Supplementary materials for:

The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex

Daniel B. Rubin, Stephen D. Van Hooser, and Kenneth D. Miller

Contents

S1 Methods **4**
S1.1 Modeling Methods in Brief 4
S1.1.1 Determination of external input shape $h(x)$ 4
S1.1.2 Parameters 4
S1.1.3 Simulation methods 5
S1.1.4 Tuning curve plots 5
S1.2 Experimental Methods in Brief 6
S1.2.1 Animal care, surgery, and recording 6
S1.2.2 Visual stimulus protocols 6
S1.2.3 Maximum likelihood estimation (MLE) and DOG and SSM models . 7
S1.3 Supplemental Methods Information: Modeling 7
S1.3.1 External input shape: General considerations and supplementary figures 7
S1.3.2 Parameters: General considerations 7
S1.3.3 DoG and SSM fits to model data: General considerations 8
S1.3.4 Determining size of stimuli used by Haider et al. (2010), relative to CRF size 9
S1.4 Supplemental Methods Information: Experiments 9
S1.4.1 Recording 9
S1.4.2 Maximum likelihood estimation (MLE) for denoising of tuning curves 9
S1.4.3 Fits of DoG and SSM models to tuning curves 13
S1.5 Statistical Methods 14
S1.5.1 Assumptions of Normality 14
S1.5.2 Test statistics for statistical tests of Figs. 4C,F, 5F and 7A,E 15

S2 Supplemental Text: Mathematical Analysis and Related Theoretical Issues **16**
S2.1 Network Resonant Frequencies . . . 16
S2.1.1 Inhibitory resonant frequency . . . 17
S2.1.2 Excitatory resonant frequency . . . 19
S2.1.3 Understanding spatially periodic activity in the model from the resonant frequencies . . . 20
S2.2 Critical frequency, and an experimental test . . . 21
S2.2.1 The critical frequency . . . 21
S2.2.2 Experimental Prediction: Ideal Case. . . . 23
S2.2.3 Experimental Prediction: More Realistic Case. . . . 24
S2.3 Linearization of the Nonlinear Model . . . 26

S3 Supplemental Figures **29**

S4 Supplemental Tables: Legends **50**

List of Figures

S1 MLE for denoising of tuning curves, and fits of DOG and SSM models to tuning curves (related to Supplement S1.4.2-S1.4.3) . . . 10
S2 Resonant frequencies and model linearization (related to Supplement S2.1-S2.3, and main text discussions of Figs. 3,5,8) . . . 25
S3 Increasing dominance of network input by inhibition with increasing input strength, using protocol of Shao et al. (2013) (related to Figs. 1F, 3B, and 6C) 30
S4 Normalization in the ring model of Figure 1 . . . 31
S5 Ring model results of Fig. 1 are well fit by equations of the normalization model 33
S6 Robustness of results for ring model (Figure 1) and Nonlinear 1-D model (Figure 3) . . . 35
S7 Effects of varying input shape, from sharp-edged to smoothly tapered, in the linear model of Fig. 2 . . . 36
S8 Nonlinear 1-Dimensional Spatial Model of Figure 3 . . . 37
S9 Experimental Results (related to Figs. 4 and 5) . . . 39
S10 All experimentally measured size tuning curves (related to Fig. 4) . . . 41
S10 All experimentally measured size-tuning curves (related to Fig. 4); (continued) 42
S11 All experimentally measured position-tuning curves (related to Fig. 4) . . . 43
S11 All experimentally measured position-tuning curves (related to Fig. 4); (continued) . . . 44
S12 All experimentally measured contrast-modulation tuning curves (related to Fig. 5) . . . 45
S12 All experimentally measured contrast-modulation tuning curves (related to Fig. 5); (continued) . . . 46

S13 In a nonlinear two-neuron model, shifting the balance of input towards E or I can cause a multiplicative change in the gain (related to Discussion) 47
S14 How the model could simultaneously account for results of Ozeki et al. (2009) and Haider et al. (2010) (related to Discussion) 48

List of Tables

S1 Statistics for length-tuning analyses of Fig. 4B-C 50
S2 Statistics for length-tuning analyses of Supplemental Fig. S1f-g 50
S3 Statistics for position-tuning analyses of Fig. 4E-F 50
S4 Statistics for length-tuning analyses of model units, Fig. 6C, left 50
S5 Statistics for position-tuning analyses of model units, Fig. 6C, right 51

S1 Methods

We first present in brief the information needed to replicate our results, excepting the details of the maximum likelihood estimation (MLE) procedure (Sections S1.1-S1.2). We then present supplemental modeling information (Sections S1.3-S1.4), including the MLE details in Section S1.4.2.

S1.1 Modeling Methods in Brief

We define the Gaussian $G_\sigma(x,y) = e^{-\frac{d(x,y)^2}{2\sigma^2}}$, where for positions $d(x,y) = |x-y|$, for orientations $d(x,y) =$ shortest distance around circle of circumference 180^o between x and y. We define $\sim \mathcal{N}(m,\sigma^2)$ to mean distributed as a normal distribution with mean m, variance σ^2.

S1.1.1 Determination of external input shape $h(x)$

Input of length l had shape $s_l(x) = \left(\frac{1}{1+e^{-\frac{x+l/2}{\sigma_\text{RF}}}}\right)\left(1-\frac{1}{1+e^{-\frac{x-l/2}{\sigma_\text{RF}}}}\right)$. Fig. 1: stimulus of orientation ϕ had $h(\theta) = G_{\sigma_\text{FF}}(\phi,\theta)$. Figs. 2-3: $h(x) = s_l(x)$. Figs. 6-8: stimulus centered at 2D position $\mathbf{x}'$, orientation ϕ has $h(\mathbf{x}) = s_l(|\mathbf{x}-\mathbf{x}'|)G_{\sigma_\text{FF}}(\phi,\theta(\mathbf{x}))$ ($\theta(\mathbf{x})$ is preferred orientation of units at 2D position $\mathbf{x}$). For full-field gratings, $s_l(\cdot)$ is replaced by 1.

S1.1.2 Parameters

All figs: $\tau_E = 20$ms, $\tau_I = 10$ms.
Figs. 1,2,3: there are N E/I units with grid spacing $\Delta\theta$ (Fig. 1) or Δx (Figs. 2,3).
Fig. 1: Connections are $W_{ab}(\theta-\theta') = J_{ab}G_{\sigma_\text{ori}}(\theta,\theta')$. $N = 180$, $\Delta\theta = 1^\text{o}$, $J_{EE} = 0.044$, $J_{IE} = 0.042$, $J_{EI} = 0.023$, $J_{II} = 0.018$, $\sigma_\text{ori} = 32^\text{o}$, $\sigma_{FF} = 30^\text{o}$, $k = 0.04$, $n = 2.0$.
Figs. 2,3: Excitatory projections are $W_{aE}(x,x') = J_{aE}G_{\sigma_{aE}}(x,x')$ for $a \in \{E,I\}$. Inhibitory projections W_{aI} are only to same grid position as projecting neuron. **Figure 2**: $N = 401$, $\Delta x = 0.25^\text{o}$, $\sigma_\text{RF} = 0.33\Delta x$, $J_{EE} = 0.385$, $J_{IE} = 1.0$, $W_{EI} = 0.55$, $W_{II} = 1.5$, $\sigma_{EE} = 0.5^\text{o}$, $\sigma_{IE} = 1^\text{o}$, $c = 1$. **Figure 3:** $N = 101$, $\Delta x = \frac{1}{3}^\text{o}$, $\sigma_\text{RF} = 0.125\Delta x$, $J_{EE} = 1.0$, $J_{IE} = 1.25$, $W_{EI} = 1.0$, $W_{II} = 0.75$, $\sigma_{EE} = \frac{2}{3}^\text{o}$, $\sigma_{IE} = \frac{4}{3}^\text{o}$, $k = 0.01$, $n = 2.2$.
Figs. 6,7: Grid is 75×75 E/I pairs, grid interval $\Delta x = \frac{16}{75}^\text{o}$, periodic boundary conditions. The map of preferred orientations $\theta(\mathbf{x})$ ($\mathbf{x}$ is 2D position) is randomly generated using the method of Kaschube et al. (2010) (their supp. materials, Eq. 20) with $n = 30$, $k_c = \frac{8 \text{ cycles}}{75 \text{ grid intervals}}$. Let $W_{ab}(\mathbf{x},\mathbf{x}')$ be synaptic weight from unit of type b (E or I), position $\mathbf{x}'$ to type a, position $\mathbf{x}$. Nonzero connections are sparse, chosen with probability $p\left(W_{ab}(\mathbf{x},\mathbf{x}') \neq 0\right) = \kappa_b G_{\sigma_{aE}}(x,x')G_{\sigma_\text{ori}}(\theta(\mathbf{x}),\theta(\mathbf{x}'))$. For nonzero connections, $W_{ab}(\mathbf{x},\mathbf{x}')$ is $\sim \mathcal{N}(J_{ab},(0.25J_{ab})^2)$; negative weights are set to zero. Weights of a given type b onto each unit are then scaled so that all units of a given type a receive the same total type b synaptic weight, equal to J_{ab} times the mean number of connections received under $p\left(W_{ab}(x,x') \neq 0\right)$. τ_E, τ_I, n_E, n_I, and k are $\sim \mathcal{N}(m,(0.05m)^2)$ where m is the corresponding mean value. n_E

and n_I are exponents of power law for E or I cells respectively. Parameters (mean values if stochastic): $\kappa_E = 0.1$, $\kappa_I = 0.5$, $J_{EE} = 0.10$, $J_{IE} = 0.38$, $J_{EI} = 0.089$, $J_{II} = 0.096$, $k = 0.012$, $n_E = 2.0$, $n_I = 2.2$, $\sigma_{EE} = 8\Delta x$, $\sigma_{IE} = 12\Delta x$, $\sigma_{EI} = \sigma_{II} = 4\Delta x$, $\sigma_{ori} = 45^{\rm o}$, $\sigma_{FF} = 32^{\rm o}$, $\sigma_{\rm RF} = \Delta x$.

Figs. 2,3,6: CM tuning is studied using a full-field grating multiplied by the contrast modulation $\frac{1}{2}\left(1 + \sin\left(2\pi\left(k \cdot x + \omega t\right)\right)\right)$ where x and k are 1D in Figs. 2 and 3 and 2D in Fig. 6 and $\omega = 4Hz$. Stimulus was run for 1 sec (1ms time steps). Preferred CM spatial frequency (SF), and preferred CM orientation in Fig. 6, were those producing maximum peak response over that time.

In Fig. 6, there were problems with units having response peaks of similar heights for multiple harmonics of a fundamental preferred SF. For this figure, we defined CM preferred SF as follows. We Fourier transformed the CM SF tuning curve and found its peak X, a number representing cycles/(tuning curve). Let f be the highest frequency in the CM SF tuning curve. We then successively looked at intervals of the tuning curve 0 to Nf/X, $N = 1, 2, \ldots,$ floor(X) followed by the full interval, looking for a response peak that was at least 95% of the height of the global peak of the tuning curve. The first such peak that was found was deemed the preferred SF. Visual inspection of many tuning curves confirmed that this procedure correctly isolated the fundamental of a harmonic stack of similar responses and otherwise simply found the peak of the tuning curve.

Figure 8: Parameters as in Fig. 6. All stimuli are full field. Photostimulus drifted at 3 Hz. Low input: tonic input 1, modulatory amplitude 1. High input: tonic input 40, modulatory amplitude 10. Spatial frequencies: 0.03 c/deg (low), 0.5 c/deg (high). Illustrated E/I pair chosen at random; grating at preferred orientation. Photostimulus oriented vertically across model cortex as shown in Fig. 6A.

S1.1.3 Simulation methods

For the linear model, the steady state can be explicitly determined. Using matrix/vector notation, letting $\mathbf{r}$ be the vector of firing rates of all units across the network (both E and I), $\mathbf{h}$ the vector of their external inputs, $\mathbf{W}$ the matrix of connection weights between these neurons and $\mathbf{1}$ the identity matrix of the same dimension, the steady state activities $\mathbf{r}_{\rm SS}$ are given by $\mathbf{r} = (\mathbf{1} - \mathbf{W})^{-1}\mathbf{h}$, which was numerically calculated. For the nonlinear models, steady states (and dynamics in CM-tuning simulations, Fig. 8, and Supplemental Fig. S2c) were determined by simulations using simple forward Euler method with 1 ms time step. All simulations were run from multiple initial conditions to confirm independence of final state from initial condition and tested with shorter time steps to ensure no changes in results.

S1.1.4 Tuning curve plots

All model tuning curves linearly connect points sampled at high resolution, without smoothing.

S1.2 Experimental Methods in Brief

S1.2.1 Animal care, surgery, and recording

Animal care protocols conformed to NIH guidelines and were approved by the Brandeis University Institutional Animal Care and Use Committee. Eight adult ferrets (90-110 days of age; 7 females, 1 male) were studied. They were singly or doubly housed and kept on a 12 hour light/12 hour dark cycle. Anesthesia was induced with a mix of ketamine (30 mg/kg) and xylazine (3.0 mg/kg) (IM) and maintained with 0.5–1.5% isoflurane. Atropine (0.2 ml, 0.5 mg/ml) was administered. All wound margins were infused with the long-lasting analgesic bupivicane (0.25%). A nylon tracheal tube was inserted by tracheotomy, and animals respirated on a 2:3 mixture of oxygen:nitrous oxide, switched to 1:1 after surgeries. Silicon oil was placed on the eyes to prevent corneal damage. The animal was secured in a stereotaxic frame, a small craniotomy (4×8 mm wide) was made and the dura removed. Ferrets were then paralyzed with continuous IP infusion of gallamine triethiodide (0.2mg/hr) to suppress spontaneous eye movements and respirated with a 1:1 mixture of oxygen:nitrous oxide. Heart rate, end-tidal CO_2 (maintained between 3.2% - 5.0%), and temperature were monitored continuously throughout surgeries and recordings.

Extracellular signals were recorded from upper layers of ferret V1 using carbon fiber microelectrodes (Kation Scientific).

S1.2.2 Visual stimulus protocols

Receptive field mapping: After initial coarse mapping of the CRF and preferred orientation, we determined CRF center (by switching between circular and complementary annular stimulus, progressively shrinking radius while searching for position that gave response to circular but not annular stimulus), then preferred direction (drifting gratings of 16 directions, 22.5° steps, spatial frequency 0.1 cycles/deg, temporal frequency 4 Hz), and then preferred spatial frequency (SF) (preferred-direction gratings of 7 SF's). Then studies of length, position, and CM tuning were carried out.

Tuning studies: Luminance gratings always had cell's preferred orientation, direction, and spatial frequency. For length tuning (30 stimuli, radius 1° to 30° by 1°) and position tuning (21 30°-diameter stimuli, position $-15°$ to 15° by 1.5°, moved parallel to grating orientation), stimuli were presented for 2 sec, drift rate 4 Hz, interstimulus interval (ISI) 4 sec. Response measure was mean firing rate during grating presentation. CM grating varied contrast sinusoidally from 0% contrast to luminance-grating contrast. We first found CM preferred orientation (among 8 orientations, 22.5° spacing) with 70% luminance contrast and CM SF 0.05 cycles/deg. This CM orientation was used to study CM SF tuning across luminance contrasts (24 stimuli: SFs 01, .015, .02, .03, .04, .05 c/deg; contrasts 4%, 8%, 16%, 64%). CM stimuli were presented for 4 seconds, drift rates 4 Hz (luminance), 1 Hz (CM), ISI 4 sec. Response measure was 1 Hz component of firing rate during CM grating presentation. For each tuning type, stimuli were presented in pseudorandom order; each stimulus appeared 4–5 times.

S1.2.3 Maximum likelihood estimation (MLE) and DOG and SSM models

We used MLE to improve our estimates of tuning curves by removing the effects of slow, stimulus-independent variations in excitability. Details are in Section S1.4.2.

For Fig. 4 and Fig. 6I, SSM and DOG curves were defined as follows. The error function is defined by $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dz\, e^{-z^2}$, and ranges from -1 to 1. We define the non-negative function $z(x) = (\mathrm{erf}(x) + 1)/2$. For size-tuning curves $r(x)$, the DoG model is $r(x) = a_1 z((x - a_2)\, a_3) - a_4 z((x - a_5)\, a_6)\, e^{a_7 x} + a_8$. The SSM model eliminates the a_8 term and replaces e^{a_7} with $(e^{a_7 x} \cos(a_9 x + a_{10}) + a_8)$. (Note that setting $a_9 = a_{10} = 0$ reduces the SSM model to same form as the DOG model, as becomes clear if the equations are reexpressed in terms of $\mathrm{erf}(x)$ rather than $z(x)$; that is, the models are nested.) For position-tuning curves $p(x)$, the DoG model is $p(x) = a_1 e^{\frac{-(x-a_2)^2}{2a_3^2}} - a_4 e^{\frac{-(x-a_5)^2}{2a_6^2}} + a_7$. The SSM model replaces a_4 with $a_4(\cos(a_8 x + a_9) + a_{10})$. SSM and DOG models were fit by minimizing sum-squared error using the Matlab function lsqcurvefit.

S1.3 Supplemental Methods Information: Modeling

S1.3.1 External input shape: General considerations and supplementary figures

We adopted the function $s_l(x)$ that defines "sharp-edged" stimuli (Section S1.1.1) early in our work, but it is very similar to the form that would be obtained by assuming that the spatial envelope of the CRF is given by a Gaussian function of retinotopic space and that the external input to a given cell is the integral of the product of its Gaussian CRF with the stimulus. If we let the Gaussian CRF have standard deviation σ_G and assume a sharp-edged stimulus constant over $x = -l$ to $x = l$, this integral is proportional to $f_l(x) = \frac{1}{2}\left(\mathrm{erf}\left(\frac{x+l/2}{\sigma_\mathrm{G}\sqrt{2}}\right) + \mathrm{erf}\left(\frac{l/2-x}{\sigma_\mathrm{G}\sqrt{2}}\right)\right)$. Identifying $\sigma_\mathrm{RF} = 0.6\sigma_\mathrm{G}$, $s_l(x)$ and $f_l(x)$ are essentially identical for $\sigma_\mathrm{RF} < 0.25l$, a condition that holds for almost all of our stimuli. As σ_RF increases from $0.25l$ to l, $s_l(x)$ becomes about 10% wider than $f_l(x)$ (half width at half height) and its amplitude relative to $f_l(x)$ decreases to about $0.6 f_l(x)$.

In Supplemental Figs. S7 and S8, we also considered more smoothly-tapering stimuli. For Gaussian-shaped stimuli of length l, $h(x) = g_l(x)$ where $g_l(x) = e^{-\frac{x^2}{2\sigma_l^2}}$ with $\sigma_l = \frac{l/2}{\sqrt{2*\ln(2)}}$ (l is full-width at half-height of the input). We considered varying degrees of tapering by linearly interpolating between $s_l(x)$ and $g_l(x)$ to produce input profiles $h(x) = a s_l(x) + (1 - a) g_l(x)$ with $a = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.

Throughout the paper, external input was taken for simplicity as identical for E and I units. In Ahmadian et al. (2013) we consider aspects of the more general case in which external drives to E and I are not equal.

S1.3.2 Parameters: General considerations

We did not attempt to tune parameters to quantitatively match data, but instead aimed at explicating simple, robust mechanisms underlying qualitative behaviors. For similar reasons

we ignored many biological details that should be explored in future work, notably the different types of I cells and their (and E cells') different cellular and synaptic properties, cortical layers, biological properties such as spike-rate adaptation or synaptic depression that would reduce the supralinearity of input/output functions, and the dynamic increase in the power-law exponent n and associated change in the parameter k that would be induced by the reduction in voltage noise levels with contrast (Sadagopan and Ferster 2012) (because the effective power is roughly determined by the distance from rest to threshold in units of the voltage noise standard deviation (Hansel and van Vreeswijk 2002, Miller and Troyer 2002)). We used equal numbers of E and I units, but we expect no significant changes if reduced numbers of I units are used with correspondingly stronger connections, based on similar behavior of spiking models with realistic proportions of I neurons and rate models with equal numbers of E and I units (Murphy and Miller 2009, and D. Obeid and K.D. Miller, unpublished observations of SSN models).

Since this work was largely completed we realized (Ahmadian et al. 2013) that the most strongly nonlinear behavior would likely arise for $\Omega_E < 0$, $\Omega_E < \Omega_I$ where, for equal external input to E and I units as used here, $\Omega_E \equiv \bar{W}_{II} - \bar{W}_{EI}$, $\Omega_I \equiv \bar{W}_{IE} - \bar{W}_{EE}$, and $\bar{W}_{XY}$ is the total synaptic strength from units of type Y received by a unit of type X. This condition was not met in Fig. 6, which had $0 < \Omega_E < \Omega_I$. These simulations nonetheless worked well enough for our purposes of displaying qualitative behavior.

We used $\tau_I = \frac{1}{2}\tau_E$. Using faster inhibition than excitation helps ensure network stability but does not otherwise affect steady state responses.

