Effects of optogenetic manipulations of cortical circuits in vivo

Alexandre Mahrach

To cite this version:

Alexandre Mahrach. Effects of optogenetic manipulations of cortical circuits in vivo. Neurons and Cognition [q-bio.NC]. Sorbonne Université, 2020. English. NNT : 2020SORUS153 . tel-03412742v1

HAL Id: tel-03412742
https://theses.hal.science/tel-03412742v1

Submitted on 3 Nov 2021 (v1), last revised 3 Nov 2021 (v2)

Effects of optogenetic manipulations of cortical circuits *in vivo*

Alexandre Mahrach
Sorbonne Université

A thesis submitted for the degree of
Doctor of Philosophy
Paris, March 2019

Committee

Sophie Bernard (President)	Université de Paris, France
David Golomb (Examiner)	Ben Gurion University, Israel
Thomas Boraud (Examiner)	Université de Bordeaux, France
Nicholas Priebe	University of Austin, USA
Carl van Vreeswijk	Université de Paris, France
David Hansel	Université de Paris, France

Abstract

Optogenetics is a common strategy used to manipulate the activity of specific cortical neurons *in vivo*. Different variants (ChR2, ReaChR, etc.) have been developed that allow for extended as well as very localized perturbations of cortical network activity. This work is devoted to describing the mechanisms underlying the response of mouse cortical networks to two variants of ChR-assisted photostimulation of the parvalbumin-positive interneurons (PV) - the largest subtype of inhibitory neurons in the mouse cortex. In the first part, we investigate the effects of the ReaChR optogenetic photostimulation of PV neurons in the anterolateral motor cortex (ALM) and in the barrel cortex (S1) on the population average responses of principal excitatory cells (PCs) and PV interneurons. ReaChR allows for a uniform stimulation of the neurons upon broadly extended regions (up to a few mm). After extending the theory of balanced networks to networks with multiple inhibitory populations, we investigate the connectivity patterns that are essential in accounting for these effects. We show that the average responses of PCs and PV neurons in ALM layer 2/3 and layer 5 can be accounted for by a four-population network with connectivity similar to that reported in the primary visual cortex. We propose an alternative network architecture to robustly account for these responses in S1. In the second part, we examine the effects of the ChR2 optogenetic stimulation of PV neurons in the mouse barrel cortex on the spatial profile of the responses of PCs and PV neurons. ChR2 allows for stimulation of the neurons upon very localized regions (a few hundred microns). We investigate these effects in two-population (one excitatory and one inhibitory) network models where the probability of connection between the neurons decay exponentially with distance. We show that for stimulation of strong enough intensity, the activity of the excitatory neurons is suppressed on a characteristic length that non-linearly depends on the radius of the perturbation. We derive conditions upon the length of the recurrent neuronal interactions to explain the profile of the excitatory and inhibitory responses in the surround of this suppression.

Table of contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

Optogenetic tools offer the ability to target and manipulate the activity of specific neuron subtypes *in vivo*. The technique relies on the expression into the neurons of a gene for a light-sensitive opsin (typically an ion channel or a pump). Upon light illumination ion channels open (resp. ion pumps activate) leading to a depolarization (resp. hyperpolarization) of the membrane voltage of the genetically modified neurons, and thus to an activity change.

Optogenetics has furthered the understanding of how different neurons contribute to network dynamics, cortical computations, cognitive processes, and behavior. Nevertheless, optogenetics has some drawbacks. For example, illumination with a defined intensity has variable effects on individual cells since not all the neurons express the markers at the same level. Moreover, stimulation spreads in brain tissues on a spatially extended region and variably affects neurons depending on their distance from the light source. As a consequence, the optogenetic perturbation of selective neuron subtypes can lead to important changes in the network dynamics and open the field of numerous counter-intuitive observations. Only a precise mathematical analysis can be used to separate the wheat from the chaff.

The major goal of this thesis is to evaluate the consequences of optogenetic perturbations of specific cortical populations of neurons *in vivo*, on how they shape neuronal dynamics and can provide insights on the operating regime of the cortex during spontaneous activity. In the first part of the thesis, I will describe the properties of cortical inhibitory neurons and their role in the dynamics of recurrent circuits. I will review the essential statistical properties of cortical network activity and describe the competing theoretical frameworks that provide a natural explanation for them. I will discuss recent experimental observations obtained from ChR2 optogenetic perturbations of the cortex and review the existing theories that give an account for the experimental results.

In the second part of the thesis, I present the results of a study conducted in collaboration with the experimental group of Dr. Nuo Li (Baylor College of Medicine, Texas). It deals with the mechanisms underlying the response of the cortex to the optogenetic manipulations of the parvalbumin-expressing inhibitory neurons (PV) in the anterior lateral motor cortex (ALM) and the barrel cortex (S1) of the mouse. In this study, we argue for the framework of recurrent networks with strong synapses and sparse connections. We investigate network architectures that can account for the experimental observations. We demonstrate that standard two-population models cannot explain our data either in ALM layer 2/3 and layer 5 nor in S1. In contrast, we described how the average response of PCs and PV neurons in ALM layer 2/3 and layer 5 can be accounted for by a four-population network with connectivity similar to that reported for V1. We proposed an alternative network architecture to account for the proportional decrease in the PC and PV population average activities observed in S1.

The third part deals with understanding the spatially extended responses of cortical networks to optogenetic perturbations of PV neurons *in vivo*. It provides an explanation of the mechanisms underlying the observations of (Li et al., 2019) in the mouse barrel cortex. We investigate the spatial response of two-population networks of neurons to photostimulation of their inhibitory population with a Gaussian-shaped perturbation when the probability of connection between neurons falls off exponentially with distance. We show that upon strong perturbation the balance of excitation and inhibition is disrupted on a characteristic length where excitatory neurons are silenced and derive its dependence with the intensity and radius of the perturbation.

In the last chapter, I will discuss my results and their implications and give perspectives and insights on future work to be conducted.

Inhibition in the mouse cortex

Recurrent circuits of excitatory and inhibitory neurons

One cubic millimeter of the mouse cortex comprises around 10^5 neurons recurrently connected through nearly 10^9 synapses. These neurons consist of glutamatergic excitatory principal neurons (PCs) and of local projecting GABAergic inhibitory interneurons that together form highly complex local circuits (Braitenberg and Schüz, 2013; Schüz and Palm, 1989). Although they only represent 10 to 15 $\%$ of all cortical neurons (Meyer et al., 2011), GABAergic interneurons have been reported to be highly heterogeneous in the cortex. Cortical computations (such as information processing of sensory inputs, working memory, and long term memory) are thought to rely on the dynamics of these networks.

In the previous decade, huge efforts have been dedicated to experimentally quantify the cortical connectivity graph and theoretically motivated constraints on the wiring diagram have been proposed. Electrophysiological recordings have revealed that the probability of connection between neurons is small, favoring the idea that cortical connectivity is sparse (Holmgren et al., 2003; Markram et al., 1997; Thomson et al., 2002; Thomson and Bannister, 2003). In addition, the probability of connection between neurons has been reported to fall off with anatomical distance on a range of a few $100 \ \mu m$ (Hellwig, 2000; Holmgren et al., 2003; Lee et al., 2012; Perin et al., 2011). Long-range projections ($500\ \mu m$ to $1\ mm$) between distinct neuronal subtypes are also known to exist (Stepanyants et al., 2009).

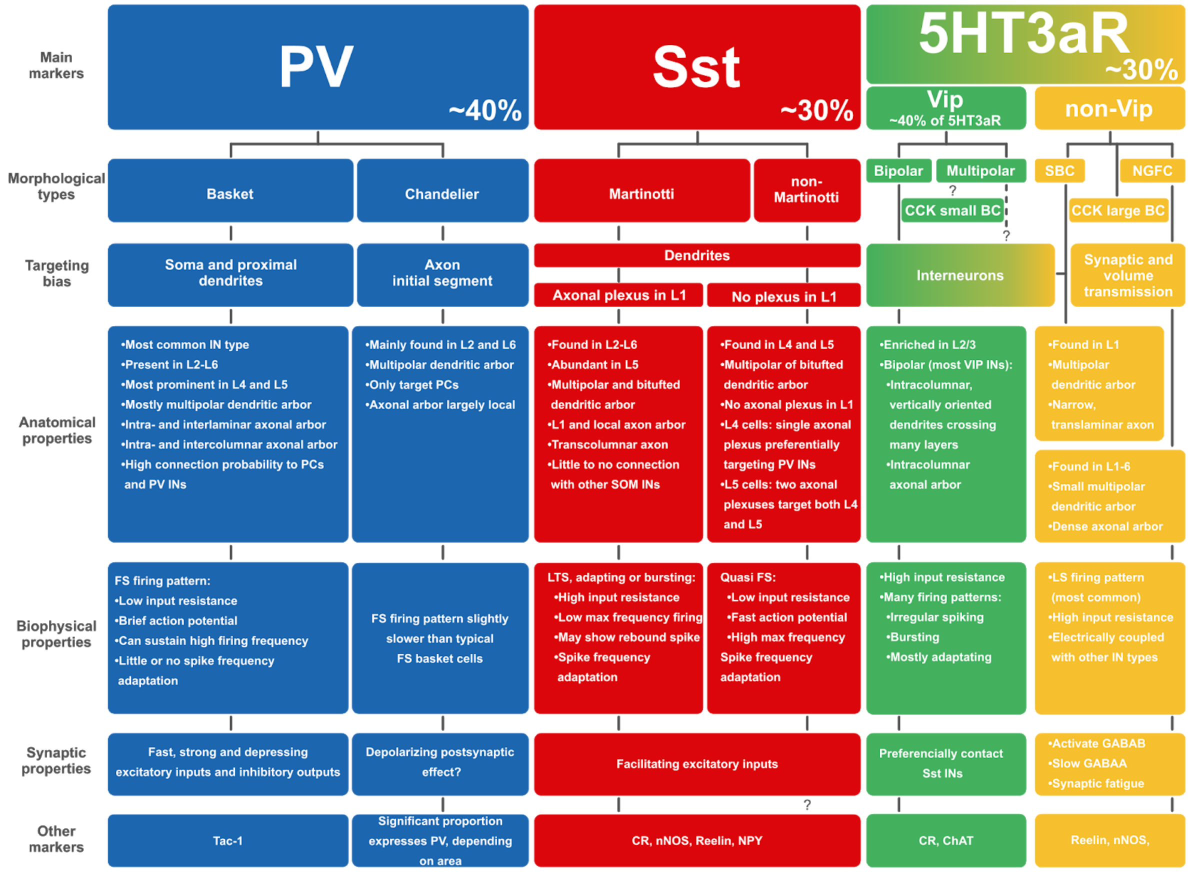
The presence of specific pre and postsynaptic targets between different inhibitory subtypes is well documented (Beierlein et al., 2003; Gibson et al., 1999; Jiang et al., 2015; Pfeffer et al., 2013). In addition, input and output synaptic efficacy, kinetics, and dynamics are subtype dependent (Beierlein et al., 2003; Gupta et al., 2000). Interneurons preferentially target distinct subcellular domains, with inhibitory neurons almost exclusively targeting either dendrites or somas (Klausberger, 2009). Moreover, interneurons exhibit a large diversity of intrinsic electrophysiological properties (*e.g.* firing patterns, synaptic dynamics, etc.) which might potentially be vital in sculpting the network dynamics.

With advances in molecular genetics, a large number of studies have focused on labeling and manipulating specific inhibitory subtypes. Experimental evidence has highlighted the specific role of distinct inhibitory subtypes in cortical functions such as gain control (Fu et al., 2014; Isaacson and Scanziani, 2011; Katzner et al., 2011), sensory feature selectivity (Kaneko et al., 2017; Wood et al., 2017), synaptic plasticity (Fu et al., 2015; Hayut et al., 2011), temporal coding, and generation of cortical rhythms (Cardin et al., 2009; Sohal et al., 2009; Stark et al., 2014, 2013).

Classification of inhibitory subtypes

The general classification of interneurons was first based on their morphological features (Klausberger and Somogyi, 2008; Somogyi and Klausberger, 2005). Nevertheless, it has been found to be inconclusive to unambiguously segregate between inhibitory subtypes (DeFelipe et al., 2013). In contrast, with the recent advances in molecular genetics, the generation of transgenic rodents expressing fluorescent proteins or enzymes (*e.g.* Cre recombinase) has allowed for a more precise categorization (Taniguchi et al., 2011). Morphological and electrophysiological analysis of the interneurons expressing distinct molecular markers as well as segregation based on developmental origin and connectivity - distinct inhibitory classes are known to specifically project to different neuronal targets - have permitted a better description of the different inhibitory populations in the cortex. Interneurons can be broadly subdivided into three molecular classes: calcium-binding protein parvalbumin-positive (PV), neuropeptides somatostatin positive (SOM) and serotonin receptor (5HT3aR) expressing neurons (Rudy et al., 2011) (Fig. 1). These three molecular markers are expressed in largely non-overlapping cell types in the cortex (Lee et al., 2013). Moreover, they have been found to account for nearly all inhibitory neurons in the primary somatosensory cortex (S1).

Neurons within each of these populations show quasi-exclusive intrinsic properties (*e.g.* morphological, electrophysiological, in vivo activity patterns, etc.) as well as gene expression patterns (Tasic et al., 2016; Zeisel et al., 2015). Within the same molecular population, differences can still be found. Each molecular subtype can be further divided into smaller ensembles of neurons (in total, around 30 subtypes have been reported, see Fig. 1) (Tremblay et al., 2016).



Main markers	PV ~40%		Sst ~30%		5HT3aR ~30%	
					Vip ~40% of 5HT3aR	non-Vip
Morphological types	Basket	Chandelier	Martinotti	non-Martinotti	Bipolar Multipolar ? CCK small BC ?	SBC NGFC CCK large BC
Targeting bias	Soma and proximal dendrites	Axon initial segment	Dendrites Axonal plexus in L1	No plexus in L1	Interneurons	Synaptic and volume transmission
Anatomical properties	•Most common IN type •Present in L2-L6 •Most prominent in L4 and L5 •Mostly multipolar dendritic arbor •Intra- and interlaminar axonal arbor •Intra- and intercolumnar axonal arbor •High connection probability to PCs and PV INs	•Mainly found in L2 and L6 •Multipolar dendritic arbor •Only target PCs •Axonal arbor largely local	•Found in L2-L6 •Abundant in L5 •Multipolar and bitufted dendritic arbor •L1 and local axon arbor •Transcolumnar axon •Little to no connection with other SOM INs	•Found in L4 and L5 •Multipolar of bitufted dendritic arbor •No axonal plexus in L1 •L4 cells: single axonal plexus preferentially targeting PV INs •L5 cells: two axonal plexuses target both L4 and L5	•Enriched in L2/3 •Bipolar (most VIP INs): •Intracolumnar, vertically oriented dendrites crossing many layers •Intracolumnar axonal arbor	•Found in L1 •Multipolar dendritic arbor •Narrow, translaminar axon •Found in L1-6 •Small multipolar dendritic arbor •Dense axonal arbor
Biophysical properties	FS firing pattern: •Low input resistance •Brief action potential •Can sustain high firing frequency •Little or no spike frequency adaptation	FS firing pattern slightly slower than typical FS basket cells	LTS, adapting or bursting: •High input resistance •Low max frequency firing •May show rebound spike •Spike frequency adaptation	Quasi FS: •Low input resistance •Fast action potential •High max frequency Spike frequency adaptation	•High input resistance •Many firing patterns: •Irregular spiking •Bursting •Mostly adaptating	•LS firing pattern (most common) •High input resistance •Electrically coupled with other IN types
Synaptic properties	Fast, strong and depressing excitatory inputs and inhibitory outputs	Depolarizing postsynaptic effect?	Facilitating excitatory inputs		Preferencially contact Sst INs	•Activate GABAB •Slow GABAA •Synaptic fatigue
Other markers	Tac-1	Significant proportion expresses PV, depending on area	CR, nNOS, Reelin, NPY	?	CR, ChAT	Reelin, nNOS,

Figure 1. Cortical interneurons diversity. Inhibitory neurons in the cortex express one of the main three non-overlapping markers: parvalbumin (PV, blue), somatostatin (SOM, red), and the ionotropic serotonin receptor 5HT3a (5HT3aR, green-yellow). Further subdivisions within each molecular group are revealed by morphological features, cellular and subcellular targeting biases, and expression of other markers, as well as some known anatomical, electrophysiological, and synaptic properties. Figure and adapted caption from (Tremblay et al., 2016).

The molecular classification of cortical inhibitory neurons mostly relies on experiments conducted in the mouse barrel cortex. However, studies comparing various cortical areas have shown that the distributions of neurons expressing the PV, SOM and 5HT3aR markers are preserved across areas (Xu et al., 2010). Moreover, the same PV, SOM and 5HT3aR expressing neurons have been described in the cortex of different mammalian species with additional interneurons subtypes discovered in higher mammals (DeFelipe et al., 2002). In mice, differences in connectivity patterns have been reported between areas (Kätzel et al., 2011; Packer et al., 2013). In addition, the density of each molecular population has been reported to vary across cortical layers (Rudy et al., 2011; Tremblay et al., 2016). Thus, cortical circuits in different areas and/or layers might exhibit very different architectures.

Inhibitory circuit motifs in mouse cortex

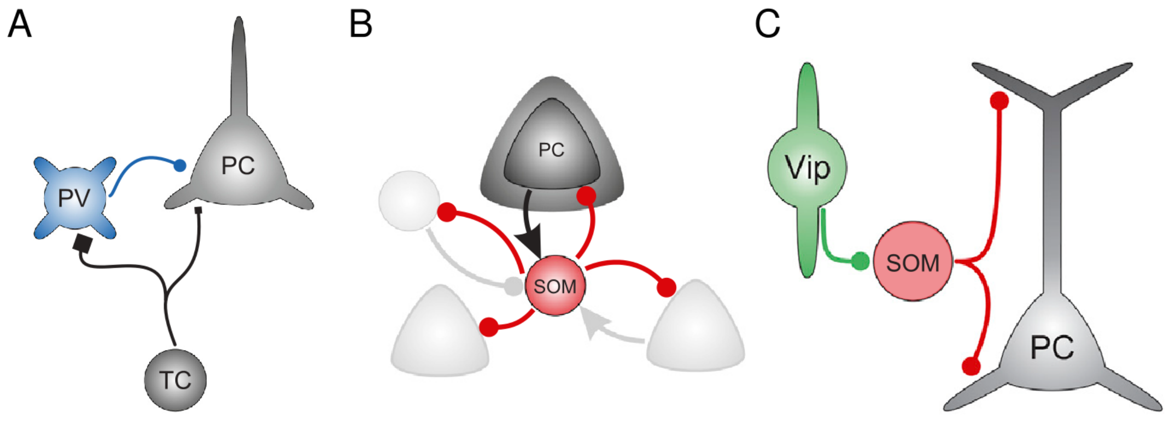


Figure 2. Circuit motifs in the mouse cortex. **A**. Thalamocortical neurons synapse onto both excitatory PCs and PV neurons allowing PV neurons to provide feedforward inhibition to the PCs. **B**. SOM interneurons provide feedback inhibition to neighboring cells. C. VIP interneurons disinhibit the PCs through the inhibition of SOM neurons. Figure adapted from (Tremblay et al., 2016).

In recent years, numerous experimental studies have investigated the rules of connectivity between inhibitory subtypes. It is thought that interneuron connectivity exhibit a few core motifs: feedforward inhibition, feedback inhibition, and disinhibition (Fig. 2).

Feedforward inhibition describes the process by which an external excitatory input drives an inhibitory subtype to provide a direct source of inhibition to PCs. It is thought to be mostly mediated by PV neurons in the cortex.

Feedback inhibition, in contrast, describes the local inhibitory input into a PC from the inhibitory neurons it projects to. It is thought of as a mechanism preventing runaway excitation. It is likely that most inhibitory subtypes participate in local feedback loops. In particular, SOM neurons lateral inhibition has been proposed as a mechanism for surround suppression in the visual cortex (Adesnik et al., 2012). Given that SOM neurons connect to both PCs and PV neurons, but do not interact with each other (Adesnik et al., 2012; Gibson et al., 1999; Hu et al., 2011), increasing the size of a visual stimulus leads to a decrease in the response of PCs and PV neurons due to an increase in the inhibition provided by the SOM population. Inhibitory neurons not only provide direct inhibition to the PCs, but they can also disinhibit the latter when inhibiting other interneurons (Freund et al., 1983; Somogyi et al., 1983). The intricate nature of the connectivity between PCs and inhibitory neurons might lead to loops in the network. Therefore, inhibitory to inhibitory connections are not always disinhibitory. It is commonly acknowledged that the VIP to SOM interaction is disinhibitory in the superficial layers of the mouse cortex. Indeed, according to recent electrophysiological studies in slice VIP neurons preferentially target SOM neurons in the layer 2/3 of primary somatosensory cortex, primary visual cortex, primary auditory cortex, and also in the prefrontal cortex (Lee et al., 2013; Pfeffer et al., 2013; Pi et al., 2013). Therefore, since SOM neurons target the PCs, VIP neurons disinhibit the PCs.

Neuron firing statistics in cortex *in vivo*

Spike train temporal irregularity

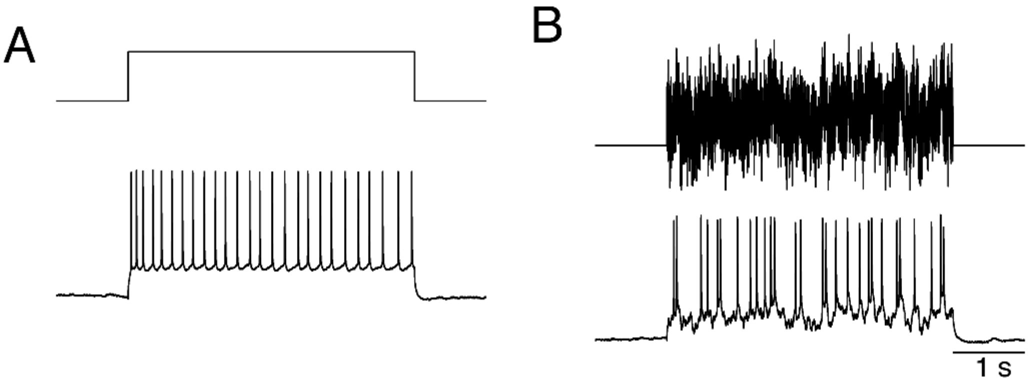


Figure 3. Irregular v.s. regular neuronal firing in cultures of neocortical neurons from the somatosensory/motor cortex. **A**. Response to conventional DC stimulation. **B**. Response evoked by a non-deterministic input current mimicking a realistic input drive of cortical networks. Figure adapted from (Giugliano et al., 2004).

Neuronal spike trains recorded in vivo exhibit significant temporal irregularity (Bair et al., 1994; Softky and Koch, 1993; Tomko and Crapper, 1974). During spontaneous as well as during stimulus-evoked activity, interspike interval distributions are right-skewed and long-tailed and resemble those generated by a Poisson process (Shinomoto et al., 2009). The irregularity of the *in vivo* neuronal activity also appears in intracellular recordings as large fluctuations in the voltage traces of the membrane potential (Giugliano et al., 2004) (Fig. 3B). In contrast in activity patterns generated *in vitro* (*e.g.* injecting neurons with constant currents Fig. 3A), neuronal discharges are highly regular and voltage fluctuations are practically nonexistent. The irregularity of the neuronal dynamics *in vivo* is likely to be independent of the intrinsic properties of the neurons. It might be due to strongly fluctuating synaptic inputs into the neurons. However, in view of a large number of nearly uncorrelated synaptic afferents to a neuron, fluctuations in its net input should average out. What can be the origin of these large temporal fluctuations?

Firing rate heterogeneity in the cortex

The mean spontaneous and evoked firing rates of individual neurons are very heterogeneous *in vivo* (Buzsáki and Mizuseki, 2014; Compte et al., 2003; Hromádka et al., 2008; Mizuseki and Buzsáki, 2013; O'Connor et al., 2010). The distributions of both spontaneous and evoked single-cell mean activities are strongly skewed, typically presenting a log-normal shape. Such distributions allow for a wide range of firing rates with a large number of low firing neurons and a small fraction of high rate cells. Skewed distributions are a ubiquitous feature across species and appear to be a general feature of neuronal dynamics.

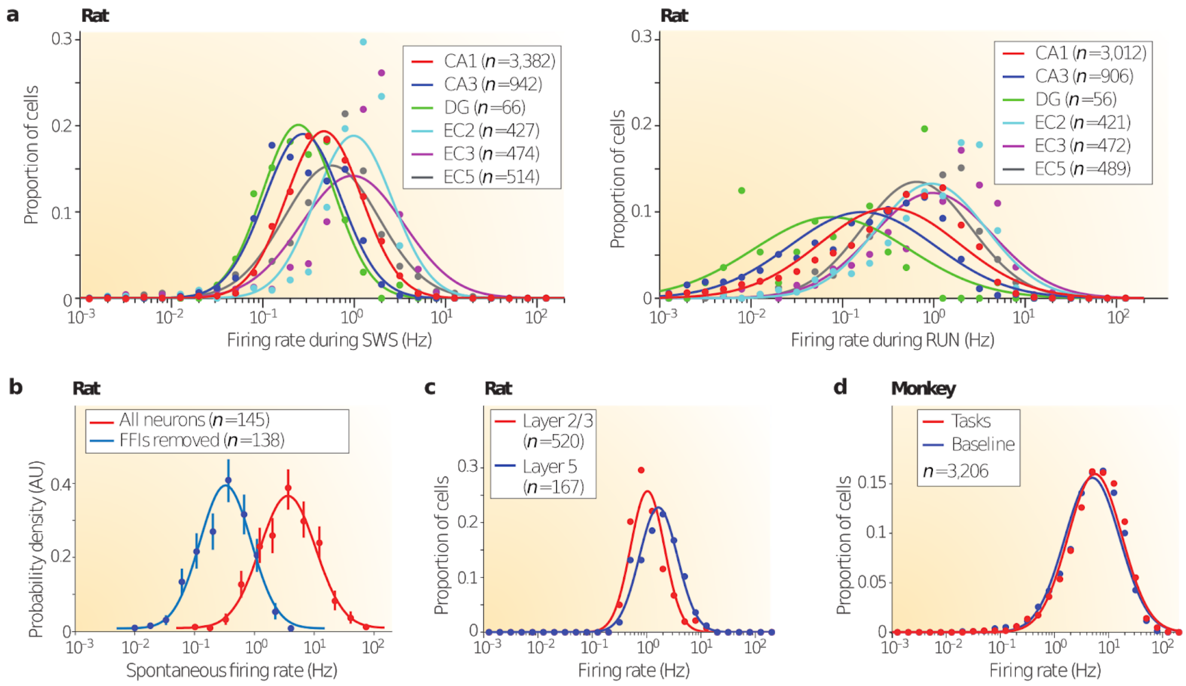


Figure 4. Log-normal distribution of firing rates in the cortex. **A**. Firing-rate distribution of principal cells in the rat hippocampus (CA1, CA3 and dentate gyrus (DG)) and the entorhinal cortex (EC; specifically, in layers 2, 3 and 5) during slow-wave sleep (SWS; left panel) and exploration (RUN; right panel). **B**. Firing-rate distribution of neurons in the auditory cortex of awake rats. **C**. Firing-rate distribution of superficial (layers 2/3) and layer 5 neurons in the prefrontal cortex of an exploring rat. **D**. Firing-rate distribution of neurons from lateral intraparietal and parietal reach region areas of the macaque cortex during a baseline condition and during the performance of a reaching task. Figure and caption adapted from (Buzsáki and Mizuseki, 2014).

Models of cortical dynamics

Models of cortical circuits usually consider two-population networks of randomly connected excitatory (E) and inhibitory (I) neurons. In most of the cases, the average number of presynaptic inputs, K_A, from neurons in population A is assumed to be sufficiently large but small in comparison with the number of neurons in this population, N_A. That is, the connectivity matrix, C_{ij}^{AB}, of interaction between neurons is sparse. Here we will consider that the connection from a neuron (B,j) in population B and a neuron (A,i) in population A is 1 with probability $\frac{K_B}{N_B}$ and 0 otherwise. There is a great variety of model neurons ranging from simple binary units (Van Vreeswijk and Sompolinsky, 2005; van Vreeswijk and Sompolinsky, 1998, 1996), rate models (Harish and Hansel, 2015; Ozeki et al., 2009), integrate and fire models (Hansel and Mato, 2013; Mongillo et al., 2012; Roxin et al., 2011) to complex conductance-based neurons (Hansel and van Vreeswijk, 2012; Pattadkal et al., 2018; Rao et al., 2019). For the purpose of this section, we will consider a simple rate model and assume here that neurons are characterized by a continuous variable, $r_i^A(t)$, commonly interpreted as their temporal activity. $r_i^A(t)$ non-linearly depends on the inputs, $u_i^A(t)$, into the neuron (A,i) through the neuronal transfer function Φ. In addition, we will assume that neurons receive feedforward external inputs from outside the network and that these are drawn from a Gaussian with mean $j_{A0}\, r_0$, total variance α_{A0} and quench variance β_{A0}.

Provided that K_A is sufficiently large, all $u_i^A(t)$ are stochastically independent and have a common Gaussian distribution,

$$u_i^A(t) = u_A + \sqrt{\beta_A}\,\xi_i + \sqrt{\alpha_A-\beta_A}\,\zeta_i(t) \qquad (1)$$

with mean

$$u_A =< [u_i^A(t)] >= j_{A0} + \sum_{B,j} j_{AB}\, [C_{ij}^{AB}]\ < r_j^B(t) > \qquad (2)$$

a total variance given by

$$\alpha_A =< [(\delta u_i^A(t))^2] >= \alpha_0 + \sum_{B,j} j_{AB}\ < [(\delta(C_{ij}^{AB}\ r_j^B(t)))^2] > \qquad (3)$$

and a quench variance due to the variability in the number of afferent connections per neuron

$$\beta_A = [< u_i^A(t)>^2] - u_A^2 \qquad (4)$$

Here $[.]$ denotes the population average, $<.>$ the temporal average and $\delta X = X - [X]$.

The activity of a neuron is then determined by

$$\frac{d\, r_i^A}{dt}(t) = -r_i^A(t) + \Phi(u_A + \sqrt{\beta_A}\,\xi_i + \sqrt{\alpha_A-\beta_A}\,\zeta_i(t)) \qquad (5)$$

where ξ_i and $\zeta_i(t)$ are Gaussian with zero mean and variance one and $< \zeta_i(t) >= 0$.

Weak interactions: Inhibitory stabilization

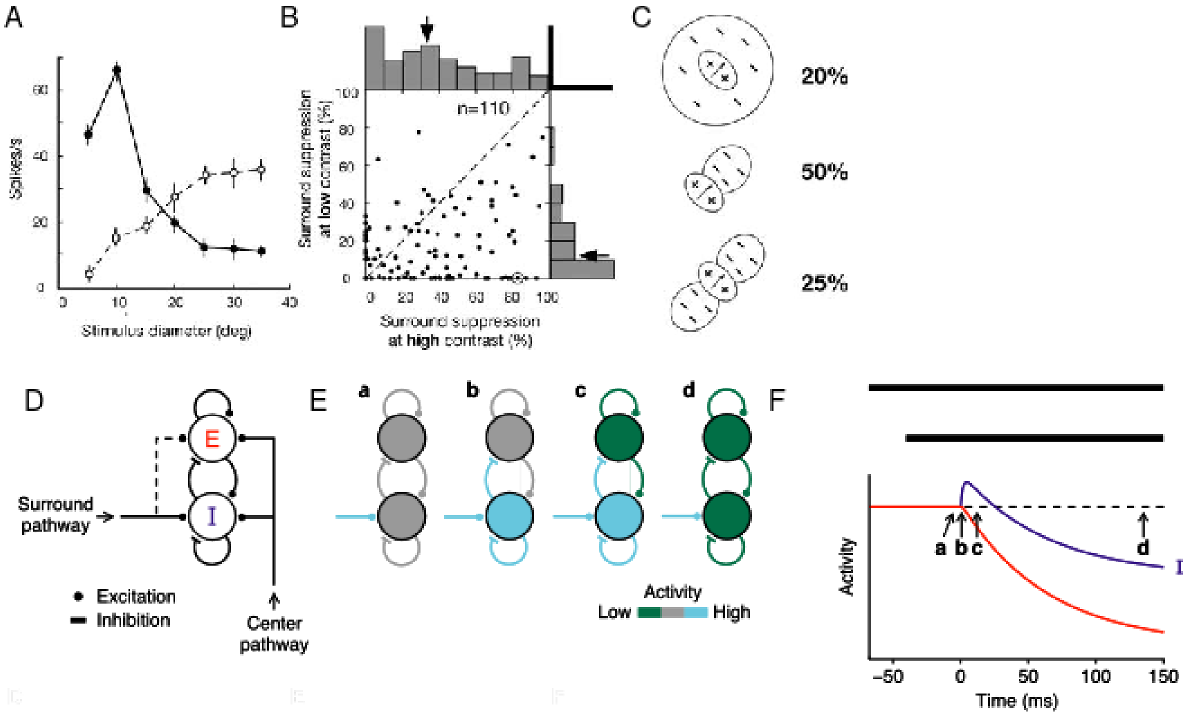


Figure 5. The mechanism of surround suppression. **A**. Effect of contrast on center-surround interactions for one MT neuron. The neuron responded optimally to a stimulus with 10 deg in diameter and was strongly suppressed by larger patterns. **B**. Strength of surround suppression measured at both high and low contrast. **A-B** Adapted from (Pack et al., 2005). **C**. Asymmetries in the spatial organization of the suppressive surround (after Xiao et al. 1997). **D**. A two populations network model. with feedforward input driven by the receptive field center and lateral excitatory input driven by the receptive field surround (I only). **E**. Events following surround stimulation after pre-existing center stimulus: after a transient increase in the activity of the inhibitory neurons, activity in both the E and I cells decreases (d) relative to the activity evoked by center stimulus alone (a). **F**. Temporal dynamics of the changes in the activity of E and I neurons. **D-E-F** adapted from (Ozeki et al., 2009).

When the strength of the synaptic interactions between neurons, j_{AB}, scale as $\frac{1}{K_B}$, *i.e.* $j_{AB} = \frac{J_{AB}}{K_B}$, the activity of the network is determined by a set of coupled nonlinear equations which depends on the neuronal transfer function,

$$r_i^A = \int D\zeta\, \Phi\left(u_A + \sqrt{\beta_{A0}}\,\xi_i + \sqrt{\alpha_{A0}-\beta_{A0}}\,\zeta\right) + O\left(\tfrac{1}{K_A}\right) \qquad (6)$$

Excitatory and inhibitory inputs into a neuron are both comparable to its rheobase even when K_A is large, but the spatial and temporal fluctuations of the neuron net recurrent input vanish as K_A increases. For large but finite K_A, the firing irregularity and heterogeneity essentially arise from the fixed irregularity and heterogeneity in the feedforward inputs into the neurons.

Inhibition stabilized networks

Numerous studies focus on a particular set of network models with weak interactions where: 1) the strength of the interactions between excitatory neurons, J_{EE} , is so strong so that the network steady-state is unstable in the absence of feedback inhibition; 2) The strength of the inhibitory recurrent feedback, J_{EI} , is sufficiently strong to stabilize the network. These networks are referred to as inhibition-stabilized networks (ISN) (Ozeki et al., 2009).

A key prediction of ISN models is that the average inhibitory activity decreases when inhibitory neurons are stimulated (Ozeki et al., 2009; Tsodyks et al., 1997). (Ozeki et al., 2009) have used this counter-intuitive “paradoxical” effect to account for the phenomenon of surround suppression in the primary visual cortex (Fig. 5). Surround suppression is the mechanism by which the relative activity of a neuron decreases when a visual stimulus is enlarged to the region surrounding its receptive field. Neurons from distinct cortical areas exhibit surround suppression (Anderson et al., 2001; Nienborg et al., 2013; Ozeki et al., 2009; Sengpiel et al., 1997; Vanni and Casanova, 2013; Wang et al., 2009). This suppression was thought to be linked to an increase in the lateral inhibitory inputs into the PCs. Instead, (Ozeki et al., 2009) have shown that surround stimulus decreases inhibition into the PCs and that suppression emerges as a property of the network recurrent dynamics.

Supralinear Stabilized networks

Recently, (Ahmadian et al., 2013) have extended ISN models toward the study of models where the neuronal transfer function present an expansive non-linearity. In their models, referred to as stabilized supralinear networks (SSNs), the response of the neurons to weak feedforward inputs is supralinear and it becomes sublinear and non-monotonic in response to strong feedforward input. (Rubin et al., 2015) have shown that SSN can account for a large ensemble of phenomena observed in the primary visual cortex.

In particular, SSN models provide a better account of the mechanisms underlying surround suppression than ISN models. It explains that the responses of the neurons in the surrounding regions are facilitating when contrast is weak (Schwabe et al., 2010; Sengpiel et al., 1997) while higher contrasts have a suppressive effect (Anderson et al., 2001; Cavanaugh et al., 2002; Sceniak et al., 1999; Song and Li, 2008).

It also accounts for the effect of response normalization observed in V1 (Carandini and Heeger, 2011) (*i.e.* the response to a sum of visual stimuli is closer to the mean than to the sum of the responses when each stimulus is presented separately). In an SSN, this summation is supralinear when stimuli are weak, consistent with experimental observations (Heuer and Britten, 2002).

Finally, SSN models give an explanation of the suppression of response variability in response to stimulation during evoked activity (Hennequin et al., 2018). Increasing the stimulus contrast leads to an effective strengthening of the recurrent interactions and therefore settles the network in a state with strong suppressive inhibitory feedback and quenched correlated variability.

Strong interactions: The balanced state

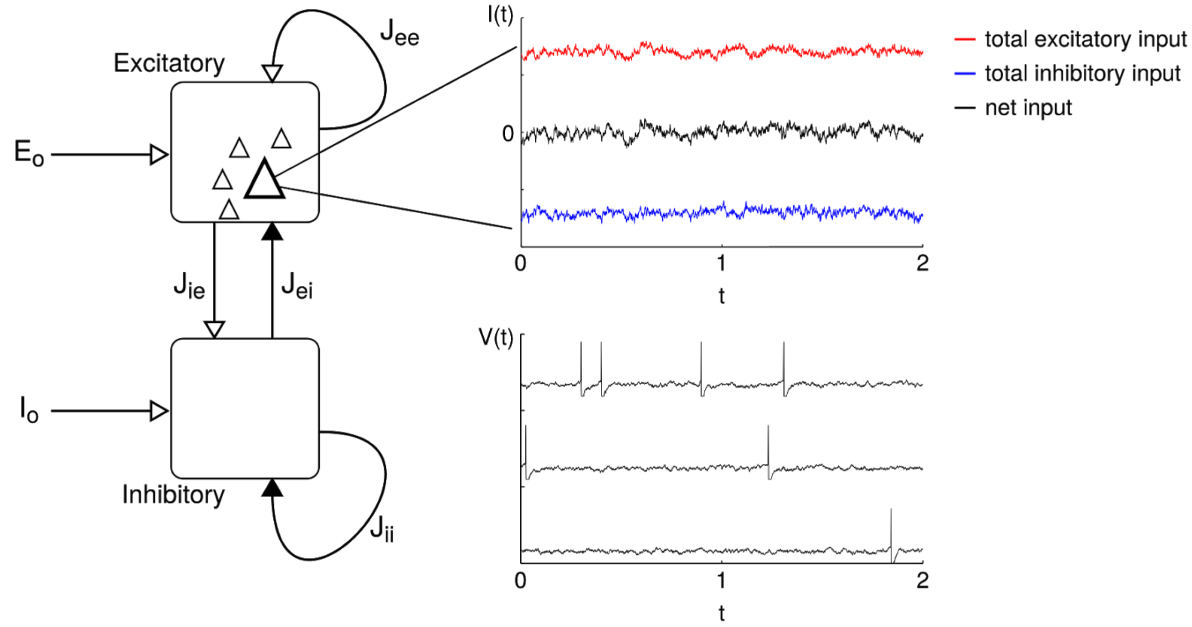


Figure 6. The balance of excitation and inhibition in a recurrent network of EI neurons with strong interactions. Figure from (Wolf et al., 2014).

Alternative models consider the case where the synaptic interactions between neurons, j_{AB}, are stronger than in the ISN and SSN scenarios by a factor $\sqrt{K_B}$ and, therefore, scale as $\frac{1}{\sqrt{K_B}}$, *i.e.* $j_{AB} = \frac{J_{AB}}{\sqrt{K_B}}$ (Van Vreeswijk and Sompolinsky, 2005; van Vreeswijk and Sompolinsky, 1998, 1996). In these models, under very general conditions the network activity self-adjusts so that excitatory and inhibitory inputs into the neurons balance each other (Fig. 6).

$$J_{A0}\, r_0 + \sum_B J_{AB}\, r_B = O(\frac{1}{\sqrt{K_A}}) \tag{7}$$

The resulting spatial and temporal fluctuations, although small in comparison with the average excitatory and inhibitory currents taken separately (*i.e.* $O(\sqrt{K_A})$) are of the same order as the total net inputs into the neurons (*i.e.* of the order of the neuronal rheobase current). Irregular firing is self-consistently generated by the collective network dynamics and does not require any fine-tuning of the parameters nor additional external noise into the neurons.

$$\alpha_A = \alpha_{A0} + \sum_B J_{AB}\, r_B = O(1) \qquad (8)$$

When K_A is sufficiently large, the dynamics linearizes and the firing rates are given by a linear combination of the strength of the external feedforward inputs into the network:

$$\vec{r} = J^{-1} \,.\, \vec{I_0} \qquad (9)$$

A necessary criterion for a network of strongly interacting neurons to operate in the balanced regime is that in the mean-field, the balanced solution is the only steady-state of the dynamics of the network. Therefore, one needs to constrain the parameters such that a solution for which $r_E = 0$ and $r_I = O(1)$ no longer exists. These constraints are summarized in the inequalities:

$$\frac{J_{E0}}{J_{I0}} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}} \qquad (10)$$

Analytical calculations (extended dynamical mean-field in the limit of large number of neurons, N, and large average number of connections, K_A) combined with numerical simulations have shown that the balanced state also provides a natural explanation for the observation that the distributions of firing rates in cortex are approximately log-normal (Roxin et al., 2011).

Balanced networks (Hansel and van Vreeswijk, 2012) can explain the emergence of strong stimulus selectivity in the absence of functionally specific connectivity. This counters what has been prevalent for many years in studies of the mouse visual cortex (Ohki and Reid, 2007).

Balanced networks can be bistable when synaptic interactions exhibit short-term plasticity such as synaptic depression and facilitation (Mongillo et al., 2012). Bistable balanced networks have been shown to account for the irregularity of persistent activity as observed in working memory tasks (Hansel and Mato, 2013). Moreover, with STP, synaptic interactions become non-linear and therefore, the balance equations become non-linear in the population firing rates just as in a SSN. In particular, (Konrad and Tchumatchenko, 2015) have recently proposed a model of balanced networks that exhibits similar properties as in a SSN in the presence of STP.

Balanced networks can also account for the quenching of neuronal variability observed during evoked activity (Churchland et al., 2010). This was recently shown in a model of mouse V1 (Rao et al., 2019) that includes circuit motifs such as bidirectional connections in the connectivity. Bidirectional connections between E and I neurons effectively act as negative feedback onto the neurons and can lead to a decrease in variability following stimulus onset.

There is experimental evidence supporting the idea that the cortex might be operating in an approximately balanced regime. Indeed, studies have reported large and opposing excitatory and inhibitory synaptic currents (Haider et al., 2006; Shu et al., 2003) as well as a high sensitivity of the network dynamics to small perturbations (London et al., 2010). Moreover, in cortical cultures, synaptic strengths have been shown to approximately scale as the inverse of the square root of the number of connections (Barral and D Reyes, 2016).

Balance of n populations

In order to balance n populations of neurons one needs to a priori discard 2^n-1 solutions and therefore consider 2^n-1 constraints on the parameters. If two or more steady states coexist, the network equilibrium could be multistable and the balanced solution could coexist with partially balanced solutions where one or more populations are silenced while the rest remain balanced.

Let us consider the case of a network consisting of two inhibitory populations (1 and 2). In the mean-field the activity of each of the two populations is

$$r_1 = \frac{1}{|J|}(J_{10}J_{22}-J_{20}J_{12})\, r_0 \tag{11}$$

$$r_2 = \frac{1}{|J|}(J_{20}J_{11}-J_{10}J_{21})\, r_0 \tag{12}$$

and a balanced solution exists as long as $r_1 > 0$ and $r_2 > 0$. There are a priori $2^2-1 = 3$ possible steady states together with the balanced solution: (1) $r_1 = O(1)$ and $r_2 = 0$; (2) $r_1 = 0$ and $r_2 = O(1)$; (3) $r_1 = 0$ and $r_2 = 0$. Because of the feedforward input into each of the populations, there is no solution for (3). (1) exists as long as $\frac{J_{10}}{J_{20}} > \frac{J_{11}}{J_{21}}$ and (2) exists as long as $\frac{J_{20}}{J_{10}} > \frac{J_{22}}{J_{12}}$. Therefore, when $|J| > 0$, the balanced solution is the unique steady state of the network dynamics. This is not the case when $|J| < 0$ where the balanced solution coexists with the partially balanced solutions (1) and (2). In this example, the balanced solution is unstable when it coexists with (1) and (2) and we are left with a bistable steady state of two partially balanced solutions.

PC-PV-SOM networks

Let us consider a more realistic network composed of three populations the PCs, the PV neurons and the SOM neurons. We consider an architecture where SOM neurons do not interact with each other (Adesnik et al., 2012; Gibson et al., 1999; Hu et al., 2011).

In the mean-field, the activity of the three populations is

$$r_E = \frac{1}{|J|}\,(J_{E0}\,J_{IS}\,J_{SI}-J_{I0}\,J_{ES}\,J_{SI}+J_{S0}\,(J_{ES}\,J_{II}-J_{EI}\,J_{IS}))\; r_0 \tag{13}$$

$$r_I = \frac{1}{|J|} \left(J_{E0} J_{IS} J_{SE} - J_{I0} J_{ES} J_{SE} + J_{S0} (J_{ES} J_{IE} - J_{EE} J_{IS})\right) r_0 \qquad (14)$$

$$r_S = \frac{1}{|J|} \left(J_{E0} (J_{SI} J_{IE} - J_{SE} J_{II}) + J_{I0} (J_{SE} J_{EI} - J_{SI} J_{EE}) + J_{S0} (J_{EE} J_{II} - J_{EI} J_{IE})\right) r_0 \qquad (15)$$

There are a priori $2^3 - 1 = 7$ possible partially balanced solutions of the dynamics:

(1) $r_E = O(1)$, $r_I = 0$ and $r_S = 0$ which never exists.

(2) $r_E = 0$, $r_I = O(1)$ and $r_S = 0$ which implies

$$\frac{J_{EI}}{J_{II}} > \frac{J_{E0}}{J_{I0}} \text{ and } \frac{J_{SI}}{J_{II}} > \frac{J_{S0}}{J_{I0}}$$

(3) $r_E = O(1)$, $r_I = O(1)$ and $r_S = 0$ which implies

$$r_E = \frac{J_{E0} J_{II} - J_{I0} J_{EI}}{J_{EI} J_{IE} - J_{EE} J_{II}} r_0 \text{ and } r_I = \frac{J_{E0} J_{IE} - J_{I0} J_{EE}}{J_{EI} J_{IE} - J_{EE} J_{II}} r_0 \text{ and } J_{S0} r_0 + J_{SE} r_E - J_{SI} r_I < 0$$

(4) $r_E = 0$, $r_I = 0$ and $r_S = O(1)$ which never exists but implies that the unbalanced solution $r_E = 0$, $r_I = 0$ and $r_S = O(\sqrt{K})$ always exists as long as the SOM population receives a strong feedforward input.

(5) $r_E = O(1)$, $r_I = 0$ and $r_S = O(1)$ which never exists.

(6) $r_E = 0$, $r_I = O(1)$ and $r_S = O(1)$ which implies

$$J_{I0} J_{SI} - J_{S0} J_{II} > 0 \text{ and } J_{E0} J_{IS} J_{SI} - J_{S0} J_{EI} J_{IS} + J_{ES} (J_{S0} J_{II} - J_{I0} J_{SI}) < 0$$

(7) $r_E = 0$, $r_I = 0$ and $r_S = 0$ which never exists.

From (4) we see that whenever the SOM population receives a strong feedforward input (of order $\sqrt{K}$), the balanced state coexists with an unbalanced solution where the activity of the SOM population saturates while the PCs and PV neurons are silenced. To avoid such pathological states, feedforward inputs into the SOM neurons must be at most of order 1. Consistent with our findings, experimental studies have actually reported that SOM neurons receive little to no inputs from external sources like the thalamus in the cortex (Beierlein et al., 2003, 2000; Cruikshank et al., 2010; Ma et al., 2006; Xu et al., 2013).

Therefore, in the absence of strong feedforward input into the SOM neurons, the balanced conditions in the PC-PV-SOM network can now be simplified.

(2) and (6) imply that $\frac{J_{ES}}{J_{IS}} > \frac{J_{E0}}{J_{I0}} > \frac{J_{EI}}{J_{II}}$. Then, the positivity of the rates of E and I imply $|J| < 0$, and the positivity of r_S that $J_{E0} (J_{SE} J_{II} - J_{SI} J_{IE}) > J_{I0} (J_{SE} J_{EI} - J_{SI} J_{EE})$.

Finally, to eliminate the existence of the partially balanced solution (3) we get the additional constraint

$$\frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}} \quad or \quad \frac{J_{EE}}{J_{IE}} > \frac{J_{E0}}{J_{I0}}$$

Altogether these constraints lead to two sets of solutions, one of which remarkably conserves the conditions required for the standard balance of two populations.

Optogenetic activation of PV inhibitory neurons in vivo

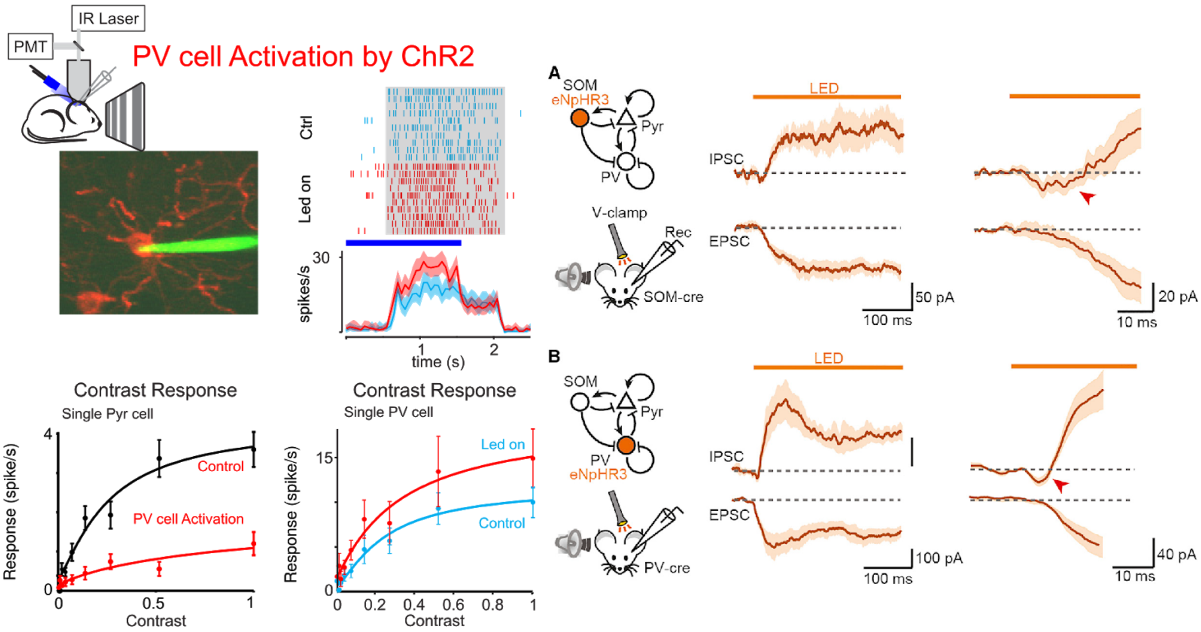


Figure 7. ChR2 photostimulation of PV inhibitory neurons *in vivo*. **A**. PV cell activation with ChR2 photostimulation. Left: Example of contrast response of a single PC in control (black) and during stimulation of PV neurons (red). Right: Example of contrast response of a single PV cell in control (cyan) and during activation with ChR2 (red). Panels adapted from (Atallah et al., 2012). **B**. The suppression of interneurons causes a paradoxical increase in PC-inhibitory synaptic current. Top: photoinactivation of SOM cells increases spontaneous EPSCs and IPSCs. Bottom: PV cell inactivation experiments. Panels adapted from (Kato et al., 2017).

