The control of plastic inhibition over excitatory synaptic plasticity leads to the joint emergence of sensory coding and contrast invariance.

René Larisch [1,*], Lorenz Gönner [1,2], Michael Teichmann [1] and Fred H. Hamker [1,3,**]

April 7, 2020

[1] *TU Chemnitz, Dept. of Computer Science, Artificial Intelligence*

[2] *TU Dresden, Faculty of Psychology, Lifespan Developmental Neuroscience*

[3] *Bernstein Center Computational Neuroscience Berlin*

[*] *rene.larisch@informatik.tu-chemnitz.de*

[**] *fred.hamker@informatik.tu-chemnitz.de*

Visual stimuli are represented by a highly efficient code in the primary visual cortex, but the development of this code is still unclear. Two distinct factors control coding efficiency: Representational efficiency, which is determined by neuronal tuning diversity, and metabolic efficiency, which is influenced by neuronal gain. How these determinants of coding efficiency are shaped during development, supported by excitatory and inhibitory plasticity, is only partially understood. We investigate a fully plastic spiking network of the primary visual cortex, building on phenomenological plasticity rules. Our results show that inhibitory plasticity is key to the emergence of tuning diversity and accurate input encoding. Additionally, inhibitory feedback increases the metabolic efficiency by implementing a gain control mechanism. Interestingly, this led to the spontaneous emergence of contrast-invariant tuning curves. Our findings highlight the role of interneuron plasticity during the development of receptive fields and in shaping sensory representations.

1 Introduction

The primary visual cortex (V1) represents visual input stimuli in a highly efficient manner (Froudarakis et al., 2014; Dadarlat & Stryker, 2017). Recent research has identified two distinct factors underlying the efficiency of visual representations: First, representational efficiency in terms of absolute information content, which is mainly determined by the receptive field tuning diversity (Goris et al., 2015). Second, metabolic efficiency in terms of the number of spikes required to represent a specific input stimulus. This aspect is strongly influenced by gain control mechanisms caused by inhibitory feedback processing (Carvalho & Buonomano, 2009; Isaacson & Scanziani, 2011). How these determinants of coding functionality are shaped during development is only partially understood. While it has long been known that excitatory plasticity is necessary for the development of an accurate and efficient input representation (Olshausen & Field, 1996; Bell & Sejnowski, 1997; Zylberberg et al., 2011), there has recently been growing interest in the role of inhibitory plasticity, fueled by recent studies demonstrating plasticity at inhibitory synapses (Khan et al., 2018). As the synaptic plasticity of inhibitory interneurons in V1 likely exerts strong effects on the outcome of excitatory plasticity (Wang & Maffei, 2014), complex circuit-level interactions occur between both types of plasticity. This notion has received further support based on recent theoretical studies (Mongillo & Loewenstein, 2018). Above all, these findings raise the question of how excitatory and inhibitory plasticity can cooperate to enable the development of an efficient stimulus code.

Network models have proposed neural-level mechanisms of sparse code formation (Olshausen & Field, 1996) based on Hebbian plasticity. However, these models typically rely on simplified learning dynamics (Savin et al., 2010; Zylberberg et al., 2011; King et al., 2013) or consider plasticity only at a subset of projections in the network (Sadeh et al., 2015; Miconi et al., 2016), not addressing the development of feedback-based gain control. As such, it remains unclear how functional input encoding can emerge in a more detailed model of V1 circuit development.

We here propose that a single underlying mechanism - the influence of inhibitory plasticity on excitatory plasticity - is sufficient to explain both, the observed feed-forward tuning and neuronal gain-control by feedback processing, and we investigate this influence in a spiking network model of V1 layer 4. Our key finding is that inhibitory plasticity supports the joint development of feed-forward tuning and balances inhibitory feedback currents. Importantly, this balance leads to the spontaneous emergence of contrast-invariant tuning curves, as an inherent phenomenon of the network and its plasticity dynamics. Our results link both representational efficiency and metabolic efficiency to synaptic plasticity mechanisms.

2 Results

To investigate the interaction between excitatory and inhibitory plasticity, we designed a spiking network model of V1-layer 4 consisting of an excitatory and inhibitory population, stimulated with natural image patches (**Fig. 1a**) (see **Methods**). The circuit of our neuronal network implements both feed-forward and feedback inhibition, in agreement with anatomical findings (Isaacson & Scanziani, 2011). Although different kinds of inhibitory neurons have been found in the neocortex (Markram et al., 2004; Priebe & Ferster, 2008), our network contains only one population of inhibitory neurons, as a simplification. The size of the inhibitory population was chosen to match the 4:1 ratio between excitatory and inhibitory neurons found in striate cortex (Beaulieu et al., 1992; Markram et al., 2004; Potjans & Diesmann, 2014). The plasticity of the excitatory synapses follows the voltage-based triplet spike timing-dependent plasticity (STDP) rule proposed by Clopath et al. (2010). The strength of the inhibitory synapses changes according to the symmetric inhibitory STDP rule described by Vogels et al. (2011), which achieves homeostasis by maintaining a constant postsynaptic firing rate (ρ). This allows us to vary the strength of inhibitory synapses in the network, to investigate how the balance between excitation and inhibition influences the emergence of neuronal gain-control and feed-forward tuning.

For this purpose, we compare a network with a $2:1$ ratio of excitation to inhibition to a model version with a $3:1$ excitation to inhibition ratio, averaged on natural scene patches (**Fig. 1b**). Additionally, we blocked inhibitory synapses after learning to investigate the dynamic effects of inhibition on network coding (called $blockInh$) Each of the three model configurations was repeated 10 times, initialized with randomly chosen weight values, to test the stability and reproducibility of the observed outcomes. To analyze the influence of inhibition during learning after all, our fourth model configuration contains no inhibitory synapses (called $noInh$ model).

Emergence of diversely tuned receptive fields The receptive fields of V1 simple cells are often described by Gabor functions (Jones & Palmer, 1987a; Ringach, 2002; Spratling, 2012). We observe the emergence of Gabor-like receptive fields in our network for the excitatory and inhibitory population with the spike triggered average method (STA, see **Methods**). Without inhibition, most of the receptive fields have a similar orientation and position (**Fig. 2a**). In contrast, the presence of inhibition during learning resulted in a higher diversity of receptive fields with a more complex structure for the excitatory population (**Fig. 2b**) and the inhibitory population (**Fig. 2c**). The measured receptive fields showed a strong tendency for weight values to cluster around the minimum or the maximum value. This is a known characteristic of the learning rule chosen for excitatory synapses, which enforces strong synaptic competition (Clopath et al., 2010; Miconi et al., 2016).

We fitted the learned receptive fields with Gabor functions (see Methods) and calculated the normalized mean-square error (NMSE, see **Eq. 14**) to quantify the fit (Spratling, 2012). Fits with an error greater than 0.5 were excluded from further evaluations about spatial properties, which occurred for around 25% of neurons for the $EI2/1$ model, and around 2% of all neurons for the $noInh$ model, averaged across 10 runs for every network configuration.

A broader range of orientations emerged in the networks with inhibition (**Fig. 2e**). Without inhibition, most receptive fields converge to a preferred orientation around 0° or 180° (**Fig. 2d**). In the model with weaker inhibition ($EI3/1$), receptive fields converge to a very similar orientation distribution than in the $EI2/1$ model (see Supplementary **S1**). In addition, the inhibitory cells in the $EI2/1$ models also become selectively tuned, with a clear preference at 0° and 180° (**Fig. 2f**). This is in line with recent experiments on mouse V1, in which tuned inhibition is found (Bock et al., 2011; Hofer et al., 2011; Liu et al., 2011).

Emergence of structured feed-forward and recurrent connectivity As both, the excitatory and inhibitory cells in our network developed a tuning for orientation and position, we expected that their modifiable synaptic connections developed a specific pattern reflecting activity correlations (King et al., 2013; Sadeh et al., 2015). Our analysis confirmed that excitatory neurons developed strong connections to inhibitory neurons with similar orientation tuning (**Fig. 3a**, top). Inhibitory weights to the excitatory layer showed a similar pattern, although with somewhat reduced specificity (**Fig. 3a**, bottom). This implements an indirect inter-neuron connection between two excitatory neurons via mutually connected inhibitory neurons, to inhibit each other maximally. The development of strong recurrent inhibitory synapses between similarly tuned inhibitory cells can be observed as well (**Fig. 3b**).

Inhibition controls response decorrelation We observed that the different levels of inhibition in the $EI2/1$ and $EI3/1$ models led to similar orientation distributions. To investigate if response correlations between neurons only depend on the orientation similarity or whether lateral inhibition has an additional decorrelation effect (as mentioned in previous modeling approaches of Wiltschut & Hamker (2009); Savin et al. (2010); Zylberberg et al. (2011); King et al. (2013)), we analyzed the development of correlations during receptive field learning (**Fig. 4a**). During the first $250,000$ of all $400,000$ input stimuli, a weak reduction of the correlation can be observed in the $noInh$ model. The $EI2/1$ model showed a pronounced decrease of correlations across learning, with the highest reduction occurring in the early phase of learning showing the highest amount of changes of the feed-forward weights. Weaker feedback inhibition ($EI3/1$ model) led to weaker decorrelation of neuronal activity. This confirms that the level of inhibition determines the degree of decorrelation of pairwise responses.

Smith & Kohn (2008) recorded the neuronal activity in V1 of macaque monkeys during the presentation of drifting sinusoidal gratings and reported a dependence of pairwise response correlations on orientation tuning similarity. We performed a similar analysis of our model data, to analyze the effect of feedback inhibition on the response correlation with respect to the orientation selectivity. We sorted all cell pairs by similarity, grouped them into 30 equally-spaced bins, and averaged their response correlation values within each bin, based on natural scene stimuli (details see **Methods**) (**Fig. 4b**). Without inhibition, we observed a mean response correlation of ≈ 0.95 for cell pairs with highly similar receptive fields. With inhibition, this value dropped to ≈ 0.8. By contrast, cell pairs with dissimilar receptive fields showed average correlation values of around 0.4 for the *noInh* and the *blockInh* model. Here, inhibitory processing substantially reduced the mean correlation to near zero-values for the $EI2/1$ model. A comparison between the $EI2/1$ model and its counterpart with blocked inhibition shows that dissimilarly tuned neuron pairs are more strongly decorrelated than pairs with highly similar tuning. At a first glance, this pattern contrasts with the emergent connectivity structure: The connectivity pattern favors strong mutual inhibitory connections between inhibitory neurons which receive projections from (and project back to) excitatory neurons with similar tuning, creating strong reciprocal inhibition (**Fig. 3a** and **Fig. 3b**). However, our observation of target-specific decorrelation is best understood by considering that correlated mean responses can arise both through a similarity of tuning and through unspecific baseline activity. Natural image patches are likely to evoke broad excitation among many cells, similar to sinusoidal grating stimuli. The correlation between dissimilarly tuned neurons is most likely caused by the activity baseline, which is strongly reduced by inhibition. Besides, similarly tuned cells will retain strongly overlapping tuning curves even after reduction of unspecific activity, associated with strong correlation of their mean response (Averbeck et al., 2006).

Inhibitory feedback shapes tuning curves To quantify the effect of inhibition on the magnitude of individual neuronal responses, we measured orientation tuning curves of each neuron by sinusoidal gratings. For all approaches and model variants, the maximum firing rate in the input was set to $\approx 85Hz$ to obtain sufficiently high activity levels. We observed high baseline and peak activity in both model variants without inhibition (**Fig. 5a**). However, activity levels in the *blockInh* model were lower than in the *noInh* model, likely owing to its smaller and more dispersed receptive fields. As expected, the model with active inhibitory feedback showed a strong reduction of firing rates. To obtain a measure of tuning sharpness, we next estimated the orientation bandwidth (OBW) of the excitatory population, based on the measured tuning curves. As expected, and consistent with previous observations (Isaacson & Scanziani, 2011; Stringer et al., 2016), we observed a sharpening effect through inhibition (**Fig. 5b**).

Spontaneous emergence of contrast-invariant tuning curves Besides the sharpening of tuning curves, previous models suggest a role of inhibition in the invariance to input contrast changes (Troyer et al., 1998; Ferster & Miller, 2000; Priebe & Ferster, 2008). However, those models assume hard-wired connectivity, and propose push-pull or anti-phase inhibition (Troyer et al., 1998; Ferster & Miller, 2000). Contrast-invariant V1 simple cells have been found in different mammals such as, cats (Skottun et al., 1987; Finn et al., 2007) or ferrets (Alitto & Usrey, 2004), based on sinusoidal gratings with different contrast strength. We use the same approach (see **Methods**) to measure the tuning curves and calculated the averaged OBW over all excitatory cells for the different contrast levels (**Fig. 6a**). Interestingly, the OBW is constant only for the $EI2/1$ model. For both models without inhibition and for the model with weaker inhibition, the OBW increases for higher input contrast values. To understand this effect, we compared the $EI2/1$ with the $EI3/1$ model with regard to their spike count, average membrane potential, and the average of the summed synaptic input current, for different contrast levels. At any contrast level, the activity of neurons in the $EI2/1$ model remains strongly suppressed at non-preferred orientations and increases around the preferred orientation (**Fig. 7a**). By contrast, the $EI3/1$ model shows increased activity for high input contrast at all orientations (**Fig. 7b**). This results in increased OBW values for higher input contrast. Interestingly, for the non-preferred orientation, the average membrane potential the $EI2/1$ model is less hyperpolarized for lower contrast than for higher contrast. For higher contrast, the average membrane potential increases at the preferred orientation and is substantially stronger than for lower contrast. Both curves intersect around $-50mV$, close to the resting state spiking threshold ($-50.4mV$) (**Fig. 7c**). This can be explained with the average input current: At higher contrast levels and non-preferred orientations, the feedback inhibitory current increases more strongly than the excitatory current and nearly compensates it (**Fig. 7e** and **S3 a**), providing hyperpolarization of the membrane potential. This compensation of excitation decreases around the preferred stimulus, where the membrane potential exceeds the spiking threshold. In comparison, the membrane potential for the $EI3/1$ model increases proportionally with the total input current caused by higher input contrast (**Fig. 7d**, **Fig. 7f** and **S3 b**). This suggests that the contrast-invariant tuning of the $EI2/1$ model depends on an appropriate balance between excitation and inhibition.

