Neuron

Supplemental Data

Inhibitory stabilization of the cortical network underlies visual surround suppression

Hirofumi Ozeki, Ian M. Finn, Evan S. Schaffer, Kenneth D. Miller, and David Ferster

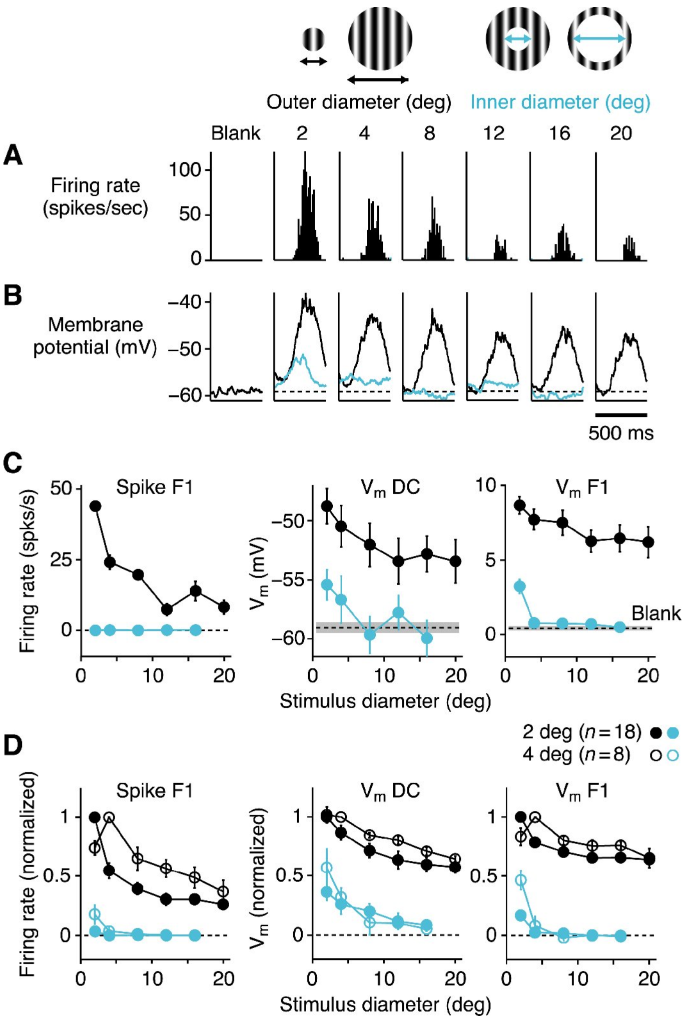

Figure S1. Membrane potential and firing rate responses underlying size tuning

(A and B) Cycle-averaged firing rate and membrane potential responses for a simple cell. Responses to stimuli of different outer diameters are shown in black; responses to annular stimuli of different inner diameters (20 degrees outer diameter) are shown in cyan. Dashed lines in (B) represent the mean of the blank response.

(C) Size-tuning curves for firing rate (F1 component) and membrane potential (DC and F1 components). Colors correspond to those in (A) and (B). Error bars and shading indicate the s.e.m. Increasing the grating size beyond the classical receptive field (2 degrees diameter) decreased the membrane potential modulation by approximately 30%, the mean depolarization by 40-50%, and the spike responses by 80-90%. Stimulation of the surround alone with an annular grating of 2 degrees inner diameter evoked no hyperpolarization, but instead evoked a small, modulated subthreshold depolarization, which is consistent with that receptive field defined by synaptic input and membrane potential response being larger than that defined by spike response (Bringuier et al., 1999; Moore and Nelson, 1998; Tan et al., 2004).

(D) Normalized and averaged size tuning curves for the population of 26 simple cells. The cells were divided into two groups: 18 cells that responded maximally to stimuli of 2 degrees diameter (closed symbols) and 8 cells that responded maximally to stimuli of 4 degrees diameter (open symbols). The 20-degree diameter, center-plus-surround stimulus caused a reduction of center-only response by 71% ± 4% in firing rate F1 (mean ± s.e.m.), 41% ± 4% in membrane potential DC, and 36% ± 3% in membrane potential F1.

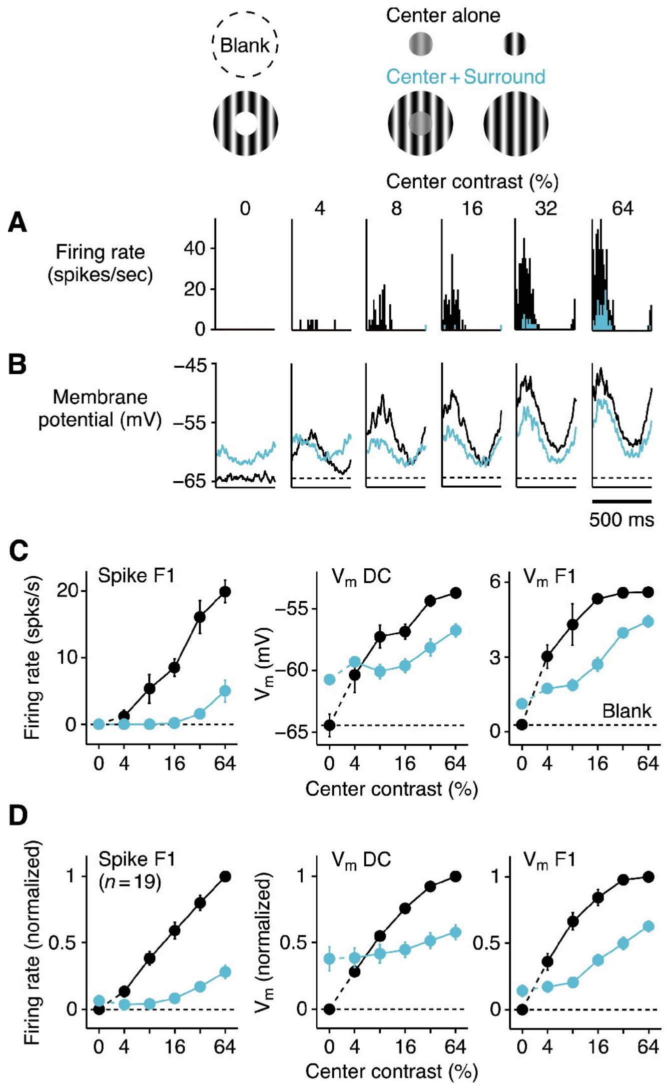

Figure S2. The effects of surround stimulation on the contrast-response function

(A-C) Firing rate and membrane potential responses to center gratings of different contrasts are shown for a simple cell, with (cyan) and without (black) an annular grating of 64% contrast present. Dashed lines represent the mean of the blank response. At lower center contrasts, the surround grating caused a slight mean depolarization and reduction in modulation amplitude of the membrane potential responses to the center stimulus. At higher center contrasts, the surround grating suppressed both the mean and the modulation components of the center responses with a consequent downward shift of the entire sinusoidally-modulated membrane potential response.

