Propagation of orientation selectivity in a spiking network model of layered primary visual cortex

Abbreviated title: Propagation of orientation selectivity

Benjamin Merkt[1,2], Friedrich Schüßler[1] and Stefan Rotter[1,2]

[1] Bernstein Center Freiburg, University of Freiburg, 79104 Freiburg, Germany
[2] Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany

Corresponding author:
Stefan Rotter
Bernstein Center Freiburg
Hansastraße 9a
79104 Freiburg
Germany
stefan.rotter@bio.uni-freiburg.de

- Number of pages: 56
- Number of figures: 12
- Number of tables: 2
- Number of words: Abstract: 246
  Introduction: 607
  Discussion: 1480

Conflict of interests: The authors declare no competing financial interests.

Acknowledgments: This work was partially supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 600925 (NeuroSeeker), by the BMBF (grant BFNT 01GQ0830) and DFG (grant EXC 1086). The HPC facilities are funded by the state of Baden-Württemberg through bwHPC and DFG grant INST 39/963-1 FUGG. In addition, our work was promoted by the German Academic Exchange Service (DAAD) and the Carl Zeiss Stiftung. We thank Uwe Grauer from the Bernstein Center Freiburg as well as Bernd Wiebelt and Michael Janczyk from the Freiburg University Computing Center for their assistance with HPC applications.

Code Accessibility: Simulation and analysis code is available upon request.

Abstract

We studied the propagation of orientation selectivity over layers in a model of rodent primary visual cortex in terms of the underlying connectivity. While significant progress has been made in measuring the activity in the different sub-populations as well as the connectivity between those in experiments, a comprehensive theoretical explanation of how network structure and its dynamics are related is still missing. We suggest a model of layered mouse visual cortex by extending the model suggested by Potjans and Diesmann (2014) with thalamic input that has an orientation bias according to Sadeh et al. (2014). We then studied the response properties of the network to such non-homogeneous input. We found that, without further assumptions, the connectivity derived from experimental data leads to layer-specific distributions of orientation selectivity very similar to what has been observed in mouse experiments. Interestingly, the network settles in a dynamical operating point, in which the efficacy of the different projections and their contribution to orientation tuning deviate strongly from the expectations based on the underlying anatomical connectivity. To understand the processes that shape the observed dynamics, it is essential to perform the analysis of the microcircuit on the system level. With that perspective, we find in particular that the difference in tuning of L2/3 and L4 neurons is an immediate, albeit unexpected consequence of the specific network connectivity. Furthermore, we introduce a novel method for predicting the effects of optogenetic stimulation of specific neuronal sub-populations and demonstrate its power in network simulations.

Significance Statement

Understanding the precise roles of neuronal sub-populations in shaping the activity of neuronal networks is a fundamental objective of neuroscience research. To this end, our work makes three important contributions. First, we show that the experimentally extracted connectivity suffices to explain the degree of selectivity of sub-populations in mouse visual cortex to visual stimulation. Second, we introduce a novel system-level approach for the analysis of input-output relations of recurrent networks, which lead to distinct activity patterns. Third, we present a method for the design of optogenetic experiments that can be used to devise specific stimuli which result in a desired and predictable change of neuronal activity.

1 Introduction

Understanding the complex computations performed by neural networks in the nervous system is a central challenge in neuroscience. In recent years, research in rodent sensory systems has provided access to many new facts, but has also raised new theoretical questions. In particular, the role of cortical layers (Niell and Stryker, 2008; Potjans and Diesmann, 2014) and other idiosyncratic subpopulations (Durand et al., 2016; Pakan et al., 2016) has received increasing attention. While the anatomical path of sensory input signals from the thalamus through the cortical layers has been particularly well characterized for the visual system, this provides no explanation yet how the information is processed and transformed in transit. Here, we apply computational neuroscience techniques to gain new insight into the various computation steps within this highly recurrent neural network.

From a theoretical perspective, the cortical network constitutes a dynamical system, which can operate on different types of sensory input. To study basic properties of such networks, the input can be assumed to be identical for all neurons (Amit and Brunel, 1997; Brunel, 2000). In a sensory system, however, the input varies across neurons, as sensors extract different aspects of the stimulus, and the primary visual cortex (V1) represents an interesting example to study such systems (Hansel and van Vreeswijk, 2012; Sadeh et al., 2014). In this case, non-homogeneous input carries information about the orientation of moving gratings or light bars presented to the eye of an animal (Hubel and Wiesel, 1959).

A major difference of rodent visual cortex compared to carnivores and primates is the absence of orientation columns and maps (Ohki et al., 2005). Although the "salt-and-pepper" organization implies that lateral interactions are quite unspecific (Jia et al., 2010; Ko et al., 2011), neurons nevertheless show strong output tuning with respect to oriented light bars and gratings (Niell and Stryker, 2008; Hofer et al., 2011; Durand et al., 2016). Even more interesting, this output tuning is already present at eye opening in juvenile mice, where connectivity appears to be functionally random (Ko et al., 2013). In recent theoretical studies, this effect could be explained by a cancellation of the untuned mode of the distributed input due to the dominance of inhibition in the recurrent network (Hansel and van Vreeswijk, 2012; Sadeh et al., 2014; Sadeh and Rotter, 2014, 2015). While these previous works identified the mechanism underlying strong output tuning in the primary visual cortex of rodents, they do not explain the different degrees of orientation selectivity in different sub-populations, like cortical layers.

In the present work, we apply computational network modeling techniques to study the emergence of orientation selectivity in a layered cortical network model, with a focus on the differences between layers and populations. As a starting point, we adopt the anatomically and physiologically founded network model developed by Potjans and Diesmann (2014) and extend it with orientation selective input similar to Sadeh et al. (2014). We model an experiment in which an animal is shown a sinusoidal grating as a visual stimulus. Without any adaptation or fine-tuning of parameters, the resulting model matches the orientation selectivity distributions found in cortical networks, including the weak tuning of inhibitory neurons, as well as L5 pyramidal neurons. We then show that these specific properties cannot be attributed to any particular parameter of the circuit, but rather is an emergent network feature of the entire microcircuit. Furthermore, we suggest a new concept for the design of optogenetic stimulation experiments. Applying this concept to the V1 cortical microcircuit, we are in the position to make specific predictions for experiments to manipulate the orientation selectivity of L5 principal cells, which eventually would confirm our analysis of the circuit.

2 Material and Methods

2.1 Modeling layered primary visual cortex

The network model considered throughout our study is, in essence, the recurrent network model developed by Potjans and Diesmann (2014) extended by thalamic input with a weak orientation bias, as introduced by Sadeh et al. (2014). Potjans and Diesmann (2014) set out to create a full-scale neural network model, matching as close as possible to the nervous tissue underneath $1\,\mathrm{mm}^2$ of neocortex. They incorporated a large set of experimental studies, most prominently the ones by Thomson et al. (2002) and Binzegger

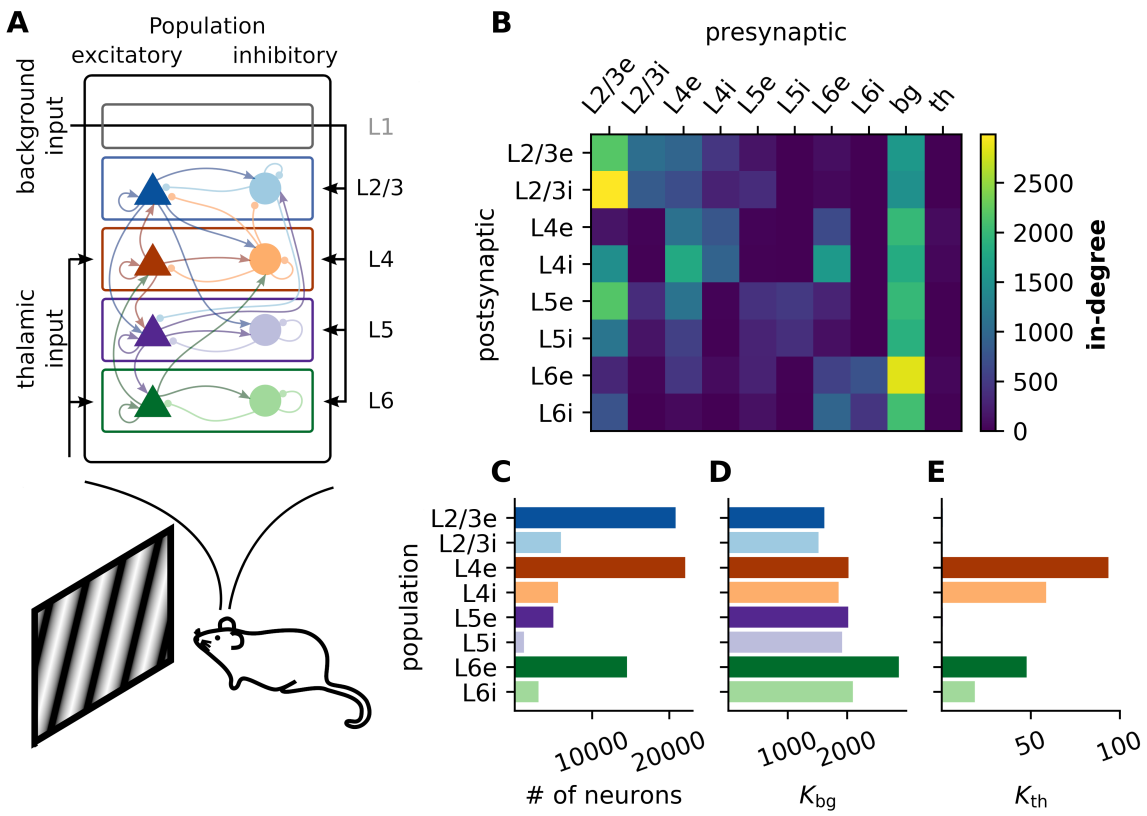

Figure 1: Model parameters. **A** Schematic of the model with eight populations in four layers, thalamic and other background input. Triangles and circles denote excitatory and inhibitory populations, respectively. **B** Total number of synapses each neuron receives (in-degree) from any other population, including input. **C** Total number of neurons in each of the eight populations. **D-E** Input in-degrees for each neuron in each of the eight populations. Numbers are identical to the two right-most columns of A.

et al. (2004), to derive a generic connectivity map between eight populations of neurons situated in four layers (L2/3, L4, L5, L6), one excitatory (e) and one inhibitory (i) population in each layer (see Figure 1A for an illustration of the model). The model network consists of almost 80 000 leaky integrate-and-fire neurons, where the sizes of individual subpopulations vary drastically across layers. While the two largest populations contain approximately 20 000 neurons, L5e comprizes only about 5 000 neurons (Figure 1C). Throughout all layers, inhibitory populations are considerably smaller than the respective excitatory populations, by a factor of four on average. In our work, the leaky integrate-and-fire neurons are equipped with delta synapses, i.e. with each incoming spike, the membrane potential is instantaneously deflected by a fixed voltage J, decaying with the membrane time constant τ. The synaptic efficacy J varies depending on the presynaptic and postsynaptic populations, respectively.

Our model assumes a fixed number of synapses for each projection, i.e. each neuron of the postsynaptic population P receives the same number of synaptic inputs K_{PT} from the presynaptic population T. For any target neuron, the source neurons in the presynaptic population are drawn randomly. The numbers K_{PT}, which are the essential parameters of

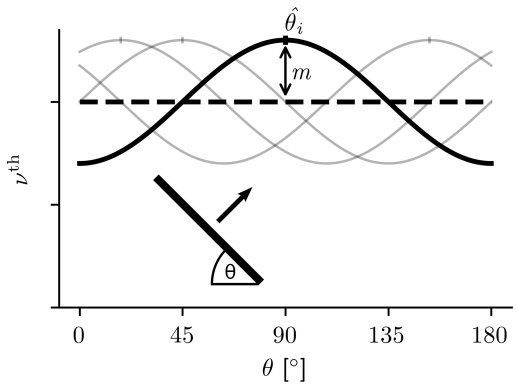

Figure 2: Thalamic input modulation. Illustration of the input rate variation over stimulation angles for single neurons in L4 and L6. The angle corresponding to the peak of the curve is the preferred orientation of the input to that neuron.

our model, were derived by Potjans and Diesmann (2014) and are illustrated in Figure 1B.

In addition to recurrent synaptic connections, neurons of the model network receive external input from two different sources (Figure 1A). The first type of input, termed "background input" throughout this paper, is targeting all populations. It represents axons from other cortical and subcortical regions, including gray matter and white matter projections (Figure 1B/D), but excludes those originating from the thalamus. These background inputs are independent of the stimulus. In all simulations, they are modeled as Poisson processes with rate ν_{bg}.

The second type of input is represented by thalamocortical projections that originate from the lateral geniculate nucleus (LGN) (Figure 1B/E). In primary visual cortex, these projections are the main source of input that conveys information about a stimulus. Consequently, this is the component where we made the central addition to the original model of Potjans and Diesmann (2014). We replaced the unspecific thalamic input with an input that depends in a specific way on the stimulus. Throughout this paper, the stimuli considered were oriented moving gratings, which are commonly used in visual neuroscience. Similar to Sadeh et al. (2014), each neuron in one of the input populations (i.e. L4 and L6) is randomly assigned a preferred orientation $\hat{\theta}_i$. The input to each neuron then depends on the orientation θ of the stimulus and varies according to

$$\nu_i^{\mathrm{th}}(\theta) = K_P^{\mathrm{th}}\,\nu_0^{\mathrm{th}}\big[1 + m\,\cos(2(\theta - \hat{\theta}_i))\big]. \tag{1}$$

An illustration of the input variation is provided in Figure 2. Note that orientations vary between 0° and 180° as we do not consider direction selectivity in this study. Similar to the background input, the thalamic input was conceived as a homogeneous Poisson process.