Based on the observation of contrast invariant tuning curves, we conclude that feedback inhibition modulates the neuronal gain controlled by input orientation and contrast. **Fig. 6b** shows the average response gain for the excitatory population, averaged across the whole population (see **Methods** for more details). We show the response gain curves for low and high contrast stimuli. For the model with blocked inhibition ($blockInh$), the gain curve is unaffected by contrast and follows the activation function defined by the neuron model. The firing rates of the $EI2/1$ model are strongly reduced relative to the $blockInh$

model. Further, this gain modulation is contrast-dependent, as the highest reduction of firing rates is observed for high contrast. This shows that the effect of inhibition on the neuronal gain function not only depends on the amount of excitatory input, but also on the stimulus orientation and contrast strength.

Sparseness is increased by both inhibition and tuning diversity As we observed that inhibitory processing led to an increase in the selectivity to artificial stimuli, we asked whether inhibition contributed to a sparser population code for natural images. We first compared the overall spiking behavior based on raster plots of network responses to five example image patches, for the $EI2/1$ (**Fig. 8a**) and the $blockInh$ model (**Fig. 8c**). The model with active inhibition showed sparser firing and a less synchronous spiking behavior than the model with blocked inhibition. Second, to quantify this effect, we measured the population sparseness for all model configurations, based on the responses to 10.000 natural image patches (**Fig. 8b**). The highest sparseness value (0.62) was observed in the $EI2/1$ model, 0.54 for the $blockInh$ model and the lowest sparseness value (0.43) in the $noInh$ model. Interestingly, the development of a higher diversity of receptive fields had a stronger influence on the population sparseness than inhibitory processing: Sparseness values differed more strongly between the model configurations without inhibition, the $noInh$ and $blockInh$ model, than between the $EI2/1$ and its blocked counterpart, which share the same feed-forward receptive fields.

Metabolic efficiency benefits from feedback inhibition The efficiency of information transmission, or metabolic efficiency, is associated with the observed increase of the population sparseness (Spanne & Jörntell, 2015). To quantify the efficiency, we calculated the mutual information between input and response (**Sec.Methods**). This analysis revealed a strong impact of inhibition on transmission efficiency (**Fig. 8d**), normalized by spike count. The $EI2/1$ model shows the highest amount of information per spike (0.96 $bits/spike$). Both models without inhibition were associated with the least efficient population coding, with a lower value for the of the $blockInh$ model, caused by a more diverse receptive field structure. To analyze further how the increase in information transmission was achieved, we calculated the discriminability index d' on 500 randomly chosen natural scene patches to quantify the trial-to-trial fluctuation. We observed that higher d' values were associated to both high tuning diversity and the presence of inhibition(see Supplementary **S2**). The improvement in discriminability is likely caused by a reduction of unspecific activity by inhibition, associated with more reliable stimulus representations, as observed in cat V1 (Haider et al., 2010) and mouse V1 (W. Zhu et al., 2009). In summary, our results show that the inhibitory processes in our models suppress redundant spikes which convey little information about the current stimulus (Kremkow et al., 2016).

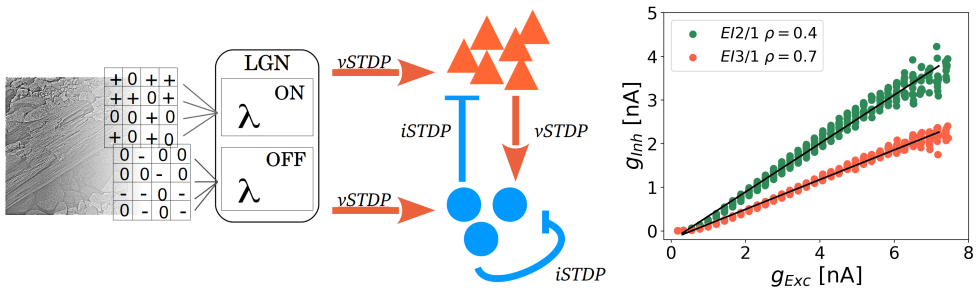
Input encoding quality benefits from plastic inhibition A fundamental purpose of sensory systems is to provide reliable information about relevant environmental stimuli. To compare our model with existing sparse coding models, in terms of stimulus encoding, we calculated the image reconstruction error (IRE), which measures the mean-square error between the input image and its reconstruction obtained by linear decoding (see **Methods**). The $EI2/1$ model with active inhibition during learning showed the lowest reconstruction error value (0.74) (**Fig. 9a**). By contrast, a substantially higher reconstruction error was observed for the $noInh$ model (1.06). Blocking inhibitory currents after circuit development caused a slight increase in the IRE to a value of 0.79 for the $blockInh$ model. Together, these results indicate that the diversity of receptive field shapes and orientations contribute to the average reconstruction accuracy.

Despite our observation about the role of feedback inhibition for the emergence of tuning diversity, the necessity of plastic inhibition compared to fixed inhibition during learning remains unclear. To analyze if plastic inhibition has a measurable effect during learning, we used shuffled weight matrices from a successfully learned $EI2/1$ model for all connections as a new initial condition, and deactivated plasticity selectively at specific connections for four model variations: Only in the inhibitory feedback connections (**9b 2**), in the two possible excitatory feed-forward connections to the inhibitory population (**9b 3**), and in the lateral inhibitory and excitatory to inhibitory connections (**9b 4**). To verify that learning is successful with the shuffled pre-learned weights, we trained one model variation where all connections are plastic (**9b 1**).

Our results show that if only the feedback inhibitory to excitatory connections are fixed, the reconstruction error increases from 0.70 (see (**9b 1**), where every connection is plastic) to 0.95 (**Fig. 9b 4**). We observe a similar error (0.92) when the excitatory connection from the LGN input to the inhibitory population is fixed (see **Fig. 9b 3**). This shows that the plasticity of both the inhibitory feedback connections and the excitatory feed-forward connection to the inhibitory population leads to a better input representation.

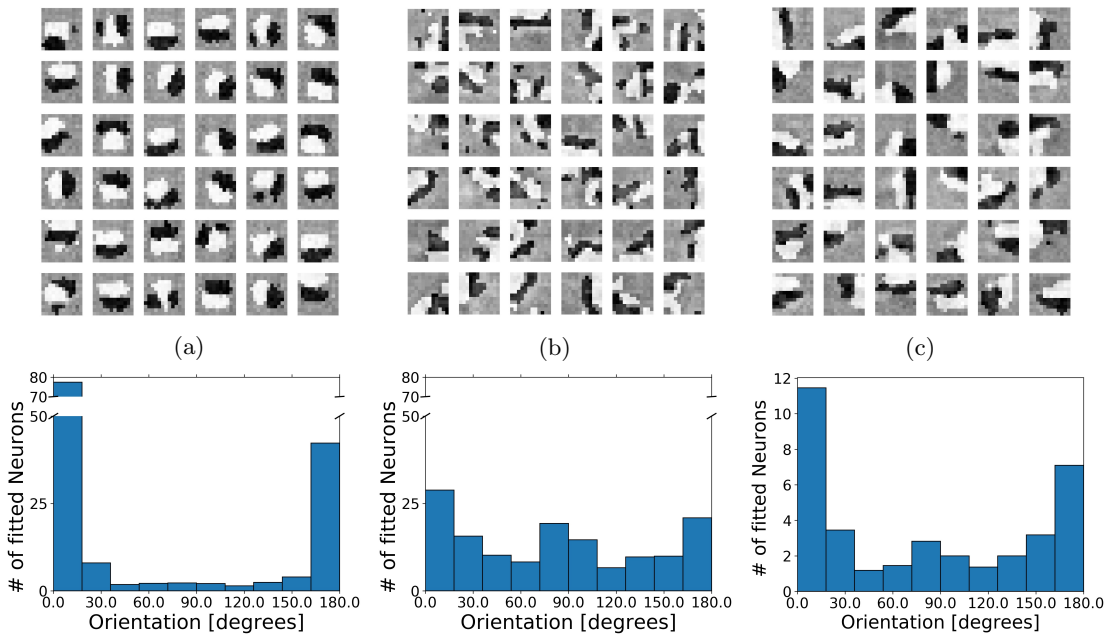
Interestingly, the reconstruction error remains small (0.71) if both, the connection from the excitatory to the inhibitory population and the lateral inhibition are fixed (see **Fig. 9b 2**). This shows that, even with fixed lateral inhibition, plasticity at the feed-forward path from LGN to inhibitory and from the inhibitory to the exitatory neurons is sufficient for the emergence of selective interneuron activity, which is essential for a reliable input representation. As an additional control to evaluate the effect of lateral inhibition, we completely deactivated the lateral inhibitory synapses during learning in a model where all other connections are plastic and measured an IRE of 0.83 (averaged across five simulations).

As explained above the input encoding benefits mainly from the distribution of the receptive fields. Therefore, we conclude that plastic feed-forward and feedback inhibition is essential for the process of developing receptive fields with diverse shapes and orientations, to improve input encoding.



(a) (b)

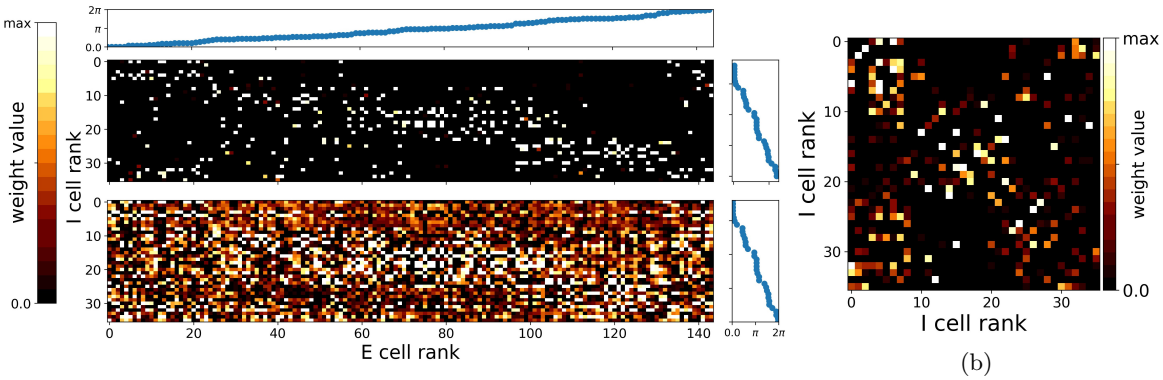
Figure 1: **Network with excitatory and inhibitory plasticity rules. (a)** Whitened image patches of size 12x12 were converted to Poisson spike trains by setting the firing rates of LGN ON- and OFF-populations to the positive and the negative part of the pixel values, respectively. Feed-forward inputs from LGN project both onto excitatory and inhibitory V1 populations, which are mutually connected. The circuit therefore implements both feed-forward and feedback inhibition. Inhibitory interneurons receive additional recurrent inhibitory connections. All excitatory synapses (orange) changes via the voltage-based STDP rule (vSTDP) (Clopath et al., 2010). All inhibitory synapses (blue) changes via the inhibitory STDP rule (iSTDP) (Vogels et al., 2011). Connectivity patterns are all-to-all. Population sizes are: LGN, 288 neurons; V1 excitatory, 144 neurons; V1 inhibitory, 36 neurons. Neurons in the LGN population showing Poisson activity and are split into ON- and OFF- subpopulations. **(b)**: Post-synaptic target firing rate of the iSTDP rule (ρ) controls the excitation to inhibition ratio at excitatory cells. For the $EI2/1$ model (green dots) a value of $\rho = 0.4$ leads to a higher inhibitory current than $\rho = 0.7$ for the $EI3/1$ model.



(a) (b) (c)

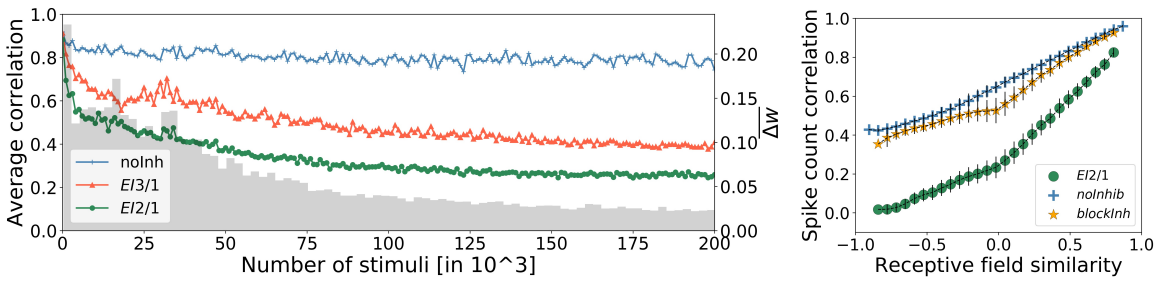
(d) (e) (f)

Figure 2: **Tuning diversity requires inhibition during learning.** Learned response profile of 36 excitatory neurons from the *noInh* model **(a)**, of 36 excitatory neurons from the $EI2/1$ model **(b)**, and of all 36 inhibitory neurons from the $EI2/1$ model **(c)**, measured with the spike triggered average method. Lighter pixels represent positive values and darker values represent negative values. Histogram of the spatial orientation across 10 model runs, of the *noInh* model's excitatory population **(d)**, the $EI2/1$ model's excitatory population **(e)**, and the the $EI2/1$ model's inhibitory population **(f)**. The spatial orientation are measured by fitting the neuronal response profile with a Gabor function (see **Methods**).