(D) Averaged contrast-response curves for the population of 19 simple cells, normalized to the center-only response at 64% contrast. For a center stimulus at 64%, the surround caused a reduction of the center response by 72% ± 5% in firing rate F1 (mean ± s.e.m.), 42% ± 6% in membrane potential DC, and 37% ± 4% in membrane potential F1. The net effect of surround stimulation on firing rate responses was an almost complete suppression at lower center contrasts and a scaling at higher center contrasts (Cavanaugh et al., 2002; Sengpiel et al., 1998), whereas the surround stimulus on its own caused a noticeable depolarization that could be a cortical mechanism underlying a contrast-dependent increase in spatial summation (Anderson et al., 2001; Sceniak et al., 1999).

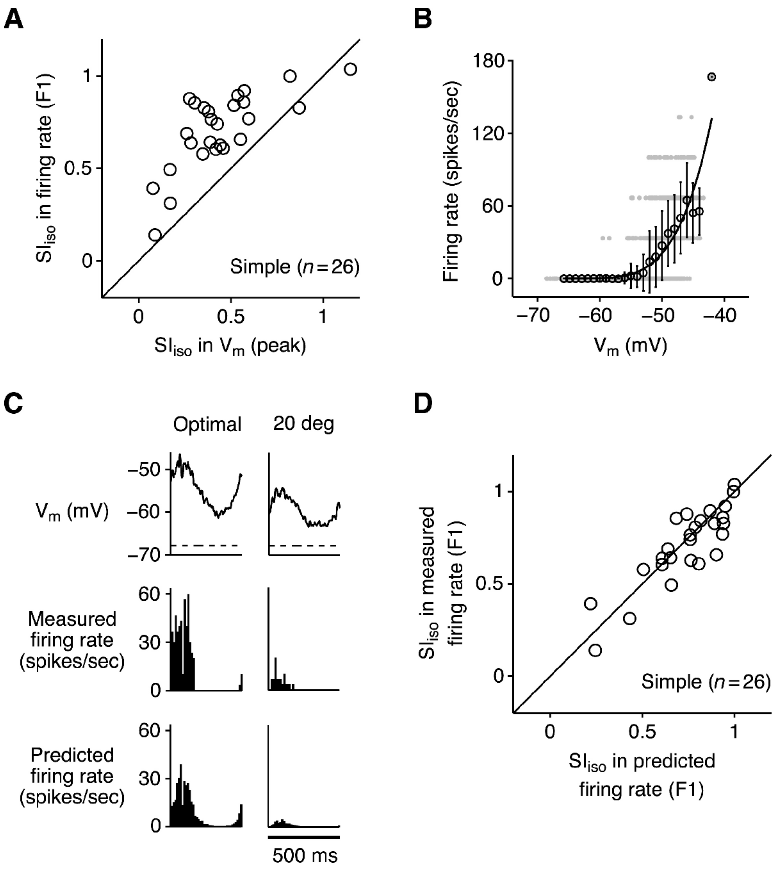

Figure S3. Amplification of surround suppression by the threshold nonlinearity

(A) Suppression index (SI = 1 – $R_{center+surround}/R_{center}$) measured for firing rate (F1 component) plotted against that for membrane potential (peak depolarization) for 26 simple cells. SI is a measure of the percentage by which the surround stimulus decreases the response to the center stimulus (SI = 1 represents complete suppression; SI = 0 represents no suppression). These 26 cells (same as Figure S1) showed statistically significant suppression in firing rate F1. The mean SI (mean ± s.e.m.): 0.44 ± 0.05 for membrane potential and 0.71 ± 0.04 for firing rate.

(B) The relationship between mean membrane potential and mean firing rate is shown for 30-ms epochs of the responses to a variety of visual stimuli from one example cell (gray points). Open symbols show the averages of the points in 1-mV bins (mean ± s.e.m.). Black line is a power-law fit to the individual points by $R(V_m) = a\left[V_m - V_{rest}\right]_+^p$, where R is firing rate, V_m is membrane potential, V_{rest} is resting potential, and the subscript + indicates a rectification at 0 (Anderson et al., 2000; Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002). The power law fit (p = 4.5) lies close to the averaged points.

(C) Prediction of the firing rate responses to the center and center-plus-surround gratings by applying the power law to the membrane potential at each point in time during the responses. The cycle averages of membrane potential (top), recorded firing rate (middle), and predicted firing rate (bottom) are shown for the same cell in (B). The SI for membrane potential, measured firing rate, and predicted firing rate: 0.35, 0.83, and 0.89.

(D) For the population of 26 cells, SI derived from the recorded firing rate is plotted against that derived from the predicted firing rate. The points cluster around a line of slope 1, indicating that threshold is likely the major mechanism accounting for the increase in suppression between membrane potential and firing rate.