It is instructive to analyze and compare to different modes of operation. The *stimulation* condition is the situation described above, emphasizing the presentation of a structured visual stimulus (here, a moving oriented grating) to the animal. In contrast, the *spontaneous* condition emulates the presentation of a homogeneous gray screen to the animal. In this case, the intensity of thalamic inputs equals the rate of the background inputs. Moreover, the angular modulation m is set to zero, removing any information

about stimulus orientation from the input.

2.2 Implementations of the model

The results of this study are based on three different mathematical implementations of the same network model. While all models are based on the same connectivity parameters and the same single neuron model, they use different numerical techniques to solve for the neuron activity. In the following, the three models will be briefly described.

Model A: spiking neuronal network model. The starting point for our analysis is constituted by a *spiking neuronal network*, very similar to the implementation provided by Potjans and Diesmann (2014). Here, for each of the almost 80 000 neurons, the temporal evolution of the membrane potential $V_i(t)$ for each neuron i is simulated using numerical methods. The interaction in the network is provided by discrete spike events emitted by pre-synaptic neurons that alter the membrane potential of the post-synaptic cells. The output of the network is then given in terms of the spike trains of all neurons in the recurrent system. Similar to the original implementation of Potjans and Diesmann (2014), we use the simulation tool NEST (Gewaltig and Diesmann, 2007) to perform all the necessary numerical simulations.

Model B: nonlinear firing rate model. Based on the spiking neural network model, we also developed a neuron-by-neuron *firing rate model*, applying the theory developed by Brunel (2000) to single neurons (Ostojic, 2014; Sadeh and Rotter, 2015). The network is characterized by the input-output transfer functions F_i for all neurons, leading to the self-consistency equation

$$\nu_i = F_i(\nu, \nu^{\mathrm{bg}}, \nu^{\mathrm{th}}), \tag{2}$$

where $i = 1, \ldots, N$ is the neuron index and ν, ν^{bg} and ν^{th} are the vectors of recurrent neuron rates and input rates, respectively. Throughout this work, variables denoted by th and bg refer to quantities originating from the thalamus or other external inputs, respectively. Consequently, the model results in a high-dimensional system of algebraic equations, which need to be solved numerically, yielding the firing rate ν_i of each individual neuron.

Model C: linearized network model. Finally, based on Model B, we also derive a *linearized network model* of the microcircuit. Fixing a point of linearization, for any input perturbation $\Delta\nu^{\mathrm{th}}$, it predicts the perturbation of the single neuron firing rates $\Delta\nu$. In the linear model, these can be explicitly calculated by

$$\Delta\nu = (\mathbb{1} - W)^{-1}\, B\, \Delta\nu^{\mathrm{th}}.$$

The matrices W and B are given by appropriate partial derivatives of the transfer function (Equation (2)). As the linearization point, we choose the activity into which the network settles when thalamic neurons are completely silenced. The perturbation $\Delta\nu^{\mathrm{th}}$ has different values for the two simulated conditions which, for the chosen linearization point, are identical to the absolute thalamic rate ν^{th}. In the spontaneous condition, it is given by a homogeneous vector with the thalamic baseline rate in all entries. In the stimulated condition, the entries are given by $\Delta\nu_i^{\mathrm{th}} = \nu_i^{\mathrm{th}}$ as defined in Equation (1). Note that this vector is non-homogeneous and contains the specific, orientation dependent perturbation for each neuron.

While the spiking network model (Model A) is the most realistic version from a biological perspective, accounting for irregular spike trains and activity fluctuations, it is at the same time computationally expensive and analytically intractable. The firing rate model (Model B) makes additional assumptions about the network activity, like negligible correlations, but it has the advantage of being numerically tractable. Finally, while the linear network model (Model C) ignores non-linear effects by construction, it has the great benefit of allowing the application of a powerful linear algebra toolbox and enabling an explicit solution of the system (Sadeh and Rotter, 2015).

Importantly, applying several different implementations of the same model is only justified if the various model predictions are consistent. Therefore, whenever possible throughout this work, we cross-check the results of the different models.

2.3 Data analysis

Measuring orientation selectivity In order to quantify orientation selectivity, we adopt a measure from circular statistics (Batschelet, 1981; Piscopo et al., 2013). If $\nu_k(\theta)$ is the firing rate of neuron k for stimulation angle θ, the orientation selectivity vector (OSV) is calculated by

$$OSV_k = \frac{\sum_\theta \nu_k(\theta) e^{i2\theta}}{\sum_\theta \nu_k(\theta)},$$

which yields a complex number. Here, the two sums are running over all stimulation angles θ, or a uniform sample of angles in a simulation. The orientation selectivity index (OSI) is then defined as the absolute value of the OSV, i.e.

$$OSI_k = |OSV_k|.$$

The preferred orientation (PO) of single neurons can be calculated from the angle of the OSV,

$$PO_k = \arg(OSV_k).$$

Note that the definition of OSI is identical to

$$OSI = \frac{F1}{F0}$$

where $F0$ and $F1$ are the zeroth and first Fourier components of the tuning curve, giving a somewhat more intuitive interpretation of the OSI. Here, $F0$ accounts for the mean firing rate over all stimulation angles, and $F1$ is equivalent to a cosine function accounting for the lowest-order modulation over all angles. Applying the definition to a perfect cosine-tuned curve as assumed for the tuned input provided by thalamic neurons yields $OSI = \frac{m}{2}$, where m is the modulation strength of the tuning curve (Equation (1)). The OSI as defined here provides a robust measure of tuning strength, in particular also for neurons with low rates and noisy responses.

Input currents Since our model uses a simple integrate-and-fire neuron model with delta-synapses, the mean current received by neuron i through synaptic input from neuron j is given by

$$I_{ij} = C_\mathrm{m} \, J_{ij} \, \bar{\nu}_j, \tag{3}$$

population	L2/3e	L2/3i	L4e	L4i	L5e	L5i	L6e	L6i
size	20683	5834	21915	5479	4850	1065	14395	2948

in-degrees		presynaptic									
		L2/3e	L2/3i	L4e	L4i	L5e	L5i	L6e	L6i	bg	th
postsynaptic	L2/3e	2199	1079	979	467	159	0	109	0	1600	0
	L2/3i	2990	860	703	289	380	0	60	0	1500	0
	L4e	159	34	1117	794	32	0	667	0	2100	93
	L4i	1480	16	1813	953	16	0	1608	0	1900	57
	L5e	2188	374	1135	31	420	496	296	0	2000	0
	L5i	1165	159	570	12	300	404	124	0	1900	0
	L6e	325	38	467	91	285	21	581	752	2900	47
	L6i	766	5	74	2	136	8	979	459	2100	17

Table 1: Network layout parameters.

where $\bar{\nu}_j$ is the mean rate of neuron j. In order to quantify the information about the stimulus contained in the input current received from any neuron in the network, we also define a tuning vector (TV) of each synaptic connection. For a given pre-synaptic neuron j and post-synaptic neuron i, it is defined by

$$\mathrm{TV}_{ij} = C_\mathrm{m}\, J_{ij}\, \bar{\nu}_j\, OSV_j. \tag{4}$$

where $\bar{\nu}_j$ is the firing rate of neuron j averaged over all angles. Note that this is identical to the first Fourier component (F1) of the input current tuning curve. The single neuron tuning vectors TV_{ij} can then be used to quantify the combined tuning information each neuron in the network receives from a given source population P by calculating

$$TV_{iP} = \sum_{j \in P} K_{ij}\, \mathrm{TV}_{ij},$$

where $K_{ij} = 1$ if neuron j makes a synapse onto neuron i, and $K_{ij} = 0$ otherwise.

2.4 Detailed model descriptions

In the following, the mathematical model implementations are described in detail. While this is certainly helpful for the mathematically interested reader, it is not essential for the comprehension of the results of the study. In the main text of the manuscript we do not refer to these details, and the reader may choose to directly jump to Section 3, Results.

Model A: Spiking neural network model

We start by describing the spiking neural network model in detail. To a large extent, it is identical to the original model introduced by Potjans and Diesmann (2014). It consists of $N = 77\,169$ neurons divided into eight populations, one excitatory and one inhibitory population in each out of four layers, respectively (Figure 1A). The individual population sizes are summarized in Figure 1C and Table 1. As a neuron model, the leaky integrate-and-fire model is used. In this point neuron approximation, the membrane potential of

parameter	value	unit
τ_{m}	10	ms
τ_{ref}	2	ms
R	40	MΩ
C_{m}	250	pF
V_{thr}	-50	mV
V_{r}	-65	mV
J_{e}	0.15 ± 0.015	mV
g	4	
d_{exc}	1.5 ± 0.75	ms
d_{inh}	0.7 ± 0.35	ms
ν_{bg}	8	Hz
ν_0^{th}	30	Hz
m	0.3	
θ	$0, 15, ..., 165$	°
$\hat{\theta}_i$	$[0, 180)$	°
T	100	s
Δt	0.1	ms

Table 2: Single neuron parameter values. Shown are variable names, numerical parameter values and, if applicable, the standard deviation of the Gaussian distribution, as well as physical units.

each individual neuron follows the differential equation

$$\tau_{\mathrm{m}} \dot{V}_i(t) = -\big[V_i(t) - V_r\big] + RI_i(t). \tag{5}$$

Here, $V_i(t)$ is the membrane potential of neuron i at time t, τ_{m} is the membrane time constant and V_{r} is the resting potential. The total input current can be divided into the three sources $I_i(t) = I_i^{\mathrm{rec}}(t) + I_i^{\mathrm{bg}}(t) + I_i^{\mathrm{th}}(t)$. The recurrent input is given by

$$RI_i^{\mathrm{rec}}(t) = \sum_{j=1}^{N} \sum_{k} J_{ij}\, \delta(t_{jk} + d_{ij} - t),$$

where t_{jk} is the time of the k-th spike of neuron j, d_{ij} is the delay of the connection, and J_{ij} is the synaptic efficacy from neuron j to neuron i. In other words, each time the pre-synaptic neuron j fires a spike at time t_{jk}, the membrane potential of neuron i is deflected by J_{ij} at time $t_{ik} + d_{ij}$. Each time the membrane potential reaches the threshold voltage V_{thr}, the neuron emits a spike to all its post-synaptic neurons. Following each spike, the membrane potential is clamped at the reset potential $V_i = V_r$ for a refractory period τ_{ref}. In our work, the synapse model slightly differs from the original model of Potjans and Diesmann (2014). While they used exponential post-synaptic currents, in our model all of the charge is transferred instantaneously, as expressed by the δ-function in Equation (5). Numerical values for the single neuron parameters are summarized in Table 2.

The connectivity of the model was compiled in Potjans and Diesmann (2014) based on a large number of electrophysiological and anatomical measurements, most prominently

by Thomson et al. (2002) and Binzegger et al. (2004). The central result of their detailed analysis is the connectivity map shown in Figure 1B and Table 1. In our implementation, these numbers describe the exact number of synapses each post-synaptic neuron receives from randomly chosen pre-synaptic neurons in each of the eight populations. All connections are unitary, and self-connections are excluded. This fixed in-degree connectivity deviates from the original model, where a fixed number of synapses was distributed by randomly drawing both target and source neuron, resulting in approximately binomially distributed in- and out-degrees. Our modification results in a somewhat lower variability of rates and orientation selectivities in the populations.

Synaptic efficacies J_{ij} are drawn from a normal distribution with a mean which depends on the type of connection, and a standard deviation of 10% of the mean. The mean is J_e for excitatory and $J_i = -g\, J_\mathrm{e}$ for inhibitory connections. One exception is the projection from L4e to L2/3e neurons, which has a mean efficacy of $2\, J_\mathrm{e}$. For all unconnected pairs, we have $J_{ij} = 0$. In order to obey Dale's principle, the normal distributions are clipped at zero. Similarly, delays d_{ij} are also drawn from two normal distributions with fixed means and variances, d_exc for excitatory and d_inh for inhibitory connections, respectively. The delay distributions are clipped at the simulation time step Δt. Numerical values for the distributions of connection strengths and delays can be found in Table 2.

In addition to recurrent connections, neurons receive two types of external input. The constant background input represents connections from other brain regions as well as from neurons within V1 which are not part of the model network. It is realized as a homogeneous Poisson input for each neuron. The input current to the single neurons is

$$RI_i^\mathrm{bg}(t) = \sum_k J_\mathrm{bg}\, \delta(t - t_{ik}).$$

In this case $J_\mathrm{bg} = J_\mathrm{e}$ is the connection strength of background synapses and t_{ik} is the time of k-th spike of the Poisson process of neuron i. The rate of each Poisson process depends on the population P each neuron belongs to and is given by $\nu_P^\mathrm{bg} = K_P^\mathrm{bg} \nu_\mathrm{bg}$. Here, K_P^bg is the number of background synapses connecting to individual neurons in the respective target population (Figure 1B/D and Table 1) and ν_bg is the rate of individual background neurons.

