isoflurane or with a mixture of Fentanyl (West-Ward Pharmaceuticals, 0.05 mg/kg), Midazolam (Hospira, 5.0 mg/kg) and Dexmedetomidine (Zoetis, 0.5 mg/kg), injected subcutaneously. Atropine (Abcam, 0.1 mg/kg) and dexamethasone (Sigma-Aldrich, 2.0 mg/kg) were injected subcutaneously to reduce inflammation and secretion. Mice's body temperature was monitored and kept constant. To prevent the eyes from drying, a layer of lubricant ointment (Rugby) was applied. The skin above the skull was disinfected with povidone iodine (Purdue). A craniotomy was made over the right visual cortex (3 to 4.5 mm in diameter) and viruses were injected with a micropump (UMP-3, World Precision Instruments) at a rate of 2 nl/s. The craniotomy was then sealed with a glass coverslip using cyanoacrylate glue and a headplate was attached with dental cement (Lang Dental or Parkell). To reverse the anesthesia induced by the Fentanyl-Midazolam-Dexmedetomidine mixture, a mixture of Naloxone (Hospira, 1.2 mg/kg), Flumazenil (West-Ward Pharmaceuticals, 0.5 mg/kg), and Atipamezol (Zoetis, 2.5 mg/kg) was injected subcutaneously after the surgical procedures.

**Viruses.**—Viruses were typically diluted to use titers of approximately  $5 \times 10^{12}$  genome copies/ml and 50 nl were injected at each injection site (3 to 5 sites per mouse) and each depth (2 from 350 to 200  $\mu$ m below the pial surface). We injected the following viruses:



AAV2/1.ef1a.GCaMP6f.WPRE (FMI Vector Core Facility), AAV2/1.ef1a.DIO.GCaMP6f.WPRE (FMI Vector Core Facility), AAV2/1.CAG.CCaMP6f (Janelia Vector Core), AAV2/9.syn.GCaMP7f (Addgene), AAV2/1.ef1a.fDIO.GCaMP6s (Janelia Vector Core), AAV2/5.CBA.Flex.ArchT-tdTomato.WPRE.SV40 (University of Pennsylvania Vector Core), AAV2/1.CAG.Flex.rc[Jaws-KGC-GFP-ER2] (Janelia Vector Core), AAV2/9.CAG.Dio.eNpHre3.0.mRuby3.WPRE.SV40 (H. Adesnik), and AAV2/9.ef1a.F-Flex.tdTomato (Xue et al., 2014).

**Visual stimulation.**—Visual stimuli were generated using the open-source Psychophysics Toolbox based on Matlab (MathWorks). Stimuli were presented at 15 cm to the left eye on a gamma-corrected LED-backlit LCD monitor (DELL) with a mean luminance of 20 cd/m<sup>2</sup>. For experiments using a resonant scanner, the power source of the monitor's LED backlight was synchronized to the resonant scanner turnaround points (when data were not acquired) to minimize light leak from the monitor (Leinweber et al., 2014). We presented drifting sinusoidal gratings (2 Hz, 0.04 cycles/°, 100% contrast) unless stated otherwise. The trial structure of all stimulus sessions (receptive field mapping, orientation tuning, *et cetera*) was block randomized (the block size was given by the total number of parameter combinations).

**Receptive field mapping:** Stimuli consisted of a circular patch of drifting sinusoidal gratings on a gray background (typically set to 20° in diameter) presented with 15° spacing on a 5 by 5 grid covering 80 by 80° of visual space. Stimuli were presented for 1 s at a single direction or for 2 s at the 4 cardinal directions (0.5 s each). Stimulation periods were interleaved by 2 s of gray screen. We recorded 5 to 10 trials per stimulus condition.

**Orientation tuning:** We presented drifting sinusoidal gratings of at least 10° diameter drifting in 8 directions (5 to 10 trials). Stimulus time was 2 s interleaved with 4 s of gray screen.

**Size tuning:** Patches of drifting sinusoidal gratings were displayed at up to 9 different sizes, linearly spaced from 5° up to 85° in diameter (10 trials per size) centered on the classical feedforward receptive field (ffRF). Stimulation time was 2 s interleaved by 4 s of gray screen. Stimuli were either presented at a single direction or at the 4 cardinal directions (0.5 s each).

**Contextual modulation:** We presented patches of drifting sinusoidal gratings on a gray background (center stimulus), full-field drifting sinusoidal gratings (iso stimulus), and patches of drifting sinusoidal gratings on cross-oriented full-field drifting sinusoidal gratings (cross stimulus). Stimulation time was 2 s interleaved by 4 s of gray screen. The diameter of the central patches (for center and cross stimuli) was set to 20° (the average preferred diameter of excitatory neurons) for all non-optogenetic experiments in order to ease the comparison across neuron types. For optogenetic experiments, we optimized the diameter of the patches (ranging from 10 to 30°) for the population of excitatory neurons that we were recording from to facilitate the interpretation of the manipulation. Stimuli were presented at a single direction except for the cross stimulus where the surround stimulus was presented at two orthogonal directions. Responses to both directions were then averaged during the analysis. Trials with optogenetic stimulation had an additional 1 s pre-stimulus and post-



stimulus gray screen during which the optogenetic light source was turned on and the total number of trials was doubled (**Optogenetics** below).

**Behavioral monitoring.**—All mice were habituated (3 to 5 days) to the experimental setup before starting experiments. Mice were head-attached and free to run or rest on an air-supported polystyrene ball. The motion of the ball was limited to the forward-backward axis by pinning the ball. The running speed of the animals was tracked with an optical mouse (G500, Logitech). During all experiments, we recorded the positions of the left eye using a CMOS camera (DMK23UM021, Imaging Source) with a 50-mm lens (M5018-MP, Moritex), and monitored its general behavior using a webcam (LifeCam Cinema 720p HD, Microsoft).

**Two-photon calcium imaging.**—Imaging was performed using either a galvanometric-scanner based MOM (Sutter) or a resonant-scanner based (8 kHz) Bergamo II two-photon microscope (Thorlabs), both controlled by ScanImage (Vidrio). Using the MOM system, we acquired images of  $128 \times 128$  pixels at a single depth at 5.92 Hz frame rate. With the Bergamo II, we acquired images of  $380 \times 512$  pixels at 1 or 4 depths at 40 Hz or 8 Hz frame rate, respectively. We obtained similar results with both systems, so all data were pooled. The illumination light source was a Ti:sapphire laser (Chameleon Ultra II, Coherent) used at an excitation wavelength of 910 nm. The laser power under the objective (16 $\times$ , Nikon) was typically set to 30 mW and never exceeded 50 mW (laser pulse width 140 fs at a repetition rate of 80 MHz).

**Optogenetics.**—To silence VIP neurons, we used a 594-nm laser (OBIS 594 LS 100 mW, Coherent). We modified the Bergamo II microscope (Thorlabs) to combine optogenetic manipulation with two-photon calcium imaging. A lens (LA1805-B, Thorlabs) was placed in the optogenetic stimulation light path to defocus the light at the imaging plane. We used a dichroic mirror (DMBP740B, Thorlabs) to combine two-photon laser and optogenetic stimulation light, and a second dichroic mirror (FF555-Di03–25 $\times$ 36, Semrock) to split the green fluorescent protein (GFP) emission from both the two-photon and optogenetic light sources. The laser for optogenetic stimulation was synchronized to the resonant scanner turnaround points (when data were not acquired) to minimize light leak from the monitor (Attinger et al., 2017) and therefore flickered at twice the frequency of the resonant scanner. Trials with optogenetic stimulation, the laser was turned on 1 s before the visual stimulus and turned off 1 s after the offset of the visual stimulus. For all opsins, the average 594-nm laser power under the objective was typically set to a constant value of 15 mW (not exceeding 18 mW) throughout the trial.

**Computational model.**—We developed a model reproducing the responses of the 5 different neuronal types that we recorded from (L2/3 excitatory, VIP, SOM, and PV inhibitory neurons and L4 excitatory neurons). Each unit of the circuit represented the average activity of a given neuron type, integrated in a ‘subnetwork’ with the other unit types. Four such subnetworks were each assigned to one of two spatial locations (each considered the ‘surround’ of the other) and one of two preferred orientations (orthogonal to each other). We consequently obtained a total of 20 units, 5 unit-types in 4 subnetworks. We

optimized the synaptic strengths between these model units to match their responses to those observed experimentally. We obtained many solutions by using many sets of pseudo data, obtained by perturbing the experimental data by random noise with standard deviation proportional to the measurement standard error.

**Experimental data:** To model the average activity of our 20 units split across the 4 subnetworks, we divided our experimental data set into 4 subgroups: ‘Centered and preferred-orientation’, ‘centered and orthogonal-orientation’, ‘surround and preferred-orientation’, and ‘surround and orthogonal-orientation’ (for details, see Data analysis). The population-averaged responses of the 5 neuron types within each subgroup were the targets for the fitting of the corresponding 5 units within the 4 subnetworks of our model. ‘Centered’ neurons were those with fRFs aligned with the location of the center stimulus, i.e. fRFs centers no more than  $10^\circ$  from the stimulus center. ‘Surround’ neurons were those with fRF offset from the location of the center stimulus, i.e. fRF centers were at least  $15^\circ$  from the stimulus center. ‘Preferred orientation’ neurons were those with preferred orientation within  $45^\circ$  of the center stimulus orientation. ‘Orthogonal orientation’ neurons were those with preferred orientation more than  $45^\circ$  from the center stimulus orientation. Normalized population-averaged responses of the neurons of each of the 5 types within each of the 4 subgroups were obtained for each of 4 stimulus conditions (spontaneous activity or presentation of the center, iso, or cross stimuli). Hence, the goal of the model was to fit model responses to the matrix of these experimentally observed mean responses,  $\tilde{X}$  whose elements were  $\tilde{x}_{is}$ , where  $i$  corresponded to one of the 20 units and corresponded to one of the 4 stimulus conditions (Figure S2A).