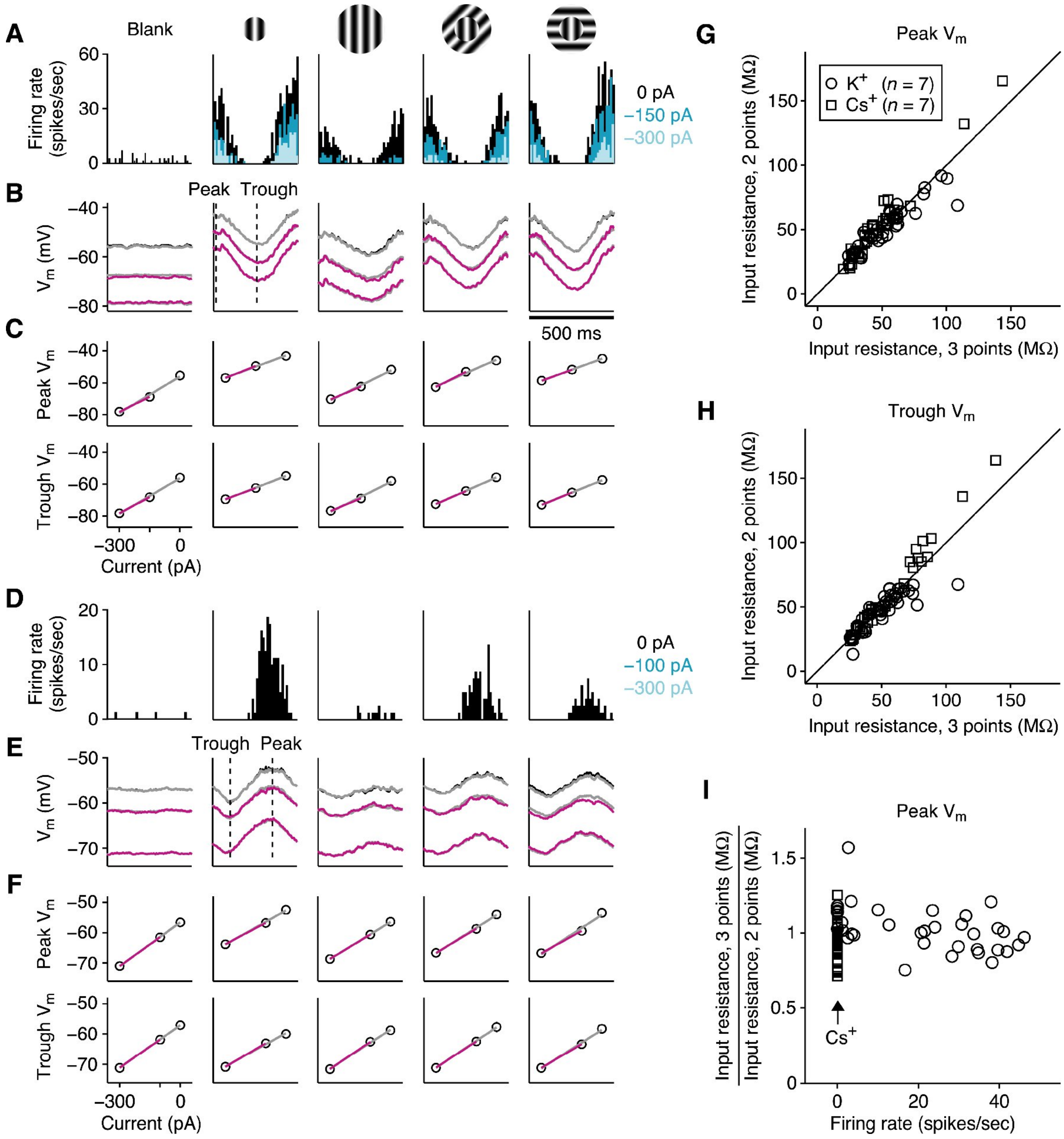

Figure S4. The effects of spiking on conductance estimates

Stimulus-evoked changes in conductance were measured by injecting steady currents of 3 different amplitudes during repeated visual stimulation (Anderson et al., 2000; Boudreau and Ferster, 2005). Electrode resistance, measured by injection of brief current pulses (–0.1 nA; 250 ms), was compensated off-line. At each point during the visual responses, membrane conductance was derived from the slope of the I-V curve. We derived the excitatory and inhibitory components of the visually-evoked conductance from the membrane equation:

$$V_{visual}(t) = \left[g_e(t)\cdot V_e + g_i(t)\cdot V_i + g_{rest}\cdot V_{rest}\right]/g(t),$$

where $V_{visual}(t)$ is the response without injected current, V_{rest} is the resting potential, $g(t)$ is the total conductance, and g_{rest} is the resting conductance. $g_e(t)$ and $g_i(t)$ are the visually-evoked changes in excitatory and inhibitory conductances relative to the resting, unstimulated level, and can be either positive or negative. V_e and V_i are reversal potentials for excitatory and inhibitory conductances. V_i is assumed to arise from GABA$_A$- and GABA$_B$-mediated inhibition. For K$^+$-gluconate solution, V_e and V_i are 0 mV and –80 mV; for Cs$^+$-based solution (which blocks GABA$_B$ receptors), V_e and V_i are 0 mV and –70 mV.

This method has been strongly criticized on the grounds that action potentials and their underlying voltage-gated conductance changes could distort somatic measurement of synaptic conductance (Guillamon et al., 2006). In real cells, however, spikes likely originate in the first node of the axon and spike-related conductances in the soma are likely small (Yu et al., 2008). The similarity of the results with and without QX-314 and Cs$^+$ (which block voltage-gated Na$^+$ and K$^+$ channels) support this conclusion (see Figure 3 in the main text). As an additional test of the possibility that spiking conductances distort our measurements, in 14 simple cells we have made two separate estimates of conductance from two sets of records. Estimate 1 was made from the full set of records at three current levels (0 pA and two different hyperpolarizing currents). Estimate 2 was made only from the records at the two hyperpolarizing currents, which contained few or no action potentials. As shown in this figure, these two estimates are effectively identical, suggesting strongly that the presence of action potentials does not distort our estimates of synaptic conductance.

(A) Cycle-averaged firing rate responses of a simple cell (same as Figure 2 in the main text) to a blank stimulus and 4 different visual stimuli, recorded with three different levels of injected current (0 pA, black; –150 pA, cyan; –300 pA, light blue).

(B) Corresponding membrane potential responses. For each of the 3 currents, two or three sets of traces are superimposed, but are so closely identical that only one trace is easily visible. The first set of traces are the recordings themselves, color coded as in (A). These are lying underneath the second set of traces (gray), which are the reconstructions of membrane potential made from Estimate 1 of synaptic conductances (using all currents). The third set of traces (magenta) are reconstructions of membrane potential made from Estimate 2 of the synaptic conductances (excluding the 0-current recordings). If the conductances derived from Estimate 1 had been distorted by the presence of spikes in the 0-current records, the two reconstructions of membrane potential would not overlap so closely.

(C) I-V relationship constructed at the peak (top) and trough (bottom) of membrane potential responses for each stimulus. Symbols show mean and s.e.m. for the three injected currents. Linear fits are made using either all three currents (gray), or only the two hyperpolarizing currents (magenta).

(D-F) Similar to (A)-(C) for a second simple cell, which did not spike during the injection of hyperpolarizing currents (cyan and light blue).

(G) Input resistance (slope of the I-V relationship) derived from the two hyperpolarizing currents alone plotted against input resistance derived from all three currents. The 70 points represent 5 stimulus conditions for 14 cells; 7 cells were recorded with Cs$^+$ plus QX-314 in the recording electrode (squares) and 7 cells with K$^+$ (circles). Measurements were made at the time of peak membrane potential for each stimulus. There is little difference between the two estimates of resistance, again indicating that the presence of spikes has little effect on conductance measurements.

(H) Same as (G), but for trough of the membrane potential response when no action potentials were evoked for any stimulus.

(I) Ratios of 3-point and 2-point measurements of input resistance plotted against firing rate in 0-current records at the peak membrane potential. Large firing rates are not associated with a strong decrease in measured input resistance. That is, ranges of input resistance recorded with and without spikes present (that is, with K$^+$ and with Cs$^+$/QX-314) are comparable.