In the 2D simulations (Figs. 6-8), degrees can be converted to distance across cortex by assuming a cortical magnification factor of 0.6 mm/deg, a typical figure for $5-10^o$ eccentricity in the cat (Albus 1975), giving $\sigma_{EE} = 1.02$mm, $\sigma_{IE} = 1.54$mm, $\sigma_{EI} = \sigma_{II} = 0.51$mm, orientation map period 1.2mm. We used slightly different exponents, $n_I > n_E$, to increase stability despite variability (experiments suggest much larger differences: Supplemental Fig. S3 of Haider et al. 2010). Variability of τ's, n's, k was limited because larger variability tended to yield instability. Biologically, large variability can probably be tolerated without instability because of various forms of homeostatic compensation (Turrigiano 2011, Vogels et al. 2011) as well as mechanisms such as adaptation/depression in excitatory cells and synapses or facilitation in inhibitory synapses, not modeled here.

S1.3.3 DoG and SSM fits to model data: General considerations

Model length-tuning curves have 61 data points (evenly spaced from 0.5 to 75 grid units). In Fig. 6I, mean/median SSE's per data point for position tuning are 1.4/1.8 (DOG) or 3.5/4.6 (SSM) times those of experiments; for length tuning, 0.09/0.08 (DOG) or 0.10/0.08 (SSM). These differences, of unknown cause, are similar for DOG and SSM and so do not indicate differences in periodicity of model vs. real cells.

S1.3.4 Determining size of stimuli used by Haider et al. (2010), relative to CRF size

In the Discussion we discuss the size of the stimuli used by Haider et al. (2010), based on the following calculation. We assume a Gaussian CRF profile (Jones and Palmer 1987), which we take for simplicity to be circularly symmetric. We define the standard deviation of the Gaussian to be length 1. Thus the CRF is proportional to $G(r) \equiv \frac{1}{2\pi} e^{-r^2/2}$ where r is the radial coordinate in 2D polar coordinates. In response to a stimulus of strength 1 uniformly stimulating the CRF out to radius x (measured in units of the standard deviation), the output of the CRF filter applied to the stimulus is $\int_0^x r dr \int_0^{2\pi} d\theta\, G(r) = 1 - e^{-x}$. We assume the CRF size is 2 standard deviations of the Gaussian, so maximal filter output is $1 - e^{-2}$ (*e.g.* beyond that radius surround suppression would set in). Finally we assume the neuron's response is the filter output raised to a power n. Thus, the condition that the stimulus radius x gives half-maximal response is $\left(\frac{1-e^{-x}}{1-e^{-2}}\right)^n = \frac{1}{2}$, or $x = -\ln\left(1 - 2^{-1/n}\left(1 - e^{-2}\right)\right)$. For n ranging from 2 to 3, this gives x ranging from 0.94 to 1.16 (in standard deviations), or from .47 to .58 of CRF size.

S1.4 Supplemental Methods Information: Experiments

S1.4.1 Recording

Spikes were amplified with a preamplifier/amplifier system by Multichannel Systems (Germany) and acquired and clustered using a Micro1401 acquisition board and Spike2 software (Cambridge Electronic Design, LLC). Visual stimuli were created in Matlab using the Psychophysics Toolbox (Kleiner et al. 2007) on a Macintosh Pro (OS10) and displayed on a Sony monitor (GDM-520) (100 Hz refresh and frame rates). Spike tuning curves were analyzed with custom software in Matlab.

S1.4.2 Maximum likelihood estimation (MLE) for denoising of tuning curves

To improve our estimates of tuning curves, which represent the stimulus-specific component of responses, we used MLE to remove the effects of slow, stimulus-independent variations in excitability (Supplemental Figs. S1a,b). This was particularly important for our data because we presented each stimulus only four or five times. Because we wished to explore a large region of stimulus parameter space with each cell, we were constrained in the number of times we could present each stimulus by the amount of time we could reasonably expect to record from a single neuron. Removing the slow stimulus-independent variations in rate improves our statistical power to estimate tuning curves. We also incorporated our prior knowledge that tuning curves are smooth by using constraints on tuning curve smoothness, which also improves our statistical power by effectively combining data across nearby stimuli.

To remove stimulus-nonspecific slow variations in response magnitude, we modeled a neuron's firing rate to stimulus type i (*i.e.*, a given stimulus size, stimulus position, or CM contrast and SF) presented at time t as $r_i(t) = g(t)s_i$ where $g(t)$ represents slow changes in

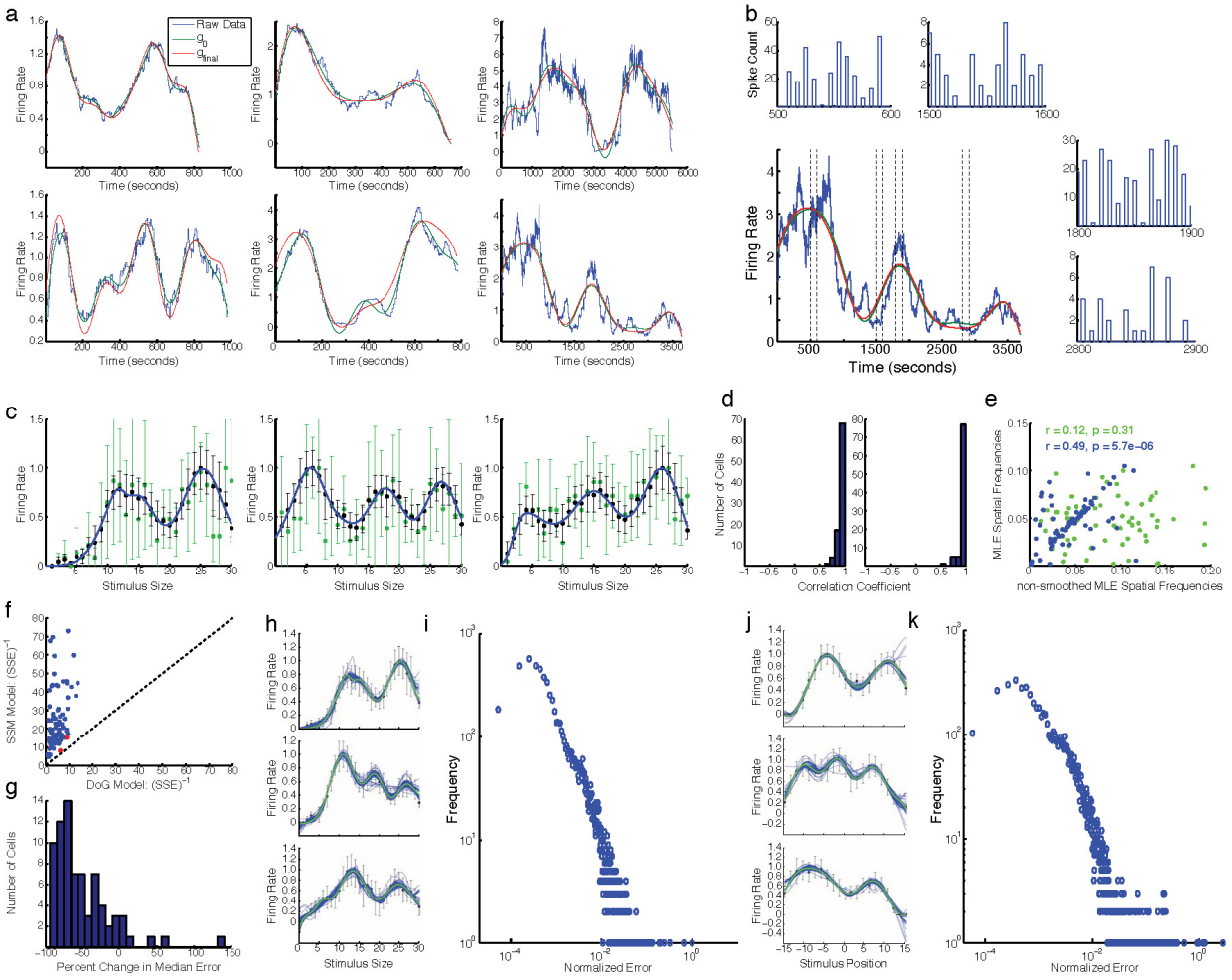


Figure S1:

MLE for denoising of tuning curves, and fits of DOG and SSM models to tuning curves (related to Supplement S1.4.2-S1.4.3). (a). Examples of slow changes in cortical excitability. The firing rate versus time for six example cells is shown in blue (calculated with a 100ms sliding window). The initial estimate of the nonspecific scaling function, $g_0(t)$, is in green. The adjusted $g(t)$ that maximizes the likelihood of both $g(t)$ and the stimulus specific response vector $s_i(t)$ is shown in red. **(b)**. Slow fluctuations in cortical excitability do not obscure stimulus selectivity. We examine one firing rate curve from **(a)** while highlighting four 100-second time epochs, marked by adjacent pairs of dashed vertical lines. Each subplot shows firing rates during one of these epochs, in bins across the 100 seconds, with each bin indicating response to a single stimulus or a single ISI (ISI responses are small or zero, so ISI bins mostly appear to be spaces between the bins showing stimulus responses). The stimuli were shown in pseudorandom order so the particular stimuli shown are different in each epoch. During periods of both high and low cortical excitability, the cell responds differentially to different stimuli. Note the changing scale on the y-axes of the subplots. **c.** Example outcomes of MLE denoising: three example size-tuning curves. The green data are the raw size-tuning curves, and the black data indicate the size-tuning responses after the MLE de-noising process. The best-fit SSM model curve for the post-processed data is in blue.
(Continued on next page.)

Figure S1:

(Continued). **d.** MLE denoising preserves response properties. Left: Histogram of the correlation coefficients between orientation tuning curves before and after the MLE denoising process. The overwhelming majority of cells show a near perfect correlation between the raw tuning curve and the post-processed tuning curve. Right: Histogram of the correlation coefficients of luminance spatial frequency tuning curves before and after the MLE denoising process. **e.** MLE smoothing eliminates high frequencies from SSM fits. We consider frequencies of the oscillation in fits of the SSM model to size-tuning curves (76 cells). The green dots show fits to the data from MLE denoising without smoothing (x-axis) and from MLE denoising with smoothing (y-axis). The two show no significant correlation ($r = 0.12$, $p = 0.31$; $DF = 74$, $t = 1.014$). We found the Gaussian spatial filter that, applied to the unsmoothed MLE size-tuning curves, gave the least mean-squared error from the smoothed MLE curves, which had standard deviation $\sigma = 0.99^{\text{o}}$. The blue dots compare frequencies from SSM fits to this Gaussian smoothing of the unsmoothed MLE curves (x-axis) to those from the smoothed MLE curves (y-axis); these are now highly correlated ($r = 0.49$, $p = 5.7 \times 10^{-6}$; $DF = 74$, $t = 4.891$). Thus the smoothing in the MLE algorithm has very similar effect to smoothing with a Gaussian filter with $\sigma = 0.99^{\text{o}}$. **f,g.** Statistics of size-tuning fits compared to the classic Difference of Gaussians (DoG) model. Same analysis as Fig. 4B-C, but here comparison is to a more classic version of the DoG model, setting parameter $a_7 = 0$ in the equation for the size-tuning DoG (Section S1.2.3). **f.** The reciprocal of the sum squared error for the DoG and SSM models, conventions and statistical tests as in Fig. 4B. 74/76 cells had significantly better fits by SSM than classic DoG by nested F-test. **g.** Results of cross-validation analysis, conventions as in Fig. 4C. For 69/76 cells the SSM model fit gave less median sum-squared error than the classic DoG model ($p = 6.4 \times 10^{-14}$, binomial test; median of distribution significantly different from zero, $p = 1.1 \times 10^{-11}$, Wilcoxon signed-rank test, SignedRank= 2775; statistical tests as in Fig. 4C). Statistics for all cells given in Supplemental Table S2. **h-i.** Cross-validation analysis of size-tuning model fits. **h.** Three examples of the cross-validation process. In each case, the original SSM curve, fit with all 30 data points, is in green. The 100 cross-validation fits, each fit with 24 randomly selected points (80% of the original data), are shown in blue. **i.** A log-log histogram of the normalized error of the cross-validation fits relative to the full-data fit shows a roughly power-law distribution, indicating that most cross-validation fits differ very little from their respective full-data fit. Letting r_i^a and r_i be the firing rate at position i for cross-validation estimate a and for the full-data fit respectively, the normalized error for estimate a is $\frac{\sum_i (r_i^a - r_i)^2}{\sum_j r_j^2}$. **j-k.** Cross-validation analysis of position-tuning model fits. **j.** Three examples of the cross-validation process. Conventions as in **h**. **k.** A log-log histogram of the normalized error between the cross-validation fits and the full-data fit. Conventions as in **i**.

overall cortical excitatability and s_i is stimulus-specific response (the tuning curve). Time was binned, one bin for each stimulus onset/offset interval and one for each ISI (ISI's are an additional stimulus type). The initial $g(t)$ was found by least-squares fitting of the spike train, smoothed with a 100ms sliding window, to the 12-parameter function, $g(t) = \sum_{i=1}^{4} A_i \sin(2\pi f_i t + \phi_i)$ where the f_i were constrained to be < 10 cycles/experiment duration (exp't duration typically 8-16 minutes) to ensure that $g(t)$ captures only slow variations corresponding to changes in cortical state rather than changes in stimuli. We then maximized the log likelihood, alternating maximizing over the s_i with $g(t)$ fixed and over the 4 A_i parameters of $g(t)$ with other parameters of $g(t)$ and the s_i fixed. To incorporate prior knowledge of the smoothness of tuning curves, for length- and position-tuning curves, the s_i maximization was constrained so that, after setting maximum s_i to 1, the second derivative of the tuning curve was always $< 0.10/\text{deg}^2$ (length-tuning) or $< 0.15/\text{deg}^2$ (position tuning). Smoothing was not used for CM-tuning because those curves had only six points, sufficiently spaced that we had no a priori reason to expect smooth curves. For length and position tuning, the likelihood of an observed k spikes in a given time bin of size Δt at time t with stimulus i was given by a Poisson distribution with mean $r_i(t)\Delta t$. For CM tuning, the log likelihood of an observed envelope F1 amplitude k in a given bin was $-(k - r_i(t)\Delta t)^2$.

To find error bounds for our estimates, we derive the Hessian matrix of the log likelihood – the matrix of second derivatives of the log likelihood (L.L.) with respect to the s_i's. The Hessian is diagonal – derivatives $\frac{\partial^2 \text{L.L.}}{\partial s_i \partial s_j}$ are zero for $i \neq j$ – and the diagonal entries are just $\frac{\partial^2 \text{L.L.}}{\partial s_i^2}$. In this case, in the neighborhood of the maximum likelihood estimate, the probability of s_i can be approximated by a Gaussian, centered at the maximum likelihood value, with variance given by the negative of the inverse of the i^{th} diagonal entry of the Hessian. We take the error bar to be the standard deviation of this distribution, *i.e.* the square root of the variance, which we refer to as the standard error (because it gives the standard deviation of the probability distribution for the mean).

The responses and error bars presented throughout the experimental results section are derived from this MLE process. Note that tuning curve amplitudes are arbitrary (replacing $g(t)$ and the s_i with $bg(t)$ and $(1/b)s_i$ would not change performance for any b), so presented tuning curves are all normalized to have peak value 1.

As controls on the MLE process, we show that (1) responses remain stimulus-selective across widely varying overall levels of cortical responsiveness (Supplemental Fig. S1b); (2) orientation tuning curves and spatial frequency tuning curves are largely unchanged by the MLE, (Supplemental Fig. S1d; the MLE process was applied to the stimulus epochs in which preferred direction and preferred spatial frequency were assessed).

We also examined the effects of the MLE fitting on the frequencies of tuning curve oscillations found by the SSM model fits to size tuning curves. The MLE procedure suppresses higher frequencies of oscillation (above about 0.1 cycle/deg) and shrinks error estimates (Supplemental Fig. S1c,e); position-tuning curves behave similarly). We believe this is due to suppression of high-frequency noise in the tuning curves. This noise arises because the random presentation times of a given stimulus may happen to occur at times of lesser excitability or of greater excitability. Since stimuli of different sizes are presented in pseudorandom or-

der, these effects are uncorrelated between different stimulus sizes, including adjacent sizes, and so induce high-frequency noise in the size tuning curve. This noise is suppressed by the use of MLE to factor out the excitability changes. High-frequency noise also arises due to trial-to-trial biological variation in responses at a fixed level of excitability. This noise, and also residual noise from excitability changes, is suppressed by the assumption of smoothness of the tuning curves we used as part of the MLE procedure.

One might worry that the MLE smoothing procedure reduced signal rather than noise. To address this, we show that the smoothing is roughly equivalent to a very limited conventional smoothing. We ran the MLE process without any smoothing constraints ("MLE-unsmoothed"). We then found the Gaussian smoothing filter that, when applied to the size-tuning curves from MLE-unsmoothed, gave the least mean-squared difference across the 76 cells with the size-tuning curves resulting from MLE with smoothing ("MLE-smoothed"). The optimal Gaussian filter had standard deviation 0.99°. The smoothness assumption in MLE gives results very much like smoothing the MLE-unsmoothed tuning curves with this optimal filter. For each cell, we calculated the sum-squared difference between the size tuning curves determined by MLE-smoothed and by this Gaussian-filtering of MLE-unsmoothed, as a percentage of the sum of the squares of the MLE-smoothed tuning curve. This percentage had median 0.0130 (5th and 95th percentiles, 0.0028 and 0.040) and mean±stdev 0.0160±0.0139. Furthermore, the frequencies of SSM fits to MLE-unsmoothed data after (but not before) the Gaussian smoothing are highly correlated to the frequencies found from MLE-smoothed results (Supplemental Fig. S1e). All of these results indicate that adding the smoothness constraint to the MLE yields tuning curves very much like those obtained simply by smoothing MLE-unsmoothed with the Gaussian filter (a conclusion also supported by visual inspection of the curves). The Gaussian filter has a width about 1° and hence represents a very limited degree of smoothing. In addition, if the suppression of frequencies above 0.1 cycle/deg were due to the smoothness constraint eliminating true signal at higher spatial frequencies, we would expect – unless cells had two distinct oscillation frequencies, one above and one below 0.1 cycle/deg – that cells with oscillation frequencies above 0.1 cycle/deg in SSM fits to the MLE-unsmoothed size-tuning data would be brought down only to 0.1 cycle/deg in MLE-smoothed data (*i.e.*, to the highest limiting frequency imposed by smoothing, as judged by the highest frequency seen in the smoothed fits), but this is not the case (Supplemental Fig. S1e).

S1.4.3 Fits of DoG and SSM models to tuning curves

For size tuning, the exponential decay $e^{a_7 x}$ was important for a good fit. We included this term in both DoG and SSM models, so that the only difference between the two models is the periodicity. Results are very similar if the SSM model is compared to a more standard DoG model without the decay term ($a_7 = 0$ in equation for DoG model; Supplemental Fig. S1f).

Fits of SSM and DOG models using Matlab function lsqcurvefit were done subject to constraints established empirically as necessary to avoid poorly-fitting local minima. In units in which the first and last points on the curves had coordinates 10 and 300, respectively, these constraints were: for size-tuning curves, $a_1 \geq 0$, $0 \leq a_3 \leq 1$, $a_4 \geq 0$, $0 \leq a_6 \leq 1$ (SSM)

or $0 \leq a_6 \leq a_3$ (DOG), $-0.1 \leq a_7 \leq 0.1$, $0 \leq a_9 \leq \pi/10$, $-\pi \leq a_{10} \leq \pi$; for position-tuning curves, $a_1 \geq 0$, $-150 \leq a_2 \leq 150$, $a_3 \geq 0$ (DOG) or $a_3 \geq 25$ (SSM), $-150 \leq a_5 \leq 150$ (DOG only), $a_6 \geq 0$ (DOG) or $a_6 \geq 50$ (SSM), $0 \leq a_8 \leq 2\pi/45$, $-\pi \leq a_9 \leq \pi$, $a_{10} \geq 0$. In size-tuning DOG model of Supplemental Fig. S1f-g ($a_7 = 0$), constraints on other parameters are unchanged except $0 \leq a_3 \leq 0.2$. The use of different constraints for the two models may cast doubt on the use of the F-test, which depends on the models being nested, *i.e.* identical after extra parameters in the SSM model are set to zero. Because fits generally became worse (greater summed-squared error) when constraints were relaxed, we think of these constraints as guidance to the error-minimizing algorithm to get it into the best parameter regime, rather than as constraints on the models. The cross-validation analyses provide an independent check on the significance of our results that does not rely on the assumption of nested models.

In Figs. 4C, 4F and Supplemental Fig. S1g, we showed results of a cross-validation analysis comparing the performance of the DoG vs. SSM models. In Supplemental Figs. S1h-i (size-tuning data) and S1j-k (positional-tuning data), we show that there is in general very little difference between the cross-validation fits of the SSM model, using 80% of the data, and the SSM fits using the full data set, indicating that the variance in our SSM estimates is very small.

S1.5 Statistical Methods

S1.5.1 Assumptions of Normality

Most of the statistical tests we used make no assumption that data is normally distributed. There are two exceptions:

- We used a nested F-test to compare SSM and DOG model fits to data. These models are nested: the SSM model contains all the parameters of the DOG model plus additional parameters; setting these additional parameters to zero gives the DOG model. On the assumption that the residuals from the fits are approximately normally distributed, and under the null hypothesis that the simpler hypothesis is correct (*i.e.*, that the extra parameters in the more complex model are all equal to zero), the F-statistic, which is derived from the ratio of the summed-squared errors of the two fits (or more generally of the Ξ^2 of the two fits) should follow an F distribution with degrees of freedom as specified in legend of Supplemental Table S1. We use this test on the plausible assumption of approximately normally distributed residuals, but we also assay the same question with a cross-validation test that requires no such assumption.

