Systems/Circuits

Inhibitory Interneurons Decorrelate Excitatory Cells to Drive Sparse Code Formation in a Spiking Model of V1

Paul D. King,[1,2] **Joel Zylberberg,**[1,3] **and Michael R. DeWeese**[1,2,3]

[1]Redwood Center for Theoretical Neuroscience, [2]Helen Wills Neuroscience Institute, and [3]Department of Physics, University of California, Berkeley, California 94720

Sparse coding models of natural scenes can account for several physiological properties of primary visual cortex (V1), including the shapes of simple cell receptive fields (RFs) and the highly kurtotic firing rates of V1 neurons. Current spiking network models of pattern learning and sparse coding require direct inhibitory connections between the excitatory simple cells, in conflict with the physiological distinction between excitatory (glutamatergic) and inhibitory (GABAergic) neurons (Dale's Law). At the same time, the computational role of inhibitory neurons in cortical microcircuit function has yet to be fully explained. Here we show that adding a separate population of inhibitory neurons to a spiking model of V1 provides conformance to Dale's Law, proposes a computational role for at least one class of interneurons, and accounts for certain observed physiological properties in V1. When trained on natural images, this excitatory–inhibitory spiking circuit learns a sparse code with Gabor-like RFs as found in V1 using only local synaptic plasticity rules. The inhibitory neurons enable sparse code formation by suppressing predictable spikes, which actively decorrelates the excitatory population. The model predicts that only a small number of inhibitory cells is required relative to excitatory cells and that excitatory and inhibitory input should be correlated, in agreement with experimental findings in visual cortex. We also introduce a novel local learning rule that measures stimulus-dependent correlations between neurons to support "explaining away" mechanisms in neural coding.

Introduction

Sparse coding has emerged as a useful principle for understanding neural representations in the cortex. In vision, computing a sparse representation of natural images identifies visual features that match the Gabor-like receptive fields (RFs) found in primary visual cortex (V1) (Olshausen and Field, 1996; Bell and Sejnowski, 1997). These models use mathematical methods such as conjugate gradient descent and independent component analysis, but how could the brain achieve this with spiking neural circuits relying solely on local synaptic plasticity rules?

Recent work has demonstrated (Zylberberg et al., 2011) how a network of spiking neurons using local Hebbian synaptic plasticity rules can learn a sparse code from natural scenes. This sparse and independent local network (SAILnet) learns Gabor-like RFs that closely match those of V1 simple cells. Like previous models of pattern learning and coding in V1 (Masquelier et al., 2009; Savin et al., 2010; Masquelier, 2012), SAILnet relies on excitatory simple cells that laterally inhibit each other directly, in conflict with the observation that cortical neurons are either excitatory (glutamatergic) or inhibitory (GABAergic) but not both ("Dale's Law") (Eccles, 1976).

Most prior network models with V1-like response properties have relied on analytically derived connection weights. For example, the locally competitive algorithm (LCA) can perform sparse "inference" (i.e., determine appropriate neural activities to sparsely represent the input for a fixed set of RFs) in a nonspiking network model (Rozell et al., 2008), and some spiking networks can as well (Shapero et al., 2011; Hu et al., 2012), but the connection weights must be precomputed using nonlocal methods in all of these cases. A recent extension of LCA makes use of separate excitatory and inhibitory network nodes, but the nodes are used for inference only, not learning, and they do not spike (Zhu et al., 2012). A network model of orientation selectivity in V1 that does use spiking excitatory and inhibitory neurons also precomputes the connection strengths rather than learning them (McLaughlin et al., 2000).

Here we present E-I Net, a spiking network model of leaky integrate and fire (LIF) neurons that extends the SAILnet model to learn a sparse code without violating Dale's Law. E-I Net contains separate populations of excitatory and inhibitory neurons that work together to learn a sparse representation. The inhibitory cells provide feedback inhibition to the excitatory cells using a novel local learning rule that modifies the synaptic weights so that the inhibitory cells send an amount of inhibitory current to the excitatory cells proportional to the expected number of spikes from those excitatory cells. Thus, the redundant part of the network activity, which can be predicted from the past, is cancelled out, similar to the function of predictive coding (Rao and Ballard,

Received Aug. 23, 2012; revised Jan. 8, 2013; accepted Jan. 10, 2013.

Author contributions: P.D.K., J.Z., and M.R.D. designed research; P.D.K. performed research; J.Z. contributed unpublished reagents/analytic tools; P.D.K. analyzed data; P.D.K., J.Z., and M.R.D. wrote the paper.

P.D.K. is grateful to Fritz Sommer and Bruno Olshausen for sponsorship at the Redwood Center, and to Jascha Sohl-Dickstein and the other members of the Redwood Center for many helpful discussions. J.Z. is an international student research fellow of the Howard Hughes Medical Institute. M.R.D. gratefully acknowledges support from the McKnight Foundation, the Hellman Family Faculty Fund, the McDonnell Foundation, and the Mary Elizabeth Rennie Endowment for Epilepsy Research.

The authors declare no competing financial interests.

Correspondence should be addressed to Paul D. King, Visiting Scholar, Helen Wills Neuroscience Institute, University of California, 132 Barker Hall MC 3190, Berkeley, CA 94720-3190. E-mail: paul@pking.org.

DOI:10.1523/JNEUROSCI.4188-12.2013

 0270-6474/13/335475-11$15.00/0

1999), or the "explaining away" feature of many sparse coding models (Rehn and Sommer, 2007; Lochmann and Deneve, 2011; Lochmann et al., 2012). E-I Net's inhibitory population performs a similar function to SAILnet's lateral inhibitory connections, which is to decorrelate the activity of the excitatory cells by suppressing redundant spiking activity (Zylberberg et al., 2011).

Materials and Methods

The model presented here is based on the SAILnet model (Zylberberg et al., 2011), but extended to include a separate population of inhibitory neurons, a form of spike-timing-dependent plasticity (STDP) (Bi and Poo, 1998; Abbott and Nelson, 2000; Dan and Poo, 2004; Feldman, 2009), and a novel local form of Hebbian learning (Hebb, 1949).

The network consists of two populations of LIF neurons. The first population, a set of 400 excitatory cells, receives sensory input from 10×10 pixel image patches drawn from Olshausen and Field's (1996) database of whitened images of natural scenes. The spike rate output of these cells constitutes the sparse representation of the image patch. The second population is a set of 49 inhibitory neurons (number chosen for display in a square grid). These cells receive input from the excitatory cells and send inhibition back to those same cells. The inhibitory cells also inhibit each other. The network is fully connected in that all excitatory cells connect to all inhibitory cells, and all inhibitory cells connect to each other.

To process an image for either training or readout, the membrane potentials are set to zero and then an image patch is presented. The network simulation runs for 50 time steps. During each time step, inputs to each neuron are weighted by synaptic strength and either added (excitatory) or subtracted (inhibitory) from the neuron's membrane potential. If the membrane potential crosses a cell-specific firing threshold, the cell spikes and the membrane potential is reset to zero. At the end of the simulation, the number of spikes generated by each excitatory cell constitutes the readout of the network.

One can think of the sum of these E cell spike counts multiplied by their respective RFs as a linear generative model of the whitened visual input to the network. We note that other interesting objectives are possible, such as maximizing the information shared between the input and the resulting network activity (Rieke et al., 1997; Karklin and Simoncelli, 2011).

Network spiking dynamics. E-I Net's simulation model is a generalization of SAILnet. Multiple neuron classes C can each receive either excitatory ($\beta = +1$) or inhibitory ($\beta = -1$) input from any or all neuron classes.

For each simulation time step t, and for each neuron i of class C, the neuron state is updated as follows:

$$u_i^{(C)}(t+1) = u_i(t)\exp(-\eta/\tau^{(C)}) + \sum_{C^*}\beta^{(C^*\to C)}\sum_j z_j^{(C^*)}(t)W_{ij}^{(C^*\to C)}$$

$$z_i^{(C)}(t+1) = \begin{cases} 1, & u_i^{(C)}(t+1) \geq \theta_i^{(C)} \\ 0, & \text{otherwise} \end{cases}$$

$$u_i^{(C)}(t+1) \Leftarrow 0 \quad \text{iff} \quad z_i^{(C)}(t+1) = 1. \tag{1}$$

The variables are as follows:

C is the neuron class and is one of the two populations in our model, excitatory cells (E) and inhibitory cells (I), or as a special case for input only, the image patch pixels values (in);

$u_i^{(C)}(t)$ is the membrane potential of neuron i of neuron class C at time t;

$z_i^{(C)}(t)$ is the spike output of neuron i of neuron class C at time t (either 0 for no spike, or 1 for spike);

η is the simulation time step size in arbitrary simulation time units (0.1 arbitrary time units here);

$\tau^{(C)}$ is the membrane time constant governing the membrane potential decay rate for neurons of class C;

$\theta_i^{(C)}$ is the spike threshold of neuron i of neuron class C;

$W_{ij}^{(C_2\to C_1)}$ is the connection weight from neuron j of class C_2 to neuron i of class C_1; and

$\beta^{(C_2\to C_1)}$ is the sign of the impact of class C_2 neurons on class C_1 neurons: $+1$ for excitatory connections and -1 for inhibitory connections.