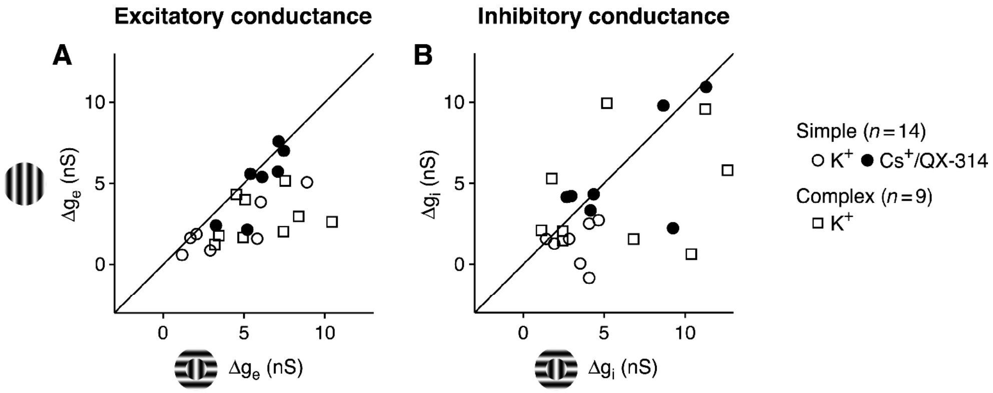

Figure S5. Orientation dependence of surround suppression of synaptic conductances

Surround suppression of both excitation and inhibition were significantly orientation tuned (one-sided paired t-test, $p < 0.0001$ for excitation; $p<0.03$ for inhibition).

(A) Change in peak excitatory conductance evoked by center-plus-surround stimulus with the surround at the preferred orientation plotted against the change in peak excitatory conductance evoked when the surround was rotated by 90 degrees. Circle and square symbols indicate simple and complex cells; open and closed symbols indicate cells recorded with K^+-gluconate and Cs^+-gluconate/methanesulfonate solution in the recording pipette.

(B) Same as (A) for inhibitory conductance.

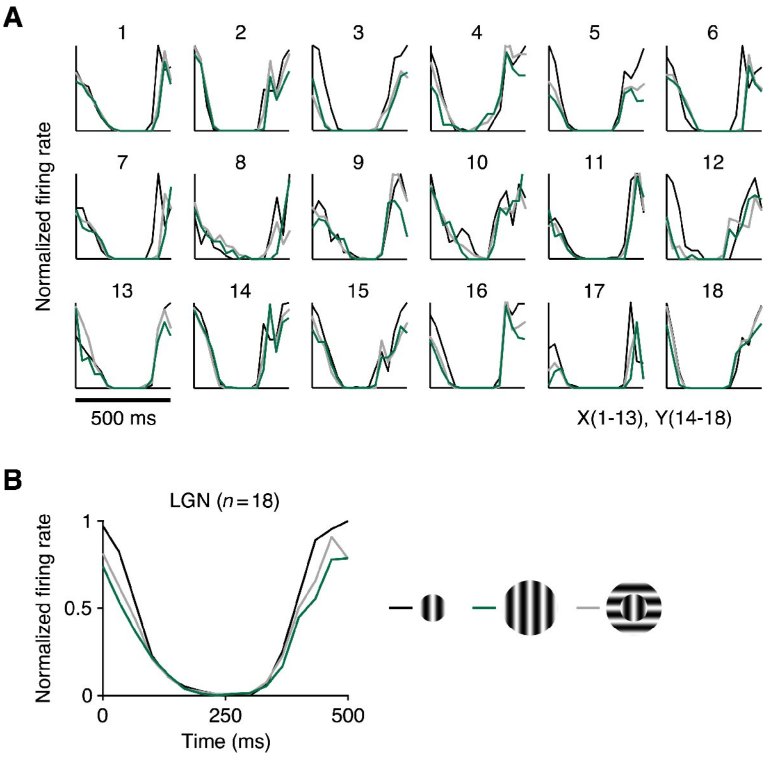

Figure S6. Surround suppression in the LGN

(A) Cycle-averaged firing rate responses of 18 LGN cells in response to center only (black) and center-plus-surround stimuli with the surround at the same orientation as the center (green) and at the orthogonal orientation (gray). Normalized to the center-only responses. The size of the center stimuli (2 or 4 degrees diameter) was optimal for cortical cells, and not for geniculate cells (see main text); the size of surround annuli (20 degrees diameter) was identical to that for cortical experiments.

(B) Responses from (A), averaged across all geniculate cells.

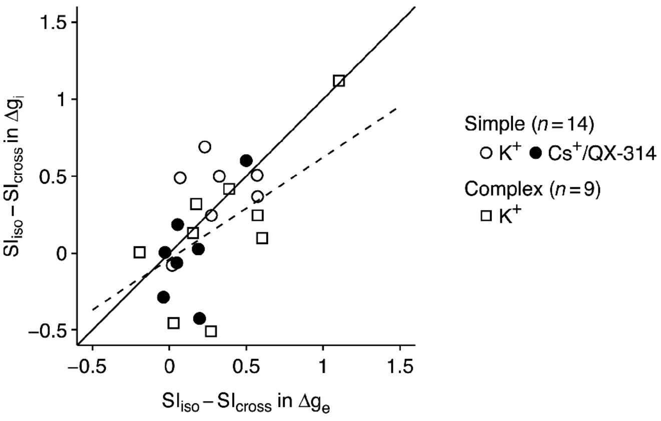

Figure S7. Orientation tuning of excitatory and inhibitory suppression

We asked whether orientation-dependent surround suppression in membrane potential is created by orientation-independent reduction of LGN input combined with orientation-dependent, suppressive intracortical input. In this scenario, the excitatory input would be maximally suppressed by the iso-oriented surround and the inhibitory input would be maximally suppressed by the cross-oriented surround. When we compared the orientation dependence of suppression in excitation and inhibition, however, the two are significantly correlated ($r = 0.46$, $p < 0.03$), rather than anti-correlated as would have been predicted by this scenario.

Here we plot the difference of suppression in the presence of the iso- and cross-oriented surround: $SI_{iso} - SI_{cross}$. Circle and square symbols indicate simple and complex cells; open and closed symbols indicate cells recorded with K^+-gluconate and Cs^+-gluconate/methanesulfonate solution in the recording pipette. Regression line: slope, 0.66; intercept, –0.04.

Supplemental Text

1 The linearized ISN model

We begin with Equations 1 and 2 of the main text (Beer, 1995, 2006; Ermentrout, 1998; Pinto et al., 1996; Tsodyks et al., 1997; Wilson and Cowan, 1972), which we repeat here for convenience:

$$\begin{aligned} \tau_E \frac{d}{dt} r_E &= -r_E + f_E\left(r_E, r_I, i_E\right) \\ \tau_I \frac{d}{dt} r_I &= -r_I + f_I\left(r_E, r_I, i_I\right) \end{aligned} . \tag{1}$$

It is worth noting that these equations can be derived starting from equations for a spiking-neuron network, and that the resulting rate-model dynamics reasonably approximate the spiking-neuron dynamics (e.g., Pinto et al., 1996; Ermentrout, 1998; Shriki et al., 2003). In the vicinity of a fixed point, these equations can in turn be well approximated by linear equations (Hirsch and Smale, 1974), which allow for more precise mathematical conclusions that apply when the perturbations away from the fixed point are sufficiently small. This section is provided for review; results are either explicit or implicit in Tsodyks et al. (1997).