- We used the Pearson correlation coefficient to determine whether two sets of data are correlated in Fig. 6K and Supplemental Figs. S9C,D,E and S1e. We use the standard test as to whether r is significantly different from zero, which is to conduct a Student's t-test with degrees of freedom $DF = n - 2$ on the statistic $t = r\sqrt{\frac{n-2}{1-r^2}}$. For an uncorrelated normal distribution, t follows the Student's t distribution with DF degrees of freedom. However, t also approximately follows this distribution for non-normal

distributions if sample size is not very small (Edgell and Noon 1984), *i.e.* ≥ 6, so no assumption of normality is required. Furthermore all but one of our tests show lack of significant correlation by a wide margin, with the two smallest non-significant p-values being 0.17 and 0.31, and the remaining test shows very strong significance ($p = 5.7 \times 10^{-6}$, one of the tests in Supplemental Fig. S1e). These margins combined with the results of Edgell and Noon (1984) alleviate any worries about non-normality.

S1.5.2 Test statistics for statistical tests of Figs. 4C,F, 5F and 7A,E

Here we give test statistics (and repeat p and r values from main text) for statistical tests other than those documented in Supplemental Tables.

- Tests of Fig. 4C,F: tests whether median of difference between sum-squared error of SSM and DOG models on withheld data in cross-validation tests is less than 0, using Wilcoxon signed rank test. Fig. 4C: $p = 1.04 \times 10^{-10}$, SignedRank= 2711; Fig. 4F: $p = 2.4 \times 10^{-7}$, SignedRank= 2347.

- Tests of Fig. 5F for differences between contrast levels in experimentally measured CM preferred SF, using Wilcoxon rank sum (WRS) test. Two middle contrasts not significantly different: $p = 0.68$, ranksum= 2584; low vs. medium: $p < 0.5 \times 10^{-4}$, ranksum= 2818; low vs. high: $p < 10^{-7}$, ranksum= 1789.5; middle vs. high: $p = 0.046$, ranksum= 7137.5.

- Tests of Fig. 7A, Student's t-test for significance of Pearson correlation coefficient (as described in previous subsection) between luminance and CM preferred orientations in model data: E units: $r = 0.098$, $p = 0.33$, $DF = 98$, $t = 0.979$; I units: $r = 0.093$, $p = 0.36$, $DF = 98$, $t = 0.923$.

- Tests of Fig. 7E for differences between high and low contrasts in model circular variance of orientation tuning of surround suppression, using WRS test. All units: $p = 2.6 \times 10^{-9}$, ranksum= 7032; E units: $p = 6.3 \times 10^{-10}$, ranksum= 1555; I units: $p = 0.034$, ranksum= 1990.

S2 Supplemental Text: Mathematical Analysis and Related Theoretical Issues

S2.1 Network Resonant Frequencies

We write the linear model equation in matrix/vector form as

$$\tau_E \mathbf{T} \frac{d}{dt}\mathbf{r}(x) = -\mathbf{r}(x) + \mathbf{W} * \mathbf{r}(x) + c\mathbf{h}(x) \tag{S1}$$

Here, $\mathbf{r}(x) = \begin{pmatrix} r_{\mathrm{E}}(x) \\ r_{\mathrm{I}}(x) \end{pmatrix}$, $\mathbf{h}(x) = \begin{pmatrix} h(x) \\ h(x) \end{pmatrix}$, $\mathbf{W}(x) = \begin{pmatrix} W_{\mathrm{EE}}(x) & -W_{\mathrm{EI}}(x) \\ W_{\mathrm{IE}}(x) & -W_{\mathrm{II}}(x) \end{pmatrix}$, $\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\tau_I}{\tau_E} \end{pmatrix}$, and the $*$ means convolution: $\mathbf{W} * \mathbf{r}(x) \equiv \sum_{x'} \mathbf{W}(x - x')\mathbf{r}(x')$. We will also refer to the full set of connections as the matrix $\mathbf{W} = \begin{pmatrix} \mathbf{W}_{\mathrm{EE}} & -\mathbf{W}_{\mathrm{EI}} \\ \mathbf{W}_{\mathrm{IE}} & -\mathbf{W}_{\mathrm{II}} \end{pmatrix}$ where the submatrix $\mathbf{W}_{\mathrm{XY}}$ has elements $(\mathbf{W}_{\mathrm{XY}})_{\mathrm{ij}} = W_{\mathrm{XY}}(x_i - x_j)$ and x_i and x_j are the i^{th} and j^{th} grid positions respectively.

Here we analyze linear network behavior, and in particular derive expressions for the excitatory and inhibitory resonant frequencies. We simplify by assuming either periodic boundary conditions for the network or that the network stretches to $\pm\infty$. For generality we will let the inputs to E and I have different magnitudes: $h_{\mathrm{E}}(x) = \alpha h(x)$, $h_{\mathrm{I}}(x) = h(x)$.

Because the connections between two units at locations x and x' depend only on the distance between the two neurons – that is, they are translation-invariant, the same at any position – the convolution in Eq. S1 is turned into a multiplication by Fourier transformation. For any function $f(x)$, we let $\tilde{f}(k)$ be its Fourier transform at what we will call spatial frequency k (on a grid of N units, k is an integer corresponding to spatial frequency k cycles per N units). We normalize the Fourier transform so that the Fourier transform of $f * g(x)$ is $\tilde{f}(k)\tilde{g}(k)$.[1] Then after Fourier transform, Eq. S1 becomes

$$\tau_E \mathbf{T} \frac{d}{dt}\tilde{\mathbf{r}}(k) = -\tilde{\mathbf{r}}(k) + \tilde{\mathbf{W}}(k)\tilde{\mathbf{r}}(k) + c\tilde{\mathbf{h}}(k) \tag{S2}$$

where $\tilde{\mathbf{r}}(k) = \begin{pmatrix} \tilde{r}_{\mathrm{E}}(k) \\ \tilde{r}_{\mathrm{I}}(k) \end{pmatrix}$, $\tilde{\mathbf{h}}(k) = \tilde{h}(k) \begin{pmatrix} \alpha \\ 1 \end{pmatrix}$, and $\tilde{\mathbf{W}} = \begin{pmatrix} \tilde{W}_{\mathrm{EE}}(k) & -\tilde{W}_{\mathrm{EI}}(k) \\ \tilde{W}_{\mathrm{IE}}(k) & -\tilde{W}_{\mathrm{II}}(k) \end{pmatrix}$. Note that the E and I activity patterns at a given spatial frequency influence one another's time evolution, but evolve independently of the activity patterns at every other spatial frequency. This follows from the linearity of the network and the translation-invariance of the weight functions.

From Eq. S2, the network fixed point, where $\frac{d}{dt}\tilde{\mathbf{r}} = 0$, satisfies $\tilde{\mathbf{r}} = (\mathbf{1} - \tilde{\mathbf{W}}(k))^{-1} c\tilde{\mathbf{h}}(k)$,

[1] See footnote 3.

where $\mathbf{1}$ is the 2-dimensional identity matrix. We can rewrite this as

$$\tilde{r}_{\mathrm{E}}(k) = \left(\frac{\left(1 + \tilde{W}_{\mathrm{II}}(k)\right)\alpha - \tilde{W}_{\mathrm{EI}}(k)}{\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right)} \right) \tilde{h}(k) \equiv \mathcal{L}_{\mathrm{E}}(k)\tilde{h}(k) \tag{S3}$$

$$\tilde{r}_{\mathrm{I}}(k) = \left(\frac{1 - \tilde{W}_{\mathrm{EE}}(k) + \alpha\tilde{W}_{\mathrm{IE}}(k)}{\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right)} \right) \tilde{h}(k) \equiv \mathcal{L}_{\mathrm{I}}(k)\tilde{h}(k) \tag{S4}$$

Here, $\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right) = \tilde{W}_{\mathrm{EI}}(k)\tilde{W}_{\mathrm{IE}}(k) - (\tilde{W}_{\mathrm{II}}(k) + 1)(\tilde{W}_{\mathrm{EE}}(k) - 1)$ is the determinant of $(\mathbf{1} - \tilde{\mathbf{W}}(k))$. Stability of the network dynamics requires that $\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right) > 0$ for all k.[2]

Equations S3-S4 show how selective amplification of spatially periodic patterns of activity arise in response to non-periodic input. The terms multiplying $\tilde{h}(k)$, which we have given the names $\mathcal{L}_{\mathrm{E}}(k)$ and $\mathcal{L}_{\mathrm{I}}(k)$, act as linear filters on the input to produce the steady-state responses. If the input is not spatially periodic, *i.e.* has no peaks at $k \neq 0$ (non-DC peaks), spatial periodicity can nonetheless arise in the firing rates if the filters have non-DC peaks within the bandwidth of $\tilde{h}(k)$, inducing nearby non-DC peaks in $r_{\mathrm{E}}(k)$ or $r_{\mathrm{I}}(k)$. By solving for these filter maxima we can find the conditions under which the network will demonstrate spatial periodicity for non-periodic input.

To find the filter maxima, we will solve for the roots of the first derivative of the network filters with respect to k. The denominator of these derivatives is $\mathrm{Det}(\mathbf{1} - \mathbf{W}(k))^2$. Since network stability requires that $\mathrm{Det}(\mathbf{1} - \mathbf{W}(k)) > 0$ for all k, this denominator remains nonzero. Thus, we can simply determine when the numerator of the derivative becomes zero.

S2.1.1 Inhibitory resonant frequency

Defining $\partial_k f(k) \equiv \frac{df(k)}{dk}$, the numerator of $\partial_k \mathcal{L}_{\mathrm{I}}(k)$ is

$$\begin{aligned} &\left(\tilde{W}_{\mathrm{EI}}(k) - \alpha\left(1 + \tilde{W}_{\mathrm{II}}(k)\right)\right)\left(\partial_k\tilde{W}_{\mathrm{IE}}(k)\left(\tilde{W}_{\mathrm{EE}}(k) - 1\right) - \partial_k\tilde{W}_{\mathrm{EE}}(k)\tilde{W}_{\mathrm{IE}}(k)\right) \\ &+ \left(1 - \tilde{W}_{\mathrm{EE}}(k) + \alpha\tilde{W}_{\mathrm{IE}}(k)\right)\left(-\partial_k\tilde{W}_{\mathrm{EI}}(k)\tilde{W}_{\mathrm{IE}}(k) + \partial_k\tilde{W}_{\mathrm{II}}(k)\left(\tilde{W}_{\mathrm{EE}}(k) - 1\right)\right) \end{aligned} \tag{S5}$$

We work in the limit in which inhibitory connectivity is very localized compared to excitatory connectivity, meaning that $\tilde{W}_{\mathrm{EI}}(k)$ and $\tilde{W}_{\mathrm{II}}(k)$ are very broad in k relative to $\tilde{W}_{\mathrm{IE}}(k)$ and $\tilde{W}_{\mathrm{EE}}(k)$. As a result, the derivatives of the two I projections are very small relative to the derivatives of the two E projections, so that the second line in Eq. S5 is very small relative to the first line. We restrict to the limit of completely localized inhibitory connectivity, as in our simulations, in which $\tilde{W}_{\mathrm{EI}}(k) \equiv \tilde{W}_{\mathrm{EI}}$ and $\tilde{W}_{\mathrm{II}}(k) \equiv \tilde{W}_{\mathrm{II}}$ are constants, independent

[2]The network is stable iff both eigenvalues of $\mathbf{1} - \tilde{\mathbf{W}}(k)$ have negative real part for all k (these eigenvalues across all k are the eigenvalues of $\mathbf{1}_{D_{\mathbf{W}}} - \mathbf{W}$, where $\mathbf{1}_{D_{\mathbf{W}}}$ represents the identity matrix of the same dimensionality as $\mathbf{W}$). The eigenvalues of a 2×2 matrix $\mathbf{M}$ have negative real part iff $\mathrm{Det}\,\mathbf{M} > 0$ and $\mathrm{Tr}\,\mathbf{M} < 0$, where $\mathrm{Tr}\,\mathbf{M}$ is the trace of $\mathbf{M}$, hence a necessary condition for stability is that $\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right) > 0$ for all k.

of k, and so their derivatives are zero; one could consider corrections for finite width of the inhibitory projections, but we do not pursue that here.

Then, the condition for a zero of the derivative is that the first line of Eq. S5 is zero. Because $W_{\mathrm{EE}}(x)$ and $W_{\mathrm{IE}}(x)$ are even functions of x, both $\tilde{W}_{\mathrm{EE}}(k)$ and $\tilde{W}_{\mathrm{IE}}(k)$ are even functions of k, so both their derivatives are zero at $k = 0$. Thus, one zero is at $k = 0$. Assuming that it is not the case that $\tilde{W}_{\mathrm{EI}} = \alpha\left(1 + \tilde{W}_{\mathrm{II}}\right)$ – a condition that, if true, would mean that $r_{\mathrm{E}}(x) \equiv 0$ for all x by Eq. S3 – the only other zeros occur for k such that

$$\tilde{W}_{\mathrm{EE}}(k) = \frac{\partial_k \tilde{W}_{\mathrm{EE}}(k)\tilde{W}_{\mathrm{IE}}(k)}{\partial_k \tilde{W}_{\mathrm{IE}}(k)} + 1 \tag{S6}$$

The first term on the right side of Eq. S6 will be positive for all k if $\partial_k \tilde{W}_{\mathrm{EE}}(k)$ and $\partial_k \tilde{W}_{\mathrm{IE}}(k)$ always have the same sign and $\tilde{W}_{\mathrm{IE}}(k) > 0$ for all k. This will be true for Gaussian connectivity functions on an infinite continuum, and should be true more generally for even connectivity functions that decay monotonically with distance sufficiently smoothly. So long as this is true, a solution of equation S6, meaning a non-zero peak for the network filter, can only occur for k with $\tilde{W}_{\mathrm{EE}}(k) > 1$, which – along with overall network stability – is the condition for the network to be an ISN. That is, given localized inhibitory projections and sufficiently smoothly decaying excitatory projections, a necessary – but not sufficient – condition for the network to show spatially periodic responses to spatially non-periodic inputs is that the network be an ISN.

We now use our choice of Gaussian excitatory connectivity functions, $\mathbf{W}_{aE}(x) = J_{aE} e^{-\frac{x^2}{2\sigma_{aE}^2}}$, and assume we are operating on an infinite continuum rather than a discrete finite grid – a continuum of units along an infinite, continuous position x and hence an infinite continuum of spatial frequencies k. Note that this means that $\mathbf{W}_{aE}(x)$ has dimensions of 1/length (so that its spatial integral with rates gives a rate); in practice, this means we must replace the J_{aE}'s from our discrete connectivity on a grid of width Δx, with $J_{aE}/\Delta x$ (so that the discrete convolution $\sum'_x \Delta x \frac{J_{aE}}{\Delta x} e^{-\frac{(x-x')^2}{2\sigma_{aE}^2}} r_E(x')$ approaches an integral in the continuum limit). The J_{aE}'s in the formulae below should all be understood numerically to correspond to this $J_{aE}/\Delta x$ and have corresponding dimension 1/length. The I connections W_{aI} apply only to the same grid location, and the continuum equivalent is that the weights are $W_{aI}\delta(x-x')$ (so that continuum and discrete operations give the same result: $\int dx' W_{aI}\delta(x-x')r_I(x') = W_{aI}r_I(x)$). This means that the Fourier transform is simply $\tilde{W}_{aI} = W_{aI}$, which is dimensionless.

The continuum limit allows us to compute the Fourier transform of the Gaussian connectivity[3], which produces analytic results that agree very well with the simulated results on

[3]We use the convention $\tilde{f}(k) = \int dx\, e^{ikx} f(x)$, $f(x) = \frac{1}{2\pi}\int dk\, e^{-ikx}\tilde{f}(k)$. With this convention, the Fourier transform of $f * g(x)$ is $\tilde{f}(k)\tilde{g}(k)$ (see text that calls footnote 1) and the Fourier transform of the weights on the continuum is $\tilde{W}_{aE}(k) = J_{aE}\sigma_{aE}\sqrt{2\pi}e^{-\frac{k^2\sigma_{aE}^2}{2}}$. If instead we use $\tilde{f}(k) = \frac{1}{\kappa}\int dx\, e^{ikx} f(x)$, $f(x) = \frac{\kappa}{2\pi}\int dk\, e^{-ikx}\tilde{f}(k)$ for some constant κ (*e.g.*, $\kappa = \sqrt{2\pi}$ is commonly used), then the convolution is multiplied by κ while the Fourier transform of the Gaussian is divided by κ. The result is that all the factors $\tilde{W}_{XY}$ in Eqs. S3-S4 are both divided by κ and multiplied by κ (the latter because they all result from convolutions), so that their values and hence the results of the analysis are unchanged.

the finite grid (*e.g.*, see Fig. 5B-C of the main text and Supplemental Fig. S2a,b). We thus determine the k that solves Eq. S6, which we call the resonant frequency of inhibition, k_I:

$$k_I = \frac{\sqrt{2\ln\left(J_{EE}\sigma_{EE}\sqrt{2\pi}\left(1-\frac{\sigma_{EE}^2}{\sigma_{IE}^2}\right)\right)}}{\sigma_{EE}} \tag{S7}$$

Spatial periodicity arises iff (if and only if) the solution k_I is real, which in turn occurs iff the argument of the logarithm is greater than 1:

$$J_{EE}\sigma_{EE}\sqrt{2\pi}\left(1-\frac{\sigma_{EE}^2}{\sigma_{IE}^2}\right) > 1 \tag{S8}$$

This in turn implies that $\sigma_{IE} > \sigma_{EE}$ and that $J_{EE}\sigma_{EE}\sqrt{2\pi} > 1$. $J_{EE}\sigma_{EE}\sqrt{2\pi} = \tilde{W}_{\mathrm{EE}}(0)$ is the maximum eigenvalue of the operation of convolving with $W_{\mathrm{EE}}(x)$ (see footnote 3), so that the latter condition means that the excitatory subnetwork by itself is unstable; that is, the network, assuming it is stable, is an ISN.

In sum, for a linear network, given sufficiently localized inhibitory connectivity and excitatory projections that are even and monotonically decay from zero sufficiently smoothly, then spatial periodicity in the inhibitory network in response to non-periodic input can only arise if (1) the network is an ISN and (2) $\sigma_{IE} > \sigma_{EE}$. The periodicity then will arise precisely when the condition in Eq. S8 is satisfied.

In addition, note that, in Eq. S7, the dependence of k_I on the items inside the logarithm is weak, so that the strongest dependence of k_I is that it is proportional to $1/\sigma_{EE}$. That is, longer-range excitatory projections lead to longer inhibitory wavelengths. In addition, we see that k_I increases more weakly with increasing J_{EE} and with increasing σ_{IE}/σ_{EE}, meaning that excitatory projections that are stronger or more widespread to I vs. E lead to shorter inhibitory wavelengths.

S2.1.2 Excitatory resonant frequency

The numerator of $\partial_k \mathcal{L}_{\mathrm{E}}(k)$ is

$$\begin{aligned}&\left(\tilde{W}_{\mathrm{EI}}(k) - \alpha\left(1+\tilde{W}_{\mathrm{II}}(k)\right)\right)\left(\partial_k\tilde{W}_{\mathrm{IE}}(k)\tilde{W}_{\mathrm{EI}}(k) - \partial_k\tilde{W}_{\mathrm{EE}}(k)\left(1+\tilde{W}_{\mathrm{II}}(k)\right)\right)\\ &+\left(1-\tilde{W}_{\mathrm{EE}}(k)+\alpha\tilde{W}_{\mathrm{IE}}(k)\right)\left(\partial_k\tilde{W}_{\mathrm{II}}(k)\tilde{W}_{\mathrm{EI}}(k) - \partial_k\tilde{W}_{\mathrm{EI}}(k)\left(1+\tilde{W}_{\mathrm{II}}(k)\right)\right)\end{aligned} \tag{S9}$$

Again, we take $\tilde{W}_{\mathrm{EI}}(k) \equiv W_{\mathrm{EI}}$ and $\tilde{W}_{\mathrm{II}}(k) \equiv W_{\mathrm{II}}$ so that the second line is zero, and find, assuming that $W_{\mathrm{EI}} \neq \alpha\,(1+W_{\mathrm{II}})$, that the zeros of $\partial_k \mathcal{L}_{\mathrm{E}}(k)$ are either at $k=0$ or where

$$W_{\mathrm{EI}} = \frac{\partial_k\tilde{W}_{\mathrm{EE}}(k)\,(1+W_{\mathrm{II}})}{\partial_k\tilde{W}_{\mathrm{IE}}(k)} \tag{S10}$$

Using the Gaussian connectivity functions and the assumption of an infinite continuum, we find that the non-DC zeros of $\partial_k \mathcal{L}_{\mathrm{E}}(k)$ occur at the excitatory resonant frequency k_E:

$$k_E = \frac{\sqrt{\frac{2}{1-(\sigma_{EE}^2/\sigma_{IE}^2)} \ln\left(\frac{W_{EI} J_{IE} \sigma_{IE}^3}{J_{EE}(1+W_{II})\sigma_{EE}^3}\right)}}{\sigma_{IE}} \tag{S11}$$

For this to be real and nonzero, and thus for excitatory spatial periodicity to exist, we either must have $\sigma_{IE} > \sigma_{EE}$ and the argument of the logarithm > 1, or else $\sigma_{IE} < \sigma_{EE}$ and the argument of the logarithm < 1

Unlike inhibition, the existence of a real k_E does not require the network to operate within the ISN regime. However, we now show that, assuming a stable network, if k_I is real and positive, then k_E is also real and positive, and furthermore $k_I < k_E$. That is, if the inhibitory network shows spatial periodicity, the excitatory network does also, with a shorter wavelength. We use the facts, which follow from k_I being real and positive, that (1) $\sigma_{IE} > \sigma_{EE}$ and (2) Eq. S8 is satisfied.