**Model parameters:** Connections between L4 excitatory units and the other model units were unidirectional, as L4 was considered as an input to the subnetworks. The L4 unit of a given subnetwork was restricted to target only excitatory and PV units of the same spatial location, but of either preferred orientation (Adesnik et al., 2012; Karnani et al., 2016). Within a subnetwork, there were 16 possible recurrent connections between L2/3 excitatory and inhibitory units. We found connection strengths that best fit the data but constrained them by the most robust features in the literature (Billeh et al., 2020; Karnani et al., 2016; Pfeffer et al., 2013; Seeman et al., 2018). In particular, we disallowed 5 connections that were negligible:  $VIP \rightarrow$  excitatory,  $VIP \rightarrow$  PV,  $VIP \rightarrow$  VIP,  $SOM \rightarrow$  SOM, and  $PV \rightarrow$  SOM. We also constrained the  $PV \rightarrow$  VIP connection to be weak relative to the other non-zero connections, as described below. Thus, each subnetwork received 11 recurrent connections and two L4 connections from within its own subnetwork, a total of 13 connections per subnetwork ( $W^{(1)}$  in Figure 3D, S2B). The same set of connections was also allowed from the opposite orientation at the same location ( $W^{(2)}$  in Figure 3D, S2B), making 26 connections to a given subnetwork from its own spatial location. Projections across spatial locations were only allowed from L2/3 excitatory units to all four L2/3 unit types, adding 8 additional connections received by each subnetwork ( $W^{(3)}$  and  $W^{(4)}$  in Figure 3D, S2B; the connections from inhibitory and L4 neurons were all set to zero and therefore not displayed in Figure 3D, S2B). In total, we thus allowed 34 non-zero connections per subnetwork.

The overall  $16 \times 20$  weight matrix was composed of the  $4 \times 5$  submatrices  $W^{(j)}$  in the following convolutional structure

$$W = \begin{pmatrix} W^{(1)} & W^{(2)} & W^{(3)} & W^{(4)} \\ W^{(2)} & W^{(1)} & W^{(4)} & W^{(3)} \\ W^{(3)} & W^{(4)} & W^{(1)} & W^{(2)} \\ W^{(4)} & W^{(3)} & W^{(2)} & W^{(1)} \end{pmatrix}$$

This structure meant that each spatial position can be considered as the surround of the other and each orientation as the orthogonal of the other. This symmetry across domains allowed us to keep the total number of parameters at 34.

The above matrix  $W$  was defined in a basis in which the 20 rates were arranged as (L2/3 excitatory, PV, SOM, VIP, L4 excitatory) of subnetwork 1, then subnetwork 2, then 3, then 4. We rearranged these weights and rates, letting  $A$  be the  $16 \times 16$  matrix of recurrent weights between the sixteen L2/3 units, found from  $W$  by keeping only the first 4 columns of each of the  $W^{(j)}$ ; and  $B$  be the  $16 \times 4$  matrix of projections from L4 units to the sixteen L2/3 units, found from  $W$  by keeping only the last column of each of the  $W^{(j)}$ . Then, in this rearranged basis,  $W$  became the concatenation  $W = (A, B)$ , and acted on a rate vector  $\begin{pmatrix} \mathbf{r} \\ \mathbf{h} \end{pmatrix}$  whose first 16 elements  $\mathbf{r}$  were the rates of the L2/3 units and whose last 4 elements  $\mathbf{h}$  were the rates of the L4 units (we use bold font to indicate vectors, and capital letters to indicate matrices).

**Rate equations:** The rate equations for the units in the network for a particular stimulus  $s$  were

$$\tau \frac{d\mathbf{r}_s}{dt} = -\mathbf{r}_s + (A\mathbf{r}_s + B\mathbf{h}_s)^n \quad (1)$$

where the element-wise operation  $(x_i)^{\cdot n} \equiv ([x_i]_+)^n$  corresponded to the input-output function, a rectified power law with exponent  $n = 2$  (Ahmadian et al., 2013). The 16-vector  $\mathbf{r}_s$  specified the activities of the L2/3 units to stimulus  $s$ , while the 4-vector  $\mathbf{h}_s$  specified the activities of the L4 units to that stimulus. The time constant was set to  $\tau = 10$  ms. We denoted the combination of L2/3 and L4 units by  $\mathbf{x}_s = \begin{pmatrix} \mathbf{r}_s \\ \mathbf{h}_s \end{pmatrix}$ . We used  $X$ ,  $R$ , and  $H$  to refer to the matrices whose columns are the vectors  $\mathbf{x}_s$ ,  $\mathbf{r}_s$ , or  $\mathbf{h}_s$ , respectively, across all stimuli  $s$ .

**Cost function:** For each stimulus  $s$ , we denoted the experimentally measured mean responses as  $\tilde{\mathbf{h}}_s$  for L4 and  $\tilde{\mathbf{r}}_s$  for L2/3. Our model found inputs (L4 responses)  $\hat{\mathbf{h}}_s$ , and synaptic weights that produced a fixed-point response denoted by  $\lim_{t \rightarrow \infty} \mathbf{r}_s(t) \equiv \hat{\mathbf{r}}_s$ . The cost function of the model demanded that the inputs and responses should have minimal summed-weighted-squared error relative to the experimental measurements, subject to certain regularization terms:

$$E_0(\hat{X}, W) = \sum_{i,s} M_{is}^2 (\hat{x}_{is} - \tilde{x}_{is})^2 + L(\hat{X}, W)$$

Here,  $m_{is}$  was a weight matrix that represented our uncertainty over the responses. More specifically  $m_{is} = \sigma_0 \beta_j / \sigma_{is}$ , where  $\sigma_{is}$  was the standard error of the responses  $\tilde{x}_{is}$  measured experimentally,  $\beta$  was a multiplicative factor to weight errors in certain unit types more than others, and  $\sigma_0 = \langle \sigma_{is} / \beta_i \rangle_{is}$  was a normalization factor, where  $\langle z_{is} \rangle_{is}$  indicated an average of  $z_{is}$  over  $i$  and  $s$ . We chose  $\beta_j = 1$  for L2/3 excitatory, PV and VIP neurons,  $\beta_j = 1.5$  for SOM and  $\beta_j = 2$  or  $2.5$  (the two parameters gave similar results) for L4 excitatory neurons. We used larger  $\beta_j$  for units that we found harder to fit, L4 and SOM units. Intuitively, this fitting difficulty might arise from the fact that L4 and SOM neurons had the most distinct response patterns compared to other neuron types.  $L$  represented the sum of all regularization terms, defined as:

$$L(\hat{X}, W) = \alpha_1 \sum_{i,S} [\epsilon_x - \hat{x}_{is}]_+ + \alpha_2 \sum_{\{i,j\} \in N} m_{ij}^{(1)} (-\log|w_{ij}|) + \alpha_3 \sum_{\{i,j\} \in N} m_{ij}^{(2)} w_{ij}^2$$

The first regularization factor, using  $\alpha_1 = 0.02$ , nudged the responses  $\hat{x}_{is}$  above a minimal threshold  $\epsilon_x = 0.01$ , since  $\hat{X}$  corresponded to estimated firing rates and were thus non-negative. The second and third factors were applied only to the 34 weights that were allowed to be non-zero, as specified above; this set of weights was designated by  $N$ . The second factor, using  $\alpha_2 = 0.05$ , nudged weights with a corresponding positive value of  $m_{ij}^{(1)}$  to prevent them from being too close to zero. The elements  $m_{ij}^{(1)} \in \{0, 1\}$  were non-zero for almost all allowed connections between units sharing orientation preference, i.e.  $W^{(1)}$  and  $W^{(3)}$ : all of those from the same spatial location of the ffRF ( $W^{(1)}$ ) with the exception of the connection from PV to VIP units (its  $m_{ij}^{(1)}$  was zero because this connection is weak according to Pfeffer et al., 2013), and those from the surround to excitatory units and to SOM units (two of the 4 non-zero elements of  $W^{(3)}$ ; these two were chosen because there is well-established evidence of substantial integration across spatial locations for excitatory and SOM neurons (Adesnik et al., 2012). The third factor, with  $\alpha_3 = 0.01$ , nudged weights with a corresponding positive value of  $m_{ij}^{(2)}$  towards zero. The elements  $m_{ij}^{(2)} \in \{0, 1\}$  were non-zero for all allowed connections between units having different preferred orientations ( $W^{(2)}$  and  $W^{(4)}$ ) and for the PV- to VIP-unit connection regardless of the orientation preference. In other words, we wanted to discourage strong connections across subnetworks with different orientation preferences and encourage strong connections among those allowed in  $W^{(1)}$  and  $W^{(3)}$ .

**Inferring connection weights from the data:** We noted that equation (1) corresponded to a recurrent neural network (RNN). This allowed us to train the RNN to find the best solutions, i.e. the weights  $w_{ij}$  and the inputs  $h_{is}$ , using backpropagation through time (BPTT; Pascanu et al., 2013; Rumelhart et al., 1986). The training of a neural network is highly sensitive to its initial weights (He et al., 2015) and in general we observed that starting from random

initial conditions would often lead to unstable solutions. This might stem from the fact that RNN training is prone to gradient vanishing and gradient explosion (Bengio et al., 1994), especially for a large number of time steps. As a first step, we therefore found stable solutions which would approximately match the data using non-negative least square (NNLS) regression, which we used as initial conditions of the BPTT. In a previous study (Dipoppa et al., 2018), we used a NNLS to infer the optimal synaptic strengths of a model evolving a dynamical equation similar to equation (1) such that the model would match the experimental data. Here we similarly inferred optimal strengths for matching the model to pseudo data  $x'_{ij} = \gamma_{ij}\tilde{x}_{ij}$ , randomly generated using the random matrix  $\gamma_{ij} \sim \text{Gamma}(\varphi^{-1}\sigma_{ij}^{-2}, \varphi\sigma_{ij}^2)$  with  $\varphi = 5$ , such that  $\langle x'_{ij} \rangle = \tilde{x}_{ij}$  and  $\text{Var}(x'_{ij}) = \varphi\sigma_{ij}^2$ . We solved the convex problem of minimizing the following cost function,  $E_1$ , which made  $X'$  as close as possible to a fixed point of equation 1 subject to regularization, as an approximation of minimizing  $E_0(X', W)$ :

$$E_1(X', W) = \sum_{i,s} M_{is}^2 \left[ (R'_{is})^{1/n} - \sum_{i,k} W_{ik} X'_{kj} \right]^2 + L(WX', W)$$

Here  $R'$  was the  $R$  component of  $X' = \begin{pmatrix} R' \\ H' \end{pmatrix}$ . We generated  $N_{NNLS} = 10,000,000$  different sets of pseudo data  $\{X'\}$  (5,000,000 with  $\beta_j = 2$  and 5,000,000 with  $\beta_j = 2.5$ ). We then used the trust region reflective algorithm to solve the problem  $\min_W E_1(X', W)$  starting from initial conditions  $W_{ij} \sim \text{Gamma}(1,1)$  and with boundaries  $0 < |W_{ij}| < 10$  with  $W_{ij}$  having positive or negative signs depending on whether they represented excitatory or inhibitory connections. After obtaining a set of optimal parameters  $\{W^{NNLS}\}$  for each set of pseudo data, we let the network evolve following equation (1) and obtained the fixed points (if they existed), discarding all solutions that had at least one of the 20 rates  $> 10$  or  $< \epsilon_x$ . This produced the set of fixed points  $\{\hat{X}^{NNLS}\}$ . Note that the  $\hat{H}^{NNLS}$  portion of  $\hat{X}^{NNLS}$  was unchanged from its original perturbed value  $H'(\hat{H} \equiv H')$ . We then recomputed the error  $\{E_0(\hat{X}^{NNLS}, W^{NNLS})\} \equiv \{E_0^{NNLS}\}$ .