The input image patch is represented as graded values rather than spikes. X_i represents the value of the whitened image patch at pixel i, which may be positive or negative. The following rule is used to convert X_i into a suitable input value $z_i^{(in)}(t)$, which can be viewed as the aggregate contrast information at that point in visual space summarized as a current injection introduced into the neuron over time as follows:

$$z_i^{(in)}(t) = \eta X_i. \tag{2}$$

To simulate temporal dynamics, several constants were used. The simulation proceeded for 5 time units of 10 time steps each, for a total of 50 time steps. To compute the moving average spike rate as input to the Hebbian plasticity rules, we used a moving average with exponential decay time constant of 1 simulation time unit (10 time steps). To match the behavior of the published SAILnet model, we used the following scaling constants on the input image patches: $X_i = (1/5)\text{pixel}_i$, where pixel_i is the value of the ith pixel after the whitened image patch has been normalized to zero mean and unit variance; and $\beta^{(in\to E)} = 5$.

Homeostatic spike rate regulation. As with SAILnet, the threshold at which a neuron fires is adjusted up or down according to a threshold adaptation rule, originally from Földiák (1990) to achieve a target spike rate over the long term that is set in advance as a network parameter:

$$\Delta\theta_i^{(C)} \propto \langle z_i^{(C)}\rangle - p^{(C)}, \tag{3}$$

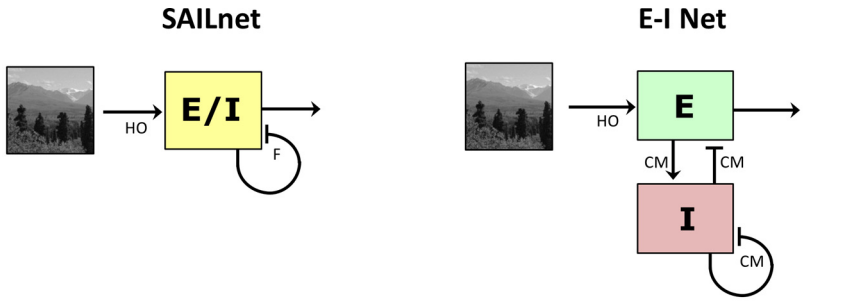
where $p^{(C)}$ is the target mean spike rate for neurons of class C. For our simulation, we used spike rates of $p^{(E)} = 0.02$ and $p^{(I)} = 0.04$ spikes per time unit. The membrane time constants controlling the decay of the membrane potential to a baseline of zero, which worked best when faster spike rates were paired with faster time constants, were $\tau^{(E)} = 1$ and $\tau^{(I)} = 0.5$.

Training and learning rules. To train the network, image patches are presented one at a time for network simulation. Both excitatory and inhibitory weights are updated according to two Hebbian plasticity rules. The weights from the image patch to the excitatory cells, $W^{(in\to E)}$, are updated according to Oja's variant of the Hebbian learning rule (Oja, 1982), labeled "HO." All remaining weights, $W^{(E\to I)}$, $W^{(I\to E)}$, and $W^{(I\to I)}$, learn using the Correlation Measuring rule introduced here and labeled "CM." The Correlation Measuring rule is inspired by Földiák's rule (Földiák, 1990) in use in SAILnet and shown here for comparison (labeled "F"). These rules are shown below for comparison as follows:

$$\begin{aligned} \text{HO: } \Delta W_{ij} &\propto y_i x_j - y_i^2 W_{ij}, \\ \text{CM: } \Delta W_{ij} &\propto y_i x_j - \langle y_i\rangle\langle x_j\rangle(1 + W_{ij}), \\ \text{F: } \Delta W_{ij} &\propto y_u x_j - \langle y_i\rangle\langle x_j\rangle. \end{aligned} \tag{4}$$

In the equations above, x_j refers to the spike rate of presynaptic (input) neuron j, and y_i represents the spike rate of postsynaptic (output) neuron i. The spike rates are a moving average of the individual spikes over time, $\langle z_i^{(C)}\rangle_{\Delta t}$, where Δt represents the temporal window of the moving average weighted with exponential decay. Weight changes are computed on each time step in a simulation of symmetric STDP (Bi and Poo, 1998; Dan and Poo, 2004; Feldman, 2009), although using sample-averaged spike rates computed once per sample produces similar results. Note that the STDP used here for inhibitory neurons is independent of pre-post spike order, a type of plasticity that, interestingly, has been observed in GABAergic neurons in hippocampus (Abbott and Nelson, 2000). Weights adjusted with the CM rule are further constrained to be non-negative. The following learning rates were used: $\alpha^{(in\to E)} = 0.008$; $\alpha^{(E\to I)} = 0.028$; $\alpha^{(I\to E)} = 0.028$; $\alpha^{(I\to I)} = 0.06$. To stabilize network behavior during training, weight changes are accumulated separately and applied in aggregate after every 100 image patch training samples.

We evaluated the performance of the network after training using primarily two measures: root mean square (RMS) reconstruction error



Learning Rules		Symbol	Equilibrium
Hebbian-Oja (HO)	$\Delta W_{ij} \propto y_i x_j - y_i^2 W_{ij}$	E —HO→	$W_{ij} = \frac{\langle y_i x_j \rangle}{\langle y_i^2 \rangle} = \frac{STA_i}{\langle y_i^2 \rangle}$
Földiák (F)	$\Delta W_{ij} \propto y_i x_j - \langle y_i \rangle \langle x_j \rangle$	I —F⊣	$\langle y_i x_j \rangle = \langle y_i \rangle \langle x_j \rangle$
Correlation Measuring (CM)	$\Delta W_{ij} \propto y_i x_j - \langle y_i \rangle \langle x_j \rangle (1 + W_{ij})$	I —CM⊣	$W_{ij} = \frac{\langle y_i x_j \rangle}{\langle y_i \rangle \langle x_j \rangle} - 1$
Threshold Adaptation	$\Delta \theta_i \propto y_i - p$	[not shown]	$\langle y_i \rangle = p$

Figure 1. Circuit diagram for SAILnet (Zylberberg et al., 2011) (top left) and the E-I Net introduced here (top right). Excitatory connections are labeled with arrows, inhibitory connections with flat ends. A letter code identifies the learning rule used for the synaptic connections as either HO, F generalized to use measured spike rates, or the CM proposed here. $\langle x \rangle$ represents the lifetime average value of x. The output of both networks is a spike train embodying a sparse representation of the input. Bottom table, Each learning rule is shown with its connection weight update equation and the equilibrium end state that the learning rule seeks out during training. In the equations, x_j is the presynaptic spike rate, y_i is the postsynaptic spike rate, and W_{ij} is the connection weight from presynaptic neuron j to postsynaptic neuron i. STA_i is the STA of postsynaptic neuron i.

and RMS of the pairwise Pearson correlation coefficients within the E cell population. The RMS of the residual reconstruction error was calculated by first reconstructing the image patch as the sum of the input→E weights multiplied by the respective E cell spike rates. This reconstructed image patch was then normalized to unit SD to match the variance of the input patches. Our RMS residual error measure is then the RMS of the difference between original image patch pixels and reconstructed pixels. The E cell correlation measure was computed as the RMS of the Pearson correlation coefficient between all cell pairs over 100 input samples.

The state of network equilibrium was reached when the RMS of the weight change across 10,000 training samples stopped decreasing and reached a steady state.

Correlation-measuring learning rule sends inhibition to suppress predicted spikes. The amount of inhibition received by an E cell is determined by the CM learning rule. This learning rule strengthens connections between positively correlated cells and weakens them if cells are anticorrelated, converging on a weight value that approximately measures the degree of spiking correlation between the neurons. The equilibrium point reached by this rule, determined by assuming $\Delta W_{jk} = 0$, is as follows:

$$W_{jk} = \frac{\langle n_j m_k \rangle}{\langle n_j \rangle \langle m_k \rangle} - 1, \quad (5)$$

where n_j and m_k are moving averages of the spike output of E cell j and I cell k, respectively, over a short temporal window Δt. The CM learning rule can be shown to perform steepest gradient descent toward this equilibrium point by assuming minimization of the mean squared error of the weight with respect to the equilibrium point. This equilibrium can be rewritten as follows:

$$W_{jk} = \frac{\langle n_j m_k \rangle - \langle n_j \rangle \langle m_k \rangle}{\langle n_j \rangle \langle m_k \rangle}. \quad (6)$$

Because the mean spike rates are homeostatically regulated to a fixed value, they can be regarded as constant and incorporated into a constant of proportionality, resulting in the following:

$$W_{jk} \propto \langle n_j m_k \rangle - \langle n_j \rangle \langle m_k \rangle. \quad (7)$$

Thus the weights learned by the CM rule are proportional to the covariance between the neurons. The weight will be zero if the neurons are uncorrelated (or anticorrelated) and will grow linearly as the degree of correlation increases.