Since $\sigma_{IE} > \sigma_{EE}$, k_E is real and positive iff the argument of the logarithm is > 1:

$$W_{EI} J_{IE} \sigma_{IE}^3 > J_{EE}\left(1 + W_{II}\right)\sigma_{EE}^3 \tag{S12}$$

As noted above, a necessary condition for network stability is that $\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right) > 0$ for all k. By solving for the extrema of this determinant (the values of k for which its derivative with respect to k is 0) and imposing the condition that the determinant at these extrema is > 0, we obtain the following requirement for stability:

$$\left(\frac{W_{EI} J_{IE} \sigma_{IE}^3}{J_{EE}\left(1 + W_{II}\right)\sigma_{EE}^3}\right)^{\left(\frac{\sigma_{EE}^2}{\sigma_{IE}^2 - \sigma_{EE}^2}\right)} > J_{EE}\sigma_{EE}\sqrt{2\pi}\left(\frac{\sigma_{IE}^2 - \sigma_{EE}^2}{\sigma_{IE}^2}\right) \tag{S13}$$

From Eq. S8, the right side of Eq. S13 is > 1. Since $\sigma_{IE} > \sigma_{EE}$, the exponent on the left side of Eq. S13 is > 0. Thus, a necessary condition for stability is that $\left(\frac{W_{EI} J_{IE} \sigma_{IE}^3}{J_{EE}(1+W_{II})\sigma_{EE}^3}\right) > 1$, which also means that the extremum is in fact a minimum.[4] This is precisely the condition of Eq. S12 that guarantees that the argument of the logarithm in Eq. S11 for k_E is > 1 and thus that the excitatory population has a real resonant frequency. Furthermore, a bit of algebra shows that the condition $k_I < k_E$ then translates precisely into the condition of Eq. S13 (which also implies that stability is lost precisely when k_I becomes greater than k_E). In sum, if the network is stable and the inhibitory network has a real, positive resonant frequency, then the excitatory network has a higher real, positive resonant frequency.

S2.1.3 Understanding spatially periodic activity in the model from the resonant frequencies

From equations S11 and S7, we see that, with increases in $\frac{\sigma_{EI}}{\sigma_{EE}}$, k_I increases, while k_E may increase or decrease. We test our formulae (derived from the continuum limit) against the

[4] The 2nd derivative at the extremum is positive iff $\left(\frac{W_{EI} J_{IE} \sigma_{IE}^3}{J_{EE}(1+W_{II})\sigma_{EE}^3}\right) > 1$.

result of simulations of our linear model on the grid, varying σ_{IE} while keeping all other parameters constant at the values given previously, and compare the preferred contrast modulation (CM) spatial frequency to the analytically predicted resonant frequencies (Supplemental Fig. S2a,b, and see also Fig. 5B-C of main text). The match is essentially perfect. The analytic solutions reveal that for σ_{IE} sufficiently wide, the network will actually be unstable (red area on the plot). This arises when the E→I connections become too broad for feedback inhibition to stabilize higher frequencies at which E→E connections are unstable, and as already noted occurs precisely where the resonant frequency of inhibition becomes greater than that of excitation.

This analysis also reveals why stimuli with sharp-edged contrast profiles produce multi-peaked length-tuning curves but those with Gaussian contrast profiles do not (Supplemental Figs. S7,S8). In the Fourier domain, a step-function is transformed into a sinc function $\frac{\sin(kl)}{k}$, where l is the width of the step function, and more generally sharp-edged stimulus profiles will have varying degrees of "ringing" in their Fourier transforms. This ringing will induce sign changes in the Fourier representation whose positions depend on stimulus length, resulting in a periodic flip in the phase of the resonant frequency with increasing stimulus size. These phase changes give rise to the periodic rise and fall of activity of the unit at the stimulus center with stimulus length, *i.e.* to periodic length-tuning curves. Gaussian stimuli, on the other hand, are simply transformed into Gaussians curves over frequency with inversely proportional width (narrow stimuli will have broad Fourier representations, and vice-versa). For sufficiently wide Gaussian stimuli, the power spectrum goes essentially to zero at a frequency below that of the peaks in the network response filters (k_E and k_I). Then the network's resonant frequencies receive sufficiently little input that the power in the firing-rate curves follows the power in the stimulus, falling monotonically with frequency. As a consequence, the firing rates show no spatial periodicity. For narrower Gaussian inputs, the stimulus may be wide enough in Fourier space to have non-zero power at the network's preferred frequencies, and in response the network does have spatially-periodic activity – see the yellow curves in the left half of Supplemental Fig. S7a. However, this does not produce periodic length tuning curves because the phase does not change with increasing stimulus length (and also the spatial periodicity largely disappears at relatively small stimulus sizes).

As the length l of a step-function stimulus is increased, there can be stimulus lengths where $\frac{\sin(kl)}{k} = 0$ for one of the peak frequencies (k_E or k_I). This will be observed across the population as a relative weakening in the amplitude of the corresponding spatial oscillation (though not a complete absence of periodicity, as spatial frequencies near the peak are also relatively amplified).

S2.2 Critical frequency, and an experimental test

S2.2.1 The critical frequency

A focus of previous theoretical work on the ISN regime has been on the "paradoxical" decrease in steady-state inhibitory firing rates that results from an increase in the external input to inhibitory units (Ozeki et al. 2009, Tsodyks et al. 1997). With the addition of a

spatial dimension in our current model, this paradoxical response depends on the stimulus spatial frequency. This can be understood from the fixed-point firing-rate solutions in equations S3 and S4. The E and I activities at each frequency k satisfy equations of the same form as the linear two-population models (one E population, one I population) studied in (Ozeki et al. 2009, Tsodyks et al. 1997). Hence the analysis of those papers applies separately to the activity at each frequency. If $\tilde{W}_{\mathrm{EE}}(k) > 1$, then the excitatory subnetwork at frequency k is unstable, so that, assuming the dynamics at frequency k are stabilized by feedback inhibition, the activity at frequency k behaves as an ISN. This includes the paradoxical response to input to inhibitory units: if external input is purely to inhibitory cells, $\tilde{h}_{\mathrm{E}}(k) \equiv 0$ or $\alpha = 0$, the fixed-point equations become

$$\tilde{r}_{\mathrm{E}}(k) = -\frac{1}{\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right)} \tilde{W}_{\mathrm{EI}}(k)\tilde{h}_{\mathrm{I}}(k) \tag{S14}$$

$$\tilde{r}_{\mathrm{I}}(k) = \frac{1}{\mathrm{Det}\left(\mathbf{1} - \tilde{\mathbf{W}}(k)\right)} \left(1 - \tilde{W}_{\mathrm{EE}}(k)\right) \tilde{h}_{\mathrm{I}}(k) \tag{S15}$$

Because the network is stable, $\det(1 - \mathbf{W}(k)) > 0$. Hence, as expected, $\tilde{r}_{\mathrm{E}}(k)$ is always decreased by an increase in input $\tilde{h}_{\mathrm{I}}(k)$ to the inhibitory units. However, $\tilde{r}_{\mathrm{I}}(k)$ is increased or decreased by such a change in input according to whether $\tilde{W}_{\mathrm{EE}}(k) < 1$ or $\tilde{W}_{\mathrm{EE}}(k) > 1$, respectively. The decrease in $\tilde{r}_{\mathrm{I}}(k)$ in response to increased input to inhibitory units for $\tilde{W}_{\mathrm{EE}}(k) > 1$ is the "paradoxical" ISN response.

Because the strength of connectivity $W_{\mathrm{EE}}(x)$ decreases monotonically and smoothly to zero as a function of distance $|x|$, in the Fourier domain the connection strength $\tilde{W}_{\mathrm{EE}}(k)$ also decreases monotonically as a function of spatial frequency $|k|$. Thus, if the network is an ISN, meaning that at least some values of $\tilde{W}_{\mathrm{EE}}(k)$ are > 1, then with increasing $|k|$ $\tilde{W}_{\mathrm{EE}}(k)$ will fall below 1 at some critical frequency $|k| = k_c$. This means that frequencies below k_c will show the paradoxical ISN behavior, while frequencies above k_c will not. The critical frequency k_c is defined by $J_{EE}\sigma_{EE}\sqrt{2\pi}e^{-\left(k_c\sigma_{EE}/\sqrt{2}\right)^2} = 1$, that is:

$$k_c = \frac{\sqrt{2\ln\left(J_{EE}\sigma_{EE}\sqrt{2\pi}\right)}}{\sigma_{EE}} \tag{S16}$$

k_c scales approximately as $\frac{1}{\sigma_{EE}}$, since the logarithmic dependence is weak. Note that if $J_{EE}\sigma_{EE}\sqrt{2\pi} < 1$, meaning that no frequency has $\tilde{W}_{\mathrm{EE}}(k) > 1$ (the network is always a non-ISN), then k_c is imaginary, meaning that there is no transition from ISN to non-ISN behavior at any real frequency.

We note from examining Eq. S7 that $k_I = \sqrt{k_c^2 + \frac{2}{\sigma_{EE}^2}\ln\left(1 - \frac{\sigma_{EE}^2}{\sigma_{IE}^2}\right)}$. Since the second term in the squareroot is negative, $k_I < k_c$. This shows again that the resonant frequency of the inhibitory elements must fall within the range of frequencies that are inhibition-stabilized.

S2.2.2 Experimental Prediction: Ideal Case.

The existence of this critical frequency yields an interesting experimental prediction, illustrated in Fig. 8 of the main text. Here and in the next section we provide the theory behind this prediction, based on the results just obtained in section S2.2.1.

We imagine we have experimentally isolated an ISN of E and I cells, for example a patch of sufficiently-stimulated layer 2/3 cells in a piece of cortex. There are many practical questions as to when an ISN has been isolated, for example: can a patch of upper layer cells be treated effectively as a closed network except for steady external input arising from other layers and other patches of cortex, *i.e.* can we neglect feedback from these other regions in response to changes in activity in the patch? What are the roles of multiple subtypes of inhibitory neurons, and which inhibitory neurons are the relevant ones for providing feedback inhibition to stabilize the otherwise unstable excitatory subnetwork? We ignore these questions, simply imagining that feedback from outside the patch is negligible, that the recurrence among the E cells in the patch is strong enough to be unstable by itself, and that we are manipulating I neurons in the patch that provide feedback inhibition that prevents this instability.

The prediction is based on selectively stimulating I cells in a manner that is sinusoidally modulated spatially across cortex. We assume this is achieved with light stimulation of channelrhodopsin (ChR) expressed only in the I cells. There must be a baseline level of stimulating light about which the sinusoidal modulation occurs. We assume the network is in the ISN regime in response to spatially unmodulated light stimulation (changes in the baseline light level). The sinusoidal modulation must drift or contrast modulate, with a temporal frequency low enough that the network has time to roughly equilibrate and so be in a near-steady-state at each moment in the cycle. We can then apply the steady-state picture of ISN and non-ISN behavior to each moment in the cycle. E- and I-cell firing rates will periodically modulate at the temporal frequency of the stimulus, and we consider the relative phases of their modulation.

Then the prediction is that, for light stimulus of increasing spatial frequencies, the responses of E and I cells undergo a roughly 180^o transition in their relative phases at the critical spatial frequency k_c (Eq. S16). We illustrate this prediction in our linear model, stimulating only the inhibitory cells in the network with a "photostimulus" with a defined spatial frequency (Supplemental Fig. S2c). Stimuli with spatial frequency below k_c drive activity that operates as an ISN, and so E and I firing rates are modulated in phase with each other and out of phase with the input (Supplemental Fig. S2c, bottom left). This is the spatial analogue of the paradoxical response described by Ozeki et al. (2009). Stimuli with spatial frequency above k_c evoke non-ISN activity, and so E and I move out of phase with each other, while I moves in phase with the input (Supplemental Fig. S2c, bottom right).

As shown in Fig. 8 of the main text, exactly the same effect holds in the more realistic, nonlinear, two-dimensional model with sparse random connectivity, with one significant change. Because of the expansive input-output nonlinearity, the paradoxical response depends not only on the I-cell network being driven at a spatial frequency below the critical frequency, but also on the combination of sensory and baseline input being sufficient to drive the network into the ISN regime.

S2.2.3 Experimental Prediction: More Realistic Case.

Even given the assumption that we have isolated a patch of cells that operates as an ISN, there are other practical problems with the proposed experiment. First, only a fraction of inhibitory neurons will be activated, *i.e.* ChR will be incompletely expressed. We assume, due to the high connection probabilities from I to E neurons in cortex (Fino and Yuste 2011, Packer and Yuste 2011), that this fraction suppresses firing across the entire E network. Thus, the resulting reduction in E→I input will be spread across all the I neurons, both those that were stimulated and those that were not. The reduction in network excitation to a stimulated I neuron may then be less than the input that was directly added to it by the stimulation, so that stimulated I neurons will respond in phase to the stimulus. Unstimulated I neurons will respond in phase with the E cells, at the opposite phase to the stimulus. Unfortunately, these behaviors also characterize the non-paradoxical response expected in a non-ISN, and so no change in this behavior will be seen as k_c is crossed. Second, there are multiple subtypes of inhibitory neuron, and in the case of parvalbumin-containing (PV) and somatostatin-containing (SOM) interneurons in rodents, the two types can show opposite responses to stimuli that suppress E cells (Adesnik et al. 2012) (further discussed in main text). The result may then depend on which class of inhibitory cells is studied, where "class" refers to both subtype and whether or not they express ChR.

These problems can be overcome and the paradoxical response and the jump in phase as k_c is crossed can be robustly studied by applying the prediction to the relative phases of the net inhibitory current and the net excitatory current received by E cells, rather than to the firing of E and I cells. Regardless of the details of varying firing across different classes of I cells, this net current received should follow the predicted ISN (low-frequency, sufficient contrast) and non-ISN (otherwise) phase relationships. To understand why, it is necessary to understand the mechanism of the paradoxical ISN effect (Ozeki et al. 2009).

Suppose that the network is at a steady state of firing rates in response to steady external inputs. The fact that recurrent excitation is strong enough to be unstable by itself means that, if E cells lower their firing rates from the previous steady-state levels, this causes withdrawal of *too much* recurrent E→E input – so much that, absent other changes, the E cell firing rates would fall still further rather than moving back to the fixed point (and similarly, if E cells raised their rates, this would cause recruitment of so much recurrent E→E input that, absent other changes, the E cell firing rates would rise still higher). In a stable network in which the external input has not changed, feedback inhibition dynamically responds to a transient decrease in E-cell firing and restores the network to the fixed point (the lowering of E firing rates causes a lowering of feedback inhibition that exceeds the loss of recurrent excitation, so E cells move back up to their fixed point levels, bringing inhibition along with them). However, when the lowering of E-cell firing is caused by an increase of external input to I cells, so that in the new steady state E-cell firing remains lower than previously, then in this new steady state the E-cells must be receiving less inhibition than previously, to compensate for the excessive loss of recurrent excitation. No matter what the detailed dynamics of the various I-cell populations may be, only such a net decrease of inhibition can stabilize the new lower firing rates given the excessive loss of recurrent excitation (they

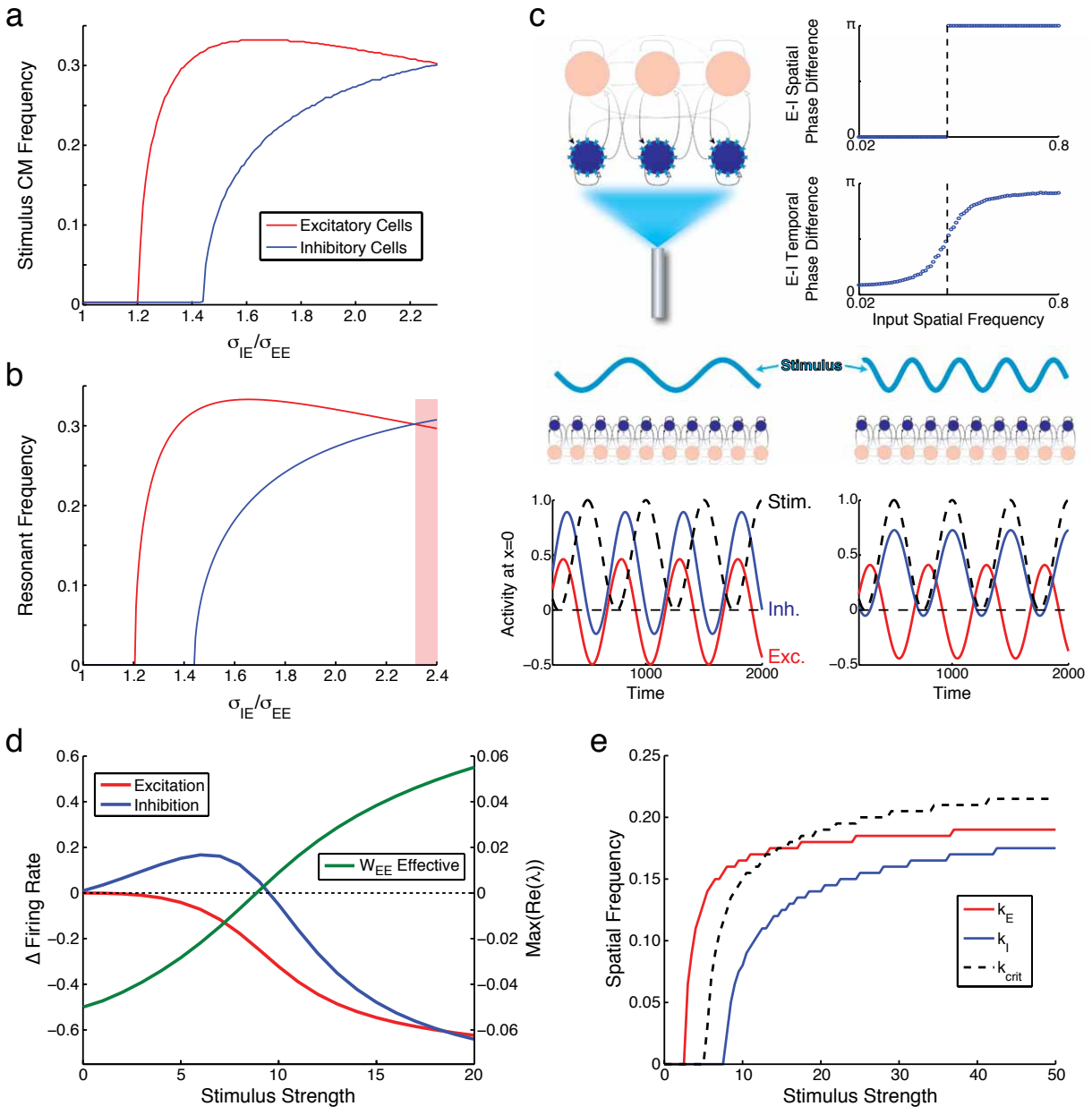


Figure S2:

Resonant frequencies and model linearization (related to Supplement S2.1-S2.3, and main text discussions of Figs. 3,5,8).

a-c: Resonant frequencies in the linear model of Fig. 2. **a,b**: Match of preferred contrast-modulation (CM) spatial frequencies (**a**) to analytically predicted resonant frequencies (**b**, computed from Eqs. S11 and S7), for E (red) and I (blue) units, vs. σ_{IE}/σ_{EE}. We vary σ_{IE} with all other parameters of the linear model fixed. In the red shaded region to the right in **(b)**, where $k_E < k_I$, the network is unstable. **c**: The ISN model predicts a 180° phase shift in the relative responses of E and I units to direct input to I cells at a critical input spatial frequency (as shown for nonlinear 2D model in Fig. 8). Top left: I cells in the network are stimulated with a photostimulus with a defined spatial frequency. The stimulus drifts at 2 Hz. Bottom: The low spatial frequency stimulus on the (continued on next page)

Figure S2:
(Continued) left has spatial frequency 0.2 cycles/degree. It drives E (red) and I (blue) cells at center of grid ($x = 0$) roughly in phase with one another, roughly out of phase with the stimulus they receive (black dashed). The higher spatial frequency stimulus on the right is at 0.5 cycles/degree. Now I cell fires roughly in phase with the stimulus and out of phase with E cell. Top right: summary plots of relative phases of E and I firing rates as a function of photostimulus spatial frequency. The vertical dashed lines are the analytically calculated critical frequency (Eq. S16). Top: phase difference between E and I activities across space at the spatial frequency of the stimulus. Bottom: phase differences in time between illustrated firing rates of E and I units at $x = 0$. **d-e**: Linearization of the nonlinear model of Fig. 3. **d**: Both E and I units are driven by a small (1^o diameter) stimulus of strength c (x-axis) centered at $x = 0$, and linearization is performed around the resulting steady state. Curves show the change in firing rates (left y-axis) of the E (red) and I (blue) units at $x = 0$ induced by giving additional input of strength 1 to stimulated I cells 1; and the maximum real part of the eigenvalues of $\hat{\mathbf{W}}_{\mathrm{EE}}$ (green, right axis). **e**: Resonant frequency increases with increasing stimulus strength. The resonant spatial frequencies of both the excitatory and inhibitory populations are calculated from the Jacobian matrix at the fixed point of response to a uniform stimulus of the given strength.

could also be stabilized if E-cells received a new source of external excitation sufficient to compensate the loss of recurrent excitation, but we have assumed that the only change to external inputs is in the input to I cells).