We selected the 20,000 best solutions  $\{\hat{X}^{NNLS}, W^{NNLS}\}$  sorted by  $\{E_0^{NNLS}\}$  as starting parameters for the BPTT (5,000 with  $\beta_j = 2$  and 15,000 with  $\beta_j = 2.5$ ). We defined the following cost function for the BPTT:

$$E_2(X(t), W) = \sum_{i,s} m_{is}^2 \langle x_{is}(t) - \tilde{x}_{is} \rangle_t^2 + \sum_{i,s} \langle x_{is}(t+1) - x_{is}(t) \rangle_t^2 + L(\langle X(t) \rangle_t, W)$$

where  $x_{is}(t)$  corresponded to the dynamics of the network at each time step  $t$  and where the average over  $t$ ,  $\langle \cdot \rangle_t$ , was computed over the last  $T = 200$  time-steps of the dynamics. The second term punished large values of the derivative of  $X$  to ensure that the network reached a fixed point. Independently of the stimulus condition, for each run of the dynamics (termed an ‘epoch’), the starting point was  $x_{is}(0) = \tilde{x}_{i0} + \delta x_{is}$ , where  $s = 0$  corresponded to the spontaneous activity and  $\delta x_{is} \sim \mathcal{N}(0, 0.01)$  was a random perturbation following a Gaussian

distribution. We used time steps of  $\tau = 2$  ms. An epoch consisted of evolving equation (1) using the Euler scheme:

$$\mathbf{r}_s(t+1) = \mathbf{r}_s(t) + \left[ -\mathbf{r}_s(t) + (\mathbf{A}\mathbf{r}_s(t) + \mathbf{B}\mathbf{h}_s)^n \right] \frac{\Delta t}{\tau} \quad (2)$$

for 500 time-steps. To compute the gradient of the loss equation  $E_2(\mathbf{X}(t), \mathbf{W})$  over  $\mathbf{W}$  and the  $H$  portion of  $\mathbf{X}$  of the discretized network given by equation (2) over the last  $T = 200$  time steps of the dynamics, we used automatic differentiation methods provided by the pytorch library in Python. Optimization was carried out by the ADAM optimizer (Kingma and Ba, 2017). To improve convergence to a solution, we employed a triangular learning rate policy (Smith, 2017) at a base learning rate of  $3 \times 10^{-4}$ , a maximum learning rate of  $3 \times 10^{-3}$ , 100 training epochs for the increasing part of the cycle, 200 training epochs for the decreasing part of the cycle. We also used a patience parameter of 1000 epochs. If the error did not improve over this length of time, the training procedure of the BPTT would stop. If not interrupted, the model was trained for 10,000 epochs. After running the BPTT for the best 20,000 starting conditions of the NNLS  $\{\hat{\mathbf{X}}^{NNLS}, \mathbf{W}^{NNLS}\}$ , we obtained a new set of inferred weights and rates  $\{\hat{\mathbf{X}}^{RNN}, \mathbf{W}^{RNN}\}$ . Note that in contrast to the NNLS, the  $\hat{H}^{RNN}$  portion of  $\hat{\mathbf{X}}^{RNN}$  was learned. Of these the BPTT procedure was able to reduce the error for 13,471 simulations (3,293 with  $\beta_j = 2$  and 10,178 with  $\beta_j = 2.5$ ).

To select the top solutions, we defined a new cost function only based on the error between the model and the data points without any extra penalty  $\beta_j$  on specific unit types or regularization factors

$$E_3(\hat{\mathbf{X}}^{RNN}) = \sum_{i,s} \sigma_{is}^{-2} (\hat{x}_{is}^{RNN} - \tilde{x}_{is})^2$$

where, as a reminder,  $\sigma_{is}$  was the standard error of the responses  $\tilde{x}_{is}$  measured experimentally. For further analysis, we selected the 115 solutions (45 with  $\beta_j = 2$  and 70 with  $\beta_j = 2.5$ ) whose error was within 1.33 of the smallest error estimated by  $E_3$ . The threshold we set to select solutions roughly divides a sparse set of lowest-error solutions from a dense set of higher-error solutions (Figure S5).

**Comparison with the data:** The model was trained to reproduce the estimated firing rate of the units  $\mathbf{X}$ , shown in Figure S2A. For better comparison to the experimental data, in Figure 3C we showed the baseline subtracted model responses  $x_{is} = x_{is} - x_{i0}$ , i.e. with the spontaneous activity  $s = 0$  subtracted. Similarly, we used  $x_{is}$  to compute CMIs and the difference between control and optogenetic conditions (Figure 5, Figure S5). Using  $x_{is}$  instead of the unit activity  $x_{is}$  did not appreciably change the results.

**Equivalence classes of solutions.**—The model was fitted to produce responses  $\hat{\mathbf{X}}^{RNN}$  that closely reproduced normalized data  $\tilde{\mathbf{X}}$ . We estimated a set of correction factors  $c_i$ , one for each of the 5 unit types, which were equal to the average firing rates in Hz of each

neuron type (for details, see Data analysis, *Contextual modulation*):  $c_{Exc} = 1.87$ ,  $c_{PV} = 6.63$ ,  $c_{SOM} = 6.86$ ,  $c_{VIP} = 2.61$ ,  $C_{LA} = 1.49$ . We wanted a model in which all the normalized activities  $\hat{X}^{RNN}$  were multiplied by these  $c_i$ 's, so that the model activities would have values comparable to observed firing rates in Hz. For equation 1, there exists an equivalence classes of solutions: given a solution with weights  $W^{RNN}$  and rates  $\hat{X}^{RNN}$ , a solution  $W^{EQ}$  and  $X^{EQ}$  that that is completely equivalent in all behaviors except for scaling of the rates,

$$x_{is}^{EQ} = c_i \hat{x}_{is}^{RNN}, \text{ is obtained by scaling the weights so that } w_{ij}^{EQ} = \left( \frac{1}{c_i^n / c_j} \right) w_{ij}^{RNN}. \text{ Although}$$

the model was trained to generate solutions  $W^{RNN}$  and  $\hat{X}^{RNN}$ , we present the solutions  $W^{EQ}$  and  $X^{EQ}$ .

**Perturbation analysis.**—We generated the perturbation analysis, in which we froze a set of inputs (Figures 4 and S4), as follows: the network started from the fixed point of the iso-response level at time  $t = 0$ ;  $r_i(0) = \hat{r}_{i, iso}$  (differently from Figures 3, 5 and S2 where we started from the spontaneous state  $r_i(0) = \hat{r}_{i, spont}$ ). We then switched the L4 excitatory inputs to their cross-response level  $\mathbf{h}_{iso} \rightarrow \mathbf{h}_{cross}$  (Figure 4A–H, S4A) or kept them at their iso-response level (Figure 4I–L, S4B–K). At the same time, we froze another input or set of inputs to the fixed point of one of the response levels (iso or cross). Each input we froze was from one unit-type  $b$  to another unit-type  $a$ , and this meant that we froze all such inputs within and between all four subnetworks to their value in the given fixed point. For a particular connection, say from unit  $j$  of type  $b$  to unit  $i$  of type  $a$ , freezing the input meant replacing the normal dynamical input,  $w_{ij}x_j(t)$ , with the fixed input that would occur in the fixed point for stimulus  $s$  (iso or cross),  $w_{ij}\hat{x}_{js}$ . We then retained the solutions which converged to a fixed point within 2s and in which no unit exploded to high values ( $> 10,000$ ).

In the perturbation analysis of Figure S3, we show the change in the fixed point introduced by a very small perturbation (excitation)  $q \ll 1$  of all VIP units in the network, starting from the iso fixed-point. For a very small perturbation, the change in the fixed point can be approximated by the dynamics linearized around the initial iso fixed-point. This yields the following equation:

$$\Delta \mathbf{r}^{(s, k)} = \left( \mathbb{1} - D^{(s)} A \right)^{-1} D^{(s)} \mathbf{p}^{(k)} \Delta q$$

Here,  $\mathbf{p}^{(k)}$  is the perturbation, a 16-element vector (one element for each unit in the model) in which each of four elements corresponding to unit type  $k$  is 1, while all other elements are set to 0. The strength of the perturbation is  $q = 0.01$ . The change in fixed-point response induced by the perturbation is  $\mathbf{r}^{(s, k)}$ , a 16-element vector giving the change in response of each unit to a perturbation of unit type  $k$  starting from the state  $s$ .  $D^{(s)}$  is a  $16 \times 16$  diagonal matrix whose elements are  $d_{ii}^{(s)} = n \hat{r}_{is}^{-1} - \frac{1}{n}$ . We focused on the effect of a perturbing all VIP units,  $k = VIP$  (Figure S3).