The net result is that when an I cell spikes, it sends more inhibition to those E cells whose firing rates are more strongly correlated with that I cell's own firing rate.

If one seeks to construct the optimal linear estimator of the firing rates of the excitatory cells (with indices j), using the concurrent inhibitory cell firing rates (with indices k), the solution is to multiply the vector of inhibitory cells' activities (with elements m_k) by the matrix W_{jk} of the (scaled) covariances between the excitatory and inhibitory cell activities, in the case where the inhibitory cell activities are uncorrelated (Salinas and Abbott, 1994). The elements of the matrix W_{jk} are the covariances between the firing rates of excitatory cell j and inhibitory cell k: $W_{jk} \propto \mathrm{cov}(n_j, m_k)$ (Eq. 7).

If the inhibitory cells' firing rates were strongly correlated with each other (i.e., inhibitory–inhibitory correlations are large), then the optimal linear estimator makes use of the inverse of the (inhibitory–inhibitory) covariance matrix. In our model, the direct recurrent inhibition between the inhibitory cells prevents them from being strongly correlated with one another, and the learning rules ensure that the weights $W_{jk}^{(I\to E)}$ are proportional to the covariance matrix between the E and I cell firing rates. Thus, our model can be interpreted in the context of predictive coding: the amount of inhibition delivered to the excitatory cells, $\Sigma_k W_{jk}^{(I\to E)}$, is approximately proportional to the optimal prediction of those cells' firing rates, gleaned from knowledge of the firing rates of the inhibitory cells (see Results, Network dynamics during inference; see Fig. 6*C*). That inhibition can then "cancel out" some of that expected (predictable) activity.

This inhibition can then act to prevent those predictable spikes from occurring, thus removing redundancy in the system. This is very similar to predictive coding models (Rao and Ballard, 1999; Spratling, 2010) in which the predictable part of the signal is suppressed, and also quite similar to "explaining away" in which activities are suppressed once their features are already accounted for (Lochmann et al., 2012), for example, via divisive feedback inhibition (Heeger, 1992).

Results

Model overview

Our spiking network model, E-I Net, extends SAILnet (Zylberberg et al., 2011) by adding a separate population of inhibitory interneurons in place of SAILnet's biologically implausible lateral inhibitory connections between ostensibly excitatory cells (Fig. 1). Both E-I Net and SAILnet are spiking neural networks that learn a sparse code with Gabor-like RF properties after exposure to image patches drawn from natural scenes. Following the precedent set by previous sparse coding studies, e.g., Olshausen and Field (1997), we used input images that were "whitened" to remove pairwise correlations in the inputs. This signal processing step is similar to what is performed on visual signals passing through the retina and the lateral geniculate nucleus (LGN) on their way to the visual cortex (Atick and Redlich, 1992; Dan et al., 1996). Both network models share the same overall dynamics allowing them to perform two important but distinct operations: inference and learning.

During inference, a network of LIF neurons is presented with a stimulus, which in our case is an image patch drawn from

whitened images of natural scenes. The exposure to the stimulus leads the neurons to spike in response over the course of simulated time. The readout of the network, and therefore the network's representation of the image patch, is the average spike rate of each neuron during the time it was exposed to the image patch. In this way, the network can be thought of as "inferring" the optimal combination of features to represent the visual input.

The network performs inference by propagating spikes within the network over the course of simulated time. The pixels of the whitened image patch correspond loosely to input from the LGN to V1. This input is integrated over simulated time and occasionally drives a neuron to spike. When a neuron spikes, it sends either excitation or inhibition (according to the neuron type) to its postsynaptic targets at the next time step in an amount proportional to the connection weight between the neurons. Over the course of the simulation (50 time steps in our case), a spike train is generated from each neuron, which constitutes that neuron's response to the stimulus.

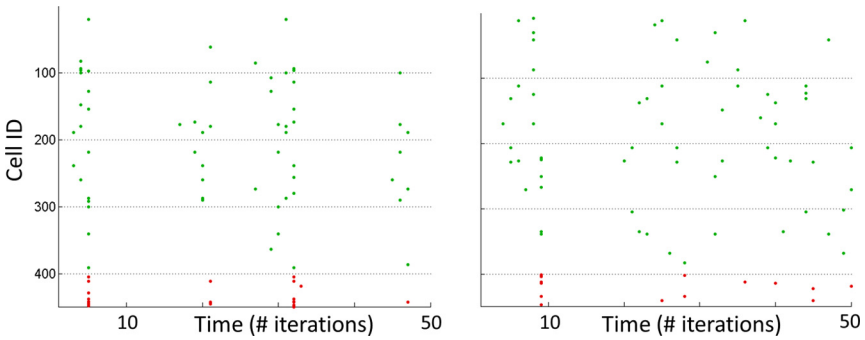


Figure 2. Representative spike raster plots generated by the E-I Net after training. Each row corresponds to a different cell. The first 400 rows are excitatory E cells (green). The last 49 rows are inhibitory I cells (red). For each plot, the network was presented with an image patch at time 0 and simulation proceeded for 50 time steps (horizontal axis). The readout of the network is the average spike rate of each excitatory neuron. Some oscillatory activity between E and I cells can be seen.

Learning proceeds on a much slower timescale using Hebbian local synaptic plasticity rules. After each image exposure, updates to the connection weights are computed locally based on the spiking activity of the presynaptic and postsynaptic neurons. Following exposure to many thousands of image patches, the neurons become tuned to specific image features, and the spiking activity of the neurons comes to resemble the Gabor-like response properties observed in simple cells of V1. Learning is driven only from network activity occurring during the inference process.

A key mechanism that shapes the dynamics of the network is the homeostatic regulation of the long-term average spike rate of the neurons. All neurons of the same class (excitatory or inhibitory) share the same long-term average spike rate, which is set in advance as a parameter. In the cortical microcircuit, numerous mechanisms contribute to homeostatic regulation (Turrigiano, 2011), which we approximate with a single spiking threshold per neuron. When a neuron's long-term average spike rate is above or below this target value, its spiking threshold is adjusted up or down using Földiák's rule for spike rate regulation (Földiák, 1990). The enforcement of a uniform long-term spike rate ensures that each neuron contributes equally to the long-term process of coding the sensory input, which is critical to achieving sparse representation stability in networks that use Hebbian learning (Perrinet, 2010).

In SAILnet (Fig. 1, left), a single population of spiking neurons learns a sparse code from image patches. Each neuron laterally inhibits all other neurons as indicated by the recurrent connection loop in the figure. Because the neurons are intended to model excitatory cells in V1, this lateral inhibition is a violation of Dale's Law.

SAILnet learns using two Hebbian local synaptic plasticity rules. The connections from the input pixels to the neurons update using Oja's variant of the Hebbian learning rule (Oja, 1982). The lateral inhibitory connections between the neurons learn using Földiák's variant of the Hebbian rule (Földiák, 1990), which strengthens connections between correlated neurons and weakens them otherwise. Critical to the functioning of SAILnet is that the neurons inhibit each other, fostering competition between the neurons, decorrelating the network, and driving the RFs to differentiate. Without lateral inhibition in a fully connected network, all neurons seek out the same principal component and converge to the same RF (Oja, 1982).

In E-I Net (Fig. 1, right), the lateral inhibitory connections of SAILnet have been replaced with a separate population of inhibitory cells. This inhibitory population laterally inhibits itself, facilitating competition and differentiating the response properties of the inhibitory population. E-I Net has four sets of connection weights instead of the two sets of SAILnet. The input connections from the whitened image patch to the excitatory cells use the Hebbian–Oja rule, as in SAILnet. All other connections use a CM inspired by Földiák's rule for correlation suppression. If persistent correlations exist within the network, which occurs with E-I Net, Földiák's rule will grow the connection weights without bound, whereas the CM rule asymptotically approaches an equilibrium value that approximately measures the activity correlation between the cells.

Figure 1, bottom, shows the weight update rules used for each connection type. Also shown are the network equilibrium states that each learning rule seeks out.

Simulation results

The dynamic properties of the neurons in E-I Net can be seen in representative spike raster plots (Fig. 2). These raster plots show the activity of the network over simulated time while being presented with an input image patch. Even though the readout of the network is the average spike rate of the excitatory cells, the spiking activity of the cells exhibits an irregular structure partly due to the decorrelating action of the inhibitory cells.