1.1 The linearized equations

If we linearize the model about the fixed point, Equation 1 becomes:

$$\tau_E \frac{d}{dt} r_E = -r_E + j_{EE} r_E - j_{EI} r_I + i_E \ , \tag{2}$$

$$\tau_I \frac{d}{dt} r_I = -r_I + j_{IE} r_E - j_{II} r_I + i_I \ , \tag{3}$$

where $j_{EE} = \frac{\partial f_E}{\partial r_E}$, $-j_{EI} = \frac{\partial f_E}{\partial r_I}$, $j_{IE} = \frac{\partial f_I}{\partial r_E}$, $-j_{II} = \frac{\partial f_I}{\partial r_I}$, with all partial derivatives taken at the fixed point. The conditions on f_E and f_I – that they are increasing functions of excitatory input and decreasing functions of inhibitory input – ensure that the four j_{XY} are all >0. Note the meanings of some quantities have changed somewhat relative to Equation 1: r_E and r_I are now defined to be zero at the fixed point and so represent deviations from the fixed-point rates, rather than absolute rates. The inputs i_X ($X = E$ *or* I) are now defined to be $\partial f_X / \partial i_X$ times the deviation of the input from the fixed-point input.

Equations 2 and 3 can be re-expressed as the matrix equations:

$$\tau \mathbf{T} \frac{d}{dt} \mathbf{r} = -\mathbf{r} + \mathbf{J}\mathbf{r} + \mathbf{i} = -\left(\mathbf{1} - \mathbf{J}\right)\mathbf{r} + \mathbf{i} \ , \tag{4}$$

or

$$\tau \frac{d}{dt}\mathbf{r} = -\mathbf{T}^{-1}(\mathbf{1}-\mathbf{J})\mathbf{r} + \mathbf{T}^{-1}\mathbf{i} \ , \qquad (5)$$

where the vector of firing rates $\mathbf{r} = \begin{pmatrix} r_E \\ r_I \end{pmatrix}$, the vector of external inputs $\mathbf{i} = \begin{pmatrix} i_E \\ i_I \end{pmatrix}$, the connectivity matrix $\mathbf{J} = \begin{pmatrix} j_{EE} & -j_{EI} \\ j_{IE} & -j_{II} \end{pmatrix}$, $\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\tau_I = k\tau_E$ for some constant $k > 0$, $\tau = \tau_E$, and $\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}$.

1.2 Conditions under which the model operates as an ISN

A network is an ISN if the following two criteria are satisfied:

(a) *Excitatory instability*: the network, linearized about the fixed point, would be unstable in the absence of feedback inhibition.

(b) *Overall stability*: the strength of feedback inhibition is sufficient to stabilize the network. The network is stable if and only if both eigenvalues of $-\mathbf{T}^{-1}(\mathbf{1}-\mathbf{J})$ have real parts less than 0, and is unstable if and only if at least one eigenvalue has real part >0.

Condition (a) requires that $j_{EE} > 1$: Without feedback inhibition, the weight matrix would be $\mathbf{J} = \begin{pmatrix} j_{EE} & 0 \\ j_{IE} & -j_{II} \end{pmatrix}$, and the eigenvalues of $-\mathbf{T}^{-1}(\mathbf{1}-\mathbf{J})$would be $j_{EE} - 1$ and $-(j_{II} + 1)/k$. For a linear network, instability is equivalent to having an eigenvalue greater than 0, which is true precisely when $j_{EE} > 1$. The condition $j_{EE} > 1$ is in turn equivalent to the condition found in the main text for instability of the excitatory subnetwork, namely that the excitatory nullcline have a positive slope at the fixed point: from Equation 2, the equation for the excitatory nullcline (dr_E/dt = 0), omitting the external input (which does not alter the slope), is $(1 - j_{EE})r_E + j_{EI}r_I = 0$, giving the slope $r_I / r_E = (j_{EE} - 1)/ j_{EI}$.

Condition (b) can be restated as the eigenvalues of $\mathbf{T}^{-1}(\mathbf{1}-\mathbf{J})$ having positive real part, and, because **this matrix** is 2x2, it in turn is equivalent to two conditions: the determinant of this matrix is positive, and the trace of this matrix is positive. The determinant condition reduces to $\det(\mathbf{1}-\mathbf{J}) > 0$, or $(1 - j_{EE})(1 + j_{II}) + j_{EI}j_{IE} > 0$. The trace condition is $1 + j_{II} > k(j_{EE} - 1)$.

The equation for the inhibitory nullcline, omitting the external input, is $-j_{IE}r_E + (1+j_{II})r_I = 0$, so its slope is $r_I / r_E = j_{IE} / (1+j_{II})$. Rearranging the determinant condition, one finds $j_{IE} / (1+j_{II}) > (j_{EE}-1)/ j_{EI}$, that is, the determinant condition is equivalent to the condition that the slope of the inhibitory nullcline is greater than the slope of the excitatory nullcline. Thus, as stated in the main text, a requirement for stability of the overall network is that the inhibitory nullcline have greater (more positive) slope than the inhibitory nullcline at the fixed point, where they cross.

The trace condition is always met if the inhibitory time constant is sufficiently fast relative to the excitatory time constant, that is, if *k* is sufficiently small. In this regard, it is worth noting that the time constant in a rate model typically corresponds to the synaptic time constant rather than the cellular *RC* time constant (Ermentrout, 1994; Shriki et al., 2003). Given the strength of NMDA receptors, which have a slow time course, at intracortical excitatory synapses (Gil and Amitai, 1996), and the paucity of the slow GABA$_B$ form of inhibition in somatically-targeted inhibition (Tamas et al., 2003), the inhibitory time constant may be considerably faster than the excitatory time constant.

1.3 The fixed point for a given input

From Equation 4, the equation for the fixed point $\mathbf{r}^{\mathrm{FP}}$, where $d\mathbf{r}/dt = 0$, is $\mathbf{r}^{\mathrm{FP}} = (\mathbf{1}-\mathbf{J})^{-1}\mathbf{i}$. We can compute $(\mathbf{1}-\mathbf{J})^{-1} = \frac{1}{\det(\mathbf{1}-\mathbf{J})}\begin{pmatrix} 1+j_{II} & -j_{EI} \\ j_{IE} & 1-j_{EE} \end{pmatrix}$. As just noted, stability requires that $\det(\mathbf{1}-\mathbf{J}) > 0$. The two entries in the left column of the matrix are always positive, meaning that an increase in external input to excitatory cells raises both r_E and r_I. The upper right entry is always negative, meaning that an increase in external input to inhibitory cells lowers r_E. The sign of the lower right entry depends on whether $j_{EE} > 1$ or $j_{EE} < 1$, that is, whether the network is an ISN or a non-ISN. If it is an ISN, this term is negative, so an increase in external input to inhibitory cells lowers r_I.