If the cortex operates in an inhibition stabilized regime, the theory predicts that stimulation of the GABAergic neurons should lead to a paradoxical response in which the activity of both excitatory and inhibitory neurons decreases (see the previous section). However, experimental evidence for inhibition stabilization in the cortex is ambiguous since both paradoxical and non-paradoxical effects have been reported. (Atallah et al., 2012) reported that optogenetic stimulation of PV interneurons in the layer 2/3 of the mouse primary visual cortex resulted in an increase in the inhibitory current to PC through an augmentation of the total inhibitory activity (Fig. 7A). More recently, (Kato et al., 2017) found that the optogenetic inactivation of PV or SOM interneurons in mouse auditory cortex layer 2/3 strengthens inhibitory currents to PCs (Fig. 7B). The studies mentioned above used Channelrhodopsin to activate the different inhibitory subtypes, it is likely that only a subset of the targeted neurons was activated. Theoretical studies ((Sadeh et al.,

2017) with networks operating in an ISN regime, (Gutnisky et al., 2017) with networks operating in an approximately balanced regime) have shown that when a relatively small fraction of the inhibitory population is stimulated, the stimulated neurons increase their activity upon light onset and therefore exhibit a non-paradoxical response (Fig. 8). More recently, (Sanzeni et al., 2019) have provided data consistent with the balanced theory, and observed a non-paradoxical response in the upper layers of visual, somatosensory, and motor cortex of the mouse when expressing Channelrhodopsin in PV neurons but in addition found that when using transgenic mouses in which almost all neurons expressed the markers, stimulating PV neurons produced a population paradoxical effect.

Modeling the ChR2 activation of PV neurons in a standard balanced network

We give here a detailed analysis of the effect of the partial activation of the inhibitory population in strongly recurrent networks of neurons with one excitatory and one inhibitory population operating in the balanced regime.

Activation of a small proportion of I neurons

We want to investigate the response of the network to the stimulation of a proportion p of its inhibitory population. Let us denote by $+$ (resp. $-$) the subpopulation of stimulated (resp. non-stimulated) inhibitory neurons. We model the effect of the stimulation as an additional large external input, $\sqrt{K}\,\delta I$ into the neurons in population $+$. Let us start with the stimulation of a very small fraction of the inhibitory population, namely $p = \frac{\varepsilon}{\sqrt{K}}$, where ε is $O(1)$, and describe the response of the network with increasing the proportion p of stimulated inhibitory cells.

The average inputs into the neurons are

$$u_E = \sqrt{K}\,(J_{E0}\,r_0 + J_{EE}\,r_E - J_{EI}\,r_-) - \varepsilon\,J_{EI}(r_+ - r_-) \qquad (16)$$

$$u_+ = \sqrt{K}\,(J_{I0}\,r_0 + \delta I + J_{IE}\,r_E - J_{II}\,r_-) - \varepsilon\,J_{II}(r_+ - r_-) \qquad (17)$$

$$u_- = \sqrt{K}\,(J_{I0}\,r_0 + J_{IE}\,r_E - J_{II}\,r_-) - \varepsilon\,J_{II}(r_+ - r_-) \qquad (18)$$

where u_+ (resp. u_-) is the net input into the perturbed population (resp. non-perturbed population), and r_+ (resp. r_-) is the average rate of the perturbed population (resp. non-perturbed population).

Because the stimulation is of strength $O(\sqrt{K})$, there is no solution in which u_E, u_+ and u_- can remain finite and non-zero. Therefore, even stimulating a very small proportion of the inhibitory neurons breaks the global balance of the network. Nevertheless, we can consider solutions for which either the stimulated population saturates and the non-stimulated population is still balanced or for which the stimulated population remains balanced and the non-stimulated population is silenced.

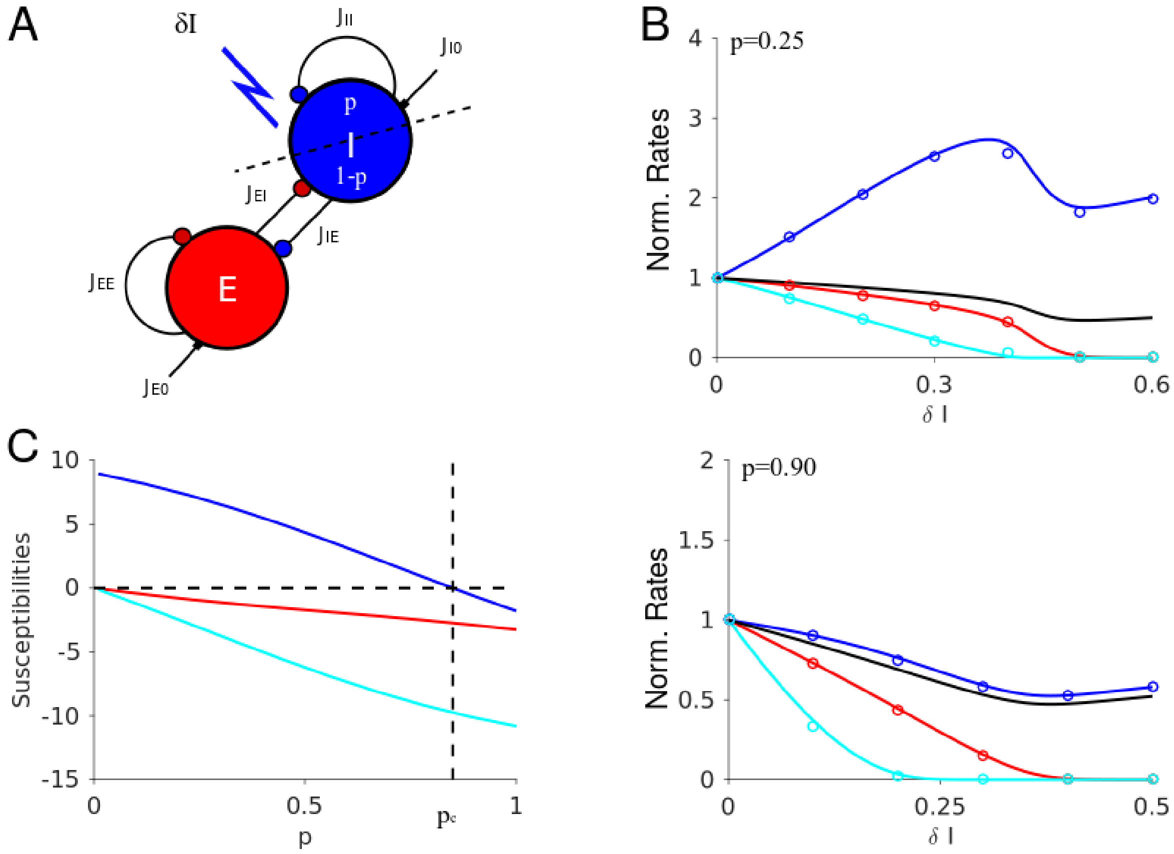


Figure 8. Modeling the ChR2 activation of PV neurons. **A.** Scheme of the EI network. **B.** Top: Activation of a small proportion of I cells. Bottom: Activation of a large proportion of I cells. **C.** Evolution of the normalized slope of PCs and PVs' normalized activities (evaluated for $\delta I \to 0$) as a function of the proportion of stimulated neurons, p.

Because we only consider the stimulation of a very small fraction of the I population, in the limit where K goes to infinity, the only consistent solution is the one where $r_E = O(1)$, $r_+ \to +\infty$ and $r_- = O(1)$.

Let us write $r_E = r_E^\infty + \frac{\nu_E}{\sqrt{K}}$, $r_+ = \Lambda_+ \sqrt{K} + \lambda_+$ and $r_- = r_-^\infty + \frac{\nu_-}{\sqrt{K}}$.

The balance of the inputs into the excitatory cells and into the non-stimulated I neurons leads to

$$J_{E0}\, r_0 + J_{EE}\, r_E^\infty - J_{EI}\,(r_-^\infty + \varepsilon\Lambda_+) = 0 \qquad (19)$$

$$J_{I0}\, r_0 + J_{IE}\, r_E^\infty - J_{II}\,(r_-^\infty + \varepsilon\Lambda_+) = 0 \qquad (20)$$

Thus, $r_E^\infty = \frac{J_{II}\,J_{E0} - J_{EI}\,J_{I0}}{|J|}\, r_0$ and $r_-^\infty = \frac{J_{IE}\,J_{E0} - J_{EE}\,J_{I0}}{|J|}\, r_0 - \varepsilon\,\Lambda_+$, and to leading order,

$$r_E^\infty = \Psi[J_{EE}\,\nu_E + J_{EI}\,(\varepsilon\,(r_-^\infty - \lambda_+) - \nu_-), \beta, \alpha] \tag{21}$$

$$r_-^\infty = \Psi[J_{IE}\,\nu_E + J_{II}\,(\varepsilon\,(r_-^\infty - \lambda_+) - \nu_-), \beta\ , \alpha] \tag{22}$$

$$r_+ = \Psi[\sqrt{K}(\delta I - \varepsilon\, J_{II}\,\Lambda_+) + J_{IE}\,\nu_E + J_{II}\,(\varepsilon\,(r_-^\infty - \lambda_+) - \nu_-), \beta, \alpha\,] \tag{23}$$

Given that $\Psi(x,\beta,\alpha) \approx \frac{x}{\tau(V_{th}-V_R)}$ when $x \to \infty$, equation [23] leads to

$$\Lambda_+ = \frac{\delta I}{\tau\,(V_{th}-V_r) + \varepsilon\, J_{II}} \tag{24}$$

$$\lambda_+ = \frac{J_{IE}\,\nu_E + J_{II}\,(\varepsilon\, r_-^\infty - \nu_-)}{\tau\,(V_{th}-V_r) + \varepsilon\, J_{II}} \tag{25}$$

and the corrections ν_E and ν_- can be computed using equations [21] and [22].

Remarkably, Λ_+ and r_-^∞ decrease with ε so for sufficiently large K , r_+ and r_- will also decrease with ε.

Depending on the parameters, namely on the value of the baseline rate r_{BL} of the inhibitory population with respect to the quantity $\frac{\delta I}{J_{II}}$, r_-^∞ or Λ_+ becomes zero. When $\frac{\delta I}{J_{II}} > r_{BL}$, r_-^∞ becomes 0 for $\varepsilon = \varepsilon_C = \frac{\tau(V_{th}-V_r)\, r_{BL}}{\delta I - J_{II}\, r_{BL}}$.

Activation of a large proportion of I neurons

When the strength of the perturbation is $O(\sqrt{K})$ and the proportion of perturbed neurons, p , is $O(1)$, the average inputs into the neurons are given by

$$u_E = \sqrt{K}\,(E + J_{EE}\, r_E - J_{EI}\,(p\, r_+ + (1-p)\, r_-) \tag{26}$$

$$u_+ = \sqrt{K}\,(I + \ \delta I + J_{IE}\, r_E - J_{II}\,(p\, r_+ + (1-p)\, r_-)) \tag{27}$$

$$u_- = \sqrt{K}\,(J_{I0}\, r_0 + J_{IE}\, r_E - J_{II}\,(p\, r_+ + (1-p)\, r_-)) \tag{28}$$

In the limit where K goes to infinity, the only consistent solution is the one where $r_E^\infty = O(1)$, $r_+^\infty = O(1)$ and $r_-^\infty = 0$. r_E^∞ and r_+^∞ are determined by the equations imposing the balance of the net inputs into the PCs and into the perturbed inhibitory population:

$$J_{E0}\, r_0 + J_{EE}\, r_E^\infty - J_{EI}\, p\, r_+^\infty = 0 \tag{29}$$

$$J_{I0}\, r_0 + \ \delta I + J_{IE}\, r_E^\infty - J_{II}\, p\, r_+^\infty = 0 \tag{30}$$

Thus, $r_E^\infty = \frac{J_{II}\, J_{E0} - J_{EI}\, J_{I0}}{|J|}\, r_0 - \frac{J_{EI}\,\delta I}{|J|}$ and $r_+^\infty = \frac{J_{IE}\, J_{E0} - J_{EE}\, J_{I0}}{p\,|J|}\, r_0 - \frac{J_{EE}\,\delta I}{p\,|J|}$.

Remarkably, the rate of the perturbed population decreases as $\frac{1}{p}$. Its response becomes paradoxical for a value of p such that $r_+^\infty(p = p_c^\infty) = r_{BL}$, namely,

$$p_c^\infty = 1 - \frac{J_{EE}\,\delta I}{r_{BL}\,|J|} \tag{31}$$

For finite but large K, let us write $r_E = r_E^\infty + \frac{\nu_E}{\sqrt{K}}$ and $r_I = p\, r_+^\infty + \frac{\nu_I}{\sqrt{K}}$,

$$r_E = \Psi[J_{EE}\,\nu_E - J_{EI}\,\nu_I, \beta, \alpha] \tag{32}$$

$$r_+ = \Psi[J_{IE}\,\nu_E - J_{II}\,\nu_I, \beta, \alpha] \tag{33}$$

$$r_- = \Psi[-\delta I\sqrt{K} + J_{IE}\,\nu_E - J_{II}\,\nu_I, \beta, \alpha] \tag{34}$$

The first two equations Eq. 32-33 determine the first-order corrections of the rates of the excitatory and inhibitory populations. The third one Eq. (34) gives the order of the correction of the rate of the non-perturbed inhibitory population which gives an exponentially small function of K.

References

Adesnik H, Bruns W, Taniguchi H, Josh Huang Z, Scanziani M. 2012. A neural circuit for spatial summation in visual cortex. *Nature*. doi:10.1038/nature11526

Ahmadian Y, Rubin DB, Miller KD. 2013. Analysis of the Stabilized Supralinear Network. *Neural Computation*. doi:10.1162/neco_a_00472

Anderson JS, Lampl I, Gillespie DC, Ferster D. 2001. Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. *J Neurosci* **21**:2104–2112.

Atallah BV, Bruns W, Carandini M, Scanziani M. 2012. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**:159–170.

Bair W, Koch C, Newsome W, Britten K. 1994. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J Neurosci* **14**:2870–2892.

Barral J, D Reyes A. 2016. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nat Neurosci* **19**:1690.

Beierlein M, Gibson JR, Connors BW. 2003. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *J Neurophysiol* **90**:2987–3000.

Beierlein M, Gibson JR, Connors BW. 2000. A network of electrically coupled interneurons drives synchronized inhibition in neocortex. *Nat Neurosci* **3**:904–910.

Braitenberg V, Schüz A. 2013. Anatomy of the Cortex: Statistics and Geometry. Springer Science & Business Media.

Buzsáki G, Mizuseki K. 2014. The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci* **15**:264–278.

Carandini M, Heeger DJ. 2011. Normalization as a canonical neural computation. *Nat Rev Neurosci* **13**:51–62.

Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, Moore CI. 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**:663–667.

Cavanaugh JR, Bair W, Movshon JA. 2002. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J Neurophysiol* **88**:2547–2556.

Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, Bradley DC, Smith MA, Kohn A, Movshon JA, Armstrong KM, Moore T, Chang SW, Snyder LH, Lisberger SG, Priebe NJ, Finn IM, Ferster D, Ryu SI, Santhanam G, Sahani M, Shenoy KV. 2010. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci* **13**:369–378.

Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang X-J. 2003. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol* **90**:3441–3454.

Cruikshank SJ, Urabe H, Nurmikko AV, Connors BW. 2010. Pathway-specific feedforward circuits between thalamus and neocortex revealed by selective optical stimulation of axons. *Neuron* **65**:230–245.

DeFelipe J, Alonso-Nanclares L, Arellano JI. 2002. Microstructure of the neocortex: comparative aspects. *J Neurocytol* **31**:299–316.

DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Cauli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, Hof PR, Huang J, Jones EG, Kawaguchi Y, Kisvárday Z, Kubota Y, Lewis DA, Marín O, Markram H, McBain CJ, Meyer HS, Monyer H, Nelson SB, Rockland K, Rossier J, Rubenstein JLR, Rudy B, Scanziani M, Shepherd GM, Sherwood CC, Staiger JF, Tamás G, Thomson A, Wang Y, Yuste R, Ascoli GA. 2013. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* **14**:202–216.

Freund TF, Martin KAC, Smith AD, Somogyi P. 1983. Glutamate decarboxylase-immunoreactive terminals of Golgi-impregnated axoaxonic cells and of presumed basket cells in synaptic contact with pyramidal neurons of the cat’s visual cortex. *J Comp Neurol* **221**:263–278.

Fu Y, Kaneko M, Tang Y, Alvarez-Buylla A, Stryker MP. 2015. A cortical disinhibitory circuit for enhancing adult plasticity. *Elife* **4**:e05558.

Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, Stryker MP. 2014. A cortical circuit for gain control by behavioral state. *Cell* **156**:1139–1152.

Gibson JR, Beierlein M, Connors BW. 1999. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature* **402**:75–79.

Giugliano M, Darbon P, Arsiero M, Lüscher H-R, Streit J. 2004. Single-Neuron Discharge Properties and Network Activity in Dissociated Cultures of Neocortex. *J Neurophysiol* **92**:977–996.

Gupta A, Wang Y, Markram H. 2000. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**:273–278.

Gutnisky DA, Yu J, Hires SA, To M-S, Bale MR, Svoboda K, Golomb D. 2017. Mechanisms underlying a thalamocortical transformation during active tactile sensation. *PLoS Comput Biol* **13**:e1005576.

Haider B, Duque A, Hasenstaub AR, McCormick DA. 2006. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* **26**:4535–4545.

Hansel D, Mato G. 2013. Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. *Journal of Neuroscience*. doi:10.1523/jneurosci.3455-12.2013

Hansel D, van Vreeswijk C. 2012. The mechanism of orientation selectivity in primary visual cortex without a functional map. *J Neurosci* **32**:4049–4064.

Harish O, Hansel D. 2015. Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Comput Biol* **11**:e1004266.

Hayut I, Fanselow EE, Connors BW, Golomb D. 2011. LTS and FS Inhibitory Interneurons, Short-Term Synaptic Plasticity, and Cortical Circuit Dynamics. *PLoS Computational Biology*. doi:10.1371/journal.pcbi.1002248

Hellwig B. 2000. A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol Cybern* **82**:111–121.

Hennequin G, Ahmadian Y, Rubin DB, Lengyel M, Miller KD. 2018. The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron* **98**:846–860.e5.

Heuer HW, Britten KH. 2002. Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol* **88**:3398–3408.

Holmgren C, Harkany T, Svennenfors B, Zilberter Y. 2003. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J Physiol* **551**:139–153.

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

Hu H, Ma Y, Agmon A. 2011. Submillisecond firing synchrony between different subtypes of cortical interneurons connected chemically but not electrically. *J Neurosci* **31**:3351–3361.

Isaacson JS, Scanziani M. 2011. How inhibition shapes cortical activity. *Neuron* **72**:231–243.

Jiang X, Shen S, Cadwell CR, Berens P, Sinz F, Ecker AS, Patel S, Tolias AS. 2015. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**:aac9462.

Kaneko M, Fu Y, Stryker MP. 2017. Locomotion Induces Stimulus-Specific Response Enhancement in Adult Visual Cortex. *J Neurosci* **37**:3532–3543.

Kato HK, Asinof SK, Isaacson JS. 2017. Network-Level Control of Frequency Tuning in Auditory Cortex. *Neuron* **95**:412–423.e4.

Kätzel D, Zemelman BV, Buetfering C, Wölfel M, Miesenböck G. 2011. The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. *Nat Neurosci* **14**:100–107.

Katzner S, Busse L, Carandini M. 2011. GABAA inhibition controls response gain in visual cortex. *J Neurosci* **31**:5931–5941.

Klausberger T. 2009. GABAergic interneurons targeting dendrites of pyramidal cells in the CA1 area of the hippocampus. *Eur J Neurosci* **30**:947–957.

Klausberger T, Somogyi P. 2008. Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations. *Science*. doi:10.1126/science.1149381

Konrad S, Tchumatchenko T. 2015. Surround suppression and normalization in a model of coupled balanced cortical networks with short-term synaptic plasticity. *BMC Neurosci* **16**:P249.

Lee S-H, Kwan AC, Zhang S, Phoumthipphavong V, Flannery JG, Masmanidis SC, Taniguchi H, Huang ZJ, Zhang F, Boyden ES, Deisseroth K, Dan Y. 2012. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**:379–383.

Lee S, Kruglikov I, Huang ZJ, Fishell G, Rudy B. 2013. A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nat Neurosci* **16**:1662–1670.

Li N, Chen S, Guo ZV, Chen H, Huo Y, Inagaki HK, Chen G, Davis C, Hansel D, Guo C, Svoboda K. 2019. Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *Elife* **8**. doi:10.7554/eLife.48622

London M, Roth A, Beeren L, Häusser M, Latham PE. 2010. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature* **466**:123–127.

Markram H, Lübke J, Frotscher M, Roth A, Sakmann B. 1997. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J Physiol* **500 (Pt 2)**:409–440.

Ma Y, Hu H, Berrebi AS, Mathers PH, Agmon A. 2006. Distinct subtypes of somatostatin-containing neocortical interneurons revealed in transgenic mice. *J Neurosci* **26**:5069–5082.

Meyer HS, Schwarz D, Wimmer VC, Schmitt AC, Kerr JND, Sakmann B, Helmstaedter M. 2011. Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc Natl Acad Sci U S A* **108**:16807–16812.

Mizuseki K, Buzsáki G. 2013. Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell Rep* **4**:1010–1021.

Mongillo G, Hansel D, van Vreeswijk C. 2012. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys Rev Lett* **108**:158101.

Nienborg H, Hasenstaub A, Nauhaus I, Taniguchi H, Huang ZJ, Callaway EM. 2013. Contrast Dependence and Differential Contributions from Somatostatin- and Parvalbumin-Expressing Neurons to Spatial Integration in Mouse V1. *Journal of Neuroscience*. doi:10.1523/jneurosci.5320-12.2013

O'Connor DH, Peron SP, Huber D, Svoboda K. 2010. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* **67**:1048–1061.

Ohki K, Reid RC. 2007. Specificity and randomness in the visual cortex. *Curr Opin Neurobiol* **17**:401–407.

Ozeki H, Finn IM, Schaffer ES, Miller KD, Ferster D. 2009. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**:578–592.

Pack CC, Hunter JN, Born RT. 2005. Contrast Dependence of Suppressive Influences in Cortical Area MT of Alert Macaque. *J Neurophysiol* **93**:1809–1815.

Packer AM, Roska B, Häusser M. 2013. Targeting neurons and photons for optogenetics. *Nat Neurosci* **16**:805–815.

Pattadkal JJ, Mato G, van Vreeswijk C, Priebe NJ, Hansel D. 2018. Emergent Orientation Selectivity from Random Networks in Mouse Visual Cortex. *Cell Reports*. doi:10.1016/j.celrep.2018.07.054

Perin R, Berger TK, Markram H. 2011. A synaptic organizing principle for cortical neuronal groups. *Proc Natl Acad Sci U S A* **108**:5419–5424.

Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M. 2013. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat Neurosci* **16**:1068–1076.

Pi H-J, Hangya B, Kvitsiani D, Sanders JI, Josh Huang Z, Kepecs A. 2013. Cortical interneurons that specialize in disinhibitory control. *Nature*. doi:10.1038/nature12676

Rao S, Hansel D, van Vreeswijk C. 2019. Dynamics and orientation selectivity in a cortical model of rodent V1 with excess bidirectional connections. *Sci Rep* **9**:3334.

Roxin A, Brunel N, Hansel D, Mongillo G, van Vreeswijk C. 2011. On the distribution of firing rates in networks of cortical neurons. *J Neurosci* **31**:16217–16226.

Rubin DB, Van Hooser SD, Miller KD. 2015. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**:402–417.

Rudy B, Fishell G, Lee S, Hjerling-Leffler J. 2011. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev Neurobiol* **71**:45–61.

Sadeh S, Silver RA, Mrsic-Flogel TD, Muir DR. 2017. Assessing the Role of Inhibition in Stabilizing Neocortical Networks Requires Large-Scale Perturbation of the Inhibitory Population. *J Neurosci* **37**:12050–12067.

Sanzeni A, Akitake B, Goldbach HC, Leedy CE. 2019. Inhibition stabilization is a widespread property of cortical networks. *bioRxiv*.

Sceniak MP, Ringach DL, Hawken MJ, Shapley R. 1999. Contrast's effect on spatial summation by macaque V1 neurons. *Nat Neurosci* **2**:733–739.

Schüz A, Palm G. 1989. Density of neurons and synapses in the cerebral cortex of the mouse. *J Comp Neurol* **286**:442–455.

Schwabe L, Ichida JM, Shushruth S, Mangapathy P, Angelucci A. 2010. Contrast-dependence of surround suppression in Macaque V1: experimental testing of a recurrent network model. *Neuroimage* **52**:777–792.

Sengpiel F, Sen A, Blakemore C. 1997. Characteristics of surround inhibition in cat area 17. *Exp Brain Res* **116**:216–228.

Shinomoto S, Kim H, Shimokawa T, Matsuno N, Funahashi S, Shima K, Fujita I, Tamura H, Doi T, Kawano K, Inaba N, Fukushima K, Kurkin S, Kurata K, Taira M, Tsutsui K-I, Komatsu H, Ogawa T, Koida K, Tanji J, Toyama K. 2009. Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Comput Biol* **5**:e1000433.

Shu Y, Hasenstaub A, McCormick DA. 2003. Turning on and off recurrent balanced cortical activity. *Nature*. doi:10.1038/nature01616

Softky WR, Koch C. 1993. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* **13**:334–350.

Sohal VS, Zhang F, Yizhar O, Deisseroth K. 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459**:698–702.

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

Somogyi P, Klausberger T. 2005. Defined types of cortical interneurone structure space and spike timing in the hippocampus. *J Physiol* **562**:9–26.

Song X-M, Li C-Y. 2008. Contrast-dependent and contrast-independent spatial summation of primary visual cortical neurons of the cat. *Cereb Cortex* **18**:331–336.

Stark E, Eichler R, Roux L, Fujisawa S, Rotstein HG, Buzsáki G. 2013. Inhibition-induced theta resonance in cortical circuits. *Neuron* **80**:1263–1276.

Stark E, Roux L, Eichler R, Senzai Y, Royer S, Buzsáki G. 2014. Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron* **83**:467–480.

Stepanyants A, Martinez LM, Ferecsko AS, Kisvarday ZF. 2009. The fractions of short- and long-range connections in the visual cortex. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.0810390106

Taniguchi H, He M, Wu P, Kim S, Paik R, Sugino K, Kvitsiani D, Fu Y, Lu J, Lin Y, Miyoshi G, Shima Y, Fishell G, Nelson SB, Huang ZJ. 2011. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**:995–1013.

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, Koch C, Zeng H. 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**:335–346.

Thomson AM, Bannister AP. 2003. Interlaminar connections in the neocortex. *Cereb Cortex* **13**:5–14.

Thomson AM, Bannister AP, Mercer A, Morris OT. 2002. Target and temporal pattern selection at neocortical synapses. *Philos Trans R Soc Lond B Biol Sci* **357**:1781–1791.

Tomko GJ, Crapper DR. 1974. Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Res* **79**:405–418.

Tremblay R, Lee S, Rudy B. 2016. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* **91**:260–292.

Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1997. Paradoxical effects of external modulation of inhibitory interneurons. *J Neurosci* **17**:4382–4388.

Vanni MP, Casanova C. 2013. Surround suppression maps in the cat primary visual cortex. *Front Neural Circuits* **7**:78.

Van Vreeswijk C, Sompolinsky H. 2005. Irregular activity in large networks of neuronsLes Houches. Elsevier. pp. 341–406.

van Vreeswijk C, Sompolinsky H. 1998. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*. doi:10.1162/089976698300017214

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

Wang C, Bardy C, Huang JY, FitzGibbon T, Dreher B. 2009. Contrast dependence of center and surround integration in primary visual cortex of the cat. *Journal of Vision*. doi:10.1167/9.1.20

Wolf F, Engelken R, Puelma-Touzel M, Weidinger JDF, Neef A. 2014. Dynamical models of cortical circuits. *Curr Opin Neurobiol* **25**:228–236.

Wood KC, Blackwell JM, Geffen MN. 2017. Cortical inhibitory interneurons control sensory processing. *Curr Opin Neurobiol* **46**:200–207.

Xu H, Jeong H-Y, Tremblay R, Rudy B. 2013. Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron* **77**:155–167.

Xu X, Roby KD, Callaway EM. 2010. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J Comp Neurol* **518**:389–404.

Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S. 2015. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**:1138–1142.

Chapter 2

Mechanisms underlying the response to the optogenetic photostimulation of PV inhibitory neurons in vivo

Introduction

Loss-of-function is a common strategy used to infer the relation between network dynamics and neural computation and behavior. Classically, it relies on irreversible lesions of brain areas (Lashley, 1931; Mishkin and Ungerleider, 1982; Newsome and Wurtz, 1988), on the use of chemicals to permanently disrupt excitatory neuronal activity (Guo et al., 2017; Hikosaka and Wurtz, 1985; Krupa et al., 1999), or on the cooling of brain tissues (Long and Fee, 2008; Ponce et al., 2008).

In recent years, optogenetic tools have allowed the reversible silencing of local excitation in the mouse cortex (Deisseroth, 2015). This photo-inhibition can be achieved when a light-sensitive protein (opsin involving light-gated Cl-/H+ hyperpolarizing pumps) is expressed in the excitatory neurons (Yizhar et al., 2011). Upon light stimulation, these proteins cause the hyperpolarization of the membrane potential of the targeted neurons, leading to their inactivation (Chow et al., 2010; Wiegert et al., 2017; Zhang et al., 2007).

Alternatively, photo-suppression can also be achieved indirectly when an excitatory opsin (channelrhodopsin, ChR) is expressed in the inhibitory neurons (Asrican et al., 2013; Madisen et al., 2012; Zhao et al., 2011). Specific Cre lines have been developed that enable the stimulation of specific inhibitory subtypes (*e.g.* PV, SOM or VIP neurons) *in vivo*. Different variants of ChR (*e.g.* with blue-shifted light, ChR2, and with red-shifted light, ReaChR; Hooks et al., 2015; Klapoetke et al., 2014; Lin et al., 2013) also exists allowing for the stimulation of local as well as very extended cortical regions.

The effects of the ChR-assisted suppression of cortical excitation can lead to counter-intuitive observations. Indeed, the optogenetic stimulation of a given inhibitory subtype can paradoxically lead to a decrease in the activity of these targeted neurons. The mechanisms underlying these effects are likely to depend on the ChR variant as well as the targeted inhibitory subtype and are still to be investigated.

We have described in the introduction how the effects of the ChR2 stimulation of PV neurons can be accounted for in two-population network models when only a subset of the inhibitory neurons is effectively stimulated. Here, we will characterize the effects of the optogenetic stimulation of PV inhibitory neurons in ReaChR mice in the anterior lateral motor cortex (ALM) and in the barrel cortex (S1). ReaChR is a red-shifted variant of ChR that enables the stimulation of a large fraction of the neurons in a very wide area.

Manuscript

Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation

Alexandre Mahrach[1], Guang Chen[2], Nuo Li[2], Carl van Vreeswijk[1], David Hansel[1]

[1] CNRS-UMR 8002, Integrative Neuroscience and Cognition Center, 45 Rue des Saints-Pères, 75270 Paris, France

[2] Department of Neuroscience, Baylor College of Medicine, Houston TX 77030, USA

Correspondence:
David Hansel
CNRS-UMR 8002
Integrative Neuroscience and Cognition Center
45 Rue des Saints-Pères, 75270
Paris, France
david.hansel@parisdescartes.fr

Abstract

GABAergic Interneurons can be subdivided into three subclasses: parvalbumin positive (PV), somatostatin positive (SOM) and serotonin positive neurons. With principal cells (PCs) they form complex networks. We examine PCs and PV responses in mouse anterior lateral motor cortex (ALM) and barrel cortex (S1) upon PV photostimulation *in vivo*. In ALM layer 5 and S1, the PV response is paradoxical: photoexcitation reduces their activity. This is not the case in ALM layer 2/3. We combine analytical calculations and numerical simulations to investigate how these results constrain the architecture. Two-population models cannot explain the results. Four-population networks with V1-like architecture account for the data in ALM layer 2/3 and layer 5. Our data in S1 can be explained if SOM neurons receive inputs only from PCs and PV neurons. In both four-population models, the paradoxical effect implies not too strong recurrent excitation. It is not evidence for stabilization by inhibition.

Introduction

Local cortical circuits comprise several subclasses of GABAergic interneurons which together with the excitatory neurons form complex recurrent networks (Goldberg et al., 2004; Jiang et al., 2015; Karnani et al., 2016; Markram et al., 2004; Moore et al., 2010; Pfeffer et al., 2013; Tasic et al., 2018; Tremblay et al., 2016). The architecture of these networks depends on the cortical area and layer (Beierlein et al., 2003; Jiang et al., 2013; Rudy et al., 2011; Xu et al., 2013; Xu and Callaway, 2009).

Optogenetics is now classically used to reversibly inactivate a particular cortical area or neuronal population to get insights into their functions (Atallah et al., 2012; Guo et al., 2014b; Lee et al., 2012; Li et al., 2015; Svoboda and Li, 2018). Optogenetics has also been applied to isolate the different components (*e.g.* feedforward *vs.* recurrent) of the net input into cortical neurons (Lien and Scanziani, 2018, 2013). It can also be used to experimentally probe the architecture of local cortical circuits (Moore et al., 2018; Xu et al., 2013). However, because of the complexity of these networks and of their nonlinear dynamics, qualitative intuition and simple reasoning (*e.g.* ‘box-and-arrow’ diagrams) are of limited use to interpret the results of these manipulations.

“Paradoxical effect” designates the phenomenon that stimulation of a GABAergic interneuron population not only decreases the average activity of the principal cells (PCs) but also *decreases* the activity of the stimulated population (Murphy and Miller, 2009; Ozeki et al., 2009; Tsodyks et al., 1997). Intuitively, paradoxical effect arises when the stimulation induces a strong activity suppression in the PCs (Kato et al., 2017; Moore et al., 2018), such that the overall (synaptic+stimulus) excitation to the stimulated population decreases. However, the precise conditions under which the paradoxical effect occurs are difficult to establish without mathematical modeling.

In simple models consisting of only two populations (one excitatory and one inhibitory) these conditions have been mathematically derived. The paradoxical effect occurs when the networks operates in the regime known as *inhibition stabilized* (inhibition stabilized networks, ISN) in which the total the total recurrent excitation is so strong that inhibition is necessary to prevent a blow up in the activity (Murphy and Miller, 2009; Ozeki et al., 2009; Tsodyks et al., 1997). Networks, with several inhibitory populations have been recently investigated (Del Molino et al., 2017; Litwin-Kumar et al., 2016; Sadeh et al., 2017). These studies considered network models with synaptic currents small compared to neuronal rheobase currents (Gerstner et al., 2014; Lapicque 1909). However, interactions in cortex are stronger than what is assumed in these studies (Shadlen and Newsome, 1994).

Simple networks with strong interactions comprising one excitatory and one inhibitory population have been studied extensively. In a broad parameter range not requiring fine-tuning, such networks dynamically evolve into a state in which strong excitation is balanced by strong inhibition such that the *net* input into the neurons is comparable to their rheobases (van Vreeswijk and Sompolinsky, 1998, 1996). The theory of balanced networks has been developed for a variety of single neuronal models including binary neurons (van Vreeswijk and Sompolinsky, 1998, 1996), rate models (Harish and Hansel, 2015; Kadmon and Sompolinsky, 2015), leaky-integrate-and fire neurons (Hansel and Mato, 2013; Mongillo et al., 2012; Rosenbaum and Doiron, 2014; Roxin et al., 2011; Van Vreeswijk and Sompolinsky, 2005) and conductance-based models (Hansel and van Vreeswijk, 2012; Pattadkal et al., 2018).

In the present study we investigate experimentally the effects of the photostimulation of PV interneurons on the anterior lateral motor cortex (ALM) and barrel cortex (S1) of the mouse. We show that two-population network models do not suffice to account for these effects. To overcome this limitation, we develop a theory for the paradoxical effect in balanced networks that takes into account the multiplicity of GABAergic neuronal populations. Combining analytical calculations and numerical simulations, we study the responses of these networks at population and single neuron level. For two-population balanced networks it has been shown that the paradoxical effect only occurs when the network is inhibition stabilized (Pehlevan and Sompolinsky, 2014; Wolf et al., 2014). Here we show that in contrast, in four-population networks, the paradoxical effect can occur even if the network is not inhibition stabilized. We conclude with prescriptions for experiments that according to the theory can be informative about network architectures in cortex.

Results

ALM layer 5 and S1 exhibit paradoxical effect but not ALM layer 2/3

We expressed a red-shifted channelrhodopsin (ReaChR) in PV interneurons to optogenetically drive local inhibition in the barrel cortex (S1) and anterior lateral motor cortex (ALM) of awake mice (Hooks et al., 2015). We used orange light (594nm) to illuminate a large area of ALM or S1 (2 mm diameter), photostimulating a large proportion of PV interneurons (Fig. 1A). We measured the light-induced effects on neural activity using silicon probe recordings. In both brain areas, putative PCs and putative PV neurons were identified based on spike width (Methods). Neurons with wide spikes were likely mostly PCs. Units with narrow spikes were fast spiking (FS) neurons and likely expressed parvalbumin (Cardin et al., 2009; Guo et al., 2014b; Olsen et al., 2012; Resulaj et al., 2018). We investigated the responses of these neurons as a function of the photostimulation intensity in ALM layer 2/3 and layer 5, and in S1.

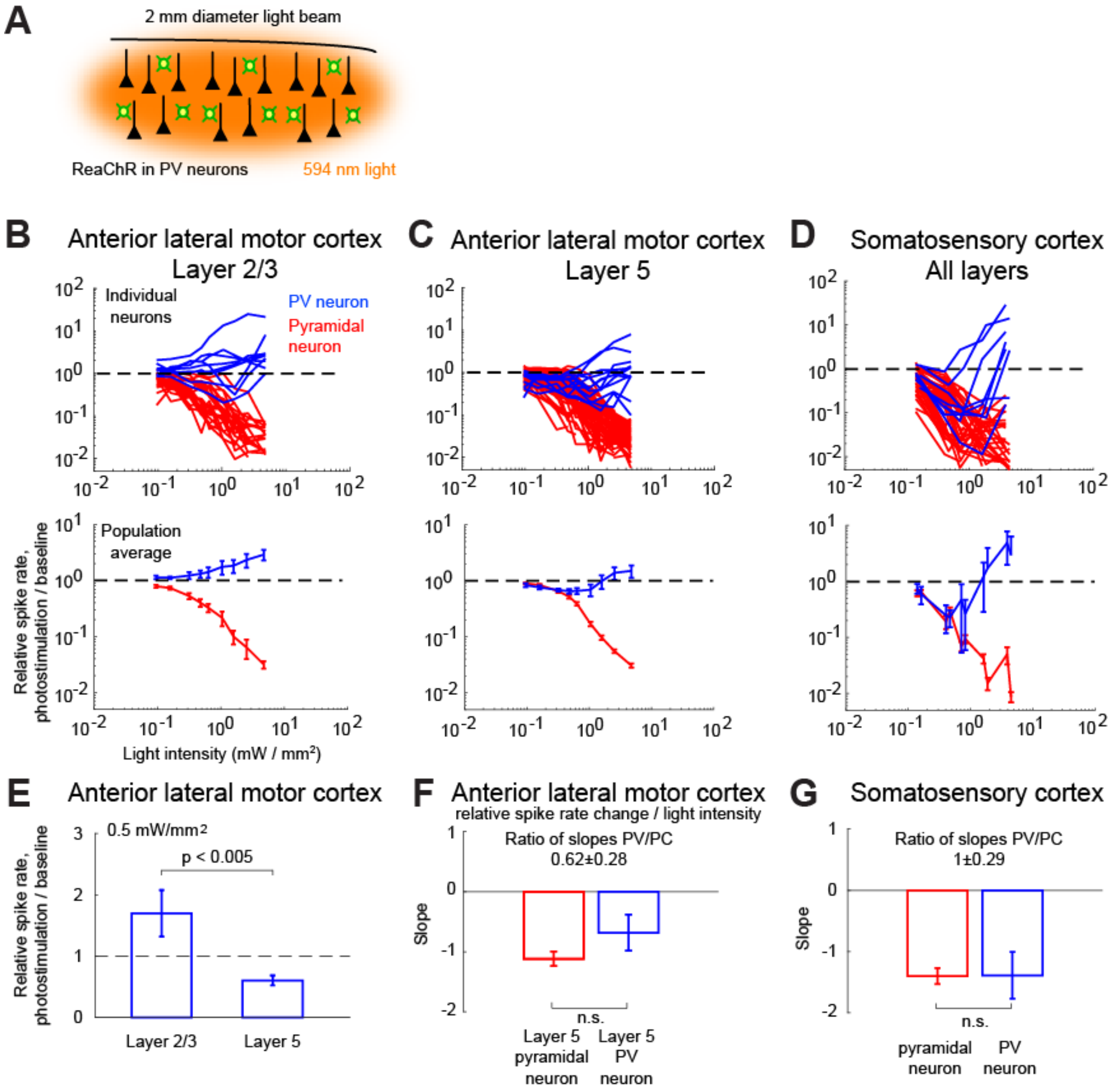


Figure 1. Effects of photostimulation of PV-positive interneurons in the mouse neocortex. A. Scheme of the experiment. B-C. Normalized spike rate as a function of laser intensity in different layers and brain areas. Top, individual neuron responses of the PCs (red) and PV (blue) neurons; bottom, population average responses. B. ALM: layer 2/3: n=26 (PCs), n=9 (PV); C. ALM layer 5: n=62 (PCs), n=12 (PV). D. S1: n=52 (PCs), n=8 (PV). Mean ± s.e.m. across neurons, bootstrap. E. Comparison of PV neurons' normalized spike rates between ALM Layer 2/3 and Layer 5 at laser intensity 0.5 mW/mm^2. F. Slope of PCs and PVs' normalized spike rate as a function of laser intensity. Data from ALM layer 5. Slopes are computed using data from 0.3 mW/mm^2 and below, before the spike rate of PV neurons begin to increase. Mean±SEM, bootstrap (Methods). G. Same as F but for data from S1. In F and G the difference between the slopes for the PC and PV populations is not significant.

We found that in all recorded layers and areas, the population average activity of the PCs decreased with the optogenetic drive (Fig. 1B, Fig. 2). In contrast in ALM, the PV population exhibited a behavior which depended on the recorded layer.

In ALM layer 2/3, the population average firing rate of PV neurons monotonically increased with the photostimulation intensity. However, individual neuron responses were heterogeneous. Most PV neurons increased their spike rates from baseline with increased photostimulation intensity. Some PV neurons initially decreased their spike rates below baseline for low light intensity.

In ALM layer 5, the response of the PV population was non-monotonic. For low laser intensity, their activity paradoxically decreased with the optogenetic drive. The slope of the normalized firing rate *v.s.* laser intensity was significantly different from zero for both the PC and PV populations (Fig. 1F). The ratio of their slopes was 0.62 ± 0.28. At high photostimulation intensity, the activity of the PV population increased. At intermediate photostimulation intensity (0.5 mW/mm^2), the response of the PV neurons was significantly different between layer 2/3 and layer 5 (Fig. 1E, $p<0.005$, unpaired t-test, two-tailed test).

Paradoxical decrease in PV neurons activity with the optogenetic drive was also observed in S1. Remarkably, the concomitant decrease of the PC and the PV population activities was proportional (Fig. 1G, ratio of slopes PV/PC, mean ± SEM; S1, 1 ± 0.29).

In both ALM layer 5 and S1, there was also a large diversity of responses. Most PV neurons decreased their activity at low photostimulation intensity. At high laser intensity (5 mW/mm^2), a fraction of PV neurons (6/12 in ALM layer 5 and 6/10 in S1) had a larger response than baseline, while the rest remained suppressed. Figure 2 shows the spike rates of PCs and PV neurons at an intermediate light intensity $(0.5\ mW.mm^{-2})$.

Network models

To assess the network mechanisms which may account for the experimental data from ALM and S1 we first considered models consisting of one excitatory and one inhibitory population. Since it is well established that cortical circuits involve a variety of inhibitory subpopulations, we later extended the theory to network models of four populations of neurons representing PCs and three subtypes of GABAergic interneurons in cortex. In all our models, neurons are described as integrate-and-fire elements. The data we seek to account for, were obtained in optogenetic experiments in which the laser diameter was substantially larger than the spatial range of neuronal interactions and comparable to the size of the area in which activity was recorded. Therefore, in all our models, we assume for simplicity that the connectivity is unstructured. We modeled the ReachR-optogenetic stimulation of the PV population as an additional external input, I_{opto}, into PV neurons. We assumed that it depends on the intensity of the laser, Γ_{opto}, as $I_{opto} = I_0 log\left(1 + \frac{\Gamma_{opto}}{\Gamma_0}\right)$ where I_0 and Γ_0 are parameters (Fig. 3-figure supplementary 1, (Hooks et al., 2015)).

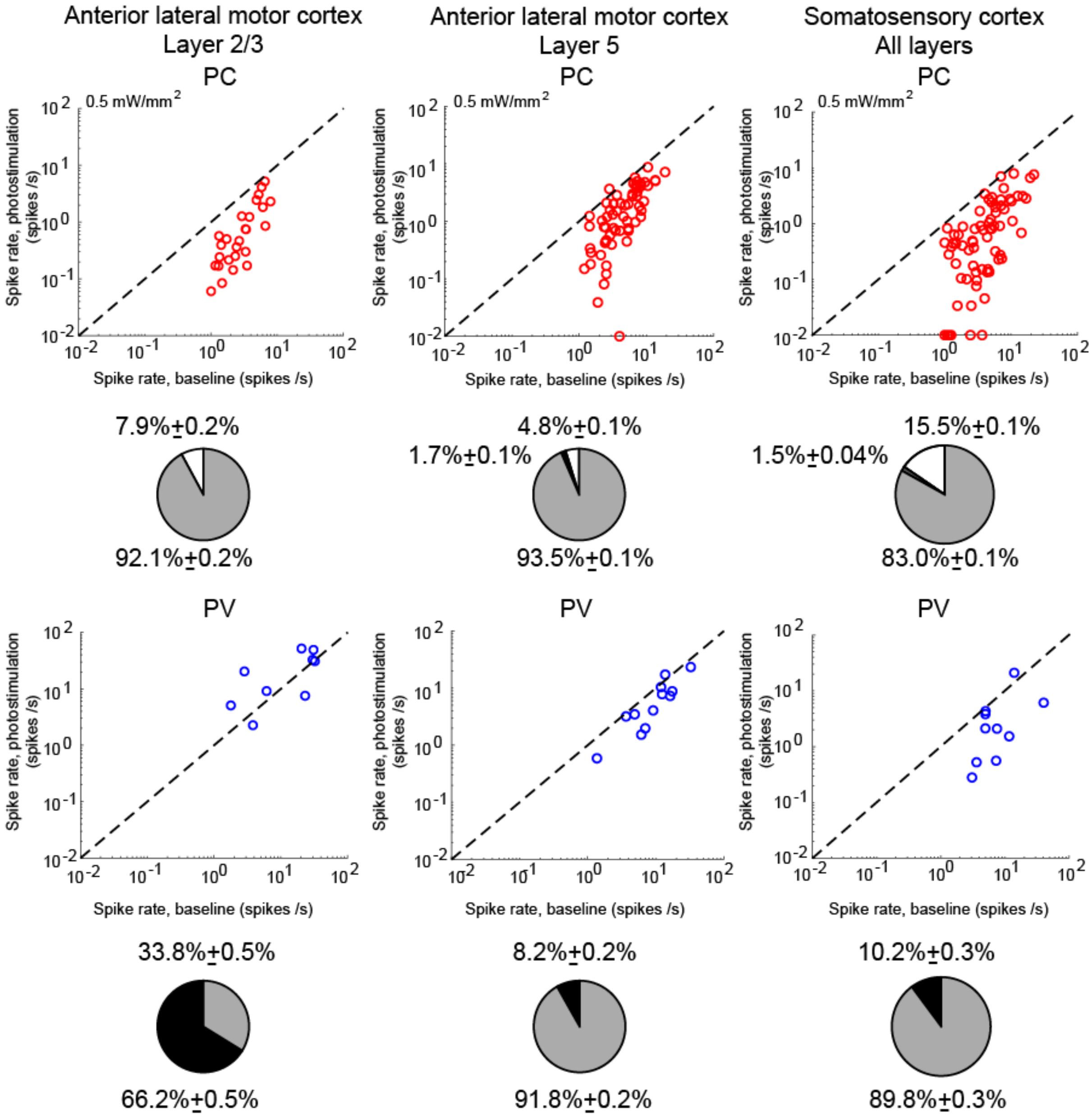


Figure 2. Spike rates of PCs (top) and PV neurons (bottom). Dots correspond to individual neurons. Laser intensity is 0.5 mW/mm^2 . Pie charts represent the fraction of neurons with different types of changes. Mean ± s.e.m. bootstrap. Black, fraction of neurons with activity increase larger than 0.1 Hz. Light gray, fraction of neurons with activity decrease larger than 0.1 Hz. Dark gray, fraction of neurons with activity change smaller than 0.1 Hz. White, fraction of neurons with activity smaller than 0.1 Hz upon PV photostimulation.

Two-population model

The two-population network is depicted in Fig. 3A. It is characterized by four recurrent interaction parameters, $J_{\alpha\beta}$, and two feedforward interaction parameters,

$J_{\alpha 0}$, $\alpha, \beta \in \{E, I\}$ (see Materials and Methods).

Results from numerical simulations of the model are depicted in Fig. 3B and C where, the dependence of the population activities normalized to baseline, are plotted against the intensity of the laser, Γ_{opto}. Figure 3B shows the response of the network where the recurrent excitation, J_{EE}, is non zero. The activity of the PV population, r_I varies non-monotonically with the laser intensity. For small intensities, r_I paradoxically decreases together with the activity of the PCs, r_E. This paradoxical effect stems from the fact that the decrease in the activity of the PCs yields a reduction in the excitation to PV neurons which is not compensated for by the optogenetic drive. As a result, the net excitation to PV neurons diminishes yielding a decrease in r_I. When r_E becomes very small, this mechanism does not operate anymore and consequently, r_I increases as Γ_{opto} is increased further. In Figure 3C, J_{EE} is zero, r_I monotonically increases with the light intensity whereas r_E monotonically decreases. For small intensities, r_I is close to a constant. It starts to increase appreciably only when $r_E \simeq 0$. Therefore, the PV response is not paradoxical.