Similar to background input, the thalamocortical input is also conceived as a homogeneous Poisson process. The current from the connections is given by

$$RI_i^\mathrm{th}(t) = \sum_k J_\mathrm{th}\, \delta(t - t_{ik})$$

with $J_\mathrm{th} = J_\mathrm{e}$. In contrast to the homogeneous background input to each individual neuron, the rate of the thalamic input depends on the orientation θ of a grating that represents the visual stimulus (Figure 2). The resulting orientation-tuned input is given by

$$\nu_i^\mathrm{th} = K_P^\mathrm{th}\, \nu_0^\mathrm{th} \big[1 + m\, \cos(2(\theta - \hat{\theta}_i))\big], \tag{6}$$

where K_P^th denotes the number of thalamocortical synapses per neuron in population P (Figure 1B/E), ν_0^th is the fixed mean rate of thalamic neurons, m is the modulation strength and $\hat{\theta}_i$ the preferred orientation of the input to this neuron. The latter is drawn from a uniform distribution on $[0°, 180°)$. The network is simulated for twelve stimulation angles θ from 0° to 165°. Again, numerical values for all parameters can be found in

Table 2 and Table 1.

The non-homogeneous thalamic input is the core modification with respect to the Potjans and Diesmann (2014) model to account for the flow of signals in the visual system. The original model used a population of 902 thalamic neurons. However, in order to include orientation tuning into the projection, it would be necessary to make assumptions on how the input tuning is generated by the thalamocortical afferents. Because our work concentrates on the cortical processing of tuning information, we replaced the thalamic neurons by the simpler Poisson type input tuning.

Each of the twelve stimulation angles θ is simulated for a total of $T = 100\,\mathrm{s}$, with a simulation time step of $\Delta t = 0.1\,\mathrm{ms}$. In addition, before the actual simulation time, the network is simulated for a warm-up time of 200 ms to exclude input transients from analysis. Spike times and the exact connectivity are extracted for all neurons for further analysis.

Model B: nonlinear firing rate model

In order to obtain a firing rate model for the system under study, we apply the approach of Brunel (2000) to each single neuron in all eight populations. We assume that the total input to each neuron i amounts to a mean current μ_i and additive current fluctuations (Gaussian white noise) $\eta(t)$ with variance σ_i^2

$$RI_i(t) = \mu_i + \sigma_i \sqrt{\tau_\mathrm{m}} \eta(t).$$

This way, we can characterize the stationary state, in which μ_i and σ_i are constant parameters. Similar to the total input current, the mean and variance can be separated into three sources

$$\begin{aligned} \mu_i &= \mu_i^\mathrm{rec} + \mu_i^\mathrm{bg} + \mu_i^\mathrm{th} \\ \sigma_i^2 &= \sigma_i^{\mathrm{rec}2} + \sigma_i^{\mathrm{bg}2} + \sigma_i^{\mathrm{th}2}. \end{aligned}$$

For each source, the respective mean and variance are approximately given by

$$\begin{aligned} \mu_i^\mathrm{rec} &= \tau_\mathrm{m} \sum_{j=1}^{N} J_{ij}\, \nu_j, & \sigma_i^{\mathrm{rec}2} &= \tau_\mathrm{m} \sum_{j=1}^{N} J_{ij}^2\, \nu_j \\ \mu_i^\mathrm{bg} &= \tau_\mathrm{m}\, J_\mathrm{bg}\, \nu_i^\mathrm{bg}, & \sigma_i^{\mathrm{bg}2} &= \tau_\mathrm{m}\, J_\mathrm{bg}^2\, \nu_i^\mathrm{bg} \\ \mu_i^\mathrm{th} &= \tau_\mathrm{m}\, J_\mathrm{th}\, \nu_i^\mathrm{th}, & \sigma_i^{\mathrm{th}2} &= \tau_\mathrm{m}\, J_\mathrm{th}^2\, \nu_i^\mathrm{th}, \end{aligned}$$

where independence of all inputs has been assumed. As for the spiking model, the in-degrees of the external input has been accounted for in ν_i^bg and ν_i^th. Furthermore, ν_i^th accounts for the non-homogeneous input each individual neuron receives due to the oriented stimulus according to Equation (6). The self-consistent solutions of the associated first-passage time problem (Siegert, 1951) yield the steady-state firing rates of all neurons

$$\nu_i = F_i(\nu, \nu_i^\mathrm{th}), \tag{7}$$

with the transfer function F_i being defined by

$$F_i(\nu, \nu_i^\mathrm{th}) = \frac{1}{\Phi_i(\nu, \nu_i^\mathrm{th})}$$

where

$$\Phi_i(\nu, \nu_i^{\mathrm{th}}) = \tau_{\mathrm{ref}} + \tau_{\mathrm{m}}\sqrt{\pi} \int_{\frac{V_r-\mu_i}{\sigma_i}}^{\frac{V_{\mathrm{thr}}-\mu_i}{\sigma_i}} e^{x^2}(1 + erf(x))\, dx,$$

where erf is the error function. Solving this non-linear, algebraic set of equations yields the desired estimates for the firing rate of each single neuron.

The challenge is now to solve this 77 169-dimensional system of highly non-linear self-consistency equations. This can be achieved by re-formulating the problem as a Wilson-Cowan type model (Wilson and Cowan, 1972). Since we are only interested in the fixed point describing the steady-state of the system, and not in its temporal evolution, we choose a unit time constant and introduce a pseudo-time s. This allows to write the problem as a system of ordinary differential equations

$$\frac{d\nu(s)}{ds} = -\nu(s) + F_i\big(\nu(s), \nu^{\mathrm{th}}\big).$$

Assuming stability, these differential equations converge to a steady-state rate vector ν^* which solves the system in Equation (7) characterized by $\frac{d\nu(s)}{ds} = 0$. For the numerical solution, we applied an explicit runge-kutta method of order (4)5 including adaptive step size control (Hairer et al., 1993).

Model C: linearized network model

The linear model is based on the rate coding model described above. We start by calculating the total derivative of the transfer function Equation (7)

$$\Delta\nu_i = \sum_{j=1}^{N} \frac{\partial F_i}{\partial \nu_j}\Delta\nu_j + \frac{\partial F_i}{\partial \nu_i^{\mathrm{th}}}\Delta\nu_i^{\mathrm{th}}\Bigg|_{\nu=\nu_{\mathrm{OP}}}. \tag{8}$$

The partial derivatives of the transfer function are given by

$$\begin{aligned}
\frac{\partial F_i}{\partial \nu_j} &= -\frac{1}{\Phi_i^2}\frac{\partial \Phi_i}{\partial \nu_j} \\
\frac{\partial \Phi_i}{\partial \nu_j} &= \frac{\partial \Phi_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \nu_j} + \frac{\partial \Phi_i}{\partial \sigma_i}\frac{\partial \sigma_i}{\partial \nu_j}
\end{aligned}$$

where the partial derivatives of Φ_i are

$$\begin{aligned}
\frac{\partial \Phi_i}{\partial \mu_j} &= erfcx\left(-\frac{V_{\mathrm{thr}}-\mu_i}{\sigma_i}\right)\frac{-1}{\sigma_i} - erfcx\left(-\frac{V_{\mathrm{r}}-\mu_i}{\sigma_i}\right)\frac{-1}{\sigma_i} \\
\frac{\partial \Phi_i}{\partial \sigma_j} &= erfcx\left(-\frac{V_{\mathrm{thr}}-\mu_i}{\sigma_i}\right)\frac{V_{\mathrm{thr}}-\mu_i}{\sigma_i^2} - erfcx\left(-\frac{V_{\mathrm{r}}-\mu_i}{\sigma_i}\right)\frac{V_{\mathrm{r}}-\mu_i}{\sigma_i^2}.
\end{aligned}$$

Here, the Leibniz integral rule was used, and $erfcx(x) = e^{x^2}(1 - erf(x))$ is the scaled complementary error function. Finally, the partial derivatives of the mean and standard

deviation of the input are given by

$$\frac{\partial \mu_i}{\partial \nu_j} = \tau_{\mathrm{m}} J_{ij}$$
$$\frac{\partial \sigma_i}{\partial \nu_j} = \frac{1}{2\sigma_i} \tau_{\mathrm{m}} J_{ij}^2.$$

Note that in all partial derivatives above, an explicit mention of the evaluation point $\nu = \nu_{\mathrm{OP}}$ was omitted for the sake of a compact notation.

The operating point ν_{OP} at which the derivatives are evaluated can be chosen freely. However, the precision of the linear approximation depends on how non-linear the activity of the network actually is and, how far the linearization point is from the region of interest. Here, it is chosen as the activity with zero thalamic input, i.e. $\nu^{\mathrm{th}} \equiv 0$. This choice provides the greatest flexibility in the predictions which are derived from the model. In this case, for both conditions considered here, the thalamic input perturbation is identical to the thalamic input rates, i.e. $\Delta\nu_i^{\mathrm{th}} = \nu_i^{\mathrm{th}}$. Note that while thalamic neurons do not fire at the linearization point, the network does still receive background input from other sources, leading to non-zero network activity at this point.

The derivatives of the transfer function are recast in a matrix W

$$W_{ij} = \left.\frac{\partial F_i}{\partial \nu_j}\right|_{\nu=\nu_{\mathrm{OP}}}$$

and the diagonal matrix

$$B_{ij} = \left.\frac{\partial F_i}{\partial \nu_j^{\mathrm{th}}}\right|_{\nu=\nu_{\mathrm{OP}}}.$$

Equation (8) can then be compactly written as

$$0 = (W - \mathbb{1})\,\Delta\nu + B\,\Delta\nu^{\mathrm{th}},$$

where we also summarized the input and output rate perturbations into the vectors $\Delta\nu$ and $\Delta\nu^{\mathrm{th}}$. Assuming that $(W - \mathbb{1})$ is invertible, which is almost always the case for the large random matrices considered here, and defining the scaled effective input $\Delta\beta = B\,\Delta\nu^{\mathrm{th}}$, we obtain the explicit expression for the output perturbation

$$\Delta\nu = (\mathbb{1} - W)^{-1}\,\Delta\beta. \tag{9}$$

In a next step, the effective input perturbation can be decomposed into

$$\Delta\beta = \Delta\beta_B + \Delta\beta_M,$$

where $\Delta\beta_B$ is the baseline and $\Delta\beta_M$ is the modulation component, respectively (Sadeh et al., 2014). The former is defined by $\Delta\beta_{B_i} = \mathbb{E}[\Delta\beta_i]$, the vector of expected effective input rates. Note that the expectation values are identical across all neurons of each population. With this, we define $\Delta\beta_M = \Delta\beta - \Delta\beta_B$ as the vector containing the deviations from the population means.

Along the same lines, the matrix W can also be decomposed into two matrices such

that

$$W = Q + S$$

where Q is a block-wise constant matrix containing expectation values $Q_{ij} = \mathbb{E}[W_{ij}]$, and S is composed of the zero-mean deviances $S = W - Q$. Combining the decomposition of the input perturbation and the matrix, we can define two linear systems

$$\begin{aligned} \Delta\nu_B &= (\mathbb{1} - Q)^{-1}\,\Delta\beta_B \\ \Delta\nu_M &= (\mathbb{1} - S)^{-1}\,\Delta\beta_M. \end{aligned} \tag{10}$$

In analogy to what was suggested by Sadeh et al. (2014), these two systems can be interpreted as two separate processing pathways of the network. In the baseline system, the matrix Q defines the processing of the untuned baseline input, while in the modulation system, the matrix S is responsible for the input modulation. However, this characterizes the system only, if there is no interference between the two pathways: The output $\Delta\nu_B$ should not contain a modulation component, and the output $\Delta\nu_M$ should not contain a baseline component. Furthermore, it is not obvious that the combination $\Delta\nu = \Delta\nu_B + \Delta\nu_M$ solves the original W-system described in Equation (9).

We start verifying these requirements by showing that $\Delta\nu_B$ is constant across populations and thus does not contain a modulation component. To this end, a symmetry argument can be applied: Let $\Delta\nu_{Bi} = x$ and $\Delta\nu_{Bj} = y$ for two neurons i and j of the same population, then $\Delta\nu_{Bi} = y$, $\Delta\nu_{Bj} = x$ must also be a solution of the system, as Q is block-wise constant. As the operator $(\mathbb{1} - Q)^{-1}$ is invertible and the system has only one solution, it follows that $x = y$ and $\Delta\nu_B$ must be constant for all neurons from the same population.

In order to show that $\Delta\nu_M$ does not have a baseline component, we look at the modulation system (Equation (10)) in its element-wise, implicit form

$$\sum_{j=1}^{N} (\delta_{ij} - S_{ij})\,\Delta\nu_{Mj} = \Delta\nu_{Mi} - \sum_{j=1}^{N} S_{ij}\,\Delta\nu_{Mj} = \Delta\beta_{Mi},$$

where δ_{ij} is the Kronecker delta function. Changing to the expectation value of the expression and assuming independence of S_{ij} and $\Delta\nu_{Mj}$ leads to

$$\mathbb{E}[\Delta\nu_{Mi}] - \sum_{j=1}^{N} \mathbb{E}[S_{ij}]\,\mathbb{E}[\Delta\nu_{Mj}] = \mathbb{E}[\Delta\beta_{Mi}].$$

Noting that $\mathbb{E}[S_{ij}] = 0$ and $\mathbb{E}[\Delta\beta_{Mi}] = 0$ by construction, this shows that

$$\mathbb{E}[\Delta\nu_{Mi}] = 0 \tag{11}$$

and thus that $\Delta\nu_M$ has no baseline component.