On the other hand, in the non-ISN regime, a drop in E firing from the old steady-state levels causes too little withdrawal of E→E input to allow the E firing rates to remain so low, so that these firing rates would bounce back up to the old steady-state levels absent other changes. The lower firing rates at the new steady state in the non-ISN regime must therefore be stabilized by an *increase* in the inhibition the E cells receive (or else by withdrawal of external excitatory input, which does not occur).

In sum, for stimuli below k_c (ISN regime), both the inhibition and the excitation received by E cells will decrease when the external input to I cells is increased, while for stimuli above k_c (non-ISN regime), the inhibition they receive will increase while the excitation they receive will decrease. Thus, for low-temporal-frequency stimuli to I cells, the relative phase of the inhibition and the excitation received by E cells should jump by 180^o as the spatial frequency of the stimulus crosses k_c. This conclusion is robust to the details of the I-cell network.

S2.3 Linearization of the Nonlinear Model

With the addition of an expansive nonlinearity, we predicted that our network would transition from non-ISN to ISN regimes with increasing input strength. This is because the effective synaptic strength increases with increasing input strength, until at some point the effective $E \to E$ connections become strong enough to be unstable in the absence of feedback inhibition. The effective synaptic strength – the change in postsynaptic firing rate for a

given change in presynaptic firing rate – increases because it is the product of the biophysical synaptic strength (*e.g.*, the current injected in the postsynaptic cell) and the postsynaptic gain (the change in postsynaptic firing rate for a given current injection). Given the expansive nonlinearity, postsynaptic gain increases with postsynaptic firing rate and thus with input strength.

This input dependence of the dynamic regime can be understood by linearizing the dynamics around the fixed-point firing rates. We let $\mathbf{r}(x) = \mathbf{r}^{\mathrm{FP}}(x) + \delta\mathbf{r}(x)$ where $\mathbf{r}^{\mathrm{FP}}(x) \equiv \begin{pmatrix} r_E^{\mathrm{FP}}(x) \\ r_I^{\mathrm{FP}}(x) \end{pmatrix}$ is the vector of fixed-point firing rates and $\delta\mathbf{r}(x)$ is the deviation of the firing rates from the fixed point. Defining $\mathbf{T}$ and $\mathbf{W}(x)$ as in Section S2.1 and $\mathbf{1}$ as the two-dimensional identity matrix, the linearized dynamics are $\tau_E \frac{d\delta\mathbf{r}(x)}{dt} = \sum_{x'} \hat{\mathbf{W}}(x,x')\delta\mathbf{r}(x')$ where $\hat{\mathbf{W}}(x,x')$, the Jacobian matrix of the dynamical system, is given by $\hat{\mathbf{W}}(x,x') = \mathbf{T}^{-1}\left(\mathbf{\Phi}(x)\mathbf{W}(x-x') - \mathbf{1}\right)$ with $\mathbf{\Phi}(x) = nk^{1/n}\begin{pmatrix} r_E^{\mathrm{FP}}(x)^{\frac{n-1}{n}} & 0 \\ 0 & r_I^{\mathrm{FP}}(x)^{\frac{n-1}{n}} \end{pmatrix}$. The linearized dynamics will accurately describe model behavior so long as $\mathbf{r}$ is sufficiently close to $\mathbf{r}^{\mathrm{FP}}$. As in Section S2.1 for $\mathbf{W}$, we write $\hat{\mathbf{W}} = \begin{pmatrix} \hat{\mathbf{W}}_{\mathrm{EE}} & -\hat{\mathbf{W}}_{\mathrm{EI}} \\ \hat{\mathbf{W}}_{\mathrm{IE}} & -\hat{\mathbf{W}}_{\mathrm{II}} \end{pmatrix}$ to refer to the full matrix of linearized connections, where the submatrix $\hat{\mathbf{W}}_{\mathrm{XY}}$ has elements $\left(\hat{\mathbf{W}}_{\mathrm{XY}}\right)_{\mathrm{ij}} = \hat{\mathbf{W}}_{\mathrm{XY}}(x_i, x_j)$.

We can numerically calculate the eigenvalues and eigenvectors of the effective linear weight matrix, $\hat{\mathbf{W}}$, at a given steady-state and use them to predict the effective dynamic regime of the nonlinear network. For the fixed point to be stable, all eigenvalues of $\hat{\mathbf{W}}$ must have negative real part. We roughly equate[5] the ISN regime with at least one pattern of excitatory activity being unstable in the absence of feedback inhibition, meaning that at least one eigenvalue of $\hat{\mathbf{W}}_{\mathrm{EE}}$ has positive real part; this will occur when the values of $r_E^{\mathrm{FP}}(x)$ become sufficiently large.

As a simple test, we stimulated E and I cells in the 1-D nonlinear model of Fig. 3 with a small, centrally-located stimulus of strength c and considered the linearization about the corresponding fixed point. All eigenvalues of $\hat{\mathbf{W}}$ have negative real part for all stimulus strengths tested (up to $c = 100$). We then perturbed only the stimulated inhibitory cells with a small additional excitatory input, and recorded the responses of the E and I cells at the stimulus center. We observe, as expected, that the stimulus strength at which the largest real part of an eigenvalue of $\hat{\mathbf{W}}_{\mathrm{EE}}$ becomes positive coincides closely with the stimulus strength at which additional drive to inhibitory cells causes a decrease in steady-state inhibitory firing rate (Supplemental Fig. S2d).

We also consider linearization about the steady-state response to a full-field stimulus of

[5]Why this is a rough rather than exact correspondence: The condition that at least one eigenvalue of $\hat{\mathbf{W}}_{\mathrm{EE}}$ has positive real part ensures that there is at least one pattern of E-unit activity that is unstable in the linear regime. By the reasoning of section S2.2.3, this ensures that a pattern of small input to I cells that causes a new steady state in which the only change is a decrease in amplitude of this unstable E-unit pattern must result in the E cells receiving less inhibition in that new steady state. This is the paradoxical signature of the ISN. However this does not ensure that any particular stimulus to I cells, such as the uniform input over a 1^{o} region used in Supplemental Fig. S2d, will yield a paradoxical response.

strength c. We use the linearized weights to calculate the network resonant and critical frequencies.[6] All three frequencies monotonically increase with stimulus strength, and once stimulus strengths reach moderate levels both E and I resonant frequencies are below the critical frequency, meaning that the network shows ISN behavior for stimuli at those frequencies (Supplemental Fig. S2e). The analytically calculated frequencies from the linearization describe most-amplified (resonant) frequencies for small sinusoidal deviations about the uniform stimulus of strength c, but they reasonably approximate the preferred CM frequencies in the actual network (Fig. 5D), which involve responses to stimuli that oscillate sinusoidally between strength 0 and strength c.

[6]Because the stimulus is spatially uniform, steady-state firing rates are spatially uniform except for boundary effects. Therefore, away from the boundaries, the linearized weight submatrices $\hat{\mathbf{W}}_{\mathrm{XY}}$, $X, Y \in E, I$, remain translation-invariant, of Gaussian shape about each position but with size scaled by steady state firing rates according to $\mathbf{\Phi}$ (which does not vary with position). As a result, we expect the eigenvectors of the submatrices to closely approximate the Fourier modes, with real eigenvalues that decrease with frequency. Accordingly, we diagonalize each linearized submatrix with eigenvalues ordered from largest to smallest real parts with imaginary part discarded, and make the ansatz that the weight matrix $\hat{\mathbf{W}}$ in the Fourier basis is well approximated by the matrix of these four diagonal submatrices (with minus signs for the submatrices $\hat{\mathbf{W}}_{\mathrm{XY}}$ for $y = I$). That is, we equate $\hat{W}_{\mathrm{XY}}(k)$ with the k^{th}-largest eigenvalue in the real diagonal submatrix $\hat{\mathbf{W}}_{\mathrm{XY}}$. We then can calculate the equivalent of $\mathcal{L}_{\mathrm{E}}(k)$ and $\mathcal{L}_{\mathrm{I}}(k)$ in this representation (Eqs. S3-S4), with replacements $(1 + \tilde{W}_{\mathrm{II}}(k)) \to \hat{W}_{\mathrm{II}}(k)$, $(1 - \tilde{W}_{\mathrm{EE}}(k)) \to -\hat{W}_{\mathrm{EE}}(k)$, $\tilde{W}_{\mathrm{EI}}(k) \to \hat{W}_{\mathrm{EI}}(k)$, $\tilde{W}_{\mathrm{IE}}(k) \to \hat{W}_{\mathrm{IE}}(k)$, $\mathrm{Det}\,(\mathbf{1} - \tilde{\mathbf{W}}(k)) \to \mathrm{Det}(-\hat{W}(k))$. We find the resonant network frequencies from the peaks of $\mathcal{L}_{\mathrm{E}}(k)$ and $\mathcal{L}_{\mathrm{I}}(k)$ vs. k. The critical frequency is found as the first frequency for which the corresponding eigenvalue of $\hat{\mathbf{W}}_{\mathrm{EE}}$ is < 0.

S3 Supplemental Figures

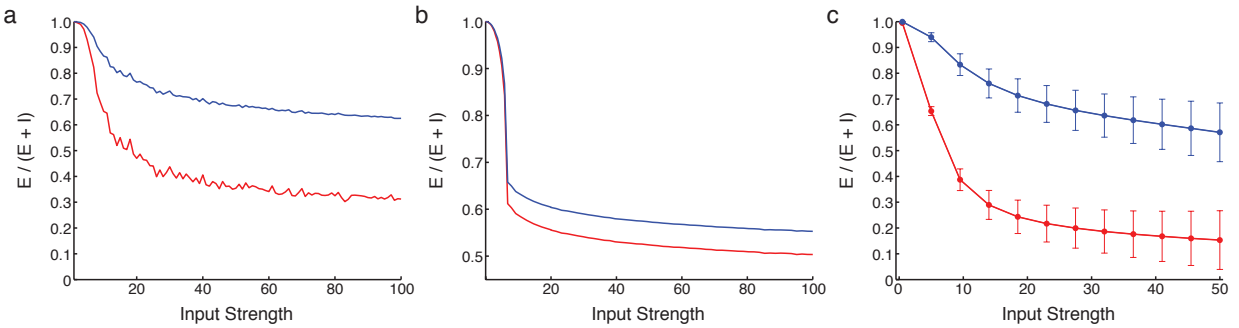


Figure S3:

Increasing dominance of network input by inhibition with increasing input strength, using protocol of Shao et al. (2013) (related to Figs. 1F, 3B, and 6C). In Figs. 1F, 3B, and 6C, we showed that the ratio $E/(E+I)$ of network-driven excitation E to network-driven inhibition I grows with increasing stimulus strength, as in Fig. 3E of Shao et al. (2013). However, in our model figures, the input was given equally to E and I cells. Here we show that the same result holds when we more closely model the protocol of Shao et al. (2013). In Shao et al. (2013), channelrhodopsin (ChRh) was expressed in L2/3 of mouse S1 in around 25% of E cells (ChRh was expressed in $21.1 \pm 0.9\%$ of all neurons, and about 80% of neurons are E cells). Then in cortical slices, intracellular recordings were made from non-ChRh-expressing E cells to assay the strength of excitatory and inhibitory input, E and I, evoked by light stimulation of the ChRh-expressing E cells. To model this, we chose E cells randomly with probability 0.25 to receive input (input identical to that in main-text figures, except set to zero for all I cells and for E cells with probability 0.75), and recorded from the E cells not receiving input (red curves) or from the I cells (blue curves). Results corresponding to Fig. 1F (**a**), Fig. 3B (**b**), or Fig. 6C (**c**) all show qualitatively the same behavior as the original figures: increasing dominance of network input by inhibition with increasing stimulus strength.

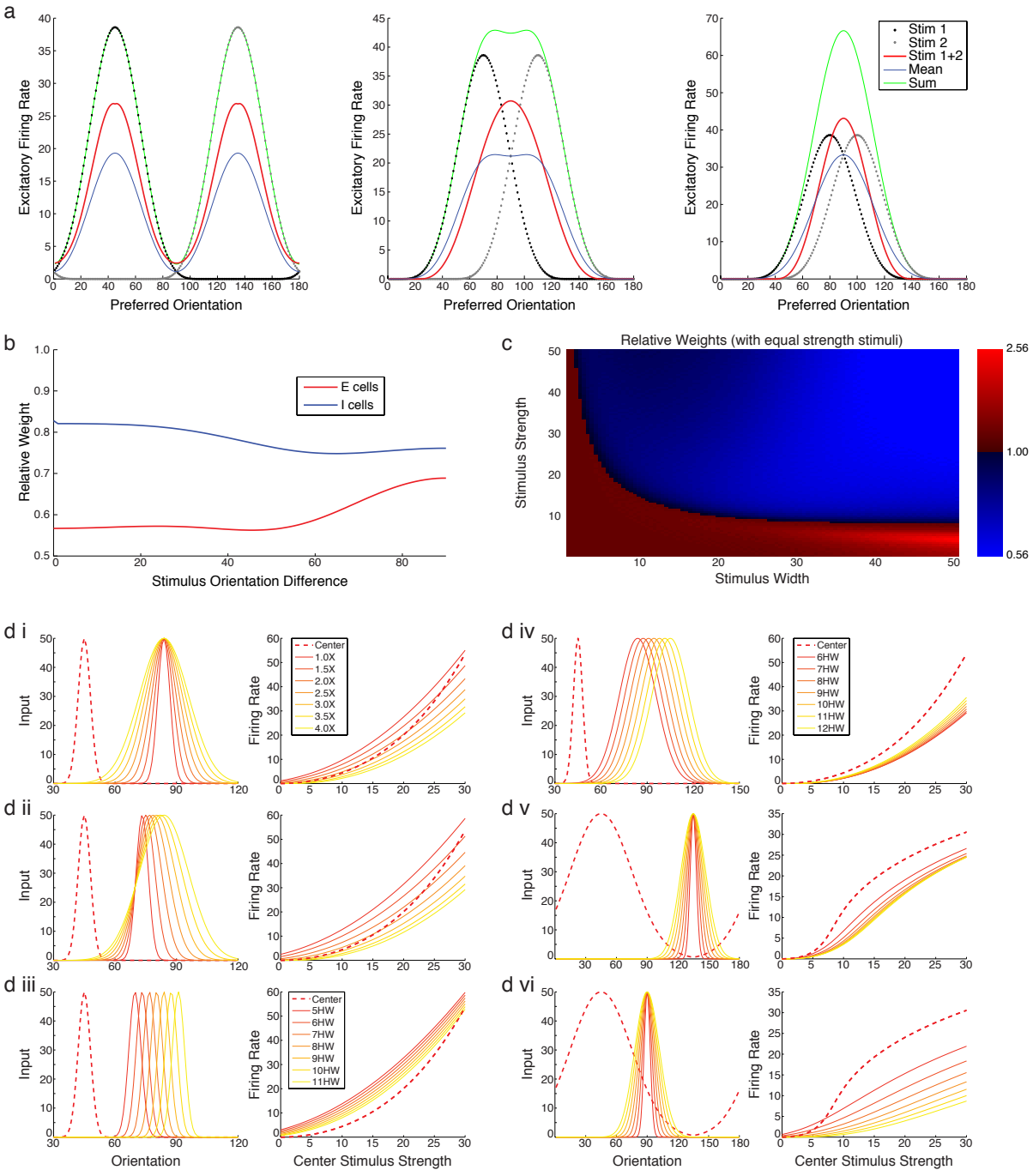


Figure S4:

Normalization in the ring model of Figure 1. a,b: Non-orthogonal grating pairs also show sublinear addition. **a**: E-cell responses to the two stimuli presented individually (dotted) and to the simultaneous presentation of both stimuli (red), as well as the sum (green) and mean (blue) of the individual responses, for orientation differences of 90°, 40°, and 20°. **b**: The sublinear weights w for E and I cells vs. stimulus orientation difference. (Continued on next page.)

Figure S4:

(Continued). w is defined as in Fig. 1I. Strength of each stimulus is $c = 50$ for all panels. **c**: The switch from supra- to sub-linear addition occurs at lower stimulus strengths for wider stimuli, as suggested by the fact that larger stimulus widths switch from being facilitating to suppressive at lower stimulus strengths (**c**, and Figs. 3E, 1J, 6H and Supplemental Fig. S8d). The relative additive weights for excitatory cells in response to two equal-strength, equal-width orthogonal stimuli (orientation difference 90°) are plotted as a function of stimulus width and stimulus strength. For comparison, simulations in **b** and in Fig. 1G used stimulus strength $c = 50$ and stimulus width $\sigma_{FF} = 30°$. **d.** Small but not large surround stimuli, relative to summation field size, can facilitate response to a low-contrast center. **d(i-vi)**. Left panels show stimuli: red dashed curve, center stimulus for $c = 50$, always presented at orientation 45° with strengths varying from $c = 0$ to $c = 50$ by 1; yellow through red solid curves, surround stimuli varying as indicated in legend, $c = 50$ in all cases. Right panels show response vs. center contrast: red dashed curve, response to center stimulus alone; yellow through red solid curves, response to center plus correspondingly colored surround stimulus. **d(i-iv)**. Small center stimulus: center size is summation field size for $c = 50$, which is $\sigma_{FF}^{ctr} = 3$ (Fig. 1J). In **(i-iv)** we will describe separation of center and surround in units of half-widths (HW), meaning the separation between the largest orientation corresponding to half width at half height (HWHH ORI) of the center grating, and the smallest HWHH ORI of the surround grating. **d(i)**. Surround stimulus is varied in size from 3 to 12 in steps of 1.5. Surround gratings have common center position, set so that largest surround is separated by 6 HW. **d(ii)**. Surround stimuli vary in size as in **(i)**, but their positions are varied so that all surround stimuli have separation 6 HW. **d(iii)-(iv)**. Qualitative effects of **(i)-(ii)** are insensitive to position. **d(iii)**. Surround stimulus equal in size to center stimulus, varied in position from 5 to 11 HW in steps of 1 HW. All curves have wide region of facilitation. **d(iv)**. Large surround stimulus (size 12) is varied in position from 6 to 12 HW in steps of 1 HW. All curves show only suppression. **d(v)-(vi)**. Sufficiently small and near surrounds can facilitate even large center stimuli. Here, center size is large ($\sigma_{FF}^{ctr} = 30$, as in all other ring model simulations) and surround is small (varying from 3 to 12 in units of 1.5). Distances are between center positions of the two gratings. **d(v)**. Surrounds orthogonal to center (90° distance). At this distance surrounds are only suppressive. **d(vi)**. Surround are 45° from center. Here, small surrounds can facilitate low-contrast center. Note, the case of center and surround both of size 30 produces only suppression, as illustrated in **a**.

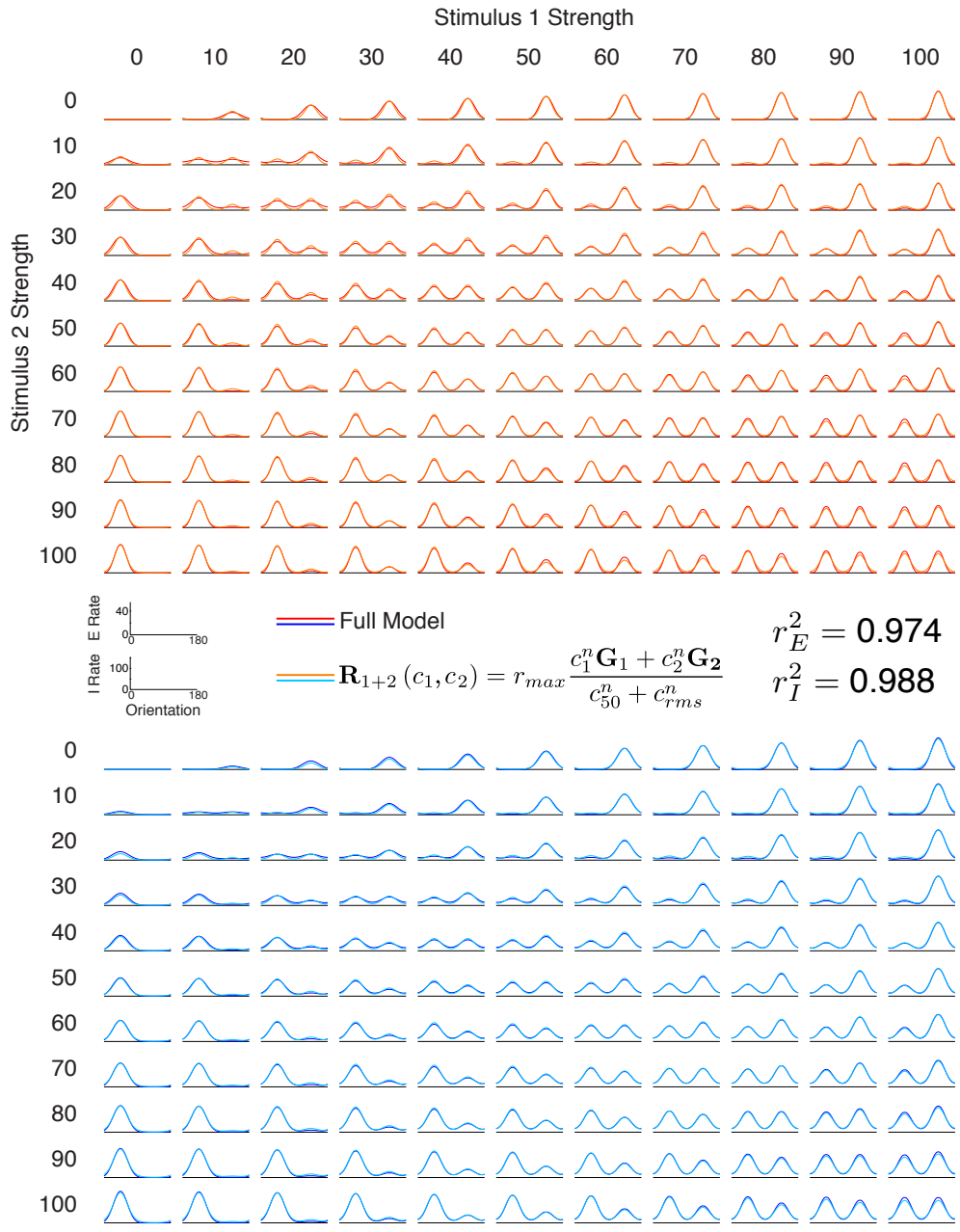


Figure S5:

Ring model results of Fig. 1 are well fit by equations of the normalization model. We presented pairs of orthogonal stimuli (at 45° and 135°). Each stimulus was presented at 21 stimulus magnitudes c, from 0 to 100 in steps of 5. This yields $21 \times 21 = 441$ firing rate curves for both E and I, of which $11 \times 11 = 121$ are illustrated. Each curve shows firing rate (y axis) vs. ring coordinate from $\theta = 0°$ to $\theta = 180°$ (x axis). Separately for E and I, these 441 curves were used to fit equations of the phenomenological normalization model (Busse et al. 2009), shown in figure. Red (E) and blue (I) curves show responses of full model, while orange (E) and cyan (I) curves show fit of normalization model (discrepancies become visible when figure is expanded). (Continued on next page.)