2 The changes in firing rates and synaptic inputs to E and I populations after changes in external inputs

In this section, we will derive in general the changes in E cell and I cell firing rates and the changes in synaptic input (total excitation and total inhibition) received by E cells and by I cells, as a function of arbitrary changes in external inputs, and determine the key ways in which ISN and non-ISN behavior are distinct. This is an extension of the analysis presented in the main text in the section “Alternative models of the surround input”.

As in the main text, we now use a more specific form of nonlinearity, defining the nonlinear functions as $f_E\left(r_E, r_I, i_E\right)= g_E\left(w_{EE}r_E - w_{EI}r_I + i_E\right)$ and $f_I\left(r_E, r_I, i_E\right)= g_I\left(w_{IE}r_E - w_{II}r_I + i_I\right)$, where g_E and g_I are nonlinear, monotonically increasing functions of a scalar argument. We let $D_E = w_{EE}r_E - w_{EI}r_I + i_E$ be the argument of g_E , i.e. the drive to the E cells, and similarly $D_I = w_{IE}r_E - w_{II}r_I + i_I$ is the drive to the I cells. We let E_E and I_E represent the external excitatory and inhibitory input to the E cells, and similarly let E_I and I_I represent external excitatory and inhibitory input to I cells, with $i_E = E_E - I_E$ and $i_I = E_I - I_I$. We assume the network is at the center-only fixed point, and we add surround input. Changes induced by the surround input are indicated with a prefix of Δ.

In the main text, we separated Δr_I into the sum of two components (each dependent on changes in inputs to, and firing rates of, the E cells): the change associated with the vertical movement of the excitatory nullcline, and the change associated with the movement along the excitatory nullcline from the value of r_E at the center-only fixed point to the new, center-plus-surround fixed point. We can similarly determine Δr_E in terms of inputs to and firing rates of the I cells. In response to a change in external inputs to I cells, the inhibitory nullcline is moved horizontally by $\Delta r_{E1} = -\frac{\Delta E_I - \Delta I_I}{w_{IE}}$, which is the change in r_E that preserves the value of D_I for fixed r_I and thus preserves the relationship that defines the nullcline, $r_I = g_I(D_I)$. The change in r_E associated with movement along the new inhibitory nullcline from the center-only value of r_I to the new fixed point is $\Delta r_{E2} = \frac{\Delta r_I}{\kappa_I}$ where κ_I is the average slope of the inhibitory nullcline between the two values of r_I (the values at the old and new fixed points). Letting κ_E be the average slope of the excitatory nullcline between the two values of r_E (which was simply called κ in the main text), we arrive at the two equations

$$\Delta r_E = -\frac{\Delta E_I - \Delta I_I}{w_{IE}} + \frac{\Delta r_I}{\kappa_I} \tag{6}$$

$$\Delta r_I = \frac{\Delta E_E - \Delta I_E}{w_{EI}} + \kappa_E \Delta r_E \tag{7}$$

These equations in turn can be solved to give Δr_E and Δr_I as functions of the changes in external inputs to the circuit:

$$\Delta r_E = \frac{\kappa_I}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{\kappa_I w_{EI}} - \frac{\Delta E_I - \Delta I_I}{w_{IE}}\right] \tag{8}$$

$$\Delta r_I = \frac{\kappa_I}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{w_{EI}} - \frac{\kappa_E(\Delta E_I - \Delta I_I)}{w_{IE}}\right] \tag{9}$$

We know that $\kappa_I > 0$, while κ_E is positive for an ISN and negative for a non-ISN. All of the w's are positive. Thus, we see the expected results, that in any network, a decrease in external drive to E cells (that is, in $\Delta E_E - \Delta I_E$) lowers both r_E and r_I, and an increase in external drive to I cells (in $\Delta E_I - \Delta I_I$) lowers r_E; whereas an increase in external drive to I cells raises r_I in a non-ISN and lowers r_I in an ISN.

Note also that the factor $\frac{\kappa_I}{\kappa_I - \kappa_E}$ is >1 in an ISN and <1 in a non-ISN. From this relationship we can derive the two "paradoxical" ISN synaptic input properties (and non-paradoxical non-ISN properties) discussed in the text. Equation 8 implies that an increase in external excitation to I cells of ΔE_I causes a decrease of feedback excitation to I cells, $w_{IE}\Delta r_E$, that is larger in magnitude than ΔE_I in an ISN (i.e., the total excitation received by I cells is reduced), but smaller in magnitude than ΔE_I in a non-ISN (total excitation received by I cells is increased). This underlies the paradoxical result that addition of external excitation to I cells lowers r_I in an ISN, while it raises it in a non-ISN (Tsodyks et al., 1997). Similarly, Equation 9 shows that an increase in external inhibition to E cells, ΔI_E, causes a decrease of feedback inhibition to E cells, $w_{EI}\Delta r_I$, that is larger in magnitude than ΔI_E in an ISN, reducing the total inhibition received by E cells, but is smaller in magnitude in a non-ISN, raising the total inhibition received by E cells. In both ISN and non-ISN, however, such an increase in ΔI_E causes a decrease in r_E. (In an ISN, r_E can decrease despite a net decrease of inhibition, because this decrease is balanced by the excess withdrawal of recurrent excitation, beyond the level that would just support the decrease in r_E, as discussed in the main text.)

Equations 8 and 9 are not closed-form solutions for Δr_E and Δr_I, because κ_E depends on Δr_E (which defines the region over which the nullcline slope is averaged) and similarly κ_I depends on Δr_I. Rather, they are self-consistent equations for Δr_E and Δr_I. Our conclusions, however, only depend on the signs of κ_E and κ_I. Thus, as long as we assume that the sign of the slope of the E nullcline does not change between the two fixed-point values of r_E, our conclusions will hold. Alternatively, we can allow arbitrary changes in the E nullcline slope, and simply define an ISN as a network in which κ_E, the average E nullcline slope between the two fixed-point values of r_E, is positive and a non-ISN as one in which it is negative, and our conclusions will hold for networks so defined.