Finally, we show that $\Delta\nu = \Delta\nu_B + \Delta\nu_M$ yields an approximate solution of the W-system. For this, we substitute this ansatz and the decomposition $W = Q + S$ into

Equation (9)

$$
\begin{aligned}
(\mathbb{1} - W)\,\Delta\nu &= (\mathbb{1} - Q - S)\,(\Delta\nu_B + \Delta\nu_M) = \\
&= \Delta\nu_B + \Delta\nu_M - Q\,\Delta\nu_B - \underbrace{Q\,\Delta\nu_M}_{(a)} - \underbrace{S\,\Delta\nu_B}_{(b)} - S\,\Delta\nu_M. \qquad (12)
\end{aligned}
$$

We will now analyze the two interference terms (a) and (b) in more detail

(a) The matrix Q is block-wise constant with value $Q_{ij} = q_{PT}$ for neurons i and j in populations P and T, respectively. For each element in the vector $Q\,\Delta\nu_M$, we then have

$$(Q\,\Delta\nu_M)_i = \sum_T \sum_{j\in T} Q_{ij}\,\Delta\nu_{M\,j} = \sum_T q_{PT} \sum_{j\in T} \Delta\nu_{M\,j}.$$

Using Equation (11), we know that $\sum_{j\in T} \Delta\nu_M \approx 0$ and therefore also $Q\,\Delta\nu_M \approx 0$.

(b) Similarly, for each element of the second interference term, we can write

$$(S\,\Delta\nu_B)_i = \sum_T \Delta\nu_{BT} \sum_{j\in T} S_{ij},$$

where now $\Delta\nu_{B\,j} = \Delta\nu_{BT}$ are the values of the population-wise constant vector $\Delta\nu_B$. The matrix S is defined as the zero-mean deviations between W and Q. Therefore, for sufficiently large population sizes and small variances in the connectivity, we have $\sum_{j\in T} S_{ij} \approx 0$ which leads to $S\,\Delta\nu_B \approx 0$.

Combining these findings with Equation (12), the expression can be reduced to

$$(\mathbb{1} - W)\,\Delta\nu \approx (\mathbb{1} - Q)\,\Delta\nu_B + (\mathbb{1} - S)\,\Delta\nu_M = \Delta\beta_B + \Delta\beta_M = \Delta\beta,$$

by definition of $\Delta\nu_B$ and $\Delta\nu_M$ (Equation (10)), meaning that $\Delta\nu = \Delta\nu_B + \Delta\nu_M$ indeed provides an approximate solution to the system.

To summarize, we showed that instead of solving the linear model via the W-system directly, it can also be studied in terms of the two pathways defined by Equation (10). Provided the two interference terms in Equation (12) can indeed be neglected for the network at hand, this is approximately equivalent and potentially more informative than the direct solution.

External stimulation In addition to the synaptic input provided by the background and thalamus, neurons can also receive additional external stimulation I^{stim}. Such an input could, for example, model optogenetic microstimulation. In the single neuron rate model (Model B), this can be accounted for by an additional term in the total mean input μ_i given by $\mu_i^{\text{stim}} = RI^{\text{stim}}$. As we consider here only constant stimulation, the variance ${\sigma_i^{\text{stim}}}^2$ of this input is zero.

In the linear model (Model C), applying a current results in an additional input perturbation and an extra term in the total derivative (Equation (8)). In the explicit expression Equation (9), this can be accounted for by

$$\Delta\nu = (\mathbb{1} - W)^{-1}\,(\Delta\beta + \Delta\gamma).$$

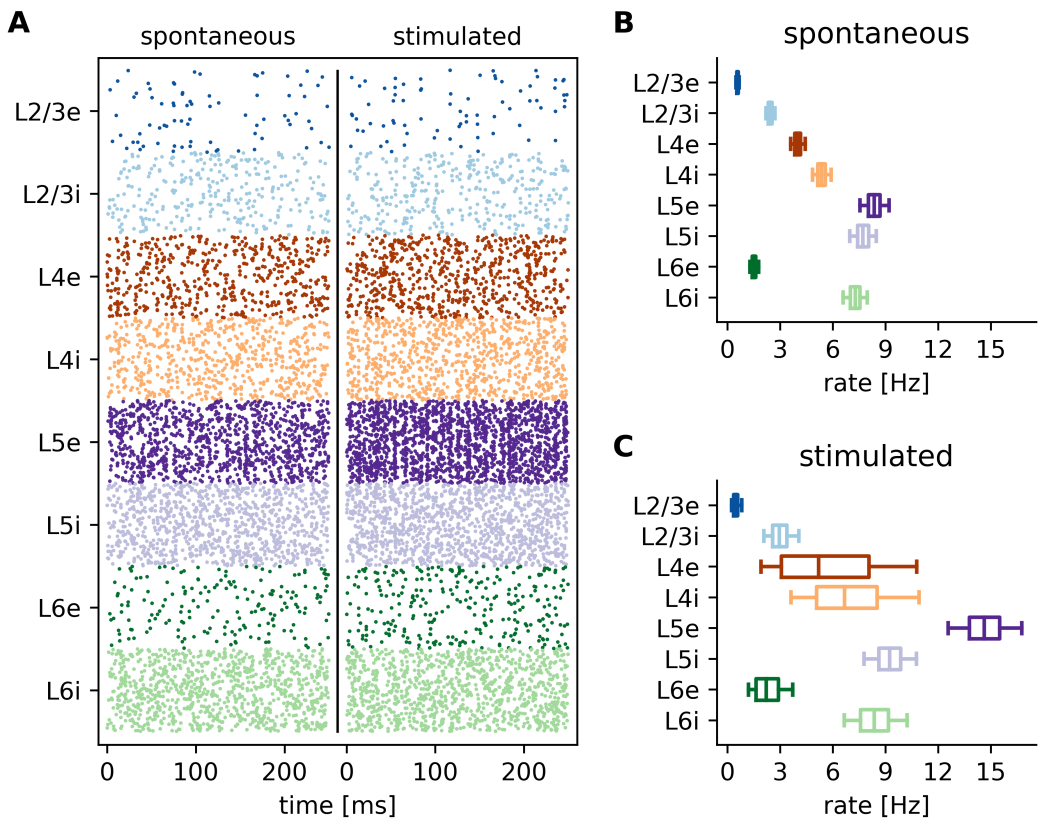

Figure 3: Spiking neuronal network. A 250 ms raster plot for both spontaneous and stimulated conditions, showing 500 neurons of each of the eight subpopulations. **B-C** Box plots of the firing rate distribution within each population, for spontaneous (B) and stimulated (C) conditions, respectively. Shown are the median (central line), lower and upper quartiles (box) as well as 5% and 95% percentiles (whisker bars).

The components of $\Delta\gamma$ are defined in analogy to the thalamic perturbation by

$$\Delta\gamma_i = \frac{\partial F_i}{\partial I_i}\,\Delta I_i\bigg|_{\nu=\nu_{\mathrm{OP}}}.$$

3 Results

3.1 Spiking neuronal network simulations

The results of the spiking neural network simulations performed in NEST, and of the detailed data analysis performed on the simulated data, are summarized in Figure 3, both for the spontaneous (weak, untuned thalamic input) and stimulated (tuned thalamic input) condition. As expected, the activity is very similar to the original results of Potjans and Diesmann (2014). As informally demonstrated by the raster plot in Figure 3A, for both conditions the network operates in the asynchronous irregular regime (Brunel, 2000; Renart et al., 2010). The single neuron firing rates match the values reported by Potjans and Diesmann (2014) (Figure 3B/C) and thus also fit well to experimental observations.

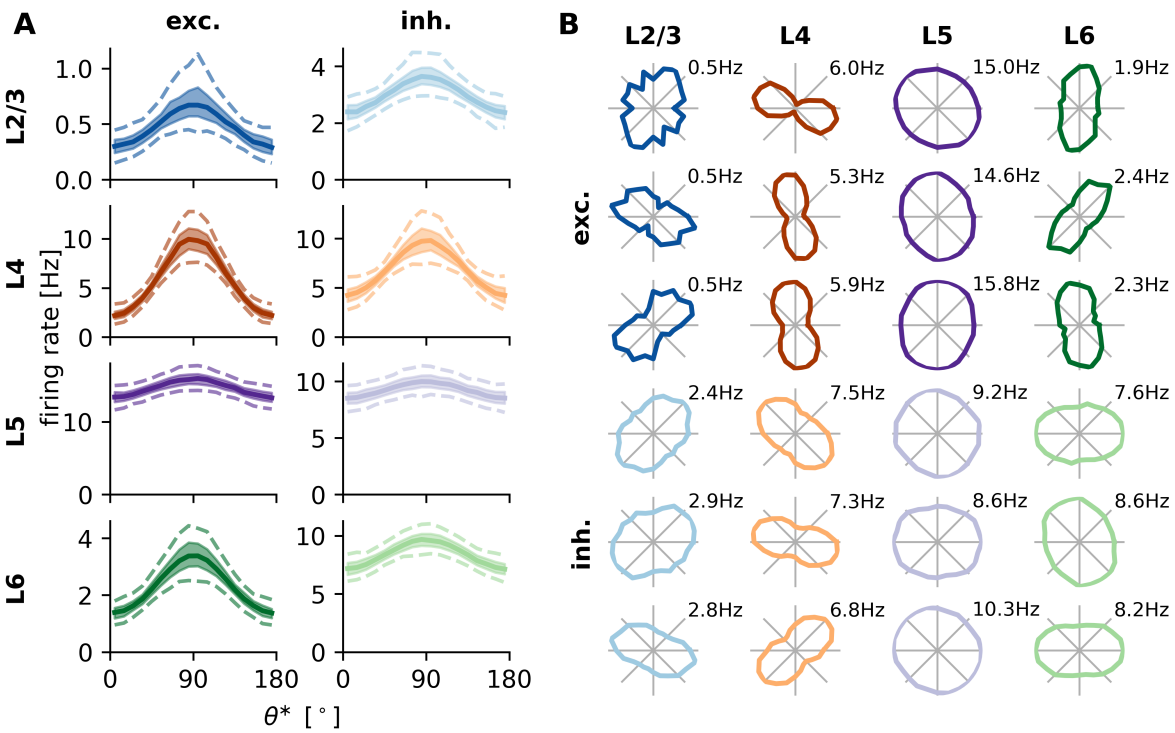

Figure 4: Single neuron tuning curves. A Single neuron tuning curves are centered about their respective preferred orientation and then binned over angles (bin size 10°). 5%, 25%, 50% (median), 75% and 95% percentiles are then calculated for each bin independently and plotted using different line styles. **B** Sample tuning curves from each population in polar representation.

In particular, we found very low activity in L2/3 and higher rates in the granular and infragranular layers, with exception of population L6e, where rates are also low (Niell and Stryker, 2008; Durand et al., 2016).

Our model extends the work of Potjans and Diesmann (2014) by giving the thalamic input a weak orientation bias, similar to Hansel and van Vreeswijk (2012); Sadeh et al. (2014). The macroscopic connectivity pattern of the network (Figure 1A) already gives a rough idea of the signal flow through the network. Thalamic afferents project mainly to L4 and to a smaller extent also to L6. From L4, the signal is transmitted to L2/3, which is thought to be the central processing unit of the microcircuit. Signals are then forwarded to higher brain areas as well as to L5, which is the major output unit for projections to sub-cortical brain areas.

In order to dissect the computations performed by the network, we extract tuning curves of all neurons in the model network. As expected from previous studies (Hansel and van Vreeswijk, 2012; Sadeh et al., 2014; Sadeh and Rotter, 2014, 2015), during stimulation, the weak orientation bias of the thalamic input is amplified to strongly orientation selective responses in the input population (Figure 4). Interestingly, L2/3e also shows strong tuning although these neurons only receive input via random connections from L4 neurons. This phenomenon was previously described by Hansel and van Vreeswijk (2012), where projections from L4e to L2/3 were studied independently of other parts of the network. The strong tuning in L2/3e can be explained by a recurrent tuning amplification

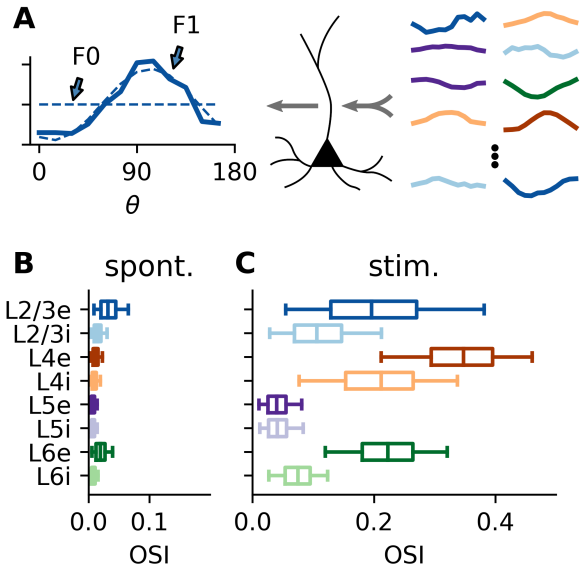

Figure 5: Orientation selectivity in the spiking neural network. A Every single neuron integrates its thalamic input with a large number of differently tuned inputs from various cortical populations (right) and generates tuned output (left). The OSI of the neuron is measured by the relative strength of the first (F1) as compared to the zeroth (F0) Fourier component of the output tuning curve (circular variance). **B-C** Orientation selectivity index (OSI) quantified for neurons in all layers, measured both under spontaneous and stimulated conditions. Percentiles are the same as in Figure 3.

in inhibition dominated networks (Hansel and van Vreeswijk, 2012; Sadeh et al., 2014). This effect is enough to create strong output tuning from a small bias in the input, which in turn is due to random sampling of preferred orientations in L4. In this context it may seem even more surprising that orientation selectivity of L5 neurons appears to be weaker compared to the other layers. While this is consistent with experimental findings (Niell and Stryker, 2008; Sun et al., 2016; Durand et al., 2016), it cannot be easily explained by inspecting the connectivity matrix. In fact, L5e and L2/3e neurons receive a comparable amount of input from L4e (Figure 1B, Table 1). Therefore, as the output tuning of L2/3e neurons is essentially inherited from these neurons, one would naively expect a similar tuning strength also in L5e.