In the course of being exposed to thousands of image patches drawn from whitened images of natural scenes, the network learns Gabor-like RFs (Fig. 3*B*), which are similar to those learned by SAILnet (Fig. 3*A*). The spike-triggered average (STA) responses of the inhibitory neurons (Fig. 3*B*, bottom) exhibit slightly broader (lower spatial frequency) tuning, in agreement with experimental findings for inhibitory neurons in V1 of mouse (Kerlin et al., 2010; Liu et al., 2011), although not in cat (Anderson et al., 2000; Hirsch et al., 2003). Importantly, only a small number of inhibitory cells is required to enable the network to learn a sparse representation code, well within the excitatory-to-inhibitory neuron ratios observed in visual cortex.

To determine sparseness, we measured both population sparseness and lifetime sparseness (Vinje and Gallant, 2000). The neural activities, and thus the representation learned, has a high

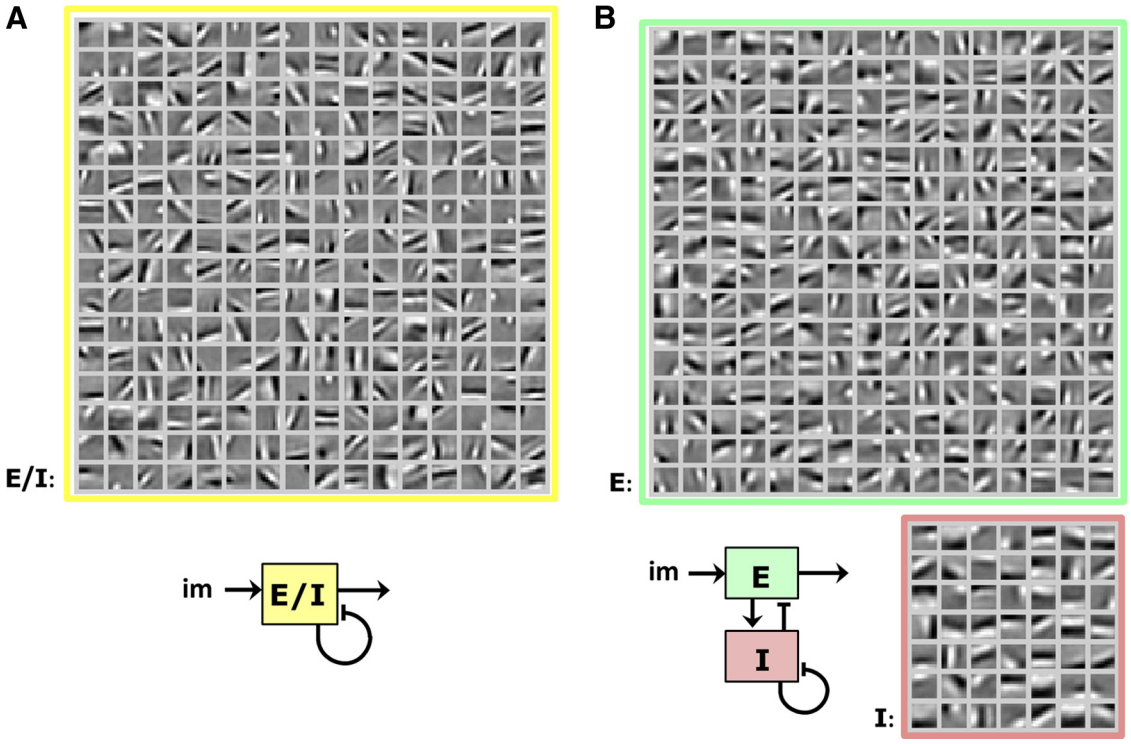


Figure 3. Representative RFs learned by the neurons in both the earlier SAILnet model (Zylberberg et al., 2011) and the new E-I Net model are described here. Both networks have 400 excitatory cells (256 randomly selected cells are shown). Each square represents the RF of a single neuron. RFs are computed as the STA of the whitened input image patches. ***A***, In SAILnet, putative excitatory neurons also laterally inhibit each other. ***B***, In E-I Net, a separate population of inhibitory neurons (bottom) decorrelates the excitatory projection neurons (top). The network in this example contained 49 inhibitory neurons. Consistent with experimental evidence, we find that the inhibitory neurons exhibit broader (lower spatial frequency) tuning curves than the excitatory cells.

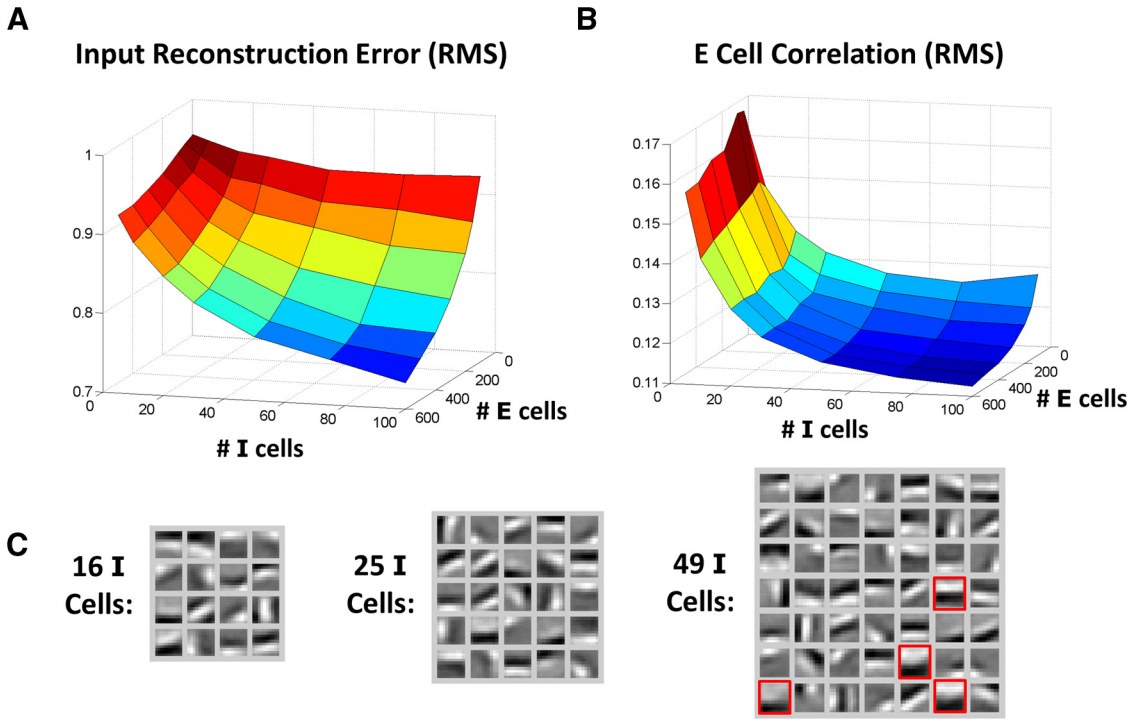


Figure 4. Adding more inhibitory (I) cells to the network improves sparse code formation as measured by reduced input reconstruction error. ***A***, Plot of input reconstruction error for varying numbers of I cells and E cells after learning a sparse code for 10 × 10 pixel image patches. Adding I cells (horizontal axis) drives the E cells to differentiate from each other in their RF response properties, leading to an improved representation code and lower reconstruction error. Increasing the number of E cells (depth axis) also improves the code by providing a larger set of differentiated cells from which to reconstruct the image. ***B***, Plot of E cell correlation (RMS of Pearson linear correlation coefficient between all pairs of E cells) for networks with different numbers of I and E cells. Adding I cells reduces E cell correlations, enabling them to learn a better sparse representation code. ***C***, Plots of the RF (spike-triggered average) of the I cells for networks with 400 E cells and 16, 25, or 49 I cells (cell counts chosen to fit a square grid). While adding I cells does improve E cell decorrelation, a point of diminishing returns is reached in which the I cells start exhibiting redundant response properties (e.g., red squares).