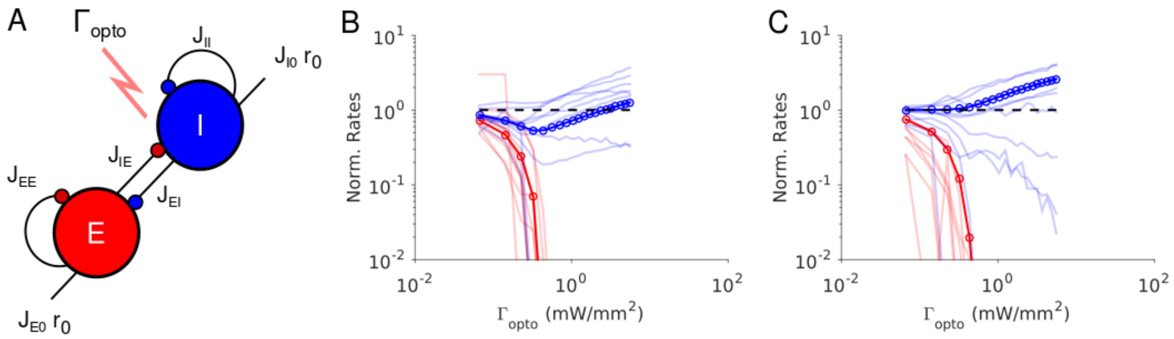


Figure 3. Paradoxical effects in the two-population model. A. The network. B-C. Responses of PCs and PV neurons normalized to baseline *vs*. the laser intensity, Γ_{opto}, for different values of the recurrent excitation, j_{EE}. B: $j_{EE} = J_{EE} / \sqrt{K} = 0.98\ \mu A\ .\ ms\ .\ cm^{-2}$, the network exhibits the paradoxical effect. C: $j_{EE} = 0$, the population activity of PV neurons is almost insensitive to small laser intensities. Red: PCs. Blue: PV neurons. Thick lines: population averaged responses. Thin lines: responses of 10 neurons randomly chosen in each population. Firing rates were estimated over $100\, s$. Parameters: $N_E = 57600$, $N_I = 19200$, $K = 500$. Other parameters as in Table 1-2. Baseline firing rates are: $r_E = 5.7\, Hz$, $r_I = 11.7\, Hz$ (B) and $r_E = 1.5\, Hz$, $r_I = 5.7\, Hz$ (C). At the minimum of r_I in (B), $r_E = 0.06\, Hz$.

Qualitatively this model seems to account for our experimental data from ALM layer 2/3, ALM layer 5 and S1. It would imply that in layer 5, J_{EE} is sufficiently large to generate the paradoxical effect, while in layer 2/3 this is not the case. On closer inspection however, there are major discrepancies between the simulation results and the experimental data. In our recordings in both ALM layer 5 and S1, the PV population activity reaches a minimum while the PCs are still significantly active: relative to baseline the activity is 40% in ALM and 25% in S1. In contrast, in the two-population model, the minimum of the PV activity is reached (Appendix 1B) when excitatory neurons are virtually completely silenced (Fig. 3B, Fig. 3-figure supplementary 2A). In fact one can show that for sufficiently large K, when r_I is minimum, the activity of the excitatory population is exponentially small in K. As a result, to account for the data one needs to assume that $K \simeq 10$.

In addition, in the experimental data the activities of the PC and PV populations in S1 decrease in equal proportions before the minimum of the PV activity (Fig. 1B). This cannot be accounted for in a two-population model unless parameters are fine-tuned (Fig. 3-figure supplementary 3). Analytical calculations (Appendix 1B) supplemented with numerical simulations show that this proportional decrease only happens when the determinant of the interaction matrix, $J_{\alpha\beta}$, is close to zero. Moreover, the external input must also be fine-tuned so that the neurons have biologically realistic firing rates (Fig. 3-figure supplementary 3).

The experimental data from ALM layer 2/3 show that for already small light intensity the activity of PV neurons increases appreciably. This is in contrast with Fig. 3C. In Fig. 3-figure supplementary 2B, we show that the two-population model can account for this feature only if the recurrent excitation is very weak in that layer and the connectivity is extremely sparse.

These discrepancies prompted us to investigate whether models with several populations of inhibitory neurons can account for our experimental data without fine-tuning. We focus on two four-population network models. Both consist of three populations representing PCs, PV and SOM neurons and a fourth population representing other inhibitory neurons. The main difference between the two models lies in the inhibitory populations from which SOM neurons receives inputs.

A four-population model with V1-like architecture (Model 1)

We first investigated the dynamics of a four-population network with an architecture that is similar to the one reported in layer 2/3 in V1 (Pfeffer et al., 2013) and S1 (Lee et al., 2013) (Fig. 4A). The model consists of four populations representing PCs, PV, SOM and VIP neurons. SOM neurons do not interact with each other (Adesnik et al., 2012; Gibson et al., 1999; Hu et al., 2011). VIP neurons only project to the SOM population (Jiang et al., 2015; Pfeffer et al., 2013). All neurons except SOM receive inputs from sources external to the network (*e.g.* thalamus) (Beierlein et al., 2003, 2000; Cruikshank et al., 2010; Ma et al., 2006; Xu et al., 2013). The same architecture was considered in (Litwin-Kumar et al., 2016).

Following (Pfeffer et al., 2013), the PV population does not project to the SOM population. Other studies have reported such a connection (Jiang et al., 2015). However, adding such a connection to Model 1 does not qualitatively affect the PC and PV responses (see Appendix 1C).

We considered parameter sets such that: 1) At baseline, the network is operating in the balanced state with all populations active; 2) the activity of the PC population decreases with the laser intensity as observed in our experiments.

Theory in the large N, K limit.

It is instructive to consider the limit in which the number of neurons in the network, N, and the average number of connections per neuron, K, go to infinity. In this limit, the analysis of the stationary state of the network simplifies (see Materials and Methods). This stems from the fact that when interactions are numerous, excitatory and inhibitory inputs are strong and only populations for which excitation is balanced by inhibition have a finite and non-zero activity. The average activities of the four populations are then completely determined by four linear equations, *the balance equations*, which reflect this balance. Solving this system of equations yields the population activities, r_α, $\alpha = E, I, S, V$, as a function of the external inputs to the network. In particular, when the laser intensity is sufficiently small, the four populations are active and their firing rates vary linearly with the current induced by the photostimulation (Appendix 1C).

Figure 4 plots the activities of the populations *vs*. the optogenetic input into PV neurons, I_{opto}, for two sets of interaction parameters. In Fig. 4B, the activity of the PV population, r_I, increases with I_{opto}. In contrast, in Fig. 4C, r_I decreases with I_{opto}: the response of the PV population is paradoxical.

To characterize for which interaction parameters the PV response is paradoxical, we consider the 4×4 susceptibility matrix $[\chi_{\alpha\beta}]$. The element $\chi_{\alpha\beta}\,(\alpha, \beta = E, I, S, V)$ is the derivative of the population activity, r_α, with respect to a small additional input, into population β, I_β. Evaluated for small I_β, $\chi_{\alpha\beta}$ characterizes by how much r_α varies with an increasing but weak extra input into population β. Its sign indicates whether r_α increases or decreases with I_β. The elements of the susceptibility matrix can be decomposed in several terms corresponding to the contributions of different recurrent loops embedded in the network (Appendix 1C). Using this decomposition one can show whether the PV response is paradoxical or not depends on the interplay between two terms. One is the gain of the disinhibitory feedback loop PC-VIP-SOM-PC and the other is the product of the recurrent excitation, J_{EE}, with the gain of the disinhibitory feedback loop VIP-SOM-VIP (Fig. 4-figure supplementary 1). Remarkably, PV neurons are not involved in these two terms. A straightforward calculation (Eq. A37) then shows that the response of PV neurons increases with I_{opto} if the recurrent excitation is sufficiently strong, namely if

$$J_{EE} > J^{*}_{EE} = J_{VE}\frac{J_{ES}}{J_{VS}} \qquad (1)$$

The denominator in J^{*}_{EE} is the strength of the connection from the SOM population to the VIP population. The numerator is the gain of the pathway which connects these two populations via the PCs. When $J_{EE} > J^{*}_{EE}$ the negative contribution of the disinhibitory loop PC-VIP-SOM-PC dominates in the expression of χ_{II}. It is the opposite when $J_{EE} < J^{*}_{EE}$. The stability of the balanced state provides other necessary conditions that the interactions must satisfy (see Materials and Methods). In particular, the determinant of the interaction matrix, J, must be positive.

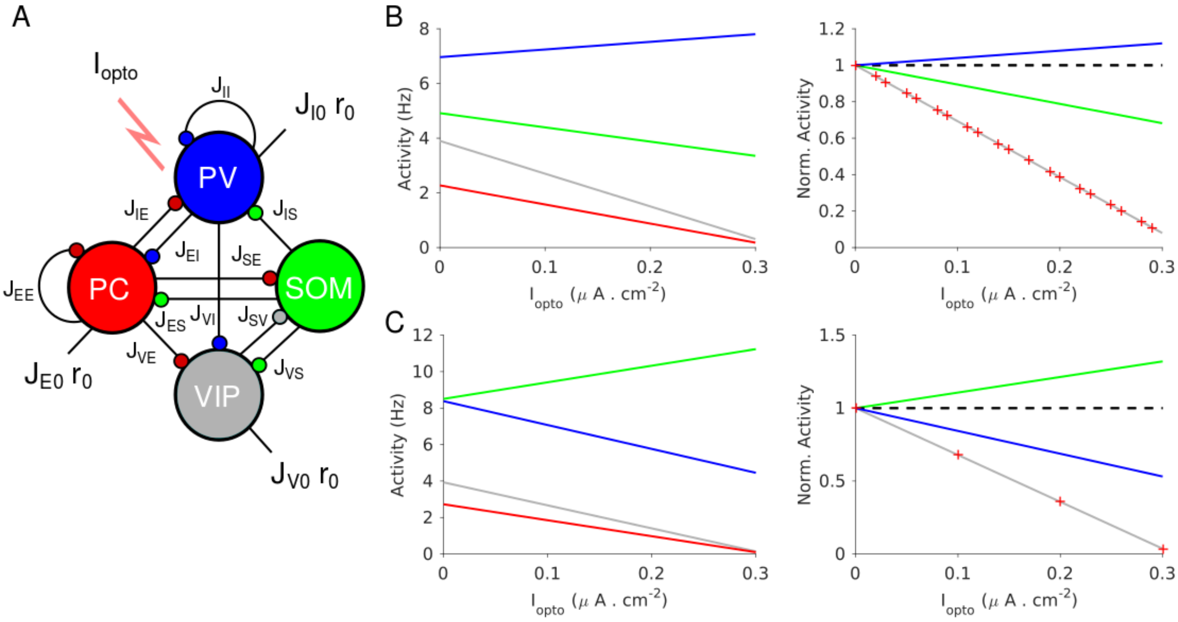


Figure 4. Population activities *vs.* I_{opto} in Model 1 in the large N, K limit. A. The network is composed of four populations representing PCs, PV, SOM and VIP neurons. The connectivity is as in (Pfeffer et al., 2013). B. Parameters as in Table 4. The activity of PV cells increases with I_{opto} while for the three other populations it decreases. C. Parameters as in Table 5. The activity of SOM neurons increases with I_{opto} while for the three other populations it decreases. Right panels in B and C: the activities are normalized to baseline.

The difference between the behaviors in Fig. 4B and 4C can now be understood as follows: in Fig. 4B, $J_{EE} > J^{*}_{EE}$ and $\chi_{II} = 1.6 > 0$, thus, r_I increases with I_{opto}; in Fig. 4C, $J_{EE} < J^{*}_{EE}$ and $\chi_{II} = -5.1 < 0$ and thus, r_I decreases. Remarkably, in both cases the activities of the PC and VIP populations normalized to baseline, are always equal (Fig. 4B-C, right panel). This is a consequence of the balance of excitatory and inhibitory inputs into the SOM population which implies that r_E and r_V are proportional (see Materials and Methods, Eq. 15.3).

In Fig. 4B, the activity of the SOM population decreases with the laser intensity. This also stems from the fact that $J_{EE} > J^*_{EE}$ (Appendix 1C, Eq. A31-A34). This qualitative behavior is therefore independent of parameter sets, provided that inequality (1) is satisfied. In contrast, for parameters for which $J_{EE} < J^*_{EE}$ the activity of the SOM population either decreases or increases with I_{opto} depending on other parameters. Moreover, it is straightforward to prove that if $J_{EE} > J^*_{EE}$, the product $\chi_{EI}\chi_{IE}$ is positive (Appendix 1C). Since we assumed that r_E decreases upon photostimulation of PV neurons, namely $\chi_{EI} < 0$, this implies that χ_{IE} is also negative. In other words, in Model 1, a non-paradoxical response of the PV population upon PV photostimulation implies that the PV activity *decreases* when PCs are photostimulated.

When I_{opto} is sufficiently large, the solution of the four balance equations will contain one or more populations for which $r_\alpha < 0$. Obviously such a solution is inconsistent. Instead, other solutions should be considered where at least one population has a firing rate which is zero and the firing rates of the other populations is determined by a new system of linear equations with lower dimensions (see Materials and Methods, Appendix 1C). Consistency requires that in these solutions the net input is hyperpolarizing for the populations with $r_\alpha = 0$. As a consequence, the network population activities are in general piecewise linear in I_{opto} (Fig. 4-figure supplementary 2).

The large N, K analysis provides precious insights into the dynamics of networks with reasonable size and connectivity. In particular, we will show that the criterion for the paradoxical effect, Eq. (1), remains valid up to small corrections. Although it is possible to treat analytically the dependence of r_α on I_{opto} for finite K, these calculations are very technical and beyond the scope of this paper. Instead here, we proceed with numerical simulations.

Numerical simulations for $J_{EE} > J^*_{EE}$

Figure 5 depicts the results of our numerical simulations of Model 1 for the same parameters as in Fig. 4B (see Materials and Methods, Table 3-4). The response of PV neurons is non-paradoxical: the activity of the PV population increases monotonically with Γ_{opto} in the whole range (Fig. 5A). Concurrently, the population activities of PC, SOM and VIP neurons monotonically decrease with Γ_{opto} (Fig. 5A-B). For sufficiently large Γ_{opto}, PCs become very weakly active and the SOM and VIP populations dramatically reduce their firing rates. The variations with Γ_{opto} of r_E, r_I, r_S and r_V are robust to changes in the average connectivity, K (Fig. 5-figure supplementary 1) and in qualitative agreement with the predictions of the large N, K limit (Fig. 4B, Appendix 1C, Fig. 4-figure supplementary 2).

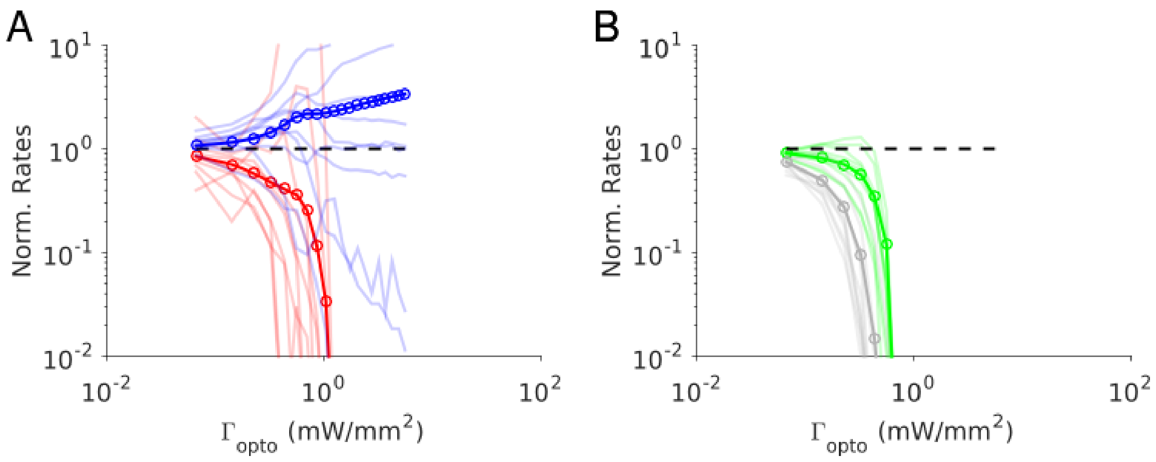


Figure 5. Numerical simulations of Model 1 for $J_{EE} > J^*_{EE}$. Responses of the neurons normalized to baseline *vs*. the intensity of the laser, Γ_{opto}. A. Activities of PCs and PV neurons: the PV response is not paradoxical. B. Activities of SOM and VIP neurons. Color code as in Fig. 4. Thick lines: population averaged responses. Thin lines: responses of 10 neurons randomly chosen in each population. Firing rates were estimated over $100\,s$. Parameters: $K = 500$, $N = 76800$. Other parameters as in Table 3-4. The baseline activities are: $r_E = 3.3\ Hz$, $r_I = 6.5\ Hz$, $r_S = 5.9\ Hz$, $r_V = 3.5\ Hz$.

To test the robustness of our results with respect to changes in the interaction strengths, we generated 100 networks with $J_{\alpha\beta}$ chosen at random within a range of $\pm 10\%$ of those of Fig. 4B. All the networks exhibited a balanced state which was stable with respect to slow rates fluctuations in the large N, K limit. We simulated those networks with $K = 500$ and computed the population activity at baseline and for $\Gamma_{opto} = 0.07 mW.mm^{-2}$. For all these networks, the results were consistent with the one of the control set: for $\Gamma_{opto} = 0.07 mW.mm^{-2}$, r_I was larger and r_E, r_S, r_V were smaller than baseline (Fig. 5-figure supplementary 2). However, a small percentage of these networks (10%) exhibited oscillations with at most an amplitude 20% of their mean in the firing rates. Apart from that, the results were robust to changes in $J_{\alpha\beta}$.

In contrast to what happens in the large N, K limit (Fig. 4B, right panel), in the results depicted in Fig. 5 the activity of the PC and VIP populations are not proportional. Moreover, in the large K limit, PC and VIP neurons are inactivated before the SOM population is. For $K = 500$, VIP is the first population to be silenced followed by the SOM and finally the PC population. Simulations with increasing values of K show that these differences are due to substantial finite K effects (Fig. 5-figure supplementary 1).

Figure 5 also depicts the changes in the firing rates (normalized to baseline) with Γ_{opto} for several example neurons. These changes are highly heterogeneous across neurons within each population. Whereas the population average varies monotonically, individual cells activity can either increase or decrease and the response can even be non-monotonic with Γ_{opto}.

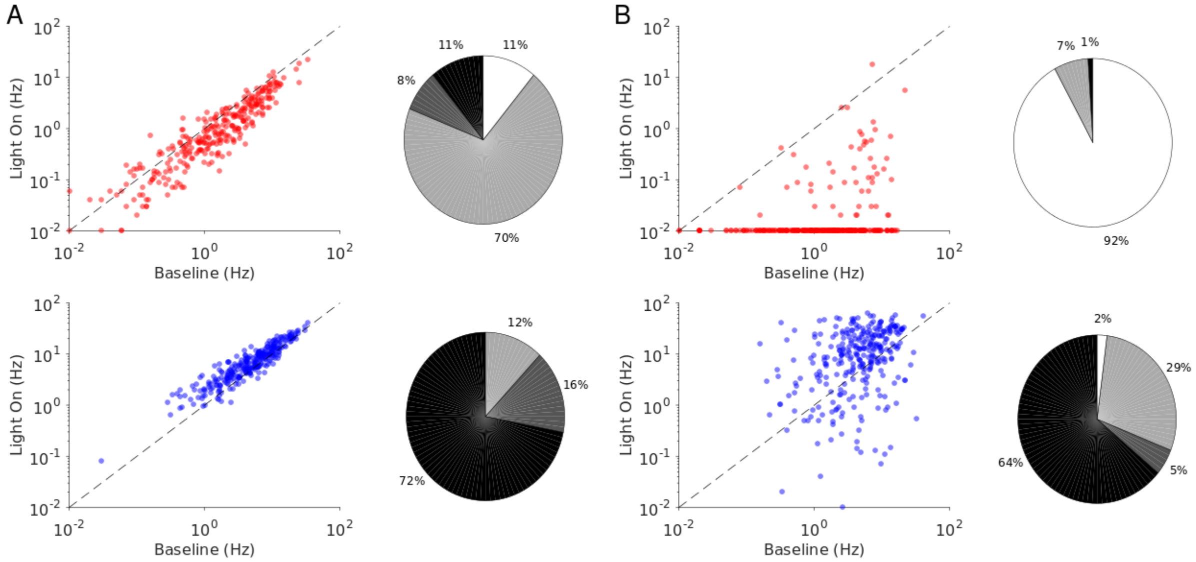


Figure 6. Single neuron firing rates in the PC and PV populations upon PV activation for two values of the light intensity (Model 1 with $J_{EE} > J^*_{EE}$). A. Single neuron firing rates at baseline *vs*. at $\Gamma_{opto} = 0.5\, mW\,.\,mm^{-2}$. B. Same for $\Gamma_{opto} = 1\, mW\,.\,mm^{-2}$. Top: PCs (red). Bottom: PV neurons (blue). Scatter plots of 300 randomly chosen PC and PV neurons. Pie charts for the whole population. The pie charts show the fraction of neurons which increase (black) or decrease (light gray) their activity compared to baseline. Dark gray: Fraction of neurons with relative change smaller than $0.1\, Hz$. White: fraction of neurons with activity smaller than $0.1\, Hz$ upon PV photostimulation.

The heterogeneity in the single neuronal responses are also clear in Fig. 6A-B that plots, for two different light intensities, the perturbed firing rate *vs*. baseline for PCs and PV neurons. Remarkably, in both populations a significant fraction of neuron exhibits a response which is incongruous with the population average. The pie charts in Fig. 6 depict the fraction of PCs and PV neurons which increased, decreased, or did not change their firing rates. The fraction of neurons whose activity is almost completely suppressed, is also shown. Remarkably, even for $\Gamma_{opto} = 1.0 mW.mm^{-2}$, some of the PCs show an activity increase. Moreover, the fraction of PV neurons whose firing rate increases is less for $\Gamma_{opto} = 1.0 mW.mm^{-2}$ than $\Gamma_{opto} = 0.5 mW.mm^{-2}$. It should be noted that in the model all PV neurons receive the same optogenetic input, therefore, the heterogeneity in the response is not due to whether or not the PV neurons were "infected". This heterogeneity is solely due to the randomness in the connectivity.

Numerical simulations for $J_{EE} < J^*_{EE}$

Figure 7 depicts the results of our numerical simulations of Model 1 when $J_{EE} < J^*_{EE}$. Parameters are the same as in Fig. 4C (see Materials and Methods, Table 3-5). The population activities of PCs and VIP neurons, r_E and r_V, decrease monotonically with the laser intensity, Γ_{opto}. Conversely, the variations of the activities of the PV and SOM populations, r_I and r_S, are non-monotonic with Γ_{opto}. For small light intensities, r_I decreases and then abruptly increases with larger Γ_{opto}; r_S exhibits the opposite behavior. Remarkably, when r_I is minimum, r_S is maximum for nearly the same value of Γ_{opto}. We show in Fig. 7 S4 that this proportional decrease only happens in a small region of parameter space when the determinant of the interaction matrix, $J_{\alpha\beta}\,\epsilon_\beta$, is close to zero.

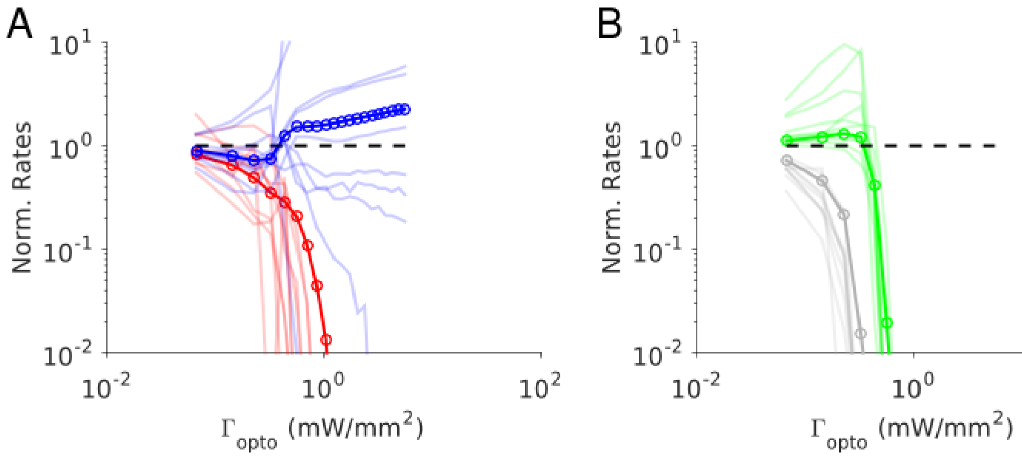


Figure 7. Numerical simulations of Model 1 for $J_{EE} < J^*_{EE}$. Responses of the neurons normalized to baseline *vs.* the intensity of the laser, Γ_{opto}. A. Activities of PCs and PV neurons: the PV response is paradoxical. B. Activities of SOM and VIP neurons. Color code as in Fig. 4. Thick lines: population averaged responses. Thin lines: responses of 10 neurons in each population. Firing rates were estimated over $100\,s$. Parameters: $K = 500$, $N = 76800$. Other parameters as in Table 3-5. The baseline activities are: $r_E = 4.8\,Hz$, $r_I = 11.2\,Hz$, $r_S = 7.1\,Hz$, $r_V = 5.3\,Hz$.

This behavior is qualitatively similar to the one derived in the large N, K limit (Fig. 4-figure supplementary 3). As suggested by the large N, K analysis, the paradoxical response of the PV neurons in the simulations, is driven by the positive feedback loop PC-VIP-SOM-PC (Fig. 4-figure supplementary 1). Remarkably, when the activity of the PV neurons is minimum, the PCs are still substantially active (40% of baseline level). This is due to finite K corrections to the large N, K predictions (Fig. 7-figure supplementary 1). These corrections are strong and scale as $\frac{1}{\sqrt{K}}$ (Appendix 1C). Indeed, even for K as large as 2000, r_E is still 25% of the baseline when r_I is minimum.

We checked the robustness of these results with respect to changes in the interaction parameters as we did for $J_{EE} > J^*_{EE}$. We found that for small light intensity all the 100 simulated networks were operating in the balanced state and exhibited the paradoxical effect (Fig. 7-figure supplementary 2).

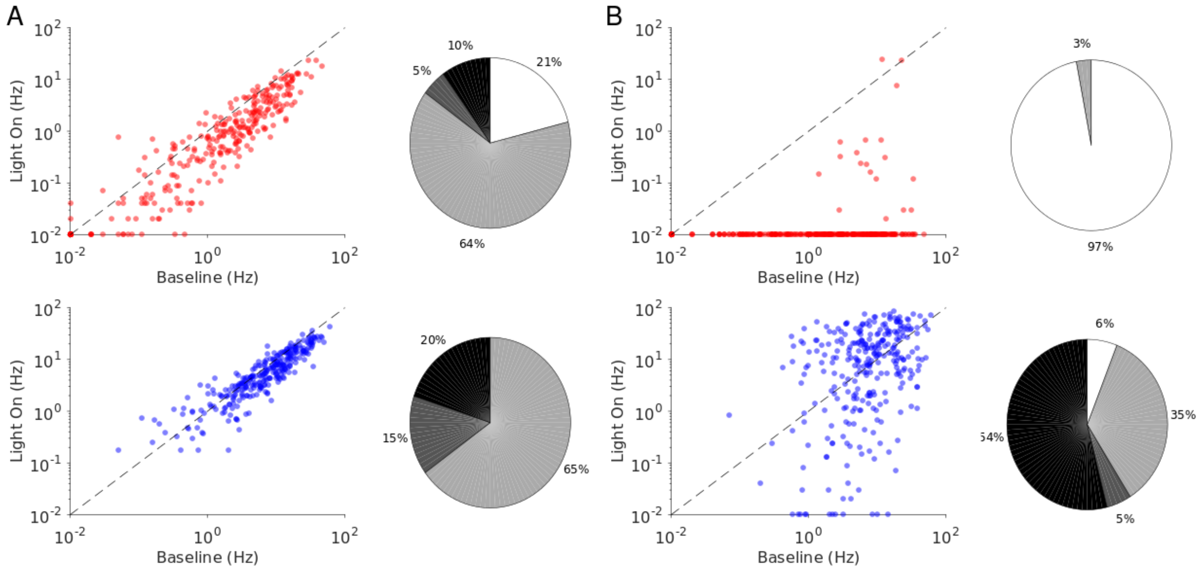


Figure 8. Single neuron firing rates in the PC and PV populations upon PV activation for two values of the light intensity (Model 1 with $J_{EE} < J_{EE}^*$). A. Single neuron firing rates at baseline *vs*. at $\Gamma_{opto} = 0.5\ mW\ .\ mm^{-2}$. B. Same for $\Gamma_{opto} = 1\ mW\ .\ mm^{-2}$. Top: PCs. Bottom: PV neurons. Scatter plots of 300 randomly chosen PC and PV neurons. Pie charts for the whole population. Firing rates were estimated over $100\ s$ simulation time. Neurons with rates smaller than $0.01\ Hz$ are plotted at $0.01\ Hz$. Color code as in Fig. 6. Parameters as in Fig. 7.

Finally, the single neuron responses are highly heterogeneous. Figure 8 plots the perturbed activities of PCs and PV neurons *vs*. their baseline firing rates for two light intensities. In Fig. 8A, the PV response is paradoxical. This is not the case in Fig. 8B. Interestingly, the fraction of PV neurons incongruous with the population activity is larger for $\Gamma_{opto} = 0.5 mW.mm^{-2}$ than for $\Gamma_{opto} = 1.0 mW.mm^{-2}$. For both light intensities the activity of almost all the PCs is decreased.

Four-population network: Model 2

In S1, in the range of laser intensities in which the PV response is paradoxical, the decrease of the PC and PV activity is proportional. This feature of the data can be accounted for in Model 1 but only with a fine-tuning of the interaction parameters (Fig. 7-figure supplementary 3 and Fig. 7-figure supplementary 4). This prompted us to investigate whether a different architecture could account robustly for this remarkable property. Our hypothesis is that this property is a direct consequence of the balance of excitation and inhibition.

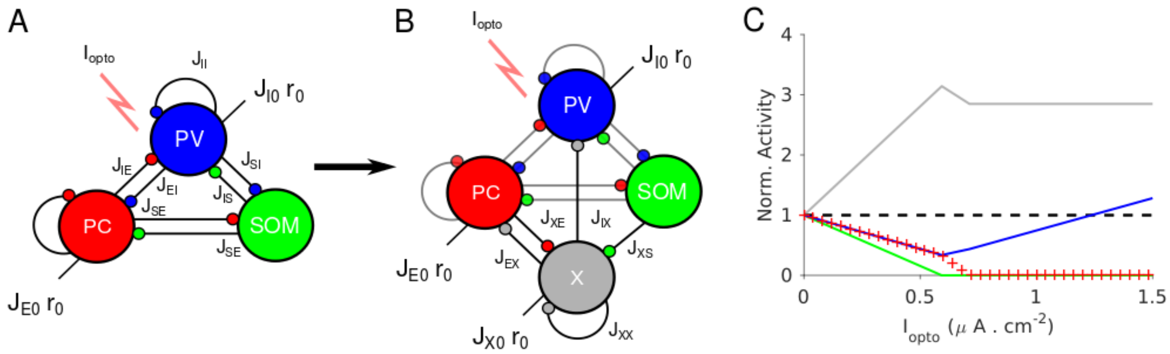


Figure 9. Network models with proportional change in the PC and PV activities upon photostimulation of the PV population. A. A three-population network consisting of PCs, PV and SOM neurons. SOM neurons only receive projections from the PC and PV populations. B. Model 2 consists of four populations: PC, PV, SOM and an unidentified inhibitory population, X. The population X projects to the PC, the PV population and to itself. The PC population projects to X. C. Population activities normalized to baseline *vs.* I_{opto} in the large N,K limit. PC and PV populations decrease their activity with I_{opto} in a proportional manner. Parameters as in Table 7. Baseline firing rates are: $r_E = 3.0\,Hz$, $r_I = 6.7\,Hz$, $r_S = 6.4\,Hz$, $r_X = 3.8\,Hz$.

Theory in the large N,K limit

We first considered the three-population model depicted in Fig. 9A. It consists of the PC, PV and SOM populations. SOM neurons receive strong inputs from PCs and PV neurons, but do not interact with each other and do not receive feedforward external inputs. In the large N,K limit, the balance of excitation and inhibition of the SOM population reads (see Materials and Methods, Eq. 16.2)

$$J_{SE}r_E - J_{SI}r_I = 0 \qquad (3)$$

Therefore, the activities of the PC and PV populations are always proportional. However, as we show in (Appendix 1D) a three-population network with such an architecture cannot exhibit the paradoxical effect.

We therefore considered a network model in which a third inhibitory population, referred to as 'X', is added without violating Eq. (3) (Fig. 9B). This requires that SOM neurons do not receive inputs from X neurons (Appendix 1D). This network exhibits the paradoxical effect if and only if $J_{SE}J_{EX}J_{XS} > J_{XX}J_{ES}J_{SE}$, that is if the gain of the positive feedback loop, SOM-X-PC-SOM, is sufficiently strong (Appendix 1D). Obviously, this condition simplifies and reads

$$J_{EX}J_{XS} > J_{XX}J_{ES} \qquad (4)$$

Remarkably, this inequality does not depend on J_{EE}. This is in contrast to what happens in Model 1 where the paradoxical effect occurs only if J_{EE} is small enough (see Eq. (2)).

As in Model 1, we further required that the activity of the PC population increases with its feedforward external input. This adds the constraint (Appendix 1D):

$$J_{IX}J_{XS} > J_{XX}J_{IS} \tag{5}$$

Eqs. (3-5) do not depend on J_{XI}. For simplicity, we take $J_{XI} = 0$ and refer to the resulting architecture as Model 2.

In Fig. 9C, the slope of the PV population activity changes from negative to positive while PCs are still active. This is because if SOM neurons are completely suppressed, the loop SOM-X-PC-SOM which is responsible for the paradoxical effect, is not effective anymore. Interestingly, the analytical calculations also show that, when the SOM population activity vanishes, the activity of the X population is maximum. Since the SOM population is inactive before PCs, there is a range of laser intensities where the activity of the latter keeps decreasing while the activity of the PV population increases. Once PCs are inactive, the activity of the X population do not vary with I_{opto}. This is because then they only receive a constant feedforward excitation from outside the network which is balanced by their strong recurrent mutual coupling, J_{XX}.

Simulations for finite K

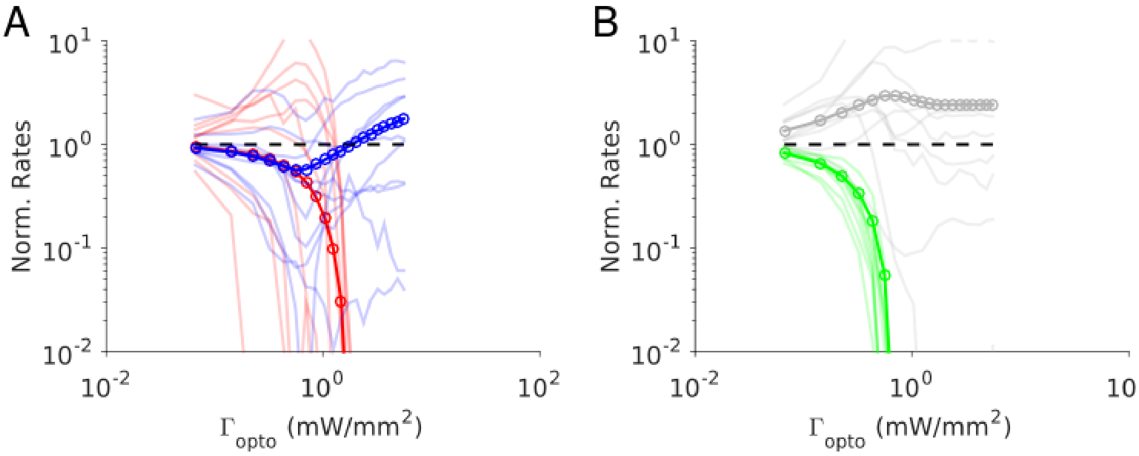


Figure 10. Numerical simulations of Model 2. Responses of the neurons normalized to baseline *vs.* the intensity of the laser, Γ_{opto}. A. Activities of PCs and PV neurons: for small Γ_{opto}, the PV response is paradoxical and the suppression of the PC and PV population activities relative to baseline are the same. B. Activities of SOM and X neurons. Color code as in Fig. 9. Thick lines: population averaged responses. Thin lines: responses of 10 neurons randomly chosen chosen in each population. Firing rates were estimated over $100\ s$. Parameters: $K = 500$, $N = 76800$. Other parameters as in Table 6-7. The baseline activities are: $r_E = 4.2\ Hz$, $r_I = 6.8\ Hz$, $r_S = 7.0\ Hz$, $r_X = 3.9\ Hz$.

These features are also observed in our simulations depicted in Fig. 10. For small laser intensities, the network exhibits a paradoxical effect where the activities of the PC and PV populations decrease with Γ_{opto} and in a proportional manner (Fig. 10A), until the SOM neurons become virtually inactive (Fig. 10B). At that value, r_I is minimum and r_X is maximum. For larger Γ_{opto}, r_I increases while r_E keeps decreasing and is still substantial. After r_E has vanished, r_X saturates but r_I continues to increase. All these results are robust to changes in the connectivity, K (Fig. 10-figure supplementary 1) as well as to changes in the interaction parameters (Fig. 10-figure supplementary 2). Single neuron responses are more heterogeneous than in the experimental data. It should be noted however that we did not tune parameters to match the experimental heterogeneity.

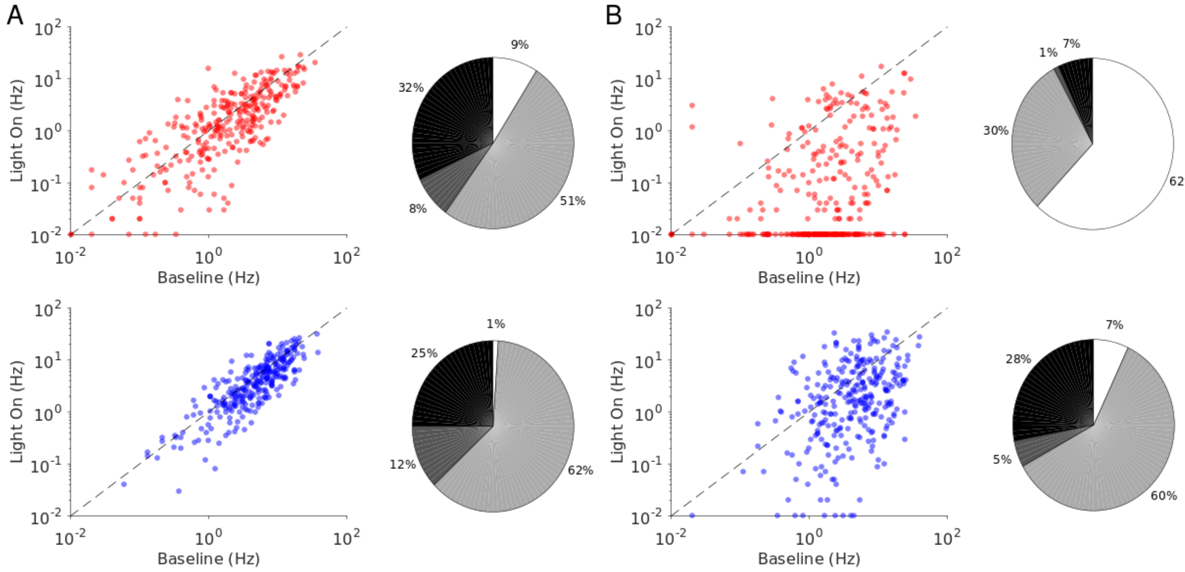


Figure 11. Single neuron firing rates in the PC and PV populations upon PV activation for two values of the light intensity (Model 2). A. Single neuron firing rates at baseline *vs.* at $\Gamma_{opto} = 0.5\, mW \, . \, mm^{-2}$. B. Same for $\Gamma_{opto} = 1\, mW \, . \, mm^{-2}$. Top: PCs. Bottom: PV neurons. Scatter plots of 300 randomly chosen PC and PV neurons. Pie charts for the whole population. Firing rates were estimated over $100\, s$. Neurons with rates smaller than $0.01\, Hz$ are plotted at $0.01\, Hz$. Color code as in Fig. 6. Parameters as in Fig. 10.

Discussion

We studied the response of cortex to optogenetic stimulation of parvalbumin positive (PV) neurons and provided a mechanistic account for it. We photostimulated the PV interneurons in layer 2/3 and layer 5 of the mouse anterior motor cortex (ALM). In layer 2/3 photostimulation increased PV activity and decreased the response of the principal cells (PCs) on average. In contrast, in layer 5 the response of the PV population was paradoxical: *both* PC and PV activity decreased on average. This is similar to what we found in the mouse somatosensory cortex (S1) (Li et al., 2019) . To account for these results, we first investigated the dynamics of networks of one excitatory and one inhibitory population of spiking neurons. We showed that two-population network models of strongly interacting neurons do not fully account for the experimental data. This prompted us to investigate the dynamics of networks consisting of more than one inhibitory population.

We considered two network models both consisting of one excitatory and three inhibitory populations. Interneurons are known to be unevenly distributed throughout the cortex. For instance, SOM neurons have been reported to be most prominent in layer 5 whereas VIP neurons are mostly found in layer 2/3 (Tremblay et al., 2016). Instead of giving a complete description of these layers and all neuronal populations they include, we propose here models with the minimal number of inhibitory populations that can account for the data.

The three inhibitory populations in Model 1 represent PV, somatostatin positive (SOM) and vasoactive intestinal peptide (VIP) interneurons with a connectivity similar to the one reported in primary visual cortex (Pfeffer et al., 2013) and S1 layer 2/3 (Lee et al., 2013). In Model 2, the first two inhibitory populations likewise represent PV and SOM neurons and the third population, denoted as X, represents an unidentified inhibitory subtype. The main difference with Model 1 is that here, the third population does not project to SOM neurons.

Depending on network parameters, the response of PV neurons in Model 1 can be paradoxical or not. To have equal relative suppression of the PCs and PV activities, however, interaction parameters have to be fine-tuned. In Model 2, the relative changes in the PC and PV activity are the same independent of interaction parameters.

For a two-population network, the paradoxical effect only occurs when it is inhibition stabilized (Pehlevan and Sompolinsky, 2014; Wolf et al., 2014). This is because the mechanism requires strong recurrent excitation. In the four-population networks we studied, however, the mechanism responsible for paradoxical effect is different. It involves a disinhibitory loop. In fact, strong recurrent excitation prevents the paradoxical effect in these networks. Therefore, the observation of the paradoxical effect upon PV photo-excitation is not a proof that the network operates in the ISN regime.

Strong vs. weak interactions

Cortical networks consist of a large number (N) of neurons each receiving a large number of inputs (K). Because N and K are large, one expects that a network behaves similar to a network where N and K are infinite. In this limit the analysis is simplified and the mechanisms underlying the dynamics are highlighted. When taking the large K limit one needs to decide how the interaction strengths scale with K. Two canonical scalings can be used: in one the interactions scale as $\frac{1}{K}$ (Hansel and Sompolinsky, 1992; Hennequin et al., 2018; Knight, 1972; Rubin et al., 2015), in the other as $\frac{1}{\sqrt{K}}$ (Darshan et al., 2017; Renart et al., 2010; Rosenbaum et al., 2017;

van Vreeswijk and Sompolinsky, 1996). These differ in the strength of the interactions. For instance, for $K = 900$ interactions are weaker by a factor 30 in the first scaling than in the second. Importantly, these two scalings give rise to qualitatively different dynamical regimes.

When interactions are strong, the excitatory and inhibitory inputs are both very large (of the order of $K.\frac{1}{\sqrt{K}} = 1$). They, however, dynamically *balance* so that the temporal average of the net input and its spatial and temporal fluctuations are comparable to the rheobase (Van Vreeswijk and Sompolinsky, 2005; van Vreeswijk and Sompolinsky, 1998), Appendix 1A). In this *balanced regime,* the average firing rates of the populations are determined by a set of linear equations: the "balance equations". These do not depend on the neuronal transfer function. For large but finite K, the network operates in an approximately balanced regime. In this regime, the population activities are well approximated by the balance equations, interspike intervals are highly irregular and firing rates are heterogeneous across neurons.

When the interactions are weak, excitatory and inhibitory inputs are both comparable to the rheobase even when K is large, but their spatial and temporal fluctuations vanish as K increases. The activity of the network is determined by a set of coupled non-linear equations which depends on the neuronal transfer function. For large but finite K, the firing of the neurons is weakly irregular and heterogeneities mostly arise from differences in the intrinsic properties of the neurons.

In which of these regimes does cortex operate *in-vivo*? This may depend on the cortical area and on whether the neuronal activity is spontaneous or driven (*e.g.* sensory, associative, or motor related). There are, however, several facts indicating that the approximate balanced regime may be ubiquitous. Many cortical areas exhibit highly irregular spiking (Shinomoto et al., 2009) and heterogeneous firing rates (Hromádka et al., 2008; Roxin et al., 2011). Excitatory and inhibitory postsynaptic potentials (PSPs) are typically of the order of 0.2 to $2mV$ or larger (Levy and Reyes, 2012; Ma et al., 2012; Pala and Petersen, 2015; Seeman et al., 2018). Model networks with PSPs of these sizes and reasonable number of neurons and connections exhibit all the hallmarks of the balanced regime (Amit and Brunel, 1997; Argaman and Golomb, 2018; Hansel and Mato, 2013; Hansel and van Vreeswijk, 2012; Lerchner et al., 2006; Pattadkal et al., 2018; Pehlevan and Sompolinsky, 2014; Rao et al., 2019; Roudi and Latham, 2007; Roxin et al., 2011; Van Vreeswijk and Sompolinsky, 2005). Moreover, there is experimental evidence of co-variation of excitatory and inhibitory inputs into cortical neurons (Haider et al., 2006; Shu et al., 2003). Finally, in cortical cultures synaptic strengths have been shown to approximately scale as $\frac{1}{\sqrt{K}}$ (Barral and D Reyes, 2016). Therefore in this paper we focused on cortical network models in which interactions are strong, *i.e.* of the order of $\frac{1}{\sqrt{K}}$.

Model 1 accounts for the responses in ALM layer 2/3 and layer 5

In Model 1, whether the network exhibits a paradoxical effect depends on the value of the ratio $\rho = \frac{J_{EE}}{J_{EE}^*}$ where $J_{EE}^* \equiv J_{VE}\frac{J_{ES}}{J_{VS}}$. Here, $J_{\alpha\beta}$, $\alpha, \beta \in \{E, S, V\}$, is the strength of the connection from population β to population α. When $\rho > 1$, the PV response is non-paradoxical and its activity increase can be substantial well before suppression of the PC activity. On the other hand when $\rho < 1$, the PV response is paradoxical and the PV activity reaches its minimum for light intensities at which the PCs are still substantially active.

In ALM layer 2/3, the activity of the PV population increases with the light intensity while the activity of the PC decreases on average. Remarkably, our experiments showed that the increase in the PV activity was already substantial for small light intensities, where the PCs were still significantly active. In ALM layer 5 the activity of the PV population initially decreased with the light intensity together with the activity of the PC population. As the light intensity is further increased, the PV activity reaches a minimum after which it increases. At this minimum, the PC activity is still substantial.

Thus, Model 1 accounts for our experimental findings in ALM layer 2/3 provided that J_{EE} is sufficiently large. It accounts for the paradoxical effect in layer 5 provided that J_{EE} is sufficiently small. Note that this does not mean that J_{EE}, is larger in the former layer as compared to the latter. The interactions J_{VE}, J_{ES} and J_{VS} are likely to be layer dependent (Jiang et al., 2015) and therefore so is the value of J_{EE}^*.

Model 2 accounts for the paradoxical effect in S1 while Model 1 would require fine-tuning

Similar to ALM layer 5, the PV response in S1 is paradoxical. Remarkably however, in S1 the relative suppression of the PC and PV activities is the same for low light intensity. Model 1 can account for this feature only when the interaction parameters are fine-tuned. In contrast, in Model 2 the co-modulation of the PC and PV activities stems from the architecture and therefore occurs in a robust manner. Furthermore, it can equally well account for the fact that in S1 the PV activity reaches its minimum when the PC population is active.

Note that in ALM layer 5 the difference between the slopes of the PC and PV population activities is not significantly different (p>0.05). Therefore, we cannot exclude that Model 2 describes ALM layer 5.

The main difference between Model 1 and Model 2 is that in Model 1, the third inhibitory population (VIP) projects to SOM neurons while in Model 2, the third population (X) does not. This suggests that population X is not the VIP population. For example, X could be chandelier cells that do not express the PV marker (Jiang et al., 2015) Alternatively, population X could describe the *effective* interaction of several inhibitory populations with PC and PV neurons.

Models 1 and 2 account for the heterogeneity of single neuron responses

The responses of PCs and PV neurons in the experimental data are highly heterogeneous across cells. Indeed in ALM layer 5 and S1, PV neurons on average show a paradoxical response but at the single neuron level the effect of the laser stimulation is very diverse. Moreover, the firing rate of a neuron can vary monotonically or non-monotonically with the laser intensity. For instance, when stimulated, the firing rates of many PV neurons increase, although, on average the activity is substantially smaller than baseline. Conversely, for some PV neurons the paradoxical effect is so strong that the laser completely suppresses their activity.

We observed an even larger diversity in single neuron responses in our simulations of Model 1 and 2. We should emphasize that in the simulated networks all the neurons were identical and the cells in the same population received the same feedforward constant external input. The only possible source of heterogeneity therefore comes from the randomness in the network connectivity. The effect of this randomness on the network recurrent dynamics is however non-trivial: one may think that the effect of the fluctuations in the number of connections from neuron to neuron should average out since in the models the number of recurrent inputs per neuron is large ($K = 500$ or more). This is not what happens because in our simulations populations which are active operate in the balanced excitation/inhibition regime (Roxin et al., 2011; van Vreeswijk and Sompolinsky, 1998, 1996). In this state, relatively small homogeneity in the number of connections per neuron is amplified to a substantial inhomogeneity in the response. Thus, strong heterogeneity in the response of neurons is not a prima facie evidence for the heterogeneity of the level of Channelrhodopsin expression in the cells nor is it for the diversity of the single neuron intrinsic properties.

Limitations

We give here a qualitative account for the mechanisms underlying the responses of different cortical areas to optical stimulation. A *quantitative* analysis of the data, in particular of the heterogeneity is beyond our scope. Such an analysis would require a much larger number of PV neurons. Moreover, it would necessitate the use of more complicated neuronal models making the mathematical analysis intractable, limiting the investigation to simulations only and thus obscuring the mechanisms.

In our experiments we expressed ReaChR in all PV neurons and in all layers in ALM. In particular, all PV neurons in layer 2/3 and layer 5 were simultaneously affected by the photostimulus. Principal cells in layer 2/3 project to layer 5 and receive feedback from the latter (Hooks et al. 2011, Naka and Adesnik, 2016). Interlaminar interactions are likely to also contribute to the effect of the photostimulation.

In our models we did not take into account such interactions. Including strong connections from layer 2/3 PCs to neurons in layer 5 and/or feedback connections from layer 5 neurons to layer 2/3, could alter our interpretations. In the absence of data that reveal the nature of interlaminar interactions, extending our model to incorporate these is impractical given the large number of parameters to vary. Experiments in ALM and S1 where the optogenetic marker is expressed in only one layer at a time would constraint models which include interlaminar interactions and facilitate their analysis (Moore et al., 2018).