When computing trajectories of excitatory and SOM units of subnetwork 1 during a small perturbation of VIP units (Figure S3), we added optogenetic excitation by using the following linearized dynamics with  $q = 0.01$  and initial condition  $r^{(s,k)}(t = 0) = 0$ :

$$\tau \frac{d\Delta r^{(s,k)}}{dt} = -\Delta r^{(s,k)} + D^{(s)}(A\Delta r^{(s,k)} + p^{(k)}\Delta q)$$

## Quantification and Statistical Analysis

**Data analysis.**—All data were analyzed using custom-written code in Matlab (MathWorks).

**Two-photon calcium imaging:** We analyzed two-photon calcium imaging data as described previously (Keller et al., 2020). Briefly, data were full-frame registered using custom-written software (<https://sourceforge.net/projects/iris-scanning/>). We selected the neurons semi manually, based on mean and maximum projection images. We calculated the raw fluorescence traces as the average fluorescence of all pixels within a selected region of interest for each frame. Fluorescence changes ( $\Delta F/F$ ) were calculated as described elsewhere (Dombeck et al., 2007). All stimulus evoked responses were baseline subtracted (1 s pre-stimulus interval). To estimate firing rates for the model, we inferred spikes using a previously published algorithm (Deneux et al., 2016).

**Response amplitude:** The response amplitude to a stimulus was computed as the average response over the duration of the stimulus presentation (excluding the first 0.5 s of each trial due to the delay and slow rise of calcium indicators). Responses were defined as significant if they exceeded a z-score of 3.29 (corresponding to  $p < 10^{-3}$ ) or 5.33 (corresponding to  $p < 10^{-7}$ ; for experiments in L4).

**Receptive field mapping:** To estimate the center of the receptive field, we fitted the responses to patches of gratings with a two-dimensional Gaussian. We excluded neurons if they failed to have at least one significant trial-averaged response within  $10^\circ$  of their estimated fRF centers. Additionally, except for the ‘surround group’ (see Computational model), we excluded neurons if their estimated fRF centers were not within  $10^\circ$  of the stimulus centers of the stimuli used for estimating size tuning, orientation tuning, *et cetera*. Neurons of the ‘surround group’ had estimated receptive field centers that were at least  $15^\circ$  away from the centers of the stimuli.

**Size tuning:** We fitted the integral over a difference of Gaussians. This fit was used to estimate the neurons’ preferred sizes. We approximated the fRF size by the size of the patch of gratings evoking the largest response (size tuning fits were bound to the interval 0.1 to  $90.1^\circ$ ).

**Orientation tuning:** We fitted a circular sum of Gaussians with a peak offset of  $180^\circ$  and equal tuning width (full width at half maximum of the Gaussian fit). When the preferred orientations of neurons were relevant, we excluded neurons with an  $R^2$  goodness-of-fit of 0.3 or below (Figure S1A; 145 of 727 excitatory neurons were excluded; Figure S2A; 327 of

1261 excitatory, 23 of 107 PV, 27 of 333 SOM, 40 of 116 VIP, and 71 of 169 L4 neurons were excluded). Otherwise, neurons were studied without regard for their orientation tuning properties. Note that we found no systematic relationship between the CMI and the orientation tuning width in any neuronal subtype (data not shown).

**Contextual modulation:** To estimate the contextual modulation of excitatory, VIP, SOM, and PV neurons, we used a center patch diameter of 20°. We calculated a contextual modulation index defined as the difference between the activity to cross and iso stimuli divided by the sum of the two. To estimate the effect of silencing VIP neurons on the contextual modulation of putative excitatory neurons, neurons were only considered if their preferred size was within 10° of the center-patch diameter. Note that for these experiments, the center-patch diameter was set to a size between 10° and 30°. When comparing to the model, population-averaged responses to center, iso and cross stimuli were estimated as spike rates (Figures 3C, S2A). To this end, trial-averaged responses of every neuron were first normalized by the maximum responses across center, iso, cross, and receptive field mapping stimuli. Then these responses were scaled by a constant factor, one for every neuron type (L2/3 excitatory: 1.87; PV: 6.63; SOM: 6.86; VIP: 2.61; and L4 excitatory: 1.49). To determine these constant factors, we used estimated firing rates in response to the stimulus evoking the largest response (on average) based on a previously published algorithm (Deneux et al., 2016). Due to the normalization, this approach reduced the impact of outliers. Note that directly using estimated firing rates, rather than the approach described above, produced similar results.

**Surround suppression:** Surround suppression was computed as one minus the responses to iso (or cross) divided by the responses to center stimuli. Neurons with a negative response to center were excluded from this analysis (Figure 1C; 62 of 727 L2/3 and 3 of 43 L4 neurons were excluded; Figure S1B; 62 of 727 excitatory, 1 of 87 PV, 3 of 279 SOM, and 9 of 49 VIP neurons were excluded).

**Baseline:** We estimated the baseline activity as the difference between the average fluorescence change during baseline periods (averaged over all 1 s pre-stimulus intervals) and the lower quartile of the overall trace of fluorescence changes. To compute the population-averaged baseline activity, we excluded neurons with an estimated baseline activity of more than 3 standard deviations above the median (Figure S2A; 23 of 934 L2/3 excitatory, 4 of 84 PV, 3 of 306 SOM, 2 of 76 VIP, and 2 of 98 L4 neurons were excluded).

**Running and resting trials:** Trials were classified as running if the median running speed during the stimulus presentation exceeded 1 cm/s and classified as resting otherwise.

**Large and small pupil trials:** To determine how arousal affects neuronal responses independent of locomotion, we only considered resting trials. Of these, trials were classified as large pupil-diameter trials if the median pupil-diameter during the stimulus presentation was above the median pupil-diameter over the corresponding recording session and classified as small otherwise.

**Eye movement and eye-movement-free trials:** Trials were classified as eye-movement trials if mice made at least one eye movement during the stimulus presentation and classified as eye-movement-free otherwise. To determine eye movements, we used a similar approach as in previous studies (Keller et al., 2020; Roth et al., 2016). Briefly, we first computed pupil position offline by smoothing and thresholding the images and fitting a circle to the pupil. Pupil position was filtered using a median filter. Eye movements were detected automatically by applying an adapted threshold. This method was cross validated in several experiments using manual detection of eye movements.

**Statistics.:** We used two-sided Wilcoxon rank-sum tests for independent group comparisons, and two-sided Wilcoxon signed-rank tests for paired tests and single group analysis. No statistical methods were used to pre-determine experimental sample sizes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements.

We thank M. Mukundan, B. Wong, and L. Bao for technical support, J.I. Glaser for technical advice on the model optimization procedure, and the members of the Scanziani laboratory for helpful discussions of this project. We thank H. Adesnik for the AAV2/9.CAG.Dio.eNpHre3.0.mRuby3.WPRE.SV40 virus and M. Rio for software support. This project was supported by NIH grant U19NS107613 (K.D.M., M.S., M.D., and A.I.), the Howard Hughes Medical Institute (M.S.), the Swiss National Science Foundation grants P300PA\_177882 and P2EZP3\_162284 to A.J.K and P300PA\_177898 to M.M.R, the Gatsby Charitable Foundation (K.D.M. and A.I.), and NSF NeuroNex Award DBI-1707398 (K.D.M. and M.D.). We acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010.

## REFERENCES

- Adesnik H (2017). Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. *Neuron* 95, 1147–1159.e4. [PubMed: 28858618]
- Adesnik H, Bruns W, Taniguchi H, Huang ZJ, and Scanziani M (2012). A neural circuit for spatial summation in visual cortex. *Nature* 490, 226–231. [PubMed: 23060193]
- Ahmadian Y, Rubin DB, and Miller KD (2013). Analysis of the Stabilized Supralinear Network. *Neural Comput.* 25, 1994–2037. [PubMed: 23663149]
- Alexander DM, and Van Leeuwen C (2010). Mapping of contextual modulation in the population response of primary visual cortex. *Cogn. Neurodyn.* 4, 1–24. [PubMed: 19898958]
- Alitto HJ, and Usrey WM (2008). Origin and Dynamics of Extraclassical Suppression in the Lateral Geniculate Nucleus of the Macaque Monkey. *Neuron* 57, 135–146. [PubMed: 18184570]
- Angelucci A, Bijanzadeh M, Nurminen L, Federer F, Merlin S, and Bressloff PC (2017). Circuits and Mechanisms for Surround Modulation in Visual Cortex. *Annu. Rev. Neurosci* 40, 425–451. [PubMed: 28471714]
- Arroyo S, Bennett C, and Hestrin S (2014). Nicotinic modulation of cortical circuits. *Front. Neural Circuits* 8.
- Attinger A, Wang B, and Keller GB (2017). Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. *Cell* 169, 1291–1302.e14. [PubMed: 28602353]
- Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, and Friston KJ (2012). Canonical Microcircuits for Predictive Coding. *Neuron* 76, 695–711. [PubMed: 23177956]
- Bengio Y, Simard P, and Frasconi P (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw* 5, 157–166. [PubMed: 18267787]