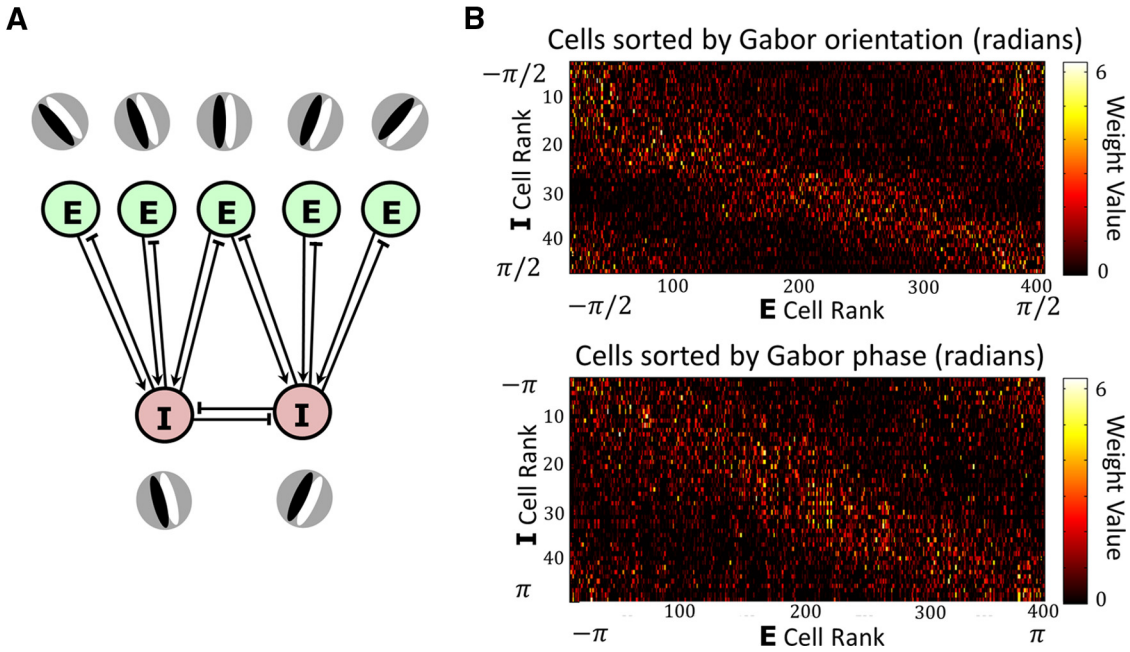


Figure 5. Connections between excitatory and inhibitory cells are strongest for pairs with similar preferred orientation and phase. ***A***, An illustration of the relationship between strongly connected E and I cells. The strongest connections are between cells with similar tuning, which are also the cells with the most correlated spiking activity. ***B***, The matrix of connection weights between E and I cells, with E and I cells sorted by orientation tuning (top) and phase tuning (bottom). Each row represents an I cell in rank order by orientation ($-\pi/2$ to $\pi/2$) or phase ($-\pi$ to π), and each column represents an E cell similarly ordered. The brightness of the point indicates the strength of either the E⟶I or I⟶E connection weight (they are the same). The strong connections (brighter points) along the diagonal show that cells with similar orientation and phase tuning have the strongest connections. The tuning of each cell's RF was computed by exposing the network to a collection of sine gratings at different orientations, phases, and spatial frequencies. A cell's preferred orientation and phase is the circular weighted average of the neural responses (spike rate) to all test stimuli.

lifetime sparseness (0.96) due to the homeostatically regulated low spike rate, and high population sparseness (0.96) due to the low correlation between the E cells (RMS Pearson correlation coefficient between cell pairs of less than 0.13).

Functional role of inhibitory neurons

In E-I Net, the excitatory and inhibitory neurons work together to compute a sparse representation code. Figure 4 shows the result of separate training simulations conducted on networks with different numbers of excitatory and inhibitory cells. These networks were trained on 10 × 10 pixel image patches drawn from whitened images' natural scenes. Learning performance was evaluated by measuring input reconstruction error, determined by reconstructing the input patch from the spike rate readout and subtracting the original pixel values (see Materials and Methods).

Networks with more neurons performed better in terms of reduced image reconstruction error than networks with fewer neurons (Fig. 4*A*). Larger numbers of excitatory cells improve performance by providing a larger and more diverse set of response properties with which to represent the input. Adding excitatory cells improved coding performance up to a point of diminishing returns at around 400 E cells (4× overcomplete relative to the 100 pixel input). Adding inhibitory cells also improved performance. However, the point of diminishing returns was reached much sooner at around 50 I cells (0.5× overcomplete). With parameter tuning, it was possible to get reasonable coding performance (reconstruction error 5% worse than optimal) with only nine inhibitory cells or 2% of the total population. This excitatory-to-inhibitory division is well within the 80/20 ratio observed in cortex (Markram et al., 2004).

The computational role of E-I Net's inhibitory neurons is to decorrelate the activity of the excitatory population, consistent with evidence for this function in visual cortex (Haider et al., 2010). As inhibitory neurons are experimentally added to the network, the degree of correlation among the E cells decreases (Fig. 4*B*). This decorrelation effect also reaches the point of diminishing returns at around 0.5× overcomplete relative to the image input, but mostly independent of the number of E cells. As I cells continue to be added, their response properties become increasingly redundant with each other (Fig. 4*C*). Note that adding E cells reduces total E cell correlations (Fig. 4*B*), contrary to what might be expected. This is a side effect of the impact on network dynamics of increased excitatory input into the I cells, which results in a stronger inhibitory feedback signal back to the E cells to drive decorrelation. If the total amount of excitatory input to the I cells is held constant via weight amplification as the E cell count is reduced, then this effect mostly goes away (data not shown).

Relationship between excitatory and inhibitory cells

Each E cell and I cell can be viewed as forming a pair with one E⟶I feedforward excitatory connection and one I⟶E feedback inhibitory connection. Both connections use the same learning rule in our model, and the learning rule is symmetric with regard to the spiking activity of the neurons. As a result, both connection weights quickly converge to the same value during learning, even if the weights start off at different random values. Thus, a strong E⟶I weight implies a strong I⟶E weight and vice versa. This correlation between strong E⟶I and I⟶E connections is consistent with electrophysiology findings for inhibitory–excitatory cell pairs in L2/3 in visual cortex of rat (Yoshimura et al., 2005).

When the network reaches training equilibrium (RMS weight change stops decreasing), the connection weight between each E and I cell is proportional to the degree of correlation between

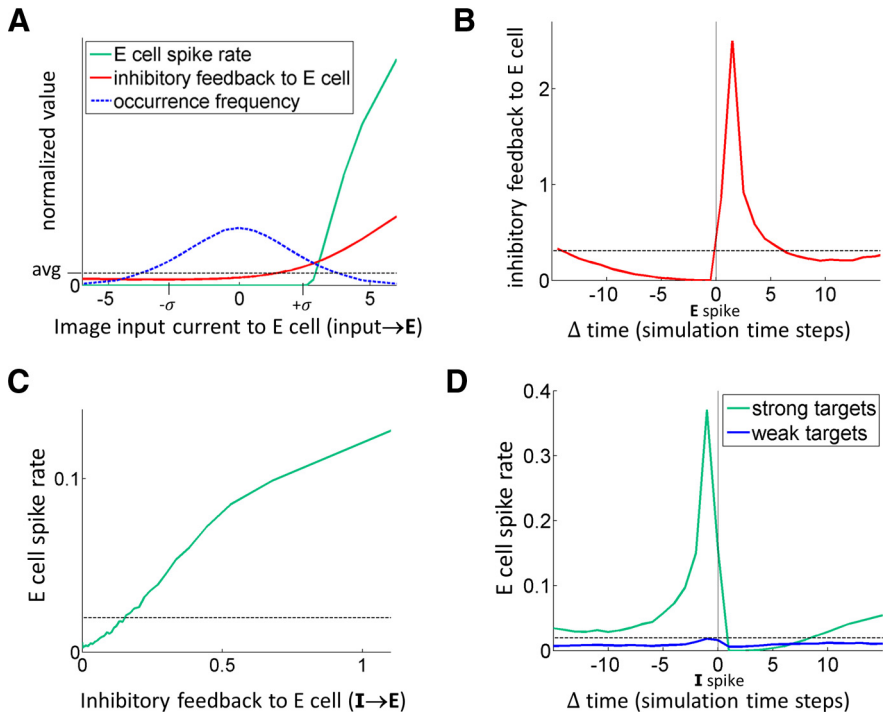


Figure 6. Network dynamics during inference illustrating computational mechanisms and predictive spike suppression. Green lines indicate excitatory cell spike rate; red lines indicate feedback inhibition to an E cell (aggregate current received from all I⟶E connections); and dashed horizontal lines indicate the mean value. ***A***, Net input current from the image patch (*x*-axis) drives the E cell spike rate (green line), resulting in inhibitory feedback through the I cell population and back to the E cells (red line). E cell spiking does not occur until the input current reaches a sufficiently high positive deflection to cross the spiking threshold. In contrast, the total inhibitory feedback from the I cells to the E cell (red line) rises smoothly as the match between the image patch and the E cell RF improves. Measurements are scaled to their mean value (dashed black line). The image input currents to the E cells forms a Gaussian distribution across image patch samples (dashed blue line, σ indicates SD). All values are averages over the full simulation period (50 time steps). ***B***, A perispike time histogram (PSTH) triggered by E cell spikes shows that I cells tend to spike after the E cells. Feedback inhibition is maximal just after spiking, when the I spike(s) facilitated by this E cell spike propagate back to the E cell. The *x*-axis shows time, in simulation time steps, relative to the E cell spike. ***C***, The total inhibitory feedback to an E cell (I⟶E) during image patch presentation is correlated with the E cell's spike rate. The dashed horizontal line indicates the mean E cell spike rate. All values are averages over the full simulation period (50 time steps). ***D***, This PSTH triggered by I cell spikes compares the E cells that are strongly connected to the I cell (E⟶I and I⟶E connection weight in the top 20%, green line) with those that are weakly connected (connection weight in the bottom 50%, blue line). The strongly connected E cells fire maximally just before the I cell spike, as they caused it via strong E⟶I connections. This same E cell population is then maximally suppressed just after the I cell spikes via strong I⟶E connections, preventing "predictable" redundant spikes from being emitted. The weakly connected E cells are mostly uncorrelated with the I cell spike. The *x*-axis shows time, in simulation time steps, relative to the I cell spike. The mean E cell spike rate across cells and images is shown for reference (dashed black line).