In order to understand in more detail the differences in orientation selectivity between layers, we extracted the orientation selectivity index (OSI) of all neurons in the model network. We employ a measure based on the circular variance to quantify orientation selectivity (see **Material and Methods** for details). For each neuron, the OSI is calculated as the fraction of first (F1) and zeroth (F0) Fourier component of the tuning curve (Figure 5B/C). A stronger modulation of the firing rate over angles (higher F1), leads to a higher OSI, while a higher mean rate (higher F0) reduces the OSI.

Consistent with experimental findings and throughout all layers, inhibitory neurons have a lower OSI than the corresponding excitatory neurons (Niell and Stryker, 2008; Scholl et al., 2015; Durand et al., 2016). While this might be expected for L4 and L6 due to the lower number of thalamic afferents, it is not at all obvious for L2/3. In this population, it is a consequence of the recurrent network connectivity, as our analysis will

show.

Note that during spontaneous activity, neurons also exhibit weak orientation selectivity (Figure 5B). This is due to fluctuations in the spiking process. Neurons may randomly fire a few action potentials more for one orientation than for another, which results in weak but nonzero orientation selectivity. The magnitudes found here are consistent with a Poisson process with the same rate and recording time (data not shown).

Keeping in mind that orientation selectivity is measured as the ratio between zeroth and first Fourier component of the tuning curves (Figure 5A), the tuning curves also reveal that the weaker OSI of L5e is mostly due to the relatively high firing rates of the neurons (Figure 4A). This raises two questions: First, which projections onto L2/3e and L5e lead to the observed difference in activity? Second, why is the tuned (F1) component of L5e not as strongly amplified as the mean rate (F0)? In the following, we will address these questions, employing suitable mathematical and computational methods. While the first question can be answered by analyzing the input currents received by the different populations, the second question requires an approach, which takes the strongly recurrent nature of the circuit across layers into consideration.

3.2 Operating point of the network

Each neuron in the eight populations of the network receives input currents from several pre-synaptic populations, possibly all with different tuning curves, and forms its own output tuning from those (Figure 5A). It has been shown in experiments that this scenario also reflects the situation in animals (Jia et al., 2010). Which projections are most potent for driving the activity of the target neuron is not obvious from the connectivity between neurons (Figure 1B). The activity of pre-synaptic neurons is also relevant for the total input current of a given neuron. In Figure 6A, we plot the mean input current for each projection between populations (Equation (3)). While the difference between inputs to L2/3e and L5e appeared small when looking at the underlying connectivity (Figure 1B), that picture changes if activity is taken into consideration (Figure 6A). The input to L2/3e neurons is mostly determined by L2/3i, both L4 populations and the constant background. Input to L5e, on the other hand, is dominated by L5i and background input, but only to a lesser extent by inputs from L2/3i, L4e and L5e. As a result, due to the different excitation-inhibition (EI) balance, the net currents imply a higher mean input to L5e compared to L2/3e (Figure 6C).

The striking difference between pure connectivity (Figure 1B) and the observed total input currents (Figure 6A) highlights the dynamic contribution of the EI balance in recurrent networks (Van Vreeswijk and Sompolinsky, 1996, 1998; Brunel, 2000). While the balance appears to be very similar when counting the number of excitatory and inhibitory synapses that terminate on L5e and L4e, the actual drive they receive is quite different. Since the activity autonomously settles at an operating point where the EI-balance is dynamically maintained, this balance can be very different for the various populations.

Having shown how the operating point of the layered network explains the high mean firing rate of L5 neurons, we can now ask the same question with regard to the orientation tuning in each population. In combination, these two aspects fully determine the orientation selectivity of all neurons. To achieve this, we calculate for each neuron a tuning vector, which summarizes the tuning information contained in the input currents to that neuron. Its magnitude measures the tuning strength of the combined input to the

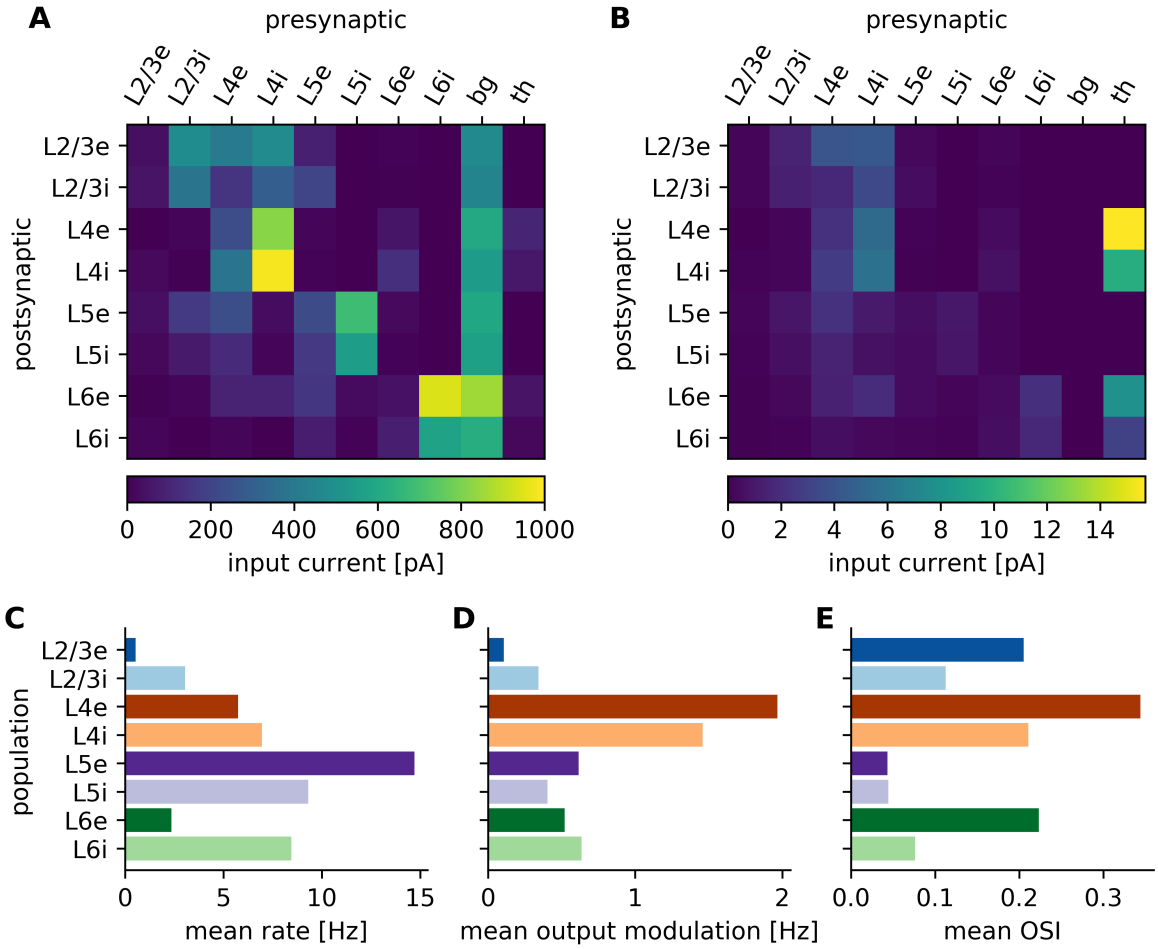

Figure 6: Neuronal input analysis. A Mean input currents for all projections between all the involved neuronal populations. **B** Mean projection tuning vector length for all recurrent projections (see text for explanation). **C-E** Mean output firing rates (F0 of tuning curves), mean output modulations (F1 of tuning curves) and resulting OSI averaged over each population.

post-synaptic neuron, whereas its direction indicates its preferred orientation (Figure 7 inset, Equation (4)). The total tuning input from one specific source population to a single neuron can then be calculated as the sum of all tuning vectors of neurons in the pre-synaptic population which connect to the post-synaptic neuron. If all tuning curves are cosine-like, this concept exactly corresponds to linear summation of inputs. In any case are the magnitudes of the tuning vectors identical to the first Fourier component (F1) of the input current tuning curves.

Figure 6B summarizes the mean lengths of all projection tuning vectors, summarizing the information flow in the system. The tuning information enters the cortical network in L4 and L6 and then spreads to the other populations. Although both L4 and L6 neurons show significant tuning to stimulus orientation, due to the low firing rates of L6e neurons it is mostly L4, which provides the tuning information to the other populations. In contrast to what would be expected from the connectivity alone (Figure 1B), also L5e neurons receive most of the tuned current from L4. Comparing the input to L2/3e and

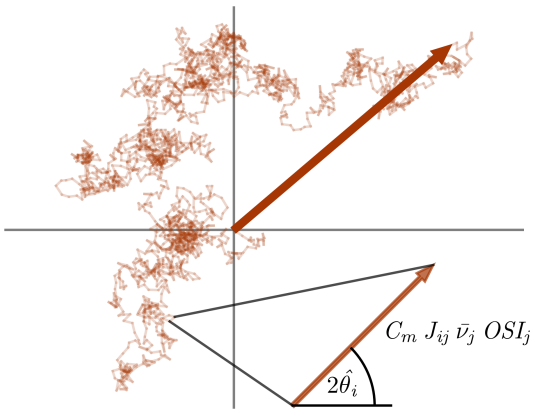

Figure 7: Tuning vectors. Illustration of single neuron input tuning vector addition. Each small arrow represents the tuning vector of a single pre-synaptic neuron. Linearly adding the vectors of all pre-synaptic neurons of one population results in the tuning vector that characterizes the projection to one post-synaptic neuron.

L5e neurons in more detail shows that L2/3e neurons mostly integrate tuned input from both populations in L4, while L4e neurons receive less tuned input almost exclusively from other cells in the same population.

When looking at the average output modulation of the different populations (Figure 6D), which is defined as the first Fourier component of the tuning curves (Figure 5A), it is surprising that despite the less tuned input to L5e, these neurons still shows a stronger output modulation than L2/3e. This highlights that, besides the input strength, the input sensitivity of neurons are relevant as well. However, the large modulation component of the output of L5e neurons cannot compensate their high firing rates, leading to a lower OSI (Figure 6C-E). Comparing the distribution of firing rates and output modulations over the eight populations (Figure 6C/D), it is evident that these two components are subject to different gain factors, leading to different orientation selectivities depending on the size of these gains. In the following, this will be studied in more detail.

3.3 Two processing pathways: baseline and modulation

Our analyses so far disentangled the input to neurons for the different populations, which result in the observed neuron output features. However, the multi-population model is highly recurrent. Therefore, it is not clear yet why the input to the neurons settled at the observed operating point. In other words, we face a chicken-and-egg problem, because in the recurrent network, the mean output of all populations are simultaneously the input to the same populations, eventually settling in the dynamical operating point. In order to tackle this problem, we now change our perspective from analyzing the input-output relation of single neurons to input-output behavior of the entire network. The goal is to manage the problem by considering the network as a distributed system, which can perform its function only as a whole.

This approach is best realized by linearizing the network dynamics about its operating point. We obtain our almost 80 000-dimensional linearized network model (Model C) by

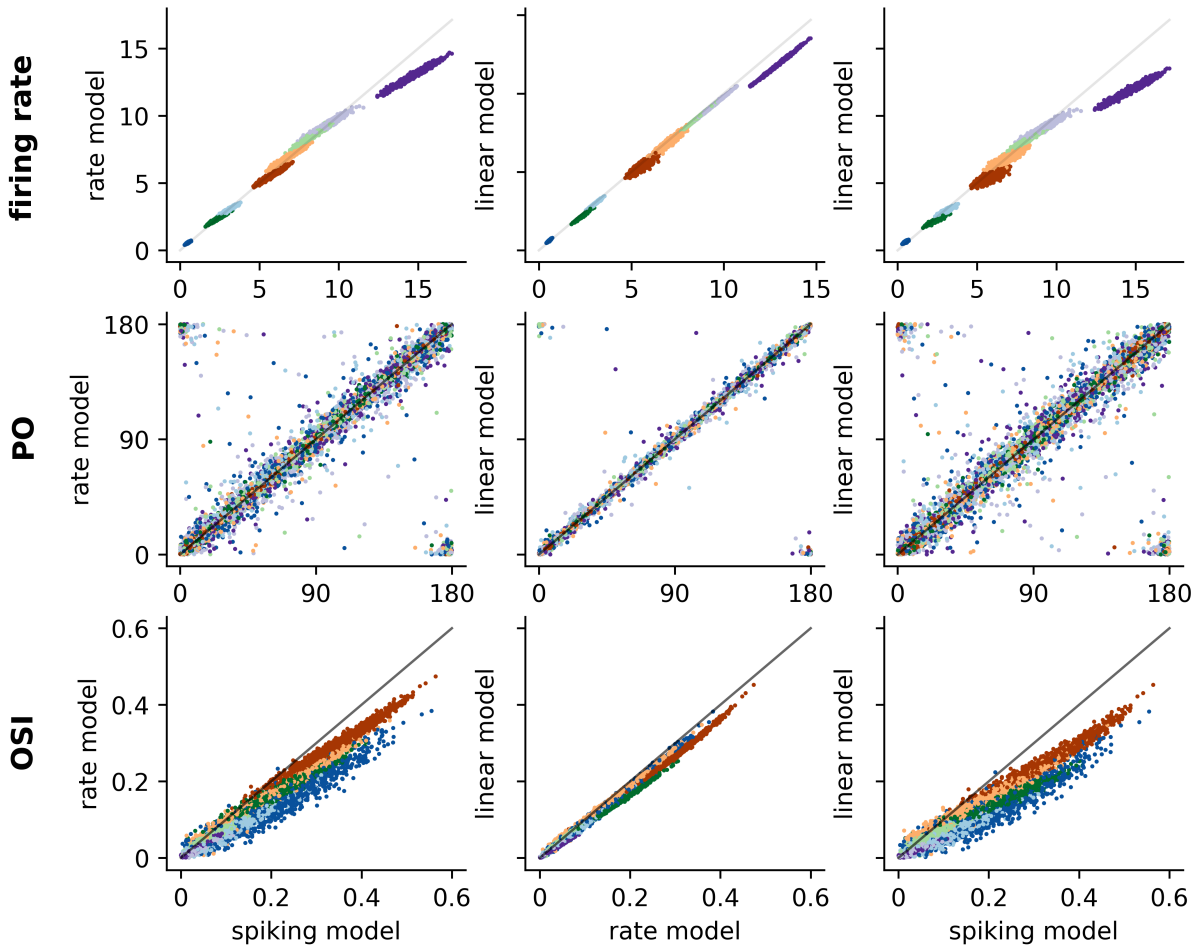

Figure 8: Modeling results. Comparison of the three central observables of the different network models. Top row: single neuron firing rates, middle row: preferred orientation (PO), bottom row: orientation selectivity index (OSI). The first column compares the spiking network model (Model A) with the firing rate model (Model B), the second column compares the firing rate model with linearized network model (Model C) and the third column compares the spiking network model with the linearized network model. Same color code is used as in previous figures to indicate the eight different populations.