There is a large amount of experimental evidence indicating that different synapses can exhibit diverse dynamics depending on their pre and postsynaptic populations (Ma et al., 2012). For instance, recent studies have shown that PCs to PV synapses are depressing while the PCs to SOM synapses are highly facilitating (Karnani et al., 2016; Xu et al., 2013). Synaptic facilitation and depression mechanisms could give rise to dynamics which will make the network responses depend on the duration of the photostimulation. Here, we did not take into account short term plasticity. Mice neocortex mostly comprises PV, SOM and 5HT3aR expressing interneurons. There is a growing amount of experimental evidence indicating that these populations include different subtypes which may have distinct connectivity patterns (Naka and Adesnik, 2016; Nigro et al., 2018; Tremblay et al., 2016). In the present work, we only considered three populations of identical interneurons: PV, SOM and VIP or X. As the number of populations increases, the number of interaction parameters increases quadratically, making it a great challenge to uncover even simple mechanisms that could underlie the network responses.

Comparison with previous theoretical work

The paradoxical effect was first described in (Tsodyks et al., 1997) and (Ozeki et al., 2009) for weak interactions using coarse grained two-population rate models (Wilson and Cowan, 1972). These models were extended in (Rubin et al., 2015) to a spatially structured network to explain center-surround interactions and other contextual effects in primary visual cortex. They found that these effects can be accounted for if the neuronal transfer function is supralinear and the network is operating in the inhibition stabilized regime (ISN). With supralinear transfer functions, whether or not the network exhibits a paradoxical effect depends on the background rate of the inhibitory neurons. These models were further extended by (Litwin-Kumar et al., 2016) to networks consisting of PC, PV, SOM and VIP neurons with an architecture similar to (Pfeffer et al., 2013). They studied the effect of photostimulation of the different inhibitory populations on the responses and orientation tuning properties of the neurons. In a recent study (Sadeh et al., 2017) have investigated the effects of partial activation of PV neurons upon photostimulation in an ISN. They argued that depending on the degree of viral expression, the average response of the infected neurons can decrease or increase with the light intensity: it decreases only if a large proportion of the population is infected. (Del Molino et al. 2017) showed that due to the non-linearity in the neuronal transfer function, the response of the network to stimulation can be different for different background rates. In particular, they showed that it can reverse the response of SOM neurons to VIP stimulation.

All these works considered inhibition stabilized networks in which the *total* recurrent excitation is so strong that the activity would blow up in the absence of inhibitory feedback. With our notations, this means that $G_E j_{EE} > \frac{1}{K}$, where G_E is the gain of the noise average transfer function (f-I curve) of the excitatory neurons. In fact, in these models all the interactions $j_{\alpha\beta}$ are of order $\frac{1}{K}$ so they are weak in our sense. Moreover, these studies considered networks that are so small that it is impossible to extrapolate their results to mouse cortex size networks. Here we studied large network models $(N = 76800)$ with strong interactions, *i.e.*, $j_{\alpha\beta}$ are of order $\frac{1}{\sqrt{K}}$ operating in the balanced regime. Note that such networks are ISNs provided that $j_{EE} \neq 0$. We showed that paradoxical effect can be present or not depending on the interaction parameters.

Since we used static synapses, changes in the background rates cannot reverse the paradoxical effect in our models. This is because with static synapses the balance equations are linear. One can recover this reversal if one introduces short-term plasticity which will make the balance equations nonlinear. We did not consider partial expression of channelrhodopsin in the PV population because our goal was to account for experimental data where virtually all neurons were infected. These effects have been studied in (Gutnisky et al., 2017; Sanzeni et al., 2019) in strongly coupled networks of two populations yielding to the same conclusions as (Sadeh et al., 2017).

Predictions

Our theory (Model 1) predicts that in ALM layer 2/3 the activity of the SOM and VIP populations will decrease upon PV photostimulation (Fig. 4B). It also predicts that upon PC photoinhibition, the PV activity will increase whereas the activity of the SOM and VIP populations will decrease (Fig. 12A). This is because in Model 1 when the PV response is non-paradoxical ($\chi_{II} > 0$) the product $\chi_{EI}\chi_{IE}$ is also positive (see Appendix 1C). Furthermore, in ALM layer 2/3 the population activity of PCs decreases upon PV photostimulation, $\chi_{EI} < 0$. Hence, χ_{IE} is negative. The balance of the PC and the VIP inputs into SOM neurons implies that VIP and PC activity covary. Finally, in Appendix 1C we show that if $\chi_{EE} > 0$ and $\chi_{IE} < 0$ then necessarily $\chi_{SE} > 0$. Thus, in ALM layer 2/3, the SOM population activity should decrease upon PC photoinhibition (Fig. 12A).

In auditory and prefrontal cortex (Pi et al., 2013) as well as in S1 (Lee et al., 2013), photostimulation of VIP neurons, activates them ($\chi_{VV} > 0$) and disinhibits the PCs ($\chi_{EV} > 0$) through an inhibition of the SOM population ($\chi_{SV} < 0$). If this is also true in ALM layer 2/3, our model predicts that photostimulation of VIP neurons should increase the PV activity ($\chi_{IV} > 0$) (Appendix 1C, Fig. 12B).

In S1 our theory (Model 2) predicts that the PC and PV activities will proportionally decrease upon PC photoinhibition (Eq. (3), Appendix 1D, Fig. 12C). Photostimulation of the SOM neurons modifies Eq. (3) and consequently, the changes in PC and PV activity no longer covary (Fig. 12D). Thus, our theory can be tested by photostimulating PV neurons as in our experiment, while also photostimulating SOM neurons with a second laser with constant power. In this case, the model predicts that S1 will still exhibit the paradoxical effect but that the responses of the PC and PV populations will no longer be proportional (Fig. 12E).

Perspectives

We only considered response of the neurons for a large radius of the laser beam. In a recent study (Li et al., 2019), Li et al. have investigated the spatial profile of the response and its dependence on the light intensity. Our theory can be extended to incorporate spatial dependencies. Studying the interplay between the connectivity pattern and laser beam width in the response profile of the networks will provide further constraints on cortical architectures.

Due to the strong interactions in our models, the nonlinearity of the *single* neuron f-I curves hardly affects the population average responses. However, it influences the response heterogeneity that naturally arises in our theory (Figs. 6-8-11). An alternative model for the paradoxical effect is the supralinear stabilized network (SSN) (Rubin et al., 2015) which relies on an expansive non-linearity of the input-output transfer function of the inhibitory *populations*. Whether this mechanism can account for our experimental data is an issue for further study. In particular, it would be interesting to know whether the SSN scenario can account for the strong heterogeneity in the responses and for the proportionality of the PC and PV population activities in S1. Answering these questions may provide a way to discriminate between the balance network and SSN theory.

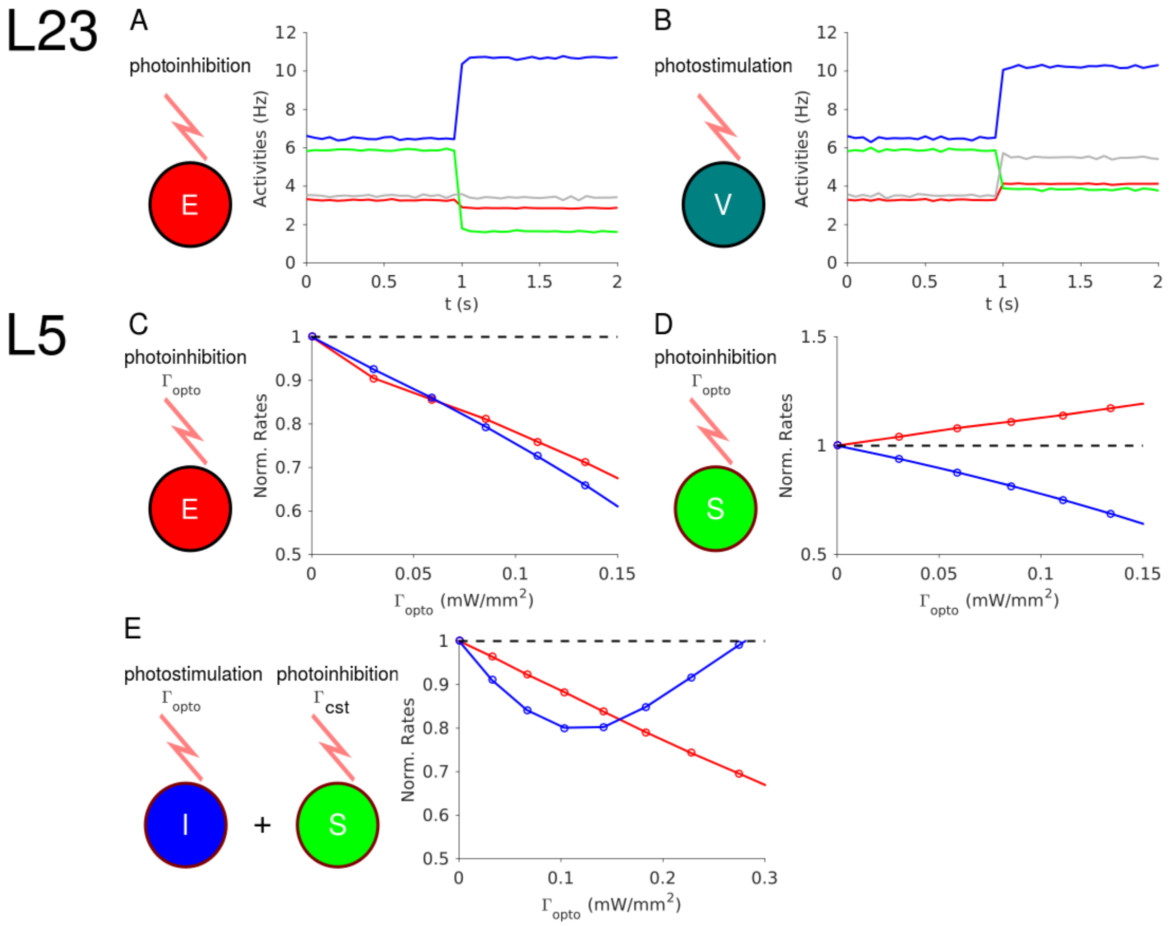


Figure 12. Predictions of the theory. **A.** In ALM layer 2/3, the activity of the PV population decreases upon photoinhibition of the PCs. **B**. In ALM layer 2/3, photostimulation of VIP neurons increases the activity of the PV population. **C**. In S1, PV and PC activity decrease proportionally upon photoinhibition of the latter. **D**. In S1, the PC and PV responses are not proportional upon photoinhibition of the SOM population. **E**. In S1, upon photostimulation of PV neurons and photoinhibition of the SOM population with a constant input, the PV response is paradoxical but PC and PV responses are no longer proportional.

Acknowledgements: We thank Karel Svoboda for illuminating discussions and comments on the manuscript. We are also thankful to Ran Darshan and Tohar Yarden for discussions. D.H thanks Svoboda's lab. and Janelia Research Campus for their warm hospitality. This work was supported by ANR grants ANR-14-NEUC-0001-01 (Cv.V & D.H), ANR-13-BSV4-0014-02 (D.H, Cv.V), the ANR-09-SYSC-002-01 (DH, CvV), the Janelia Research Campus visiting program (D.H), the Helen Hay Whitney Foundation fellowship (N.L.), the Robert and Janice McNair Foundation (N.L.), Whitehall Foundation (N.L.), Alfred P. Sloan Foundation (N.L.), Searle Scholars Program (N.L.), NIH NS104781 (N.L.), the Pew Charitable Trusts (N.L.), and Simons Collaboration on the Global Brain (#543005, N.L.). Work performed in the framework of the France-Israel Center for Neural Computation (CNRS/Hebrew University of Jerusalem).

Author Contributions: DH and AM conceived the project. AM performed the simulations. AM, DH and CvV performed the analytical calculations. GC collected and GC and NL analyzed the in vivo data. AM, DH and CvV wrote the paper with comments from other authors.

Materials and Methods

Key Resources Table				
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Transgenic mouse *Pvalb*-IRES-Cre x R26-LSL-CAG-LSL-ReaChR-mCit	PV-Ires-Cre x R26-CAG-LSL-ReaChR-mCitrine mice	The Jackson Laboratory	JAX #008069 JAX #026294	Red-shifted channelrhodopsin (ReaChR) in PV neurons

Animals and Surgery

The experimental data is from 9 PV-Ires-Cre x R26-CAG-LSL-ReaChR-mCitrine mice (age > P60, both male and female mice) (Hooks et al., 2015). 3 mice were used for photoinhibition in somatosensory cortex (S1). 6 mice were used for photoinhibition in anterior lateral motor cortex (ALM). All procedures were in accordance with protocols approved by the Janelia Research Campus and Baylor College of Medicine Institutional Animal Care and Use Committee.

Mice were prepared for photostimulation and electrophysiology with a clear-skull cap and a headpost (Guo et al., 2014a, 2014b). The scalp and periosteum over the dorsal surface of the skull were removed. A layer of cyanoacrylate adhesive (Krazy glue, Elmer’s Products Inc) was directly applied to the intact skull. A custom made headbar was placed on the skull (approximately over visual cortex) and cemented in place with clear dental acrylic (Lang Dental Jet Repair Acrylic; Part# 1223-clear). A thin layer of clear dental acrylic was applied over the cyanoacrylate adhesive covering the entire exposed skull, followed by a thin layer of clear nail polish (Electron Microscopy Sciences, Part# 72180).

Photostimulation

Light from a 594 nm laser (Cobolt Inc., Colbolt Mambo 100) was controlled by an acousto-optical modulator (AOM; MTS110-A3-VIS, Quanta Tech; extinction ratio 1:2000; 1μs rise time) and a shutter (Vincent Associates), coupled to a 2D scanning galvo system (GVA002, Thorlabs), then focused onto the brain surface (Guo et al., 2014a). The laser at the brain surface had a diameter of 2 mm. We tested photoinhibition in barrel cortex (bregma posterior 0.5 mm, 3.5 mm lateral) and ALM (bregma anterior 2.5mm, 1.5 mm lateral).

To prevent the mice from detecting the photostimulus, a ‘masking flash’ pulse train (40 1ms pulses at 10 Hz) was delivered using a LED driver (Mightex, SLA-1200-2) and 590 nm LEDs (Luxeon Star) positioned near the eyes of the mice. The masking flash began before the photostimulus started and continued through the end of the epoch in which photostimulation could occur.

The photostimulus had a near sinusoidal temporal profile (40 Hz) with a linear attenuation in intensity over the last 100-200 ms (duration: 1.3 s including the ramp). The photostimulation was delivered at ~7 s intervals. The power (0.5, 1.2, 2.2, 5, 12 mW for S1 photostimulation; 0.3, 0.5, 1, 1.5, 2, 3.3, 5, 8, 15 mW for ALM photostimulation) were chosen randomly. Because we used a time-varying photostimulus, the power values reported here reflect the time-average.

Electrophysiology

All recordings were carried out while the mice were awake but not engaged in any behavior. Extracellular spiking activity was recorded using silicon probes. We used 32-channel NeuroNexus silicon probes (A4x8-5mm-100-200-177) or 64-channel Cambridge NeuroTech silicon probes (H2 acute probe, 25 μm spacing, 2 shanks). The 32-channel voltage signals were multiplexed, digitized by a PCI6133 board at 400 kHz (National Instruments) at 14 bit, demultiplexed (sampling at 25,000 Hz) and stored for offline analysis. The 64-channel voltage signals were amplified and digitized on an Intan RHD2164 64-Channel Amplifier Board (Intan Technology) at 16 bit, recorded on an Intan RHD2000-Series Amplifier Evaluation System (sampling at 20,000 Hz) using Open-Source RHD2000 Interface Software from Intan Technology (version 1.5.2), and stored for offline analysis.

A 1 mm diameter craniotomy was made over the recording site. The position of the craniotomy was guided by stereotactic coordinates for recordings in ALM (bregma anterior 2.5mm, 1.5 mm lateral) or barrel cortex (bregma posterior 0.5 mm, 3.5 mm lateral).

Prior to each recording session, the tips of the silicon probe were brushed with DiI in ethanol solution and allowed to dry. The surface of the craniotomy was kept moist with saline. The silicon probe was positioned on the surface of the cortex and advanced manually into the brain at ~ 3 μm/s, normal to the pial surface. The electrode depth was inferred from manipulator depth and verified with histology. For ALM recordings, putative layer 2/3 units were above 450 μm and putative layer 5 units were below 450 μm (Hooks et al., 2013). For S1, our recording did not distinguish layers.

Data analysis

The extracellular recording traces were band-pass filtered (300-6 kHz). Events that exceed an amplitude threshold (4 standard deviations of the background) were subjected to manual spike sorting to extract single units (Guo et al., 2014a).

Our final data set comprised of 204 single units (S1, 95; ALM, 109). For each unit, its spike width was computed as the trough to peak interval in the mean spike waveform

(Guo et al., 2014a). We defined units with spike width <0.35 ms as FS neurons (31/204) and units with spike width >0.45 ms as putative pyramidal neurons (170/204). Units with intermediate values (0.35 - 0.45 ms, 3/204) were excluded from our analyses.

To quantify photoinhibition strength, we computed “normalized spike rate” during photostimulation. For each neuron, we computed its spike rate during the photostimulus (1 second time window) and its baseline spike rate (500 ms time window before photostimulus onset). The spike rates under photostimulation were divided by the baseline spike rate. The “normalized spike rate” thus reports the total fraction of spiking output under photostimulation. For normalized spike rate of individual neurons, each neuron’s spike rate with photostimulation was normalized by dividing its baseline spike rate (Fig 1B-D, top). For normalized spike rate of the neuronal population (Fig 1B-D, bottom), the spike rates with photostimulation were first averaged across the population (without normalization) and then normalized by dividing the averaged baseline spike rate.

Bootstrap was performed over neurons to obtain standard errors of the mean. For each round of bootstrapping, repeated 1000-10000 times, we randomly sampled with replacement neurons in the dataset. We computed the means of the resampled datasets. The standard error of the mean was the standard deviation of the mean estimates from bootstrap.

Network models

All the models we consider consist of strongly interacting leaky integrate-and-fire neurons. We first study networks of one excitatory (E) and one inhibitory (I) population. We then investigate two models comprising three inhibitory populations, namely parvalbumin positive (PV or I), somatostatin positive (SOM or S) and a third population either corresponding to the vasoactive intestinal peptide positive (VIP or V) neurons (Model 1) or to an unidentified population denoted by X (Model 2).

In all models the total number of neurons is $N = 76800$. In the two population model, 75% are excitatory and 25% inhibitory. In the four-population networks, 75% are excitatory and the number of cells is the same, $\frac{N}{12}$, for all GABAergic inhibitory population.

The data we seek to account for, were obtained in optogenetic experiments in which the laser diameter was substantially larger than the spatial range of neuronal interactions and comparable to the size of the cortical area were the recordings were performed. Therefore, in all models we assume for simplicity that the connectivity is unstructured: neuron (i, α), $\left(\alpha = E, I, S, \frac{V}{X}\right)$, is postsynaptically connected to neuron (j, β) with probability

$$P_{ij}^{\alpha\beta} = \frac{K_{\alpha\beta}}{N_\beta} \qquad (1)$$

For simplicity, we take $K_{\alpha\beta}$ the same for all populations, $K_{\alpha\beta} = K$.

Neuron dynamics: The dynamics between spikes of the membrane potential of the neuron (i, α) is given by

$$C_M \frac{dV_i^\alpha(t)}{dt} = - g_{leak}^\alpha (V_i^\alpha(t) - V_R) + I_{rec}^{\alpha i}(t) + \Lambda_{ext}^\alpha + \Lambda_{opto}^{\alpha i} \qquad (2)$$

Here, $I_{rec}^{\alpha i}(t)$ is the net recurrent input into neuron (i, α), Λ_{ext}^α represents inputs from outside the circuit (*e.g.* thalamic excitation) to population α, and $\Lambda_{opto}^{\alpha i}$ is the optogenetic input into neuron (i, α).

We assumed that the capacitance, C_M, is identical for all neurons and the leak conductance, g_{leak}^α, is identical for all the cells in the same population. We take $C_M = 1 \mu F.cm^{-2}$, $g_{leak}^I = 0.1 mS.cm^{-2}$ and $g_{leak}^E = g_{leak}^S = g_{leak}^{\frac{V}{X}} = 0.05 mS.cm^{-2}$.

Equation (2) has to be supplemented by a reset condition: if at time t the membrane potential of the neuron (i, α) crosses the threshold $V_i^\alpha(t^-) = V_{th} = -50mV$, the neuron fires a spike and its voltage is reset to the resting potential $V_i^\alpha(t^+) = V_R = -70mV$.

Recurrent inputs: The net recurrent input into neuron (i, α) is

$$I_{rec}^{\alpha i}(t) = \sum_{\beta, j} j_{\alpha\beta}\, \varepsilon_\beta\, C_{ij}^{\alpha\beta}\, S_j^{\alpha\beta}(t) \qquad (3)$$

where $C^{\alpha\beta}$ is the connectivity matrix between (presynaptic) population β and (postsynaptic) population α, such that $C_{ij}^{\alpha\beta} = 1$ if neuron (j, β) projects to neuron (i, α) and $C_{ij}^{\alpha\beta} = 0$ otherwise. The parameter $j_{\alpha\beta}$, is the strength of the interaction from neurons in population β to neurons population α. We assumed it to depend on the pre and postsynaptic populations only. The polarity (excitation or inhibition) of the interaction is denoted by ε_β. Therefore if $\beta = E$, $\varepsilon_\beta = 1$ and $\varepsilon_\beta = -1$ otherwise.

The function $S_j^{\alpha\beta}(t)$ is

$$S_j^{\alpha\beta}(t) = \sum_k f_{\alpha\beta}\left(t - t_{\beta j}^k\right) \qquad (4)$$

where $t_{\beta j}^k$ is the time at which neuron (j, β) has emitted its k^{th} spike, the sum is over all the spikes emitted by neuron (j, β) prior to time t and

$$f_{\alpha\beta}(t) = \frac{1}{\tau_{\alpha\beta}} e^{\frac{-t}{\tau_{\alpha\beta}}} \tag{5}$$

where $\tau_{\alpha\beta}$ is the synaptic time constant of the interactions between neurons in population β and α.

External and optogenetic inputs: The feedforward input, Λ^{α}_{ext}, into the neurons in population α is described by inputs from $2K$ external neurons with constant firing rate $r_0 = 5Hz$ and an interaction strength $j_{\alpha 0}$, therefore, $\Lambda^{\alpha}_{ext} = 2Kj_{\alpha 0} r_0$.

We model the ReachR photostimulation as an additional external constant input to the stimulated population. For simplicity, we assume that this input, $\Lambda^{\alpha i}_{opto} = \Lambda^{\alpha}_{opto}$, is the same for all stimulated neurons. Unless specified otherwise, we only consider $\Lambda^{I}_{opto} = \Lambda_{opto}$ and $\Lambda^{\alpha}_{opto} = 0$ for $\alpha \neq I$.

In qualitative agreement with Fig. 3, and Figs. 5, 7, 7-supplementary figure 3, 10 (Hooks et al., 2015) we take

$$\Lambda_{opto} = \Lambda^{\alpha}_{0} log\left(1 + \frac{\Gamma_{opto}}{\Gamma^{\alpha}_{0}}\right) \tag{6}$$

where Γ_{opto} is the laser intensity and Λ_0 and Γ_0 are parameters.

Architectures of the four-population models: The network of Model 1 is depicted in Fig. 4A. In line with the results of (Pfeffer et al., 2013), there are no connections from PV to SOM, VIP to PC and VIP to PV neurons. There is no mutual inhibition between SOM as well as between VIP neurons. All the populations except SOM receive feedforward external input.

The interaction matrix of the network is

$$[j_{AB}\varepsilon_B] = \begin{bmatrix} j_{EE} & -j_{EI} & -j_{ES} & 0 \\ j_{IE} & -j_{II} & -j_{IS} & 0 \\ j_{SE} & 0 & 0 & -j_{SV} \\ j_{VE} & -j_{VI} & -j_{VS} & 0 \end{bmatrix} \tag{7}$$

The network of Model 2 is depicted in Fig. 9B. SOM only receives projections from PCs and PV neurons. X neurons are recurrently connected and project to PCs and PV neurons. The PC and SOM populations project to the population X. All the populations except SOM receive feedforward external input.

The interaction matrix is

$$[j_{AB}\varepsilon_B] = \begin{bmatrix} j_{EE} & -j_{EI} & -j_{ES} & -j_{EX} \\ j_{IE} & -j_{II} & -j_{IS} & -j_{IX} \\ j_{SE} & -j_{SI} & 0 & 0 \\ j_{XE} & 0 & -j_{XS} & -j_{XX} \end{bmatrix} \quad (8)$$

Numerical simulations: The dynamics of the models was integrated numerically using a second-order Runge-Kutta scheme (Press et al., 1986) without spike time interpolation. Unless specified otherwise the time step was $\Delta t = 0.01ms$ and the temporally averaged firing rates were estimated over $100s$.

The balance equations

We consider recurrent networks of strongly interacting neurons (van Vreeswijk and Sompolinsky, 1996) in which order $\sqrt{K}$ excitatory synaptic inputs are sufficient to bring the voltage above threshold. To understand the behavior of such networks, it is imperative to analyse how it behaves when K goes to infinity. To this end, we scale the interactions as

$$j_{\alpha\beta} = \frac{J_{\alpha\beta}}{\sqrt{K}} \quad (9)$$

where $J_{\alpha\beta}$ does not depend on K. Since a neuron receives on average K inputs from each of its presynaptic populations, the total interaction from population β to a neuron in population α is $J_{\alpha\beta}\sqrt{K}$. To keep the relative strength of the optogenetic input, Λ^{α}_{opto}, as K increases we take

$$\Lambda^{\alpha}_{opto} = I^{\alpha}_{opto}\sqrt{K} \quad (10)$$

where I^{α}_{opto} depends on the intensity of the laser:

$$I^{\alpha}_{opto} = I^{\alpha}_{0} log\left(1 + \frac{\Gamma_{opto}}{\Gamma^{\alpha}_{0}}\right) \quad (11)$$

We take: $I^{\alpha}_{0} = I_0 = 8nA$ and $\Gamma^{\alpha}_{0} = \Gamma_0 = 0.5mW.mm^{-2}$.

The net input into the neurons must remain finite in the infinite K limit. This implies that up to corrections which are of the order of $\frac{1}{\sqrt{K}}$,

$$2\,J_{\alpha 0}\,r_0 + I^{\alpha}_{opto} + \sum_{\beta} J_{\alpha\beta}\,\varepsilon_{\beta}\,r_{\beta} = 0 \quad (12)$$

In a n-population network, these n equations determine the n firing rates, $r_{\alpha},\ \alpha \in \{1, ..., n\}$.

This set of linear equations express the fact that, for the population activities to be finite, excitatory and inhibitory inputs to the neurons must compensate. These "balance" equations have a unique solution (unless the determinant of the matrix $J_{\alpha\beta}\varepsilon_\beta$ is zero). To be meaningful the solution must be such that all population activities are positive. This constrains the feedforward and recurrent interaction parameters.

The stability of this balanced solution further constraints the interaction parameters and synaptic time constants. A necessary condition for the stability is that $det\left[J_{\alpha\beta}\varepsilon_\beta\right] > 0$. This condition guarantees that the "balanced state" is stable with respect to divergence of the firing rates. A complete study of these constraints for our LIF networks is beyond the scope of this paper.

In all the models, we study parameter ranges in which, at baseline ($I^{\alpha}_{opto} = 0$), the network operates in a stable balanced state where distributions of rates exhibit a quasi-lognormal shape and spikes are emitted irregularly as in a Poisson process (Fig. 5-figure supplementary 3, Fig. 7-figure supplementary 5 and Fig. 10-figure supplementary 3). For I^{α}_{opto} sufficiently large, it may happen that one or more population activity reaches zero. In this case, the network evolves to a partially balanced state in which the rates of the populations that remain active satisfy a reduced set of balanced equations. For example, if we consider a solution were the rate of population γ, r_γ is zero and all other rates are positive, the reduced balance equations are

$$2\,J_{\alpha 0}\,r_0 + I^{\alpha}_{opto} + \sum_{\beta\neq\gamma} J_{\alpha\beta}\,\epsilon_\beta\,r_\beta = 0\text{, for }\alpha\neq\gamma. \tag{13}$$

Consistency of this solution leads to the requirement that the input into population γ is hyperpolarizing.

$$2\,J_{\gamma 0}\,r_0 + I^{\gamma}_{opto} + \sum_{\beta\neq\gamma} J_{\gamma\beta}\,\varepsilon_\beta\,r_\beta < 0 \tag{14}$$

Note that they may be multiple self-consistent solutions which are partially balanced.

Upon photostimulation of PV, in Model 1, the balanced equations are

$$2\,J_{E0}\,r_0 + J_{EE}\,r_E - J_{EI}\,r_I - J_{ES}\,r_S = 0 \tag{15.1}$$

$$2\,J_{I0}\,r_0 + I^{I}_{opto} + J_{IE}\,r_E - J_{II}\,r_I - J_{IS}\,r_S = 0 \tag{15.2}$$

$$J_{SE}\,r_E - J_{SV}\,r_V = 0 \tag{15.3}$$

$$2\,J_{V0}\,r_0 + J_{VE}\,r_E - J_{VI}\,r_I - J_{VS}\,r_S = 0 \tag{15.4}$$

In particular, Eq. 15.3 implies that r_E and r_V are always proportional ($J_{SE}, J_{SV} > 0$).

Similarly, in Model 2, the balanced equations are

$$2\,J_{E0}\,r_0 + J_{EE}\,r_E - J_{EI}\,r_I - J_{ES}\,r_S - J_{EX}\,r_X = 0 \qquad (16.1)$$

$$2\,J_{I0}\,r_0 + I^{I}_{opto} + J_{IE} r_E - J_{II}\,r_I - J_{IS}\,r_S - J_{IX}\,r_X = 0 \qquad (16.2)$$

$$J_{SE}\,r_E - J_{SI}\,r_I = 0 \qquad (16.3)$$

$$2\,J_{X0}\,r_0 + J_{VE}\,r_E - J_{VS}\,r_S - J_{XX}\,r_X = 0 \qquad (16.4)$$

Equation 16.3 implies that in this network r_E and r_I are always proportional $(J_{SE}, J_{SI} > 0)$.

Parameters of the two-population model

Table 1. Connection strength matrix (rows: postsynaptic populations; columns: presynaptic populations)

$J_{\alpha\beta}\ (\mu A.ms.cm^{-2})$	Feedforward	PC	PV
PC	17	29	30
PV	17	36	36

Table 2. Synaptic time constants

$\tau_{\alpha\beta}\ (ms)$	E	I
E	4	2
I	2	2

Default parameters of Model 1

Table 3. Synaptic time constants

$\tau_{\alpha\beta}\,(ms)$	PC	PV	SOM	VIP
PC	4	2	2	N/A
PV	2	2	4	N/A
SOM	2	N/A	N/A	4
VIP	4	2	4	N/A

Table 4. Connection strength matrix for $J_{EE} > J^{*}_{EE}$ (rows: postsynaptic populations; columns: presynaptic populations)

$J_{\alpha\beta}\,(\mu A.ms.cm^{-2})$	Feedforward	PC	PV	SOM	VIP
PC	34	20	26.4	41	0
PV	27	44	28	35.6	0
SOM	0	24	0	0	14
VIP	39	12	35.2	35	0

Table 5. Connection strength matrix for $J_{EE} < J^{*}_{EE}$ (rows: postsynaptic populations; columns: presynaptic populations).

$J_{\alpha\beta}\,(\mu A.ms.cm^{-2})$	Feedforward	PC	PV	SOM	VIP
PC	52	17.4	34.4	32.8	0
PV	39	36.6	29.2	28.8	0
SOM	0	24.2	0	0	16.8
VIP	30	31.2	31	14.6	0

Default parameters of Model 2

Table 6. Synaptic time constants in Model 2

$\tau_{\alpha\beta}$ (ms)	PC	PV	SOM	X
PC	4	2	2	4
PV	2	2	4	4
SOM	2	2	N/A	N/A
X	2	N/A	4	2

Table 7. Connection strength matrix (rows: postsynaptic populations; columns: presynaptic populations).

$J_{\alpha\beta}$ $(\mu A.ms.cm^2)$	Feedforward	PC	PV	SOM	X
PC	48	20	30	32	36
PV	29	40	28	16	32
SOM	0	26	12	0	0
X	24	24	0	36	22

References

Adesnik H, Bruns W, Taniguchi H, Josh Huang Z, Scanziani M. 2012. A neural circuit for spatial summation in visual cortex. *Nature*. doi:10.1038/nature11526

Amit DJ, Brunel N. 1997. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb Cortex* 7:237–252.

Argaman T, Golomb D. 2018. Does layer 4 in the barrel cortex function as a balanced circuit when responding to whisker movements? *Neuroscience* 368:29–45.

Atallah BV, Bruns W, Carandini M, Scanziani M. 2012. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* 73:159–170.

Barral J, D Reyes A. 2016. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nat Neurosci* 19:1690.

Beierlein M, Gibson JR, Connors BW. 2003. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *J Neurophysiol* 90:2987–3000.

Beierlein M, Gibson JR, Connors BW. 2000. A network of electrically coupled interneurons drives synchronized inhibition in neocortex. *Nat Neurosci* 3:904–910.

Capocelli, R. M., and L. M. Ricciardi. 1971. “Diffusion Approximation and First Passage Time Problem for a Model Neuron.” *Kybernetik* 8 (6): 214–23.

Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, Moore CI. 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459:663–667.

Cruikshank SJ, Urabe H, Nurmikko AV, Connors BW. 2010. Pathway-specific feedforward circuits between thalamus and neocortex revealed by selective optical stimulation of axons. *Neuron* 65:230–245.

Darshan R, Wood WE, Peters S, Leblois A, Hansel D. 2017. A canonical neural mechanism for behavioral variability. *Nat Commun* 8:15415.

Garcia Del Molino LC, Yang GR, Mejias JF, Wang X-J. 2017. Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types. *Elife* 6. doi:10.7554/eLife.29742

Gerstner W, Kistler WM, Naud R, Paninski L. 2014. Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition. Cambridge University Press.

Gibson JR, Beierlein M, Connors BW. 1999. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature* 402:75–79.

Goldberg JH, Lacefield CO, Yuste R. 2004. Global dendritic calcium spikes in mouse layer 5 low threshold spiking interneurones: implications for control of pyramidal cell bursting. *J Physiol* 558:465–478.

Guo ZV, Hires SA, Li N, O’Connor DH, Komiyama T, Ophir E, Huber D, Bonardi C, Morandell K, Gutnisky D, Peron S, Xu N-L, Cox J, Svoboda K. 2014a. Procedures for behavioral experiments in head-fixed mice. *PLoS One* 9:e88678.

Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K. 2014b. Flow of Cortical Activity Underlying a Tactile Decision in Mice. *Neuron*. doi:10.1016/j.neuron.2013.10.020

Gutnisky DA, Yu J, Hires SA, To M-S, Bale MR, Svoboda K, Golomb D. 2017. Mechanisms underlying a thalamocortical transformation during active tactile sensation. *PLoS Comput Biol* 13:e1005576.

Haider B, Duque A, Hasenstaub AR, McCormick DA. 2006. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* 26:4535–4545.

Hansel D, Mato G. 2013. Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. *Journal of Neuroscience*. doi:10.1523/jneurosci.3455-12.2013

Hansel D, Sompolinsky H. 1992. Synchronization and computation in a chaotic neural network. *Phys Rev Lett* 68:718–721.

Hansel D, van Vreeswijk C. 2012. The mechanism of orientation selectivity in primary visual cortex without a functional map. *J Neurosci* 32:4049–4064.

Harish O, Hansel D. 2015. Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Comput Biol* 11:e1004266.

Hennequin G, Ahmadian Y, Rubin DB, Lengyel M, Miller KD. 2018. The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron* 98:846–860.e5.

Hooks BM, Lin JY, Guo C, Svoboda K. 2015. Dual-channel circuit mapping reveals sensorimotor convergence in the primary motor cortex. *J Neurosci* 35:4418–4426.

Hooks BM, Mao T, Gutnisky DA, Yamawaki N, Svoboda K, Shepherd GMG. 2013. Organization of Cortical and Thalamic Input to Pyramidal Neurons in Mouse Motor Cortex. *Journal of Neuroscience*. doi:10.1523/jneurosci.4338-12.2013

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

Hu H, Ma Y, Agmon A. 2011. Submillisecond firing synchrony between different subtypes of cortical interneurons connected chemically but not electrically. *J Neurosci* 31:3351–3361.

Jiang X, Shen S, Cadwell CR, Berens P, Sinz F, Ecker AS, Patel S, Tolias AS. 2015. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350:aac9462.

Jiang X, Wang G, Lee AJ, Stornetta RL, Zhu JJ. 2013. The organization of two new cortical interneuronal circuits. *Nat Neurosci* 16:210–218.

Kadmon J, Sompolinsky H. 2015. Transition to Chaos in Random Neuronal Networks. *Phys Rev X* 5:041030.

Karnani MM, Jackson J, Ayzenshtat I, Tucciarone J, Manoocheri K, Snider WG, Yuste R. 2016. Cooperative Subnetworks of Molecularly Similar Interneurons in Mouse Neocortex. *Neuron* 90:86–100.

Knight BW. 1972. Dynamics of encoding in a population of neurons. *J Gen Physiol* 59:734–766.

LAPICQUE, L. 1909. Definition experimentale de l'excitabilite. *Soc Biol* 77:280–283.

Lee S-H, Kwan C, Zhang S, Phoumthipphavong V, Flannery JG, Masmanidis SC, Taniguchi H, Huang ZJ, Zhang F, Boyden ES, Deisseroth K, Dan Y. 2012. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* 488:379–383.

Lee S, Kruglikov I, Huang ZJ, Fishell G, Rudy B. 2013. A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nat Neurosci* 16:1662–1670.

Lerchner A, Ursta C, Hertz J, Ahmadi M, Ruffiot P, Enemark S. 2006. Response variability in balanced cortical networks. *Neural Comput* 18:634–659.

Levy RB, Reyes 1D. 2012. Spatial Profile of Excitatory and Inhibitory Synaptic Connectivity in Mouse Primary Auditory Cortex. *Journal of Neuroscience*.

doi:10.1523/jneurosci.5158-11.2012

Lien 1D, Scanziani M. 2018. Cortical direction selectivity emerges at convergence of thalamic synapses. *Nature* 558:80–86.

Lien 1D, Scanziani M. 2013. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat Neurosci* 16:1315.

Li N, Chen S, Guo ZV, Chen H, Huo Y, Inagaki HK, Davis C, Hansel D, Guo C, Svoboda K. Spatiotemporal limits of optogenetic manipulations in cortical circuits. eLife 2019;8:e48622

Li N, Chen T-W, Guo ZV, Gerfen CR, Svoboda K. 2015. A motor cortex circuit for motor planning and movement. *Nature* 519:51–56.

Litwin-Kumar A, Rosenbaum R, Doiron B. 2016. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *J Neurophysiol* 115:1399–1409.

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

Ma Y, Hu H, Agmon A. 2012. Short-term plasticity of unitary inhibitory-to-inhibitory synapses depends on the presynaptic interneuron subtype. *J Neurosci* 32:983–988.

Ma Y, Hu H, Berrebi AS, Mathers PH, Agmon A. 2006. Distinct subtypes of somatostatin-containing neocortical interneurons revealed in transgenic mice. *J Neurosci* 26:5069–5082.

Mongillo G, Hansel D, van Vreeswijk C. 2012. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys Rev Lett* 108:158101.

Moore AK, Weible AP, Balmer TS, Trussell LO, Wehr M. 2018. Rapid Rebalancing of Excitation and Inhibition by Cortical Circuitry. *Neuron* 97:1341–1355.e6.

Moore CI, Carlen M, Knoblich U, Cardin JA. 2010. Neocortical interneurons: from diversity, strength. *Cell* 142:189–193.

Murphy BK, Miller KD. 2009. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* 61:635–648.

Naka A, Adesnik H. 2016. Inhibitory Circuits in Cortical Layer 5. *Front Neural Circuits* 10:35.

Nigro MJ, Hashikawa-Yamasaki Y, Rudy B. 2018. Diversity and Connectivity of Layer 5 Somatostatin-Expressing Interneurons in the Mouse Barrel Cortex. *J Neurosci* 38:1622–1633.

Olsen SR, Bortone DS, Adesnik H, Scanziani M. 2012. Gain control by layer six in cortical circuits of vision. *Nature* 483:47–52.

Ozeki H, Finn IM, Schaffer ES, Miller KD, Ferster D. 2009. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* 62:578–592.

Pala A, Petersen CCH. 2015. In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission in layer 2/3 mouse barrel cortex. *Neuron* 85:68–75.

Pattadkal JJ, Mato G, van Vreeswijk C, Priebe NJ, Hansel D. 2018. Emergent Orientation Selectivity from Random Networks in Mouse Visual Cortex. *Cell Reports*. doi:10.1016/j.celrep.2018.07.054

Pehlevan C, Sompolinsky H. 2014. Selectivity and sparseness in randomly connected balanced networks. *PLoS One* 9:e89992.

Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M. 2013. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat Neurosci* 16:1068–1076.

Pi H-J, Hangya B, Kvitsiani D, Sanders JI, Josh Huang Z, Kepecs A. 2013. Cortical interneurons that specialize in disinhibitory control. *Nature*. doi:10.1038/nature12676

Press WR, Flannery BP, Teulosky SA, Vetterling WT. 1986. Numerical Recipies: The Art of Super Computing.

Rao S, Hansel D, van Vreeswijk C. 2019. Dynamics and orientation selectivity in a cortical model of rodent V1 with excess bidirectional connections. *Sci Rep* 9:3334.

Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD. 2010. The asynchronous state in cortical circuits. *Science* 327:587–590.

Resulaj A, Ruediger S, Olsen SR, Scanziani M. 2018. First spikes in visual cortex enable perceptual discrimination. *eLife*. doi:10.7554/elife.34044

Rosenbaum R, Doiron B. 2014. Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections. *Phys Rev X* 4:021039.

Rosenbaum R, Smith MA, Kohn A, Rubin JE, Doiron B. 2017. The spatial structure of correlated neuronal variability. *Nat Neurosci* 20:107–114.

Roudi Y, Latham PE. 2007. A balanced memory network. *PLoS Comput Biol* 3:1679–1700.

Roxin A, Brunel N, Hansel D, Mongillo G, van Vreeswijk C. 2011. On the distribution of firing rates in networks of cortical neurons. *J Neurosci* 31:16217–16226.

Rubin DB, Van Hooser SD, Miller KD. 2015. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* 85:402–417.

Rudy B, Fishell G, Lee S, Hjerling-Leffler J. 2011. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev Neurobiol* 71:45–61.

Sadeh S, Silver RA, Mrsic-Flogel TD, Muir DR. 2017. Assessing the Role of Inhibition in Stabilizing Neocortical Networks Requires Large-Scale Perturbation of the Inhibitory Population. *J Neurosci* 37:12050–12067.

Sanzeni A, Akitake B, Goldbach HC, Leedy CE. 2019. Inhibition stabilization is a widespread property of cortical networks. *bioRxiv*.

Seeman SC, Campagnola L, Davoudian PA, Hoggarth A, Hage TA, Bosma-Moody A, Baker CA, Lee JH, Mihalas S, Teeter C, Ko AL, Ojemann JG, Gwinn RP, Silbergeld DL, Cobbs C, Phillips J, Lein E, Murphy G, Koch C, Zeng H, Jarsky T. 2018. Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *Elife* 7. doi:10.7554/eLife.37349

Shadlen MN, Newsome WT. 1994. Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4:569–579.

Shinomoto S, Kim H, Shimokawa T, Matsuno N, Funahashi S, Shima K, Fujita I, Tamura H, Doi T, Kawano K, Inaba N, Fukushima K, Kurkin S, Kurata K, Taira M, Tsutsui K-I, Komatsu H, Ogawa T, Koida K, Tanji J, Toyama K. 2009. Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Comput Biol* 5:e1000433.

Shu Y, Hasenstaub A, McCormick DA. 2003. Turning on and off recurrent balanced cortical activity. *Nature*. doi:10.1038/nature01616

Svoboda K, Li N. 2018. Neural mechanisms of movement planning: motor cortex and beyond. *Curr Opin Neurobiol* 49:33–41.

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller J, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin A, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch

C, Zeng H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563:72–78.

Tremblay R, Lee S, Rudy B. 2016. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* 91:260–292.

Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1997. Paradoxical effects of external modulation of inhibitory interneurons. *J Neurosci* 17:4382–4388.

Van Vreeswijk C, Sompolinsky H. 2005. Irregular activity in large networks of neuronsLes Houches. Elsevier. pp. 341–406.

van Vreeswijk C, Sompolinsky H. 1998. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*. doi:10.1162/089976698300017214

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

Wilson HR, Cowan JD. 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12:1–24.

Wolf F, Engelken R, Puelma-Touzel M, Weidinger JDF, Neef A. 2014. Dynamical models of cortical circuits. *Curr Opin Neurobiol* 25:228–236.

Xu H, Jeong H-Y, Tremblay R, Rudy B. 2013. Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron* 77:155–167.

Xu X, Callaway EM. 2009. Laminar Specificity of Functional Input to Distinct Types of Inhibitory Cortical Neurons. *Journal of Neuroscience*. doi:10.1523/jneurosci.4104-08.2009

Supplementary Materials: Figures

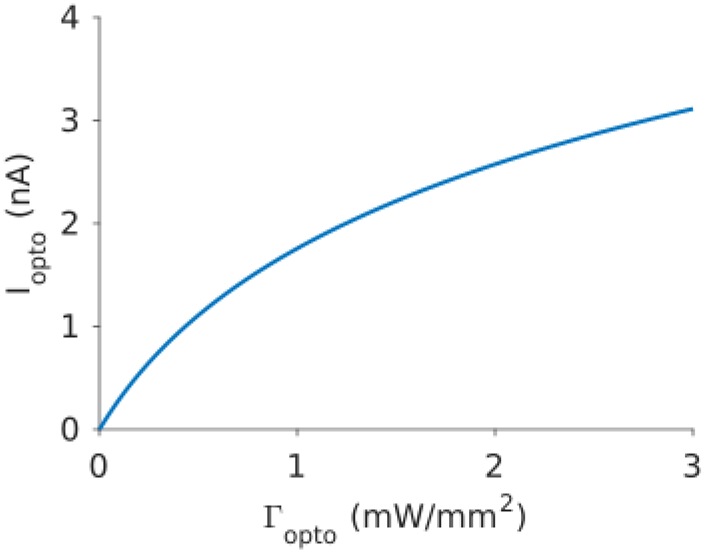


Figure 3-figure supplementary 1. Current, I_{opto}, *v.s.* laser intensity, Γ_{opto}. Parameters are $I_0 = 8nA$, $\Gamma_0 = 0.5mW.mm^{-2}$.

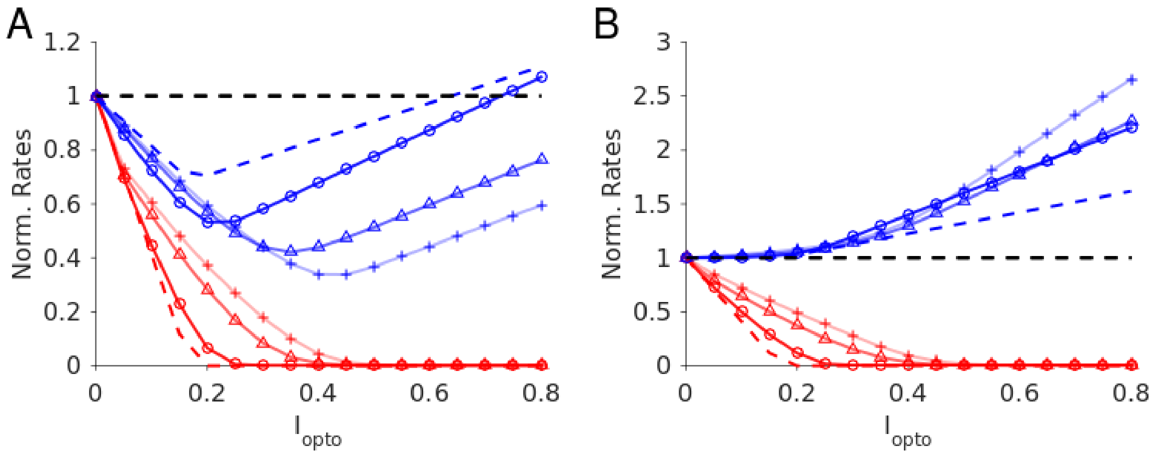


Figure 3-figure supplementary 2. Effects of K on the responses of a two-population network to photoactivation of the inhibitory population. A. $J_{EE} = 22\mu A.ms.cm^{-2}$, the inhibitory population activity always recovers when the PCs are silenced. B. $J_{EE} = 0$, as K increases, the response of the inhibitory population becomes more and more insensitive to the perturbation. Cross: $K = 50$; triangles: $K = 100$; circles: $K = 500$. Dashed line: $\to\infty$. Color code and parameters as in Fig. 3. Baseline firing rates: A. $K = 50$: $r_E = 10.8Hz$, $r_I = 16.8Hz$; $K = 100$: $r_E = 8.8Hz$, $r_I = 14.7Hz$; $K = 500$: $r_E = 5.7Hz$, $r_I = 11.7Hz$; $K = \infty$: $r_E = 3.9Hz$, $r_I = 8.5Hz$. B. $K = 500$: $r_E = 1.9Hz$, $r_I = 3.6Hz$; $K = 100$: $r_E = 2Hz$, $r_I = 4.8Hz$; $K = 500$: $r_E = 1.5Hz$, $r_I = 5.7Hz$; $K = \infty$: $r_E = 1.4Hz$, $r_I = 9.1Hz$.

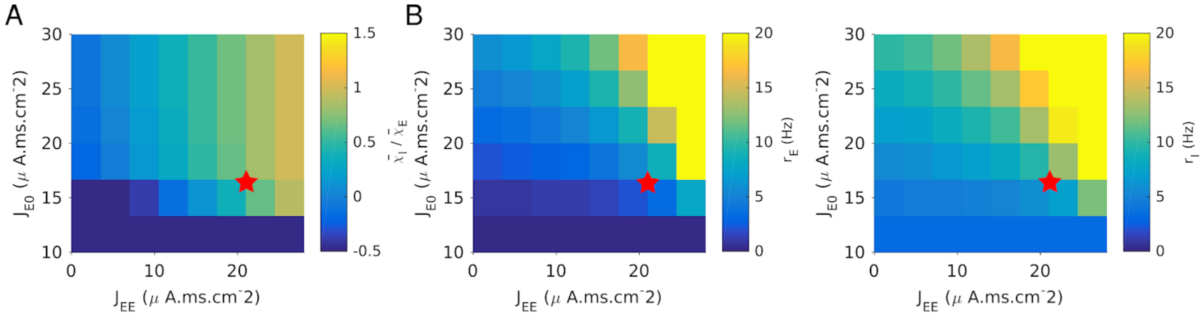


Figure 3-figure supplementary 3. Two-population model. The response of the PC and PV populations upon stimulation of the latter are proportional only if parameters are fine-tuned. A. $\frac{\bar{\chi}_I}{\bar{\chi}_E}$ where $\bar{\chi}_\alpha = \frac{\left(\frac{r_\alpha^{light\ on}}{r_\alpha}-1\right)}{\Gamma_{opto}}$ estimated for $\Gamma_{opto} = 0.5mW.mm^{-2}$. The ratio is close to one only if $J_{EE} \approx J_{EI}\frac{J_{IE}}{J_{II}} = 30\mu A.ms.cm^{-2}$. B. Red star indicates the approximate center of the region with proportionality of the responses together with reasonable activities. Parameters as in Fig. 3. $K = 500$.