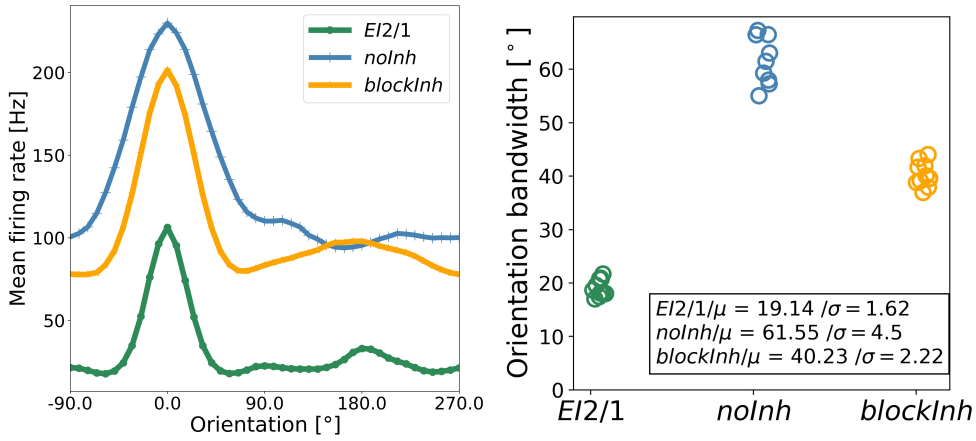
(a) (b)

Figure 3: **Synaptic connections reflect tuning similarity.** Weight matrices from excitatory to inhibitory population (and vice versa) **(a)**, sorted over the receptive field orientation, and for the lateral inhibition **(b)**. **a**,Top: Weights from excitatory to inhibitory population. **a**, Bottom: Weights from inhibitory to excitatory population. For display, all weight matrices were normalized by the maximum value. All weights from the $EI2/1$ model.



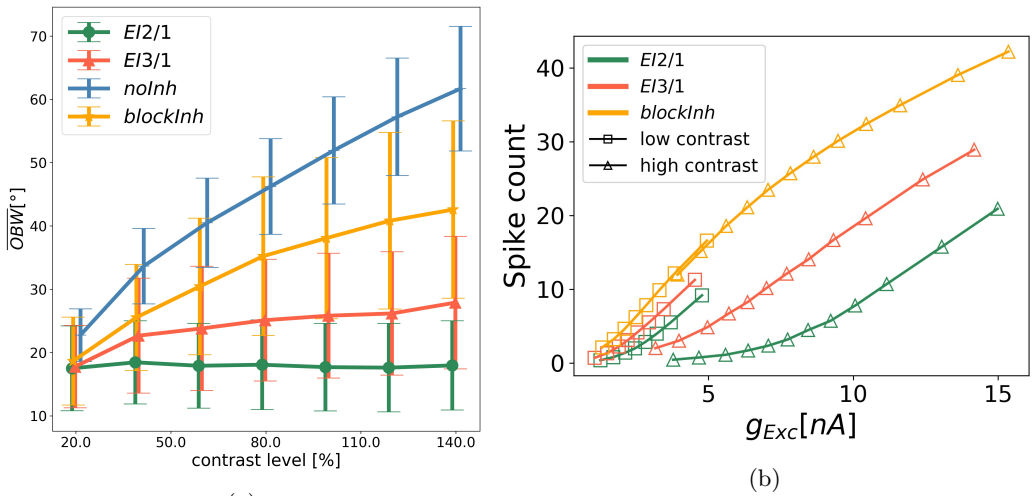
(a) (b)

Figure 4: **Inhibitory strength influence the response decorrelation. (a)** The development of mean response correlation and weight change at the LGN excitatory synapses across learning. Stronger inhibition, in the $EI2/1$ model, leads to a stronger decorrelation of the neuron responses during learning (compare green with red ($EI3/1$) line). Mean response correlation changed only very slightly without inhibition (blue line). **(b)** Response correlation is higher for neurons with more similar receptive fields. Blocking inhibition (yellow line) after learning reveals that. Inhibition leads to a overall decrease of the response correlation (green line).



(a) (b)

Figure 5: **Inhibition determines tuning curve sharpening. (a)** Average tuning curve of all excitatory cells in the $EI2/1$ model, the corresponding counterpart with blocked inhibition, and the no inhibition model. **(b)** The orientation bandwidth (OBW) of cells in all three models. Every point represents the average OBW resulting from model simulation. A smaller OBW means a more sharp tuning curve.



(a) (b)

Figure 6: **Response gain control by inhibition. (a)** Mean OBW as a function of the contrast level in the input. Whiskers represent the standard deviation. Data from the $EI2/1$ model (green line), $EI3/1$ model (red line), $noInh$ model (blue line) and $blockInh$ model (orange line). **(b)** Spike count as a function of the excitatory input current for the $EI2/1$ model (green line), the $EI3/1$ model (red line) and the $blockInh$ model (orange line). Data are taken from the sinusoidal tuning curve measurement, sorted ascending over the input current. Squares are data from low input contrast level and triangles are data from high input contrast level.

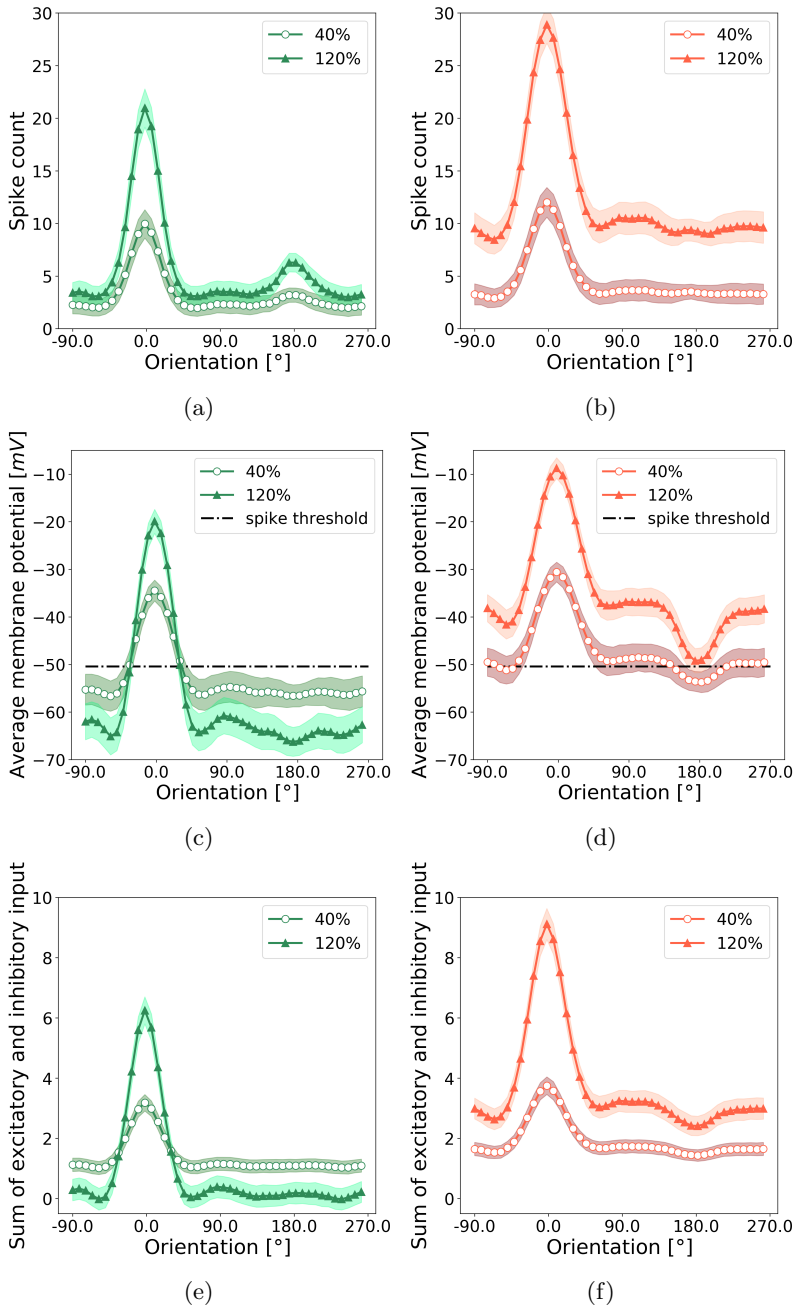


Figure 7: **Emergence of contrast-invariant responses** **(a)** Average neural tuning curve for low and high contrast stimuli in the $EI2/1$ model, **(b)** and the $EI3/1$ model. **(c)** Average membrane potential (averaged across all neurons in the excitatory population) as a function of orientation and contrast level for the $EI2/1$ model, **(d)** and the $EI3/1$ model. **(e)** Sum of the excitatory and inhibitory input current as a function of orientation and contrast level for the $EI2/1$ model, **(f)** and the $EI3/1$ model.

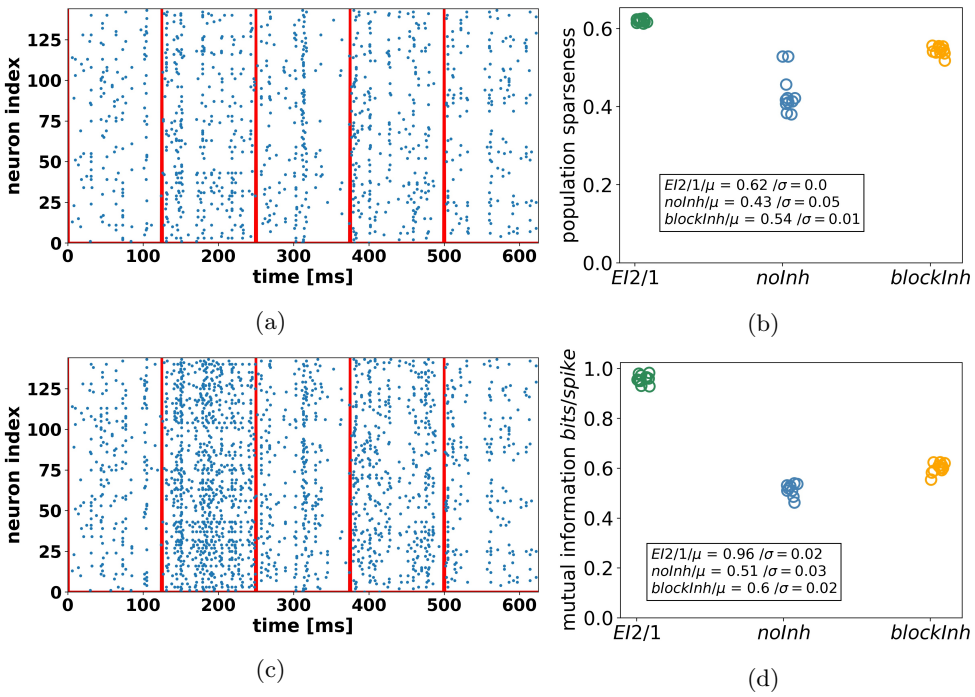


Figure 8: **Sparse and efficient input representation through the inhibitory circuit. (a)** Raster plot of the excitatory population for the $EI2/1$ model and for the *blockInh* model (**(c)**). Spikes are recorded on the same five natural image patches. The red lines show the stimulus onset. **(b)** Population sparseness for the $EI2/1$, the *blockInh*, and the *noInh* model, averaged over 10.000 natural scene patches. A higher value represents a higher sparseness of population activity. **(d)** Mutual information in *bits/spike* for the same three models as in **(b)**. **(b)**,**(d)** shows data from eleven independent simulations per model configuration.

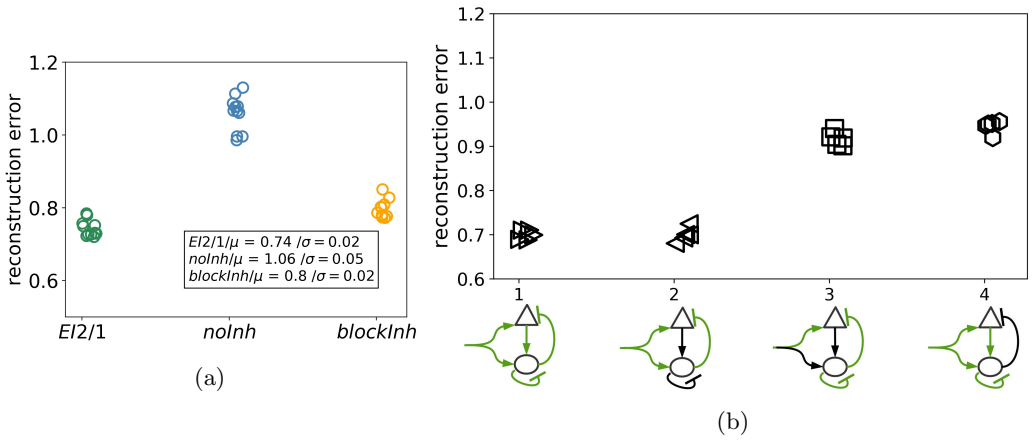


Figure 9: **Plastic inhibition during learning improves input encoding quality. (a)** Image reconstruction error (IRE) for the $EI2/1$ model (green dots), the *blockInh* model (orange dots), and the *noInh* model (blue dots). IRE is calculated as the mean-square error between input image and the reconstruction. Better reconstruction is represented by a smaller value. Data shown from eleven independent simulations per model configuration. **(b)** Image reconstruction error for model variations with different combinations of plastic and fixed excitatory and inhibitory synapses. Only in the first two variations (black triangles), the feed-forward inhibition and the inhibitory feedback are plastic. Plastic synapses indicated by green connections and fixed synapses by black connections.