first formulating a single neuron firing rate model based on the single neuron transfer function $F_i(\nu, \nu^{\text{th}})$ (Model B). The firing rate model is then linearized about its dynamical operating point, leading to the explicit input-output relation

$$\Delta\nu = (\mathbb{1} - W)^{-1}\Delta\beta. \tag{13}$$

Here, $\Delta\nu$ is a vector summarizing all single neuron rate changes due to a change in the effective input $\Delta\beta$, which in turn is the change in thalamic input rate scaled by the input sensitivities of the different populations. These sensitivities can also be understood as the feed-forward gains of the system (Sadeh et al., 2014), see **Material and Methods** for details. Furthermore, the recurrent model behavior is governed by the matrix W, which is composed of the recurrent sensitivities of all the neurons in the network. In combination

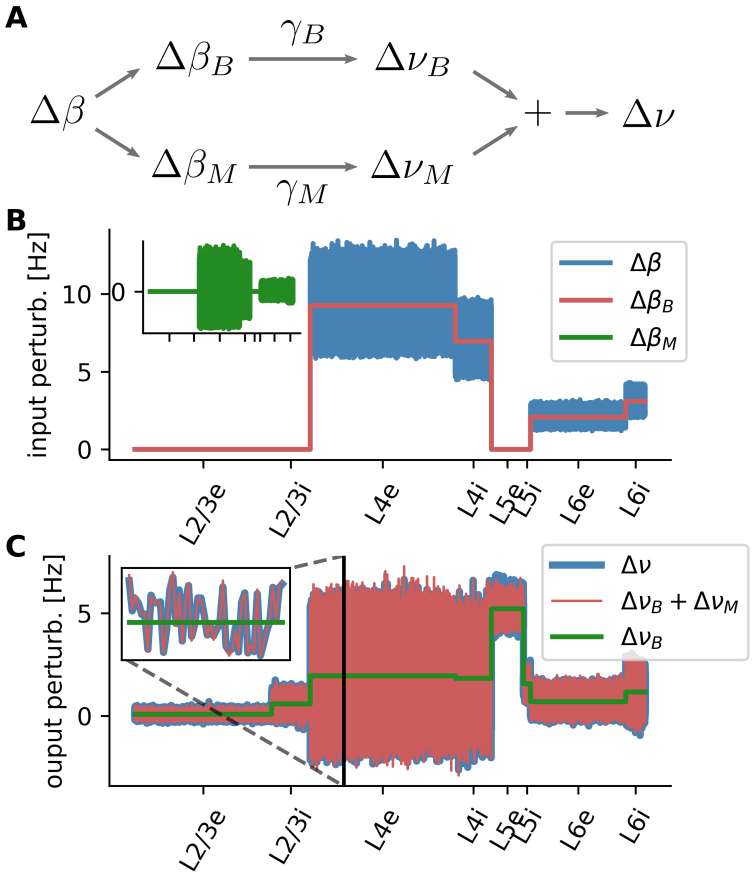

Figure 9: Separation of baseline and modulation. A The input perturbation $\Delta\beta$ can be decomposed into baseline and modulation, which are then amplified or attenuated by the network with different gains. **B** Input perturbation for one particular orientation of the oriented stimulus. For each neuron in the eight populations of the layered network, the effective change in input rate is plotted (blue). The baseline (population-wise mean) is shown in red. The modulation, which is the difference between the two, is plotted in the inset (green). **C** Solutions for the baseline and modulation system. Shown are the baseline (green) and baseline plus modulation (red) output rates. For comparison, also the solution of the full W system is shown (blue). The inset shows a magnification for a small sample of neurons in population L4e.

with the activity at the operating point ν_0, the rate change $\Delta\nu$ results in the output activity $\nu = \nu_0 + \Delta\nu$ of the linear network model.

The explicit form of Equation (13) also supports the idea of considering the network as one system with a combined (vector-valued) input-output relation. Importantly, this relation is governed by the inverse of the effective connectivity $\mathbb{1} - W$. Because the entries in W_{ij} are a function of K_{ij}, this inversion potentially distributes the influence of each single connectivity parameter K_{ij} over the entire network. This observation also explains why the effects of individual connectivity parameters can be counterintuitive (Tsodyks et al., 1997).

The firing rate model (Model B) as well as the linear network model (Model C) can be considered as simplifications of the spiking neural network (Model A), as they make

additional assumptions about the dynamics. Therefore, before analyzing these models in more detail, it is important to establish the consistency of their respective behavior. Figure 8 compares the simulation results for all three models. It shows that the single neuron firing rates of the three models are in very good agreement. Only for the large firing rates of L5e some minor discrepancies emerge. Similarly, also the preferred orientations and orientation selectivities are in good agreement. While the match is excellent for the firing rate and linear network model, the spiking network model shows slightly higher discrepancies. This can be explained by the role played by activity fluctuations in this model. In particular, the observed orientation selectivities are somewhat higher, reflecting the same positive bias as the one observed during spontaneous activity.

Now that consistency in the behavior of the different models is established, we can analyze the linearized network. In previous work, a separation into two separate, non-interfering pathways through the recurrent network with distinct gain factors was postulated (Sadeh et al., 2014) (Figure 9A). In this interpretation, the effective input perturbation is split into a baseline and a modulation component such that $\Delta\beta = \Delta\beta_B + \Delta\beta_M$. Here, we generalize this decomposition to our present multi-population model. Furthermore, we present a novel analytical approach for their analysis.

For each population, the baseline $\Delta\beta_B$ is conceived as the mean input perturbation these neurons receive during stimulation. Since the mean input is different for each population, this leads to a population-wise constant input vector (Figure 9B, red curve). Furthermore, the modulation $\Delta\beta_M$ accounts for all neuron-by-neuron deviations from the population baseline, i.e. $\Delta\beta_M := \Delta\beta - \Delta\beta_B$, leading to a vector with zero mean in each population (Figure 9B, inset). Most importantly, the modulation part includes the tuning information that each neurons receives, based on the neuron-specific orientation bias of the stimulus. In other words, the deviation of $\Delta\beta_B$ from the baseline for each single neuron is mostly determined by the orientation of the stimulus, as described by Equation (6).

Note that the baseline and modulation components are only indirectly related to mean rate (F0) and tuning strength (F1) discussed in the previous section. While those where calculated for the tuning curve of each single neuron independently, the baseline and modulation of the input are calculated over all neurons in one population, for a single stimulation angle.

Separation of baseline and modulation pathway means that these two inputs $\Delta\beta_B$ and $\Delta\beta_M$ are processed by two independent mechanisms with separate, independent gains γ_B and γ_M. As shown in **Material and Methods**, the two pathways are governed by

$$\Delta\nu_B = (\mathbb{1} - Q)^{-1}\,\Delta\beta_B$$
$$\Delta\nu_M = (\mathbb{1} - S)^{-1}\,\Delta\beta_M,$$

where the decomposition $W = Q + S$ was used with Q containing population-wise expectation values, resulting in an 8-by-8 block structure, and S are the individual modulations $S = W - Q$ of single connections in the network. Importantly, there is no cross-talk between the two pathways, i.e. a change in the input baseline strength $\Delta\beta_B$ does not lead to change in the output modulation $\Delta\nu_M$, and vice versa.

Solving these two systems independently yields two output rate changes $\Delta\nu_B$ and $\Delta\nu_M$. If the cross-terms in the calculations are negligible, the solution $\Delta\nu = \Delta\nu_B + \Delta\nu_M$ provides a good approximation to the network behavior (see Equation (12)ff). For the network

under study here, the mean magnitudes μ of the cross-terms are $\mu[|Q\Delta\nu_M|] = 0.02$ and $\mu[|S\Delta\nu_B|] = 0.09$ as compared to $\mu[|\Delta\beta|] = 3.63$, implying that the system can indeed be studied by treating the baseline and the modulation pathway separately.

The solutions of the baseline and modulation systems, $\Delta\nu_B$ and $\Delta\nu_M$, are shown in Figure 9C in comparison to the solution of the W-system. Throughout all populations, the separation solution is in excellent agreement with the direct solution. Looking in more detail at the modulation (Figure 9C inset) shows that the modulation part, which carries the tuning information of the neurons, matches almost perfectly.

In previous sections, we identified the high baseline rate of L5e as the main reason for the low orientation selectivity of its neurons. This is manifested as a high baseline output rate $\Delta\nu_B$ for this population (Figure 9C). While this still does not fully reveal the underlying reason for the high rates, it hints at the effective mean connectivity represented by Q. In the following, we will therefore study the baseline pathway in more detail.

3.4 Mode decomposition

We start by summarizing the results of Sadeh et al. (2014, 2015), where a similar scenario for a two population EI network was studied. In analogy to the present model, also the two population system can be described by a linear system of the form

$$\Delta\nu = (\mathbb{1} - W)^{-1}\Delta\beta,$$

where the dimensionality of the system equals the total number of neurons. In order to study the network behavior, it is useful to study the eigenvalue spectrum of the matrix W. In the two population case, when both populations receive identical input, the eigenvalue spectrum consists of a bulk localized at the origin of known radius and a single exceptional eigenvalue λ (Rajan and Abbott, 2006). For inhibition dominated networks, this exceptional eigenvalue is real and negative (Figure 10A). Furthermore, the corresponding eigenvector Ψ to that exceptional eigenvalue is the constant vector, i.e. all entries are identical. In this two population network model, the baseline component of the input perturbation is also proportional to the constant vector

$$\Delta\beta_B \propto \Psi \propto \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Therefore, exploiting the eigenvector characteristic of the perturbation, the exceptional eigenvalue can be transformed into the gain factor of the baseline. In mathematical terms, this can be written as

$$\Delta\nu_B = (\mathbb{1} - W)^{-1}\,\Delta\beta_B = (1-\lambda)^{-1}\,\Delta\beta_B = \tilde{\lambda}\,\Delta\beta_B,$$

For inhibition dominated networks, this gain factor $\tilde{\lambda}$ has a small magnitude and thus results in a strong attenuation of the baseline component. This, in turn, amplifies the orientation selectivity of the network.