- Bergen JR, and Julesz B (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature* 303, 696–698. [PubMed: 6855915]
- Billeh YN, Cai B, Gratiy SL, Dai K, Iyer R, Gouwens NW, Abbasi-Asl R, Jia X, Siegle JH, Olsen SR, et al. (2020). Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex. *Neuron* 106, 388–403.e18. [PubMed: 32142648]
- Blakemore C, and Tobin ElisabethA. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp. Brain Res* 15.
- Bolz J, and Gilbert CD (1986). Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature* 320, 362–365. [PubMed: 3960119]
- Chiao C-C, and Masland RH (2003). Contextual tuning of direction-selective retinal ganglion cells. *Nat. Neurosci* 6, 1251–1252. [PubMed: 14595442]
- Coen-Cagli R, Kohn A, and Schwartz O (2015). Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* 18, 1648–1655. [PubMed: 26436902]
- Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, Ko H, Hofer SB, and Mrsic-Flogel TD (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* 518, 399–403. [PubMed: 25652823]
- Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, Tsegaye G, Tsang A, Wong A, Patel R, et al. (2019). High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* 16, 649–657. [PubMed: 31209382]
- Deneux T, Kaszas A, Szalay G, Katona G, Lakner Tamas, Grinvald A, Rozsa B, and Vanzetta I (2016). Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun* 17.
- Dipoppa M, Ranson A, Krumin M, Pachitariu M, Carandini M, and Harris KD (2018). Vision and Locomotion Shape the Interactions between Neuron Types in Mouse Visual Cortex. *Neuron* 98, 602–615.e8. [PubMed: 29656873]
- Dombeck DA, Khabbaz AN, Collman F, Adelman TL, and Tank DW (2007). Imaging Large-Scale Neural Activity with Cellular Resolution in Awake, Mobile Mice. *Neuron* 56, 43–57. [PubMed: 17920014]
- Fino E, Packer AM, and Yuste R (2013). The Logic of Inhibitory Connectivity in the Neocortex. *The Neuroscientist* 19, 228–237. [PubMed: 22922685]
- Fitzpatrick D (2000). Seeing beyond the receptive field in primary visual cortex. *Curr. Opin. Neurobiol* 10, 438–443. [PubMed: 10981611]
- Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, and Stryker MP (2014). A Cortical Circuit for Gain Control by Behavioral State. *Cell* 156, 1139–1152. [PubMed: 24630718]
- Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, and McCormick DA (2010). Synaptic and Network Mechanisms of Sparse and Reliable Visual Cortical Activity during Nonclassical Receptive Field Stimulation. *Neuron* 65, 107–121. [PubMed: 20152117]
- Han X, Chow BY, Zhou H, Klapoetke NC, Chuong A, Rajimehr R, Yang A, Baratta MV, Winkle J, Desimone R, et al. (2011). A High-Light Sensitivity Optical Neural Silencer: Development and Application to Optogenetic Control of Non-Human Primate Cortex. *Front. Syst. Neurosci* 5.
- He K, Zhang X, Ren S, and Sun J (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In 2015 IEEE International Conference on Computer Vision (ICCV), (Santiago, Chile: IEEE), pp. 1026–1034.
- Heintz TG, Hinojosa AJ, and Lagnado L (2020). Opposing forms of adaptation in mouse visual cortex are controlled by distinct inhibitory microcircuits and gated by locomotion (*Neuroscience*).
- Hennequin G (2018). The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron* 21.
- Henry CA, Jazayeri M, Shapley RM, and Hawken MJ (2020). Distinct spatiotemporal mechanisms underlie extra-classical receptive field modulation in macaque V1 microcircuits. *ELife* 9, e54264. [PubMed: 32458798]
- Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, Zeng H, Lein E, Lesica NA, and Mrsic-Flogel TD (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci* 14, 1045–1052. [PubMed: 21765421]

- Huang X, Rangel M, Briggman KL, and Wei W (2019). Neural mechanisms of contextual modulation in the retinal direction selective circuit. *Nat. Commun* 10, 2431. [PubMed: 31160566]
- Hubel DH, and Wiesel TN (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol* 160, 106–154. [PubMed: 14449617]
- Hubel DH, and Wiesel TN (1965). RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE IN TWO NONSTRIATE VISUAL AREAS (18 AND 19) OF THE CAT. *J. Neurophysiol* 28, 229–289. [PubMed: 14283058]
- Jiang X, Shen S, Cadwell CR, Berens P, Sinz F, Ecker AS, Patel S, and Tolias AS (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350, aac9462–aac9462. [PubMed: 26612957]
- Jones HE, Grieve KL, Wang W, and Sillito AM (2001). Surround Suppression in Primate V1. *J. Neurophysiol* 86, 2011–2028. [PubMed: 11600658]
- Jones HE, Andolina IM, Ahmed B, Shipp SD, Clements JTC, Grieve KL, Cudeiro J, Salt TE, and Sillito AM (2012). Differential Feedback Modulation of Center and Surround Mechanisms in Parvocellular Cells in the Visual Thalamus. *J. Neurosci* 32, 15946–15951. [PubMed: 23136432]
- Jones HE, Andolina IM, Shipp SD, Adams DL, Cudeiro J, Salt TE, and Sillito AM (2015). Figure-ground modulation in awake primate thalamus. *Proc. Natl. Acad. Sci* 112, 7085–7090. [PubMed: 25901330]
- Kapadia MK, Westheimer G, and Gilbert CD (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc. Natl. Acad. Sci* 96, 12073–12078. [PubMed: 10518578]
- Kapadia MK, Westheimer G, and Gilbert CD (2000). Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception. *J. Neurophysiol* 84, 2048–2062. [PubMed: 11024097]
- Karnani MM, Jackson J, Ayzenshtat I, Tucciarone J, Manoocheri K, Snider WG, and Yuste R (2016). Cooperative Subnetworks of Molecularly Similar Interneurons in Mouse Neocortex. *Neuron* 90, 86–100. [PubMed: 27021171]
- Keller GB, and Mrsic-Flogel TD (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron* 100, 424–435. [PubMed: 30359606]
- Keller AJ, Roth MM, and Scanziani M (2020). Feedback generates a second receptive field in neurons of the visual cortex. *Nature* 582, 545–549. [PubMed: 32499655]
- Kingma DP, and Ba J (2017). Adam: A Method for Stochastic Optimization. *ArXiv* 1412.6980v9.
- Knierim JJ, and van Essen DC (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol* 67, 961–980. [PubMed: 1588394]
- Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, and Mrsic-Flogel TD (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature* 473, 87–91. [PubMed: 21478872]
- Lamme V (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci* 15, 1605–1615. [PubMed: 7869121]
- Lee JH, Koch C, and Mihalas S (2017). A Computational Analysis of the Function of Three Inhibitory Cell Types in Contextual Visual Processing. *Front. Comput. Neurosci* 11, 28. [PubMed: 28487644]
- Lee S, Hjerling-Leffler J, Zaghera E, Fishell G, and Rudy B (2010). The Largest Group of Superficial Neocortical GABAergic Interneurons Expresses Ionotropic Serotonin Receptors. *J. Neurosci* 30, 16796–16808. [PubMed: 21159951]
- Leinweber M, Zmarz P, Buchmann P, Argast P, Hübener M, Bonhoeffer T, and Keller GB (2014). Two-photon Calcium Imaging in Mice Navigating a Virtual Reality Environment. *J. Vis. Exp* 50885.
- Levick WR, Cleland BG, and Dubin MW (1972). Lateral geniculate neurons of cat: Retinal inputs and physiology. *Invest. Ophthalmol* 11, 10.
- Litwin-Kumar A, Rosenbaum R, and Doiron B (2016). Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *J. Neurophysiol* 115, 1399–1409. [PubMed: 26740531]
- Ma W. -p., Liu B. -h., Li Y. -t., Josh Huang Z, Zhang LI, and Tao HW (2010). Visual Representations by Cortical Somatostatin Inhibitory Neurons--Selective But with Weak and Delayed Responses. *J. Neurosci* 30, 14371–14379. [PubMed: 20980594]

- McIlwain JT (1964). Receptive fields of optic tract axons and lateral geniculate cells: Peripheral extent and barbiturate sensitivity. *J. Neurophysiol* 27, 1154–1173. [PubMed: 14223976]
- Miller KD, and Palmigiano A (2020). Generalized paradoxical effects in excitatory/inhibitory networks. *BioRxiv*.
- Millman DJ, Ocker GK, Caldejon S, Kato I, Larkin JD, Lee EK, Luviano J, Nayan C, Nguyen TV, North K, et al. (2020). VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli. *ELife* 9, e55130. [PubMed: 33108272]
- Nelson JJ, and Frost BJ (1978). Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Res.* 139, 359–365. [PubMed: 624064]
- Nurminen L, Merlin S, Bijanzadeh M, Federer F, and Angelucci A (2018). Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nat. Commun* 9, 2281. [PubMed: 29892057]
- Ölveczky BP, Baccus SA, and Meister M (2003). Segregation of object and background motion in the retina. *Nature* 423, 401–408. [PubMed: 12754524]
- Ozeki H, Finn IM, Schaffer ES, Miller KD, and Ferster D (2009). Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. *Neuron* 62, 578–592. [PubMed: 19477158]
- Pakan JM, Lowe SC, Dylida E, Keemink SW, Currie SP, Coutts CA, and Rochefort NL (2016). Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *ELife* 5, e14985. [PubMed: 27552056]
- Pascanu R, Mikolov T, and Bengio Y (2013). On the difficulty of training recurrent neural networks. *ArXiv* 12115063v2 9.
- Peron S, Pancholi R, Voelcker B, Wittenbach JD, Ólafsdóttir HF, Freeman J, and Svoboda K (2020). Recurrent interactions in local cortical circuits. *Nature* 579, 256–259. [PubMed: 32132709]
- Pfeffer CK, Xue M, He M, Huang ZJ, and Scanziani M (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci* 16, 1068–1076. [PubMed: 23817549]
- Pi H-J, Hangya B, Kvitsiani D, Sanders JI, Huang ZJ, and Kepecs A (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. [PubMed: 24097352]
- Priebe NJ, and Ferster D (2008). Inhibition, Spike Threshold, and Stimulus Selectivity in Primary Visual Cortex. *Neuron* 57, 482–497. [PubMed: 18304479]
- Priebe NJ, Mechler F, Carandini M, and Ferster D (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat. Neurosci* 7, 1113–1122. [PubMed: 15338009]
- Rao RPN, and Ballard DH (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci* 2, 79–87. [PubMed: 10195184]
- Rossi AF, Desimone R, and Ungerleider LG (2001). Contextual Modulation in Primary Visual Cortex of Macaques. *J. Neurosci* 21, 1698–1709. [PubMed: 11222659]
- Roth MM, Dahmen JC, Muir DR, Imhof F, Martini FJ, and Hofer SB (2016). Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. *Nat. Neurosci* 19, 299–307. [PubMed: 26691828]
- Rubin DB, Van Hooser SD, and Miller KD (2015). The Stabilized Supralinear Network: A Unifying Circuit Motif Underlying Multi-Input Integration in Sensory Cortex. *Neuron* 85, 402–417. [PubMed: 25611511]
- Rumelhart DE, Hintont GE, and Williams RJ (1986). Learning representations by back-propagating errors. *Nature* 4.
- Sanzeni A, Akitake B, Goldbach HC, Leedy CE, Brunel N, and Histed MH (2020). Inhibition stabilization is a widespread property of cortical networks. *ELife* 9, e54875. [PubMed: 32598278]
- Schnabel UH, Bossens C, Lorteije JAM, Self MW, Op de Beeck H, and Roelfsema PR (2018). Figure-ground perception in the awake mouse and neuronal activity elicited by figure-ground stimuli in primary visual cortex. *Sci. Rep* 8, 17800. [PubMed: 30542060]
- Seeman SC, Campagnola L, Davoudian PA, Hoggarth A, Hage TA, Bosma-Moody A, Baker CA, Lee JH, Mihalas S, Teeter C, et al. (2018). Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *ELife* 27.