We next consider the surround-induced changes in the total excitatory and total inhibitory input received by the cells. We define $\Delta[E \to E]$ to be the change in total excitation received by excitatory cells, given by $\Delta[E \to E] = \Delta E_E + w_{EE}\Delta r_E$; and $\Delta[I \to E]$ to be the change in total inhibition received by excitatory cells, $\Delta[I \to E] = \Delta I_E + w_{EI}\Delta r_I$ ($\Delta[I \to E]$ was referred to as ΔI^{TOT} in the main text). That is, the excitation and inhibition received are defined to be the corresponding elements inside the nonlinearity, while the nonlinearity converts this input to target firing rate. Then, using equations 6-9, we find

$$\Delta[E \to E] = \Delta E_E + w_{EE}\Delta r_E \tag{10}$$

$$= \Delta E_E + w_{EE}\frac{\kappa_I}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{\kappa_I w_{EI}} - \frac{\Delta E_I - \Delta I_I}{w_{IE}}\right] \tag{11}$$

$$\Delta[I \to E] = \Delta I_E + w_{EI}\Delta r_I = \Delta E_E + w_{EI}\kappa_E\Delta r_E \tag{12}$$

$$= \Delta E_E + \kappa_E w_{EI}\frac{\kappa_I}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{\kappa_I w_{EI}} - \frac{\Delta E_I - \Delta I_I}{w_{IE}}\right] \tag{13}$$

Comparing equations 10 and 12, or 11 and 13, we see that $\Delta[X \to E]$, where X is a variable that takes the values *E* or *I*, is given by $\Delta[X \to E] = \Delta E_E + A_X$, where the additional term A_X has the same sign for $X = E$ and $X = I$ in an ISN, but opposite signs in a non-ISN. Thus, in a non-ISN, for $\Delta[E \to E]$ and $\Delta[I \to E]$ to both be negative, as observed, it must be the case that ΔE_E is negative and is larger in absolute value than $\max(A_E, A_I)$. We also see that changes in all external inputs except E_E – that is, changes in any external drive to I cells or in inhibition to E cells – together make a contribution that, in a non-ISN, has opposite signs for $\Delta[E \to E]$ and $\Delta[I \to E]$, but in an ISN has the same sign for $\Delta[E \to E]$ and $\Delta[I \to E]$. ΔE_E always contributes with the same sign to both $\Delta[E \to E]$ and $\Delta[I \to E]$, in both ISNs and non-ISNs (for $\Delta[I \to E]$ for a non-ISN, ΔE_E is multiplied by $1 + \frac{\kappa_E}{\kappa_I - \kappa_E} = 1 - \frac{|\kappa_E|}{\kappa_I + |\kappa_E|} > 0$).

Similarly, we define $\Delta[E \to I] = \Delta E_I + w_{IE}\Delta r_E$ and $\Delta[I \to I] = \Delta I_I + w_{II}\Delta r_I$, and find

$$\Delta[E \to I] = \Delta E_I + w_{IE}\Delta r_E = \Delta I_I + \frac{w_{IE}}{\kappa_I}\Delta r_I$$

$$= \Delta I_I + \frac{w_{IE}}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{w_{EI}} - \kappa_E\frac{\Delta E_I - \Delta I_I}{w_{IE}}\right]$$

$$\Delta[I \to I] = \Delta I_I + w_{II}\Delta r_I$$

$$= \Delta I_I + \frac{w_{II}\kappa_I}{\kappa_I - \kappa_E}\left[\frac{\Delta E_E - \Delta I_E}{w_{EI}} - \kappa_E \frac{\Delta E_I - \Delta I_I}{w_{IE}}\right]$$

Here we see the expected results that decreasing the external drive to E cells decreases both the total excitation and the total inhibition that I cells receive (because it decreases both r_E and r_I, Equations 8-9), and that increasing external excitation to I cells increases both the total excitation and the total inhibition I cells receive in a non-ISN, but lowers both the total excitation and the total inhibition I cells receive in an ISN. Thus, for I cells as for E cells, addition of external excitation to the local circuit can cause a decrease in both the inhibition and the excitation the cells receive in an ISN but not in a non-ISN. Decreasing external inhibition to I cells also lowers both the total excitation and the total inhibition I cells receive in an ISN, and decreases the total excitation received by I cells in a non-ISN, but can either raise or lower the total inhibition received by I cells in a non-ISN.

3 Multi-neuron ISN model

We have thus far considered models that describe only the mean dynamics of the excitatory and inhibitory cell populations. Here, we develop a model with the same mean behavior, but in which each population has many neurons with varied connectivity. We find that the variability in behavior among the individual neurons mimics, at least qualitatively, what is seen in experiments. For a network of N excitatory and N inhibitory neurons, the dynamics are described by the matrix equation:

$$\tau \frac{d}{dt}\mathbf{r}_{\mathrm{N}} = -\mathbf{r}_{\mathrm{N}} + \mathbf{W}_{\mathrm{N}}\mathbf{r}_{\mathrm{N}} + \mathbf{i}_{\mathrm{N}} \, . \tag{21}$$

Here, $\mathbf{r}_{\mathrm{N}} = \begin{pmatrix} r_E^1 \\ \vdots \\ r_E^{\mathrm{N}} \\ r_I^1 \\ \vdots \\ r_I^{\mathrm{N}} \end{pmatrix}$ is the 2N-dimensional vector of firing rates at a given time (the superscripts indicate the neuron's identity among the N excitatory or inhibitory neurons). $\mathbf{W}_{\mathrm{N}}$ is the 2N-dimensional connectivity matrix, which is composed of 4 N×N blocks: the top left, top right, bottom left, and bottom right blocks correspond to the $E \to E$, $I \to E$, $E \to I$, and $I \to I$ connections. The Experimental Procedures in the main text explain how the matrix $\mathbf{W}_{\mathrm{N}}$ and the inputs $\mathbf{i}_N$ were generated. Note that if N = 1 and the variance in weights and inputs is zero, Equation 21 becomes equivalent to the 2-dimensional model with $\tau_E = \tau_I$. For simplicity, we make τ equal for all cells, which does not affect the fixed point (though it can affect its stability). It can be shown analytically that with no variance in the weight matrix (all entries within a block identical), the stability of this model and the mean response of each population at the fixed point is identical to that of the 2-d model whose four weights are given by N times the values of the weights in the corresponding block of the 2N-d model, and inputs to E and I populations given by the mean of inputs to E cells and to I cells, respectively, in the 2N-d model. Thus, we can put the mean model behavior roughly where we want it by finding a 2-d model with the behavior we desire, and then choosing the mean weight in each block and the mean inputs to E and I cells so that the 2N-d model, if weights were uniform, would have fixed point and mean response given by this 2-d model.

Adding variance to the weights (see Experimental Procedures in the main text) can in principle cause the model to lose stability. To help prevent this, the sum of excitatory and sum of inhibitory weights to each cell were preserved (described in Experimental Procedures), and we verified (by numerical computation of eigenvalues) that with this procedure stability was maintained.