Generalizing the two population scenario to the eight population network considered here, there are two major differences in terms of the eigenvalue spectrum of W. First, instead of a single exceptional eigenvalue, the spectrum of the effective connectivity is

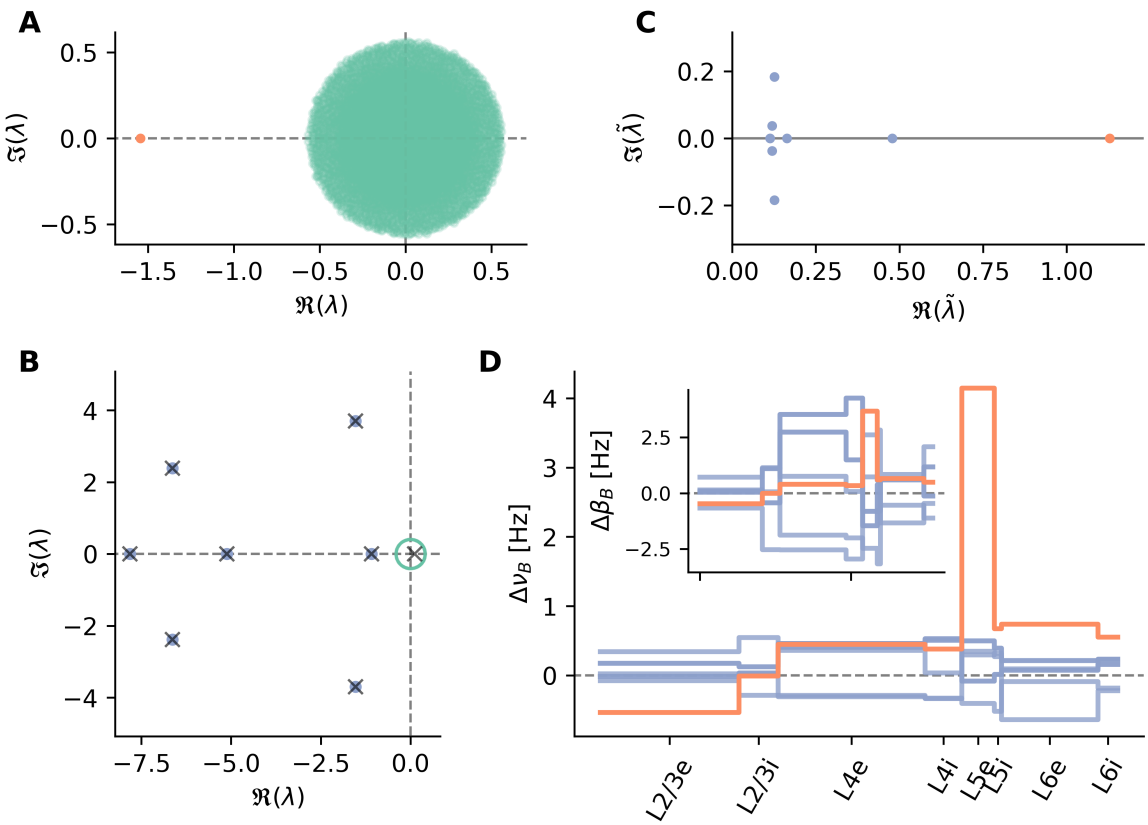

Figure 10: Eigenmode analysis of input-output relations. A Eigenvalue spectrum of the linearized model of Brunel (2000). The exceptional eigenvalue is shown in red, the bulk spectrum in green. **B** Eigenvalues of the almost 80 000-dimensional linear system. Red and green dots indicate the exceptional and bulk eigenvalues of W, respectively. Crosses mark eigenvalues of Q. **C** Eigenvalues of the matrix $(\mathbb{1} - Q)^{-1}$. **D** The eight output modes of the Q system. Shown are the change in output rate for each single neuron, due to each mode. Inset: Input perturbation decomposed into the eight modes, which are also the right eigenvectors of Q.

more complex (Figure 10B). In addition to the bulk of eigenvalues (diameter indicated in green in Figure 10B), there are seven eigenvalues with a significantly larger magnitude (blue dots in Figure 10B). In a random matrix theory context, it was previously shown that the exceptional and bulk eigenvalues are due to the expectation values and variances of the distribution of the elements of W. Therefore, the exceptional eigenvalues are asymptotically identical to the eigenvalues of Q (Aljadeff et al., 2016). Indeed, the first seven eigenvalues are in good agreement with the numerical calculations on W (crosses in Figure 10B). Furthermore, an eighth exceptional eigenvalue is revealed which lies in the middle of the bulk and could therefore not be distinguished from bulk eigenvalues in numerical calculations on W.

The second difference compared to the two population scenario is the fact that the constant vector is not any more an eigenvector of an exceptional eigenvalue. Instead, the eigenvectors of the baseline system Q are only population-wise constant (c.f. Figure 10D inset). However, also the input perturbation is not proportional to the constant vector

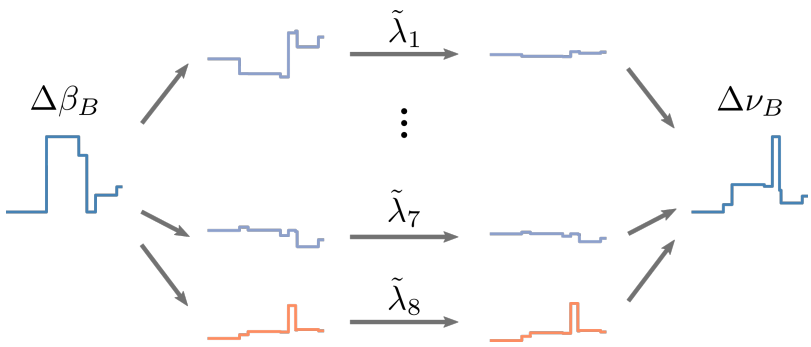

Figure 11: Illustration of mode decomposition. Left: The baseline component of the input perturbation $\Delta\beta_B$ is decomposed into the eight input modes $\xi_i\Psi_i$. Center: By the network action, each mode is multiplied by its individual gain factor $\tilde{\lambda}_i$. Right: The scaled output modes $\tilde{\lambda}_i\xi_i\Psi_i$ form the baseline component of the output perturbation $\Delta\nu_B$.

in this case. For each population, it depends on the sensitivity to the thalamic input, also resulting in a population-wise constant function. In order to study the input-output behavior, the input can therefore be decomposed into the eight eigenvectors

$$\Delta\beta_B = \sum_{i=1}^{8} \xi_i\Psi_i$$

(cf. Figure 11 left). Here, Ψ_i are the eigenvectors of Q and ξ_i are the coefficients of the modes. The components $\xi_i\Psi_i$ can be considered as the *input modes* of the input perturbation $\Delta\beta_B$. The eight modes which represent the perturbation of the stimulation condition are shown in the inset of Figure 10D. Note that the sum over these modes is identical to $\Delta\beta_B$ in Figure 9B. Applying the decomposition to the linear baseline system of Equation (14) results in a decomposition of the output rates of the network, where the eight exceptional eigenvalues define the gain factors of the individual modes by

$$\Delta\nu_B = \sum_{i=1}^{8} \tilde{\lambda}_i\xi_i\Psi_i \tag{14}$$

(Figure 11 right). The transformed gain factors $\tilde{\lambda}_i$ are shown in Figure 10C. The parts of this decomposition can be considered as *output modes* of the system, with a direct one-to-one relation to the input modes. The effect of the exceptional eigenvalues λ is quite similar to the two-population scenario described previously, where the single exceptional eigenvalue defines the gain of the homogeneous common mode. In the multi-population case, each input mode is amplified with the respective gain $\tilde{\lambda}_i$ to form the output perturbation of the network. The eight output modes are shown in Figure 10D. Note that, similar to the input perturbation $\Delta\beta_B$, the sum of the modes is identical to the baseline component $\Delta\nu_B$, which is shown in Figure 9C.

Studying the output modes in more detail, it is apparent that the high rates of L5e neurons, and thus also their low orientation selectivity, are mostly due to one specific mode (orange line in Figure 10D). On the other hand, the same mode has a similar strength as the other components in the input space (orange line, Figure 10D inset). Because

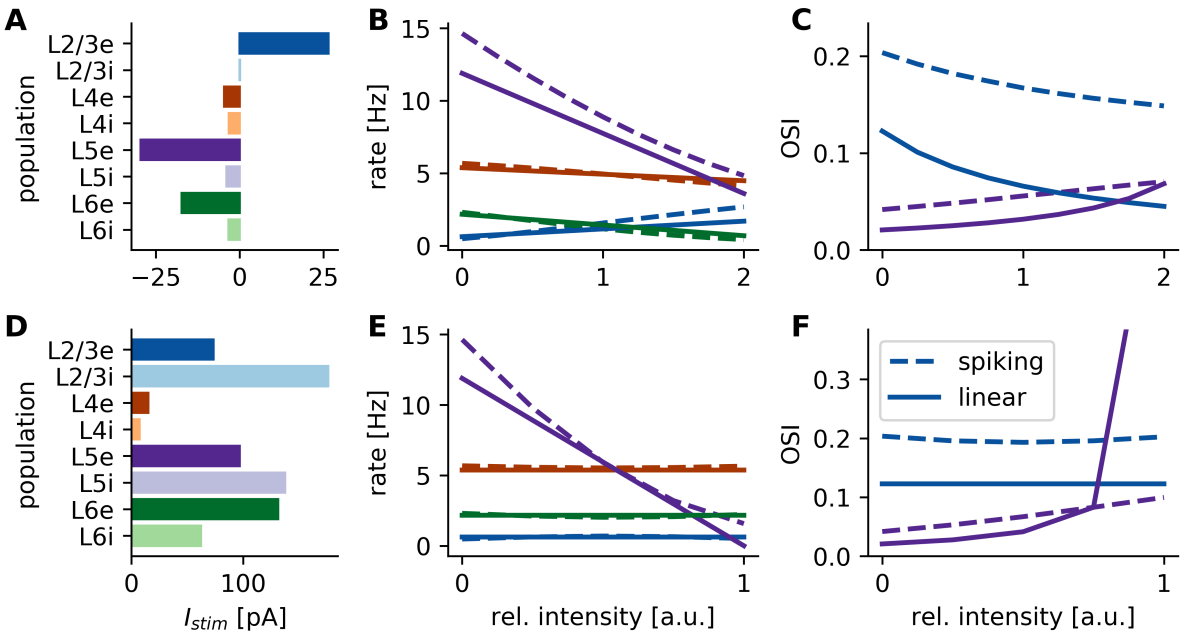

Figure 12: Predictions for future current stimulation experiments. A-C Current stimulus 1: Effective deletion of the distinct mode from the input. **D-F** Current stimulus 2: Selective suppression of L5e activity. **A/D** Single neuron current stimulation required to achieve the effect. **B/E** Change of mean population rate for excitatory populations upon current stimulation. Inhibitory populations are omitted for simplicity. x-axis shows current intensity relative to values in A/D, respectively. In all panels, lines represent the linear model (Model C), dashed lines the spiking neural network simulations (Model A). **C/F** Change of orientation selectivity in L2/3e and L5e due to the two different current stimulations.

the corresponding eigenvalue of W has a very small magnitude, however, the gain factor $\tilde{\lambda}$ of that mode is much larger compared to the other modes (Figure 10C, orange dot). Due to this gain, the mode is dominant in the output rate change, resulting in a strong rate increase in L5e. The fact that only a single mode is responsible for the large gain underlines the observation that this effect is a feature of the whole network, which cannot be pinned down to one specific projection in a meaningful way.

3.5 Predictions for new stimulation experiments

Having identified the large gain of a specific mode in the input as the source of the strong firing rate increase of L5e, we now apply our theory to predict the network behavior for a scenario where that particular mode is removed from the input. For two reaseons, the mode cannot be directly removed from the mean input firing rates. First, that mode has nonzero values for all populations, which would lead to non-zero thalamic input rates also to L2/3 and L5. While it is technically possible to implement this in a computational model, it is infeasible as a biological experiment. The second problem is more fundamental: Subtracting the mode from the input perturbation would lead to negative input firing rates, which is impossible to realize. Therefore, instead of using an altered Poisson input

to the neurons, we designed a constant external current stimulation, which is equivalent in its effect on the neuron output rates. It is conceivable to realize such an input by specific optogenetic stimulation. In the linear model (Model C), an external current results in an additional input perturbation $\Delta\gamma$ such that

$$\Delta\nu = (\mathbb{1} - W)^{-1} (\Delta\beta + \Delta\gamma).$$

In order to delete a specific mode Ψ_k from the input, we can set $\Delta\gamma = -\xi_k \Psi_k$. Therefore, when the two inputs $\Delta\beta$ and $\Delta\gamma$ are combined, the mode cancels out, resulting in altered output perturbation.

The required single neuron input currents for each population are shown in Figure 12A. They are not directly proportional to the deleted mode, due to the scaling by the individual feed-forward gain of each population, as well as due to the highly recurrent processing of the network. As expected, since the rate of L5e neurons should be suppressed, these neurons receive a strongly negative current. More surprisingly, L2/3e requires a positive input current of similar magnitude, although the desired change in rate of that population is much smaller.

When the current is applied with an intensity that exactly removes the mode from the input, the change in output exactly matches the sum of the remaining modes (Figure 12B, relative intensity 1). For stronger stimulation intensities, the mode is also partially removed from the operating point activity. The predicted mean population rates of the linearized network and spiking network model are in good agreement. Also in the spiking neural network model, the mean rates change almost linearly even if the input mode is overcompensated by a factor of two.

As expected, the reduction of L5e firing rates leads to an increase in orientation selectivity of 250% from 0.02 to 0.07 in the linearized network model (75% in the spiking network model, Figure 12C). Simultaneously, due to the increase in firing rate in L2/3e, the orientation selectivity of this population decreases by 66% from 0.12 to 0.04 in the linearized network model (25% in the spiking network model). The discrepancy between the OSI of the spiking and linearized network model can be partially explained by the bias of random tuning due to the fluctuations of the spiking process, which were also the reason for non-zero OSIs observed the spontaneous, unstimulated condition in Figure 5B. A second explanation, in particular for the stronger change of OSI with stronger current intensities, could be a weak crosstalk between the baseline and modulation pathways introduced by nonlinear effects in the spiking neural network model.

Effectively deleting the mode from the input is not the only option to reduce the activity of L5e neurons. From our analytic calculations, all feed-forward and recurrent baseline gains are known. Exploiting this knowledge, we can design a second, new current stimulus that exclusively affects L5e neurons, leaving the mean firing rates of all other populations unchanged (**Material and Methods**). The desired change of rates for this stimulus can be conveniently summarized in a vector $\Delta\nu^{\mathrm{L5e}}$. This vector is zero for all populations except L5e. For this population, it contains the negative of the mean rates in the stimulation condition such that the resulting rates vanish. The required effective external perturbation can then be calculated by

$$\Delta\gamma = (\mathbb{1} - W)\, \Delta\nu^{\mathrm{L5e}},$$

which are the external currents scaled with the known neuron sensitivities.