their stimulus-dependent firing rates, and hence to their tuning similarity. Figure 5*A* illustrates how each I cell forms strong connections to a subset of E cells that exhibit within group activity correlations and have similar tuning. Each I cell can therefore be viewed as representing a cluster of similarly tuned E cells via its strongest E⟶I and I⟶E connections.

When a few spikes arrive from members of the E cell cluster represented by an I cell, the I cell spikes and sends inhibition back to all members of the cluster. Those E cells that have not yet spiked will have their future spike transiently suppressed. In this way, the I cells implement predictive coding (Rao and Ballard, 1999; Spratling, 2010) by suppressing future E cell spikes that have been "explained away" by the E cell spikes that have already occurred (Lochmann et al., 2012). I cells also inhibit each other in proportion to their tuning similarity. Figure 5*B* shows an example set of E-to-I (also I-to-E) weights after training. The E and I cells have been sorted by tuning orientation (top) and tuning phase (bottom). The strongest connections are between neurons with similar orientation and phase tuning, as can be seen by the stronger weights (brighter points) along the matrix diagonal. The matched orientation tuning between connected excitatory and inhibitory cells is consistent with experimental findings in cat (Ferster, 1986; Hirsch et al., 2003) and mouse visual cortex (Liu et al., 2011), as well as with the observation in rat visual cortex that fast-spiking inhibitory interneurons and excitatory pyramidal cells form micronetworks in which cells with similar response properties form strong connections with each other (Yoshimura et al., 2005).

Network dynamics during inference

Figure 6 provides a view into the computational dynamics of the network during inference. The presented image patch generates a net input current to the E cell that can be viewed as a measure of the match between the image patch and the RF of the neuron (Fig. 6*A*, *x*-axis). The image input current drives the E-cell firing rate (green line), resulting in inhibitory feedback transmitted through the I-cell population (red line). The E cell does not fire until the input current reaches a positive level sufficient to cross the spiking threshold. In contrast, the total inhibitory feedback via the I cell population to the E cell rises smoothly as the match between the image patch and the E cell RF improves. The input current to the E cells has a fairly Gaussian distribution (dashed blue line).

The I cells tend to fire just after the E cells do, sending feedback inhibition to the E cell population (Fig. 6*B*). The E cells receive a total amount of feedback inhibition that is roughly proportional to their expected firing rate (Fig. 6*C*); thus, the total inhibition received by the E cell over the inference period is a good predictor of the E cell firing rate. The inhibition from the I cells is sent to all E cells that have similar response characteristics, causing those E cells that have not yet fired (but might be close to firing) to be suppressed indirectly by the E cells that have already fired. In this way, E cell spikes that have already escaped indirectly suppress potential E cell spikes that are no longer needed to complete the representation. Figure 6*D* shows the differential effect of the I-spikes on the E cells according to their RF similarity. I cell spikes are caused by (and hence preceded by) spikes from E cells that have strong E⟶I connections and similar RFs (green line). The I spike then suppresses firing in this group of similar E cells, which includes the E cells that spiked as well as others that may be close to spiking. The weakly connected E cells (blue line) are more likely to spike just after the I spike than the strongly connected E cells that have been suppressed, as these cells convey unexpected information.

Critical period plasticity and synaptic pruning

It has been proposed that GABAergic neurons mediate critical period plasticity in visual cortex (Hensch, 2005), possibly via

GABAergic plasticity of basket cells providing inhibitory feedback to pyramidal cells in layer 4 of V1 (Maffei et al., 2006), and consistent with our model here. The majority of the connections to, from, and between inhibitory cells in E-I Net after training is near zero. Deleting the 80% weakest connections to, from, and between inhibitory cells (E⟶I, I⟶I, and I⟶E connections) after training did not meaningfully affect network performance or coding behavior (0.6% increase in reconstruction error), suggestive of physiological observations of synaptic pruning in visual cortex during brain development (Bourgeois and Rakic, 1993). Importantly, this connection pruning required that synaptic plasticity be frozen—if learning was allowed to continue after pruning, performance degraded considerably (13% increase in reconstruction error). Thus, synaptic pruning was an effective connection-reduction strategy, but only after learning had completed and plasticity was suspended. This ordering is consistent with evidence for anatomical pruning following visual experience induced plasticity (Espinosa and Stryker, 2012). Surprisingly, pruning the 80% weakest input⟶E connections actually improved network coding performance (7% decrease in reconstruction error, 2% improvement in decorrelation).

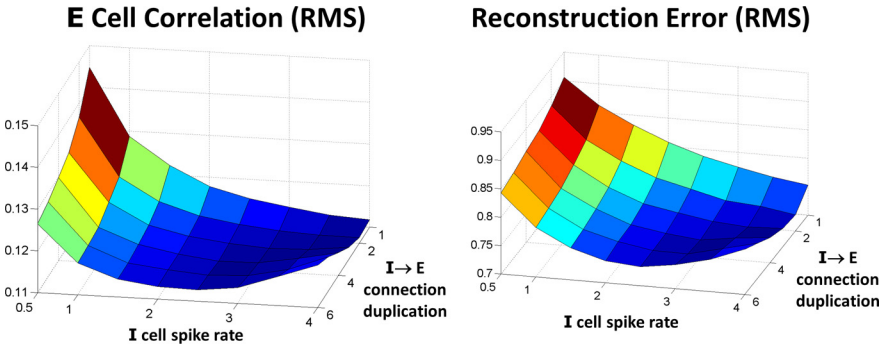


Figure 7. Performance is best for networks with I cells that spike faster than the E cells, or with duplicated I⟶E connections, or a combination of both, consistent with V1 physiology. In the surface plots above, the axes show I cell spike rate as a multiple of E cell spike rate, and synapse duplication multiplier on the I⟶E connections. Decorrelation and input reconstruction error performance was best when total inhibitory input was at the optimal level to drive decorrelation, either by increasing the I cell spike rate, increasing the number of duplicate I⟶E connections, or a combination of both. However, if total inhibitory impact was too high, network performance degraded.

Fast-spiking interneurons

Cortical neurons are often classified with regard to their temporal properties as regular spiking or fast spiking (Thomson and Lamy, 2007). We investigated the relationship between long-term average spike rate and network performance. We found that the network achieved greater decorrelation, faster convergence, and lower reconstruction error if the smaller population of inhibitory cells had a higher average spike rate and faster membrane time constant than the excitatory cells, consistent with the idea that these model neurons might be analogous to fast-spiking neurons in the cortex.

One prevalent category of fast-spiking cell in cortex is the basket cell (Thomson and Lamy, 2007), which gets its name from its tendency to form multiple duplicate connections onto its targets, for example, in V1 (Somogyi et al., 1983). We explored the effect of adding duplicate I⟶E connections in a manner similar to the fast-spiking basket cell to further increase the inhibitory impact of the smaller population of I cells. Figure 7 shows the effect of varying the I cell spike rate and also varying the number of duplicate I⟶E connections. The improved performance of the network with faster I cell spike rates is not simply a benefit of faster spike rates in general, since the excitatory cells showed only a modest spike rate effect (data not shown). Adding an I⟶E connection multiplier to simulate the reported multiple synapses made by fast-spiking basket cells onto excitatory cells also improved performance, but only up to a point. We reason that the small population of I cells relative to total E cell input needs to amplify its inhibitory impact to drive lateral competition among the E cells.

Discussion

We have introduced a model, E-I Net, which uses separate populations of excitatory (E) and inhibitory (I) spiking neurons to compute a sparse representation of image patches drawn from whitened images of natural scenes. Relying solely on synaptically local plasticity rules, the neurons in this model spontaneously learn Gabor-like RFs such as those of V1 simple cells. The inhibitory neurons in this network enable sparse code formation by facilitating competition among the excitatory cells, and only a small number of inhibitory cells is required for sparse code formation.