3 Discussion

Our model suggests that a single underlying mechanism - the interaction of excitatory and inhibitory plasticity - can explain the stable emergence of reliable and efficient input encoding. We have shown that in particular, the combination of plastic inhibitory feedback and plastic feed-forward inhibition has an influence on shaping the receptive fields. This is in line with recent physiological findings that inhibitory plasticity influences the mode of operation of excitatory neurons (for example the excitability) (Griffen & Maffei, 2014; Wang & Maffei, 2014; Khan et al., 2018; Znamenskiy et al., 2018), or influences the occurrence of LTP and LTD (Paille et al., 2013; Griffen & Maffei, 2014; Mongillo & Loewenstein, 2018). Previous models based on STDP rules, which have demonstrated the emergence of V1 simple cells, made several simplifications in terms of the learning dynamics (Savin et al., 2010; Zylberberg et al., 2011; King et al., 2013), or consider plasticity only for a subset of projections (Sadeh et al., 2015; Miconi et al., 2016). These assumptions make it difficult to investigate the influence of plastic feed-forward and feedback inhibition on network dynamics and input encoding. For example, the observation of response decorrelation is a direct consequence of the chosen learning mechanism (Zylberberg et al., 2011; King et al., 2013). Other learning rules have been designed to optimize the mutual information between input and output (Savin et al., 2010). This suggests that a more detailed model of V1 circuit development is necessary to understand the dynamics between excitation and inhibition during the developmental process. To advance our understanding of this process, we investigated a spiking network model of V1 simple cell development, based on two phenomenological learning rules implemented at all synaptic projections.

Feed-forward and feedback inhibitory plasticity improves representational efficiency Our results show that plastic inhibitory feedback as well as plastic feed-forward inhibition influence the development of V1 simple cells and improve representational efficiency. Inhibitory plasticity has been reported in numerous physiological studies (Froemke et al., 2007; Carvalho & Buonomano, 2009; Kullmann et al., 2012; Wang & Maffei, 2014; D'Amour & Froemke, 2015; Khan et al., 2018). Previous model studies suggest a role for inhibitory plasticity in controlling the balance between excitation and inhibition (Vogels et al., 2011; Litwin-Kumar & Doiron, 2014), or in enabling stability in recurrent networks (Litwin-Kumar & Doiron, 2014; Sprekeler, 2017). However, there is ongoing discussion about the necessity and role of inhibitory plasticity during learning a functional sensory code book (Griffen & Maffei, 2014; Srinivasa & Jiang, 2013; Sprekeler, 2017), and this issue has received limited attention in model studies so far.

In a model based on a combination of STDP and inhibitory STDP learning rules, Litwin-Kumar & Doiron (2014) showed that inhibitory plasticity is necessary for stable learning in a network with recurrent excitatory connections. Their study used a generic cortical network receiving non-plastic input from a set

of 20 artificially stimuli, which in turn resulted in the formation of 20 assemblies representing the input stimuli. They emphasized that inhibitory plasticity acted to equilibrate firing rates in the network, such that different assemblies (each coding for one stimulus) received different amounts of inhibition, preventing dominant activity of single assemblies. Our results of a feature-specific strength of inhibition generalize their finding of firing rate heterogeneity induced by iSTDP from an "assembly code", in which different stimuli rarely overlap, to the quasi-continuous space of natural visual stimuli. This supports the necessity of the interaction of inhibitory and excitatory plasticity during the development of the visual cortex.

Emergence of a self-organized balance of excitation and inhibition We observed in our model that the inhibitory input current to a neuron is proportional to the excitatory input, when the currents are averaged across the duration of a stimulus. However, as we did not observe an equal strength between these currents, excitation is dominant in our network. This indicates a detailed and loose balance (for definition see, Hennequin et al. (2017)) between excitation and inhibition in our network. While a detailed balance has been reported in rat auditory cortex (Dorrn et al., 2010), it is still under discussion if a more loose or tight balance exists in the primary visual cortex of higher mammals (Froemke, 2015). Recent model studies suggest a tight balance between inhibition and excitation (Denève & Machens, 2016) or rather an inhibitory dominated network for stable learning in a network with recurrent excitatory synapses (Litwin-Kumar & Doiron, 2014; Sadeh et al., 2015; Miconi et al., 2016). However, most of these models investigate excitation-inhibition balance in a singe-neuron setup (Denève & Machens, 2016), or set a subset of synaptic connections fixed (Litwin-Kumar & Doiron, 2014; Sadeh et al., 2015; Miconi et al., 2016). Interestingly, we observed that the ratio between excitation and inhibition changes in our network for different contrast levels of sinusoidal grating stimuli, up to a $1:1$ balance for the highest contrast level for the $EI2/1$ model. This shows that the balance between excitation and inhibition is input-specific.

Inhibition implements a gain control mechanism and shapes tuning curves Previous physiological studies found that parvalbumin-expressing (PV) interneurons have a divisive impact on the gain function of pyramidal neurons in the visual cortex, to implement a contrast gain control mechanism (Atallah et al., 2012; Wilson et al., 2012; Y. Zhu et al., 2015). In our model we observed that the ratio between excitatory and inhibitory currents influences the response of the neuron towards its input. Consequently, feedback inhibition implements a gain control mechanism for the excitatory neurons.

Savin et al. (2010) proposed a rapid intrinsic plasticity mechanism to adapt the neuronal gain function to optimize the information transmission between input stimuli and neuronal output. They suggested that the emergence of V1 simple cell receptive fields depends on the interplay between the adaption of the neuronal gain function and the synaptic plasticity (Savin et al., 2010). By contrast, in our network,

changes in neuronal gain curves are caused by feedback inhibition, which adapts at the fast time scale of synaptic plasticity to maintain a given target rate.

In our model, when blocking inhibition after learning, we observed an increase not only in the baseline activity, but also in the orientation bandwidth (OBW). This demonstrates a sharpening of tuning curves by inhibition, similar to the observation of Katzner et al. (2011), where inhibitory synapses in cat primary visual cortex were blocked with gabazine. Interestingly, PV cells seem not to affect the sharpening of tuning curves (Atallah et al., 2012; Wilson et al., 2012), whereas somatostatin-expressing neurons (SOM) sharpen neuronal responses (Wilson et al., 2012). This demonstrates the influences of the different inhibitory neuron types (Markram et al., 2004), which must be taken into account in future models.

Shift in the E/I balance leads to the spontaneous emergence of contrast invariant tuning curves As a consequence of the contrast gain mechanism by inhibition, our model shows the spontaneous emergence of contrast invariant orientation tuning (Skottun et al., 1987; Troyer et al., 1998; Finn et al., 2007). Early modeling studies have proposed feed-forward inhibition to implement a push-pull inhibitory mechanism for the emergence of contrast-invariant tuning curves (Troyer et al., 1998; Ferster & Miller, 2000). Despite the fact that our network contains feed-forward inhibition, we did not observe a push-pull inhibitory effect, in other words, anti-correlation of excitation and inhibition (Anderson et al., 2000). A direct comparison of the excitatory and inhibitory input current for the contrast invariance task shows a simultaneous increase and decrease of excitation and inhibition, caused by the detailed balance in our network. We have observed that for the $EI2/1$ model, inhibitory input currents increase more rapidly than excitatory currents at higher contrast levels and non-preferred orientations. This results in a shift from a two-to-one ratio of excitation to inhibition to a one-to-one ratio between excitation and inhibition, and implements a contrast-dependent modulation of the neuron's gain curve. This shows that the emergence of contrast-invariant tuning curves is an inherent effect of the ratio between excitation and inhibition in our network. A contrast-dependent shift in the balance between excitation and inhibition has been reported in the visual cortex of awake mice (Adesnik, 2017). Although the influence of inhibition on the neuronal gain function for the emergence of contrast invariance is in line with previous assumptions (Mitchell & Silver, 2003; Finn et al., 2007), recent studies have proposed that changes in the neuronal gain are caused by response variability in the afferent thalamic path (Sadagopan & Ferster, 2012; Priebe, 2016).

Sparseness and metabolic efficiency benefit from E/I balance We observed that in the $EI2/1$ model, the standard deviation of the membrane potential increases for non-preferred orientations. Together with the observed asynchronous spiking behavior, we conclude that the balance of inhibition and excitation leads to a more irregular spiking behavior. Previous work suggests that a more irregular activity and irregular membrane potential behavior is related to improved metabolic efficiency in terms of efficient

input encoding (Denève & Machens, 2016). Our observations agree with these findings, because the efficiency of information transmission in our network mainly benefits from the ratio between excitatory and inhibitory currents in the stable network.

An established approach in terms of input encoding efficiency is the concept of sparse coding (Rolls & Tovee, 1995; Vinje & Gallant, 2000; Tolhurst et al., 2009). However, in recent years, it has been discussed how the level of sparseness reported in physiological experiments is influenced by animal age and the level of anesthesia (Berkes et al., 2009), and the benefit of highly sparse codes for information processing has been questioned (Wiltschut & Hamker, 2009; Barak et al., 2013; Spanne & Jörntell, 2015). Overall, the intermediate sparseness values observed in our model are in agreement with experimental findings (Berkes et al., 2009; Froudarakis et al., 2014).

Structured connectivity caused by inhibitory and excitatory plasticity Previous physiological studies have shown that inhibitory interneurons are connected in a nonspecific manner to other cells in their surrounding (Harris & Mrsic-Flogel, 2013). However, recent studies observed that inhibitory PV cells develop strong connections to excitatory cells with similar orientations (Znamenskiy et al., 2018), and that neurons with similar preferred orientations have a higher probability for recurrent connections (Ko et al., 2011; Cossell et al., 2015).

We observed a similar connectivity pattern in our network, namely, the appearance of strong connectivity between co-tuned neurons. King et al. (2013) also obtained a structured connectivity between co-tuned excitatory and inhibitory cells in a spiking network. While King et al. (2013) achieved this goal by designing a suitable learning rule for the synaptic projections involving inhibitory neurons, we observed the appearance of strong connectivity as an emergent property of our model architecture based on detailed phenomenological rules.

Stable learning despite limitations of simultaneous excitatory and inhibitory plasticity Previous studies have mentioned the difficulty to achieve a certain level of inhibition in a network with inhibition and plastic excitatory synapses (Zenke & Gerstner, 2017; Hennequin et al., 2017). We next discuss the behavior of the selected learning rules more in detail to show some of the difficulties during the interaction of excitatory and inhibitory plasticity, and discuss the limitations of our modeling approach.

For the excitatory learning rule, Clopath et al. (2010) have shown that a lower input firing rate leads to bigger receptive fields, as a compensatory effect of the homeostatic mechanism. This mechanism is controlled by the long-term postsynaptic membrane potential in relation to a reference value. If the membrane potential is too low, less long-term depression (LTD) in relation to long-term potentation (LTP) occurs, and the weights will increase. Otherwise, if the membrane potential is too high, a higher amount of LTD will occur to decrease the weights. Consequently, for a lower input firing rate, more weights will

increase, saturating at their maximum, to achieve a specific postsynaptic activity.

The homeostatic mechanism of the inhibitory rule (Vogels et al., 2011) strengthens the inhibition if the postsynaptic activity is too high, with respect to a target firing rate (ρ), or decreases the weight otherwise. In our network, the postsynaptic membrane potential is a result of the difference between the incoming excitatory and inhibitory current, such that a reduction in the membrane potential through inhibition is comparable to a reduction through less presynaptic spikes. The operation of both homeostatic mechanisms on the postsynaptic activity leads to a competition between weight changes at excitatory and at inhibitory synapses and should lead to bigger receptive fields, or, in the worst case, to a saturation of all synapses to their maximum value.

However, we observed the emergence of stable receptive fields and stable connections between the populations. Additionally, our results show a reduction in the mean activity, caused by inhibition, without causing bigger receptive fields. We assume that in contrast to a reduction in the input, what leads to a proportional reduction on the postsynaptic neuron, the inhibitory current leads to a more irregular, or fluctuating, behavior of the membrane potential (Vogels et al., 2005). To allow LTP at excitatory synapses, the membrane potential must be higher than θ_+ ($= -45.3mV$), which is slightly above the steady-state spiking threshold ($V_{T_{rest}} = -50.4mV$). But if the membrane potential is hyperpolarized by inhibition, it falls below the LTP threshold: No LTP occurs, and the weights will not increase to the maximum. Additionally, we observed that the interplay of the excitatory and inhibitory rules are mainly influenced by the magnitude of learning rates. In particular, a higher excitatory or higher inhibitory learning rate led to the saturation of all synapses, as an effect of the competition between both homeostatic mechanisms. How fast the synaptic weight changes depends not only on the magnitude of learning rates, but also on the number of spikes, that is, the number of learning events. Therefore, the learning rates for the $noInh$ model is smaller, to compensate the higher activity in the neuron populations. Finally, the competitive pressure between learning rules is controlled by the postsynaptic target activity in the inhibitory learning rule. Smaller values of ρ enhances the inhibitory pressure on the post-synaptic neuron to achieve a lower firing rate and can also lead to an unlimited growth of synaptic weights. This limited the amount of inhibition that can emerge in the network.