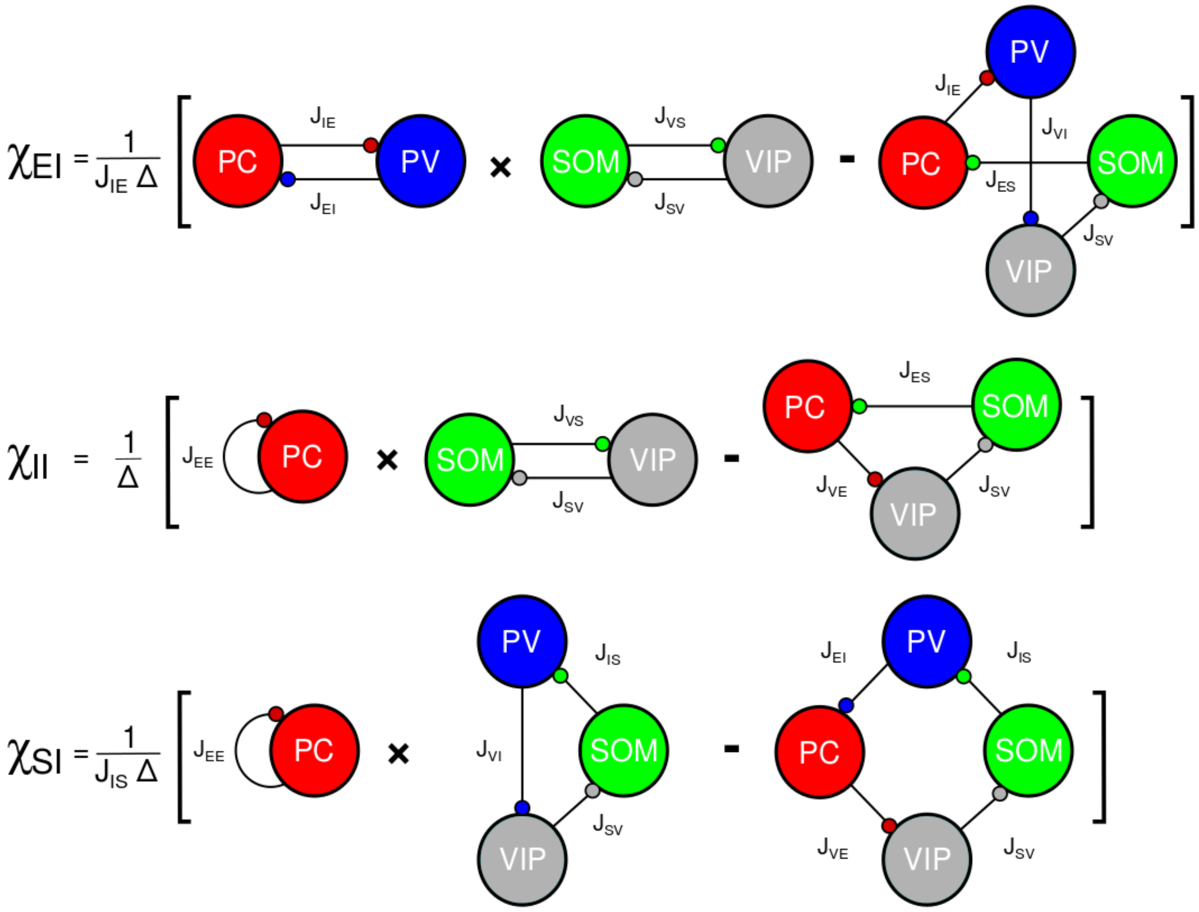


Figure 4-figure supplementary 1. Graphical representation of the population susceptibilities upon stimulation of PV in Model 1 (large N, K limit). The prefactor in front of each diagram accounts for the fact that additional terms are needed to complete the loops. Note: $\chi_{VI} = \frac{J_{SE}}{J_{SV}}\chi_{EI}$.

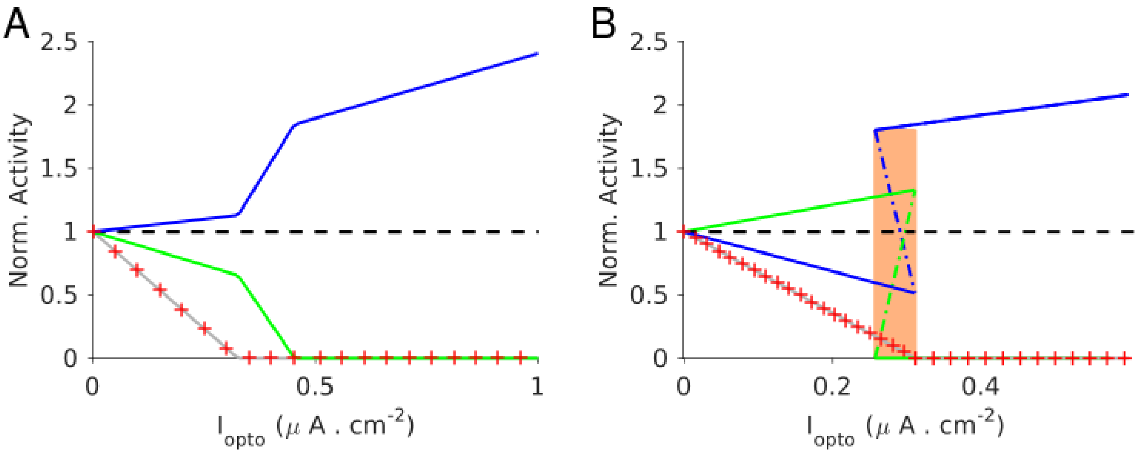


Figure 4-figure supplementary 2. Population activities *vs.* I_{opto} in Model 1 (large N, K limit). The activities are normalized to baseline. A. Parameters as in Table 4. The activity of the PV (blue) population increases with I_{opto}. For PC (red cross), SOM (green) and VIP (gray) the activity decreases. B. Parameters as in Table 5. In the shaded region, the network is bistable. In one stable state all the four populations are active. In the other stable state, only the PV population is active. A third state in which only the PV and SOM populations are active exists in this range of laser intensity (dotted-dashed line). This state is unstable. Baseline firing rates as in Fig. 4.

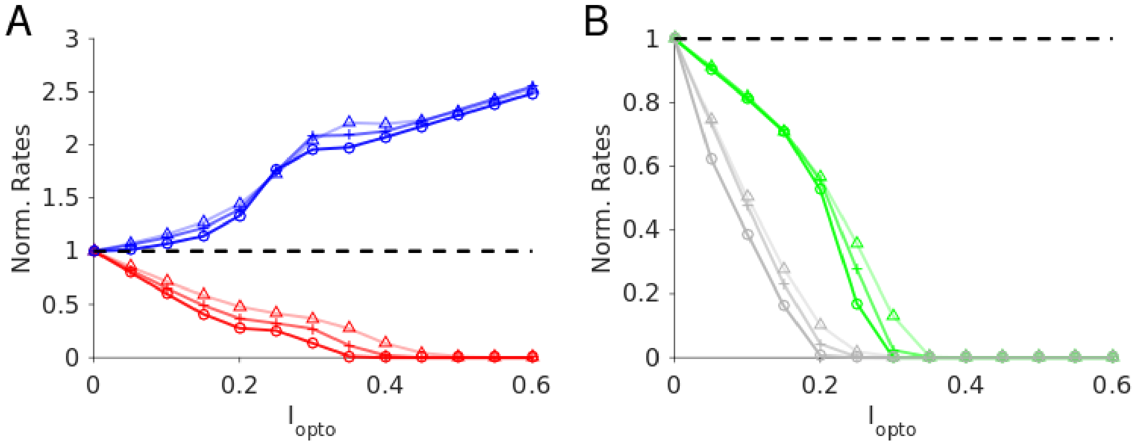


Figure 5-figure supplementary 1. Model 1 with $J_{EE} > J^{*}_{EE}$. Robustness with respect to change in the average connectivity, K. Triangles: $K = 500$; cross: $K = 1000$; circles: $K = 2000$. $N_{\alpha} = 10000$ neurons per population. Baseline firing rates: $K = 500$: $r_E = 3.3Hz$, $r_I = 6.5Hz$, $r_S = 5.9Hz$, $r_V = 3.5Hz; K = 1000$: $r_E = 3.0Hz$, $r_I = 6.6Hz$, $r_S = 5.6Hz$, $r_V = 3.7Hz$; $K = 2000$: $r_E = 2.9Hz$, $r_I = 6.7Hz$, $r_S = 5.4Hz$, $r_V = 3.8Hz$. Rates are averaged over $10s$. Color code and parameters as in Fig. 5.

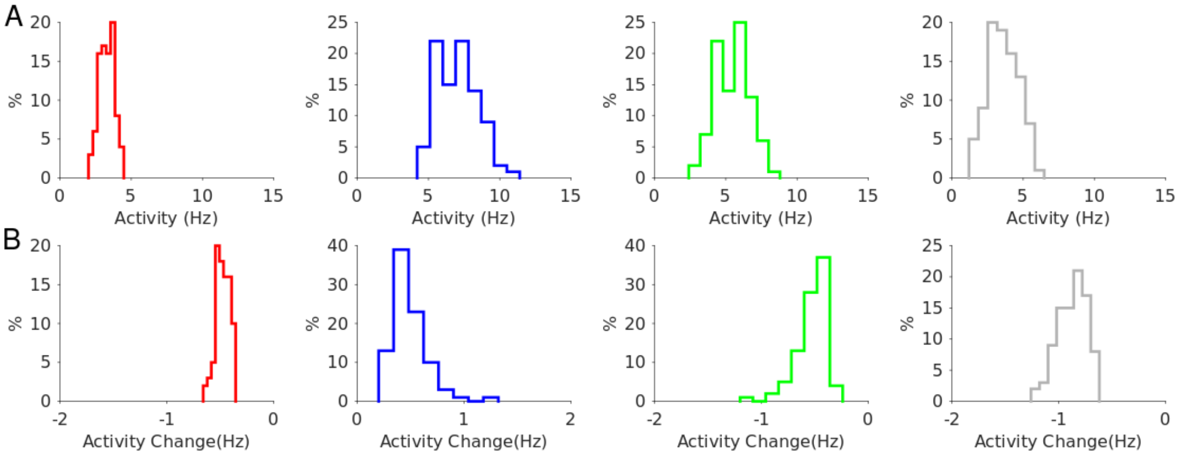


Figure 5-figure supplementary 2. Model 1 with $J_{EE} > J^*_{EE}$. Robustness to a change of $\pm 10\%$ in the interaction parameters. A. Distribution of the population activities. B. Distribution of the activity changes upon stimulation for $\Gamma_{opto} = 0.07 mW.mm^{-2}$. Color code as in Fig. 5. Rates are averaged over $10s$.

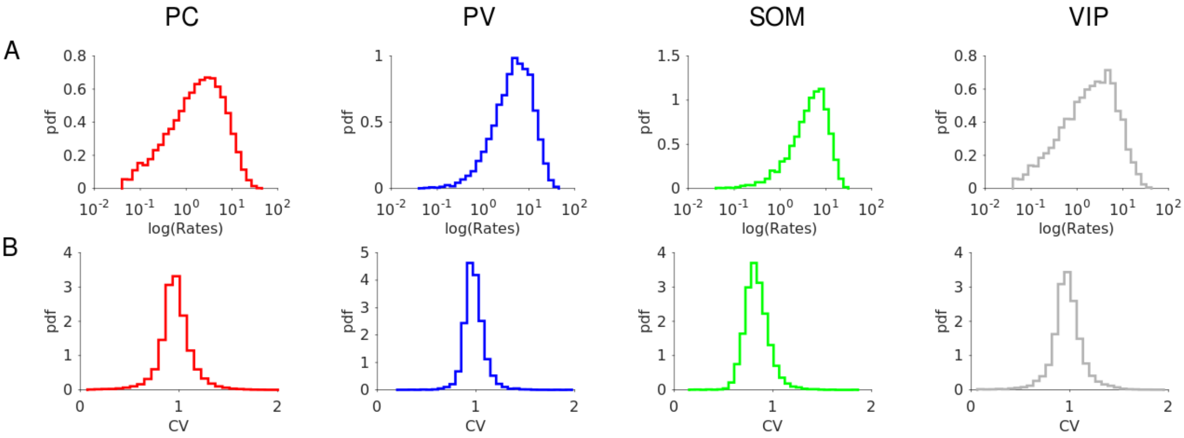


Figure 5-figure supplementary 3. Model 1 with $J_{EE} > J^*_{EE}$. Firing statistics at baseline. A. Distribution of the firing rates (mean: $r_E = 3.3Hz$, $r_I = 6.5Hz$, $r_S = 5.9Hz$, $r_V = 3.5Hz$). B. Distribution of CV. Color code as in Fig. 5. Parameters as in Fig. 5. Individual rates are averaged over $100s$ with a threshold at $0.05Hz$. CVs are computed over $30s$.

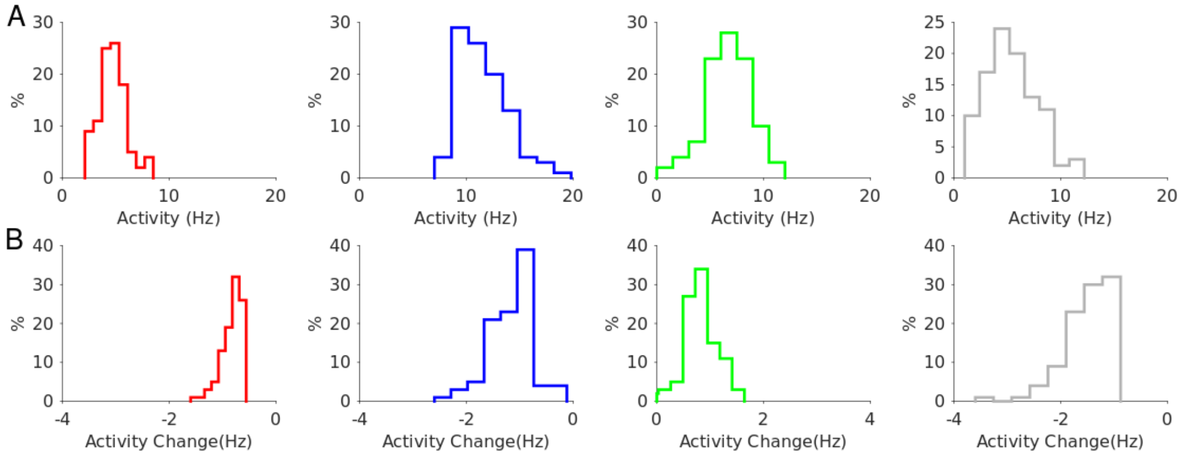


Figure 7-figure supplementary 1. Model 1 with $J_{EE} < J^*_{EE}$. Robustness to a change of $\pm 10\%$ in the interaction parameters. A. Distribution of the population activities. B. Distribution of the activity changes upon stimulation for $\Gamma_{opto} = 0.07 mW.mm^{-2}$. Rates are averaged over $10s$. Color code as in Fig. 7. Parameters as in Fig. 7.

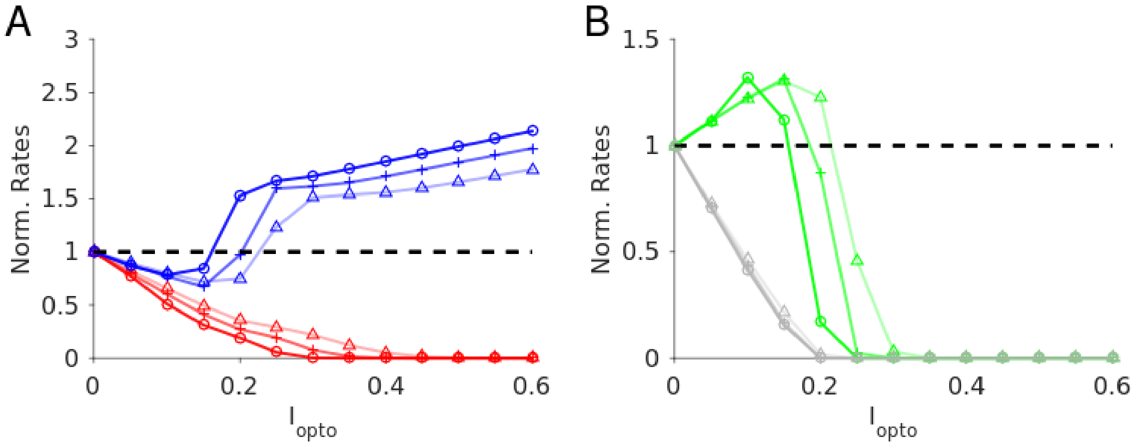


Figure 7-figure supplementary 2. Model 1 with $J_{EE} < J^*_{EE}$. Robustness with respect to change in the average connectivity, K. Triangles: $K = 500$; cross: $K = 1000$; circles: $K = 2000$. $N_\alpha = 10000$ neurons per population. Baseline firing rates: $K = 500$: $r_E = 4.7Hz$, $r_I = 11.2Hz$, $r_S = 7.1Hz$, $r_V = 5.2Hz; K = 1000$: $r_E = 4.1Hz$, $r_I = 10.3Hz$, $r_S = 7.6Hz$, $r_V = 4.7Hz$; $K = 2000$: $r_E = 3.7Hz$, $r_I = 9.7Hz$, $r_S = 7.8Hz$, $r_V = 4.4Hz$. Rates are averaged over $10s$. Color code and parameters as in Fig. 7.

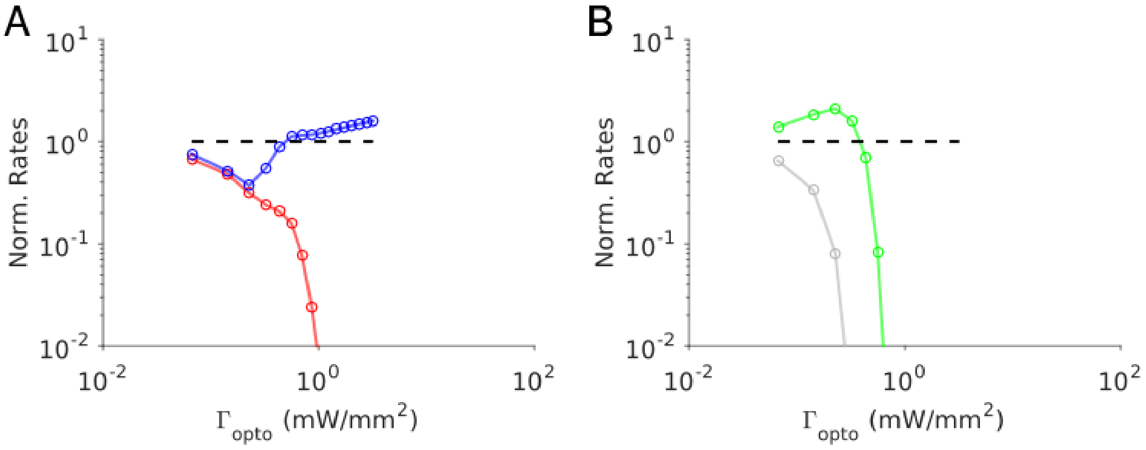


Figure 7-figure supplementary 3. Model 1 for $J_{EE} < J^*_{EE}$. Proportionality of the PC and PV activity requires fine-tuning. A. The response of the PV population is paradoxical for small Γ_{opto} and is proportional to the PC response. B. Responses of the SOM and VIP neurons. Baseline firing rates: $r_E = 6.4Hz$, $r_I = 12.2Hz$, $r_S = 6.5Hz$, $r_V = 11.0Hz$. Color code as in Fig. 7. Interaction parameters: $J_{E0} = 40\ \mu A.ms.cm^{-2}$, $J_{EE} = 20\ \mu A.ms.cm^{-2}$, $J_{EI} = 32\ \mu A.ms.cm^{-2}$, $J_{ES} = 22\ \mu A.ms.cm^{-2}$, $J_{EV} = 0$, $J_{I0} = 3$. Other parameters as in Table 3.

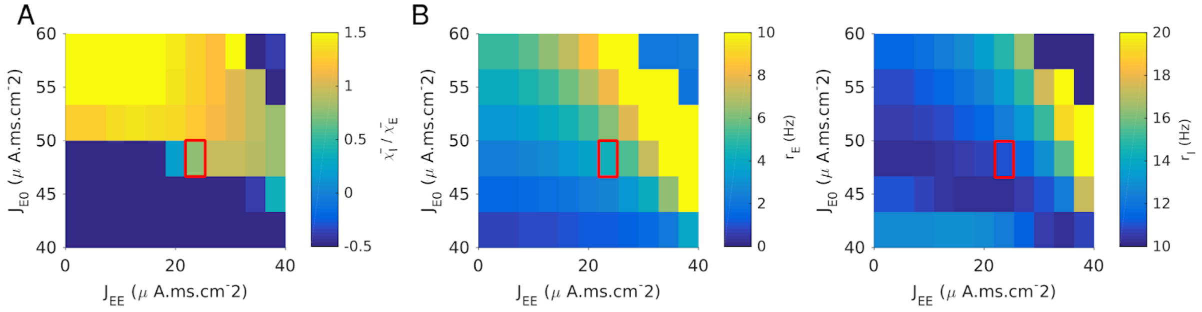


Figure 7-figure supplementary 4. Model 1. The response of the PC and PV populations upon stimulation of the latter are proportional only if parameters are fine-tuned. A. $\frac{\bar{\chi}_I}{\bar{\chi}_E}$ where $\bar{\chi}_A = \frac{(\frac{r_A^{light\ on}}{r_A} - 1)}{\Gamma_{opto}}$ estimated for $\Gamma_{opto} = 0.5mW.mm^{-2}$. B. Red square indicates the region of the parameter space for which the ratio of the PC and PV slopes 1 ± 0.3 and activities are reasonable ($r_E < 5Hz$, $5Hz < r_I < 10Hz$). Parameters as in Fig. 5. $K = 500$.

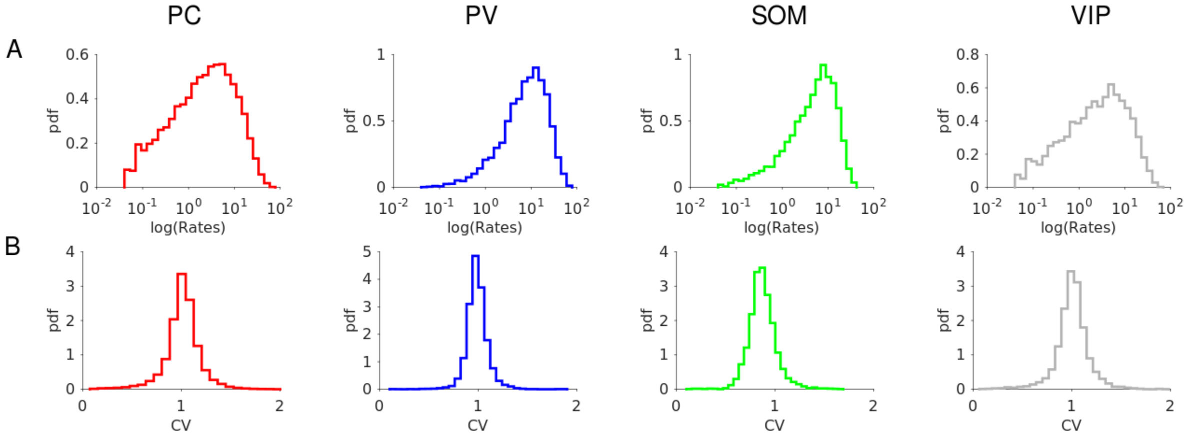


Figure 7-figure supplementary 5. Model 1 with $J_{EE} < J^{*}_{EE}$. Firing statistics at baseline. A. Distribution of the firing rates (mean: $r_E = 4.8Hz$, $r_I = 11.2Hz$, $r_S = 7.1Hz$, $r_V = 5.3Hz$). B. Distribution of CV. Individual rates are average over $100s$ with a threshold at $0.05Hz$. CVs are computed over $30s$. Color code as in Fig. 7. Parameters as in Fig. 7.

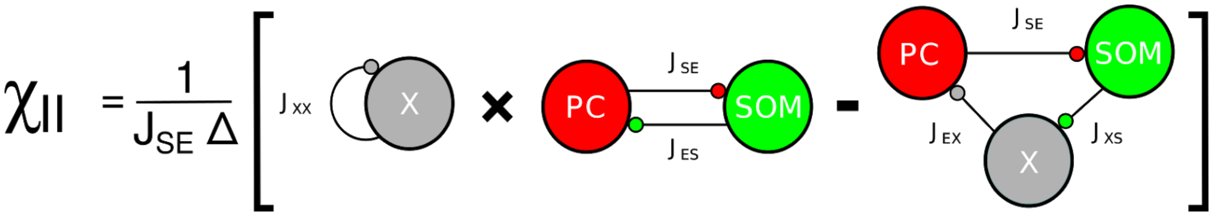


Figure 9-figure supplementary 1. Model 2. Graphical representation of χ_{II} (large N, K limit). Note: $\chi_{EI} = \frac{J_{SI}}{J_{SE}} \chi_{II}$.

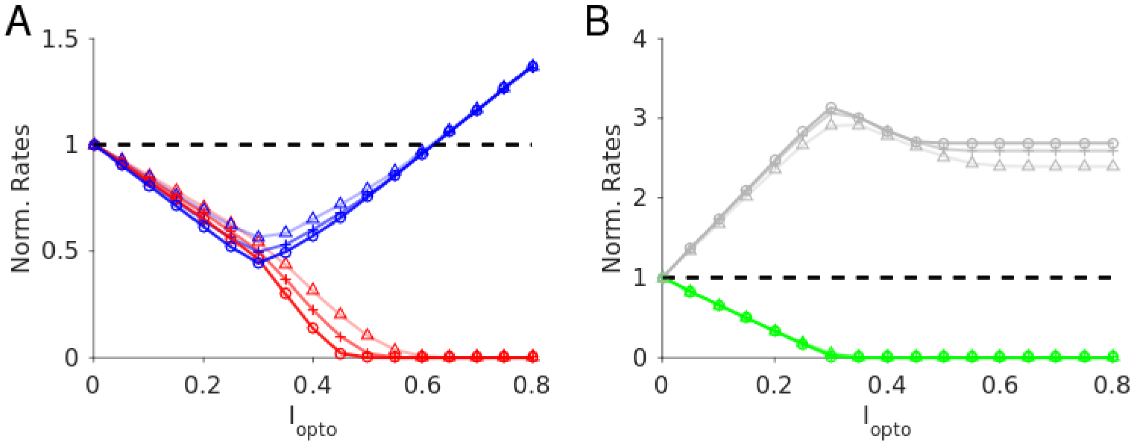


Figure 10-figure supplementary 1. Model 2. Robustness with respect to change in the average connectivity, K. Triangles: $K = 500$; cross: $K = 1000$; circles: $K = 2000$. $N_\alpha = 10000$ neurons per population. Color code and parameters as in Fig. 10. Baseline firing rates: $K = 500$: $r_E = 4.2Hz$, $r_I = 7.0Hz$, $r_S = 7.0Hz$, $r_X = 4.0Hz; K = 1000$: $r_E = 4.0Hz$, $r_I = 6.8Hz$, $r_S = 6.8Hz$, $r_X = 3.8Hz$; $K = 2000$:

$r_E = 3.7Hz$, $r_I = 6.8Hz$, $r_S = 6.7Hz$, $r_X = 3.8Hz$. Rates are averaged over $10s$.

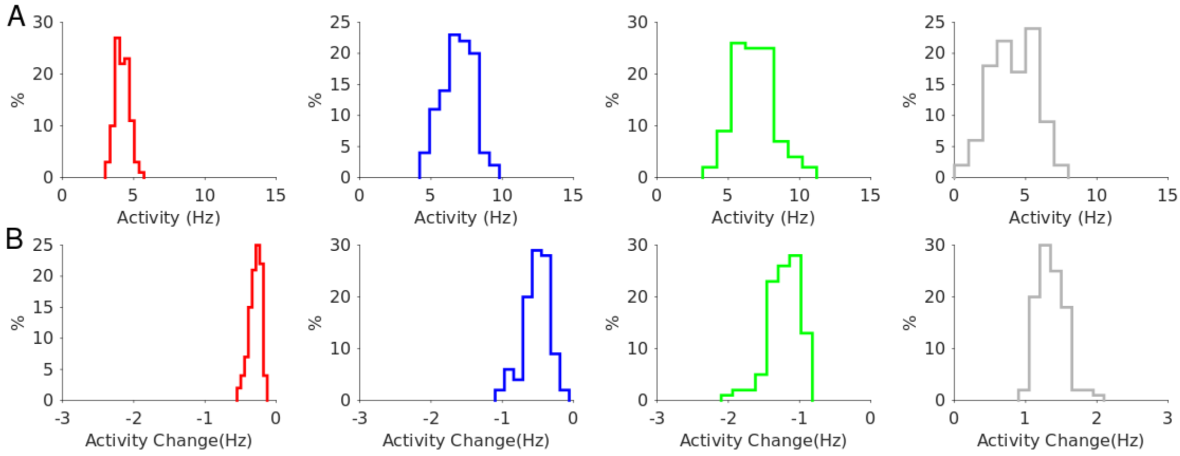


Figure 10-figure supplementary 2. Model 2. Robustness to a change of $\pm 10\%$ in the interaction parameters. A. Distribution of the population activities. B. Distribution of the activity changes upon stimulation for $\Gamma_{opto} = 0.07 mW.mm^{-2}$. Rates are averaged over $10s$. Color code as in Fig. 10.

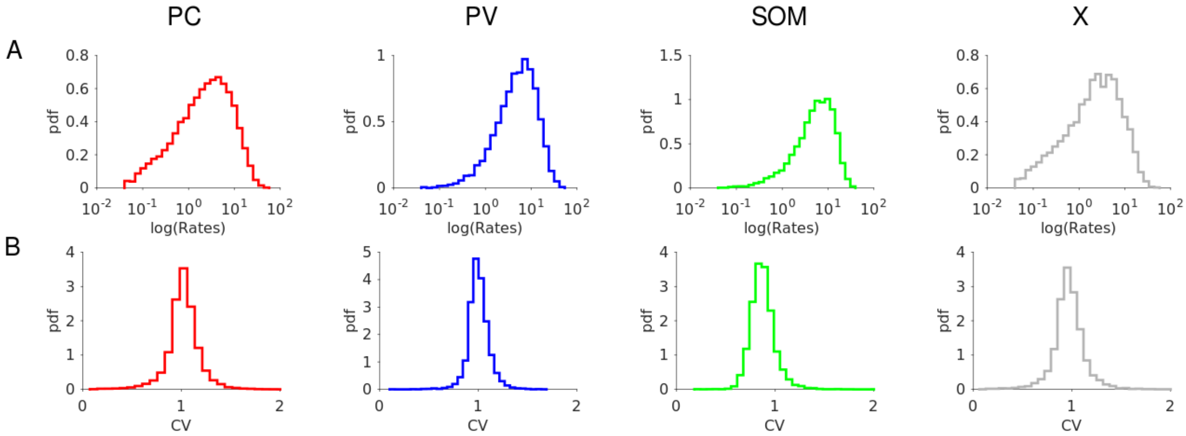


Figure 10-figure supplementary 3. Model 2. Firing statistics at baseline. A. Distribution of the firing rates (mean: $r_E = 4.5Hz$, $r_I = 10.6Hz$, $r_S = 7.2Hz$, $r_V = 4.9Hz$). B. Distribution of CV. Individual rates are average over $100s$ with a threshold at $0.05Hz$. CVs are computed over $30s$. Color code and parameters as in Fig. 10.

Appendix: Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation

Alexandre Mahrach[1], Guang Chen[2], Nuo Li[2], Carl van Vreeswijk[1], David Hansel[1]
[1] CNRS-UMR 8002, Integrative Neuroscience and Cognition Center,
45 Rue des Saints-Pères, 75270 Paris, France
[2] Department of Neuroscience, Baylor College of Medicine, Houston TX 77030, USA

Correspondence:
David Hansel
CNRS-UMR 8002
Integrative Neuroscience and Cognition Center
45 Rue des Saints-Pères, 75270
Paris, France
david.hansel@parisdescartes.fr

Appendix 1

A. Mean field theory

Let us consider a network consisting of n populations (*e.g.* $n = 4$) receiving feedforward input, Λ_{ext}^{α}, from an external population with constant firing rate, r_0, and an optogenetic input, Λ_{opto}^{α} (Materials and Methods). The total input into neuron (i, α) is

$$I_{tot}^{\alpha i}(t) = I_{rec}^{\alpha i}(t) + \Lambda_{ext}^{\alpha} + \Lambda_{opto}^{\alpha} \qquad \text{(A1)}$$

If the size of the network, N, and mean connectivity, K, are large and the synaptic time constants are sufficiently small compared to the membrane time constants, one can take the diffusion approximation and neglect the temporal correlations and write

$$I_{tot}^{\alpha i}(t) = u_{\alpha} + \sqrt{A_{\alpha}}\zeta_i^{\alpha} + \sqrt{B_{\alpha}}\eta_i^{\alpha}(t) \qquad \text{(A2)}$$

where ζ_i^{α} is an i.i.d. Gaussian with zero mean and unit variance, and $\eta_i^{\alpha}(t)$ is a Gaussian white noise with zero mean and unit variance. The mean input, u_{α}, is

$$u_{\alpha} = [< I_{tot}^{\alpha i}(t) >] = \Lambda_{ext}^{\alpha} + \Lambda_{opto}^{\alpha} + K\sum_{\beta} j_{\alpha\beta}\, \varepsilon_{\beta}\, r_{\beta} \qquad \text{(A3)}$$

where the population average firing rate of population β is $r_{\beta} = [r_j^{\beta}]$ and r_j^{β} is the firing rate of the neuron (j, β). Here $< . >$ denotes temporal average (*i.e.* over $\eta_i^{\alpha}(t)$) and $[.]$ is the average over the quenched disorder (ζ_i^{α}). The latter stems from heterogeneities in the in-degree of the inputs into the neurons.

In Eq. (A2), A_{α} is the variance of the quenched disorder which is given by

$$A_{\alpha} = [< I_{tot}^{\alpha i}(t) >^2 - u_{\alpha}^2] = K\sum_{\beta} j_{\alpha\beta}^2 q_{\beta} \qquad \text{(A4)}$$

while B_{α} is the variance of the temporal fluctuations (Van Vreeswijk and Sompolinsky 2005; Roxin et al. 2011)

$$B_{\alpha} = \frac{1}{\tau_m^{\alpha}}\left[\frac{1}{\Delta t}\int_{t}^{t+\Delta t}\left(dt' I_{tot}^{\alpha i}(t') - < I_{tot}^{\alpha i}(t') >\right)^2\right] \qquad \text{(A5)}$$

In Eq. (A4), $q_{\beta} = [(r_j^{\beta})^2]$.

Equations (A4-5) have to be supplemented with the expression of the input-output transfer function which relates the average firing rate, r_i^{α}, to the statistics of $I_{tot}^{\alpha i}(t)$,

$$r_i^{\alpha} = \Phi_{\alpha}(u_{\alpha} + \sqrt{A_{\alpha}}\,\zeta_i^{\alpha}, B_{\alpha}) \qquad \text{(A6)}$$

$$r_{\alpha} = \int D\zeta\, \Phi_{\alpha}(u_{\alpha} + \sqrt{A_{\alpha}}\zeta, B_{\alpha}) \qquad \text{(A7)}$$

$$q_{\alpha} = \int D\zeta\, \Phi_{\alpha}(u_{\alpha} + \sqrt{A_{\alpha}}\zeta, B_{\alpha})^2 \qquad \text{(A8)}$$

where $D\zeta = \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta^2}{2}}$, and Φ_α is given by (Capocelli and Ricciardi 1971)

$$\Phi_\alpha(x,y) = \left\{ \sqrt{\frac{\pi \tau_m^\alpha}{y}} \int_{X_\alpha^+}^{X_\alpha^-} dw e^{w^2} erfc(w) \right\}^{-1} \quad \text{(A9)}$$

where $X_\alpha^- = \frac{x - g_{leak}^\alpha V_R}{\sqrt{y}}$, $X_\alpha^+ = \frac{x - g_{leak}^\alpha V_{Th}}{\sqrt{y}}$ and $\tau_m^\alpha = \frac{C_M}{g_{leak}^\alpha}$ is the membrane time constant of the neurons in population α.

With $j_{\alpha\beta} = \frac{J_{\alpha\beta}}{\sqrt{K}}$, $\Lambda_{ext}^\alpha = 2\sqrt{K}$ and $\Lambda_{opto}^\alpha = I_{opto}^\alpha \sqrt{K}$ (see Materials and Methods), we obtain

$$u_\alpha = \sqrt{K}\left(2J_{\alpha 0}\, r_0 + I_{opto}^\alpha + \sum_\beta J_{\alpha\beta}\, \varepsilon_\beta\, r_\beta\right) \quad \text{(A10)}$$

$$A_\alpha = \sum_\beta J_{\alpha\beta}^2 q_\beta \quad \text{(A11)}$$

$$B_\alpha = \frac{1}{\tau_m^\alpha} \sum_\beta J_{\alpha\beta}^2 r_\beta \quad \text{(A12)}$$

For finite, but large K, the average activity of population α is

$$r_\alpha = \Psi_\alpha\left[u_\alpha, A_\alpha, B_\alpha\right] \quad \text{(A13)}$$

where Ψ_α is the right hand-side of Eq. (A7).

In the limit where $u_\alpha \to -\infty$, it can be shown that

$$\Psi_\alpha\left[u_\alpha, A_\alpha, B_\alpha\right] \sim -\frac{u_\alpha}{\tau_m^\alpha \sqrt{\pi}} \frac{B_\alpha}{(2A_\alpha + B_\alpha)^{\frac{3}{2}}} e^{-\frac{u_\alpha^2}{2A_\alpha + B_\alpha}} \quad \text{(A14)}$$

In the large K limit, the activities, r_α, have to satisfy a set of n linear balance equations (Eq. (12), Materials and Methods) and are given by

$$r_\alpha = -\varepsilon_\alpha \sum_\beta \left[J^{-1}\right]_{\alpha\beta} \left(2\, J_{\beta 0}\, r_0 + I_{opto}^\beta\right) \quad \text{(A15)}$$

We define the susceptibility matrix, $\chi_{\alpha\beta}$, as the derivative of the activity, r_α, with respect to I_{opto}^β,

$$\chi_{\alpha\beta} = -\varepsilon_\alpha \left[J^{-1}\right]_{\alpha\beta} \quad \text{(A16)}$$

At baseline $\left(I_{opto}^\beta = 0\right)$, the positivity of $r_\alpha, \forall\alpha$ imposes conditions on the recurrent and feedforward interaction strengths, $J_{\alpha\beta}$ and $J_{\alpha 0}$. The requirement that there are no "partially" balanced solutions for which one or more of the n populations is inactive or saturates and the stability of the balanced solution imposes further constraints.

B. Two-population model

Large K limit

For a two-population (one excitatory E and one inhibitory I) network, solving Eq. (A13) gives for a perturbation, I_{opto}, upon I,

$$r_E = \frac{2(J_{II}J_{E0} - J_{EI}J_{I0})r_0 - J_{EI}I_{opto}}{\Delta} \tag{A17}$$

$$r_I = \frac{2(J_{IE}J_{E0} - J_{EE}J_{I0})r_0 - J_{EE}I_{opto}}{\Delta} \tag{A18}$$

where $\Delta = J_{EI}J_{IE} - J_{EE}J_{II}$.
The requirement that at baseline the network state is fully balanced and stable implies that

$$\frac{J_{E0}}{J_{I0}} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}} \tag{A19}$$

Therefore, $\Delta > 0$.

The susceptibilities with respect to a perturbation of I are

$$\chi_{EI} = \frac{-J_{EI}}{\Delta} \tag{A20}$$

$$\chi_{II} = \frac{-J_{EE}}{\Delta} \tag{A21}$$

which both are negative. Therefore, r_E and r_I *decrease* linearly with I_{opto}, *i.e.*, the response of the I population is paradoxical.
It is useful to consider the susceptibilities normalized to baseline rate

$$\bar{\chi}_{EI} = -\frac{J_{EI}}{2(J_{II}J_{E0} - J_{EI}J_{I0})r_0} \tag{A22}$$

$$\bar{\chi}_{II} = -\frac{J_{EE}}{2(J_{IE}J_{E0} - J_{EE}J_{I0})r_0} \tag{A23}$$

Eq. (A19) implies that, $\left|\bar{\chi}_{EI}\right|$ is larger than $\left|\bar{\chi}_{II}\right|$.

Moreover, whereas $\bar{\chi}_{EI}$ is independent of J_{EE}, $\bar{\chi}_{II}$ depends on J_{EE}. When $J_{EE} = 0$, $\bar{\chi}_{II}$ is zero: the PV activity is insensitive to I_{opto}.

The identity of the two normalized susceptibilities can only be achieved with a fine-tuning of the interaction parameters such that $\Delta \simeq 0$ for

$$J_{EE} \simeq J_{EI}\frac{J_{IE}}{J_{II}} \tag{A24}$$

Concurrently, as $J_{EE} \rightarrow J_{EI}\frac{J_{IE}}{J_{II}}$, the activity of the two populations diverge as $\frac{1}{\Delta}$ with a constant ratio equal to $\frac{J_{IE}}{J_{II}}$. Thus, to keep the activities finite, $2\,(J_{II}J_{E0} - J_{EI}J_{I0})\,r_0$ and $2\,(J_{IE}J_{E0} - J_{EE}J_{I0})\,r_0$ must also tend to zero.

Finally, if $I_{opto} = I^*_{opto} \equiv 2\left(J_{E0}\frac{J_{II}}{J_{EI}} - J_{I0}\right)r_0$, r_E vanishes (Fig. 3-figure supplementary 1). When $I_{opto} > I^*_{opto}$, the balance between the total external excitatory (optogenetic+feedforward) and recurrent inhibitory inputs into I implies that r_I linearly increases with I_{opto} and the slope is $\frac{1}{J_{II}}$.

Finite K corrections to r_E and r_I near I^*_{opto}

When K is finite, r_I starts to increase with I_{opto} when r_E is exponentially small in K . To show that, we have to derive the leading order correction to the activities near I^*_{opto} .

We make the ansatz that when $I_{opto} = I^*_{opto} + \delta I\sqrt{\frac{log(K)}{K}}$, $r_E = \nu_E\frac{\sqrt{log(K)}}{K}$ and $r_I = r_I^\infty + \nu_I\sqrt{\frac{log(K)}{K}}$, where ν_E and ν_I are $O(1)$ and $r_I^\infty = 2J_{E0}\frac{r_0}{J_{EI}}$ is the inhibitory firing rate at $I_{opto} = I^*_{opto}$ in the large K limit.

To leading order:

$$r_I^\infty = \Psi_E\left[\sqrt{log\,(K)}\,(\delta I + J_{IE}\nu_E - J_{II}\nu_I)\,, A_I^\infty, B_I^\infty\right] \quad \text{(A25.1)}$$

$$\nu_E\sqrt{\frac{log(K)}{K}} = \Psi_E\left[\sqrt{log\,(K)}\,(J_{EE}\nu_E - J_{EI}\nu_I)\,, A_E^\infty, B_I^\infty\right]$$

(A25.2)

where A_α^∞ and B_α^∞, $\alpha \in \{E, I\}$, are the variance of the temporal and quenched noise in the large K limit (Eqs. A11-A12).

Equation (A25.1) implies that

$$\delta I + J_{IE}\nu_E - J_{II}\nu_I = O\left(\frac{1}{\sqrt{log(K)}}\right) \quad \text{(A26)}$$

Together with Eq (A25.2) one obtains

$$\nu_E\sqrt{\frac{log(K)}{K}} = \Psi_E\left[\left(-\,(J_{EI}\delta I + \nu_E\Delta)\,\frac{\sqrt{log(K)}}{J_{II}}\right), A_E^\infty, B_I^\infty\right] \quad \text{(A27)}$$

where $\Delta = J_{EI}J_{IE} - J_{EE}J_{II}$.

For large K,

$$\frac{\nu_E}{\sqrt{K}} = \frac{Q}{J_{II}}\,(J_{EI}\delta I + \nu_E\Delta)\,e^{-\frac{(J_{EI}\delta I + \nu_E\Delta)^2 log(K)}{\left(2A_E^\infty + B_E^\infty\right)J_{II}^2}} \quad \text{(A28)}$$

where $Q = \frac{1}{\tau_m^E\sqrt{\pi}}\frac{B_E^\infty}{(2A_E^\infty + B_E^\infty)^{\frac{3}{2}}}$.

Since ν_E must be positive, $(J_{EI}\delta I + \nu_E\Delta)$ must also be positive, Eq. (A28) then implies that to leading order

$$\nu_E = \frac{1}{\Delta}\left(J_{II}\sqrt{A_E^\infty + \frac{B_E^\infty}{2}} - J_{EI}\delta I\right) \quad \text{(A29)}$$

Hence, ν_I is

$$\nu_I = \frac{1}{\Delta}\left(J_{IE}\sqrt{A_E^\infty + \frac{B_E^\infty}{2}} - J_{EE}\delta I\right) \qquad \text{(A30)}$$

Therefore, both ν_E and ν_I decrease with δI. This holds for $\delta I \lesssim \frac{J_{II}}{J_{EI}}\sqrt{A_E^\infty + \frac{B_E^\infty}{2}}$. Beyond this range r_E is exponentially small, $\nu_I = \frac{\delta I}{J_{II}}$ and r_I increases with I_{opto}.
In conclusion, when the response of the I population is minimum the firing rate of the excitatory population is exponentially small in K.

C. Four-population model: Model 1

Large K limit

In Model 1, the population susceptibilities in response to a perturbation of the PV population are given by Eq. (A16)

$$\chi_{EI} = J_{SV}\frac{(J_{EI}J_{VS} - J_{ES}J_{VI})}{\Delta} \qquad \text{(A31)}$$

$$\chi_{II} = J_{SV}\frac{(J_{EE}J_{VS} - J_{ES}J_{VE})}{\Delta} \qquad \text{(A32)}$$

$$\chi_{SI} = J_{SV}\frac{(J_{EI}J_{VE} - J_{EE}J_{VI})}{\Delta} \qquad \text{(A33)}$$

$$\chi_{VI} = \frac{J_{SE}}{J_{SV}}\chi_{EI} \qquad \text{(A34)}$$

where $\Delta = det([J_{AB}\varepsilon_B])$.

Note, in this model we do not take into account any PV to SOM connections. Nevertheless even If one includes these, the expressions of the PC and PV susceptibility will only differ by a scaling factor from the ones in A31 and A32 (because of Δ) and therefore their sign will depends on the same conditions than A31 and A32.

Interestingly, for stable solutions $(\Delta > 0)$, then $\chi_{II} > 0$ implies that $J_{EE}\,J_{VS} > J_{ES}\,J_{VE}$. while $\chi_{EI} < 0$ implies that $J_{ES}\,J_{VI} > J_{EI}\,J_{VS}$. Therefore, $J_{EE}\,J_{VS}\,J_{VI} > J_{VE}\,J_{ES}\,J_{VI}$. and $J_{ES}\,J_{VI}\,J_{VE} > J_{EI}\,J_{VS}\,J_{VE}$. Combining the latter one has $J_{EE}\,J_{VS}\,J_{VI} > J_{EI}\,J_{VS}\,J_{VE}$. Therefore, $J_{EE}\,J_{VI} > J_{EI}\,J_{VE}$ which is equivalent to $\chi_{SI} < 0$.

Similarly one can show that if $\chi_{EE} > 0$ and $\chi_{IE} < 0$ necessarily $\chi_{SE} > 0$.

Let us consider a particular set of parameters for which a stable balanced solution exists when $J_{EE} = 0$ $(\Delta(0) > 0)$.

The susceptibility χ_{II} as a function of J_{EE} is

$$\chi_{II}(J_{EE}) = J_{SV}\frac{J_{VS}J_{EE} - J_{VE}J_{ES}}{\Delta(J_{EE})} \qquad \text{(A35)}$$

$$\Delta(J_{EE}) = -\hat{\chi}_{EE} J_{EE} + \Delta(0) \qquad \text{(A36)}$$

where $\hat{\chi}_{EE} \equiv \chi_{EE}.\Delta(J_{EE}) = J_{SV}(J_{VI}J_{IS} - J_{II}J_{VS})$, is the numerator in the susceptibility χ_{EE}.

In our models, we assumed $\chi_{EE} > 0$. When $J_{EE} = 0$, $\Delta(0)$ is positive thus, $\chi_{II}(0) < 0$. As J_{EE} increases, the sign of $\chi_{II}(J_{EE})$ depends on the order relationship between two quantities. The first one, J^*_{EE}, is the value of J_{EE} for which the numerator in Eq. (A35) changes sign

$$J^*_{EE} = \frac{J_{VE}J_{ES}}{J_{VS}} \qquad \text{(A37)}$$

The second one, J^c_{EE}, is defined by $\Delta(J^c_{EE}) = 0$

$$J^c_{EE} = \frac{\Delta(0)}{\hat{\chi}_{EE}} \qquad \text{(A38)}$$

Therefore, for $J_{EE} > J^c_{EE}$, the dynamics is unstable. Two cases can be distinguished:

1) If $J^*_{EE} < J^c_{EE}$, then χ_{II} is an increasing function of J_{EE}. It is negative if $J_{EE} < J^*_{EE}$ and becomes positive for $J_{EE} > J^*_{EE}$.
2) If $J^*_{EE} > J^c_{EE}$, χ_{II} is a decreasing function of J_{EE} and is negative in all the region where the dynamics is stable.

The derivative of χ_{II}, (Eq. (A35)), with respect to J_{EE}, has the same sign as $\chi_{EI}\chi_{IE}$. Therefore, $\chi_{EI}\chi_{IE}$ is positive in the first case and negative in the second.

Experimental data shows that the activity of the PC population decreases upon PV photostimulation, *i.e.*, $\chi_{EI} < 0$. Therefore, if $\chi_{II} > 0$ as in ALM layer 2/3, χ_{IE} must be negative, *i.e.*, the activity of the PV population decreases upon PC photostimulation.

Finite K

When I_{opto} is sufficiently strong, a fully balanced solution $(r_\alpha > 0, \forall\alpha)$ no longer exists in our case $r_E = r_V = 0$ for $I_{opto} > I^*_{opto}$ where

$$I^*_{opto} = \frac{J_{E0}(J_{IS}J_{VI} - J_{II}J_{VS}) + J_{I0}(J_{EI}J_{VS} - J_{ES}J_{VI}) + J_{V0}(J_{ES}J_{II} - J_{EI}J_{IS})}{(J_{ES}J_{VI} - J_{EI}J_{VS})} .$$

To understand the network behavior after this point we need to consider finite K corrections.

Since the PC and VIP population activities decrease with I_{opto}, when I_{opto} is sufficiently large and due to the balance of the SOM input, r_E and r_V will both be at most $O\left(\frac{1}{\sqrt{K}}\right)$. Let us write: $r_E \equiv \frac{\nu_E}{\sqrt{K}}$ and $r_V \equiv \frac{\nu_V}{\sqrt{K}}$ where ν_E and ν_V are at most $O(1)$.

One should consider four cases:

1) ν_E and ν_V are $O(1)$.

In this case, the average net input into the SOM population, $u_S = J_{SE}\nu_E - J_{SV}\nu_V$, is $O(1)$ and the temporal fluctuations, B_S, and heterogeneities, A_S, are negligible. If u_S is larger than the rheobase, $\frac{(V_{th}-V_R)}{g^S_{leak}}$, r_S is also $O(1)$. Otherwise, $r_S = 0$.

Because ν_E and ν_V are $O(1)$, u_E and u_V are $o\left(\frac{1}{\sqrt{K}}\right)$. Thus, to leading order,

$$2J_{E0}r_0 - J_{EI}r_I - J_{ES}r_S = 0 \tag{A39}$$

$$2J_{V0}r_0 - J_{VI}r_I - J_{VS}r_S = 0 \tag{A40}$$

Moreover, the balance of the PV population implies that

$$2J_{I0}r_0 + I_{opto} - J_{II}r_I - J_{IS}r_S = 0 \tag{A41}$$

Thus, there are three linear equations (Eqs. (A39-40-41)) for two unknowns $(r_I \text{ and } r_s)$. These cannot be satisfied and hence, in this case, there is no consistent solution.