Figure S5: (continued) In the normalization model equation, $\mathbf{R}_{1+2}(c_1, c_2)$ is the vector of responses across the population (E or I) to presentation of stimuli 1 and 2 at strengths c_1 and c_2, respectively, $c_{rms} = \sqrt{c_1^2 + c_2^2}$, and $\mathbf{G}_1$ and $\mathbf{G}_2$ are vectors of responses across the population given by identical circular Gaussians, with amplitude A and width parameter p, centered at 45° and 135°, respectively. This model has five free parameters: r_{max}, c_{50}, n, A, and p. These were fit to give the least squared error across the 441 curves. r_E^2 and r_I^2 are the R^2 values (% of variance explained) for the fits for E and I curves, respectively. The best-fit parameters are: E units, $n = 1.2278$, $c_{50} = 24.4622$, $r_{\max} = 47.7657$, $A = 1.6441$, $p = 3.3977$; I units, $n = 1.4213$, $c_{50} = 80.2834$, $r_{\max} = 149.9930$, $A = 2.9298$, $p = 2.3514$. Note that this model cannot fit the supralinear summation we see for stimulus strengths below 10 (Fig. 1I), because $\mathbf{G}_1$ does not contribute appreciable activation to the region activated by $\mathbf{G}_2$ and vice versa. Because 19 of 20 nonzero strengths used here were 10 or above, this did not greatly affect fit quality. Circular Gaussians have the form $G(\theta) = Ae^{p*cos(2(\theta-\theta_0))}/(2\pi\mathcal{I}_0(p))$, where θ_0 is the respective stimulus orientation and $\mathcal{I}_0$ is the modified Bessel function of the first kind ("besseli" in Matlab) of order 0.

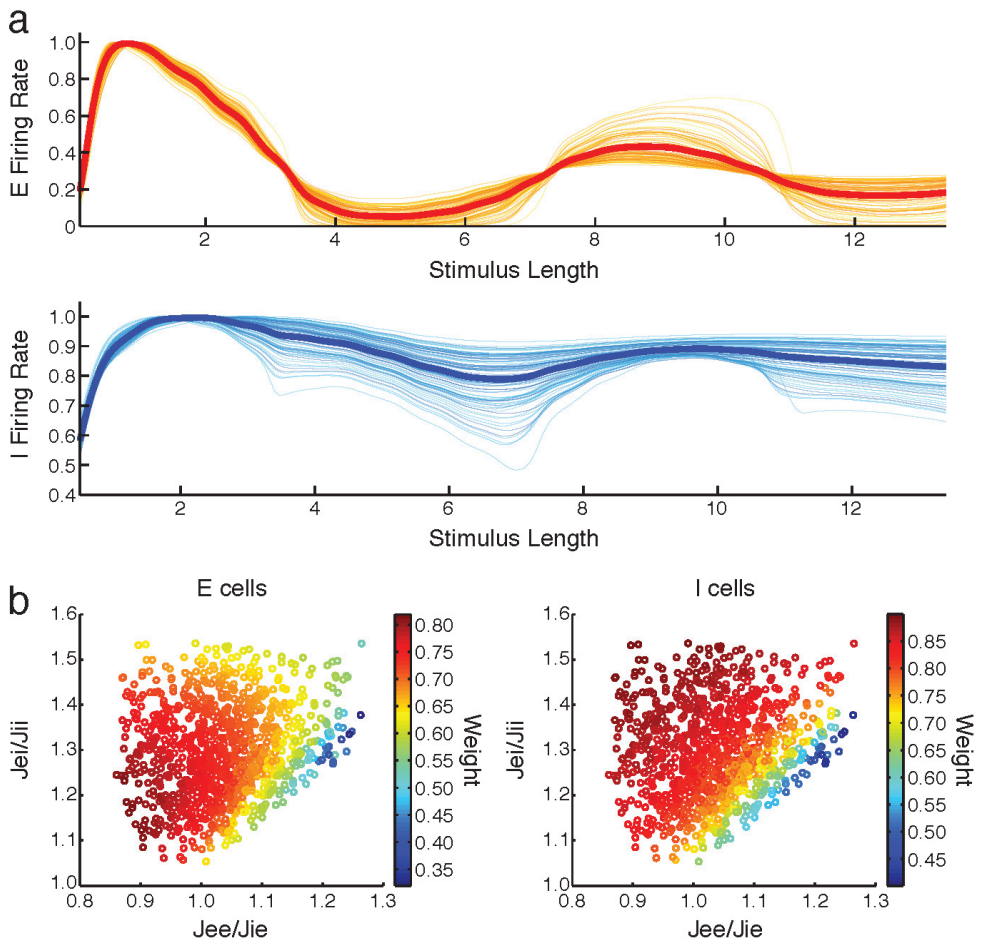


Figure S6:

Robustness of results for ring model (Figure 1) and Nonlinear 1-D model (Figure 3). For Fig. 3 (**a**) and Fig. 1 (**b**), we show that a basic feature of the model response stays qualitatively invariant, changing smoothly and continuously, as parameters are changed, indicating that model behavior does not require fine tuning of parameters. **a.** Length-tuning curves for nonlinear 1-D model of Fig. 3 (top: E cell; bottom: I cell) under weight perturbations. 100 simulations were run, identical to Fig. 3C for $c = 31$, except that in each simulation, each of the four weight-strength parameters (J_{EE}, J_{IE}, W_{EI}, W_{II}, defined in Section S1.1.2) were randomly chosen from a uniform distribution covering $\pm 10\%$ about the value used for Fig. 3. The length-tuning curves from Fig. 3C are shown in bold, and those from the 100 randomly-perturbed simulations are shown as thin lines. **b.** Summation weights w for ring model of Fig. 1 (E cells: left; I cells: right) under weight perturbations. As in Fig. 1I, two orthogonal gratings of equal strength (here, $c = 100$) were shown, and the weight w was determined as that giving least-squared-error fit to the equation $R_{12} = w(R_1 + R_2)$ where R_{12} is the vector of responses across the population of E or I cells to the simultaneous presentation of both gratings, while R_1 and R_2 are the vector responses to one or the other grating alone. 1000 simulations were run, in each of which the four weights J_{EE}, J_{IE}, J_{EI}, and J_{II} (Section S1.1.2)) were each drawn from uniform distributions covering $\pm 10\%$ about the value used in Fig. 1. Each dot presents results from one simulation, with color indicating the resulting weight w. Sublinearity of summation grows stronger (w decreases) with increasing bias of excitatory projections toward E cells (increasing J_{EE}/J_{IE}) and, to a lesser extent, with increasing bias of inhibitory projections toward I cells (decreasing J_{EI}/J_{II}).

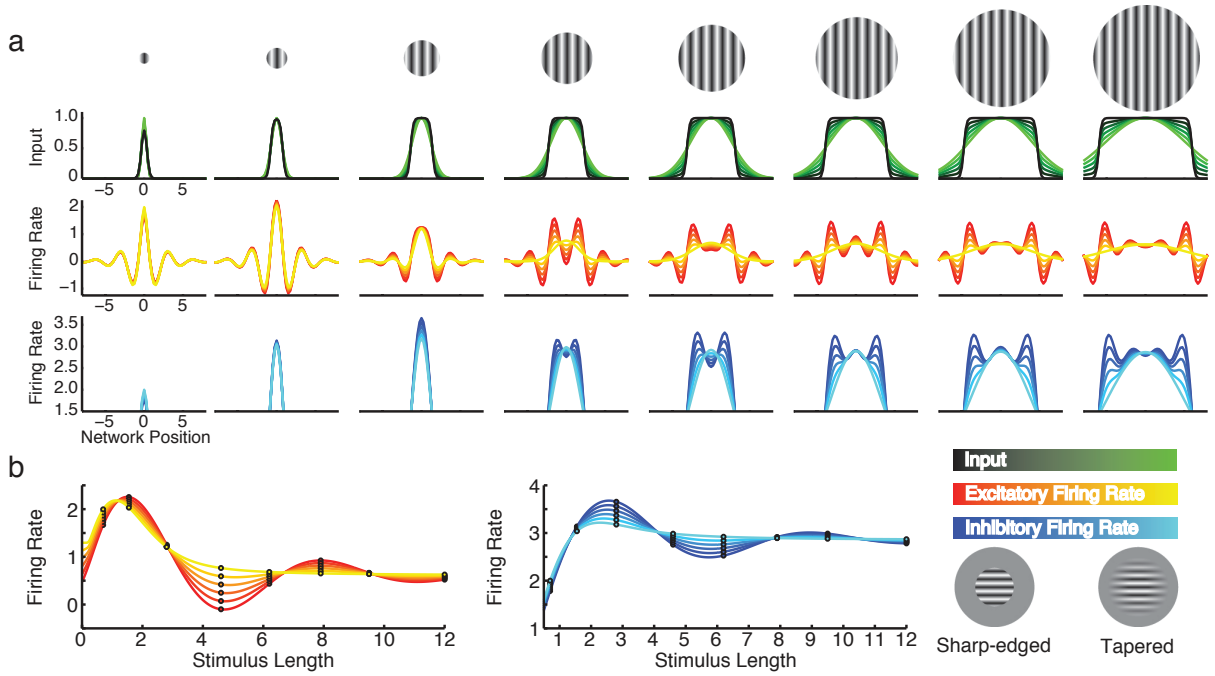


Figure S7:

Effects of varying input shape, from sharp-edged to smoothly tapered, in the linear model of Fig. 2. a. Input (top row) and corresponding E (middle row) and I (bottom row) activity vs. spatial position (x-axes). This repeats Fig. 2B, but now with input shape (top) interpolating, as line colors move from black to light green ("Input" in legend, bottom right), from sharp-edged (black lines, as used in Fig. 2) to Gaussian (lightest green lines) (see Section S1.1.1 for equations for input shapes). Corresponding E (middle) and I (bottom) firing rates are shown by lines varying (see legend) from red to yellow (E) or blue to cyan (I) as inputs vary from sharp-edged to Gaussian. For larger stimuli, more smoothly-tapering stimuli do not yield spatially periodic activity, for reasons described in Section S2.1.3. **b** E (left) and I (right) length tuning curves for cell at middle of grid. This repeats Fig. 2C with the array of input shapes from **a**. Colors as in **a** and legend. More smoothly-tapering stimuli do not yield periodic length tuning curves.

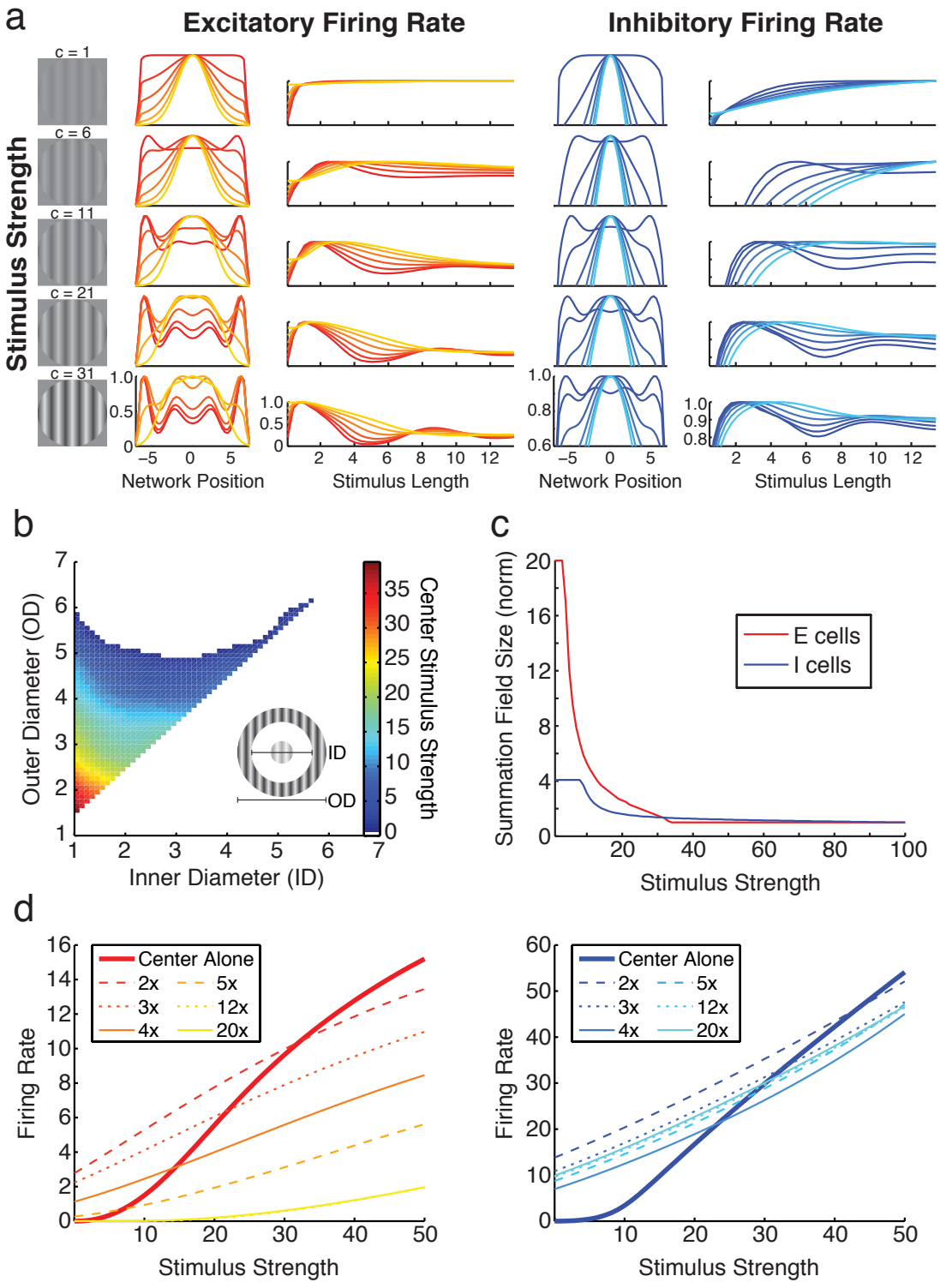


Figure S8:

Nonlinear 1-Dimensional Spatial Model of Figure 3. **a**: Periodicity in spatial activity profiles and length tuning curves depends on input shape. (Continued on next page.)

Figure S8:

(Continued). This panel repeats Fig. 3C, but with the array of input shapes of Fig. S7, ranging from sharp-edged (as used in Fig. 3) to Gaussian. Colors as in Fig. S7, with red to yellow (E) and blue to cyan (I) corresponding to sharp-edged to Gaussian-shaped stimuli. Conventions otherwise as in Fig. 3C. As in the linear model (Fig. S7), periodicity in spatial activity patterns and length tuning curves disappears as stimuli become more smoothly tapering. **b**: Dependence of facilitation on position as well as size of surround. Center stimulus fills $c = 50$ summation field, as in Fig. 3E. Surrounds ($c = 50$) are "annular", which in 1D means segments equally spaced on either side of center stimulus (inset shows inner diameter (ID) and outer diameter (OD) for 2D case; 1D stimulus is a slice through the center of the 2D stimulus). Inner diameter of annular stimulus (distance between inner edges of the two segments), in units of center stimulus diameter, given by x-axis (= 1 in Fig. 3E) and outer diameter by y-axis (=total stimulus size in Fig. 3E). Figure shows heat map of center stimulus strength at which effect of surround switches from facilitation to suppression for E units. White indicates that surround is never facilitating (lowest center strength examined, $c = 0.5$). For larger inner and outer diameters up to 20, there is no facilitation except for inner diameter 1 and outer diameter 14-17.5 for very low center strengths (switch at center c ranging from 2 to 4.5). Note that this panel explains why high-contrast surrounds can facilitate responses to a low-contrast center (**d**; Fig. 3E), whereas in the ring model a high-contrast orthogonal grating only suppressed response to a low-contrast grating (Fig. 1H). These results are seen to be consistent when surround size is expressed in terms of high-contrast summation-field size (hcSFS). Here, facilitation occurs for surrounds smaller than about 5 times the hcSFS. In the ring model, the gratings were about 10× larger than the hcSFS (compare Fig. 1J). Smaller high-contrast orthogonal gratings can instead yield facilitation (Supplemental Fig. S4 c,d). **c-d**: Contrast-dependent changes in summation field size and surround effects are insensitive to stimulus shape. The simulations from Figs. 3D-E are repeated but with Gaussian-shaped rather than sharp-edged stimuli. The results are essentially the same: (**c**) shrinking of summation fields with increasing contrast and (**d**) a switch from surround facilitation to surround suppression with increasing center contrast for smaller, but not larger, surround stimuli for E cells (left) and for all surround stimuli for I cells (right). In **d**, center stimulus is Gaussian-shaped with full-width at half-height (FWHH) 0.55° for E, 1.9° for I units, and stimulus strength as shown on x-axis; surround stimuli are sharp-edged stimuli with sizes, positions, and strengths ($c = 50$) exactly as in Fig. 3E.