Conclusion To the best of our knowledge, our simulations are the first demonstration of the parallel emergence of fundamental properties of the primary visual cortex such as sparse coding, contrast invariant tuning curves and high accuracy input representation, in a spiking network with spike timing-dependent plasticity rules. A central finding of our study is that the emergence of representational efficiency (such as tuning diversity) and metabolic efficiency (such as the numbers of spikes to represent a specific input stimuli) require plasticity at feed-forward and feedback inhibitory synapses. This emphasizes the role of

inhibition in the shaping of neuronal responses (Isaacson & Scanziani, 2011; Stringer et al., 2016; Sprekeler, 2017) and in the development of reliable and efficient input encoding.

Acknowledgment

This work has been supported by a grant from the European Social Fund (ESF) and the Free State of Saxony, and by a CRCNS US-German-Israeli collaboration on computational neuroscience grant "Multi-level neuro-computational models of basal ganglia dysfunction in Tourette syndrome" funded by the Federal Ministry of Education and Research, Germany (BMBF 01GQ1707).

Author Contributions

Conceptualization	F.H.H., M.T., L.G., R.L.
Methodology	M.T., L.G., R.L.
Software	R.L.
Validation	R.L.
Formal Analysis	R.L., L.G.
Investigation	R.L., L.G., M.T.
Resources	F.H.H.
Writing – Original Draft Preparation	R.L., L.G.
Writing – Review & Editing	F.H.H., L.G., M.T., R.L.
Visualization	R.L.
Funding Acquisition	F.H.H.

Declaration of Interest

The authors declare no competing interests.

4 Methods

The first part of this section (4.1–4.5) describes the network architecture including the neuron model and learning rules. In the second part (4.6), we explain the analysis methods used to characterize neuronal responses. The model has been implemented in Python 3.6, using the ANNarchy simulator (Vitay et al., 2015), with a simulation time step of $dt = 1ms$ (Euler integration). The neuronal simulator is available from `https://bitbucket.org/annarchy/annarchy` . The implementation of the adaptive exponential integrate-and-fire neuron model and the voltage-based triplet STDP learning rule from Clopath et al. (2010) based mainly on the re-implementation by Larisch (2019).

4.1 Network architecture

Our network model, which is inspired by the primary visual cortex and its inputs from LGN, consists of three populations of spiking neurons (**Fig.1a**): An input layer representing LGN, and excitatory and inhibitory populations of V1, each receiving feed-forward inputs from LGN. The V1 populations are mutually interconnected via excitatory or inhibitory synapses, respectively. The circuit therefore implements both feed-forward and feedback inhibition, in agreement with anatomical findings (Isaacson & Scanziani, 2011). Inhibitory interneurons receive additional recurrent inhibitory connections. All projections follow an all-to-all connectivity pattern, excluding self inhibitory feedback connections.

The LGN layer consists of 288 neurons showing Poisson activity and is split into ON- and OFF-subpopulations. For the V1 excitatory population (144 neurons) and the inhibitory population (36 neurons), we used adaptive exponential integrate-and-fire neurons (Sec. 4.3). The size of the inhibitory population was chosen to match the 4:1 ratio between excitatory and inhibitory neurons found in visual and striate cortex (Beaulieu et al., 1992; Markram et al., 2004; Potjans & Diesmann, 2014).

All synaptic connections within our model are plastic and were randomly initialized. They change their weight based on either the voltage-based STDP-rule propsed by Clopath et al. (2010) (excitatory connections) or the symmetric iSTDP-rule proposed by Vogels et al. (2011) (inhibitory connections; Sec. 4.5).

4.2 Network input

As network input, we used whitened patches from natural scenes (Olshausen & Field, 1996, 1997). Each patch was chosen randomly, with a size of 12 by 12 by 2 pixels (Wiltschut & Hamker, 2009). The third dimension corresponds to the responses of ON- and OFF-cells. To avoid negative firing rates, we mapped positive pixel values to the ON-plane, and the absolute value of negative pixels to the OFF-plane. Every patch was normalized with the maximum absolute value of the corresponding natural scene. The firing rate of each Poisson neuron represents the brightness value of the input pixels. The firing rate associated

to the (rarely occurring) maximum pixel value was set to $125Hz$. We stimulated the network with 400.000 patches during training, with a presentation time of $125ms$ per patch, corresponding to around $14h$ of simulated time. To avoid any orientation bias in the input, the patch was flipped around the vertical or horizontal axis independently with 50% probability (Clopath et al., 2010).

4.3 Poisson neuron model in LGN

For modeling convenience, we generated Poisson activity in LGN neurons by injecting brief voltage pulses, generated by a Poisson process, into a binary spiking neuron model, such that each voltage pulse input triggered a spike. This simplified the numerical calculation of a spike trace required for the learning rule, while preserving the precise timing of spikes drawn from a Poisson process. The spike trace $\overline{x}_i$ is updated whenever the presynaptic neuron i spikes, and decays exponentially: $X_i(t) = 1$ if a spike is present at time t, and $X_i(t) = 0$ otherwise.

$$\frac{du}{dt} = I_{Poisson} \tag{1}$$

$$\tau_x \frac{d\overline{x}_i}{dt} = -\overline{x}_i + X_i \tag{2}$$

4.4 Adaptive exponential integrate-and-fire neurons in V1

For the neurons in the V1 excitatory and inhibitory layer, we used a variant of the adaptive exponential integrate-and-fire model as described by Clopath et al. (2010). In this model, the membrane potential u is influenced by the following additional dynamical variables: An adaptive spike threshold, V_T, a hyperpolarizing adaptation current, w_{ad}, and a depolarizing afterpotential, z. Excitatory and inhibitory synaptic currents are denoted by I_{exc} and I_{inh}. For an explanation of constant parameter values as used by Clopath et al. (2010), see Table 1. The full equation for the membrane potential is

$$C\frac{du}{dt} = -g_L(u - E_L) + g_L \Delta_T e^{\frac{u - V_T}{\Delta_T}} - w_{ad} + z + I_{exc} - I_{inh} \tag{3}$$

As the triplet voltage STDP rule is sensitive to the precise time course of the membrane voltage, including the upswing during a spike, the magnitude of weight changes depends on the implementation details of the after-spike reset. To avoid long simulation times associated with smaller time steps, we opted for the following simplified treatment of the spike waveform which reproduced the results reported by Clopath et al. (2010): Whenever the membrane potential u exceeded the spike threshold, u was held at a constant value of $29mV$ for $2ms$, and then reset to the resting potential E_L. We obtained highly similar results from an alternative implementation, in which the after-spike reset was immediately applied when the spike threshold was crossed, with an additional update of the voltage traces by the amount expected from

a $2ms$-long spike (data not shown).

The reset value for the spike threshold is $V_{T_{max}}$, with exponential decay towards the resting value $V_{T_{rest}}$, with a time constant τ_{V_T} (Eq. 4):

$$\tau_{V_T}\frac{dV_T}{dt} = -(V_T - V_{T_{rest}}) \tag{4}$$

The afterpotential z has a reset value of I_{sp} and decays to zero (Eq.5). Further, the variable w_{ad} is incremented by the value b and decays exponentially (Eq. 6).

$$\tau_z\frac{dz}{dt} = -z \tag{5}$$

$$\tau_{w_{ad}}\frac{dw_{ad}}{dt} = a(u - E_L) - w_{ad} \tag{6}$$

The model proposed by Clopath et al. (2010) assumed excitatory synaptic input in the form of voltage pulses. For modeling convenience, we approximated this setting by current-based excitatory synapses with a short time constant of $1ms$. Inhibitory synaptic currents decayed with a slower time constant of $10ms$. Both synaptic currents are incremented by the sum of synaptic weights of those presynaptic neurons which spiked in the previous time step:

$$\begin{aligned}\tau_{I_{exc}}\frac{dI_{exc}}{dt} &= -I_{exc} + w_i^{exc}\sum_{i\in Exc}\delta(t - t_i^{'})\\ \tau_{I_{inh}}\frac{dI_{inh}}{dt} &= -I_{inh} + w_j^{inh}\sum_{j\in Inh}\delta(t - t_j^{'})\end{aligned} \tag{7}$$

where $t_i^{'}$ denotes the spike time of presynaptic neuron i, and δ is the indicator function with $\delta(0) = 1$.

4.5 Synaptic plasticity

4.5.1 Voltage-based triplet STDP at excitatory synapses

Plasticity at excitatory connections (LGN to Exc. and Exc. to Inh.) follows the voltage-based triplet STDP rule proposed by Clopath et al. (2010). We here repeat the essential features of this plasticity model. The neuronal and synaptic variables describing the development of the weight from a presynaptic neuron with index i onto a given postsynaptic neuron are: X_i, the presence of a presynaptic spike; $\overline{x}_i$, the presynaptic spike trace (Eq. 2); u, the postsynaptic neuron's membrane potential; and two running averages of the membrane potential, $\overline{u}_+$ and $\overline{u}_-$, defined as follows:

$$\tau_+\frac{d\bar{u}_+}{dt} = -\bar{u}_+ + u, \tag{8}$$

where $\bar{u}_-$ is defined analogously, with the time constant τ_-. In addition, the learning rule includes a homoeostatic term, $\bar{\bar{u}}$, which regulates the relative strength of LTD versus LTP, and which measures the

mean postsynaptic depolarization on a slower time scale:

$$\tau_{\bar{u}} \frac{d\bar{\bar{u}}}{dt} = [(u - E_L)^+]^2 - \bar{\bar{u}} \tag{9}$$

Here, $x^+ = \max(x, 0)$ denotes top-half rectification.

The full learning rule is given as the sum of the LTP term and the LTD term:

$$\frac{dw_i}{dt} = A_{LTP}\ \overline{x}_i (u - \theta_+)^+ (\overline{u}_+ - \theta_-)^+ - A_{LTD} \frac{\bar{\bar{u}}}{u_{ref}} X_i (\overline{u}_- - \theta_-)^+ \tag{10}$$

where A_{LTP} and A_{LTD} are the learning rates for LTP and LTD, θ_+ and θ_- are threshold parameters, and u_{ref} is a homeostatic parameter which controls the postsynaptic target firing rate. Clopath et al. (2010) have shown that this learning rule results in BCM-like learning dynamics (Bienenstock et al., 1982), in which a sliding metaplasticity threshold leads to the development of selectivity.

Following Clopath et al. (2010), for the LGN efferent connections, we equalized the norm of the OFF weights to the norm of the ON weights every $20s$. The weight development is limited by the hard bounds w^e_{min} and w^e_{max}.

4.5.2 Homeostatic inhibitory plasticity

Previous biological studies have observed spike timing-dependent plasticity of inhibitory synapses which differs from the well-known asymmetric STDP window (Caporale & Dan, 2008; D'Amour & Froemke, 2015). We therefore chose to implement the phenomenologically motivated, symmetric inhibitory STDP (iSTDP) rule proposed by Vogels et al. (2011) at all inhibitory synapses (Eq.11):

$$w(t + dt) = \begin{cases} w(t) + \eta(\overline{x}_{post} - \rho) & \text{if } t = t_{pre} \text{ (presynaptic spike)} \\ w(t) + \eta \overline{x}_{pre} & \text{if } t = t_{post} \text{ (postsynaptic spike)} \end{cases} \tag{11}$$

Here, η is the learning rate, and ρ is a constant which controls the amount of LTD relative to LTP. Further, Vogels et al. (2011) have shown that this learning rule has a homeostatic effect, and the parameter ρ controls the postsynaptic target firing rate. The variables $\overline{x}_{pre}$ and $\overline{x}_{post}$ are spike traces for the pre- and postsynaptic neurons, defined in analogy to Eq. (2), with time constants τ_{pre} and τ_{post}. In this plasticity rule, near-coincident pre- and post-synaptic spiking causes potentiation of weights, irrespective of their temporal order. By contrast, isolated pre- or postsynaptic spikes cause depression of weights. As for the excitatory learning rule, weights are bounded by w^i_{min} and w^i_{max}. For parameter values, see Table 1.