Our network uses a form of STDP (Dan and Poo, 2004; Feldman, 2009). Recent work on another interesting cortical model (Clopath et al., 2010) also used a form of STDP, but that network did not produce Gabor-like RFs resembling those from V1. Another recent study (Evans and Stringer, 2012) considered a network with separate E and I populations, but their inhibitory connections were not plastic, and they studied neither the shapes of the RFs learned by their excitatory cells nor the specific representation of the stimulus formed by the neuronal activities.

In our E-I Net, the primary role of the inhibitory neurons is to suppress future spiking activity that can be predicted from spikes that have already occurred, thus implementing a form of predictive coding (Rao and Ballard, 1999) to facilitate explaining away via feedback inhibition (Lochmann et al., 2012). By suppressing predicted spikes, the I cells actively decorrelate the neural population generally, consistent with observations that inhibitory cells perform decorrelation in V1 (Haider et al., 2010), which in turn increases the sparseness of the V1 representation (Vinje and Gallant, 2000).

The inhibitory cells can successfully drive decorrelation using only a relatively small number of cells. If the inhibitory neurons themselves are decorrelated, and thus orthogonal to each other, each neuron can divide the neural population in half, allowing N_I inhibitory cells to decorrelate 2^{N_I} excitatory cells, or equivalently requiring only $N_I = \log_2 N_E$ inhibitory cells to decorrelate N_E neurons. Consistent with this prediction, we were able to decorrelate 400 E cells using only 9 I cells in our network.

Our results suggest a functional role for inhibitory neurons in neural code formation—decorrelating the excitatory neurons. Many spiking models of pattern learning and sparse coding use direct inhibitory connections between ostensibly excitatory cells to facilitate competitive learning with the assumption that these connections could be replaced with inhibitory interneurons (Masquelier et al., 2009; Savin et al., 2010; Zylberberg et al., 2011; Masquelier, 2012). Here we show how

this can be achieved with a separate inhibitory population obeying local plasticity rules.

Accurate predictions of V1 RFs (Olshausen and Field, 1996; Rehn and Sommer, 2007; Olshausen et al., 2009; Zylberberg et al., 2011) provide support for longstanding ideas about various forms of coding efficiency in neural representations (Attneave, 1954; Barlow, 1961; Laughlin, 1981, 2001; Atick and Redlich, 1992; Rieke et al., 1997). In addition, sparse coding models have successfully predicted response properties at several stages in the ascending auditory pathway (Klein et al., 2003; Smith and Lewicki, 2006; Zhao and Zhaoping, 2011; Carlson et al., 2012). Firing rate statistics also provides evidence for sparse coding in visual cortex (Vinje and Gallant, 2000, 2002; Lennie, 2003; Graham and Field, 2006; Haider et al., 2010) and auditory cortex (DeWeese et al., 2003; Hromádka et al., 2008), though there are also reports of dense coding (Tolhurst et al., 2009) and mixtures of both (Sakata and Harris, 2009).

Moreover, experimental observations show that correlations among similarly tuned V1 neurons is low, suggesting the existence of a mechanism for active decorrelation in the cortical microcircuit (Ecker et al., 2010). The orientation tuning similarity between connected excitatory and inhibitory neurons matches physiological observations in V1 (Anderson et al., 2000; Hirsch et al., 2003; Alitto and Dan, 2010). The broader tuning we observed for inhibitory cells, as well as the positive activity correlation between connected excitatory and inhibitory cells, matches physiological findings in mouse (Kerlin et al., 2010; Liu et al., 2011) and rat (Yoshimura et al., 2005) V1; however, these effects may be species specific as they have not been observed in cat (Anderson et al., 2000; Hirsch et al., 2003).

Can a particular class of neuron be identified in cerebral cortex that performs the functional role of active decorrelation via feedback inhibition? One candidate is the parvalbumin-positive (PV^+) basket cell found in both L4 and L2/3 of visual cortex. These neurons have horizontal connections to each other (Tamás et al., 1998) in agreement with E-I Net's model interneurons. PV^+ basket cells furthermore form reciprocal connections with pyramidal cells in L4 of sensory cortex (Ali et al., 2007; Thomson and Lamy, 2007) and with excitatory cells in L2/3 of rat visual cortex (Yoshimura et al., 2005).

E-I Net requires only a small percentage of inhibitory neurons relative to the total neural population (2–10%) to adequately decorrelate the excitatory cells and learn a sparse code. This excitatory-to-inhibitory division is well within the 80/20 ratio seen in visual cortex. Basket cells account for about 50% of all GABAergic neurons in neocortex (Markram et al., 2004; Alitto and Dan, 2010), which in turn account for $<$20% of total neurons, implying that basket cells constitute $<$10% of the total neural population in V1.

Basket cells make numerous redundant connections onto their targets and are observed to fire at faster than average rates (Thomson and Lamy, 2007). Consistent with this, we found that network simulations achieved the highest level of E cell decorrelation and the lowest input reconstruction error levels when the inhibitory cells spike at higher average firing rates (up to 4×) than the excitatory cells, or when the inhibitory cells made multiple redundant connections onto the excitatory cells, or both. The increased spike rate or connection count compensated for the smaller number of inhibitory cells, allowing a relatively small population to facilitate competition.

It has been proposed that PV^+ basket cells mediate critical period plasticity in visual cortex (Hensch, 2005). Consistent with this, we found that the 80% weakest inhibitory connections could be deleted after training with negligible effect on network coding performance, but only if this occurred after synaptic plasticity was disabled, or if E cell RFs were "locked in" by pruning their input connections. If we allowed synaptic plasticity to continue after pruning, the thinned out inhibitory network was unable to adapt to the changing correlation patterns in the excitatory population caused by changing connection strengths, resulting in substantially degraded coding performance.

Other inhibitory cell types exhibit reciprocal within-class connections. For example in rat V1, 50% of the targets of calretinin-positive (CR^+) interneurons are other CR^+ interneurons (Gonchar and Burkhalter, 1999). Within-class reciprocal connections between inhibitory interneurons may be a general strategy in cortical circuits for facilitating pattern separation in sensory coding.

Two complementary types of inhibition have been proposed in sensory cortex: feedforward and feedback (Isaacson and Scanziani, 2011). In feedforward inhibition, inhibitory inputs would be expected to be negatively correlated with the target cell so as to suppress responses that are contradicted by the sensory input, for example, in a push–pull fashion (Ferster and Miller, 2000; Hirsch et al., 2003). We do not attempt to address feedforward inhibition here. In place of feedforward inhibition, our LGN-like inputs can adopt both positive and negative values, a relaxation of biological constraints that allows us to focus on the inhibitory feedback mechanism. Feedforward inhibition could be modeled with a separate population of inhibitory cells receiving feedforward input from the stimulus, which is an exciting direction for future work.

In feedback inhibition, the inhibitory mechanism we focus on here, sensory signals that have already been coded are fed back via inhibitory interneurons that are positively correlated with the target cell. Feedback inhibition can be understood as divisive inhibition (Heeger, 1992) to remove input that can be predicted or explained away (Lochmann et al., 2012). The output signal from the excitatory population propagates through the inhibitory interneurons to inhibit excitatory cells with similar RFs, thus suppressing redundant spiking activity.

Our primary result is that a distinct population of inhibitory interneurons can enable a biologically realistic spiking network to learn a sparse code for natural scenes from input data using only synaptically local plasticity rules. This can be achieved with a relatively small number of inhibitory cells, in agreement with excitatory-to-inhibitory neuron ratios observed in visual cortex. Moreover, performance of our network improves with increasing firing rates in our inhibitory population, but performance is largely unaffected by increased firing rates among the excitatory neurons, suggesting a computational benefit for the observed high firing rates of the smaller number of inhibitory relative to excitatory neurons in V1. We show how this can be achieved using a learning rule that inhibits spikes that can be predicted, thus suppressing redundant activity that has been explained away. We have described a mechanism for interneuron-mediated competition between neurons in the visual cortex; however, our model may apply to cortical circuits generally.

Notes

Supplemental material for this article is available at http://www.pking.org/research/EINet. Materials include MATLAB source code for the simulation available for download. This material has not been peer reviewed.

References

Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. Nat Neurosci 3[Suppl]:1178–1183. Medline

Ali AB, Bannister AP, Thomson AM (2007) Robust correlations between action potential duration and the properties of synaptic connections in layer 4 interneurones in neocortical slices from juvenile rats and adult rat and cat. J Physiol 580:149–169. CrossRef Medline

Alitto HJ, Dan Y (2010) Function of inhibition in visual cortical processing. Curr Opin Neurobiol 20:340–346. CrossRef Medline

Anderson JS, Carandini M, Ferster D (2000) Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. J Neurophysiol 84:909–926. Medline

Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? Neural Comput 4:196–210. CrossRef

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193. CrossRef Medline

Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: Sensory communication (Rosenblith WA, ed), pp 217–234: MIT.

Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. Vision Res 37:3327–3338. CrossRef Medline

Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472. Medline

Bourgeois JP, Rakic P (1993) Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. J Neurosci 13:2801–2820. Medline

Carlson NL, Ming VL, Deweese MR (2012) Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. PLoS Comput Biol 8:e1002594. CrossRef Medline

Clopath C, Büsing L, Vasilaki E, Gerstner W (2010) Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci 13:344–352. CrossRef Medline

Dan Y, Poo MM (2004) Spike timing-dependent plasticity of neural circuits. Neuron 44:23–30. CrossRef Medline

Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16:3351–3362. Medline

DeWeese MR, Wehr M, Zador AM (2003) Binary spiking in auditory cortex. J Neurosci 23:7940–7949. Medline

Eccles J (1976) From electrical to chemical transmission in the central nervous system. Notes Rec R Soc Lond 30:219–230. CrossRef Medline

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. Science 327:584–587. CrossRef Medline

Espinosa JS, Stryker MP (2012) Development and plasticity of the primary visual cortex. Neuron 75:230–249. CrossRef Medline

Evans BD, Stringer SM (2012) Transformation-invariant visual representations in self-organizing spiking neural networks. Front Comput Neurosci 6:46. CrossRef Medline

Feldman DE (2009) Synaptic mechanisms for plasticity in neocortex. Annu Rev Neurosci 32:33–55. CrossRef Medline

Ferster D (1986) Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex. J Neurosci 6:1284–1301. Medline

Ferster D, Miller KD (2000) Neural mechanisms of orientation selectivity in the visual cortex. Annu Rev Neurosci 23:441–471. CrossRef Medline

Földiák P (1990) Forming sparse representations by local anti-Hebbian learning. Biol Cybern 64:165–170. CrossRef Medline

Gonchar Y, Burkhalter A (1999) Connectivity of GABAergic calretinin-immunoreactive neurons in rat primary visual cortex. Cereb Cortex 9:683–696. CrossRef Medline

Graham DJ, Field DJ (2006) Sparse coding in the neocortex. In: Evolution of nervous systems (Kaas JH, Krubitzer LA, eds), Vol 3, pp 181–187. London: Academic.

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

Hebb DO (1949) The organization of behavior: a neuropsychological theory. New York: Wiley.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197. CrossRef Medline

Hensch TK (2005) Critical period plasticity in local cortical circuits. Nat Rev Neurosci 6:877–888. CrossRef Medline

Hirsch JA, Martinez LM, Pillai C, Alonso JM, Wang Q, Sommer FT (2003) Functionally distinct inhibitory neurons at the first stage of visual cortical processing. Nat Neurosci 6:1300–1308. CrossRef Medline

Hromádka T, Deweese MR, Zador AM (2008) Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol 6:e16. CrossRef Medline

Hu T, Genkin A, Chklovskii D (2012) Computing sparse representations using a network of integrate-and-fire neurons, Janelia Farm Research Campus. In Computational and Systems Neuroscience (COSYNE) Meeting, Salt Lake City, UT, February 27, 2012.

Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. Neuron 72:231–243. CrossRef Medline

Karklin Y, Simoncelli EP (2011) Efficient coding of natural images with a population of noisy linear-nonlinear neurons. Adv Neural Inform Proc Systems 24:999–1007.

Kerlin AM, Andermann ML, Berezovskii VK, Reid RC (2010) Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. Neuron 67:858–871. CrossRef Medline

Klein D, König P, Körding K (2003) Sparse spectrotemporal coding of sounds. EURASIP J Adv Signal Proc 2003:659–667. CrossRef

Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch C 36:910–912. Medline

Laughlin SB (2001) Energy as a constraint on the coding and processing of sensory information. Curr Opin Neurobiol 11:475–480. CrossRef Medline

Lennie P (2003) The cost of cortical computation. Curr Biol 13:493–497. CrossRef Medline

Liu BH, Li YT, Ma WP, Pan CJ, Zhang LI, Tao HW (2011) Broad inhibition sharpens orientation selectivity by expanding input dynamic range in mouse simple cells. Neuron 71:542–554. CrossRef Medline

Lochmann T, Deneve S (2011) Neural processing as causal inference. Curr Opin Neurobiol 21:774–781. CrossRef Medline

Lochmann T, Ernst UA, Denève S (2012) Perceptual inference predicts contextual modulations of sensory responses. J Neurosci 32:4179–4195. CrossRef Medline

Maffei A, Nataraj K, Nelson SB, Turrigiano GG (2006) Potentiation of cortical inhibition by visual deprivation. Nature 443:81–84. CrossRef Medline

Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C (2004) Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5:793–807. CrossRef Medline

Masquelier T (2012) Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model. J Comput Neurosci 32:425–441. CrossRef Medline

Masquelier T, Guyonneau R, Thorpe SJ (2009) Competitive STDP-based spike pattern learning. Neural Comput 21:1259–1276. CrossRef Medline

McLaughlin D, Shapley R, Shelley M, Wielaard DJ (2000) A neuronal network model of macaque primary visual cortex (V1): orientation selectivity and dynamics in the input layer 4Calpha. Proc Natl Acad Sci U S A 97:8087–8092. CrossRef Medline

Oja E (1982) A simplified neuron model as a principal component analyzer. J Math Biol 15:267–273. CrossRef Medline

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609. CrossRef Medline

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res 37:3311–3325. CrossRef Medline

Olshausen BA, Cadieu CF, Warland DK (2009) Learning real and complex overcomplete representations from the statistics of natural images. Proc SPIE 7446:74460S. CrossRef

Perrinet LU (2010) Role of homeostasis in learning sparse representations. Neural Comput 22:1812–1836. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. CrossRef Medline

Rehn M, Sommer FT (2007) A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. J Comput Neurosci 22:135–146. CrossRef Medline

Rieke F, Warland D, de Ruyter von Steveninck R, Biolek W (1997) Spikes: exploring the neural code. Cambridge, MA: MIT.

Rozell CJ, Johnson DH, Baraniuk RG, Olshausen BA (2008) Sparse coding

via thresholding and local competition in neural circuits. Neural Comput 20:2526–2563. CrossRef Medline

Sakata S, Harris KD (2009) Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron 64:404–418. CrossRef Medline

Salinas E, Abbott LF (1994) Vector reconstruction from firing rates. J Comput Neurosci 1:89–107. CrossRef Medline

Savin C, Joshi P, Triesch J (2010) Independent component analysis in spiking neurons. PLoS Comput Biol 6:e1000757. CrossRef Medline

Shapero S, Brüderle D, Hasler P, Rozell C (2011) Sparse approximation on a network of locally competitive integrate and fire neurons. In Computational and Systems Neuroscience (COSYNE) Meeting, Salt Lake City, UT, February 2011.

Smith EC, Lewicki MS (2006) Efficient auditory coding. Nature 439:978–982. CrossRef Medline

Somogyi P, Kisvárday ZF, Martin KA, Whitteridge D (1983) Synaptic connections of morphologically identified and physiologically characterized large basket cells in the striate cortex of cat. Neuroscience 10:261–294. CrossRef Medline

Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. J Neurosci 30:3531–3543. CrossRef Medline

Tamás G, Somogyi P, Buhl EH (1998) Differentially interconnected networks of GABAergic interneurons in the visual cortex of the cat. J Neurosci 18:4255–4270. Medline

Thomson AM, Lamy C (2007) Functional maps of neocortical local circuitry. Front Neurosci 1:19–42. CrossRef Medline

Tolhurst DJ, Smyth D, Thompson ID (2009) The sparseness of neuronal responses in ferret primary visual cortex. J Neurosci 29:2355–2370. CrossRef Medline

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

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276. CrossRef Medline

Vinje WE, Gallant JL (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. J Neurosci 22:2904–2915. Medline

Yoshimura Y, Dantzker JL, Callaway EM (2005) Excitatory cortical neurons form fine-scale functional networks. Nature 433:868–873. CrossRef Medline

Zhao L, Zhaoping L (2011) Understanding auditory spectro-temporal receptive fields and their changes with input statistics by efficient coding principles. PLoS Comput Biol 7:e1002123. CrossRef Medline

Zhu M, Olshausen BA, Rozell CJ (2012) Biophysically accurate inhibitory interneuron properties in a sparse coding network. In Computation and Systems Neuroscience (COSYNE) Meeting, Salt Lake City, UT, February 2012.

Zylberberg J, Murphy JT, DeWeese MR (2011) A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. PLoS Comput Biol 7:e1002250. CrossRef Medline