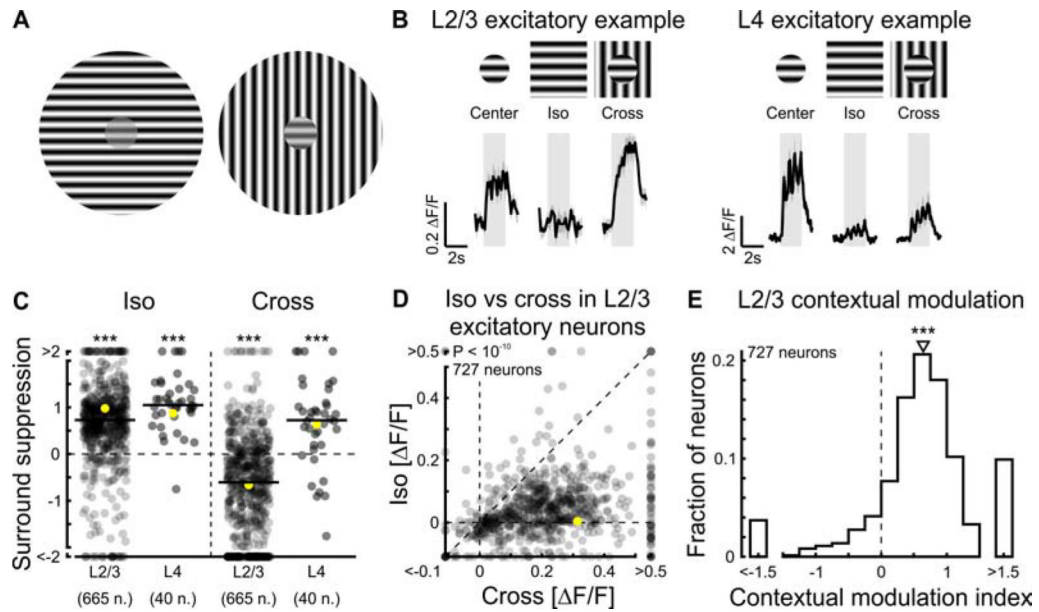


- Self MW, Lorteije JAM, Vangeneugden J, van Beest EH, Grigore ME, Levelt CN, Heimel JA, and Roelfsema PR (2014). Orientation-Tuned Surround Suppression in Mouse Visual Cortex. *J. Neurosci* 34, 9290–9304. [PubMed: 25009262]
- Seriès P, Lorenceau J, and Frégnac Y (2003). The “silent” surround of V1 receptive fields: theory and experiments. *J. Physiol.-Paris* 97, 453–474. [PubMed: 15242657]
- Sillito AM, Grieve KL, Jones HE, Cudeiro J, and Davis J (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378, 492–496. [PubMed: 7477405]
- Smith LN (2017). Cyclical Learning Rates for Training Neural Networks. ArXiv 150601186v6.
- Solomon SG (2006). Suppressing Surrounds and Contrast Gain in Magnocellular-Pathway Retinal Ganglion Cells of Macaque. *J. Neurosci* 26, 8715–8726. [PubMed: 16928860]
- Spoerer CJ, McClure P, and Kriegeskorte N (2017). Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition. *Front. Psychol* 8, 1551. [PubMed: 28955272]
- Treisman AM, and Gelade G (1980). A feature-integration theory of attention. *Cognit. Psychol* 12, 97–136. [PubMed: 7351125]
- Tremblay R, Lee S, and Rudy B (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* 91, 260–292. [PubMed: 27477017]
- Tsodyks MV, Skaggs WE, Sejnowski TJ, and McNaughton BL (1997). Paradoxical Effects of External Modulation of Inhibitory Interneurons. *J. Neurosci* 17, 4382–4388. [PubMed: 9151754]
- Vangeneugden J, van Beest EH, Cohen MX, Lorteije JAM, Mukherjee S, Kirchberger L, Montijn JS, Thamizharasu P, Camillo D, Levelt CN, et al. (2019). Activity in Lateral Visual Areas Contributes to Surround Suppression in Awake Mouse V1. *Curr. Biol* 29, 4268–4275.e7. [PubMed: 31786063]
- Vinck M, Batista-Brito R, Knoblich U, and Cardin JA (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron* 86, 740–754. [PubMed: 25892300]
- Walker GA, Ohzawa I, and Freeman RD (1999). Asymmetric Suppression Outside the Classical Receptive Field of the Visual Cortex. *J. Neurosci* 19, 10536–10553. [PubMed: 10575050]
- Xue M, Atallah BV, and Scanziani M (2014). Equalizing excitation–inhibition ratios across visual cortical neurons. *Nature* 511, 596–600. [PubMed: 25043046]
- Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang W-C, Jenvay S, Miyamichi K, Luo L, and Dan Y (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665. [PubMed: 25104383]



**Highlights**

- Visual context modulates the response of SOM oppositely to all other V1 neurons
- The VIP-SOM disinhibitory circuit controls the impact of context on V1 responses
- The VIP-SOM disinhibitory circuit controls V1 by modulating recurrent excitation
- As we predict by modelling, silencing of VIP neurons reduces contextual modulation



**Figure 1. Contextual modulation in excitatory neurons**

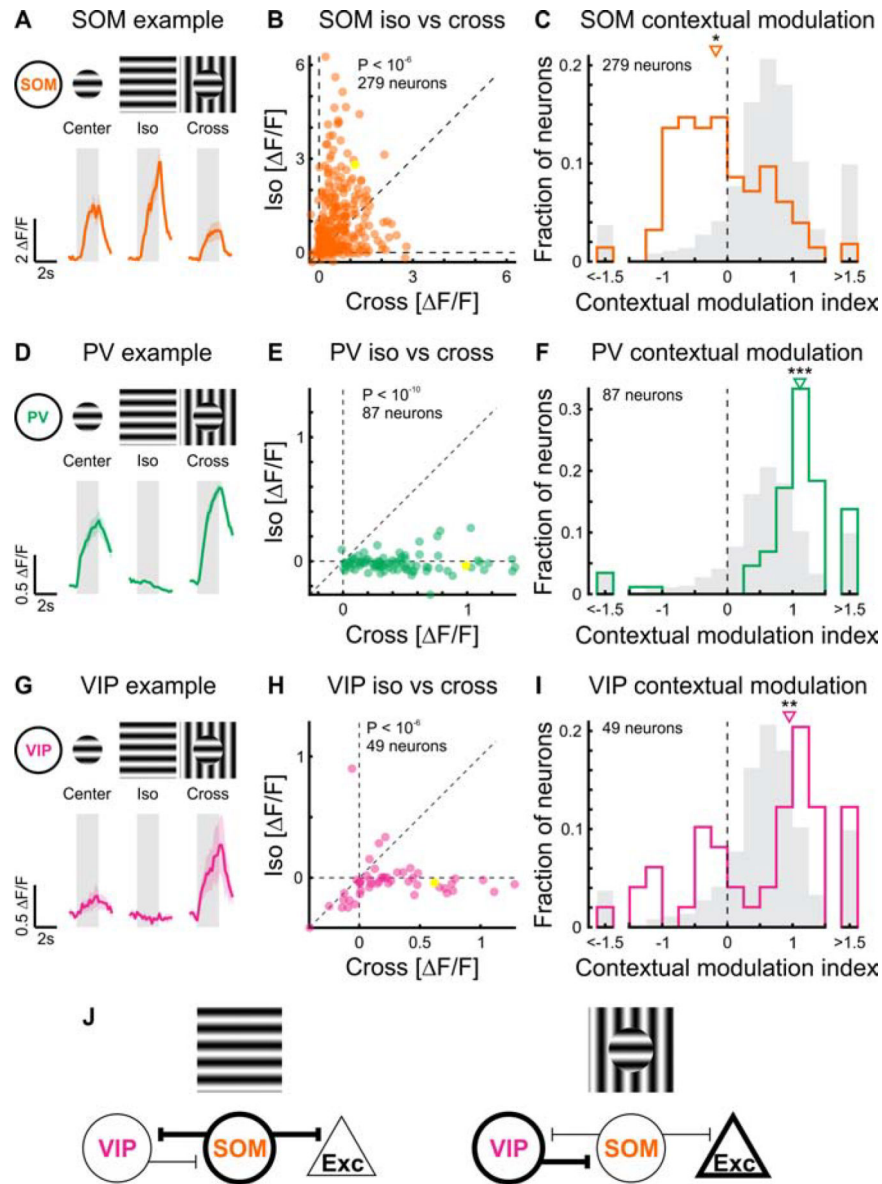
(A) The small grating patches in the centers have the same contrast but due to the distinct surround, they are perceived as more or less salient, allowing them to pop out (right) or merge with the rest of the visual scene (left).

(B) Visual stimuli were presented to awake mice while imaging calcium responses in L2/3 excitatory neurons of primary visual cortex (V1) expressing GCaMP6f or GCaMP7f. Top: Schematic of a small grating patch ( $20^\circ$  in diameter) presented alone (center), with an iso-oriented surround (iso), or with a cross-oriented surround (cross). Bottom left: Trial-averaged responses of an example L2/3 excitatory neuron to center, iso, and cross stimuli. Bottom right: Same but for an example L4 excitatory neuron. In all figures, shaded areas are periods of stimulus presentation.