Global parameter values			
Parameter (values from Clopath et al. (2010))	Value	Parameter	Value
C, membrane capacitance	$281pF$	τ_z, spike current time constant	$40ms$
g_L, leak conductance	$30nS$	τ_{V_T}, spike threshold time const.	$50ms$
E_L, resting potential	$-70.6mV$	τ_x, spike trace time constant	$15ms$
Δ_T, slope factor	$2mV$	τ_{wad}, adaption time constant	$144ms$
$V_{T_{rest}}$, spike threshold at rest	$-50.4mV$	I_{sp}, spike current after spike	$400pA$
$V_{T_{max}}$, spike threshold after spike	$30.4mV$	a, subthreshold adaptation	$4nS$
w^e_{min}, min. excitatory weight	0.0	b, spike-triggered adaption	$0.805pA$
τ_-, time constant for $\overline{u}_-$	$10.0ms$	τ_+, time constant for $\overline{u}_+$	$7.0ms$
θ_-, plasticity threshold	$-70.6mV$	θ_+, plasticity threshold (LTP)	$-45.3mV$
Parameter (added)	Value	Parameter	Value
$\tau_{I_{exc}}$, excitatory input time const.	$1.0ms$	$\tau_{I_{inh}}$, inhibitory input time const.	$10.0ms$
Projection-specific parameters			
Parameter (custom values)	*LGNtoE*	*LGNtoI*	*EtoI*
$\tau_{\overline{\overline{u}}}$	$750ms$	$750ms$	$750ms$
w^e_{max}	5.0	3.0	1.0
w_{init} (bounds of random uniform distribution)	$[0.015, 2.0]$	$[0.0175, 2.15]$	$[0.0175, 0.25]$
A_{LTP} $(EI2:1, EI3:1)$	1.35×10^{-4}	5.4×10^{-5}	1.2×10^{-5}
A_{LTD} $(EI2:1, EI3:1)$	1.05×10^{-4}	4.2×10^{-5}	1.4×10^{-5}
A_{LTP} $(noInh)$	7.2×10^{-5}	n/a	n/a
A_{LTD} $(noInh)$	5.6×10^{-5}	n/a	n/a
$\overline{\overline{u}}_{ref}$	$60.0mV^2$	$55.0mV^2$	$55.0mV^2$

Table 1: **Parameters for the neuron model and excitatory synapses.** Note that for the *noInh* model, learning rates were reduced to compensate for the increased firing rates in the absence of inhibition.

	ItoE and ItoI	ItoE	ItoI
τ_{post}	$10.0ms$		
τ_{pre}	$10.0ms$		
w^i initial	0.0		
w^i_{min}	0.0		
w^i_{max}		0.7	0.5
η		10^{-5}	10^{-5}
ρ $(EI3:1)$		0.7	0.6
ρ $(EI2:1)$		0.4	0.6

Table 2: **Parameters for inhibitory synapses.**

4.5.3 Choice of parameter configurations

As our main goal is to determine the influence of inhibitory strength both on the formation of selectivity and on the dynamics of stimulus coding, we simulated our network using different parameter and network configurations. First, we used the above presented network, where the strength of the inhibitory feedback is controlled by the homeostatic parameter ρ. With $\rho = 0.4$ for the feedback inhibitory synapses, we achieved a ratio of excitation to inhibition (E/I-ratio) of approximately $2:1$ on patches of natural scenes (abbreviated as $EI2/1$). On one hand, a lower ρ would strengthen the inhibitory feedback, but caused unstable behaviour during learning. On the other hand, a higher ρ would weaken the inhibitory feedback of the model. With $\rho = 0.7$ we achieve a E/I-ratio of approximately $3:1$ on natural scene input (abbreviated as $EI3/1$), this led to similar but weaker characteristics for most of the experiments (**Fig.1b**). Because of this, the data are only presented for experiments, where the weaker inhibitory feedback lead to a significance difference.

Second, we simulated a purely excitatory feed-forward network without any inhibitory activity (abbreviated as $noInh$), as the learning rule proposed by Clopath et al. (2010) is capable of learning distinct shapes of receptive fields given different initial weights.

Further, to control for the dynamical effects of inhibition in the steady state following receptive field development, we simulated the effects of deactivating the inhibitory synaptic transmission in the $EI2/1$ model after learning (abbreviated as $blockInh$). All three model variations are based on the same network architecture, except that inhibitory weights differ in their strength or are deactivated. The different parameters for learning the models are shown in Table 1. To test the stability and the reproducibility of our results, we performed eleven runs of each model with randomly initialized synaptic weights.

To evaluate how inhibitory plasticity interacts with plastic excitation, we deactivated the plasticity for specific synapses for three model variations. First, we deactivated the plasticity only in the inhibitory feedback connections. Second, the plasticity was deactivated in both excitatory connections the inhibitory population. And we deactivated the plasticity in the connections from the excitatory to the inhibitory population and for the lateral inhibition. Additionally, we trained one model variation where all connections are plastic to validate, that the learning is successful with pre-trained, shuffled weight matrices. To ensure, that the same average amount of excitatory or inhibitory current is conveyed by the fixed synapses, we used shuffled weight matrices from previous simulations of the $EI2/1$ model for the respective synapses. No parameter changes were needed. To test the stability and reproducibility, we performed five runs of each variation.

4.6 Analysis methods

4.6.1 Receptive field mapping

Over the course of learning, the excitatory input weights from LGN to V1 develop based on the pre- and postsynaptic activity. It is therefore possible to obtain a good approximation of the neurons' receptive fields (RFs) by taking the weight matrix and reverting the ON-OFF mapping. To do this, we subtract the OFF-synapses from the ON-synapses and receive the receptive field. This is possible as either the ON- or the OFF-synapses can be activated by the input, so that the weights will also follow this distribution.

In addition to the visualization based on weight matrices, the receptive fields can also be revealed by probing the neurons with random stimuli. This approach has been successfully used in physiological research, in form of the spike triggered average (STA) (Ringach & Shapley, 2004; Schwartz et al., 2006; Pillow & Simoncelli, 2006). In this method, a neuron's receptive field is defined as the average of white noise stimuli, weighted by the stimulus-triggered neuronal activity. We applied this method on the learned neural network. We presented noise patches drawn from a normal distribution with $\mu = 15$, $\sigma = 20$ as input image to the network, and converted these to Poisson spike trains (cf. Sec. 4.2). Negative pixel values were set to zero, and the presentation time per patch was $125ms$. For each neuron, we recorded the number of spikes per stimulus and calculated the average across all stimuli, weighted by the number of postsynaptic spikes (Eq. 12).

$$STA = \frac{1}{N} \sum_{n=1}^{N} s(t_n) \tag{12}$$

Here, $s(t_n)$ is the input stimulus at time point t_n, when the nth spike has occurred, and N is the total number of postsynaptic spikes. Accordingly, stimuli evoking more spikes are higher weighted than stimuli evoking few or no spikes.

As we observed a high similarity between each neuron's STA and its ON-OFF receptive field, we concluded that the overall receptive field shape was not significantly influenced by inhibition. Thus, for simplicity, the feed-forward weight vectors can be used for further evaluations.

4.6.2 Gabor fits of receptive fields

As a first approximation, the receptive fields (RFs) of neurons in the primary visual cortex can be well described by Gabor functions (Jones & Palmer, 1987b). This is commonly used to describe their properties (Ringach, 2002; Zylberberg et al., 2011). We calculated the RFs of V1 neurons based on their LGN input weights, as described in Sec. 4.6.1. For each excitatory and inhibitory neuron, we then fit the parameters of a 2D-Gabor function ($g(x, y)$) to this feed-forward weight matrix, using least-squares minimization. The Gabor function is defined as followed (Eq. 13) and is similar to the one used in Ringach (2002),

extended by an offset parameter o.

$$
\begin{aligned}
g(x,y) &= o + A\exp(-\frac{x_p^2}{2\sigma_x^2} - \frac{y_p^2}{2\sigma_y^2})\cos(2\pi x_p f - \phi) \\
x_p &= (x - x_0)\cos(\theta) + (y - y_0)\sin(\theta) \\
y_p &= -(x - x_0)\sin(\theta) + (y - y_0)\cos(\theta),
\end{aligned}
\tag{13}
$$

where A denotes the amplitude, θ is the angle of the spatial orientation, σ_x and σ_y are the spatial extents, f is the spatial frequency, ϕ the phase, and x_0 and y_0 denote the position of the center.

We used the normalized mean squared error (NMSE) (Eq.14) to calculate the fitting error between the Gabor-function g and the weight vector w of a neuron (Spratling, 2012). The function normalizes the quadratic fitting error by the length of the weight vector and allows to compare error rates between different models. It allows to define a threshold until a RF is accepted as Gabor-like, we define this threshold as 0.5. Neurons with higher values have been excluded from evaluations based on the Gabor fit (see Results for details).

$$
NMSE = \frac{\sum_i (g_i - w_i)^2}{\sum_i w_i^2} \tag{14}
$$

4.6.3 Receptive field similarity

As mentioned above, the feed-forward weight vector approximates the receptive field of a neuron. To measure the similarity between two receptive fields, we calculate the cosine between their feed-forward weight vectors (Eq. 15).

$$
\cos(\phi_{i,j}) = \frac{W_i \cdot W_j}{|W_i||W_j|} \tag{15}
$$

A value near $+1$ indicates high similarity, values around zero describe orthogonal weight vectors, and values near -1 indicates inverted weight vectors (i.e., maximally overlapping RFs with opposite directional preference).

4.6.4 Tuning curves and orientation selectivity

The orientation selectivity is a well-studied characteristic of simple cells in V1 of mammals (Gilbert & Wiesel, 1990; Priebe & Ferster, 2008; Niell & Stryker, 2008) and thus, also a topic of interest for models of the visual cortex (e.g., Sadeh et al., 2014; W. Zhu et al., 2010; Tao et al., 2004). One possibility to quantify the orientation selectivity of a neuron is to measure its tuning curve (Ringach et al., 2002). For simple cells in the primary visual cortex, the orientation tuning curve describes the magnitude of responses evoked by a stimulus presented at different angles. In many biological studies, the tuning curves have been measured based on two-dimensional sinusoidal gratings (Anderson et al., 2000; Smith & Kohn, 2008; Ringach et al., 2002; Katzner et al., 2011). Therefore, we measured the responses to sinusoidal

grating stimuli, rotated in steps of 8°, with different spatial phases from $0rad$ to πrad, a different spatial frequencies from 0.05 up to $0.15cycles/pixel$, centred to the input space and with a presentation time of $125ms$.

Because of Poisson activity in the input layer, neuronal activity shows trial-to-trial fluctuations. Hence, we repeated every presentation 50 times, and calculated the mean across all 50 repetitions (or $6.25s$ presentation time). In contrast to the natural scene input used for training, the maximum input firing rate was set to $85.7Hz$. This was suitable to obtain sufficiently high activity levels.

To estimate tuning curve sharpness, we calculated the orientation bandwidth (OBW) for every neuron. The OBW is defined as the half-width of the tuning curve, at an activity level of $\frac{1}{\sqrt{2}}$ (approx. 70.7%) of the maximum (Ringach et al., 2002). Higher OBW values correspond to a broader tuning curve, and vice versa. Other definitions use the height at half-maximum, which does not change the overall result of this evaluation.

4.6.5 Neuronal gain curves

A neuron's gain function describes how neuronal activity is scaled by variations in the magnitude of excitatory inputs (Katzner et al., 2011; Isaacson & Scanziani, 2011). While an integrate-and-fire neuron receiving only excitatory inputs has a relatively static gain function (also called transfer function), controlled by the parameters of the neuron model, additional inhibitory inputs can modulate the effective input-to-output relationship. To characterize these inhibitory influences on gain curves, we recorded the excitatory synaptic currents and spiking activity evoked by sinus gratings (see Sec. 4.6.4), which we rotated from the orthogonal towards the preferred orientation of each neuron. Further, we changed the contrast of the input, by changing the pixels relative to the maximum input firing from $14.25Hz$ up to $100Hz$. As before, we presented each stimulus orientation for $125ms$, repeated 50 times ($6.25s$), and determined gain curves based on the average spike count across these 50 repetitions. We measured the spike count for each input degree and contrast strength and sorted the neuronal activity to the corresponding excitatory input, in ascending order.

4.6.6 Measurement of E to I ratio

To determine the ratio between excitatory and inhibitory input current, we measure both incoming currents for the excitatory population for 1.000 randomly chosen natural scenes. Every scene was presented for $125ms$ and was repeatedly shown for 100 times. We averaged the incoming currents over the input stimuli repetitions and sorted for each neuron and stimuli the excitatory input currents ascending with the related inhibitory currents. For better visualization, the currents are summarized into bins.

4.6.7 Sparseness

The sparseness value expresses the specificity of population codes and single neurons, both in experimental studies (Rolls & Tovee, 1995; Vinje & Gallant, 2000, 2002; Weliky et al., 2003; Tolhurst et al., 2009) and in model simulations (Wiltschut & Hamker, 2009; Zylberberg et al., 2011; King et al., 2013). It quantifies either the fraction of neurons which respond to a single stimulus, called population sparseness, or the number of stimuli to which a single neuron responds, called lifetime sparseness (Tolhurst et al., 2009). In the past, many different sparseness measurements are established (Rolls & Tovee, 1995; Hoyer, 2004). To measure the specificity of our network activity, we calculated the population sparseness after Vinje & Gallant (2000) (see Eq. 16).

$$S = \frac{1 - \frac{(\sum r_i/n)^2}{\sum(r_i^2/n)}}{1 - (1/n)} \tag{16}$$

where r_i is the activity of the ith neuron to a specific input and n the number of neurons in the neuron population.

By construction, sparseness values are bound between zero and one. If the neuron population has dense activity, i.e., most neurons are active to an input stimulus, the sparseness level approaches zero. By contrast, few active neurons of the population lead to a sparseness value close to one. As input, we used 30.000 natural scene patches, and determined sparseness values based on the firing rates of each neuron on each input patch.