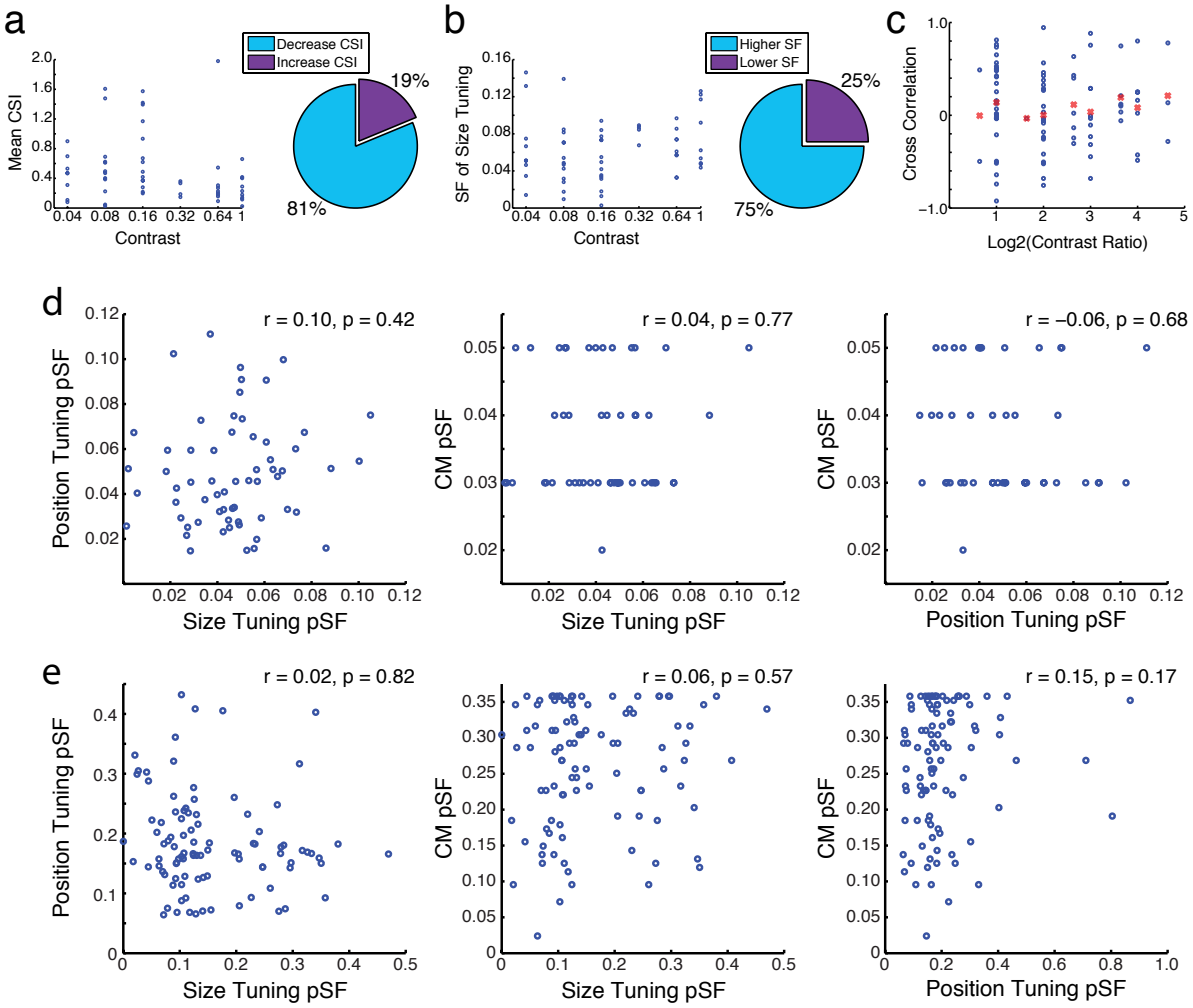


Figure S9:

Experimental Results (related to Figs. 4 and 5). a-c: Size-tuning at multiple levels of stimulus contrast. We were only able to study this in a small number of cells (16): 4 studied at 2 contrasts, 2 at 3 contrasts, 6 at 4 contrasts, and 4 at 5 contrasts. **a.** Our results are consistent with those of Wang et al. (2009). They studied "counter-suppression" (CS) – a re-emergence of activity at larger stimulus sizes after surround suppression. They defined a CS index (CSI): CSI $= \frac{R_{cs}-R_{min}}{R_{max}}$, where R_{max} is the firing rate at the summation field peak (the first peak in the size-tuning curve); R_{min} is the minimum firing rate in the surround suppressed region (the first dip); and R_{cs} is the maximum rate for still larger stimuli. They found that CSI was largest at low contrasts, which seems counter to our prediction that periodicity of length tuning curves is stronger at higher contrasts. We suggest that CSI decreases with contrast because at low contrast, size tuning curves can be fundamentally summating/facilitating – responses increase with size, but with some wiggles and dips along the way. Such continued facilitation after the first dip will produce the largest R_{cs} and hence the largest CSI. This interpretation is supported by findings of Wang et al. (2009) that responses at maximal stimulus size were larger, relative to R_{min}, and suppression indices lower at low contrast than at high contrast. (Continued on next page.)

Figure S9:

(Continued). To calculate CSI in our own data, we removed subjectivity by using all extrema in each size-tuning curve, calculating CSI for each consecutive peak-trough-peak triplet and taking the average of these as the curve's CSI. Dots represent CSI in the 16 cells for each combination of cell and contrast. Consistent with our interpretation, in our data, which by multiple other measures accords with our model's predictions, we also observe a decrease in CSI with increasing contrast, with 13/16 cells (81%) showing a higher CSI at the lowest contrast studied than at the highest contrast studied ($p < 0.022$, two-sided binomial test of null hypothesis that CSI equally likely to be highest at lowest or highest contrast). **b.** From our fit to the SSM model, we see an increase in dominant spatial frequency with increasing contrast, as the model predicts. 12/16 neurons (75%) showed a higher size-tuning spatial frequency at the highest contrast studied than at the lowest contrast studied. This trend did not reach statistical significance in a two-sided test ($p < 0.077$, two-sided binomial test of null hypothesis that frequency equally likely to be highest at lowest or highest contrast) but was significant at the 0.05 level in a one-sided binomial test ($p < 0.0385$). **c.** To determine whether spatially periodic responses might arise from static properties of the functional architecture, such as the periodicity of orientation columns, rather than arising dynamically as a contrast-dependent spatial resonance, we measured the cross-correlations of all pairs of size-tuning curves measured for a given cell at two different contrasts (after discarding the first 40% of the curve, which in general for all cells simply captures spatial summation). We plot this against the base-2 log of the ratio of the stimulus contrasts. The mean of each distribution is indicated by a red X. For all contrast ratios (except for ratio 3.125, log ratio 1.644, for which there was only 1 data point), the median was not significantly different from 0 (two-sided one-sample Wilcoxon signed rank test). (The means also were not significantly different from zero as judged by a two-sided t-test, with all p-values ≥ 0.13, but because the data may not be normally distributed these statistics are not reliable, hence we used the non-parametric Wilcoxon test). Thus, peak and trough locations for a given cell vary with contrast in a manner inconsistent with a static origin of the periodicity. For increasing contrast ratios, statistics for the Wicoxon test are: $N = (2, 28, 21, 7, 11, 6, 7, 3)$ curves; $p = (1.00, 0.11, 0.99, 0.47, 0.97, 0.094, 0.69, 0.75)$; signed rank $= (1, 273, 116, 19, 32, 19, 17, 4)$. **d,e**: Lack of correlation between the three measures of network resonant frequency (best fit frequencies to size-tuning and position-tuning curves and contrast modulation pSF) in experiments (**d**) and in model (**e**). Experiments: all cells for which we had both forms of data; model: 100 randomly selected E units; in both cases, restricted to cells or units for which SSM gave a better fit than DOG by nested F-test for length and/or position tuning if used in given panel. The calculated correlation coefficient and its p-value are indicated on each plot. In both experiments and model, there is no significant correlation between any pair of the three spatial frequencies, presumably reflecting different subnetworks of cells being recruited by each stimulus paradigm. For three panels, left to right: Experiments, $n = 62$, 50, 45; $DF = 60$, 48, 43; $t = 0.817$, 0.292, -0.417; Model, $n = 86$, 96, 88; $DF = 86$, 96, 88; $t = 0.227$, 0.575, 1.376.

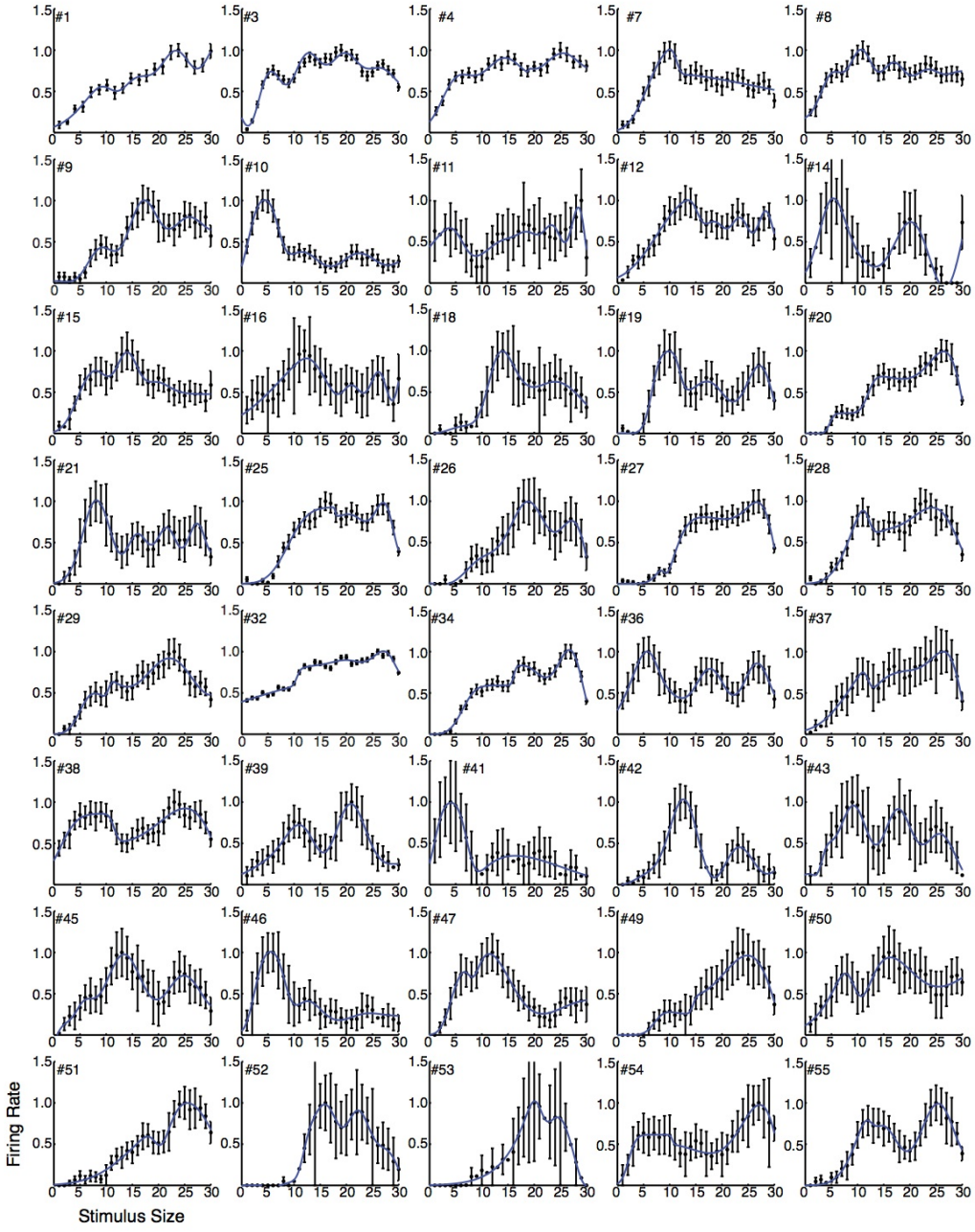


Figure S10:

All experimentally measured size-tuning curves (related to Fig. 4). Cell numbers are indicated. Data points: tuning curve and standard deviation as computed by maximum likelihood estimation (Section S1.4.2). Blue curves: best fit of SSM model curve. Statistics for goodness of fit of SSM vs. DOG model are in Supplemental Tables S1 (for 8-parameter DOG fit) and S2 (for 7-parameter DOG fit). Cells illustrated in Fig. 4A are #42 (left) and #36. (Continued on next page.)

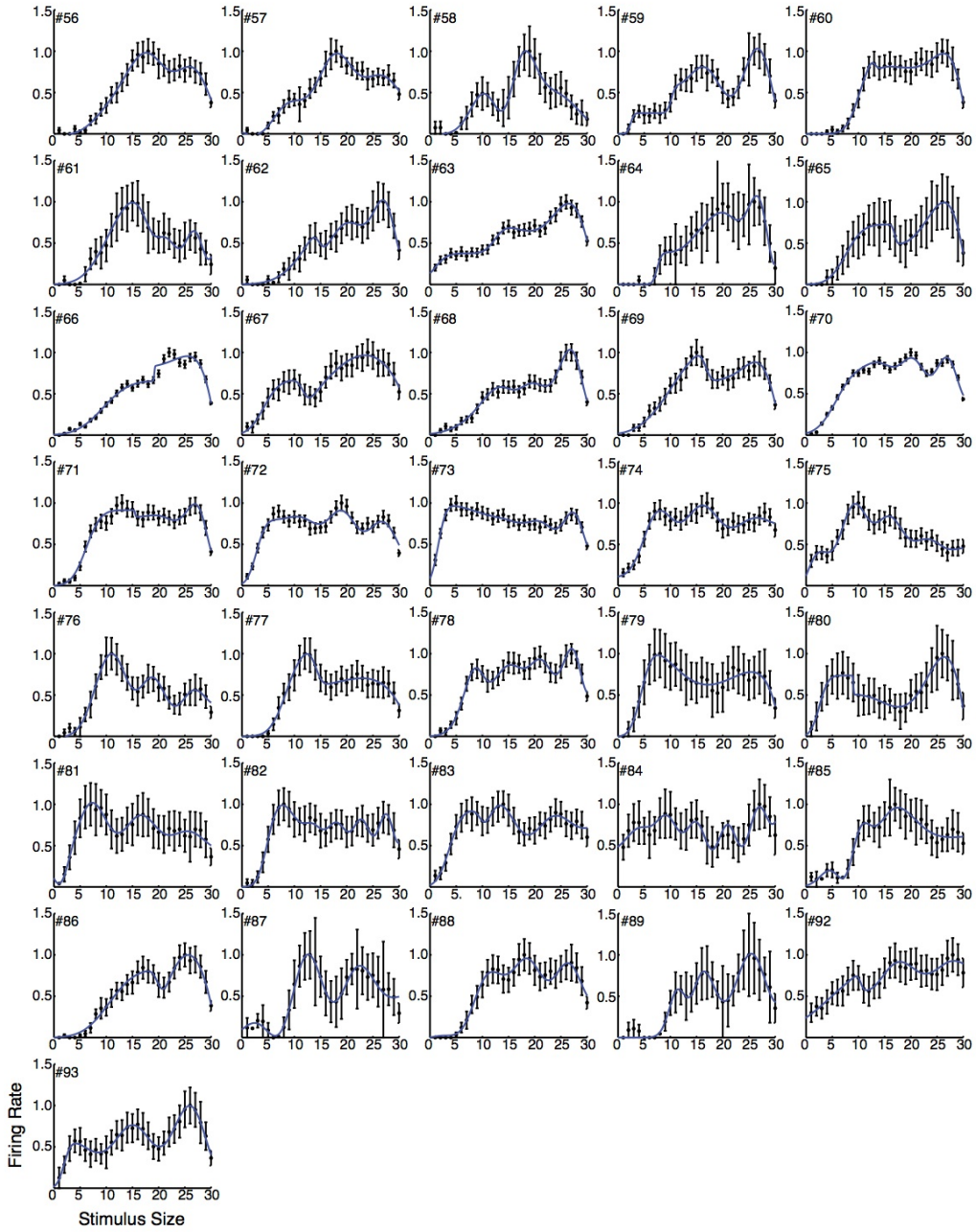


Figure S10:
All experimentally measured size-tuning curves (related to Fig. 4); (continued).

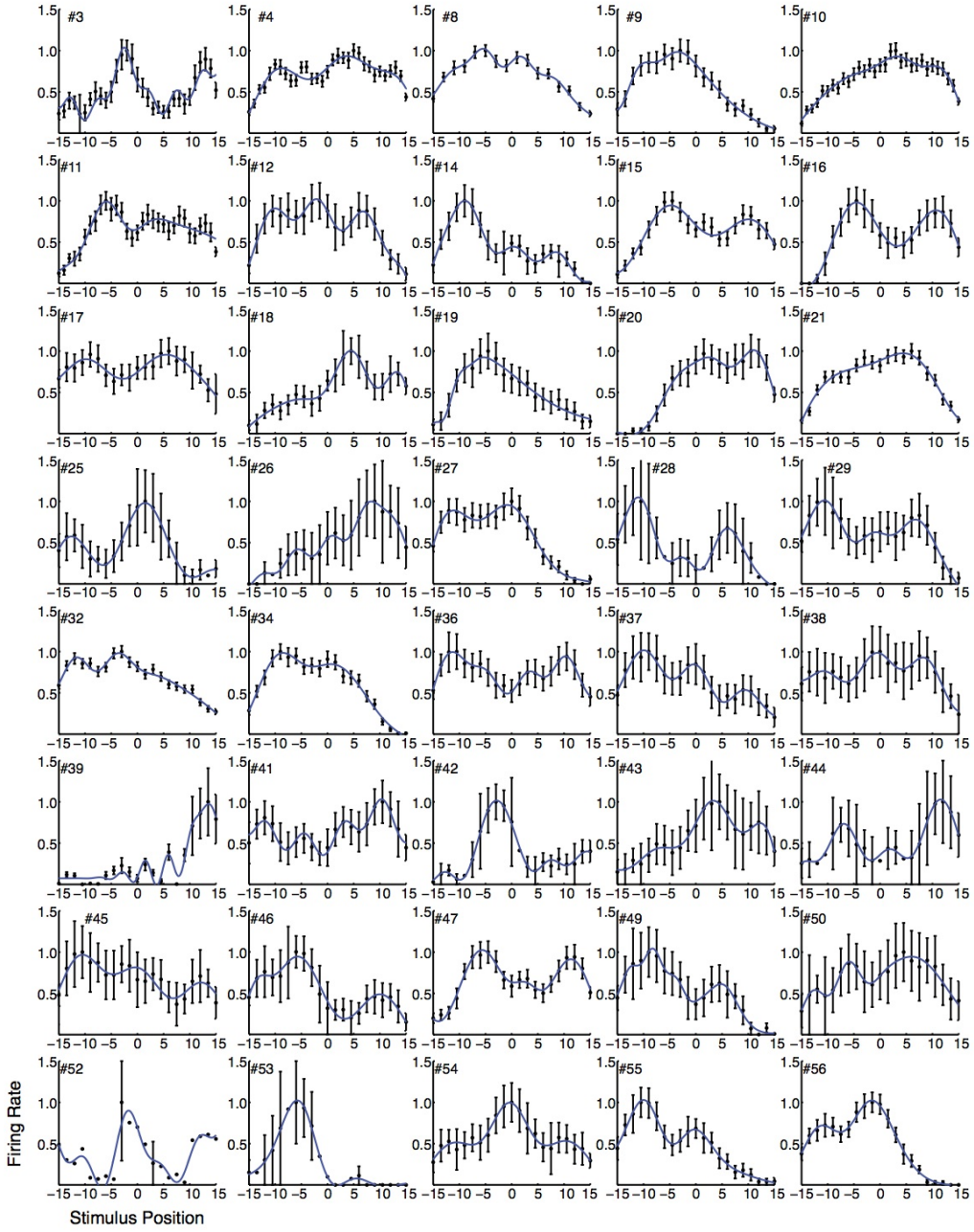


Figure S11:

All experimentally measured position-tuning curves (related to Fig. 4). Cell numbers are indicated. Data points: tuning curve and standard deviation as computed by maximum likelihood estimation (Section S1.4.2). Blue curves: best fit of SSM model curve. Statistics for goodness of fit of SSM vs. DOG model are in Supplemental Table S3. Cells illustrated in Fig. 4D are #16 (left) and #79. (Continued on next page.)

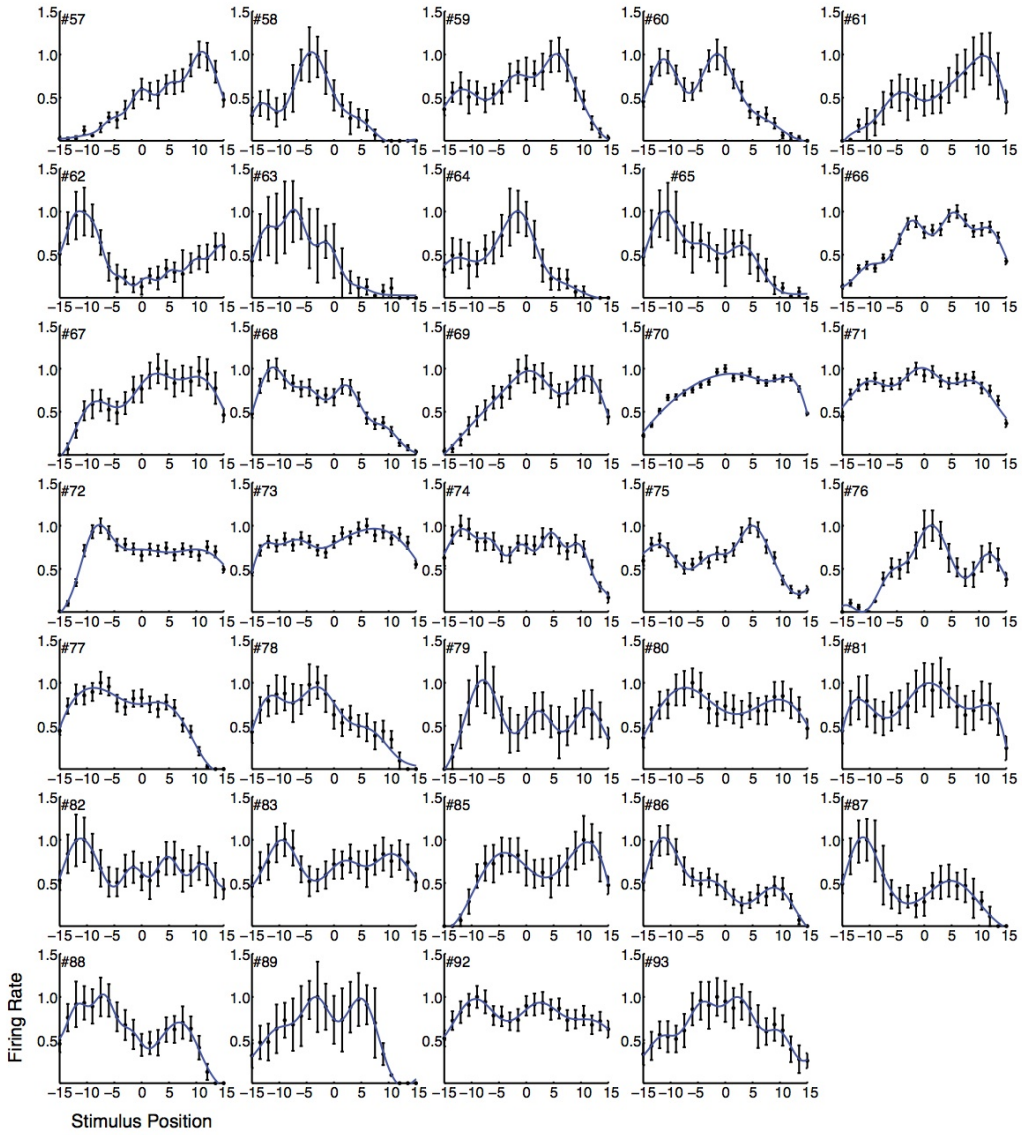


Figure S11:

All experimentally measured position-tuning curves (related to Fig. 4); (continued).