(C) Surround suppression was computed for both L2/3 and L4 neurons as the difference in responses to center stimuli and the responses to iso (or cross) stimuli, normalized by the responses to center stimuli. Single-distribution two-sided sign-rank test; iso L2/3, \*\*\*:  $p < 10^{-10}$ ; cross L2/3, \*\*\*:  $p < 10^{-10}$ , 665 neurons in 9 mice; iso L4, \*\*\*:  $p < 10^{-7}$ ; cross L4, \*\*\*:  $p = 1.9 \times 10^{-4}$ , 40 neurons in 5 mice. In all panels, yellow symbols represent the example neurons shown in (B). In all figures, horizontal black lines indicate the median of the distribution.

(D) Scatter plot of L2/3 responses to iso and cross. Paired two-sided sign-rank test;  $p < 10^{-10}$  (727 neurons in 9 mice).

(E) CMI was computed as the difference divided by the sum of the responses to cross and iso stimuli. Here and in all figures triangles above histograms indicate median. Single-distribution two-sided sign-rank test;  $p < 10^{-10}$ ; same neurons as in (D). In all figures, traces and shading represent mean  $\pm$  SEM. See also Figure S1.



**Figure 2. Contextual modulation in inhibitory neurons**

(A) Top: Schematic of visual stimuli. Bottom: Trial-averaged calcium responses to center, iso, and cross stimuli of an example SOM inhibitory neuron expressing GCaMP6f.

(B) Scatter plot of the responses to iso and cross stimuli. Paired two-sided sign-rank test;  $p < 10^{-6}$ ; 279 neurons in 13 mice. Yellow symbol represents the example neuron shown in (A).

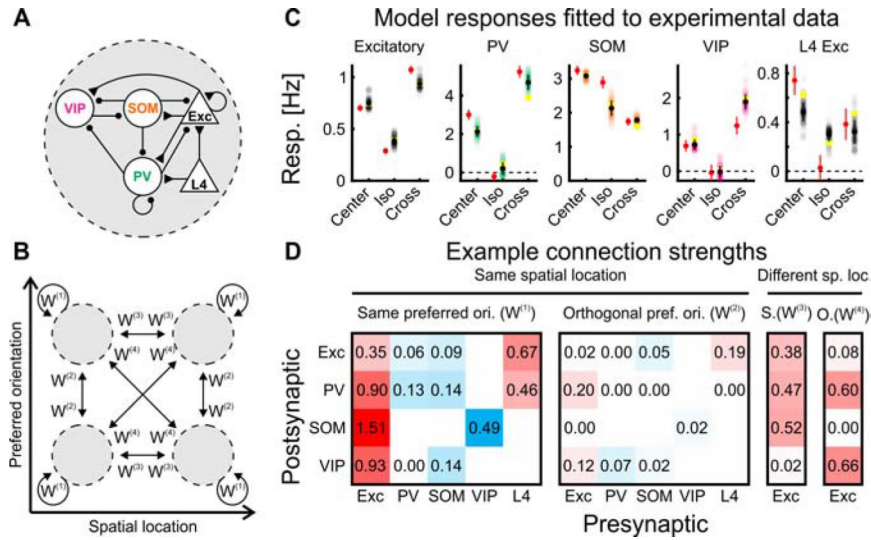
(C) CMI distribution of SOM neurons. Single-distribution two-sided sign-rank test; \*:  $p = 0.0081$ ; same neurons as in (B). Gray shading: CMI distribution of L2/3 excitatory neurons (Figure 1E)

(D-F) As above, but for PV inhibitory neurons. (E) Paired two-sided sign-rank test;  $p < 10^{-10}$ ; 87 neurons in 9 mice.

(F) Single-distribution two-sided sign-rank test; \*\*\*:  $p < 10^{-10}$ ; same neurons as in (E).

(G-I) As above, but for VIP inhibitory neurons. (H) Paired two-sided sign-rank test;  $p < 10^{-6}$ ; 49 neurons in 6 mice.

(I) Single-distribution two-sided sign-rank test; \*\*:  $p = 0.0012$ ; same neurons as in (H).  
(J) Proposed mechanism of contextual modulation of excitatory neurons through the interaction between VIP and SOM neurons. Left: In response to an iso stimulus, SOM neurons are active and inhibit both VIP and excitatory neurons. Right: In response to the cross stimulus, VIP neurons are active, inhibiting SOM neurons, which leads to relief of excitatory neurons from suppression. See also Figure S1.



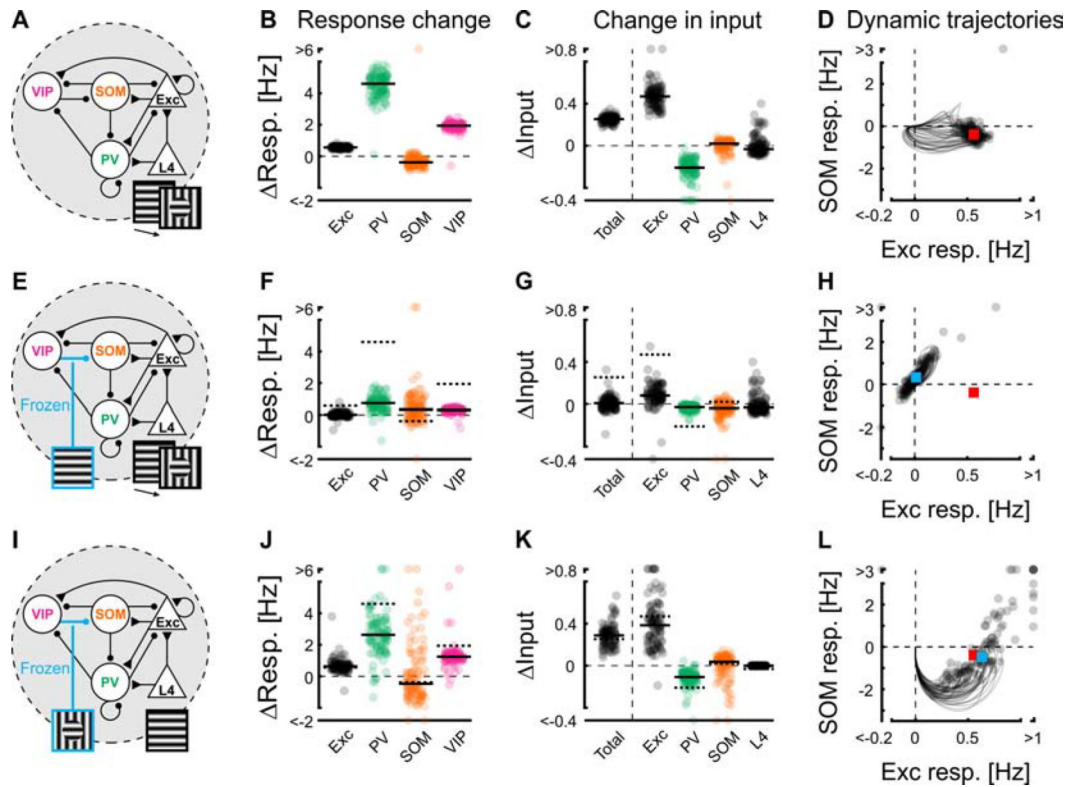
**Figure 3. A computational model trained to fit experimental data.**

(A) ‘Subnetwork’ of the model. Five unit-types, L2/3 excitatory, VIP, SOM, and PV inhibitory and L4 excitatory units form a subnetwork. Unit types were connected according to biological constraints.

(B) Four subnetworks were assigned to one of two spatial locations of the feedforward receptive field (center and surround) and one of two preferred orientations (preferred or orthogonal orientation), connected with the weight matrices  $W^{(1)}$ ,  $W^{(2)}$ ,  $W^{(3)}$ , and  $W^{(4)}$ .

(C) Responses of the different unit types in the centered and preferred-orientation subnetwork from the top 115 solutions (see Figure S2A for the unit responses in all 4 subnetworks). Each dot represents the response of a unit from a single solution. Yellow circles represent the example solution shown in (D). Black symbols represent mean  $\pm$  SD of the solutions (SD rather than SEM was used to show the range of possible solutions). Red symbols represent experimental data (mean  $\pm$  SEM; same neurons as in Figure S2A).

(D) Example connection strengths of one of the best 115 solutions. Excitatory connections are represented in red, inhibitory connections in blue, white (without numbers) indicates connections constrained to be zero. The 4 matrices correspond to  $W^{(1)}$ ,  $W^{(2)}$ ,  $W^{(3)}$ , and  $W^{(4)}$  in (B). In  $W^{(3)}$  and  $W^{(4)}$ , only excitatory projections were allowed. For medians of all connections over the 115 solutions, see Figure S2B.



**Figure 4. The VIP-SOM circuit is both necessary and sufficient for contextual modulation.**

(A) Transition from iso- to cross-response level: The activity in the network is initially at its fixed-point level in response to an iso stimulus. The network is then perturbed by switching L4 input to its cross-response level. Here and in the rest of the figure, changes are induced simultaneously in all 4 subnetworks.

(B) Change in activities of the four unit-types belonging to the centered and preferred-orientation subnetwork, after the transition from iso- to cross-response level. Zero corresponds to the iso-response levels. Here and in the rest of the figure each dot represents a solution. Here and in (C), horizontal black lines are medians across the solutions that reached a fixed point (here, 99.1%). 114 solutions reached a fixed point for all unit types.

(C) Change of inputs to excitatory units after the transition from iso- to cross-response level. Inputs were calculated as the product of pre-synaptic firing rates and corresponding connection strengths. Changes shown are total input change (black circles, left) and contribution to this change from each unit type (summed across all subnetworks). Note that, for inhibitory units, a positive change in input corresponds to a negative change in activity and *vice versa*. Same solutions as in (B).