4.6.8 Image reconstruction error

The network's coding performance following training can be measured by the difference between input images and their reconstruction from network activity. This method gives direct insight on how well visual input is represented by the network as a whole. This aspect was often not considered in previous biologically motivated circuit models of the primary visual cortex. We used the root mean-square error between one image of the natural scenes dataset from Olshausen & Field (1996) and the reconstructed one (cf. Zylberberg & DeWeese, 2013; King et al., 2013) (Eq. 17), termed image reconstruction error (IRE):

$$IRE = \sqrt{\frac{\sum_k (I_o - I_r)^2}{N}} \tag{17}$$

where N denotes the number of image pixels. To obtain the reconstructed image I_r, we subdivided the full image into patches of size 12×12, in an overlapping fashion (in increments of 3 pixels). We showed each patch 50 times for $125ms$ each, and recorded neuronal activities. We weighted the activity of each neuron by its feed-forward weights to obtain a linear reconstruction of each image patch, which we combined to reconstruct the full image. This approach is equivalent to calculating the IRE for individual patches, and calculating the root mean-square of these individual IRE values. To ensure that pixel values of the

reconstructed image were in the same range as the original image, we normalized the reconstructed as well as the original image to zero mean and unit variance (Zylberberg & DeWeese, 2013; King et al., 2013).

4.6.9 Mutual information

An information-theoretic approach to estimate the coding efficiency of the network is based on the mutual information between stimulus identity and neuronal activity (Dayan & Abbott, 2001; Dadarlat & Stryker, 2017). This measure allows to calculate the average information transmission per spike (Vinje & Gallant, 2002; Sengupta et al., 2013). To quantify information transmission, we calculated the mutual information, $I(s, r)$, between the stimulus identity and neuronal responses for each neuron, following Vinje & Gallant (2002):

$$I(s, r) = H(r) - H(r|s) \tag{18}$$

In Eq. 18, $I(s, r)$ is the mutual information carried between stimulus and response for a time bin of $125ms$ length, the duration of a single stimulus. For that purpose, we calculate the total response entropy, $H(r)$, and the conditional response entropy, also called stimulus-specific noise entropy, $H(r|s)$.

$$H(r) = -\sum_{j=0}^{\infty} p_j log_2(p_j) \tag{19}$$

$$H(r|s = k) = -\sum_{j=0}^{\infty} p_j^k log_2(p_j^k) \tag{20}$$

The total response entropy is given by Eq. 19. The variable p_j is the number of time bins containing exactly j spikes, divided by the total number of time bins, or stimuli. It follows from Eq. 19 that the total response entropy is maximal if all spike counts occur with equal probability (and, if they do, the number of possible spike counts increases the entropy). The noise entropy for a specific stimulus (see Eq. 20) describes the variability of the neuronal responses across repetitions of a single stimulus k. Every stimulus was repeated 100 times. Similar to the total response entropy, j is the number of spikes which occurred in response to a stimulus k. Here, p_j^k is the number of repetitions of stimulus k to which exactly j spikes are emitted, divided by the overall number of repetitions of that stimulus. To calculate the overall noise entropy of a neuron $H(r|s)$, we averaged the noise entropy across all stimuli. Information per spike was computed by dividing $I(s, r)$ by the mean number of spikes per stimuli, or time bins.

4.6.10 Discriminability

To evaluate how well the network responses allow to distinguish between any two input patches, in the presence of trial-to-trial (how much is the variance in the firing rate of a neuron to specific input (Shadlen & Newsome, 1998)) fluctuations induced by Poisson input, we calculated the discriminability index, d'

(e.g., Dayan & Abbott, 2001; Dadarlat & Stryker, 2017). The d' index measures the separation of two random distributions, and is closely related to the performance of a linear classifier assuming independent neuronal responses. Based on a random set of 500 natural scene patches, we calculated the d' by pairing the response on every patch to all other patches. For each pair of stimuli, s_1 and s_2, we presented each stimulus with $N = 100$ repetitions, and recorded the network responses of all $n = 144$ excitatory neurons for each repetition, obtaining the n-dimensional response vectors $s_1^{(i)}$ and $s_2^{(i)}$, $i = 1, \ldots, N$. We first calculated the mean activity of each cell in response to each stimulus, across the N repetitions (denoted by $\overline{s_1}$ and $\overline{s_2}$). We next projected each individual population response $s_1^{(i)}$ and $s_2^{(i)}$ onto the vector between these means, by taking the dot product between each response and the difference $\overline{s_1} - \overline{s_2}$:

$$\begin{aligned} \alpha_{s_1}^{(i)} &= s_1^{(i)} \cdot (\overline{s_1} - \overline{s_2}) \\ \alpha_{s_2}^{(i)} &= s_2^{(i)} \cdot (\overline{s_1} - \overline{s_2}) \quad \text{for } i = 1, \ldots, N \end{aligned} \tag{21}$$

where α_{s_1} and α_{s_2} denote the projected responses. Next, we calculated the means and variances of the projected responses α_{s_1} and α_{s_2}, denoted by $(\mu_{s_1}, \sigma_{s_1}^2)$ and $(\mu_{s_2}, \sigma_{s_2}^2)$. Finally, we calculate the discriminability d'_{s_1,s_2}, as the ratio between the separation of the means and the variances of the projected data:

$$d'_{s_1,s_2} = \frac{\mu_{s_1} - \mu_{s_2}}{\sqrt{\frac{1}{2}(\sigma_{s_1}^2 + \sigma_{s_2}^2)}} \tag{22}$$

Note that we used the same sequence of patches for all model configurations to calculate the discriminability, and every patch was presented for $125ms$. Previous research found that the variance of the response of a neuron to input stimuli is proportional to the mean (Gershon et al., 1998). Further studies demonstrated that inhibition leads to less variance in the responses to one repeatedly shown stimulus (Haider et al., 2010). The discriminability (d') increases if the response variance decreases by the same response mean. Therefore, we can measure differences in the response variance.

References

Adesnik, H. (2017). Synaptic mechanisms of feature coding in the visual cortex of awake mice. *Neuron*, *95*(5), 1147 - 1159.e4. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0896627317307055` doi: https://doi.org/10.1016/j.neuron.2017.08.014

Alitto, H. J., & Usrey, W. M. (2004). Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. *Journal of Neurophysiology*, *91*(6), 2797-2808. Retrieved from `https://doi.org/10.1152/jn.00943.2003` (PMID: 14762157) doi: 10.1152/jn.00943.2003

Anderson, J. S., Carandini, M., & Ferster, D. (2000). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of neurophysiology*, *84*, 909–926.

Atallah, B. V., Bruns, W., Carandini, M., & Scanziani, M. (2012). Parvalbumin-Expressing Interneurons Linearly Transform Cortical Responses to Visual Stimuli. *Neuron*, *73*(1), 159–170. Retrieved from `http://dx.doi.org/10.1016/j.neuron.2011.12.013` doi: 10.1016/j.neuron.2011.12.013

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(May). doi: 10.1038/nrn1888

Barak, O., Rigotti, M., & Fusi, S. (2013). The Sparseness of Mixed Selectivity Neurons Controls the Generalization-Discrimination Trade-Off. *Journal of Neuroscience*, *33*(9), 3844–3856. Retrieved from `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2753-12.2013` doi: 10.1523/JNEUROSCI.2753-12.2013

Beaulieu, C., Kisvarday, Z., Somogyi, P., Cynader, M., & Cowey, A. (1992). Quantitative distribution of gaba-immunopositive and-immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cerebral Cortex*, *2*(4), 295-309. Retrieved from `+http://dx.doi.org/10.1093/cercor/2.4.295` doi: 10.1093/cercor/2.4.295

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*(23), 3327 - 3338. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0042698997001211` doi: https://doi.org/10.1016/S0042-6989(97)00121-1

Berkes, P., White, B., & Fiser, J. (2009). No evidence for active sparsification in the visual cortex. *Neural Information Processing Systems Foundation*, *22*, 108–116.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, *2*(1), 32–48. doi: 10.1371/journal.ppat.0020109

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., ... Reid, R. C. (2011, Mar 09). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, *471*, 177 EP -. Retrieved from http://dx.doi.org/10.1038/nature09802 (Article)

Caporale, N., & Dan, Y. (2008). Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. *Annu. Rev. Neurosci*, *31*, 25–46. Retrieved from www.annualreviews.org doi: 10.1146/annurev.neuro.31.060407.125639

Carvalho, T. P., & Buonomano, D. V. (2009). Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions. *Neuron*, *61*(5), 774 - 785. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627309000804 doi: https://doi.org/10.1016/j.neuron.2009.01.013

Clopath, C., Büsing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based stdp with homeostasis. *Nature Neuroscience*, *13*(3), 344–352. Retrieved from http://www.nature.com/doifinder/10.1038/nn.2479 doi: 10.1038/nn.2479

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., ... Mrsic-flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, *000*(00), 1–5. doi: 10.1038/nature14182

Dadarlat, M. C., & Stryker, M. P. (2017). Locomotion enhances neural encoding of visual stimuli in mouse v1. *Journal of Neuroscience*, *37*(14), 3764–3775. Retrieved from http://www.jneurosci.org/content/37/14/3764 doi: 10.1523/JNEUROSCI.2728-16.2017

D'Amour, J. A., & Froemke, R. C. R. (2015). Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron*, *86*(2), 514–528. Retrieved from http://www.sciencedirect.com/science/article/pii/S089662731500210Xhttp://dx.doi.org/10.1016/j.neuron.2015.03.014 doi: 10.1016/j.neuron.2015.03.014

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience*. MIT Press.

Denève, S., & Machens, C. K. (2016). Efficient codes and balanced networks. *Nature neuroscience*, *19*(3), 375–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26906504 doi: 10.1038/nn.4243

Dorrn, A. L., Yuan, K., Barker, A. J., Schreiner, C. E., & Froemke, R. C. (2010). Developmental sensory experience balances cortical excitation and inhibition. *Nature*, *465*, 932–936. doi: 10.1038/nature09119

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, *23*(1), 441-471. Retrieved from https://doi.org/10.1146/annurev.neuro.23.1.441 (PMID: 10845071) doi: 10.1146/annurev.neuro.23.1.441

Finn, I. M., Priebe, N. J., & Ferster, D. (2007). The Emergence of Contrast-Invariant Orientation Tuning in Simple Cells of Cat Visual Cortex. *Neuron*, *54*(1), 137–152. doi: 10.1016/j.neuron.2007.02.029

Froemke, R. C. (2015). Plasticity of cortical excitatory-inhibitory balance. *Annual Review of Neuroscience*, *38*(1), 195-219. Retrieved from `https://doi.org/10.1146/annurev-neuro-071714-034002` (PMID: 25897875) doi: 10.1146/annurev-neuro-071714-034002

Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, *450*(7168), 425–429. Retrieved from `http://www.nature.com/doifinder/10.1038/nature06289` doi: 10.1038/nature06289

Froudarakis, E., Berens, P., Ecker, A. S., Cotton, R. J., Sinz, F. H., Yatsenko, D., ... Tolias, A. S. (2014). Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. *Nature Neuroscience*, *17*(6), 851–7. Retrieved from `http://bethgelab.org/publications/120/` doi: 10.1038/nn.3707

Gershon, E. D., Wiener, M. C., Latham, P. E., & RICHMOND, B. J. (1998). Coding strategies in monkey V1 and inferior temporal cortices. *Journal Of Neurophysiology*, *79*(3), 1135–1144. Retrieved from `papers3://publication/uuid/235E82AD-5CE2-4672-9E80-1F5B27476904` doi: 9497396

Gilbert, C. D., & Wiesel, T. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision research*, *30*(11), 1689–701. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/2288084` doi: 10.1016/0042-6989(90)90153-C

Goris, R. L., Simoncelli, E. P., & Movshon, J. A. (2015). Origin and Function of Tuning Diversity in Macaque Visual Cortex. *Neuron*, *88*(4), 819–831. Retrieved from `http://dx.doi.org/10.1016/j.neuron.2015.10.009` doi: 10.1016/j.neuron.2015.10.009

Griffen, T., & Maffei, A. (2014). Gabaergic synapses: their plasticity and role in sensory cortex. *Frontiers in Cellular Neuroscience*, *8*, 91. Retrieved from `https://www.frontiersin.org/article/10.3389/fncel.2014.00091` doi: 10.3389/fncel.2014.00091

Haider, B., Krause, M. R., Duque, A., Yu, Y., Touryan, J., Mazer, J. A., & McCormick, D. A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*, *65*(1), 107 - 121. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0896627309009829` doi: https://doi.org/10.1016/j.neuron.2009.12.005

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*, 51–58. doi: 10.1038/nature12654

Hennequin, G., Agnes, E. J., & Vogels, T. P. (2017). Inhibitory plasticity: Balance, control, and codependence. *Annual Review of Neuroscience*, *40*(1), 557-579. Retrieved from https://doi.org/10.1146/annurev-neuro-072116-031005 (PMID: 28598717) doi: 10.1146/annurev-neuro-072116-031005

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., ... Mrsic-Flogel, T. D. (2011, Jul 17). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, *14*, 1045 EP -. Retrieved from http://dx.doi.org/10.1038/nn.2876 (Article)

Hoyer, P. O. (2004). Non-negative Matrix Factorization with Sparseness Constraints. *The Journal of Machine Learning Research*, *5*, 1457–1469. doi: 10.1109/ICMLC.2011.6016966

Isaacson, J. S., & Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, *72*(2), 231–243. Retrieved from http://dx.doi.org/10.1016/j.neuron.2011.09.027 doi: 10.1016/j.neuron.2011.09.027

Jones, J. P., & Palmer, L. a. (1987a). An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex. *Journal of Neurophysiology*, *58*(6), 1233–1258. doi: citeulike-article-id:762473

Jones, J. P., & Palmer, L. A. (1987b). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1187-1211. Retrieved from https://doi.org/10.1152/jn.1987.58.6.1187 (PMID: 3437330) doi: 10.1152/jn.1987.58.6.1187

Katzner, S., Busse, L., & Carandini, M. (2011). Gabaa inhibition controls response gain in visual cortex. *Journal of Neuroscience*, *31*(16), 5931–5941. Retrieved from https://www.jneurosci.org/content/31/16/5931 doi: 10.1523/JNEUROSCI.5753-10.2011

Khan, A. G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T. D., & Hofer, S. B. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, *21*(6), 851–859. Retrieved from http://dx.doi.org/10.1038/s41593-018-0143-z doi: 10.1038/s41593-018-0143-z

King, P. D., Zylberberg, J., & DeWeese, M. R. (2013). Inhibitory Interneurons Decorrelate Excitatory Cells to Drive Sparse Code formation in a Spiking Model of V1. *The Journal of Neuroscience, 5475*, *33*(13), 5475–5485. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23536063 doi: 10.1523/JNEUROSCI.4188-12.2013

Ko, H., Hofer, S., Pichler, B., A Buchanan, K., Sjöström, P., & Mrsic-Flogel, T. (2011, 05). Functional specificity of local connections in neocortical networks. *Nature*, *473*, 87-91.