4 Conditions under which the cortex may operate in the ISN regime

Latham et al. (2000) (Appendix B) argued that cortical fixed points with firing rates larger than a fraction of a Hz, i.e., most cortical activity states, should exist on the positive-sloping portion of the excitatory nullcline. This calculation relied on two uncertain assumptions. First, the gain of cortical cells was set to a value observed in responses to suprathreshold DC current injections in slices (McCormick et al., 1985). Because cortical cells *in vivo* typically fire in response to voltage fluctuations starting from a mean potential below threshold (Anderson et al., 2000), their gain is lower than the gain measured above threshold, and lower still at low firing rates (Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002). Second, a value must be chosen for N_{EE}, the average number of neurons in the excitatory assembly from which a single neuron in the assembly receives excitatory input (convergence). If the assembly's firing rate is r_E, then the average rate of unitary EPSPs originating from within the assembly is $N_{EE}r_E$. Letting V_{EPSP} be the amplitude of unitary EPSPs, Latham et al. (2000) assumed that $N_{EE}V_{EPSP} \geq 1000$ mV, e.g., N_{EE} = 2000 for V_{EPSP} = 1/2 mV. While cat V1 neurons receive 1000s of excitatory synapses (Beaulieu and Colonnier, 1985), they receive excitatory inputs from a smaller number of distinct neurons and a still smaller number from within the local circuit. It is estimated that within rat whisker barrel cortex each layer 4 excitatory neuron receives input from only about 200 other layer 4 excitatory neurons (Lubke et al., 2003). Furthermore it is not clear how many of these participate in the equivalent of a single excitatory neural assembly, as in Figure 6A (main text). For example, during stimulus-driven activity in V1, such an assembly might be restricted to cells of similar preferred orientation.

We can repeat the calculation of Latham et al. (2000) with an *in vivo* gain function as follows. We assume $f_E = k(V - V_0)^\alpha$ for $V > V_0$, where V_0 is the resting potential, and k and α are constants, with $V = w_{EE} r_E - w_{EI} r_I + i_E$ (Anderson et al., 2000; Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002; Priebe et al., 2004). Assuming inhibition maintains stability, the cortex becomes an ISN when $\partial f_E / \partial r_E > 1$ (with the derivative taken at the fixed point). We compute $\partial f_E / \partial r_E = \alpha w_{EE} k^{1/\alpha} f_E^{\frac{\alpha-1}{\alpha}}$. We can replace f_E with r_E on the right-hand side, because $r_E = f_E$ on the excitatory nullcline and so in particular at the fixed point. Thus, the requirement $\partial f_E / \partial r_E > 1$ becomes $r_E > \left(\frac{1}{\alpha w_{EE} k^{1/\alpha}} \right)^{\frac{\alpha}{\alpha-1}}$. As representative numbers for cortical gain, we choose k = .0075Hz/mV^3, α = 3, which produces a 60 Hz response for a 20 mV depolarization and 104 Hz for 24 mV (compare Figure 2 of Priebe et al., 2004). w_{EE} is the change in mean voltage produced per change in r_E. If individual EPSP's are 1/2 mV in amplitude, with a time constant of 10 ms, then $w_{EE} = N_{EE}(5\text{ mV}\cdot\text{ms}) = \frac{N_{EE}\text{ mV}}{200\text{ Hz}}$. Using these numbers, we obtain the estimate that cortex should operate as an ISN for $r_E > \frac{6300\text{ Hz}}{N_{EE}^{1.5}}$. For N_{EE} = 25, 50, 100, 200, 400, or 800, cortex should operate as an ISN at fixed points that have excitatory firing rates greater than 50, 18, 6.3, 2.2, 0.8, or 0.3 Hz. It is therefore plausible that V1 operates as an ISN in some or all physiologically relevant ranges. Given our lack of knowledge of the value of N_{EE}, however, the calculation cannot go beyond demonstrating plausibility; the question must be decided by empirical evidence, such as we present here.

References

Anderson, J.S., Lampl, I., Gillespie, D.C., and Ferster, D. (2000). The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. Science *290*, 1968-1972.

Anderson, J.S., Lampl, I., Gillespie, D.C., and Ferster, D. (2001). Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. J. Neurosci. *21*, 2104-2112.

Beaulieu, C., and Colonnier, M. (1985). A laminar analysis of the number of round-asymmetrical and flat-symmetrical synapses on spines, dendritic trunks, and cell bodies in area 17 of the cat. J. Comp. Neurol. *231*, 180-189.

Beer, R.D. (1995). On the dynamics of small continuous-time recurrent neural networks. Adapt. Behav. *3*, 469-509.

Beer, R.D. (2006). Parameter space structure of continuous-time recurrent neural networks. Neural Comput. *18*, 3009-3051.

Bringuier, V., Chavane, F., Glaeser, L., and Fregnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science *283*, 695-699.

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.

Ermentrout, B. (1994). Reduction of conductance based models with slow synapses to neural nets. Neural Comput. *6*, 679–695

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. Rep. Prog. Phys. *61*, 353-430.

Gil, Z., and Amitai, Y. (1996). Adult thalamocortical transmission involves both NMDA and non-NMDA receptors. J. Neurophysiol. *76*, 2547-2554.

Guillamon, A., McLaughlin, D.W., and Rinzel, J. (2006). Estimation of synaptic conductances. J. Physiol. (Paris) *100*, 31-42.

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.

Hirsch, M.W., and Smale, S. (1974). Differential Equations, Dynamical Systems, and Linear Algebra (San Diego: Academic Press).

Latham, P.E., Richmond, B.J., Nelson, P.G., and Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. J. Neurophysiol. *83*, 808-827.

Lubke, J., Roth, A., Feldmeyer, D., and Sakmann, B. (2003). Morphometric analysis of the columnar innervation domain of neurons connecting layer 4 and layer 2/3 of juvenile rat barrel cortex. Cereb. Cortex *13*, 1051-1063.

McCormick, D.A., Connors, B.W., Lighthall, J.W., and Prince, D.A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J. Neurophysiol. *54*, 782-806.

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.

Moore, C.I., and Nelson, S.B. (1998). Spatio-temporal subthreshold receptive fields in the vibrissa representation of rat primary somatosensory cortex. J. Neurophysiol. *80*, 2882-2892.

Pinto, D.J., Brumberg, J.C., Simons, D.J., and Ermentrout, G.B. (1996). A quantitative population model of whisker barrels: re-examining the Wilson-Cowan equations. J. Comput. Neurosci. *3*, 247-264.

Priebe, N.J., Mechler, F., Carandini, M., and Ferster, D. (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nat. Neurosci. *7*, 1113-1122.

Sceniak, M.P., Ringach, D.L., Hawken, M.J., and Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. Nat. Neurosci. *2*, 733-739.

Sengpiel, F., Baddeley, R.J., Freeman, T.C., Harrad, R., and Blakemore, C. (1998). Different mechanisms underlie three inhibitory phenomena in cat area 17. Vision Res. *38*, 2067-2080.

Shriki, O., Hansel, D., and Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. Neural Comput. *15*, 1809-1841.

Tamas, G., Lorincz, A., Simon, A., and Szabadics, J. (2003). Identified sources and targets of slow inhibition in the neocortex. Science *299*, 1902-1905.

Tan, A.Y., Zhang, L.I., Merzenich, M.M., and Schreiner, C.E. (2004). Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J. Neurophysiol. *92*, 630-643.

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

Wilson, H.R., and Cowan, J.D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. *12*, 1-24.

Yu, Y., Shu, Y., and McCormick, D.A. (2008). Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics. J. Neurosci. *28*, 7260-7272.