2) $\nu_E = o(1)$ and $\nu_V = O(1)$.

Here, to leading order, $u_S = -J_{SV}\nu_V < 0$, while $A_S = B_S = 0$. As a result, to leading order, $r_S = 0$. The activity of the PV population is then

$$r_I = \frac{(2J_{I0}r_0 + I_{opto})}{J_{II}} \tag{A42}$$

Because ν_V is $O(1)$,

$$2J_{V0}r_0 - J_{VI}r_I = 0 \tag{A43}$$

Eqs. (A42, A43) cannot both be satisfied. This solution is also inconsistent.

3) $\nu_E = O(1)$ and $\nu_V = o(1)$.

In this case $u_S = J_{SE}\nu_E > 0$ and therefore r_S can be $O(1)$. Eqs. (A39) and (A41) imply

$$2J_{E0}r_0 - J_{EI}r_I - J_{ES}r_S = 0 \tag{A44}$$

$$2J_{I0}r_0 + I_{opto} - J_{II}r_I - J_{IS}r_S = 0 \tag{A45}$$

which determine r_I and r_S as $r_I = \frac{(J_{ES}J_{I0} - J_{IS}J_{E0})r_0 + J_{ES}I_{opto}}{J_{ES}J_{II} - J_{EI}J_{IS}}$ and $r_S = \frac{(J_{II}J_{E0} - J_{EI}J_{E0})r_0 - J_{EI}I_{opto}}{J_{ES}J_{II} - J_{EI}J_{IS}}$.

Provided that the parameters are such that they are positive, ν_E is given by

$$r_S = \Psi_S\left[J_{SE}\nu_E, 0, 0\right]$$

(A46)

Finally, since $\nu_V = o(1)$ consistency implies that

$2J_{V0}r_0 - J_{VI}r_I - J_{VS}r_S < 0$ (A47)

This solution is valid for a finite range of I_{opto}. It exists as long as $r_s > 0$ which implies that $J_{E0}\frac{J_{II}}{J_{EI}} - J_{I0} > I_{opto} > I^*_{opto}$.

4) $\nu_E = o(1)$ and $\nu_V = o(1)$.

Here, $u_S = A_S = B_S = 0$ and thus, $r_S = 0$. This solution exists only for sufficiently large I_{opto} such that u_E and u_V are $O\left(\sqrt{K}\right)$ and negative. Therefore, PV is the only active population and r_I is given by Eq. (A40).

In conclusion, in this model at the minimum of r_I, r_E is of order $\frac{1}{\sqrt{K}}$ in contrast to the two-population case where r_E is exponentially small in K.

D. Four-population model: Model 2

Large K limit

To get insights on the network architecture that could explain the proportional paradoxical effect observed in layer 5 of ALM and S1, we first considered a three-population network consisting of the PC, PV and SOM populations (Fig. 9A).

In this network, the population activities are

$$r_E = J_{SI}\frac{2(J_{ES}J_{I0} - J_{IS}J_{E0})r_0 + J_{ES}I_{opto}}{\Delta} \quad \text{(A48)}$$

$$r_I = \frac{J_{SE}}{J_{SI}}r_E \quad \text{(A49)}$$

$$r_S = \frac{2((J_{II}J_{SE} - J_{IE}J_{SI})J_{E0} - (J_{EI}J_{SE} - J_{EE}J_{SI})J_{I0})r_0 - (J_{EI}J_{SE} - J_{EE}J_{SI})I_{opto}}{\Delta} \quad \text{(A50)}$$

where $\Delta = (J_{II}J_{SE} - J_{IE}J_{SI})J_{ES} + (J_{EE}J_{SI} - J_{EI}J_{SE})J_{IS} > 0$.

The full balance of the network activities implies

$$\frac{J_{ES}}{J_{IS}} > \frac{2J_{E0}r_0}{2J_{I0}r_0 + I_{opto}} > \frac{J_{EI}}{J_{II}} \quad \text{(A51)}$$

The inequality on the left side stems from the positivity of the rates. The inequality on the right side stems from the fact that the balanced state is the only solution of the dynamics, namely that no partially balanced solution (in particular, $r_E = 0$, $r_I = O(1)$ and $r_S = 0$ and $r_E = 0$, $r_I = O(1)$ and $r_S = O(1)$) exists.

r_E and r_I are proportional (Eq. A49) and increase with I_{opto}. As a consequence, the network never exhibits the paradoxical effect.

In this three-population network, the proportionality of r_E and r_I stems from the balance of inputs into the SOM population. To account for the proportional *paradoxical effect*, we consider a network model with an additional inhibitory population, denoted X (Fig. 9B). Because in this network the SOM neurons only receive inputs from PCs and PV neurons, here, the balance of the SOM input also ensure the proportionality of r_E and r_I.

The susceptibilities upon PV stimulation are

$$\chi_{EI} = J_{SI}\frac{(J_{ES}J_{XX} - J_{EX}J_{XS})}{\Delta} \tag{A52}$$

$$\chi_{II} = \frac{J_{SE}}{J_{SI}}\chi_{EI} \tag{A53}$$

$$\chi_{SI} = \frac{(J_{EE}J_{SI}J_{XX} - J_{XE}J_{SI}J_{XE} - J_{EI}J_{SE}J_{XX})}{\Delta} \tag{A54}$$

$$\chi_{XI} = \frac{(J_{ES}J_{SI}J_{XE} + J_{EI}J_{SE}J_{XS} - J_{EE}J_{SI}J_{XS})}{\Delta} \tag{A55}$$

where $\Delta = det([J_{AB}\varepsilon_B])$ (see Material and Methods).

Paradoxicality implies that

$$J_{EX} > J^{*}_{EX} \equiv \frac{J_{ES}J_{XX}}{J_{XS}} \tag{A56}$$

The susceptibilities upon PC stimulation are

$$\chi_{EE} = J_{SI}\frac{(J_{IX}J_{XS} - J_{IS}J_{XX})}{\Delta} \tag{A57}$$

$$\chi_{IE} = \frac{J_{SE}}{J_{SI}}\chi_{EE} \tag{A58}$$

$$\chi_{SE} = \frac{(J_{IX}J_{SI}J_{XE} + J_{II}J_{SE}J_{XX} - J_{IE}J_{SI}J_{XX})}{\Delta} \tag{A59}$$

$$\chi_{XE} = \frac{(J_{IE}J_{SI}J_{XS} - J_{IS}J_{SI}J_{XE} - J_{II}J_{SE}J_{XS})}{\Delta} \tag{A60}$$

Therefore, the PC population activity increases upon PC stimulation if

$$J_{IX}J_{XS} > J_{IS}J_{XX} \tag{A61}$$

One can find a range of parameters (*e.g.* Fig. 9C) such that:

1) The relative decrease in the SOM population is larger than that in the E and I populations. As a consequence, as I_{opto} is increased, r_S approaches zero when the PC and PV activities are still finite.

2) As I_{opto} is increased further, the network settles into a partially balanced state where r_E, r_I and r_X are finite and r_I increases with I_{opto}, while r_E continues to decrease.

Thus, r_I reaches its minimum value when r_E is finite even in the large K limit.

Conclusion of the chapter

In this work, we investigated how the optogenetic photostimulation of the PV neurons affects the responses of PCs and PV neurons in both ALM layer 2/3 and layer 5 and S1. We inferred from our observations the properties in the network architectures that are essential to account for the responses in each of these areas.

We showed that to understand the effects of these manipulations it is crucial to take into account the diversity of the inhibitory neurons. We expect that this diversity is essential to understand the effects of other optogenetic perturbations and in particular, when all the inhibitory subtypes are stimulated. Indeed, each subtype contributes to very specific loops in the connectivity. Therefore, the network response is likely to still depend on specific disinhibitory patterns.

In the next chapter, we will investigate the effects of the optogenetic perturbation of cortical circuits on a different scale. We will consider how we can infer from optogenetic manipulations, information on the characteristic length of the interactions between cortical neurons.

References

Asrican B, Augustine GJ, Berglund K, Chen S, Chow N, Deisseroth K, Feng G, Gloss B, Hira R, Hoffmann C, Kasai H, Katarya M, Kim J, Kudolo J, Lee LM, Lo SQ, Mancuso J, Matsuzaki M, Nakajima R, Qiu L, Tan G, Tang Y, Ting JT, Tsuda S, Wen L, Zhang X, Zhao S. 2013. Next-generation transgenic mice for optogenetic analysis of neural circuits. *Front Neural Circuits* **7**:160.

Chow BY, Han X, Dobry AS, Qian X, Chuong AS, Li M, Henninger MA, Belfort GM, Lin Y, Monahan PE, Boyden ES. 2010. High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**:98–102.

Deisseroth K. 2015. Optogenetics: 10 years of microbial opsins in neuroscience. *Nature Neuroscience*. doi:10.1038/nn.4091

Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K. 2017. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545**:181–186.

Hikosaka O, Wurtz RH. 1985. Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus. *J Neurophysiol* **53**:266–291.

Hooks BM, Lin JY, Guo C, Svoboda K. 2015. Dual-channel circuit mapping reveals sensorimotor convergence in the primary motor cortex. *J Neurosci* **35**:4418–4426.

Klapoetke NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GK-S, Boyden ES. 2014. Independent optical excitation of distinct neural populations. *Nat Methods* **11**:338–346.

Krupa DJ, Ghazanfar AA, Nicolelis MA. 1999. Immediate thalamic sensory plasticity depends on corticothalamic feedback. *Proc Natl Acad Sci U S A* **96**:8200–8205.

Lashley KS. 1931. MASS ACTION IN CEREBRAL FUNCTION. *Science* **73**:245–254.

Lin JY, Knutsen PM, Muller A, Kleinfeld D, Tsien RY. 2013. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat Neurosci* **16**:1499–1508.

Long MA, Fee MS. 2008. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*. doi:10.1038/nature07448

Madisen L, Mao T, Koch H, Zhuo J-M, Berenyi A, Fujisawa S, Hsu Y-WA, Garcia AJ 3rd, Gu X, Zanella S, Kidney J, Gu H, Mao Y, Hooks BM, Boyden ES, Buzsáki G, Ramirez JM, Jones AR, Svoboda K, Han X, Turner EE, Zeng H. 2012. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat Neurosci* **15**:793–802.

Mishkin M, Ungerleider LG. 1982. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res* **6**:57–77.

Newsome W, Wurtz R. 1988. Probing visual cortical function with discrete chemical lesions. *Trends in Neurosciences*. doi:10.1016/0166-2236(88)90076-8

Ponce CR, Lomber SG, Born RT. 2008. Integrating motion and depth via parallel pathways. *Nat Neurosci* **11**:216–223.

Wiegert JS, Mahn M, Prigge M, Printz Y, Yizhar O. 2017. Silencing Neurons: Tools, Applications, and Experimental Constraints. *Neuron* **95**:504–529.

Yizhar O, Fenno LE, Davidson TJ, Mogri M, Deisseroth K. 2011. Optogenetics in neural systems. *Neuron* **71**:9–34.

Zhang F, Wang L-P, Brauner M, Liewald JF, Kay K, Watzke N, Wood PG, Bamberg E, Nagel G, Gottschalk A, Deisseroth K. 2007. Multimodal fast optical interrogation of neural circuitry. *Nature* **446**:633–639.

Zhao S, Ting JT, Atallah HE, Qiu L, Tan J, Gloss B, Augustine GJ, Deisseroth K, Luo M, Graybiel AM, Feng G. 2011. Cell type–specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat Methods* **8**:745–752.

Chapter 3

Mechanisms underlying the spatial photosupression of cortical excitatory activity

The response of networks of one excitatory (E) and one inhibitory (I) population of neurons to perturbations of their inhibitory population has been the subject of numerous recent studies. The majority of these works focused on describing the response of networks with unstructured connectivity. Here, we investigate the response of spatially structured networks of neurons with strong recurrent interactions to the stimulation of their inhibitory population with a Gaussian-shaped perturbation. When the connectivity decays exponentially with distance, we show that upon strong perturbation the balance of excitation and inhibition is disrupted on a characteristic length, x_c, where excitatory neurons are silenced. We derive the dependence of x_c with the intensity and radius of the perturbation. Our models account for the experimental observations of (Li et al., 2019) in the mouse barrel cortex. Notably, we give general predictions on the dependence of the photo-suppression length, x_c, with the intensity and radius of the laser. Finally, for finite-size networks, we propose conditions upon the length of the recurrent interactions to explain the concomitant decrease of the excitatory and inhibitory responses in the surround of the suppression.

Introduction

In the mouse cortex, excitatory connections are sparse with connection probability falling off sharply with distance (Levy and Reyes, 2012), inputs are integrated from nearby neurons (*i.e.* 250 μm to 500 μm). Long-range connections (*i.e.* 500 μm to 1500 μm) between neurons are known to exist but are thought to be made between neurons with similar functional properties (Stepanyants et al., 2009). Thalamocortical projections are known to be broader than intracortical interactions (Freund et al., 1989; Landry and Deschênes, 1981) and it is commonly admitted that excitatory connections are broader than their inhibitory counterpart (Fino and Yuste, 2011; Hioki et al., 2013; Kato et al., 2017; Packer and Yuste, 2011).

In a recent study, (Li et al., 2019) measured the effects of the ChR-assisted photoactivation of the PV neurons in the mouse barrel cortex. They limited the viral expression of ChR to a laterally confined region within a $300\,\mu m$ radius across all layers (Fig. 1A-B). They activated these localized PV neurons with a Gaussian shaped laser for different intensities. They measured the responses of the PCs and PV neurons at different distances from the stimulation center with a set of electrodes evenly spaced between the site of viral expression and $1200\,\mu m$ of cortical distance. Surprisingly, they reported that the photostimulation of PV neurons suppressed the activity of the PCs in a region that extended beyond the site of viral expression. For large laser intensities, PCs were completely silenced up to $0.4\,mm$ from the center of the light and totally recovered only at a distance of $1\,mm$ from the center of the laser beam (Fig. 1C). Conversely, PV neurons were highly activated up to $0.2\,mm$ from the center of the laser and exhibited an activity decrease in the surrounding regions up to $0.6\,mm$ from the center where they recovered to their baseline activity (Fig. 1C).

Over the range of distances where the activity of the neurons decreased, the relative suppression of the excitatory and inhibitory activity appears to be proportional at all light intensity. Figure 2 plots the relative spike rate versus the lateral distance from the laser center when the viral expression is not localized. Because of the scattering of the light through brain tissues, a larger set of neurons is affected by the light. For sufficiently large laser intensities (Fig 2A and B left), the relative spike rates (normalized to baseline) of the PCs and PV neurons in the surround of the center of the stimulation ($\sim 1\,mm$ to $3\,mm$) are nearly identical.

Previous works have investigated the dynamics of two-population networks with spatially structured connectivity (Ben-Yishai et al., 1997; Ebsch and Rosenbaum, 2018; Rosenbaum and Doiron, 2014; Van Vreeswijk and Sompolinsky, 2005). Few studies consider networks of strongly interacting neurons operating in the balanced regime. (Rosenbaum and Doiron, 2014) have investigated the conditions that recurrent and feedforward interactions have to satisfy for the existence of a balanced regime in networks where the probability of connection between neurons is Gaussian. They found that for a network to operate in a balanced state, feedforward interactions must be broader than recurrent excitatory connections that in turn must be broader than inhibitory interactions. Consequently, they concluded that mechanisms that are thought to rely on broad lateral inhibitory interactions (*e.g.* tuning curve sharpening (Shapley et al., 2003) cannot be accounted for in a balanced network.

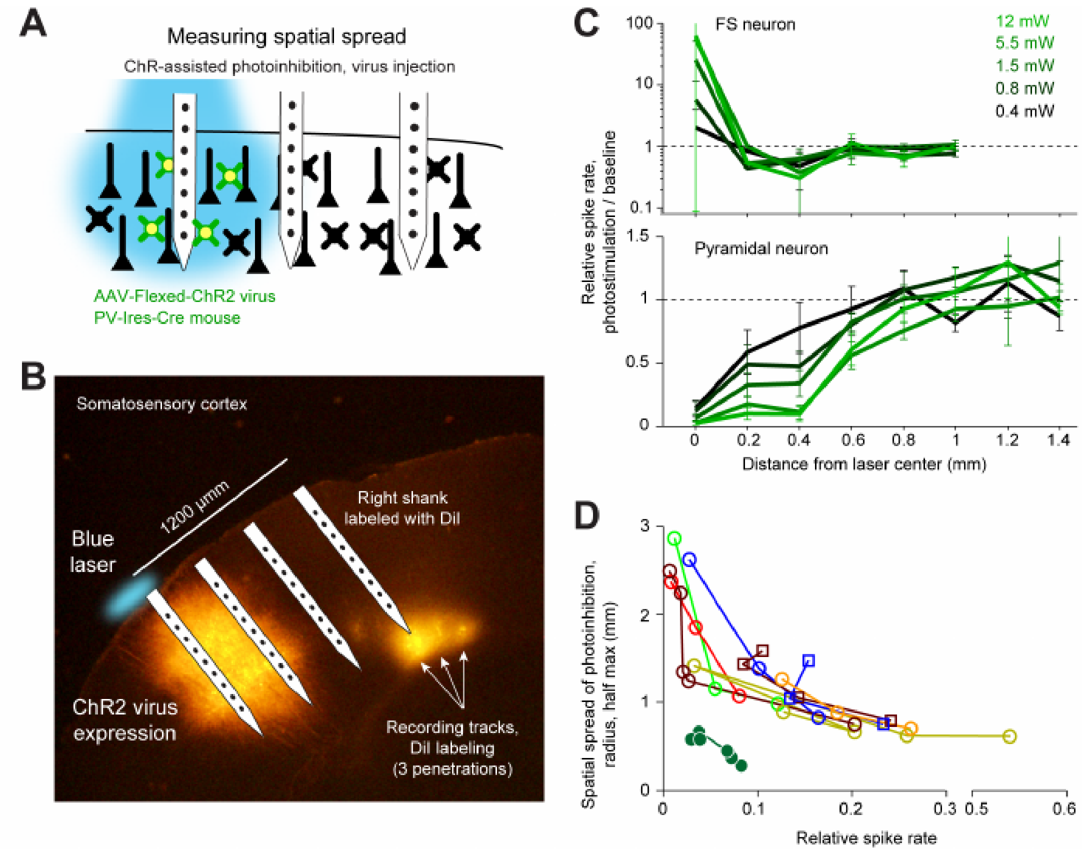


Figure 1. ChR-assisted photoinhibition using virus injection can achieve submillimeter spatial resolution. (A) Schematics, confined ChR2 expression in PV neurons and silicon probe recording at different distances from the expression site. (B) Silicon probe recording in the barrel cortex during photostimulation. The right shank of the silicon probe was painted with DiI to label the recording tracks. Coronal section showing viral expression of ChR2-tdTomato, electrode and photostimulus locations. The photostimulus was aligned to the virus injection site. (C) Relative spike rate versus lateral distance from the photostimulus center for different laser powers. Top, putative FS neurons (n = 14). Bottom, pyramidal neurons (n = 78). Neurons were pooled across cortical depths. (D) Photoinhibition strength versus spatial spread. Relative spike rate is the average across all pyramidal neurons near the laser center (<0.4 mm, all cortical depths). Spatial spread is the distance at which photoinhibition strength is half of that at the laser center (‘radius, half-max’). Each circle represents data from one photostimulation power. Lines connect all circles of one method.

Here, we consider two-population network models with exponentially decaying distance-dependent interactions to explain the observed responses of principal excitatory cells (PCs) and parvalbumin-positive (PV) neurons to optogenetic perturbations of the PV neurons in the superficial layers of the mouse barrel cortex. We investigate how local interactions account for the photo-suppression of the excitatory activity and shape the network responses. Spatially structured interactions result in a local competition between the network activities that lead to a modulation of the responses over distances that extend beyond the radius of the photostimulation.

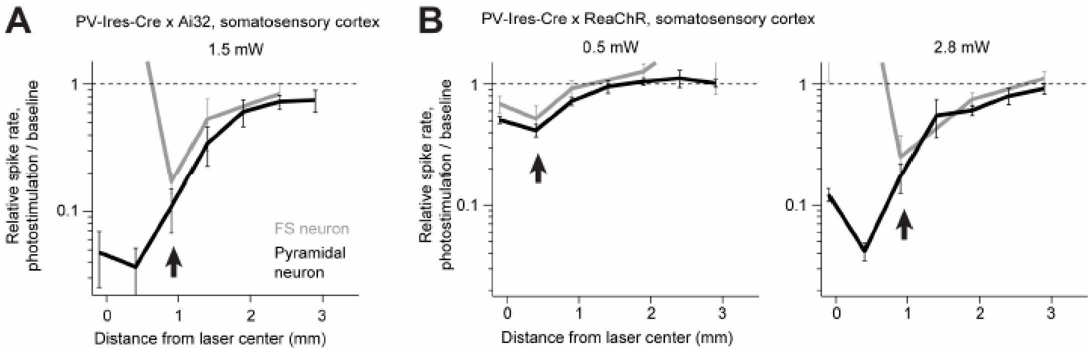


Figure 2. Proportional activity decrease in pyramidal and FS neurons during ChR-assisted photoinhibition. (A) Relative spike rate versus lateral distance from the photostimulus center for PV-IRES-Cre x Ai32. Data from Figure 5C replotted with activity shown on a log scale. PV neurons (gray) and pyramidal neurons (black). The arrows point to regions in the photostimulus surround where the activity of PV neurons and pyramidal neurons decrease in proportion (paradoxical effect). (B) Same as (A) but for PV-IRES-Cre x ReaChR.

We have shown in a previous study (Mahrach et al., 2019) that two-population balanced network models were not sufficient to describe the effects of the photostimulation of PV neurons when the radius of the laser is large. A complete understanding of the mechanisms underlying the spatial responses of the PCs and PV neurons would also require the modeling of networks with multiple inhibitory populations. We here give as a necessary prerequisite, an account of these mechanisms in two-population network models.

Network models

We consider networks of one excitatory (E) and one inhibitory (I) populations of neurons (Fig. 3A), consisting of N randomly strongly connected neurons. The neurons lie on a ring of size $L = 3\ mm$, and we denote the position of a neuron (A, i) in population A as x_i^A. The neurons are uniformly distributed across the ring, namely, $x_i^A = \frac{iL}{N_A}$. The probability of connection between the neurons is assumed to have a periodic exponential profile (Fig. 3B)

$$P^{AB}_{i,j} = \frac{Z_B}{\lambda_{AB}} \sum_{k=-\infty}^{+\infty} e^{-\frac{|x_i^A - x_j^B + k\,L|}{\lambda_{AB}}} \tag{1}$$

where Z_B is a normalization factor such that $\sum_j P_{ij}^{AB} = K$, so that the neurons on average receive K inputs per population. λ_{AB} is the characteristic length of the decay of the interaction from population B to population A with distance. We will abusively refer to it as the interaction length from population B to population A.

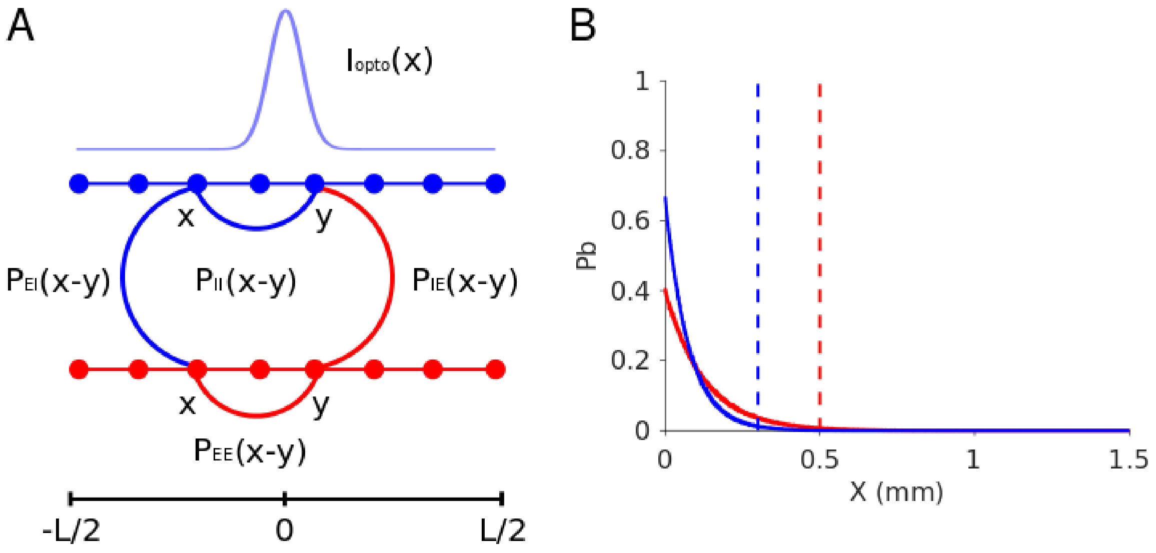


Figure 3. Network model and connection probability. A. Scheme of the network architecture (ring model). B. Profile of the probability of connections (example with $\lambda_{AB} = \lambda_B$). Dashed-line: red: $4\,\lambda_E = 0.5\ m$; blue: $4\,\lambda_I = 0.3\ m$.

We first assume that $\lambda_{AB} = \lambda_B$, in other words, that the length of the interaction between the neurons only depends on the presynaptic population.

The neurons receive homogeneous inputs from sources outside of the network (*e.g.* thalamus) with a fixed rate r_0 and strong interactions strength, $j_{A0} = J_{A0}\sqrt{K}$. We also assume that the neurons interact through strong synapses and denote by $j_{AB} = \frac{J_{AB}}{\sqrt{K}}$ the absolute strength of the connection from population B to population A.

We model the effect of the photostimulation as a strong additional current into the inhibitory neurons with a Gaussian shape of variance σ_0 and amplitude $\Gamma_0\sqrt{K}$:

$$\mathcal{I}(x) = \sqrt{K}\, I_I(x) = \sqrt{K}\, \frac{\Gamma_0}{\sigma_0\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma_0^2}} \tag{2}$$

Globally balanced solutions

The net input, $u_A(x)$, into a neuron (A, x) in population A at the position x is

$$u_A(x) = \sqrt{K}\left(I_A(x) + J_{A0}\, r_0 + \sum_B \frac{J_{AB}}{\lambda_B}\, \varepsilon_B \int_{-\infty}^{+\infty} dy\, e^{-\frac{|x-y|}{\lambda_B}}\, r_B(y)\right) \tag{3}$$

where J_{AB} is the absolute scaled strength of the interaction from population B to population A, and $\varepsilon_A = \pm 1$ depending on whether A is excitatory or inhibitory. Here, $r_A(x)$ is the activity of the neuron (A, x).

Given that the profile of the interactions is exponential, one easily shows that the inputs into the neurons are the solutions of a system of two second-order differential equations

$$u''(x) - M\,.\,u(x) = \sqrt{K}\left\{I''(x) - (J\,.\,\Lambda^{-2}\,.\,J^{-1})\,.\,I(x) - 2\,(J\,.\,\Lambda^{-2})\,.\,r(x)\right\} \tag{4}$$

where $\Lambda_{AB} = \delta_{AB}\,\lambda_B$ and δ_{AB} is the Kronecker symbol.

Each differential equation has to be provided with a proper set of boundary conditions. Assuming that the solutions are symmetric we have

$$r_A(\tfrac{L}{2}) = r_A(-\tfrac{L}{2}) \tag{5}$$

In addition, we assume that the size of the ring, L, is large in comparison with the interaction length and with the perturbation radius. Therefore, at $x = \frac{L}{2}$, the network behaves as if it was not perturbed implying that

$$\frac{dr_A}{dx}(\tfrac{L}{2}) = 0 \tag{6}$$

Analytical solutions to the set of self-consistent equations of the dynamics can then be found numerically. However, these can be tricky to solve given that they involve the resolution of nonlinear transcendental equations.

Theory in the large N, K limit

When the average number of connections K becomes infinitely large, the network operates in a regime where excitation and inhibition are balanced if the net input into a neuron (A, x) remains finite and non-zero (Van Vreeswijk and Sompolinsky, 2005). This implies that for each neuron, the right-hand side of Eq. [3] must be $O(\frac{1}{\sqrt{K}})$. Similarly, the balance of the net input into a neuron at any position x implies that the right-hand side of Eq. [4] is zero. Therefore, the activities are

$$r_E^{\infty}(x) = r_E^0 - \frac{J_{EI}}{2|J|}\frac{\Gamma_0}{\sigma_0\sqrt{2\pi}}\left(1 + \frac{\lambda_E^2}{\sigma_0^2}\left(1 - \frac{x^2}{\sigma_0^2}\right)\right) e^{-\frac{x^2}{2\sigma_0^2}} \tag{7}$$

$$r_I^{\infty}(x) = r_I^0 - \frac{J_{EE}}{2|J|}\frac{\Gamma_0}{\sigma_0\sqrt{2\pi}}\left(1 + \frac{\lambda_I^2}{\sigma_0^2}\left(1 - \frac{x^2}{\sigma_0^2}\right)\right) e^{-\frac{x^2}{2\sigma_0^2}} \tag{8}$$

where the baseline rates are $r_E^0 = \frac{1}{2}\frac{J_{E0}\,J_{II} - J_{I0}\,J_{EI}}{J_{EI}\,J_{IE} - J_{EE}\,J_{II}} r_0$ and $r_I^0 = \frac{1}{2}\frac{J_{E0}\,J_{IE} - J_{I0}\,J_{EE}}{J_{EI}\,J_{IE} - J_{EE}\,J_{II}} r_0$.

The solution is balanced as long as $r_E^\infty(x)$ and $r_I^\infty(x)$ are positive for any given x.

At the center, the responses of E and I are minimum and below their baseline value

$$r_E^\infty(0) = r_E^0 - \frac{J_{EI}}{2|J|} \frac{\Gamma_0}{\sigma_0\sqrt{2\pi}} \left(1 + \frac{\lambda_E^2}{\sigma_0^2} \right) \tag{9}$$

$$r_I^\infty(0) = r_I^0 - \frac{J_{EE}}{2|J|} \frac{\Gamma_0}{\sigma_0\sqrt{2\pi}} \left(1 + \frac{\lambda_I^2}{\sigma_0^2} \right) \tag{10}$$

Therefore, the responses at the center are always paradoxical: the activities of both E and I decrease with the intensity of the perturbation Γ_0. When Γ_0 is too large, the activity of either E or I reaches zero and the balance no longer holds. This occurs for an intensity

$$\Gamma_0^* = min\left[\frac{J_{E0}\,J_{II} - J_{I0}\,J_{EI}}{J_{EI}\,(\lambda_E^{\hat{2}} + \sigma_0^2)},\ \frac{J_{E0}\,J_{IE} - J_{I0}\,J_{EE}}{J_{EE}\,(\lambda_I^{\hat{2}} + \sigma_0^2)}\right] r_0\,\sigma_0^3\,\sqrt{2\,\pi} \tag{11}$$

When $\Gamma_0 < \Gamma_0^*$, the response of E exhibits a maximum at $x_E^{max} = \frac{\sigma_0\,\sqrt{3\,\lambda_E^2 + \sigma_0^2}}{\lambda_E}$. Respectively, the inhibitory response is maximum at $x_I^{max} = \frac{\sigma_0\,\sqrt{3\,\lambda_I^2 + \sigma_0^2}}{\lambda_I}$. The values of the normalized activities at their maximum are

$$\bar{r}_E(x_E^{max}) \equiv \frac{r_E^\infty(x_E^{max})}{r_E^0} = 1 + \sqrt{\frac{2}{\pi}}\,\frac{J_{EI}}{|J|}\,\frac{\Gamma_0\,\lambda_E^2}{\sigma_0^3}\,e^{-\frac{3}{2}}\,e^{-\frac{\sigma_0^2}{2\,\lambda_E^2}} \tag{12}$$

$$\bar{r}_I(x_I^{max}) \equiv \frac{r_I^\infty(x_I^{max})}{r_I^0} = 1 + \sqrt{\frac{2}{\pi}}\,\frac{J_{EE}}{|J|}\,\frac{\Gamma_0\,\lambda_I^2}{\sigma_0^3}\,e^{-\frac{3}{2}}\,e^{-\frac{\sigma_0^2}{2\,\lambda_I^2}} \tag{13}$$

Therefore, at their maximum, the responses of E and I increase with Γ_0.

One easily shows that a necessary condition for a balanced state to be stable is $\lambda_E > \lambda_I$. Remarkably, when this is the case $x_I^{max} > x_E^{max}$. Therefore, for $x < x_E^{max}$ the responses of E and I decrease with Γ_0. For $x_E^{max} < x < x_I^{max}$, the response of E increases with Γ_0 while the response of I decreases. Finally, for $x > x_I^{max} > x_E^{max}$, both the responses of E and I increase with Γ_0.

Partially balanced solutions

We want to investigate the response of the network to large perturbations of its inhibitory population $(\Gamma_0 > \Gamma_0^*)$. We expect that if the perturbation is large enough, the inhibitory neurons in the perturbed region will increase their activity and consequently silence their excitatory neighbors. Therefore, we look for solutions to the network dynamics where the excitatory population is silenced within a given region, namely $r_E(x) = 0$ in $[-x_c; x_c]$. For reasons of symmetry, we will only investigate the network responses in $[0;\frac{L}{2}]$, and label $\{-\}$ the solutions of the dynamics in $[0;x_c]$ and $\{+\}$ the ones in $[x_c;\frac{L}{2}]$.

The net input into neuron x in population A can be expressed in terms of the activities on the left side and right side of x_c as

$$u_A(x) = \sqrt{K}\ \{I_A(x) + J_{A0}\, r_0 - \sum_B \frac{J_{AB}}{\lambda_B}\ (\int_{x_c}^{+\infty} dy\ (e^{-\frac{|x+y|}{\lambda_B}} + e^{-\frac{|x-y|}{\lambda_B}})\ r_B^+(y)$$

$$+ \int_0^{x_c} dy\ (e^{-\frac{|x+y|}{\lambda_B}} + e^{-\frac{|x-y|}{\lambda_B}})\ r_B^-(y)\} \qquad (14)$$

Activities in $[x_c;\frac{L}{2}]$

For a neuron in $[x_c;\frac{L}{2}]$, Eq. [14] simplifies as

$$u_A^+(x) = \sqrt{K}\ \{I_A(x) + J_{A0}\, r_0 + 2\sum_B \frac{J_{AB}}{\lambda_B}\ (e^{-\frac{x}{\lambda_B}}[\int_{x_c}^{x} dy\ cosh(\frac{y}{\lambda_B})\ r_B^+(y) + \int_0^{x_c} dy\ cosh(\frac{y}{\lambda_B})\ r_B^-(y)]$$

$$+ cosh(\frac{x}{\lambda_B}) \int_x^{+\infty} dy\ e^{-\frac{y}{\lambda_B}}\ r_B^+(y))\} \qquad (15)$$

One easily shows that the input into the neurons in $[x_c;\frac{L}{2}]$ is the solutions of the same system of two second-order differential equation as in the globally balanced scenario:

$$u_+''(x) - M.\, u_+(x) = \sqrt{K}\{I''(x) - M.\,(I(x) + I_0) - 2\,(J\,.\,\Lambda^{-2})\,.\,r_+(x)\} \quad (16)$$

where $M = J\,.\,\Lambda^{-2}\,.\,J^{-1}$, $I_0 = (J_{E0}\, r_0; J_{I0}\, r_0)$ and $\Lambda_{AB} = \delta_{AB}\, \lambda_B$.

Activities in $[0;x_c]$

For the neurons in $[0;x_c]$, we are only interested in the response of the inhibitory population since we assumed $r_E^-(x) = 0, \forall x \in [0;x_c]$. Eq. [14] then simplifies as

$$u_I^-(x) = \sqrt{K}\ \{I_I(x) + J_{I0}\, r_0 + 2\, \frac{J_{IE}}{\lambda_E}\ (cosh(\frac{x}{\lambda_E}) \int_{x_c}^{+\infty} dy\ e^{-\frac{y}{\lambda_E}}\ r_E^+(y))$$

$$-2\, \frac{J_{II}}{\lambda_I}(cosh(\frac{x}{\lambda_I})[\int_{x_c}^{+\infty} dy\ e^{-\frac{y}{\lambda_I}}\ r_I^+(y) + \int_x^{x_c} dy\ e^{-\frac{y}{\lambda_I}}\ r_I^-(y)]$$

$$+ e^{-\frac{x}{\lambda_I}} \int_0^{x} dy\ cosh(\frac{y}{\lambda_I})\ r_I^-(y))\} \qquad (17)$$

One easily shows that the input into the inhibitory neurons in $[0;x_c]$ is the solution of the following second-order differential equation

$$\frac{d^2u_I^-}{dx^2}(x)-\frac{u_I^-(x)}{\lambda_I^2} = \sqrt{K}\{\, I_I''(x) - \frac{I_I(x)+J_{I0}\, r_0}{\lambda_I^2} + 2\frac{J_{II}}{\lambda_I^2}\, r_I^-(x) \; + 2\, \frac{J_{IE}}{\lambda_E}\, \left(\frac{1}{\lambda_E^2}-\frac{1}{\lambda_I^2}\right)$$

$$\times\, cosh(\tfrac{x}{\lambda_E}) \int_{x_c}^{+\infty} dy\; e^{-\frac{y}{\lambda_E}}\, r_E^+(y)\} \qquad (18)$$

Surprisingly, when $\lambda_E = \lambda_I$, the left and right solutions uncouple. On the left side, $r_I^-(x)$ behaves as if the network consisted of a single inhibitory population receiving an external input $I_I(x)$. On the right-hand side, the excitatory and inhibitory populations are balanced. However, when $\lambda_E \neq \lambda_I$, the inhibitory response in the region $[0; x_c]$, depends on the spatially averaged response of the active excitatory neurons in the region $[x_c; \frac{L}{2}]$.

Suppression length x_c

The length of the region where the excitatory activity is silenced, x_c, has to be determined self-consistently. Equations [16] and [18] have to be provided with a proper set of boundary conditions. Assuming that the solutions are continuous, the continuity of the rates and of their first derivative at $x = x_c$ implies

$$r_A^+(x_c) = r_A^-(x_c)$$

$$\frac{dr_A^+}{dx}(x_c) = \frac{dr_A^-}{dx}(x_c) \qquad (19)$$

In addition, when the size of the ring, L, is large in comparison with the length of the interactions and with the perturbation radius, we can assume that when $x = \frac{L}{2}$, the network behaves as if it was not perturbed implying that

$$\frac{dr_A^+}{dx}(\tfrac{L}{2}) = 0 \qquad (20)$$

Theory in the large N, K limit

In the limit where K goes to infinity, the steady-state of the network is given by setting the right terms of each differential equation to zero.

In the region $[x_c; \frac{L}{2}]$ the network behaves as a globally balanced one

$$r^+(x) = \tfrac{1}{2}\, (\Lambda^2 . J^{-1} . I''(x) - J^{-1} . (I(x) + I_0)) \qquad (21)$$

The activity of the inhibitory neuron x in $[0; x_c]$ is

$$r_I^-(x) = \frac{\lambda_I^2}{2\, J_{II}}(\frac{I_I(x)+J_{I0}\, r_0}{\lambda_I^2} - I_I''(x) + 2\, \frac{J_{IE}}{\lambda_E}\, (\frac{1}{\lambda_I^2}-\frac{1}{\lambda_E^2})\, cosh(\tfrac{x}{\lambda_E}) \int_{x_c}^{+\infty} dy\; e^{-\frac{y}{\lambda_E}}\, r_E^+(y)) \qquad (22)$$

To completely figure out the network response, we need to determine the value of x that separates the two solutions. x_c has to be determined self-consistently given its definition: it is the minimum value of x for which $r_I^-(x) > 0,\ r_E^-(x) = 0$ no longer stands, implying that $u_E^-(x_c^-) = 0$ and $u_E^-(x_c^+) > 0$. This leads to finding the roots of the transcendental equation that determines the net input into E at x_c,

$$J_{E0}\, r_0 + 2\ \{\tfrac{J_{EE}}{\lambda_E} (cosh(\tfrac{x_c}{\lambda_E}) \int_{x_c}^{+\infty} dy\ e^{-\frac{y}{\lambda_E}}\ r_E^+(y)) - \tfrac{J_{EI}}{\lambda_I}\ (cosh(\tfrac{x_c}{\lambda_I})[\int_{x_c}^{+\infty} dy\ e^{-\frac{y}{\lambda_I}}\ r_I^+(y)]$$

$$+ e^{-\frac{x_c}{\lambda_I}} \int_0^{x_c} dy\ cosh(\tfrac{y}{\lambda_I})\ r_I^-(y))\} = 0 \qquad (23)$$

The local susceptibility of the inhibitory neurons in $[0; x_c]$ is

$$\chi_I^-(x)\ \equiv \frac{d\, r_I^-}{d\,\Gamma_0}(x) = \frac{\lambda_I^2}{2\, J_{II} \sqrt{2\,\pi}\, \sigma_0^3}\ [\ (1 + \tfrac{\sigma_0^2}{\lambda_I^2} - \tfrac{x^2}{\sigma_0^2})\ e^{-\frac{x^2}{2\,\sigma_0^2}}\ +\ cosh(\tfrac{x}{\lambda_E})\ \frac{J_{IE}\, J_{EI}}{|J|}\ \frac{(\lambda_E^2 - \lambda_I^2)}{\lambda_E^2\, \lambda_I^2} (x_c\, \lambda_E - \sigma_0^2)\ e^{-\frac{x_c}{\lambda_E} - \frac{x_c^2}{2\,\sigma_0^2}}\,]$$
(24)

Depending on the parameters the inhibitory activity at the center of the perturbation can be paradoxical even when E is silenced. When x_c is sufficiently large $\chi_I^-(0)$ is always positive. Therefore, at the center, the inhibitory activity increases with the perturbation intensity. Conversely, when x_c is sufficiently small $\chi_I^-(0)$ is always negative. Therefore, the inhibitory activity at the center decreases with the perturbation intensity.

Figure 4A plots the phase diagram of the steady-state of the network in the (σ_0, Γ_0) plane. When $\Gamma_0 < \Gamma_0^*(\sigma_0)$, the network is balanced: excitatory and inhibitory neurons have a non-zero activity at any x. When $\Gamma_0 > \Gamma_0^*(\sigma_0)$, the network settles in a partially balanced state where excitation is silenced on a length x_c. For sufficiently large Γ_0, the inhibitory response in $[0, x_c]$ becomes zero and the network settles in a non-balanced state.

Figure 4B shows the spatial profile of the mean-field responses of the excitatory and inhibitory populations (normalized to their baseline) in the three regions. The radius of the perturbation, $4\,\sigma_0 = 0.4\ mm$, is smaller than the length of the recurrent excitatory interactions, $4\,\lambda_E = 0.5\ mm$ and than the length of the recurrent inhibitory interactions, $4\,\lambda_I = 0.3\ mm$.

Figure 4Ba shows the responses of the excitatory and inhibitory populations when the network is globally balanced and $\Gamma_0 = 0.005\ a.u.$ The excitatory and inhibitory responses exhibit a minimum at the center of the perturbation. Activities show a maximum in the surround. Far from the center of the perturbation, the network is at its baseline. Figure 4Bb shows the responses of the excitatory and inhibitory populations at the edge of the globally balanced region $(\Gamma_0 = \Gamma_0^* =\ .017\ a.u.)$. The profile of the activities is similar to the one in Fig. 4Ba except that at the center of the perturbation, excitation is zero.

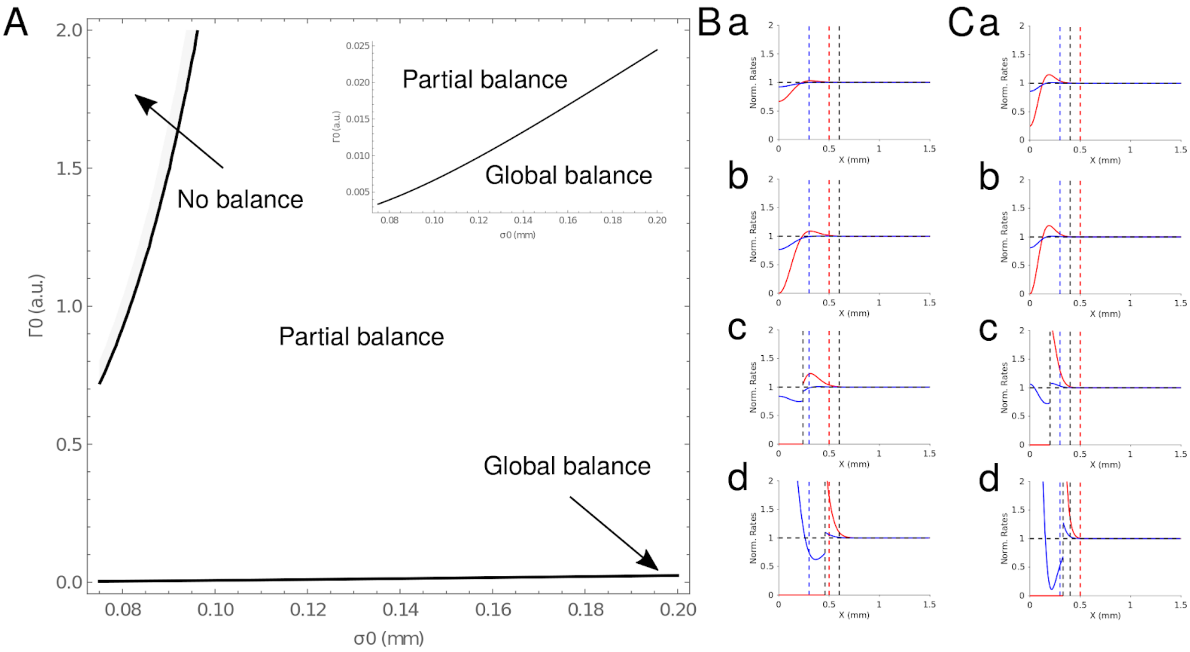


Figure 4. Gaussian stimulation of the inhibitory population in an EI network with exponential interactions (large K limit). A. Phase diagram of the steady-state of the network in the (σ_0, Γ_0) plane. B-C. Spatial profile of the activities (normalized to baseline). B. The radius of the perturbation is greater than the length of the recurrent excitatory interactions: $4\,\sigma_0 = 0.6\ mm > 4\,\lambda_E = 0.5\ mm > \ 4\,\lambda_I = 0.3\ mm$. C. The radius of the perturbation is smaller than the length of the recurrent excitatory interactions: $4\,\lambda_E = 0.5\ mm > 4\,\sigma_0 = 0.4\ mm >$ $4\,\lambda_I = 0.3\ mm$. a. Global balance, $\Gamma_0 = 0.1\ a.u.$. b. Edge of the global balance, $\Gamma_0 = 2\ a.u.$. c. Partial balance, $\Gamma_0 = 0.1\ a.u.$. d. Partial balance, $\Gamma_0 = 2\ a.u.$. Dashed-line: interaction length ($4\,\lambda_A$), Red: excitatory neurons; blue: inhibitory neurons; black: $4\,\sigma_0$; gray: suppression length, x_c. Parameters in Table 1.

Figure 4 Bc-d shows the responses of the excitatory and inhibitory populations when the network is partially balanced. In Fig. 4Bc, the intensity of the perturbation is rather small, namely, $\Gamma_0 = 0.1\ a.u.$, but sufficiently large so that the global balance is disrupted and excitation is partially silenced on a suppression length $x_c = 0.21\ mm$. In this region, the activity of the inhibitory population, $r_I^-(x)$, decays from a value slightly greater than its baseline at $x = 0$ to a minimum reached before x_c. In the region $[x_c; \frac{L}{2}]$, excitation and inhibition are balanced. The activities of the excitatory and inhibitory neurons, $r_E^+(x)$ and $r_I^+(x)$, decay exponentially from large non zero values to their baseline values. In Fig. 4Bd, the intensity of the perturbation is larger ($\Gamma_0 = 2\ a.u.$). Here, excitation is silenced in a region of greater length. However, the profile of activities is similar as in Fig. 4Bc: (1) in a region $[0; x_c]$, $r_I^-(x)$ exponentially decays from a non-zero value much greater than its baseline to a minimum below its baseline before x_c and then increases exponentially until x_c to a value below its baseline; (2) in $[x_c; \frac{L}{2}]$, both $r_E^+(x)$ and $r_I^+(x)$ decay exponentially from large non-zero values to their baseline values.

Remarkably, in Fig. 4Bc-d, the network responses on the left and right-hand side of x_c, are discontinuous.

Figure 4C shows the spatial profile of the mean-field responses of the excitatory and inhibitory populations (normalized to their baseline) in the three regions when the radius of the perturbation is greater than the length of the recurrent interactions, ($4\sigma_0 = 0.6\ mm > 4\lambda_E = 0.5\ mm > 4\lambda_I = 0.3\ mm$).

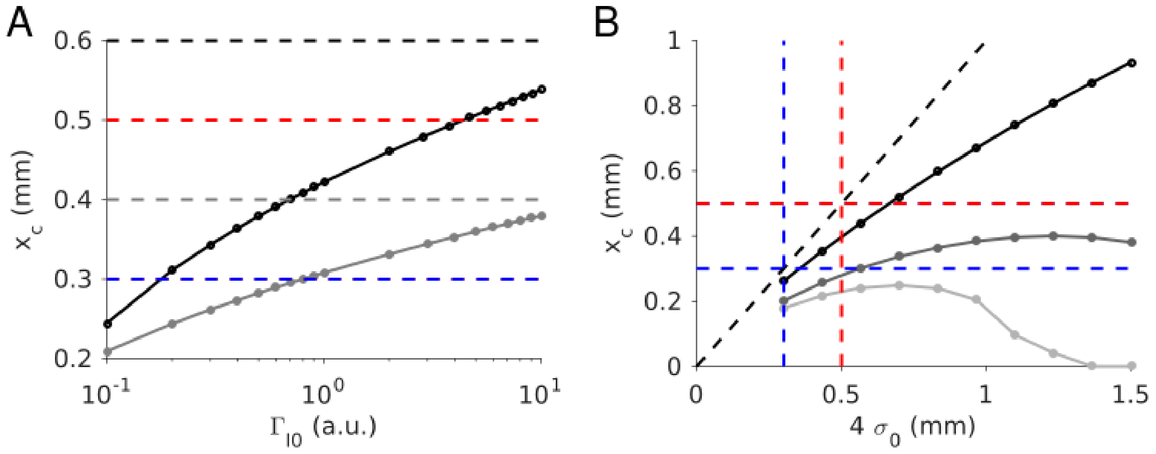


Figure 5. Suppression length, x_c, *v.s.* perturbation intensity and perturbation radius (large K limit). A. Suppression length, x_c *v.s.* photostimulation intensity, Γ_0. Black: $4\sigma_0 = 0.6\ mm$. Gray: $4\sigma_0 = 0.4\ mm$. B. Suppression length, x_c *v.s.* photostimulation radius, $4\sigma_0$. Black: $\Gamma_0 = 2\ a.u.$. Dark-gray: $\Gamma_0 = 0.2\ a.u.$. Gray: $\Gamma_0 = 0.1\ a.u.$. Dashed-line same as in Fig. 4. Parameters in Table 1.