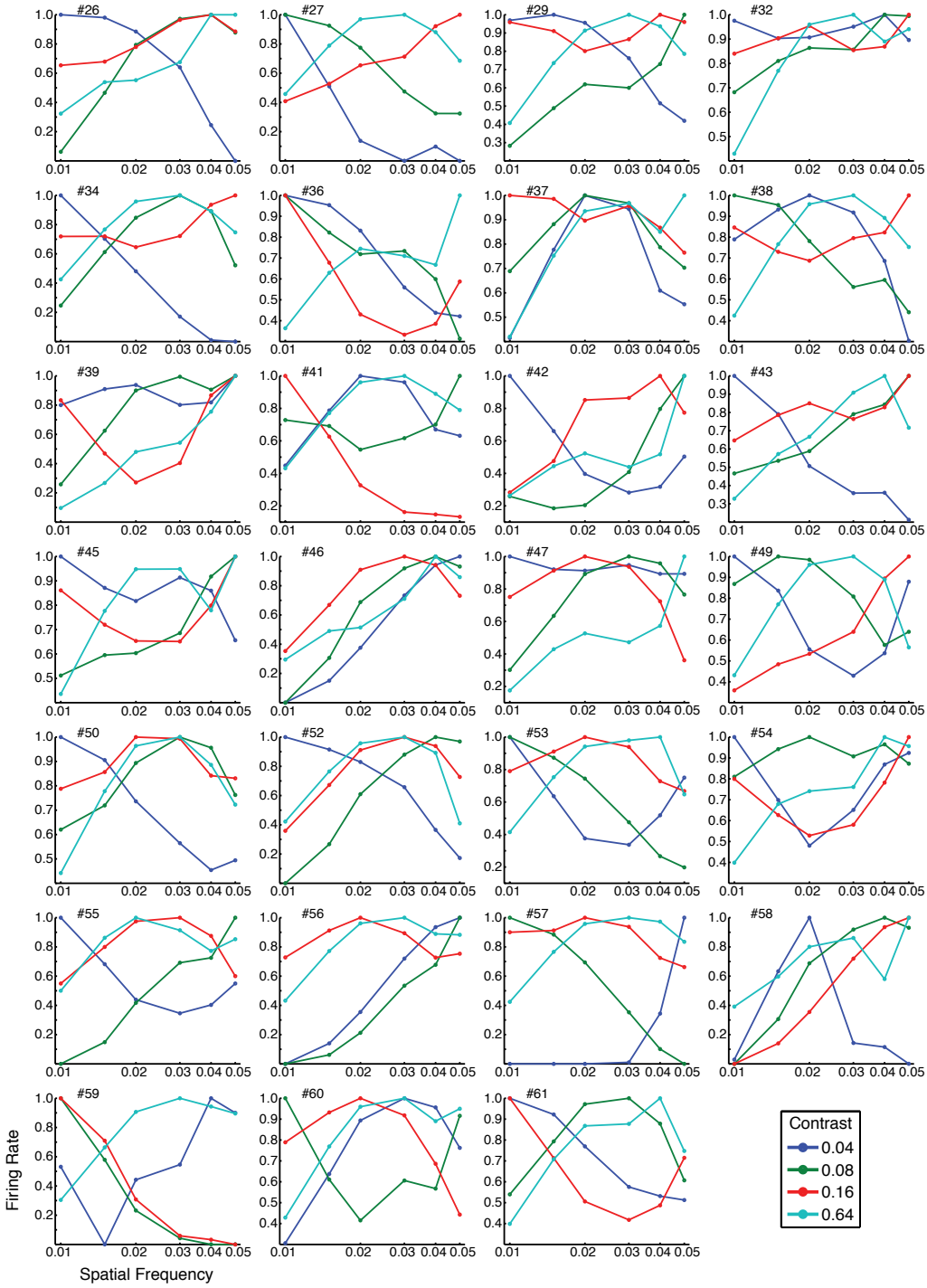


Figure S12:

All experimentally measured contrast-modulation curves (related to Fig. 5). Cell numbers are indicated. Data points: tuning curve for each contrast as indicated in legend, computed by maximum likelihood estimation (Section S1.4.2). Error bars omitted for visibility given multiple curves. Cells illustrated in Fig. 5E (left to right) are #89, #79, and #66. (Continued on next page.)

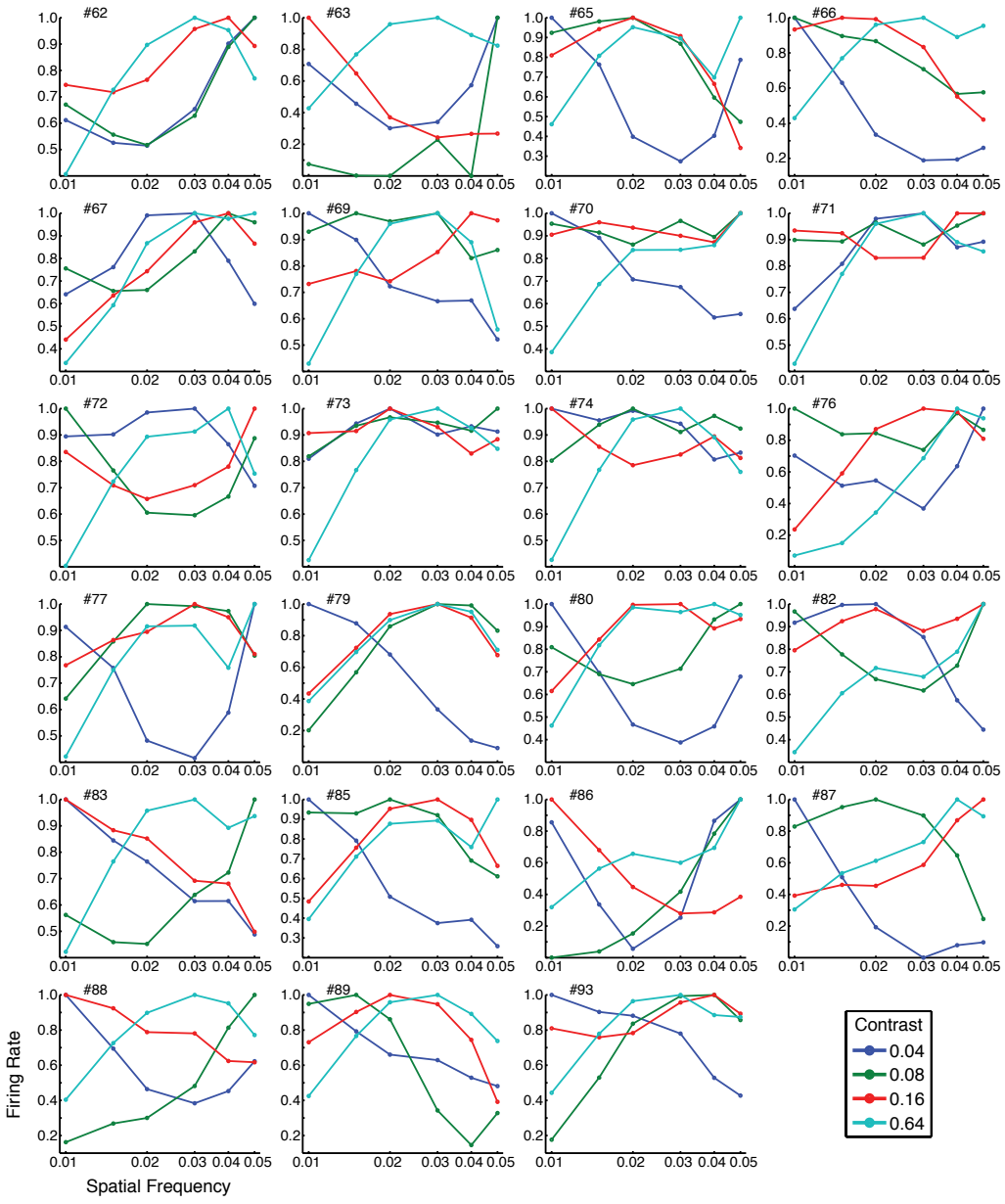


Figure S12:

All experimentally measured contrast-modulation tuning curves (related to Fig. 5); (continued).

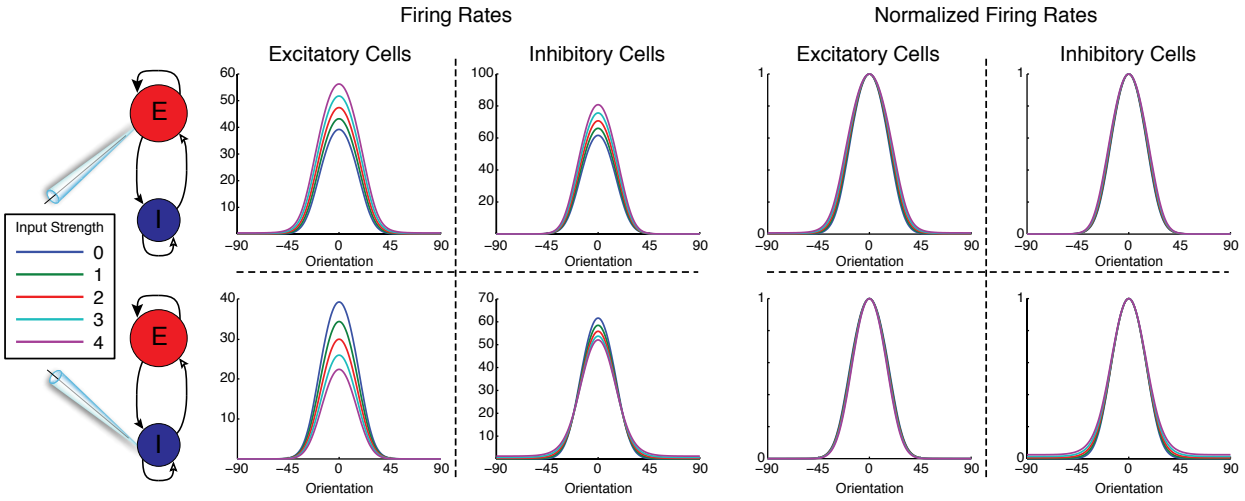
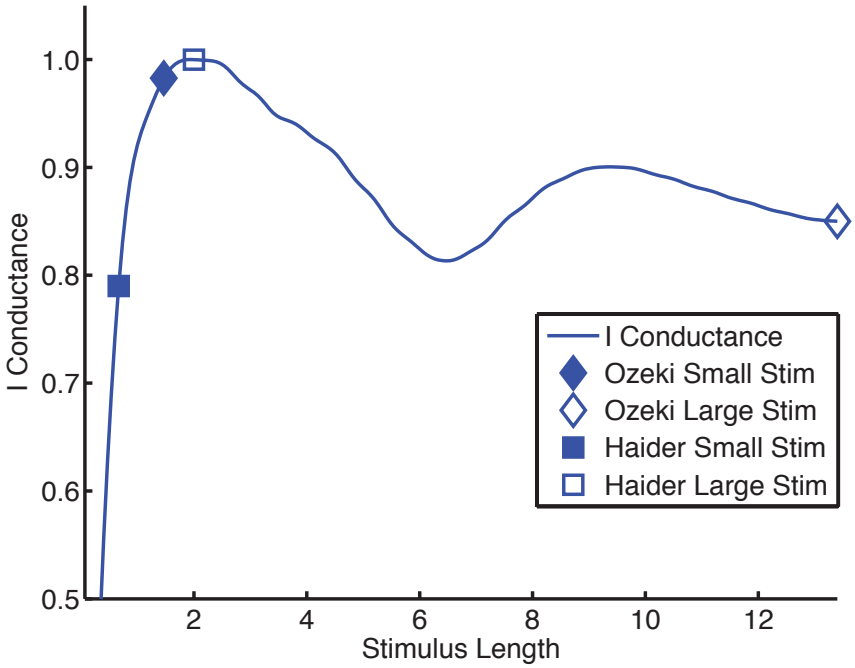


Figure S13:

In a nonlinear two-neuron model, shifting the balance of input towards E or I can cause a multiplicative change in the gain (related to Discussion). The rate model of balanced amplification studied in Murphy and Miller (2009) was a linear model, so responses to different stimuli necessarily added linearly. Thus, a modulatory input added a fixed amount to the responses to any other input. Here we show that balanced amplification in the present model with a power-law input/output nonlinearity can yield multiplicative gain modulation. "Tuning-curves" for both the E and I cell in this simple model were generated by varying the strength of a stimulus of equal magnitude delivered to both cells. At the same time, a constant modulatory input (relative input strength shown in key at left) was added to either the E cell (top row) or I cell (bottom row). Left two plots: adding an input to the E cell boosted the gain of both E and I, whereas adding an input to the I cell reduced the gain. Right two plots: curves replotted normalized so that maximal firing rate in each curve equals 1; the modulatory input induces a nearly multiplicative change in gain. Parameters: $n = 2.2$, $k = 0.01$, $W_{\mathrm{EE}} = 1.0$, $W_{\mathrm{IE}} = 1.25$, $W_{\mathrm{EI}} = 0.75$, $W_{\mathrm{II}} = 0.75$, $g_E = g_I = 1$. Baseline input at orientation θ: $50e^{-\frac{\theta^2}{2\sigma^2}}$, $\sigma = 20^{\mathrm{o}}$. Modulatory input: to I cells, from 0 to 10 in steps of 2.5; to E cells, from 0 to 5 in steps of 1.25.



]

Figure S14:

How the model could simultaneously account for results of Ozeki et al. (2009) and Haider et al. (2010) (related to Discussion). We consider the model of Fig. 3. The figure shows the net I input received by an E cell ("I conductance") at the stimulus center as a function of stimulus length, for $c = 40$. Symbols mark possible sizes of small or center stimuli (filled symbols) and large or center+surround stimuli (open symbols) in experiment of Ozeki et al. (2009) (diamonds) or of Haider et al. (2010) (squares). This shows how Haider et al. (2010) could have seen an increase in inhibition received with increasing stimulus size, while Ozeki et al. (2009), using larger stimuli, saw a decrease in inhibition received. The indicated stimulus sizes are determined as follows. In Ozeki et al. (2009), the small stimulus size, meant to cover the CRF, was determined as the size of a shrinking annular stimulus, centered on the CRF center, that first evoked spikes, typically around 2° diameter; the large stimulus size was 20° diameter, typically about 10× larger. The annulus used to measure the CRF had a large outer diameter, and the inner diameter was shrunk until spikes were evoked. In our model for the parameters we use, an annulus with a large outer diameter never facilitates E-cell responses, and hence would not evoke spikes, down to an inner diameter of the E-cell summation-field size (Supplemental Fig. S8b), which is about 0.5° (Fig. 3C). Thus, if we used a CRF size as determined by a large annulus, it would be equal to the summation-field size. We instead here defined the CRF size to be the inner diameter of an annulus of 1° width – outer diameter 2° larger than inner diameter – that first evoked a response. This is quite simply a "hack", but we don't know how general across parameters is the model behavior that annuli with large outer diameter do not facilitate, and it at least illustrates how the model might produce behaviors like those seen experimentally. The key point is that the CRF should be larger than the E-cell summation field size, so that it more closely corresponds to the I-cell summation field size (which in the model for the parameters we use is larger than the E-cell summation field size, Figs. 3C, 7C) and thus to the peak of inhibition received. (Continued on next page.)

Figure S14:
(Continued). Indeed, in monkeys it has been reported that receptive fields determined by the annular procedure are on average 47% larger than the summation field size (Cavanaugh et al. 2002). While 47% is much less than the nearly 4-fold difference between summation field size and annular receptive field size used here, it may also be that for other parameters the average I-cell summation field size would be closer to 50% larger than the E-cell summation field size. We took the large stimulus for Ozeki et al. (2009) simply to be a large stimulus where suppression, though oscillating, had reached a plateau in overall level, here 13°. In Haider et al. (2010), the small stimulus was defined as the circular fit to the half-maximal iso-response contour of the RF as mapped by sparse noise stimuli, while the large stimulus was three times larger. In the model, the size of the small stimulus, so defined, is somewhat dependent on the size of the stimuli used for sparse mapping of the receptive field. We used a 0.5°-width stimulus, moved over 500 positions in steps of 0.0667°. A wider or narrower mapping stimulus would produce slightly larger or smaller sizes, respectively, for the small stimulus. We took the large stimulus for Haider et al. (2010) to be 3× the size of the small stimulus.

S4 Supplemental Tables: Legends

Supplemental Tables are provided in separate Excel files. Here we provide the legends for those tables.

Table S1:

Statistics for length-tuning analyses of Fig. 4B-C. Columns show: cell ID (as in Supplemental Figs. S10-S12); SSE_DOG and SSE_SSM, the sum-squared error (SSE) for DOG model and for SSM model respectively; the value of the statistic F, and the corresponding p-value, for the nested F-test; and CV Percent Change, the % change in median sum-squared error from DOG to SSM models. For F-test, degrees of freedom (DF) are $(\Delta p, B)$ where Δp is the difference in the number of parameters in the two models, and $B = n - (1 + p)$ where n is the number of data points in the tuning curves being compared and p is the number of parameters in the more complex model. For Fig. 4B these numbers are: $n = 30$, $p = 10$, $\Delta p = 2$, $DF = (2, 19)$.

Table S2:

Statistics for length-tuning analyses of Supplemental Fig. S1f-g. Same as Supplemental Table S1, but for comparison of fits of SSM model to fits of 7-parameter DOG model ($a_7 = 0$) rather than 8-parameter DOG model as in Supplemental Table S1. All conventions as in Supplemental Table S1. For F-tests of Supplemental Fig. S1f: $n = 30$, $p = 10$, $\Delta p = 3$, $DF = (3, 19)$.

Table S3:

Statistics for position-tuning analyses of Fig. 4E-F. Same as Supplemental Table S1, but for position-tuning analyses of Fig. 4E-F. All conventions as in Supplemental Table S1. For F-tests for Fig. 4E: $n = 21$, $p = 10$, $\Delta p = 3$, $DF = (3, 10)$.

Table S4:

Statistics for length-tuning analyses of model units, Fig. 6C, left. Same as Supplemental Table S1, but for comparison of fits to length-tuning curves of SSM vs. DOG model for 100 randomly selected model E units, Fig. 6C. All conventions as in Supplemental Table S1, except that to save space we did not discuss cross-validation tests for model cells, so those statistics are not included here. For F-tests for Fig. 6C, left: $n = 61$, $p = 10$, $\Delta p = 2$, $DF = (2, 50)$.

Table S5:

Statistics for position-tuning analyses of model units, Fig. 6C, right. Same as Supplemental Table S1, but for comparison of fits to position-tuning curves of SSM vs. DOG model for 100 randomly selected model E units, Fig. 6C, right. All conventions as in Supplemental Table S1, except that to save space we did not discuss cross-validation tests for model units, so those statistics are not included here. For F-tests for Fig. 6C, right: $n = 30$, $p = 10$, $\Delta p = 3$, $DF = (3, 19)$.

References

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

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

Albus, K. (1975). A quantitative study of the projection area of the central and the paracentral visual field in area 17 of the cat. I. The precision of the topography. *Exp. Brain Res.*, 24:159–179.

Busse, L., Wade, A. R., and Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, 64:931–942.

Cavanaugh, J. R., Bair, W., and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.*, 88:2530–2546.

Edgell, S. E. and Noon, S. M. (1984). Effect of violation of normality on the t test of the correlation coefficient. *Psych. Bull.*, 95:576–583.

Fino, E. and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron*, 69:1188–1203.

Haider, B., Krause, M. R., Duque, A., Yu, Y., Touryan, J., Mazer, J. A., and McCormick, D. A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*, 65:107–121.

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

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58:1233–1258.

Kaschube, M., Schnabel, M., Löwel, S., Coppola, D. M., White, L. E., and Wolf, F. (2010). Universality in the evolution of orientation columns in the visual cortex. *Science*, 330:1113–1116.

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, 36:ECVP Abstract Supplement.

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

Murphy, B. K. and Miller, K. D. (2009). Balanced amplification: A new mechanism of selective amplification of neural activity patterns. *Neuron*, 61:635–648.

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

Packer, A. M. and Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *J. Neurosci.*, 31:13260–13271.

Sadagopan, S. and Ferster, D. (2012). Feedforward origins of response variability underlying contrast invariant orientation tuning in cat visual cortex. *Neuron*, 74:911–923.

Shao, Y. R., Isett, B. R., Miyashita, T., Chung, J., Pourzia, O., Gasperini, R. J., and Feldman, D. E. (2013). Plasticity of recurrent l2/3 inhibition and gamma oscillations by whisker experience. *Neuron*, 80:210–222.

Tsodyks, M. V., Skaggs, W. E., and Sejnowski, T. J.and McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.*, 17:4382–4388.

Turrigiano, G. (2011). Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annu. Rev. Neurosci.*, 34:89–103.

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334:1569–1573.

Wang, C., Bardy, C., Huang, J. Y., FitzGibbon, T., and Dreher, B. (2009). Contrast dependence of center and surround integration in primary visual cortex of the cat. *J Vis.*, 9:1–15.