The required stimulation currents are shown in Figure 12D. Surprisingly, in this case, the input currents to L5e neurons all need to be positive, although the desired effect on the output rates is negative, i.e. the rates of these neurons are reduced. Also interesting is that input currents to all neurons except L4 are of similar strength. It is clear that the other populations also require some stimulation currents to maintain their original rate since the altered rate of L5e neurons needs to be compensated. However, the magnitude of these currents is surprising since L5 does not serve as a major input to any other population within the network (Figure 1B, Figure 6A). This highlights again the strongly recurrent and often unintuitive nature of such networks.

Figure 12E shows that the new current stimulus successfully reduces the rate of L5e neurons while keeping all other mean rates constant. Similar to the first current stimulus, the two models match quite well. Only for strong stimulation intensities, when the rates of L5e neurons approach zero, the two models show discrepancies. In this regime, the rates of L5e begin to be rectified. Since this is a non-linear effect, it can of course not be captured by the linearized version of the model.

Analogous to the first prediction for current stimulation, L5e neurons show a strong increase in orientation selectivity (Figure 12F). In the linear model, the OSI rises very sharply for large current intensities. This is an artifact of the neglected rectification of neuron rates, as already observed above. In the more realistic spiking network model (Model A), the OSI also increases strongly by about 150% (from 0.04 to 0.1). However, in contrast to the first current stimulus discussed above, in this case orientation selectivity of L2/3e neurons is not compromised by the additional input currents. Similarly, also the OSIs of the other excitatory and inhibitory populations remain unchanged (data not shown).

4 Discussion

Model The goal of this work was to study signal propagation across layers in primary visual cortex of rodents, with a particular focus on the emergence of orientation tuning in those layers, which have no direct access to thalamic input. We decided to utilize a model that was developed previously by Potjans and Diesmann (2014). The connectivity parameters of the model were taken from anatomical measurements, and since the availability of such data is still limited, the authors combined data from different species (cat and rat) into one single generic model circuit. In addition, the measurements were performed in several different brain regions, so it is open whether this model can account for information processing in rodent visual cortex. In view of this, it may seem surprising that the results of our study are in very good agreement with measurements performed in mouse visual cortex. This success is compatible with the idea of a universal connectivity scheme which is found in different species and in different cortical regions.

In our work, the original model proposed by Potjans and Diesmann (2014) was extended by adding neuron-specific thalamic input, which is tuned to the orientation of a visual stimulus (grating or light bar). This input was provided to both the excitatory and inhibitory populations in L4 and L6. Via random connections, a very weak orientation bias in the input was already enough to induce strong orientation tuning in L2/3, see Hansel and van Vreeswijk (2012), Sadeh et al. (2014) and Sadeh and Rotter (2015) for an analysis of this generic effect. Note, however, that we did not make any assumptions about the origin of the orientation bias in the input. It could possibly stem from the

alignment of the receptive fields of afferent neurons (Hubel and Wiesel, 1962), or it could be inherited from already tuned thalamic neurons (Marshel et al., 2012; Piscopo et al., 2013; Scholl et al., 2013). While the input was assumed to convey information about a visual stimulus throughout the present study, our analysis of the layered network described here does not rely on this interpretation. Our approach could therefore also be applied to study feature selectivity in other sensory modalities, like in somatosensory or auditory cortex.

The distribution of orientation selectivity that emerges across the different layers and neuronal populations is qualitatively very similar to experimental recordings (Niell and Stryker, 2008; Durand et al., 2016). This is a remarkable result as these recordings were performed in adult mice while our model entirely lacks specific connectivity, possibly more resembling the situation at eye opening (Ko et al., 2013). Our analysis demonstrates that cortical connectivity is tuned to function even without feature-specific connections between neurons.

In our model, the mouse which is subject to visual stimulation is assumed to be non-moving. As shown in a recent study (Durand et al., 2016), neuronal responses in primary visual cortex of rodents are very similar in awake and anesthetized animals. However, the responses are known to dramatically change during locomotion behavior (Niell and Stryker, 2010). Different inhibitory neuronal subtypes play a central role in this (Pakan et al., 2016). A future enhanced version of our model might account for such effects as well.

Operating point. While analyzing network activity in terms of firing rates and orientation tuning, it becomes apparent that plain inspection of connectivity cannot fully explain all features of the dynamics observed in network simulations. For example, although afferent connectivity from L4e is very similar, firing rates and orientation selectivity in L2/3e and L5e are quite different. As the input is comprised of both excitatory and inhibitory projections which can partially cancel in their effect, a small difference in the input may have strong effects on the operating point the two populations eventually settle in (Van Vreeswijk and Sompolinsky, 1996, 1998). Such effects are not straight-forward to predict, since in a highly recurrent network neuronal output is simultaneously also input to the same network. Sometimes this leads to non-intuitive effects like effective inhibitory influence of excitatory neuron populations (Tsodyks et al., 1997). To master such difficulties in the analysis, a system-level view of complex microcircuits is inevitable (Yuste, 2015; Bassett and Sporns, 2017).

Baseline and modulation pathways. Previously, inhibition dominated recurrent networks were shown to exhibit two distinct processing pathways (Sadeh et al., 2014). While the untuned and uninformative component of the input is strongly attenuated by the recurrent network, the modulated component, which conveys all the information about the stimulus, has a higher gain. This results in a net amplification of feature selectivity through the input-output transformation exerted by the network.

By separating the (linearized) layered model into two systems, we were able to analytically calculate the feed-forward and recurrent gains of the baseline pathway also for a multi-population system. In combination, these two gains describe how input to one population affects the output of any other population. In contrast to our model, neurons in the model of Potjans and Diesmann (2014) have binomially distributed in-degrees, lead-

ing to some degree of cross-talk between the two pathways by non-vanishing interference terms in Equation (12) . While this introduces some discrepancies between the direct and separate solutions of the linear model, it does not impair the described mechanism.

It should be emphasized that the separation between baseline and modulation pathways, its different gains, and thus also tuning amplification are purely linear effects. This is possible as a linear transformation may impose different scaling factors to different stimulus components.

Mode decomposition. Instead of analyzing the operation of a network from the viewpoint of neuronal populations, we chose a basis that is more natural for the network in question. We decomposed the input into eigenmodes of the system, where the eigenvalues of the effective connectivity correspond to the gains of the individual input modes. Interestingly, the high gain of L5e can be attributed to a single mode, which has a much higher gain than all the other modes. Due to this mode, L5e neurons exhibit a high firing rate and weak selectivity for stimulus orientation.

What is the functional significance of having a mode with this exceptionally high gain? For a preliminary aswer, it is important to realize that the different populations in the layered cortical network play different roles for information processing. L2/3e neurons project to higher brain areas and need to transmit information in a reliable way (Felleman and Van, 1991). Therefore, it is desirable that this population has a strong attenuation of the baseline and, thus, strong tuning amplification. L5e neurons, on the other hand, mainly project to sub-cortical brain regions, including thalamus (Sherman, 2016). It is likely that this population is involved in feedback control mechanisms, where tuning information is not as relevant.

Predictions for optogenetic intervention. Using the theory developed in this work, we are in the position to devise external (e.g. optogenetically applied) stimuli which effectively manipulate the orientation tuning of L5e neurons. We designed two different stimuli which reduce the mean firing rates of these neurons, thus increasing their tuning. Interestingly, although the effect on the rate of L5e is identical, one stimulus requires a positive current, while the other utilizes a negative current. This can be explained by the effect of the stimulus on other populations, which is propagated via the recurrent network.

These calculations constitute strong predictions of our model, which can be tested with some effort in experiments. Different subpopulations can be stimulated by a combination of channelrhodopsin and halorhodopsin with different wavelength sensitivities, as described by Klapoetke et al. (2014), possibly in combination with locally confined optical stimulation.

Throughout our study, and in particular for the derived predictions, the consistency of results between the spiking network and the linear model are remarkable. Only for very strong stimulation, when non-linear rectification sets in, some deviations become evident. This confirms previous findings that a network of highly nonlinear neurons can be effectively linearized about its dynamical operating point, which represents the balanced state of the excitatory-inhibitory network (Van Vreeswijk and Sompolinsky, 1996, 1998).

In our work, the newly developed mathematical tools were applied to design current stimuli which effectively and selectively suppress L5e activity. However, because in our model all rate gains are known from the decomposition into baseline and modulation pathways, other types of experiments can be performed as well, even for complex multi-

population networks different from the primary visual cortex. This makes the suggested method an ideal tool for the design of external current stimuli in a large variety of experiments.

In summary, we presented a novel method for analyzing the dynamics of multi-population networks in terms of their input-output relations. We applied the approach to the analysis of a layered model of rodent visual cortex, revealing interesting and to some degree also unexpected and non-intuitive consequences of the underlying connectivity. In addition, we derived predictions for future optogenetic experiments, which could be performed to test the power of our computational analysis.

References

Aljadeff, J., Renfrew, D., Vegué, M., and Sharpee, T. O. (2016). Low-dimensional dynamics of structured random networks. *Physical Review E*, 93(2):022302.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252.

Bassett, D. S. and Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20(3):353.

Batschelet, E. (1981). Circular statistics in biology. *ACADEMIC PRESS, 111 FIFTH AVE., NEW YORK, NY 10003, 1981, 388*.

Binzegger, T., Douglas, R. J., and Martin, K. A. (2004). A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience*, 24(39):8441–8453.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8(3):183–208.

Durand, S., Iyer, R., Mizuseki, K., de Vries, S., Mihalas, S., and Reid, R. C. (2016). A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *Journal of Neuroscience*, 36(48):12144–12156.

Felleman, D. J. and Van, D. E. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47.

Gewaltig, M.-O. and Diesmann, M. (2007). Nest (neural simulation tool). *Scholarpedia*, 2(4):1430.

Hairer, E., Nørsett, S. P., and Wanner, G. (1993). Solving ordinary differential equations. i, springer series in computational mathematics.

Hansel, D. and van Vreeswijk, C. (2012). The mechanism of orientation selectivity in primary visual cortex without a functional map. *Journal of Neuroscience*, 32(12):4049–4064.

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., and Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, 14(8):1045.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.

Jia, H., Rochefort, N. L., Chen, X., and Konnerth, A. (2010). Dendritic organization of sensory input to cortical neurons in vivo. *Nature*, 464(7293):1307.

Klapoetke, N. C., Murata, Y., Kim, S. S., Pulver, S. R., Birdsey-Benson, A., Cho, Y. K., Morimoto, T. K., Chuong, A. S., Carpenter, E. J., Tian, Z., et al. (2014). Independent optical excitation of distinct neural populations. *Nature Methods*, 11(3):338.

Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., and Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96.

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., and Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87.

Marshel, J. H., Kaye, A. P., Nauhaus, I., and Callaway, E. M. (2012). Anterior-posterior direction opponency in the superficial mouse lateral geniculate nucleus. *Neuron*, 76(4):713–720.

Niell, C. M. and Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience*, 28(30):7520–7536.

Niell, C. M. and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472–479.

Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., and Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433(7026):597.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience*, 17(4):594.

Pakan, J. M., Lowe, S. C., Dylda, E., Keemink, S. W., Currie, S. P., Coutts, C. A., and Rochefort, N. L. (2016). Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *Elife*, 5.

Piscopo, D. M., El-Danaf, R. N., Huberman, A. D., and Niell, C. M. (2013). Diverse visual features encoded in mouse lateral geniculate nucleus. *Journal of Neuroscience*, 33(11):4642–4656.

Potjans, T. C. and Diesmann, M. (2014). The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. *Cerebral Cortex*, 24:785–806.

Rajan, K. and Abbott, L. (2006). Eigenvalue spectra of random matrices for neural networks. *Physical Review Letters*, 97(18):188104.

Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.

Sadeh, S., Cardanobile, S., and Rotter, S. (2014). Mean-field analysis of orientation selectivity in inhibition-dominated networks of spiking neurons. *SpringerPlus*, 3(1):148.

Sadeh, S., Clopath, C., and Rotter, S. (2015). Processing of feature selectivity in cortical networks with specific connectivity. *PloS one*, 10(6):e0127547.

Sadeh, S. and Rotter, S. (2014). Distribution of orientation selectivity in recurrent networks of spiking neurons with different random topologies. *PloS one*, 9(12):e114237.

Sadeh, S. and Rotter, S. (2015). Orientation selectivity in inhibition-dominated networks of spiking neurons: effect of single neuron properties and network dynamics. *PLoS Computational Biology*, 11(1):e1004045.

Scholl, B., Pattadkal, J. J., Dilly, G. A., Priebe, N. J., and Zemelman, B. V. (2015). Local integration accounts for weak selectivity of mouse neocortical parvalbumin interneurons. *Neuron*, 87(2):424–436.

Scholl, B., Tan, A. Y., Corey, J., and Priebe, N. J. (2013). Emergence of orientation selectivity in the mammalian visual pathway. *Journal of Neuroscience*, 33(26):10616–10624.

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature Neuroscience*, 19(4):533.

Siegert, A. J. (1951). On the first passage time probability problem. *Physical Review*, 81(4):617.

Sun, W., Tan, Z., Mensh, B. D., and Ji, N. (2016). Thalamus provides layer 4 of primary visual cortex with orientation-and direction-tuned inputs. *Nature Neuroscience*, 19(2):308.

Thomson, A. M., West, D. C., Wang, Y., and Bannister, A. P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cerebral Cortex*, 12(9):936–953.

Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., and McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *Journal of Neuroscience*, 17(11):4382–4388.

Van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.

Van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10(6):1321–1371.

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

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, 16(8):487.