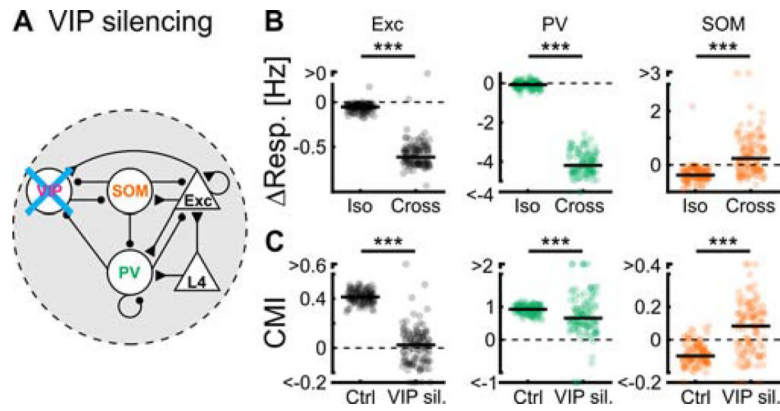
(D) Trajectories of firing rates of excitatory and SOM units starting from the iso-response level (the origin) during the transition from iso- to cross-response level. Single dots are the fixed point of the trajectories for a given solution. For clarity, we only showed the 50% of the trajectories with the shortest duration to reach their fixed point. The red square is the median activity across solutions after reaching their fixed point. Same solutions as in (B).

(E-H) As in (A-D) but, in addition to switching L4 inputs to their cross-response level, the VIP to SOM unit input has been frozen to its iso-response level. Blue square is the median

across the solutions that reached a fixed point (here, 97.4%). Red square is median from (D). Black lines in (F) and (G) are medians; dotted lines or black lines are medians from (B) and (C). 112 solutions reached a fixed point for all unit types.

(I-L) As in (E-H), but L4 inputs remain at their iso-response levels, and instead the VIP- to SOM-unit input has been switched to, and frozen at, its cross-response level. 88.7% of the solutions reached a fixed point. 102 solutions reached a fixed point for all unit types. See also Figures S3 and S4.



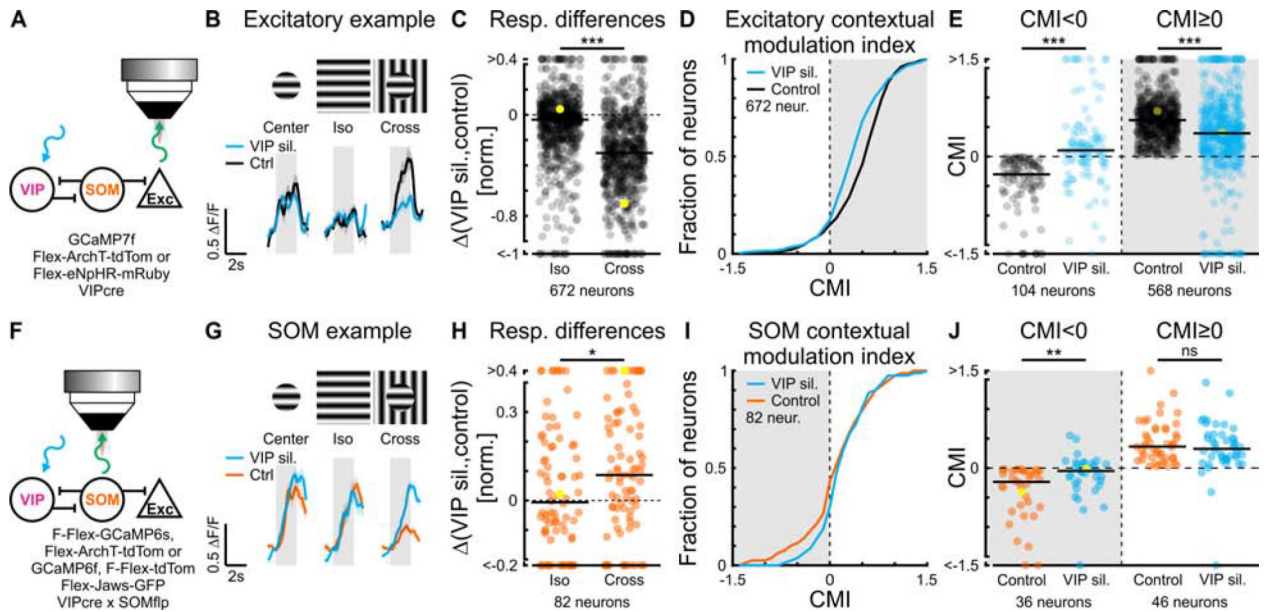


**Figure 5. The model predicts the impact of silencing VIP units on contextual modulation.**

(A) VIP units across all 4 subnetworks were silenced by fixing their activities to zero.

(B) Changes in response to iso and cross stimuli, upon silencing VIP units, of L2/3 excitatory, PV and SOM unit types in the centered and preferred-orientation subnetwork for the best 115 solutions. For all unit types: paired two-sided sign-rank test;  $p < 10^{-10}$ ; 115 solutions.

(C) CMI under control conditions compared to CMI during silencing of VIP units for the same unit types. For L2/3 excitatory and SOM units: paired two-sided sign-rank test;  $p < 10^{-10}$ ; 115 solutions. For PV units: paired two-sided sign-rank test;  $p < 10^{-3}$ ; 115 solutions. See also Figure S5.



**Figure 6. VIP and SOM neurons cooperatively contribute to contextual modulation in excitatory neurons.**

(A) Experimental setup (see STAR Methods).

(B) Trial-averaged calcium responses of a putative L2/3 excitatory neuron with and without silencing VIP neurons. Here, stimuli were presented at 50% contrast (similar responses to 100% stimuli, Figure S6E–G).

(C) Iso- and cross-response differences between silencing VIP neurons and control conditions for putative excitatory neurons. Paired two-sided sign-rank test; \*\*\*,  $p < 10^{-10}$ ; 672 neurons in 6 mice. Yellow symbol represents the example neuron shown in (B).

(D) Cumulative sum of CMI in putative excitatory neurons. Paired two-sided sign-rank test;  $p < 10^{-4}$ . Same neurons as in (C).

(E) Upon silencing VIP neurons, putative L2/3 excitatory neurons with a negative CMI increased their CMI and those with positive CMI decreased their CMI. Paired two-sided sign-rank; CMI<0 and CMI 0, \*\*\*,  $p < 10^{-10}$ ; 104 and 568 neurons, respectively, in 6 mice. Yellow symbols represent the example neuron shown in (B).

(F) Experimental setup (see STAR Methods).

(G–J) Same as (B–D), but for SOM neurons. (H) Paired two-sided sign-rank test; \*,  $p = 0.027$ ; 82 neurons in 8 mice. Yellow symbol represents the example neuron shown in (G).

(I) Paired two-sided sign-rank test;  $p = 0.12$ . Same neurons as in (H).

(J) Paired two-sided sign-rank test; CMI<0, \*\*,  $p = 0.0016$ ; 36 neurons in 6 mice; CMI 0, ns;  $p = 0.27$ ; 46 neurons in 8 mice. See also Figures S7 and S8.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
AAV2/1.ef1a.GCaMP6f.WPRE	FMI Vector Core	N/A
AAV2/1.ef1a.DIO.GCaMP6f.WPRE	FMI Vector Core	N/A
AAV2/1.CAG.CGaMP6f	Janelia Vector Core	N/A
AAV2/9.syn.GCaMP7f	Dana et al., 2019	Addgene Cat#: 104488-AAV9
AAV2/1.ef1a.fDIO.GCaMP6s	Janelia Vector Core	N/A
AAV2/5.CBA.Flex.ArchT-tdTomato.WPRE.SV40	Han et al., 2011	Addgene Cat#: 28305-AAV5
AAV2/1.CAG.Flex.rc[Jaws-KGC-GFP-ER2]	Janelia Vector Core	N/A
AAV2/9.CAG.Dio.eNpHre3.0.mRuby3.WPRE.SV40	H. Adesnik lab	N/A
AAV2/9.ef1a.F-Flex.tdTomato	Xue et al., 2014	N/A
Chemicals, Peptides, and Recombinant Proteins		
Fentanyl	West-Ward Pharmaceuticals	NDC 0641-6029-01
Midazolam	Hospira	NDC 0409-2308-01
Dexmedetomidine	Zoetis	Cat#: 122692-5
Povidone iodine	Purdue	NDC 67618-150-17
Naloxone	Hospira	NDC 0409-1219-01
Flumazenil	West-Ward Pharmaceuticals	NDC 0143-9783-01
Atipamezol	Zoetis	Cat#: 107204-6
Atropine	Abcam	Cat#: ab145582-1
Dexamethasone	Sigma-Aldrich	Cat#: 1176007-125MG
Orthojet Dental Cement	Lang Dental	Cat#: 1334
Super Bond CB Self Curing Dental Cement	Parkell	Cat#: S380
Experimental Models: Organisms/Strains		
Mouse: GAD2 <sup>tm2(cre)Zjh</sup>	The Jackson Laboratory	RRID: IMSR_JAX:010802
Mouse: Gt(ROSA)26Sor <sup>tm14(CAG-tdTomato)Hze</sup>	The Jackson Laboratory	RRID: IMSR_JAX:007914
Mouse: Tg(Scnn1a-cre)3Aibs/J	The Jackson Laboratory	RRID: IMSR_JAX:009613
Mouse: Igs7 <sup>tm148.1(tetO-GCaMP6f,CAG-tTA2)Hze</sup>	The Jackson Laboratory	RRID: IMSR_JAX:030328
Mouse: Sst <sup>tm2.1(cre)Zjh</sup>	The Jackson Laboratory	RRID: IMSR_JAX:028864
Mouse: Pvalb <sup>tm1(cre)Arbr</sup>	The Jackson Laboratory	RRID: IMSR_JAX:017320
Mouse: Vip <sup>tm1(cre)Zjh</sup>	The Jackson Laboratory	RRID: IMSR_JAX:010908
Mouse: Sst <sup>tm3.1(flpo)Zjh</sup>	The Jackson Laboratory	RRID: IMSR_JAX:028579
Software and Algorithms		
LabVIEW	National Instruments	RRID: SCR_014325
MATLAB	The MathWorks	RRID: SCR_001622
Calliope: Image processing software	N/A	<a href="https://svn.code.sf.net/p/iris-scanning/calliope/">https://svn.code.sf.net/p/iris-scanning/calliope/</a>
Spike estimation algorithm	Deneux et al., 2016	<a href="https://github.com/MLspike">https://github.com/MLspike</a>