Kremkow, J., Perrinet, L. U., Monier, C., Alonso, J.-M., Aertsen, A., Frégnac, Y., & Masson, G. S. (2016). Push-Pull Receptive Field Organization and Synaptic Depression: Mechanisms for Reliably Encoding Naturalistic Stimuli in V1. *Frontiers in Neural Circuits*, *10*(May), 37. Retrieved from http://journal.frontiersin.org/Article/10.3389/fncir.2016.00037/abstract doi: 10.3389/fncir.2016.00037

Kullmann, D. M., Moreau, A. W., Bakiri, Y., & Nicholson, E. (2012). Plasticity of Inhibition. *Neuron*, *75*(6), 951–962. Retrieved from http://dx.doi.org/10.1016/j.neuron.2012.07.030 doi: 10.1016/j.neuron.2012.07.030

Larisch, R. (2019, November). [Re] Connectivity reflects coding a model of voltage-based STDP with homeostasis. *ReScience C*, *5*(3). Retrieved from https://doi.org/10.5281/zenodo.3538217 doi: 10.5281/zenodo.3538217

Litwin-Kumar, A., & Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications*, *5*(May), 1–12. Retrieved from http://dx.doi.org/10.1038/ncomms6319 doi: 10.1038/ncomms6319

Liu, B.-h., Li, Y.-t., Ma, W.-p., Pan, C.-j., Zhang, L. I., & Tao, H. W. (2011). Broad inhibition sharpens orientation selectivity by expanding input dynamic range in mouse simple cells. *Neuron*, *71*(3), 542–554. Retrieved from http://dx.doi.org/10.1016/j.neuron.2011.06.017 doi: 10.1016/j.neuron.2011.06.017

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature reviews. Neuroscience*, *5*(10), 793–807. Retrieved from http://dx.doi.org/10.1038/nrn1519 doi: 10.1038/nrn1519

Miconi, T., McKinstry, J. L., & Edelman, G. M. (2016). Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. *Nat Commun*, *7*, 13208. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27796298{%}5Cnhttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5095518/pdf/ncomms13208.pdf doi: 10.1038/ncomms13208

Mitchell, S. J., & Silver, R. (2003). Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron*, *38*(3), 433 - 445. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627303002009 doi: https://doi.org/10.1016/S0896-6273(03)00200-9

Mongillo, G., & Loewenstein, Y. (2018). Inhibitory connectivity defines the realm of excitatory plasticity. *Nature Neuroscience*, *21*(January), 1463 – 1470. doi: 10.1038/s41593-018-0226-x

Niell, C. M., & Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience*, *28*(30), 7520–7536. Retrieved from https://www.jneurosci.org/content/28/30/7520 doi: 10.1523/JNEUROSCI.0623-08.2008

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609. Retrieved from http://dx.doi.org/10.1038/381607a0 doi: 10.1038/381607a0

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*(23), 3311–3325. doi: 10.1016/S0042-6989(97)00169-7

Paille, V., Fino, E., Du, K., Morera-Herreras, T., Perez, S., Kotaleski, J. H., & Venance, L. (2013). Gabaergic circuits control spike-timing-dependent plasticity. *Journal of Neuroscience*, *33*(22), 9353–9363. Retrieved from https://www.jneurosci.org/content/33/22/9353 doi: 10.1523/JNEUROSCI.5796-12.2013

Pillow, J. W., & Simoncelli, E. P. (2006). Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of vision*, *6*(4), 414–428. doi: 10.1167/6.4.9

Potjans, T. C., & Diesmann, M. (2014). The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. *Cerebral Cortex*(March), 785–806. doi: 10.1093/cercor/bhs358

Priebe, N. J. (2016). Mechanisms of Orientation Selectivity in the Primary Visual Cortex. *Annual Review of Vision Science*, *2*(1), 85–107. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev-vision-111815-114456 doi: 10.1146/annurev-vision-111815-114456

Priebe, N. J., & Ferster, D. (2008). Inhibition, Spike Threshold, and Stimulus Selectivity in Primary Visual Cortex. *Neuron*, *57*(4), 482–497. doi: 10.1016/j.neuron.2008.02.005

Ringach, D. L. (2002). Spatial Structure and Symmetry of Simple-Cell Reveptive Fields in Macaque Primary Visual Cortex. *Journal of Neurophysiology*, *88*(1), 455–463. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12091567 doi: 10.1152/jn.00881.2001

Ringach, D. L., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166. doi: 10.1016/j.cogsci.2003.11.003

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque v1: Diversity and laminar dependence. *Journal of Neuroscience*, *22*(13), 5639–5651. Retrieved from https://www.jneurosci.org/content/22/13/5639 doi: 10.1523/JNEUROSCI.22-13-05639.2002

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the Neuronal Representation of Stimuli in the Primate Temporal Visual Cortex. *Journal of Neurophysiology*, *73*(2), 713–726. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7760130

Sadagopan, S., & Ferster, D. (2012). Feedforward Origins of Response Variability Underlying Contrast Invariant Orientation Tuning in Cat Visual Cortex. *Neuron*, *74*(5), 911–923. Retrieved from http://dx.doi.org/10.1016/j.neuron.2012.05.007 doi: 10.1016/j.neuron.2012.05.007

Sadeh, S., Cardanobile, S., & Rotter, S. (2014). Mean-field analysis of orientation selectivity in inhibition-dominated networks of spiking neurons. *SpringerPlus*, *3*(1), 148. Retrieved from http://www.springerplus.com/content/3/1/148 doi: 10.1186/2193-1801-3-148

Sadeh, S., Clopath, C., & Rotter, S. (2015). Emergence of Functional Specificity in Balanced Networks with Synaptic Plasticity. *PLoS Computational Biology*, *11*(6), 1–27. doi: 10.1371/journal.pcbi.1004307

Savin, C., Joshi, P., & Triesch, J. (2010). Independent Component Analysis in Spiking Neurons. *PLoS Comput Biol*, *6*(4), e1000757. doi: 10.1371/journal.pcbi.1000757

Schwartz, O., Pillow, J. W., Rust, N. C., & Simoncelli, E. P. (2006). Spike-triggered neural characterization. *Journal of Vision, 6,*, *6*(4), 484–507. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16889482 doi: 10.1167/6.4.13

Sengupta, B., Laughlin, S. B., & Niven, J. E. (2013). Balanced Excitatory and Inhibitory Synaptic Currents Promote Efficient Coding and Metabolic Efficiency. *PLoS Computational Biology*, *9*(10). doi: 10.1371/journal.pcbi.1003263

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *18*(10), 3870–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9570816 doi: 0270-6474/98/183870-27$05.00/0

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of Neurophysiology*, *57*(3), 773–786. doi: 10.1109/ACSAC.2009.36

Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *The Journal of Neuroscience*, *28*(48), 12591–603. Retrieved from http://www.jneurosci.org/content/28/48/12591.abstract doi: 10.1523/JNEUROSCI.2929-08.2008

Spanne, A., & Jörntell, H. (2015). Questioning the role of sparse coding in the brain. *Trends in Neurosciences*, *38*(7), 417–427. doi: 10.1016/j.tins.2015.05.005

Spratling, M. W. (2012). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Computation*, *24*(1), 60–103.

Sprekeler, H. (2017). Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. *Current Opinion in Neurobiology*, *43*, 198–203. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28500933 doi: 10.1016/j.conb.2017.03.014

Srinivasa, N., & Jiang, Q. (2013). Stable learning of functional maps in self-organizing spiking neural networks with continuous synaptic plasticity. *Frontiers in Computational Neuroscience*, *7*, 10. Retrieved from https://www.frontiersin.org/article/10.3389/fncom.2013.00010 doi: 10.3389/fncom.2013.00010

Stringer, C., Pachitariu, M., Steinmetz, N. A., Okun, M., Bartho, P., Harris, K. D., ... Lesica, N. A. (2016, dec). Inhibitory control of correlated intrinsic variability in cortical networks. *eLife*, *5*, e19695. Retrieved from https://doi.org/10.7554/eLife.19695 doi: 10.7554/eLife.19695

Tao, L., Shelley, M., McLaughlin, D., & Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences*, *101*(1), 366–371. Retrieved from http://www.pnas.org/cgi/doi/10.1073/pnas.2036460100 doi: 10.1073/pnas.2036460100

Tolhurst, D. J., Smyth, D., & Thompson, I. D. (2009). The sparseness of neuronal responses in ferret primary visual cortex. *Journal of Neuroscience*, *29*(8), 2355–2370. Retrieved from https://www.jneurosci.org/content/29/8/2355 doi: 10.1523/JNEUROSCI.3869-08.2009

Troyer, T. W., Krukowski, a. E., Priebe, N. J., & Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *Journal of Neuroscience*, *18*(15), 5908–5927.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276. Retrieved from http://science.sciencemag.org/content/287/5456/1273 doi: 10.1126/science.287.5456.1273

Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in v1. *Journal of Neuroscience*, *22*(7), 2904–2915. Retrieved from https://www.jneurosci.org/content/22/7/2904 doi: 10.1523/JNEUROSCI.22-07-02904.2002

Vitay, J., Dinkelbach, H., & Hamker, F. (2015). Annarchy: a code generation approach to neural simulations on parallel hardware. *Frontiers in Neuroinformatics*, *9*, 19. Retrieved from https://www.frontiersin.org/article/10.3389/fninf.2015.00019 doi: 10.3389/fninf.2015.00019

Vogels, T. P., Rajan, K., & Abbott, L. (2005). Neural Network Dynamics. *Annual Review of Neuroscience*, *28*(1), 357–376. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev.neuro.28.061604.135637 doi: 10.1146/annurev.neuro.28.061604.135637

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., & Gerstner, W. (2011). Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science*, *334*(6062), 1569–1573. Retrieved from http://science.sciencemag.org/content/334/6062/1569.abstract doi: 10.1126/science.1211095

Wang, L., & Maffei, A. (2014). Inhibitory Plasticity Dictates the Sign of Plasticity at Excitatory Synapses. *Journal of Neuroscience*, *34*(4), 1083–1093. Retrieved from http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4711-13.2014 doi: 10.1523/JNEUROSCI.4711-13.2014

Weliky, M., Fiser, J., Hunt, R. H., & Wagner, D. N. (2003). Coding of natural scenes in primary visual cortex. *Neuron*, *37*(4), 703–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12597866 doi: 10.1016/S0896-6273(03)00022-9

Wilson, N. R., Runyan, C. A., Wang, F. L., & Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, *488*(7411), 343–348. Retrieved from http://dx.doi.org/10.1038/nature11347 doi: 10.1038/nature11347

Wiltschut, J., & Hamker, F. H. (2009). Efficient coding correlates with spatial frequency tuning in a model of V1 receptive field organization. *Visual Neuroscience*, *26*(1), 21–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19203427 doi: 10.1017/S0952523808080966

Zenke, F., & Gerstner, W. (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1715), 20160259. Retrieved from http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2016.0259 doi: 10.1098/rstb.2016.0259

Zhu, W., Shelley, M., & Shapley, R. (2009). A neuronal network model of primary visual cortex explains spatial frequency selectivity. *Journal of Computational Neuroscience*, *26*(2), 271–287. doi: 10.1007/s10827-008-0110-x

Zhu, W., Xing, D., Shelley, M., & Shapley, R. (2010). Correlation between spatial frequency and orientation selectivity in V1 cortex: Implications of a network model. *Vision Research*, *50*(22), 2261–2273. Retrieved from http://dx.doi.org/10.1016/j.visres.2010.01.007 doi: 10.1016/j.visres.2010.01.007

Zhu, Y., Qiao, W., Liu, K., Zhong, H., & Yao, H. (2015). Control of response reliability by parvalbumin-expressing interneurons in visual cortex. *Nature Communications*, *6*, 1–11. Retrieved from http://dx.doi.org/10.1038/ncomms7802 doi: 10.1038/ncomms7802

Znamenskiy, P., Kim, M.-H., Muir, D. R., Iacaruso, M. F., Hofer, S. B., & Mrsic-Flogel, T. D. (2018). Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv*. Retrieved from `https://www.biorxiv.org/content/early/2018/04/04/294835.1` doi: 10.1101/294835

Zylberberg, J., & DeWeese, M. R. (2013). Sparse Coding Models Can Exhibit Decreasing Sparseness while Learning Sparse Codes for Natural Images. *PLoS Computational Biology*, *9*(8). doi: 10.1371/journal.pcbi.1003182

Zylberberg, J., Murphy, J. T., & DeWeese, M. R. (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*(10), e1002250. doi: 10.1371/journal.pcbi.1002250