Figure 5 depicts the effect of changing the perturbation intensity and radius on the length, x_c of the region where excitation is suppressed. Figure 5A plots the suppression length, x_c, *v.s.* the perturbation intensity Γ_0 for two fixed perturbation radius ($4\sigma_0 = 0.4\ mm$ and $4\sigma_0 = 0.6\ mm$). For both perturbation radius ($\sigma_0 > \lambda_E > \lambda_I$ and $\lambda_E > \sigma_0 > \lambda_I$), x_c increases monotonically with Γ_0 spanning continuously across the interaction length of the inhibitory and excitatory populations. Interestingly, for sufficiently large intensities, x_c slowly increases as the logarithm of the perturbation intensity. Figure 5B plots the suppression length, x_c *v.s.* the perturbation radius, $4\sigma_0$, for three fixed perturbation intensities ($\Gamma_0 = 0.1\ a.u.$, $\Gamma_0 = 0.2\ a.u.$ and $\Gamma_0 = 2\ a.u.$). When the perturbation intensity is large, x_c monotonically increases with σ_0 with a slope smaller than 1. Therefore, as σ_0 increases, x_c sweeps away from the value of the perturbation radius. Surprisingly, as Γ_0 becomes smaller, x_c becomes non-monotonic in σ_0. For sufficiently small Γ_0 and sufficiently large σ_0, x_c can even become zero, and the network is globally balanced.

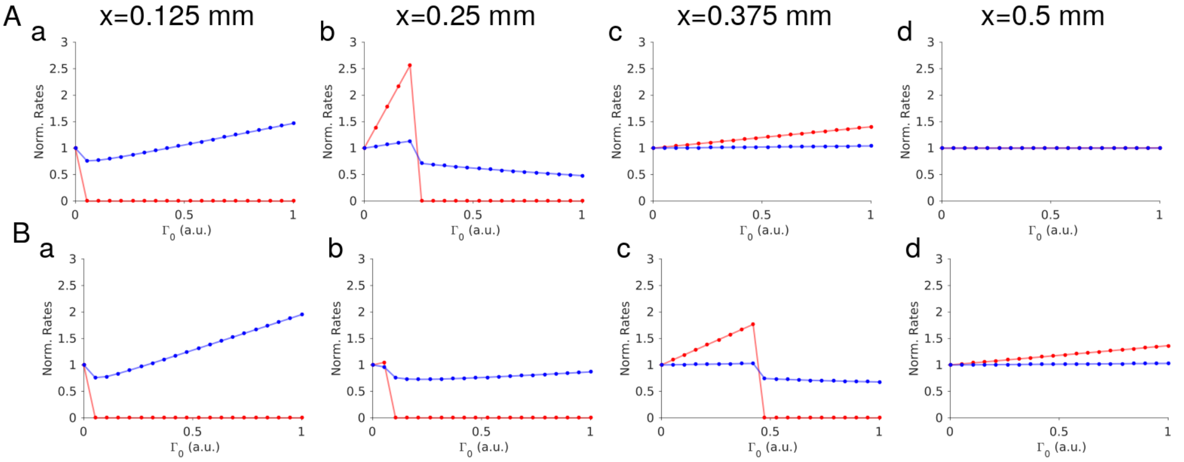


Figure 6. Responses of the neurons normalized to baseline *v.s.* the intensity of the perturbation, Γ_0 (large K limit). A. $4\,\sigma_0 = 0.4\,mm$, $x_E^{max} = 0.19\,mm$, $x_I^{max} = 0.22\,mm$. B. $4\,\sigma_0 = 0.6\,mm$, $x_E^{max} = 0.32\,mm$, $x_I^{max} = 0.40\,mm$. Parameters in Table 1.

Figure 6 plots the normalized activities *v.s.* the perturbation intensity for two different perturbation radius and at different positions on the ring. At the center and in its near neighbors, the responses of E and I are paradoxical for small Γ_0. When the excitatory neurons are silenced, the inhibitory response increases linearly with Γ_0. When x is sufficiently far from the center but smaller than x_I^{max}, for small perturbation intensities the inhibitory response decreases with Γ_0 while the excitatory response increases. When x is greater than x_I^{max} but smaller than x_E^{max}, for small perturbation intensities both the inhibitory and excitatory responses increase with Γ_0. Far from the center of the perturbation, the excitatory and inhibitory responses always increase with Γ_0.

The large N, K analysis provides precious insights into the understanding of the response of networks with reasonable size and connectivity to photostimulation of their inhibitory population. In particular, we will show that the properties exhibited in the large K limit, remains valid up to small corrections. Although it is possible to treat analytically the dependence of the activities on the perturbation for finite K, these calculations are very technical and beyond the scope of this paper.

Numerical simulations

We provide in the following numerical simulations of network models of leaky integrate and fire neurons to account for eventual discrepancies between the theory in the large K limit and the finite connectivity case. Unless mentioned otherwise, we will consider networks of $N_E = 15000$ excitatory and $N_I = 15000$ inhibitory cells interacting strongly with on average $K = 500$ neurons, and the length of the interactions is $4\,\lambda_E = 0.5\,mm$ and $4\,\lambda_I = 0.3\,mm$.

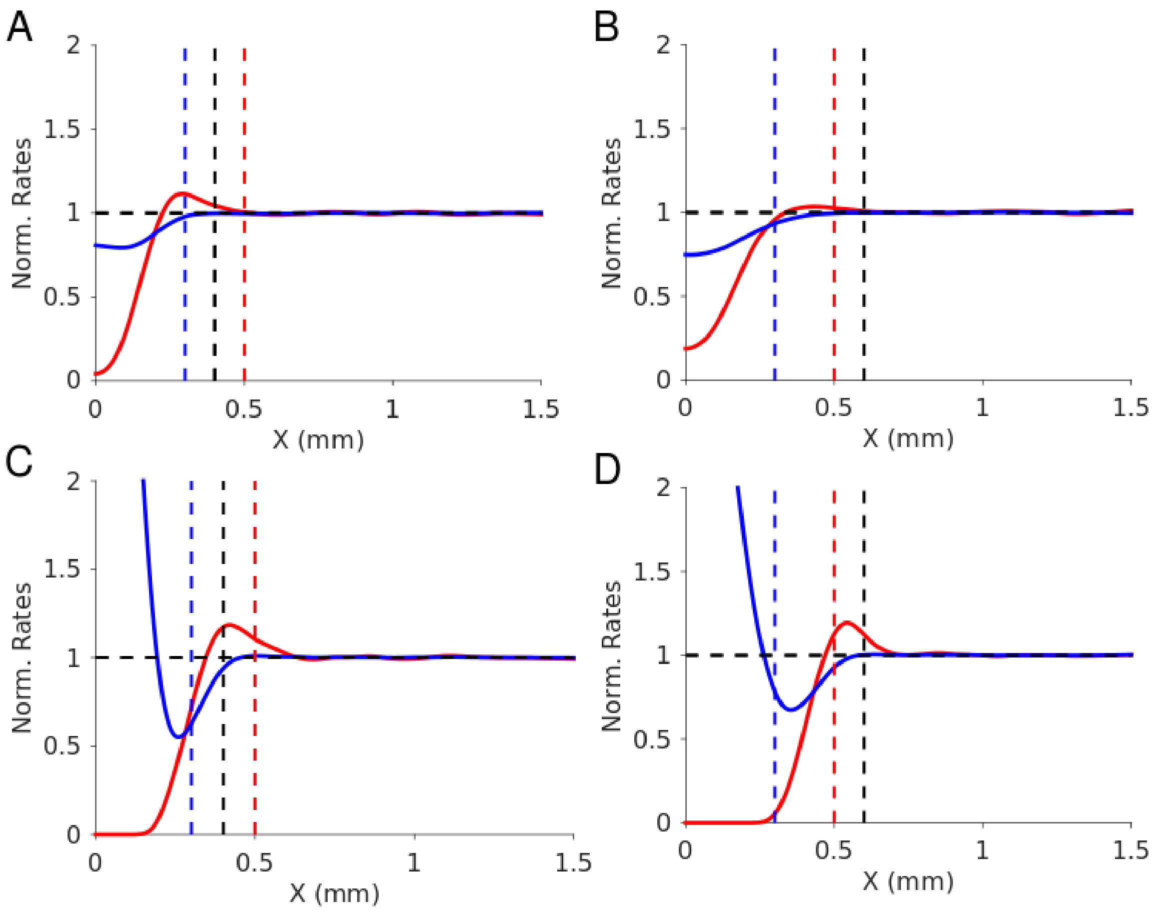


Figure 7. Gaussian photo-stimulation of the inhibitory population in EI network with exponential interactions. A. $\Gamma_0 = 0.1\ a.u.$. B. $\Gamma_0 = 2\ a.u.$. Left: the radius of the perturbation is smaller than the length of the recurrent excitatory interactions: $4\lambda_E = 0.5\ mm > 4\sigma_0 = 0.4\ mm > 4\lambda_I = 0.3\ mm$. Right: the radius of the perturbation, is greater than the length of the recurrent excitatory interactions: $4\sigma_0 = 0.6\ mm > 4\lambda_E = 0.5\ mm > 4\lambda_I = 0.3\ mm$. Dashed-line: interaction length ($4\lambda_A$), Red: excitatory neurons; blue: inhibitory neurons. black: $4\sigma_0$. Parameters in Table 1. $N_E = N_I = 15000$, $K = 500$.

Figure 7 depicts the results of our numerical simulations for the same parameters as in Fig. 4. Figure 7A plots the spatial profile of the population activities (normalized to baseline) for a small laser intensity ($\Gamma_0 = 0.1\ a.u.$) and a perturbation radius smaller than the range of the recurrent excitatory interactions ($4\lambda_E > 4\sigma_0 = 0.4\ mm > 4\lambda_I$). In Fig. 7B, $\sigma_0 > \lambda_E > \lambda_I$ and the laser intensity is the same. In contrast to the predictions of the theory in the large K limit, here, when $\Gamma_0 = 0.1\ a.u.$, the excitatory population is always active. Moreover, the inhibitory response is always below its baseline value even at the center of the laser. However, $r_E(x)$ is non-monotonic: it is highly suppressed near the laser center and increases exponentially to a maximum above its baseline value in the surround before exponentially decreasing to its baseline. Conversely, the response of the inhibitory population is always paradoxical, it is slightly non-monotonic when $\lambda_E > \sigma_0 > \lambda_I$ near the laser center where the activity at zero is greater than in its neighboring surround and then recovers to its baseline. In particular, when $\sigma_0 > \lambda_E > \lambda_I$, $r_I(x)$ is minimum at the center of the stimulation and monotonically increases to its baseline value in the surround.

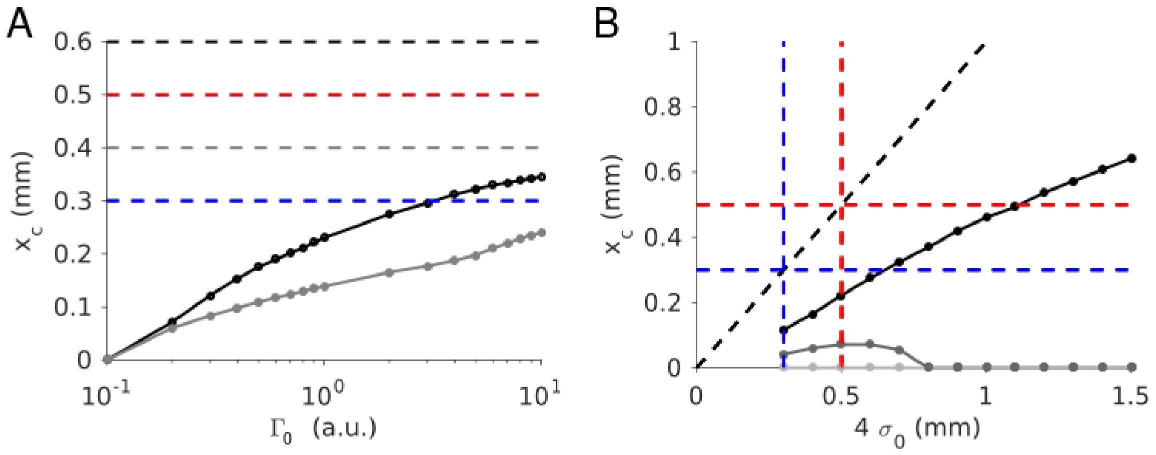


Figure 8. Photo-suppression length *v.s.* laser intensity and laser radius. A. Suppression length, x_c *v.s.* photostimulation intensity, Γ_0. Black: $4\,\sigma_0 = 0.6\,mm$. Gray: $4\,\sigma_0 = 0.4\,mm$. B. Suppression length, x_c *v.s.* photostimulation radius, $4\,\sigma_0$. Black: $\Gamma_0 = 2\,a.u.$. Dark-gray: $\Gamma_0 = 0.2\,a.u.$. Gray: $\Gamma_0 = 0.1\,a.u.$. Dashed-line same as in Fig. 7. Parameters in Table 1.

Figure 7C-D shows similar network responses but for a larger perturbation intensity, namely, $\Gamma_0 = 2a.u.$. Here, Γ_0 is sufficiently large so that excitatory neurons are silenced in a given region near the center of the perturbation. In this region inhibitory neurons are highly activated at the laser center, then the inhibitory activity decreases exponentially to a minimum value below its baseline but at a value of x for which the activity of the excitatory population is non-zero. Finally, $r_I(x)$ recovers exponentially to its baseline value. Conversely, the excitatory neurons are silenced in a small region near the laser center, then the excitatory activity exponentially increases to a maximum value above its baseline and finally, $r_E(x)$ exponentially decreases to its baseline value.

The discrepancies exhibited by the simulations with regard to the theory in the large K limit are probably due to strong size effects as suggested by simulations of larger networks (see SM Fig. 1).

Figure 8 depicts the effect of changing the photo-stimulation intensity and radius on the length, x_c of the region where excitation is suppressed. Figure 8A plots the suppression length, x_c, *v.s.* the photostimulation intensity Γ_0 for two fixed photostimulation radius ($4\,\sigma_0 = 0.4\,mm$ and $4\,\sigma_0 = 0.6\,mm$). Consistently with the theory in the large K limit, for both perturbation radius ($\sigma_0 > \lambda_E > \lambda_I$ and $\lambda_E > \sigma_0 > \lambda_I$), x_c increases monotonically with Γ_0. Moreover, for sufficiently large intensities, x_c slowly increases as the logarithm of the perturbation intensity, as predicted by the theory. Figure 8B plots the suppression length, x_c *v.s.* the photostimulation radius,

$4\sigma_0$, for three fixed photo-stimulation intensities ($\Gamma_0 = 0.1\ a.u.$, $\Gamma_0 = 0.2\ a.u.$ and $\Gamma_0 = 2\ a.u.$). Here again, the simulations are in agreement with the prediction of the theory. When the photostimulation intensity is large, x_c monotonically increases with σ_0 with a slope smaller than 1 and for small Γ_0, x_c becomes non-monotonically in σ_0. For sufficiently small Γ_0, x_c can even always be zero.

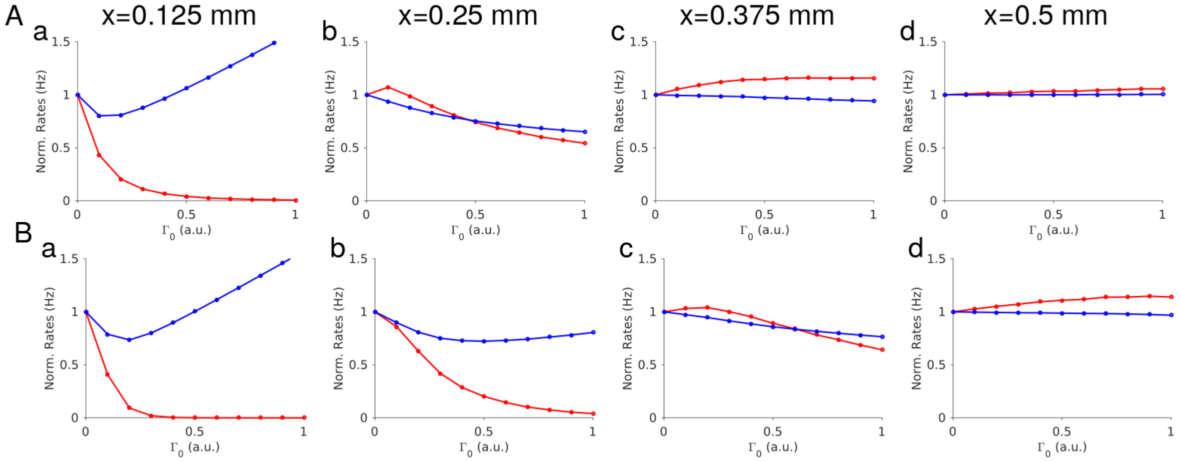


Figure 9. Responses of the neurons normalized to baseline *v.s.* the intensity of the perturbation, Γ_0 . A. $4\sigma_0 = 0.4\ mm$. B. $4\sigma_0 = 0.6\ mm$. Parameters in Table 1.

Figure 9 plots the normalized responses versus the laser intensity at different positions from the laser center. The averaged responses of E and I neurons are in agreement with the ones in the large K limit.

For now, our model accounts for only parts of the experimental data presented by (Li et al., 2019). In particular, there is no sign in the recordings of a rebound of the excitatory response in the region surrounding the suppression. We postulate that the rebound of activity could be due to the assumption in the model that the interaction lengths depend only on the presynaptic population, namely, $\lambda_{AB} = \lambda_A$. Unfortunately, when this is not the case, network models with exponentially decaying interactions can no longer be reduced to simple differential equations. Instead, one is left with a set of coupled Fredholm equations that have no explicit solutions. However, it is possible to get some insights about the role of which specific interaction lengths shape the responses of the network in the limit when $K \to \infty$.

Partially balanced solutions when $\lambda_{BA} \neq \lambda_{CA}$

We want to investigate the response of the network to a strong perturbation of its inhibitory population when the length of an interaction depends on both the pre and postsynaptic populations ($\lambda_{BA} \neq \lambda_{CA}$). Similarly to the previous case, we look for solutions to the dynamics of the network where the excitatory population is silenced within a given region $[-x_c; x_c]$ and we only investigate the network's responses in

$[0;\frac{L}{2}]$. The solutions of the dynamics in $[0;x_c]$ are labeled $\{-\}$ and the ones in $[x_c;\frac{L}{2}]$ are labeled $\{+\}$.

The net input into neuron x in population A can be expressed in terms of the activities on the left and right side of x_c as

$$u_A(x) = \sqrt{K}\,\{\, I_A(x) + J_{A0}\, r_0 + \sum_B \frac{J_{AB}}{\lambda_{AB}}\, \varepsilon_B [\int_{-\infty}^{+\infty} dy\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^+(y) - \int_{-\frac{x_c}{2}}^{+\frac{x_c}{2}} dy\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^+(y) + \int_{-\frac{x_c}{2}}^{+\frac{x_c}{2}} dy\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^-(y)]\} \qquad (25)$$

Network activities in $[x_c,\frac{L}{2}]$ in the large N, K limit

The balance of the excitatory and inhibitory inputs into the neurons in the region $[x_c,\frac{L}{2}]$ implies that the right-hand side of Eq. [25] must remain finite as K becomes large. In the limit where K goes to infinity, it implies that

$$I_A(x) + J_{A0}\, r_0 + \sum_B \frac{J_{AB}}{\lambda_{AB}}\, \varepsilon_B [\int_{-\infty}^{+\infty} dy\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^+(y) - \int_{-\infty}^{+\infty} dy\, \Pi_{x_c}(y)\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^+(y) + \int_{-\infty}^{+\infty} dy\, \Pi_{x_c}(y)\, e^{-\frac{|x-y|}{\lambda_{AB}}}\, r_B^-(y)] = 0 \qquad (26)$$

where $\Pi_{x_c}(x) = \Theta(x+x_c) - \Theta(x-x_c)$ and Θ is the Heaviside function.

Let us rewrite Eq. [18] in the Fourier domain:

$$\hat{I}_A(k) + J_{A0}\, r_0\, \delta(k) + \sum_B \hat{G}_{AB}(k)\, [\hat{r}_B^+(k) - \hat{H}_B^{+}(k) + \hat{H}_B^{-}(k)] = 0 \qquad (27)$$

where $\hat{H}_B^{\pm}(k) = \pm\, FT[\Pi_{x_c}] \circledast FT[r^{\pm}]$ and $\hat{G}_{AB}(k) = \frac{J_{AB}\, \varepsilon_B}{1 + k^2 \lambda_{AB}^{\;2}}$.

Because of the convolution integrals in Eq. [26] all include the interaction kernel, $\frac{1}{\lambda_{AB}} e^{-\frac{|x-y|}{\lambda_{AB}}}$, Eq. [27] is linear in the terms that include the activities. Provided that G is invertible one can write

$$\hat{r}^+(k) + \hat{H}^+(k) + \hat{H}^-(k) =- \hat{G}^{-1}(k)\,.\,(\hat{I}(k) + I_0\, \delta(k)\,) \qquad (28)$$

The left-hand side of Eq. [28] can be easily Fourier inverse noticing that

$$FT^{-1}[FT[\Pi_{x_c}] \circledast FT[r^{\pm}]](x) = \Pi_{x_c}(x)\, r^{\pm}(x) = 0 \;\mathit{since}\; x \geq x_c \qquad (29)$$

Therefore, provided that the right-hand side of Eq. [28] has a defined inverse Fourier transform the activity in the region $[x_c;\frac{L}{2}]$ is

$$r^{+}(x) = -FT^{-1}[\hat{G}^{-1} . (\hat{I} + I_0)](x) \tag{30}$$

Equation [30] can be analytically solved in general, nevertheless its derivation is very technical and will be published somewhere else. To get an understanding of which interaction is essential to shape the excitatory response we will only consider the case where $J_{EE} = 0$ and compare it to numerical simulations with $J_{EE} \neq 0$.

When $J_{EE} = 0$, $FT^{-1}[\hat{G}^{-1} . \hat{I}]$ is well defined and the activities of each population are

$$r_E^{+}(x) = r_E^{0} - \Gamma_0 \frac{\lambda_{IE}^2 \sigma_0^2 + \sigma_0^4 - x^2 \lambda_{IE}^2}{2 J_{IE} \sigma_0^5} e^{-\frac{x^2}{2\sigma_0^2}} \tag{31}$$

$$r_I^{+}(x) = r_I^{0} \tag{32}$$

where $r_E^0 = \frac{J_{E0} J_{II} - J_{I0} J_{EI}}{2 J_{EI} J_{IE}} r_0$ and $r_I^0 = \frac{J_{E0}}{2 J_{EI}} r_0$ are the baseline activities.

Remarkably, $r_E^{+}(x)$ is a non-monotonic function of x that always exhibits a maximum at $x_* = \frac{\sigma_0 \sqrt{\sigma_0^2 + 3 \lambda_{IE}^2}}{\lambda_{IE}}$. Remarkably, x_* is determined by σ_0 and λ_{IE} only and is independent of the rest of the interaction length (λ_{EI} and λ_{II}).

Moreover, at x_* the normalized excitatory activity is equal to

$$F(\lambda_{IE}) \equiv \frac{r_E^{+}(x_*)}{r_E^{0}} = (1 + \frac{2 e^{-3/2} e^{-{\sigma_0}^2/2 {\lambda_{IE}}^2} \Gamma_0 {\lambda_{IE}}^2}{{\sigma_0}^3 (J_{E0} J_{II} - J_{I0} J_{EI}) r_0}) \tag{33}$$

One easily shows that the function F monotonically increases with λ_{IE}. In other words, the maximum value of the normalized activity of E is minimized for small λ_{IE}. Note that F does not depend on λ_{EI} and λ_{II}.

Inhibitory activity in $[x_c, \frac{L}{2}]$ in the large N, K limit

For the neurons in $[0; x_c]$, we are only interested in the response of the inhibitory population, since we assumed $r_E^{-}(x) = 0, \forall x \in [0; x_c]$.

One easily shows that the activity of the inhibitory neuron x in the region $[0; x_c]$ is

$$r_I^{-}(x) = \frac{\lambda_{II}^2}{2 J_{II}} (\frac{I_I(x) + J_{I0} r_0}{\lambda_{II}^2} - I_I''(x) + 2 \frac{J_{IE}}{\lambda_{IE}} (\frac{1}{\lambda_{II}^2} - \frac{1}{\lambda_{IE}^2}) \cosh(\frac{x}{\lambda_{IE}}) \int_{x_c}^{+\infty} dy\, e^{-\frac{y}{\lambda_{IE}}} r_E^{+}(y)) \tag{34}$$

Suppression length x_c in the large N, K limit

To completely figure out the network response, we need to establish the value of x that separates these two solutions. x_c has to be determined self-consistently given its definition: it is the minimum value of x for which $r_I^{-}(x) > 0,\ r_E^{-}(x) = 0$ no longer stands, implying that either $u_E^{-}(x_c^{-}) = 0$ and $u_E^{-}(x_c^{+}) > 0$. This leads to the determination of the roots of the transcendental equation that determines the net input into E:

$$J_{E0}\, r_0 - 2\, \frac{J_{EI}}{\lambda_{EI}} (cosh(\frac{x_c}{\lambda_{EI}}) [\int_{x_c}^{+\infty} dy\, e^{-\frac{y}{\lambda_{EI}}}\, r_I^+(y)] + e^{-\frac{x_c}{\lambda_{EI}}} \int_0^{x_c} dy\, cosh(\frac{y}{\lambda_{EI}})\, r_I^-(y)) = 0 \quad (35)$$

The evolution of the suppression length, x_c, with the interaction lengths, λ_{AB}, cannot be derived analytically. However, we show in SM Fig. S2 that x_c is an increasing function of λ_{IE} and λ_{EI}.

The excitatory activity in the region surrounding the suppression will present a rebound if x_* is sufficiently large in comparison with x_c. However, this rebound of activity is small when λ_{IE} is small and its position is independent of the inhibitory interaction lengths. Moreover, x_c is an increasing function of λ_{EI}, as a consequence for sufficiently small λ_{IE} and sufficiently large λ_{EI}, the excitatory response will not present any rebound of activity.

We will show that the properties exhibited here remain valid in networks with finite size. Then, we will investigate how the previous results change in numerical simulations of networks with strong recurrent excitatory connections.

Numerical simulations

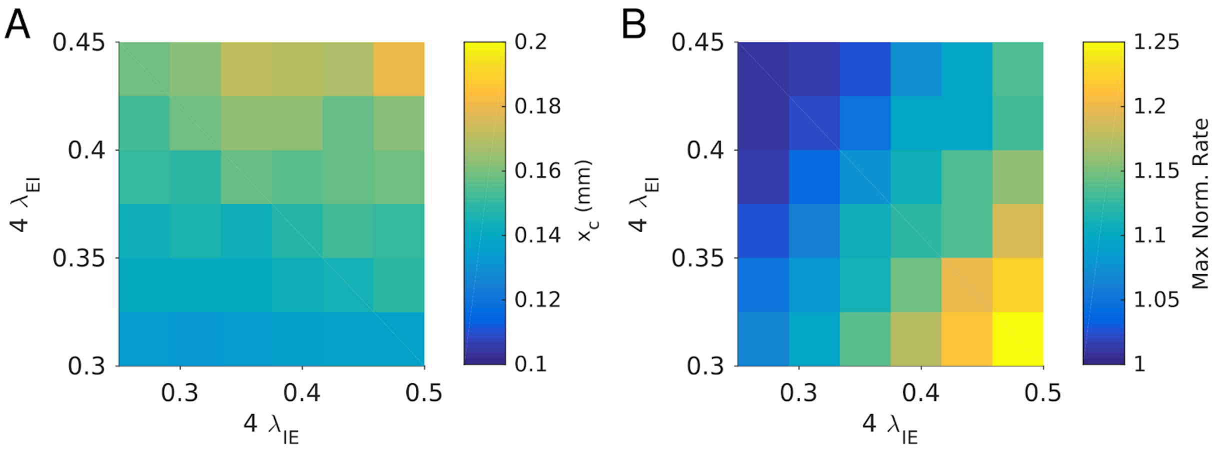


Figure 10. Effect of the length of the E to I and I to E recurrent interactions on the photo-suppression length and the maximum of the normalized E response. $J_{EE} = 0$, $\Gamma_0 = 1\, a.u.$, $4\,\sigma_0 = 0.4\, mm$. Other parameters in Table 1. $N_E = N_I = 15000$, $K = 500$.

Figure 10 depicts the results of our numerical simulations for a network without strong recurrent interaction ($J_{EE} = 0$). Figure 10A shows the evolution of the maximum of the normalized excitatory activity, $\bar{r}_E$, for different E to I and I to E recurrent interaction lengths. Consistent with the large N, K theory, for a given λ_{EI}, the maximum value of $\bar{r}_E$ decreases as λ_{IE} becomes smaller. For sufficiently large λ_{IE} and sufficiently small λ_{IE}, this maximum is equal to 1 and the excitatory activity exhibit no rebound activity. Figure 8B shows how the change in λ_{IE} and λ_{EI} affect the suppression length, x_c. As predicted in the large N, K limit, at given λ_{EI}, x_c decreases as λ_{IE} becomes smaller. Conversely, for fixed λ_{IE}, x_c increases as λ_{EI} becomes larger.

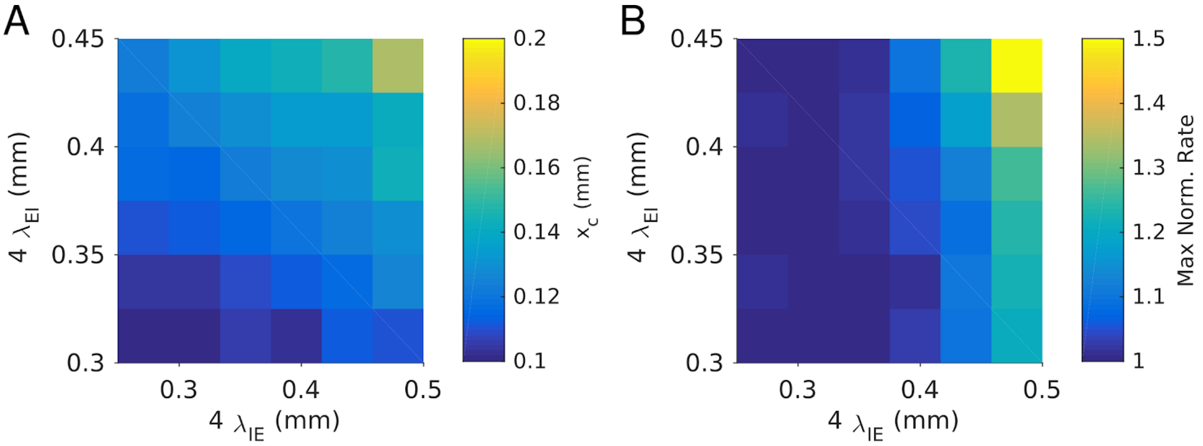


Figure 11. Effect of the length of the E to I and I to E recurrent interactions on the photo-suppression length and the maximum of the normalized E response. $J_{EE} \neq 0$. $\Gamma_0 = 1\ a.u.$, $4\ \sigma_0 = 0.4\ mm$. Other parameters in Table 1. $N_E = N_I = 15000$, $K = 500$.

Figure 11 depicts the results of our numerical simulations for a network with strong recurrent interaction ($J_{EE} = O(\frac{1}{\sqrt{K}})$). Figure 11A depicts the evolution of the maximum of the normalized excitatory activity, $\bar{r}_E$, for different E to I and I to E recurrent interaction lengths. For a given λ_{EI}, the maximum value of $\bar{r}_E$ decreases as λ_{IE} becomes smaller. For sufficiently small λ_{IE}, this maximum becomes smaller or equal to 1, therefore the excitatory activity exhibit no rebound activity. Remarkably and in contrast with Fig. 10A, when λ_{IE} is fixed, the maximum value of $\bar{r}_E$ increases as λ_{EI} becomes larger. Figure 11B shows how the change in λ_{IE} and λ_{EI} affect the suppression length, x_c. Similarly to Fig. 10B, for a given λ_{EI}, x_c decreases as λ_{IE} becomes smaller. Conversely, for fixed λ_{IE}, x_c increases as λ_{EI} becomes larger.

Figure 12 depicts the result of the simulations of the photo-stimulation of the inhibitory population in a network where $\lambda_{EE} > \lambda_{EI} > \lambda_{II} > \lambda_{IE}$. Figure 10A plots the spatial profile of the population activities (normalized to baseline) for a small laser intensity ($\Gamma_0 = 0.1\ a.u.$) and a perturbation radius smaller than the range of the recurrent excitatory interactions ($\lambda_{EE} > \sigma_0 = 0.4\ mm$) but greater than the length of the other interactions. In Fig. 10B, $\sigma_0 > \lambda_{EE}$ and the laser intensity is the same. The excitatory population is always active, however, in contrast to the responses observed Fig. 6, the excitatory response is always below its baseline value even in the surround of the center of the laser: there is no rebound of the excitatory activity in the surrounding region. $r_E(x)$ is monotonic: it is highly suppressed near the laser center and increases exponentially to its baseline value in the surround. The response of the inhibitory population is always paradoxical and very similar to the one in Fig. 6A-B: it is slightly non-monotonic when $\lambda_{EE} > \sigma_0$ near the laser center where the activity at zero is greater than in its neighboring surround and then recovers to its baseline. When $\sigma_0 > \lambda_{EE}$, $r_I(x)$ is minimum at the center of the stimulation and monotonically increases to its baseline value in the surround. Figure 12C-D shows similar network responses but for a larger photo-stimulation intensity, namely, $\Gamma_0 = 2\ a.u.$.

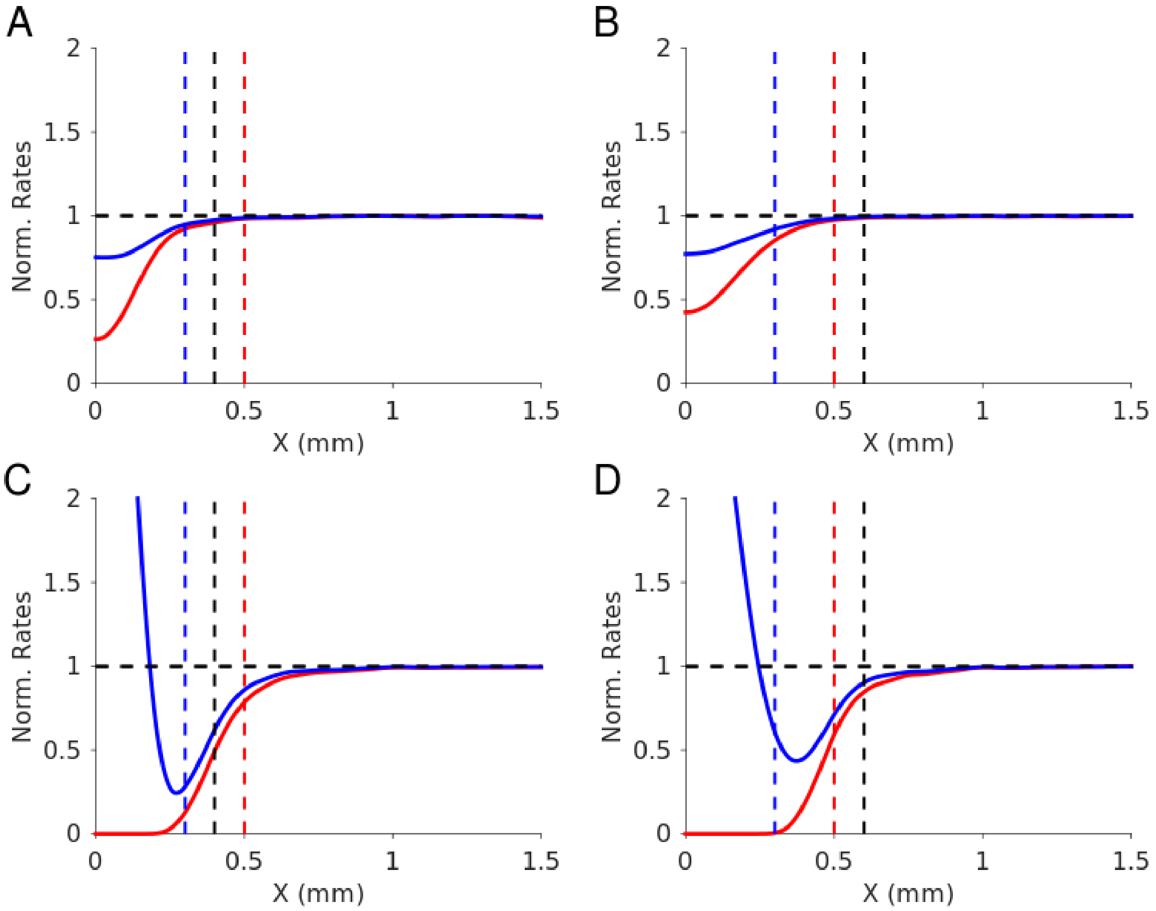


Figure 12. Gaussian photo-activation with Exponential Interactions when $\lambda_{EE} > \lambda_{EI} > \lambda_{II} > \lambda_{IE}$. A. $\Gamma_0 = 0.1\ a.u.$. B. $\Gamma_0 = 2\ a.u.$. Left: the radius of the perturbation is smaller than the length of the recurrent excitatory interactions: $4\lambda_{EE} = 0.5\ mm > 4\sigma_0 = 0.4\ mm$. Right: the radius of the perturbation, is greater than the length of the recurrent excitatory interactions: $4\sigma_0 = 0.6\ mm > 4\lambda_{EE} = 0.5\ mm$. Dashed-line: interaction length ($4\lambda_A$), Red: excitatory neurons; blue: inhibitory neurons. black: $4\sigma_0$. Parameters in Table 1. $N_E = N_I = 15000$, $K = 500$. $\lambda_{EE} = 0.125\ mm$, $\lambda_{EI} = 0.125 mm$, $\lambda_{II} = 0.075\ mm$, $\lambda_{IE} = 0.0625\ mm$.

Here, Γ_0 is sufficiently large so that excitatory neurons are silenced in a given region near the center of the perturbation. In this region inhibitory neurons are highly activated at the laser center, then the inhibitory activity decreases exponentially to a minimum value below its baseline but at a value of x for which the activity of the excitatory population is non-zero. Finally, $r_I(x)$ recovers exponentially to its baseline value. Conversely, the excitatory neurons are silenced in a small region near the laser center, then the excitatory activity exponentially increases to its baseline value. In particular, the responses of E and I are almost identical in the region from the minimum of I and the recovery to their baseline. Remarkably, the lengths of the region where E is suppressed are similar to the one in Fig. 7.

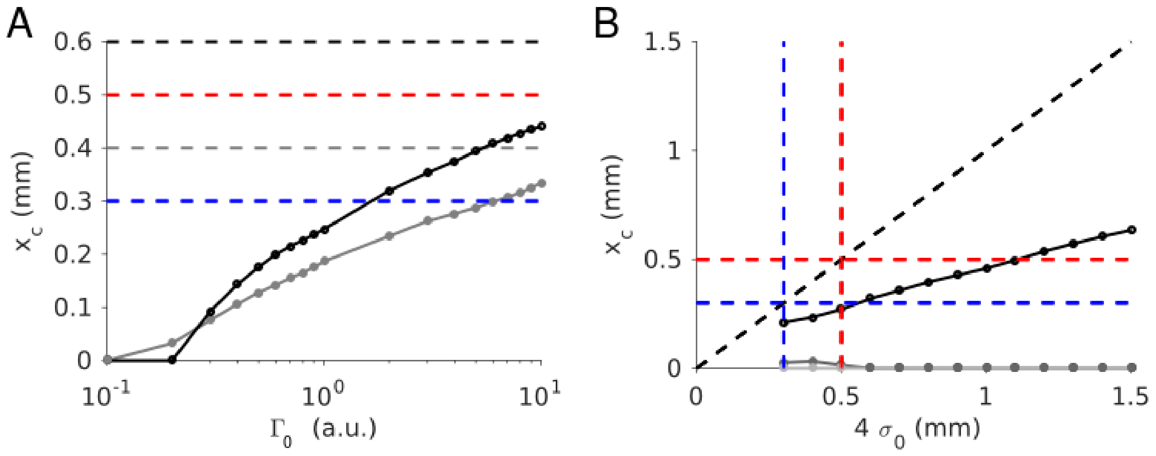


Figure 13. Photo-suppression length v.s. laser intensity and laser radius when $\lambda_{EE} > \lambda_{EI} > \lambda_{II} > \lambda_{IE}$. A. Suppression length, x_c *v.s.* photostimulation intensity, Γ_0. Black: $4\,\sigma_0 = 0.6\ mm$. Gray: $4\,\sigma_0 = 0.4\ mm$. B. Suppression length, x_c *v.s.* photostimulation radius, $4\,\sigma_0$. Black: $\Gamma_0 = 2\ a.u.$. Dark-gray: $\Gamma_0 = 0.2\ a.u.$. Gray: $\Gamma_0 = 0.1\ a.u.$. Dashed-line same as in Fig. 12. Parameters as in Fig. 6. Parameters as in Fig. 10. $N_E = N_I = 15000$, $K = 500$.

Figure 13 depicts the effect of the photo-stimulation intensity and radius on the length, x_c of the region where excitation is suppressed. Figure 13A plots the suppression length, x_c, *v.s.* the photo-stimulation intensity Γ_0 for two fixed photo-stimulation radius ($4\,\sigma_0 = 0.4mm$ and $4\,\sigma_0 = 0.6mm$). For both perturbation radius ($\sigma_0 > \lambda_{EE}$ and $\lambda_{EE} > \sigma_0$), x_c increases monotonically with Γ_0 spanning continuously across the interaction length of the inhibitory and excitatory populations. Moreover, for sufficiently large intensities, x_c slowly increases as the logarithm of the perturbation intensity. Figure 13B plots the suppression length, x_c *v.s.* the photostimulation radius, $4\,\sigma_0$, for three fixed photostimulation intensities ($\Gamma_0 = 0.1\ a.u.$, $\Gamma_0 = 0.2\ a.u.$ and $\Gamma_0 = 2\ a.u.$). When the photostimulation intensity is large, x_c linearly increases with σ_0 with a slope smaller than 1. As σ_0 increases, x_c becomes more and smaller than the photo-stimulation radius. When Γ_0 becomes smaller, x_c becomes non-monotonic in σ_0. For sufficiently small Γ_0 and sufficiently large σ_0, x_c becomes zero. These results are qualitatively similar to that one of Fig. 7. However, here, the region of (Γ_0, σ_0) such that the region of global balance is bigger than in Fig. 7.

Discussion

We provided a mechanistic account for the spatial profile of cortex to the optogenetic stimulation of parvalbumin-positive (PV) neurons. Photostimulation enhanced the PV activity and suppressed the response of the principal cells (PCs) in a region of the order of the laser radius. In the surround, excitatory and inhibitory activity was suppressed in a relatively proportional manner independently of the light intensity (Li et al., n.d.) 2019). To account for these results, we investigated the dynamics of networks of one excitatory and one inhibitory population where the probability of interaction decays exponentially with the distance between the neurons. We showed that when a perturbation is strong enough, the balance of excitation and inhibition is disrupted on a finite length, x_c, where the excitatory population is silenced. Conversely, the balance of excitation and inhibition is preserved in the surrounding regions. We showed that for large perturbation intensities the length of the suppression, x_c, is linear with the radius of the perturbation. This is not the case for small perturbation intensities where x_c is non-linear with the radius of the perturbation: it increases with small perturbation sizes and then decreases for sufficiently large radiuses.

In network models where the length of the interactions only depends on the presynaptic population (homogeneous interaction lengths, $\lambda_{AB} = \lambda_A$), we showed that the excitatory population exhibits a large rebound of activity in the surround of the suppression. Therefore, these models cannot fully account for the experimental data. This prompted us to investigate the dynamics of networks where the length of the interactions depends on both their pre and postsynaptic targets (heterogeneous interaction lengths, $\lambda_{AB} \neq \lambda_{AC}$). These models can account for the relative similar suppression of the excitatory and inhibitory activity in the surround of the perturbation provided that the excitatory to inhibitory interaction length λ_{IE} is small in comparison with the radius of the perturbation. Moreover, we showed that then, to counter the consequent reduction in the length of suppression of the PC activity, inhibitory to excitatory interactions must be broader.

Limitations

In our models, we did not take into account the diversity of inhibitory neurons in the cortex. For example, somatostatin expressing interneurons have been shown to exhibit interactions on a length similar to the PC to PC connection. The study of spatially extended network models with multiple inhibitory populations will precise the mechanisms that underlie cortical response to locally restricted optogenetic manipulations and provide insights on the relative length of cortical neuron interactions.

Interlaminar interactions are likely to also contribute to the spatial effects of optogenetic manipulations. For example, in cortex thalamocortical projections are thought to be broader than intracortical interactions. Future modeling work should investigate to what extent recurrent interactions and/or feedforward projections shape the spatial response of cortical neurons.

Comparison with previous works

Previous studies have investigated how spatially extended inhibition could sharpen tuning curves (Somers et al., 1995) or promote pattern formation (Coombes, 2005; Kilpatrick and Ermentrout, 2013). Nevertheless, these works consider neural fields with weak interactions. We show here that broad inhibitory to excitatory interactions are compatible with the balanced state.

(Rosenbaum and Doiron, 2014) have studied the conditions upon which a balanced state exists in strongly recurrent networks of neurons with periodic Gaussian probability of connections. They have shown in the case where interaction lengths are homogeneous ($\sigma_{AB} = \sigma_B$, where σ_{AB} is the interaction length of the connection from population B to population A) that balanced states required external inputs broader than recurrent excitatory connections which in turn must be broader than the inhibitory ones ($\sigma_0 > \sigma_E > \sigma_I$). We show here that when interactions have an exponential profile balanced states can exist even when inhibitory to excitatory connections are broader than the excitatory to inhibitory connections ($\lambda_{EI} > \lambda_{IE}$). This is also the case with Gaussian interactions. We give more detail about the necessary conditions upon which it occurs in the Appendix.

In a recent study, (Ebsch and Rosenbaum, 2018) have investigated the mechanism of amplification and suppression from a local imbalance of excitation and inhibition in recurrent circuits of strongly interacting neurons with periodic Gaussian interactions and homogeneous interactions. They model the effect of visual stimulation in a model where a layer of excitatory neurons representing layer 4 of the mouse primary visual cortex sends feedforward projections to an EI network representing its layer 2/3. When the length of the feedforward interaction was small, excitatory neurons exhibited a response profile similar to a 2D Mexican hat with a sharp peak at the center, a minimum response in the surround and recovery to baseline far from the center. They showed that to observe a non-linear response with stimulus size, the length of the interactions from a population of neurons must be smaller than the stimulus radius. We show here a more complete picture that accounts for heterogeneous interactions in the connectivity.

Perspectives

Li et al. only considered the response of the neurons for a large radius of the laser beam. Investigating the dependence of the spatial response of PCs and PV neurons on the laser radius will provide further constraints on the length of the interactions in the cortex. In particular, future work should investigate the effect of photostimulation of the cortex on a scale smaller or equal to the length of the inhibitory to inhibitory interactions. In particular, on how the balance state is then disrupted and spatial correlations build up in the network responses.

Parameters

Table 1. Connection strength matrix (rows: postsynaptic populations; columns: presynaptic populations)

$J_{\alpha\beta}$ $(\mu A\ .\ ms\ .\ cm^{-2})$	Feedforward	PC	PV
PC	34	29	30
PV	34	36	36

Table 2. Synaptic time constants

$\tau_{\alpha\beta}(ms)$	E	I
E	3	2
I	3	2

Supplementary Materials

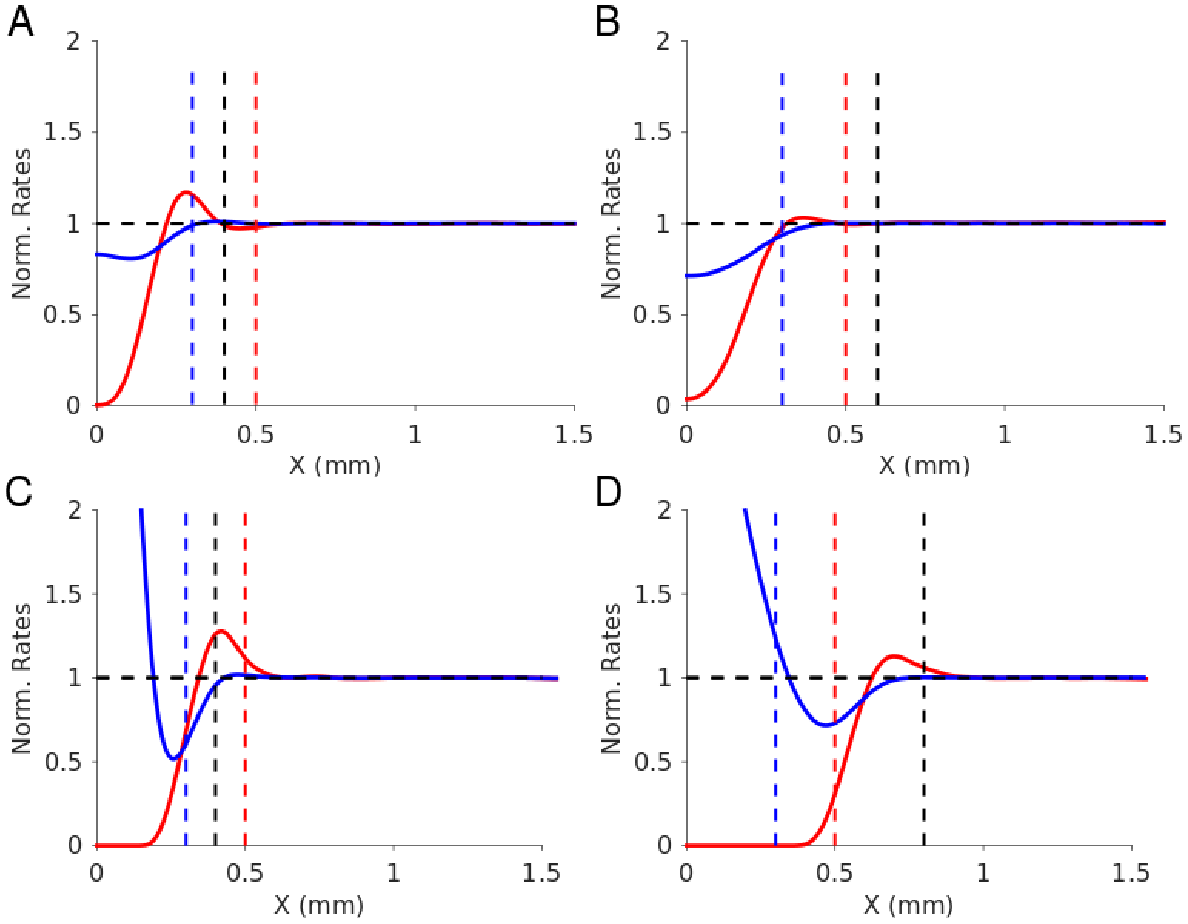


Figure S1. Gaussian photo-activation with Exponential Interactions. Interaction lengths are homogeneous. $N = 120000$, $K = 2000$. Parameters as in Fig. 7.

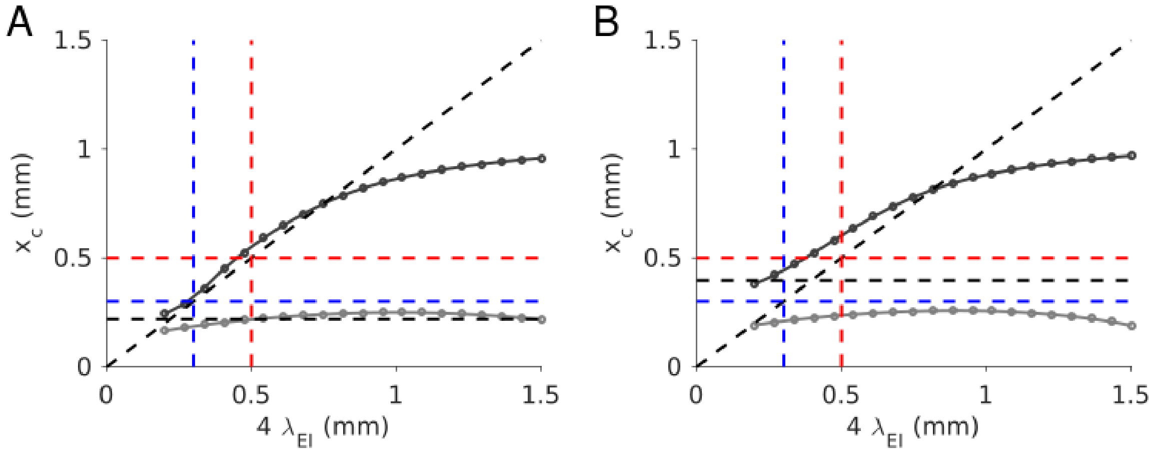


Figure S2. Suppression length v.s. λ_{EI} in the large K limit ($J_{EE} = 0$) for two different values of Γ_0. A. $4\sigma_0 = 0.4mm$. B. $4\sigma_0 = 0.6\ mm$. Parameters as in Fig. 8.

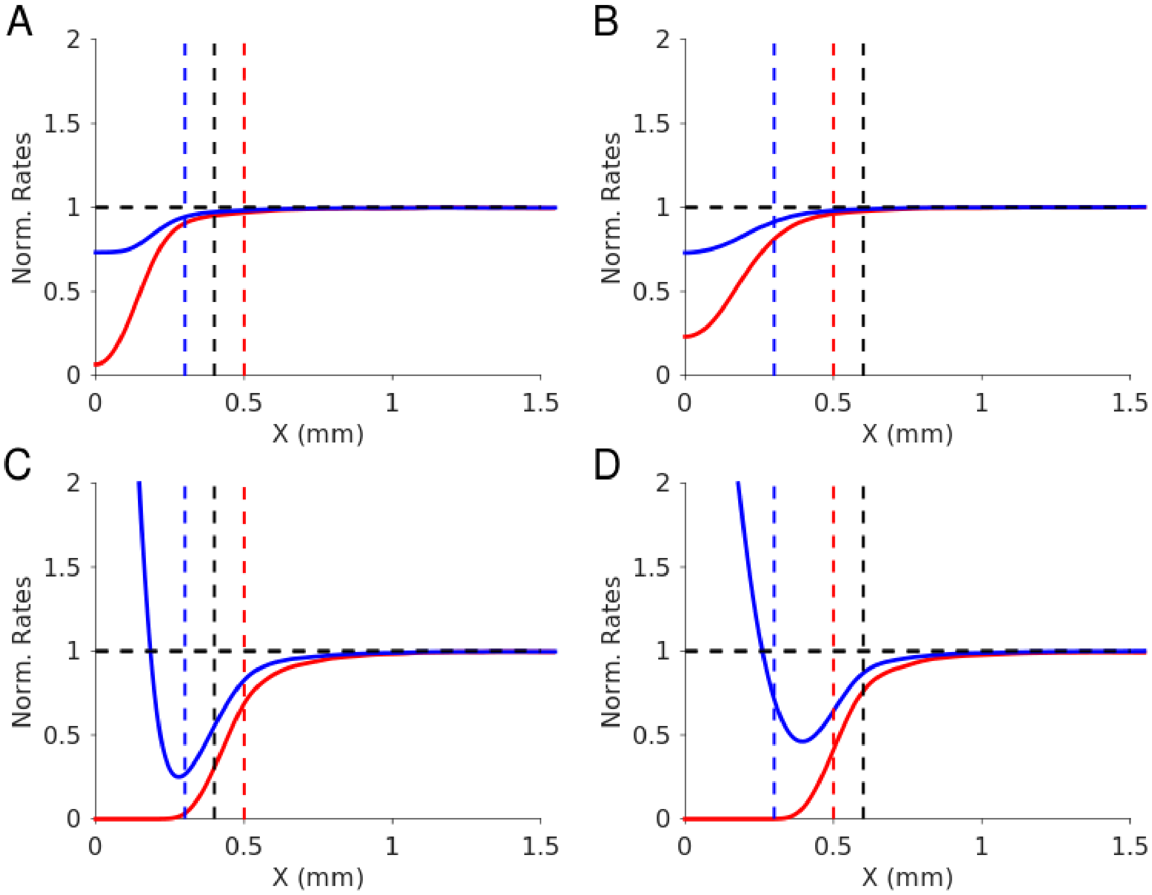


Figure S3. Gaussian photo-activation with Exponential Interactions. Interaction lengths are heterogeneous. $N = 120000$, $K = 2000$. Parameters as in Fig. 12.

References

Ben-Yishai R, Hansel D, Sompolinsky H. 1997. Traveling waves and the processing of weakly tuned inputs in a cortical network module. *J Comput Neurosci* **4**:57–77.

Coombes S. 2005. Waves, bumps, and patterns in neural field theories. *Biol Cybern* **93**:91–108.

Ebsch C, Rosenbaum R. 2018. Imbalanced amplification: A mechanism of amplification and suppression from local imbalance of excitation and inhibition in cortical circuits. *PLoS Comput Biol* **14**:e1006048.

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

Freund TF, Martin KA, Soltesz I, Somogyi P, Whitteridge D. 1989. Arborisation pattern and postsynaptic targets of physiologically identified thalamocortical afferents in striate cortex of the macaque monkey. *J Comp Neurol* **289**:315–336.

Hioki H, Okamoto S, Konno M, Kameda H, Sohn J, Kuramoto E, Fujiyama F, Kaneko T. 2013. Cell type-specific inhibitory inputs to dendritic and somatic compartments of parvalbumin-expressing neocortical interneuron. *J Neurosci* **33**:544–555.

Kato HK, Asinof SK, Isaacson JS. 2017. Network-Level Control of Frequency Tuning in Auditory Cortex. *Neuron* **95**:412–423.e4.

Kilpatrick ZP, Ermentrout B. 2013. Wandering Bumps in Stochastic Neural Fields. *SIAM J Appl Dyn Syst* **12**:61–94.

Landry P, Deschênes M. 1981. Intracortical arborizations and receptive fields of identified ventrobasal thalamocortical afferents to the primary somatic sensory cortex in the cat. *J Comp Neurol* **199**:345–371.

Levy RB, Reyes AD. 2012. Spatial Profile of Excitatory and Inhibitory Synaptic Connectivity in Mouse Primary Auditory Cortex. *Journal of Neuroscience*. doi:10.1523/jneurosci.5158-11.2012

Li N, Chen S, Guo ZV, Chen H, Huo Y, Inagaki HK, Chen G, Davis C, Hansel D, Guo C, Svoboda K. 2019. Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *Elife* **8**. doi:10.7554/eLife.48622

Li N, Chen S, Guo ZV, Chen H, Huo Y, Inagaki HK, Davis C, Hansel D, Guo C, Svoboda K. n.d. Spatiotemporal limits of optogenetic manipulations in cortical circuits. doi:10.1101/642215

Mahrach A, Chen G, Li N, van Vreeswijk C, Hansel D. 2019. Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *arXiv [q-bioNC]*.

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

Rosenbaum R, Doiron B. 2014. Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections. *Phys Rev X* **4**:021039.

Shapley R, Hawken M, Ringach DL. 2003. Dynamics of orientation selectivity in the primary visual cortex and the importance of cortical inhibition. *Neuron* **38**:689–699.

Somers DC, Nelson SB, Sur M. 1995. An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci* **15**:5448–5465.

Stepanyants A, Martinez LM, Ferecsko AS, Kisvarday ZF. 2009. The fractions of short- and long-range connections in the visual cortex. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.0810390106

Van Vreeswijk C, Sompolinsky H. 2005. Irregular activity in large networks of neuronsLes Houches. Elsevier. pp. 341–406.

Appendix

Networks with Gaussian interactions

We consider a second model where the probability of connection has a Gaussian profile. Similar studies were conducted by (Ebsch and Rosenbaum, 2018; Rosenbaum and Doiron, 2014) in the case where the interaction length only depended on the presynaptic population. We will show that upon photostimulation of I, these network models behave qualitatively in a similar fashion than in the case where the interactions were exponentials (at least for finite K). We will then give some necessary conditions upon which the networks can operate in the balanced regime when the probability of connection depends on both the pre and postsynaptic populations and investigate the response of such networks with photostimulation.

Balance of excitation and inhibition

The probability of connection between the neurons is now assumed to have a periodic Gaussian profile,

$$P_{i,j}^{AB} = Z_B \sum_{k=-\infty}^{+\infty} e^{-\frac{(x_i^A - x_j^B + k\,L)^2}{2\sigma_{AB}^2}} \tag{1}$$

where Z_B is a normalization factor such that $\sum_j P_{ij}^{AB} = K$, so that the neurons on average receive K inputs per population. σ_{AB} is the characteristic length of the interaction from population B to population A.

In the large K limit, the balance of the net input into the neurons in each population implies that

$$J_{E0}\, r_0 + I_{opto}^E(x) + J_{EE} \circledast r_E(x) - J_{EI} \circledast r_I(x) = 0 \tag{2}$$

$$J_{I0}\, r_0 \; + I_{opto}^I(x) + J_{IE} \circledast r_E(x) - J_{II} \circledast r_I(x) = 0 \tag{3}$$

Therefore, after taking the Fourier transform of the previous expressions, one can write the Fourier moments of the population activities as

$$\hat{r}_E(k) = r_E^0\, \delta(k) \; + \frac{\Gamma_{E0}\, J_{II}\, e^{-(\sigma_{E0}^2 + \sigma_{II}^2)k^2} - \Gamma_{I0}\, J_{EI}\, e^{-(\sigma_{I0}^2 + \sigma_{EI}^2)k^2}}{J_{EI}\, J_{IE}\, e^{-(\sigma_{EI}^2 + \sigma_{IE}^2)k^2} - J_{EE}\, J_{II}\, e^{-(\sigma_{EE}^2 + \sigma_{II}^2)k^2}} \tag{4}$$

$$\hat{r}_I(k) = r_I^0\, \delta(k) + \frac{\Gamma_{E0}\, J_{IE}\, e^{-(\sigma_{E0}^2 + \sigma_{IE}^2)k^2} - \Gamma_{I0}\, J_{EE}\, e^{-(\sigma_{I0}^2 + \sigma_{EE}^2)k^2}}{J_{EI}\, J_{IE}\, e^{-(\sigma_{EI}^2 + \sigma_{IE}^2)k^2} - J_{EE}\, J_{II}\, e^{-(\sigma_{EE}^2 + \sigma_{II}^2)k^2}} \tag{5}$$

where $r_E^0 = \frac{J_{E0}\,J_{II} - J_{I0}\,J_{EI}}{J_{EI}\,J_{IE} - J_{EE}\,J_{II}} r_0$ and $r_I^0 = \frac{J_{E0}\,J_{IE} - J_{I0}\,J_{EE}}{J_{EI}\,J_{IE} - J_{EE}\,J_{II}} r_0$.

When $\sigma_{EE} = \sigma_{IE} = \sigma_E$ and $\sigma_{EI} = \sigma_{II} = \sigma_I$, these expression simplifies and the solution is balanced as long as the range of the feedforward interactions is larger than the range of the recurrent ones, namely $\sigma_{E0} > \sigma_E, \sigma_I$ and $\sigma_{I0} > \sigma_E, \sigma_I$.

When all the interactions are taken into account, the balance of the solution imposes that in the limit where k is large, the moments of the population activities remain finite. Therefore one can a priori consider two cases: 1) the case where $\sigma_{EE}^2 + \sigma_{II}^2 > \sigma_{EI}^2 + \sigma_{IE}^2$; 2) the case where $\sigma_{EI}^2 + \sigma_{IE}^2 > \sigma_{EE}^2 + \sigma_{II}^2$. Nevertheless, when $\sigma_{EI}^2 + \sigma_{IE}^2 > \sigma_{EE}^2 + \sigma_{II}^2$, one can show that the network steady-state is unstable.

When $\sigma_{EE}^2 + \sigma_{II}^2 > \sigma_{EI}^2 + \sigma_{IE}^2$, as $k \to \infty$ the moments can be reduced to

$$\hat{r}_E(k) \sim \frac{e^{(\sigma_{EI}^2 + \sigma_{IE}^2)k^2}}{J_{EI}\,J_{IE}} \left(\Gamma_{E0}\,J_{II}\,e^{-(\sigma_{E0}^2 + \sigma_{II}^2)k^2} - \Gamma_{I0}\,J_{EI}\,e^{-(\sigma_{I0}^2 + \sigma_{EI}^2)k^2}\right) \tag{6}$$

$$\hat{r}_I(k) \sim \frac{e^{(\sigma_{EI}^2 + \sigma_{IE}^2)k^2}}{J_{EI}\,J_{IE}} \left(\Gamma_{E0}\,J_{IE}\,e^{-(\sigma_{E0}^2 + \sigma_{IE}^2)k^2} - \Gamma_{I0}\,J_{EE}\,e^{-(\sigma_{I0}^2 + \sigma_{EE}^2)k^2}\right) \tag{7}$$

Therefore, the moments remain finite when

$$\sigma_{E0}^2 + \sigma_{II}^2 > \sigma_{EI}^2 + \sigma_{IE}^2 \ , \sigma_{I0}^2 > \sigma_{IE}^2 \ , \ \sigma_{E0}^2 > \sigma_{EI}^2 \text{ and } \sigma_{I0}^2 + \sigma_{EE}^2 > \sigma_{EI}^2 + \sigma_{IE}^2 \tag{8}$$

In particular, when the perturbation is limited to the inhibitory population one gets

$$\hat{r}_E(k) \sim \ - \frac{\Gamma_{I0}\,J_{EI}}{J_{IE}}\ e^{-(\sigma_{I0}^2 - \sigma_{IE}^2)k^2} \tag{9}$$

$$\hat{r}_I(k) \sim \ - \frac{\Gamma_0\,J_{EE}}{J_{EI}\,J_{IE}}\ e^{-(\sigma_{I0}^2 + \sigma_{EE}^2 - \sigma_{EI}^2 - \sigma_{IE}^2)k^2} \tag{10}$$

and the necessary conditions

$$\sigma_{I0}^2 > \sigma_{IE}^2 \text{ and } \sigma_{I0}^2 + \sigma_{EE}^2 > \sigma_{EI}^2 + \sigma_{IE}^2 \tag{11}$$

This set of inequalities gives an ensemble of necessary conditions upon which a network could operate in a balanced state. Unfortunately, one cannot derive the general conditions for which such a state would exist. Therefore, we will rely on simulations to give a qualitative description of the network responses and test the robustness of our predictions with N and K .

We can guess from these conditions that in order to maintain a balanced state when slightly perturbing the inhibitory population σ_{IE} must be smaller than the perturbation radius. Moreover, since σ_{IE} shapes the Fourier moments $r_E(k)$ when $k \to \infty$, small values of σ_{IE} should ensure that $r_E(x)$ stays below its baseline. Similarly, σ_{EI} controls the rate upon which the Fourier moments of $r_I(x)$ decay to zero, we expect that the larger σ_{EI} is, the larger a perturbation would affect the inhibitory response and consequently, the suppression length, x_c.

Numerical simulations

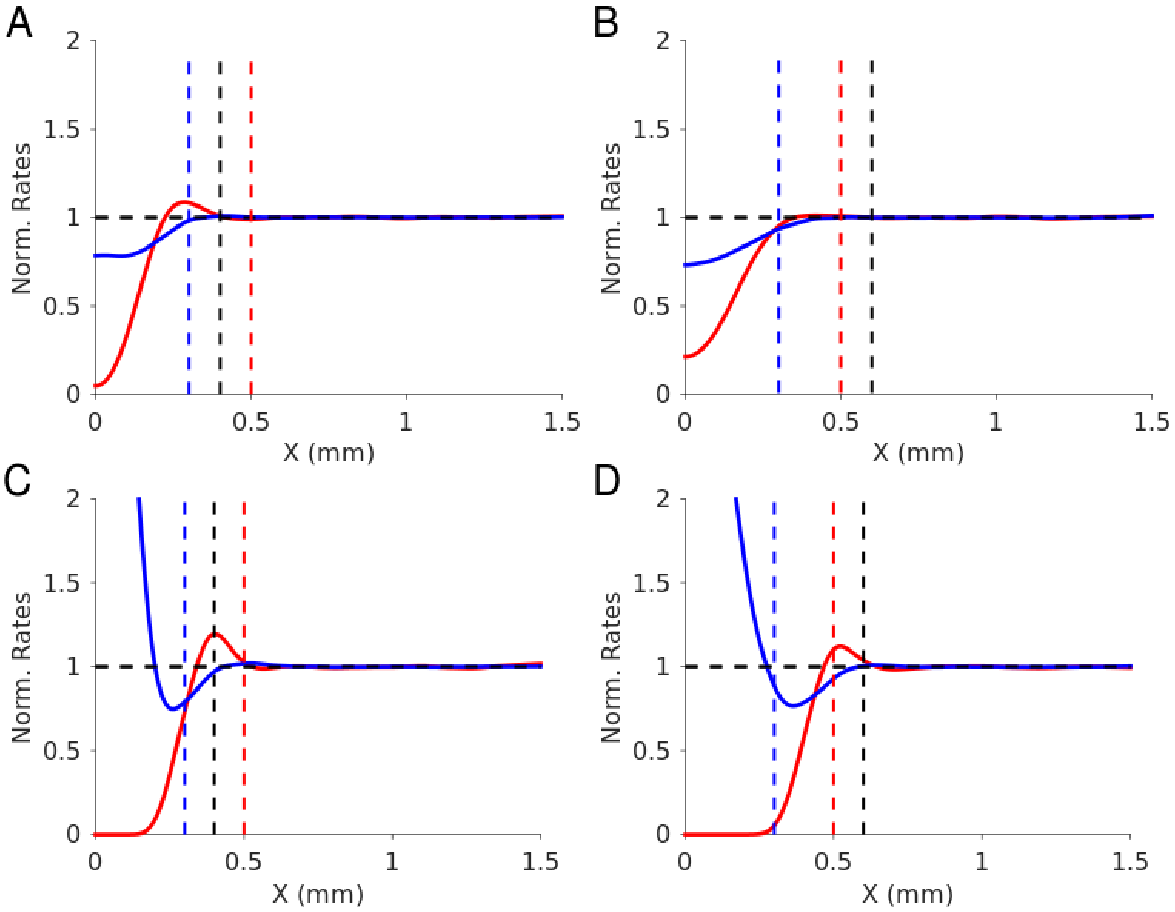


Figure A1. Gaussian photo-stimulation of the inhibitory population in EI network with Gaussian interactions. A. $\Gamma_0 = 0.1\ a.u.$. B. $\Gamma_0 = 2\ a.u.$. Left: the radius of the perturbation is smaller than the length of the recurrent excitatory interactions: $4\,\sigma_E = 0.5\ mm > 4\,\sigma_0 = 0.4\ mm > 4\,\sigma_I = 0.3\ mm$. Right: the radius of the perturbation, is greater than the length of the recurrent excitatory interactions: $4\,\sigma_0 = 0.6\ mm > 4\,\sigma_E = 0.5\ mm > 4\,\sigma_I = 0.3\ mm$. Dashed-line: interaction length ($4\,\sigma_A$), Red: excitatory neurons; blue: inhibitory neurons. black: $4\,\sigma_0$. Parameters in Table 1. $N_E = N_I = 15000$, $K = 500$.

Figure A1 shows the spatial profile of the responses of the excitatory and inhibitory populations (normalized to their baseline) upon strong photostimulation of the inhibitory population with a Gaussian perturbation for two laser intensities, Γ_0 and two perturbation radius, $4\,\sigma_0$. Here, the interactions are Gaussian and the interaction lengths only depend on the presynaptic population. For a given perturbation, the response of the network is qualitatively identical to the response shown in Fig. 6.

Therefore, for finite K, whether one considers exponentially decaying interaction profiles or Gaussian interaction profile has little to no effect on the network response.

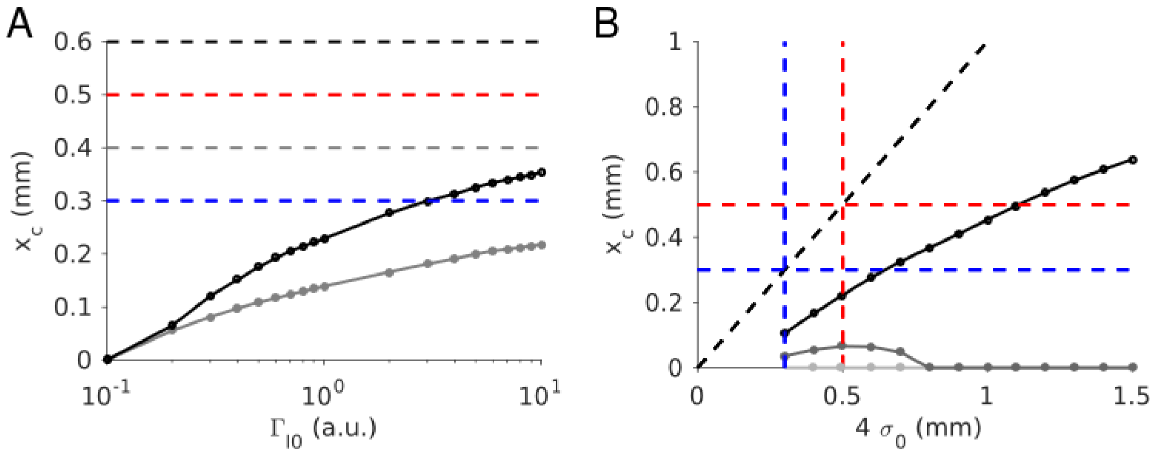


Figure A2. Photo-suppression length v.s. laser intensity and laser radius. A. Suppression length, x_c *v.s.* photostimulation intensity, Γ_0 (Gaussian interactions). Black: $4\,\sigma_0 = 0.6\,mm$. Gray: $4\,\sigma_0 = 0.4\,mm$. B. Suppression length, x_c *v.s.* photostimulation radius, $4\,\sigma_0$. Black: $\Gamma_0 = 2\ a.u.$. Dark-gray: $\Gamma_0 = 0.2\ a.u.$. Gray: $\Gamma_0 = 0.1\ a.u.$. Dashed-line same as in Fig. 6. Parameters as in Fig. A1.

Figure A2 depicts the effect of changing the photostimulation intensity and radius on the length, x_c of the region where excitation is suppressed. Here again, the general properties of the dependence of x_c with Γ_0 and σ_0 derived in Fig. A1 seems to be preserved. Moreover, Figure A2 A plots the suppression length, x_c, *v.s.* the photostimulation intensity Γ_0 for two fixed photo-stimulation radius. For both photostimulation radius ($\sigma_0 > \sigma_E > \sigma_I$ and $\sigma_E > \sigma_0 > \sigma_I$) x_c increases monotonously with Γ_0 spanning continuously across the interaction length of the inhibitory and excitatory populations. Figure 9B plots the suppression length, x_c *v.s.* the photostimulation radius, $4\,\sigma_0$, for three fixed photo-stimulation intensities. When the photostimulation intensity is large, x_c monotonously increases with σ_0 with a slope smaller than 1. As σ_0 increases, x_c becomes more and more smaller than the photostimulation radius. Γ_0 becomes smaller, x_c becomes non-monotonous in σ_0. For sufficiently small Γ_0 and sufficiently large σ_0, x_c can even become zero.

Figure A3 shows the result of the numerical simulations for a set of interaction length such that $\sigma_{EE} > \sigma_{EI} > \sigma_{II} > \sigma_{IE}$ and Eq. [11] is satisfied. Similarly to the case when the interactions decay exponentially, when σ_{IE} is small, there is no rebound of the excitatory activity in the surround of the suppression. Moreover, for sufficiently large σ_{EI} the dependence of the suppression length with the intensity and radius of the perturbation are identical to the one in Fig. A2.

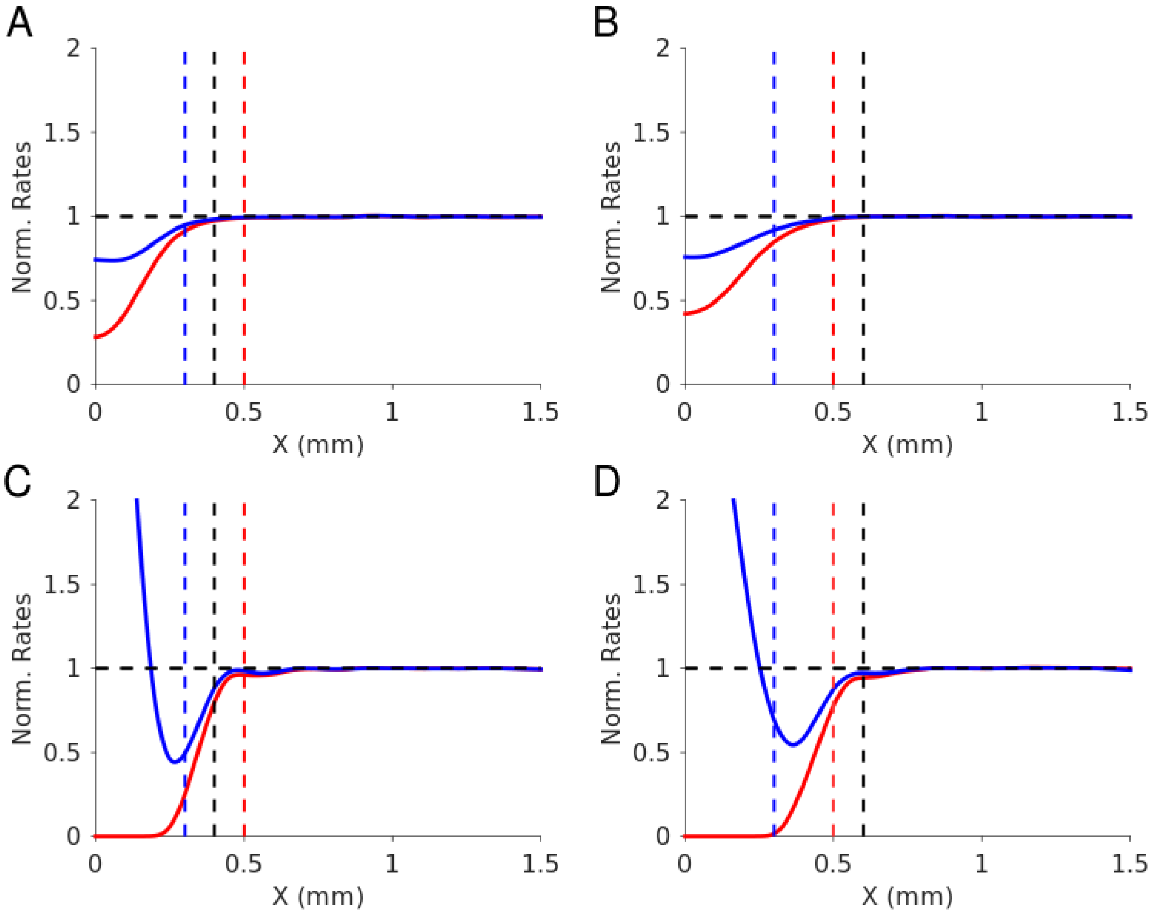


Figure A3. Gaussian photo-activation with Gaussian Interactions when $\sigma_{EE} > \sigma_{EI} > \sigma_{II} > \sigma_{IE}$. A. $\Gamma_0 = 0.1\ a.u.$. B. $\Gamma_0 = 2\ a.u.$. Left: the radius of the perturbation is smaller than the length of the recurrent excitatory interactions: $4\ \sigma_{EE} = 0.5\ mm > 4\ \sigma_0 = 0.4\ mm$. Right: the radius of the perturbation, is greater than the length of the recurrent excitatory interactions: $4\ \sigma_0 = 0.6\ mm > 4\ \sigma_{EE} = 0.5\ mm$. Dashed-line: interaction length ($4\ \sigma_A$), Red: excitatory neurons; blue: inhibitory neurons. black: $4\ \sigma_0$. Parameters in Table 1. $N_E = N_I = 15000$, $K = 500$. $\sigma_{EE} = 0.125\ mm,\ \sigma_{EI} = 0.125mm,\ \sigma_{II} = 0.075\ mm,\ \sigma_{IE} = 0.0625\ mm$.

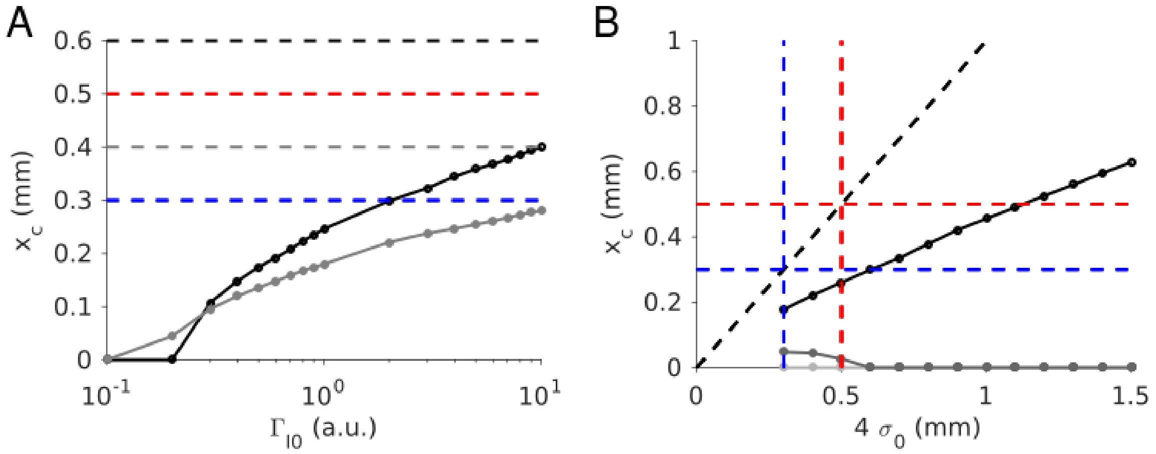


Figure A4. Photo-suppression length v.s. laser intensity and laser radius when $\sigma_{EE} > \sigma_{EI} > \sigma_{II} > \sigma_{IE}$. A. Suppression length, x_c *v.s.* photostimulation intensity, Γ_0. Black: $4\,\sigma_0 = 0.6\ mm$. Gray: $4\,\sigma_0 = 0.4\ mm$. B. Suppression length, x_c *v.s.* photostimulation radius, $4\,\sigma_0$. Black: $\Gamma_0 = 2\ a.u.$. Dark-gray: $\Gamma_0 = 0.2\ a.u.$. Gray: $\Gamma_0 = 0.1\ a.u.$. Dashed-line same as in Fig. A2. Parameters as in Fig. A3. Parameters as in Fig. 10. $N_E = N_I = 15000$, $K = 500$.

Chapter 4

Discussion

The standard model of the balanced state with one excitatory and one inhibitory population (van Vreeswijk and Sompolinsky, 1998, 1996) has been very successful in accounting for some of the basic features of cortical dynamics observed experimentally and is supported by some in-vitro (Shu et al., 2003) and in-vivo (Haider et al., 2006) experimental studies. However, it is a simplified description of reality and thus suffers from several significant limitations. Whether cortical neurons do operate in this regime remains to a large extent an open question. In this thesis, we investigate this question by extending the theory of standard balanced hypotheses. Our strategy is to infer from the cortical responses to specific optogenetic perturbations the state in which the cortex operates.

Summary

The role of inhibitory interneuron diversity

Cortical inhibitory neurons comprise several populations that differ in their intrinsic properties, their patterns of connectivity, and the feedforward inputs they receive. The recent experimental developments combining molecular, genetic and optical techniques have provided new tools to target certain neuronal subpopulations (excitatory, PV, SOM, VIP) and perturb their activity *in vivo*. In this study, we extended the balanced state to models that take into account the diversity of interneurons. We then investigated how these different populations contribute to the global balance in the network and we investigated the effects of optogenetic manipulation of some of these populations on the network dynamics. Given a network connectivity scheme, we predicted the network responses to various types of optogenetic perturbations. These need to be experimentally tested in future works.

Spatial spread of inhibition

The spatial range and the intensity of the perturbations can be manipulated by changing the size and the power of the laser beam. Here we studied how changing these parameters affect the state of the network, *e.g.* the spatial profile of the activity. Our models account for the spatial spread of photo-inhibition in the experimental data of Li et al. We found a significant dependence of the photo-suppression length, with the intensity and radius of the laser. For large stimulus intensities, it increases with the stimulus radius. This is not the case when the stimulus intensity is small where the response is non-linear. In order to explain the concomitant decrease of the excitatory and inhibitory responses in the surround of the suppression, we derived conditions upon the length of the recurrent interactions for finite-size networks. We found that in order to account for the experimental observations, excitatory to inhibitory connections must be small in comparison with the stimulus radius. Conversely, to obtain a sufficiently large suppression in the excitatory response, the inhibitory to excitatory projections must be broadened. Future experimental and theoretical work would be required to understand the effect of photostimulation of the cortex on a scale of the length of the inhibitory to inhibitory interactions. Finally, extending the theory to networks with multiple inhibitory populations would be necessary in order to understand which inhibitory interactions are crucial in shaping the network response.

Beyond the cortex

Interneuron diversity is not restricted to cortical areas. Indeed, interneurons have been reported to also exhibit a large amount of heterogeneity outside of the cortex in regions like the hippocampus, the cerebellum and the basal ganglia (*e.g.* in the cerebellum Purkinje cells, the sole output of the cerebellar cortex and main computational cell type, receive input from several classes of interneurons that are thought to play an essential role in controlling cerebellar cortical output during motor behavior (Bower, 2010; Dizon and Khodakhah, 2011)). Several lines of evidence suggest that interneuron dysfunction may contribute to cognitive abnormalities in several brain diseases. For example, when PV interneurons are dysfunctional in the hippocampus, the network exhibits hypersynchrony and oscillatory rhythmic activity that lead to cognitive abnormalities in Alzheimer's disease (Hijazi et al., 2019; Verret et al., 2012). The precise causes and physiological consequences of this dysfunction still remain to be understood from a physiological as well as from a mathematical point of view. Indeed, investigating the mechanisms by which interneurons could alter network dynamics and function could be of particular interest to the development of specific therapeutic treatments. In particular, with the advance of optogenetics, manipulation of specific interneurons could be a good strategy to improve brain functions in pathological conditions.

(Mastro et al., 2017) have reported that optogenetic activation of different neuronal populations in the basal ganglia, can restore movement and attenuate pathological activity even after stimulation offset in parkinsonian mice (Fig. 1B). We will here give a brief account of how simple mathematical reasoning can explain the contributions of these different interneurons to the mechanism underlying parkinsonian behavior in a model of the basal ganglia.

Interneurons in the basal ganglia

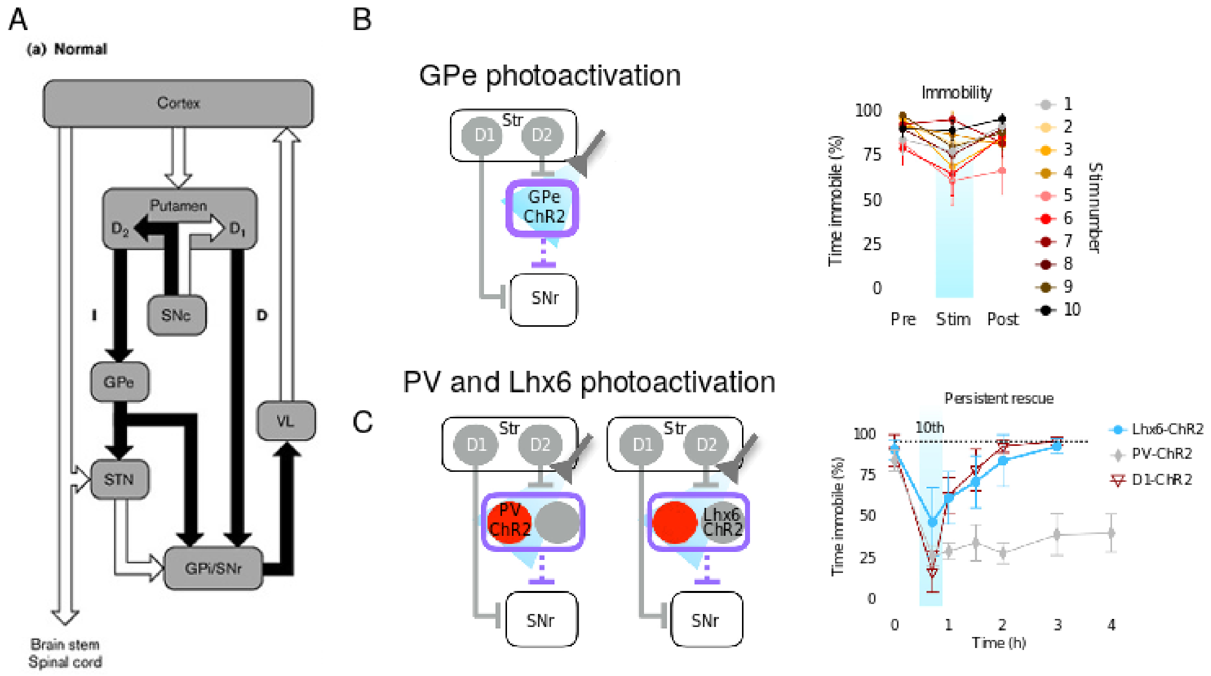


Figure 1. Optogenetic manipulation of the basal ganglia. **A.** Nuclei and pathway of the basal ganglia. Panel from (Wichmann and DeLong, 1996)**. B.** Global GPe stimulation does not rescue movement in dopamine deprived mice. **C.** Stimulation of specific interneurons rescue movement in dopamine deprived mice. **B-C** panels adapted from (Mastro et al., 2017).

Basal Ganglia (BG) consist of a set of interconnected nuclei intensely connected with other cerebral structures, including the cerebral cortex, the thalamus, and dopaminergic nuclei (Fig. 1A). BG are involved in motor control and its dysfunction is known to be a key component underlying movement disorders.

BG main input structure, the striatum, receives topographically ordered excitatory projections from cortex (McGeorge and Faull 1989; Wiesendanger et al., 2004) and provides inhibitory inputs into the substantia nigra pars reticulata (SNR), its main output structure, and into the globus pallidus pars externa (GPe), one of BG internal nuclei. GPe mostly sends GABAergic projections to the subthalamic nucleus (STN) which in turn provides glutamatergic projections to the SNR. Last, SNR provides feedback inhibition to the VentroAnterior (VA), VentroLateral (VL) and VentroMedial (VM) thalamus. Dopaminergic SNc neurons project to several BG nuclei (for review, see Graybiel, 2000), but these projections and the dopaminergic receptors are more numerous in the striatum (Haber and Fudge, 1997).

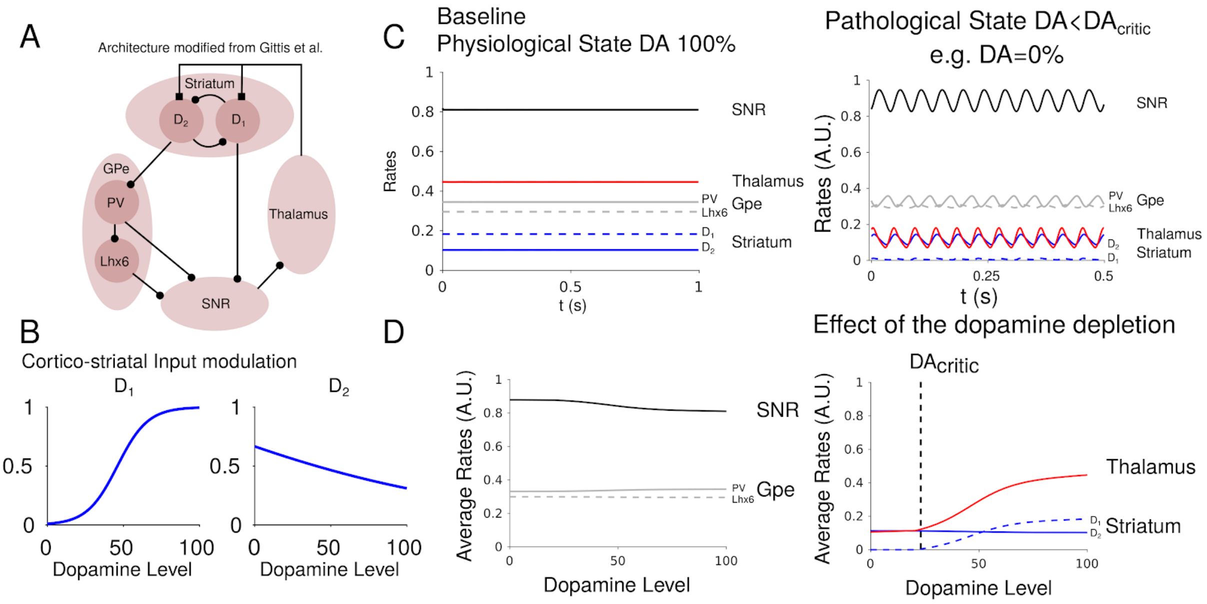


Figure 2. A model of the BG. **A.** Architecture of the network. **B.** Cortico-striatal input modulation with the dopamine level (DA) **C.** Dopamine depletion leads to oscillatory activity. **D.** Left: Activities of SNR and Gpe neurons present little change with DA level. Right**:** Thalamus and D2 neurons decrease their activity with the DA level.

The Cortex-BG-Thalamic loop is known to decompose into three parallel pathways: (1) a direct pathway through the cortex-striatum-SNR-cortex loop (Albin, Young, and Penney 1989; Alexander and Crutcher 1990); (2) an indirect pathway through the cortex-striatum-GPe-STN-SNR-cortex loop; (3) an hyperdirect pathway through the cortex-STN-SNR-cortex (Nambu, Tokuno, and Takada, 2002).

Previous theoretical studies have investigated how the competition between these different loops are responsible for a variety of functions and dysfunctions, and in particular for the generation of oscillatory dynamics that is thought to be a key component of Parkinson's disease. We propose here some preliminary results of a model where oscillatory neural activity can arise from the BG itself and spread towards the cortex and the thalamus when dopamine is depleted. Our model places GPe interneuronal heterogeneity as a key component of the BG dysfunction.

We consider a simple rate model of the Basal Ganglia consisting of the striatum composed of two inhibitory populations D_1 and D_2, with D_1 and D_2 interconnected. Following (Mastro et al., 2017) (Fig. 1B), our GPe consists of two inhibitory populations PV and Lhx6, with projections from PV neurons to Lhx6 neurons only. The SNR receives inhibition from all the populations in the GPe and from D_1 in the striatum. Finally, the thalamus receives inhibitory feedback from the SNR. For simplicity, we omit the thalamus-cortex-striatum pathway that provides excitation to the striatum and considers that the thalamus directly provides excitatory inputs to D_1 and D_2 in the striatum.

We can, therefore, distinguish between two distinct pathways: (1) a direct pathway consisting of the Striatum-SNR-Thalamus-Striatum loop; (2) an indirect pathway consisting of the Striatum-GPe-SNR-Thalamus-Striatum loop. Because we introduced interneuronal diversity in the GPe, the indirect pathway (2) can be decomposed into two sub-loops one involving only the PV neurons, and the other including all the inhibitory populations in the GPe (Fig. 2A).

Cortico-striatal transmission is thought to be altered by DA depletion (Calabresi et al., 2000). Moreover, the signal-to-noise ratio in the striatum has been reported to depend on the DA level (O'Donnell, 2003; Nicola et al., 2004). We, therefore, model the effect on the dopamine level (DA) received by the neurons in the striatum as follow: the excitatory I input to D_1 neurons is a sigmoid function of the dopamine level (0 when $DA = 0$ and 1 when $DA = 100\%$); the excitatory input to D_2 neurons is a slightly decreasing function of the dopamine level (Fig. 2B).

The network goes from a stable steady-state to an oscillatory state when dopamine concentration is depleted below a given level $DA = DA_{critic}$ (Fig. 2C). This is because as dopamine is depleted, the gain of the positive loops involving D_1 in the circuit is diminished. For $DA = DA_{critic}$, these loops can no longer counter the negative loops involving D_2, namely the loop D_2-PV-SNR-Thalamus-D_2, and the network exhibits synchronous oscillations. In fact, the loop D_2-PV-SNR-Thalamus-D_2 could in principle be balanced by the contribution of the loop that involves the Lhx6 neurons (namely the loop D_2-PV-Lhx6-SNR-Thalamus-D_2). This is the case when the strength of the projections from PV neurons in GPe to the SNR would be smaller than the gain of the indirect pathway from GPe to SNR namely, the product of the strength of the interaction from PV neurons to Lhx6 neurons and the strength of the interaction from Lhx6 neurons to the SNR. In fact, that $J_{PV \to SNR} > J_{PV \to Lhx6} \, J_{Lhx6 \to SNR}$ is also required to ensure that in the normal physiological state (when dopamine is not depleted) the steady-state is stable. As a postdiction, we observed in the model, that even when dopamine level is depleted beyond DA_{critic}, small activation of PV neurons or large activation of Lhx6 suppress the synchronous oscillatory network activity.

Perspectives

Effects of optogenetic manipulation on feature selectivity

Several recent studies have investigated the contribution of different inhibitory neurons to stimulus selectivity in the mouse visual cortex (Atallah et al., 2012; El-Boustani and Sur, 2014; Lee et al., 2012; Wilson et al., 2012). Cortical inhibitory cells have been reported to be more broadly tuned than PCs. Between inhibitory subtypes tuning properties are known to vary. VIP neurons have been reported to be more selective for orientation than SOM and PV neurons in mouse V1 (Kerlin et al., 2010). A few Inhibitory neurons have been reported to exhibit strong stimulus selectivity that is similar to that of PCs (Runyan et al., 2010). Similarly, inhibitory neurons in other sensory areas have also been reported to be more broadly tuned than PCs (Kato et al., 2017). What factors underlie tuning properties of interneurons, and why are inhibitory neurons generally more broadly tuned than PCs? How do distinct interneurons contribute to the shaping of the tuning response of PCs?

There is no consensus regarding the role of distinct inhibitory populations in shaping the response of the PCs. Some studies found that SOM neurons sharpen orientation tuning (Wilson et al., 2012), while that activation of PV neurons have little to no effect on tuning but changed response gain ((Atallah et al., 2012), Fig. 3A). Others reported that PV neurons sharpened tuning, but not SOM neurons ((Lee et al., 2012), Fig. 3B).

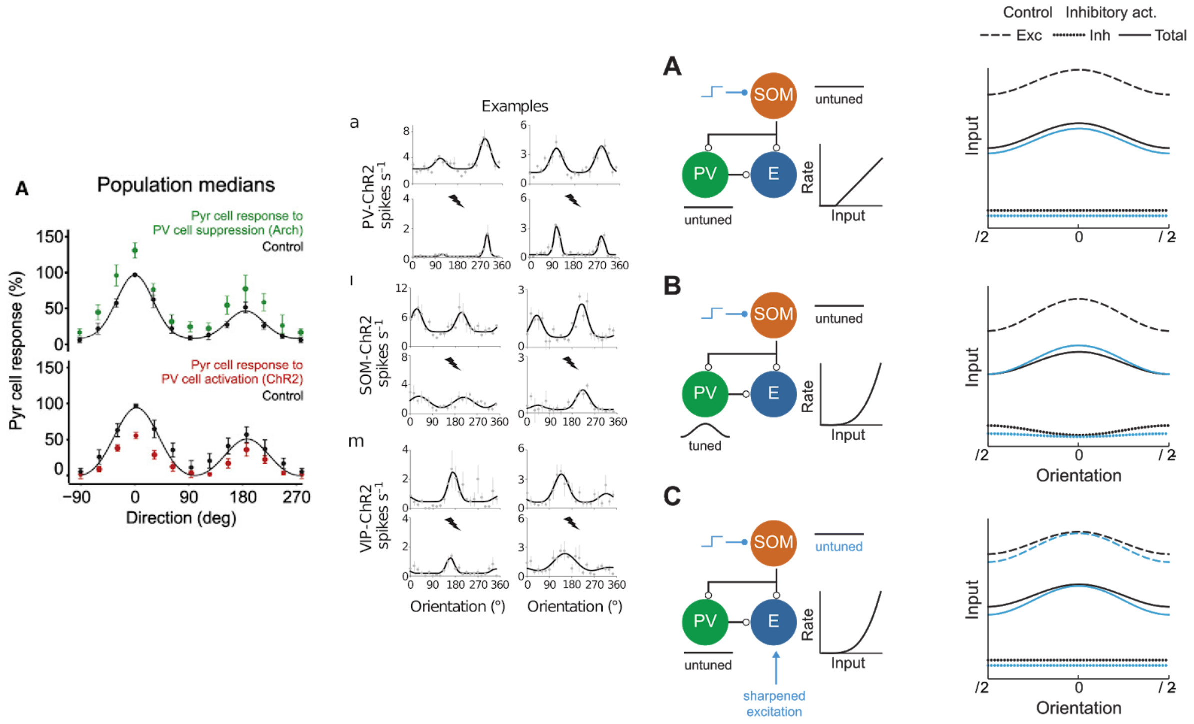


Figure 3. Interneurons shape the tuning properties of PCs. **A**. PV photostimulation decrease PCs stimulus selectivity. Panel adapted from (Atallah et al., 2012). **B**. PV photostimulation enhances V1 stimulus selectivity. ChR2-mediated changes in stimulus selectivity for PV stimulation (top), SOM stimulation (middle) and VIP stimulation (bottom). Panel adapted from (Lee et al., 2012). **C**. Three hypotheses for the mechanism of subtraction. Top: untuned inhibition leads to subtraction if excitatory f-I curves are linear. Middle: withdrawal of tuned PV inhibition leads to a sharpening of subthreshold input and approximate subtraction if excitatory f-I curves are nonlinear. Bottom: nonlinear dendritic integration allows untuned SOM inhibition to sharpen tuned excitatory input, leading to subtraction if excitatory f-I curves are nonlinear. Panel and caption adapted from (Litwin-Kumar et al., 2016).

In a recent study, (Litwin-Kumar et al., 2016) investigated the dynamics of recurrent networks of weakly connected neurons with multiple inhibitory subtypes (namely, the PV, SOM and VIP neurons) and connectivity matching the one reported in mouse primary visual cortex. They studied the role of each interneuron subtype in the mechanisms of disinhibition, surround suppression, and modulation of orientation tuning in networks operating in an inhibition stabilized regime (Fig. 3C). To this date, no similar study was conducted in a balanced network model with multiple inhibitory populations.

Response of cortical networks to deterministic noise

The dynamics of fluctuations is at the core of the balanced state regime and the use of stochastic inputs could refine the characterization of its parameters. A particular type of noise could be used to test how fluctuations propagate within networks. Indeed, it is possible to generate a stimulus that has all characteristics of noise but that is deterministically equal from trial to trial (frozen noise). It has been shown that such a signal may produce highly precise spike timings in single cells in vitro (Mainen and Sejnowski, 1995; Nowak et al., 1997). By measuring the trial to trial variability of neural responses, it is possible to measure the propagation of fluctuations within the network. It is now possible to reproduce such inputs experimentally with optogenetic tools. One could analyze the dependence of the network variability on the frequency of the stimulus. This will be used to determine whether slow channel dynamics, motifs in the connectivity, or fluctuations in the external input are the dominant source of fluctuations in the network activity on short time scales (tens of milliseconds).

Diversity of short-term plasticity

In addition to specific patterns of connectivity, inhibitory interneurons also exhibit preferential patterns of short-term synaptic plasticity (STP) (Gupta et al., 2000; Karnani et al., 2014). For instance, synapses on and from PV neurons are typically depressing while synapses on and from SOM neurons are typically facilitating. While the experimental characterization of these STP motifs is rapidly progressing, the understanding of their function has been lagging behind. STP has been shown to be a robust source of non-linearity for networks operating in the balanced regime. In particular, it enables the co-existence of multiple balanced states (Hansel and Mato, 2013; Mongillo et al., 2012). Further work should investigate the impact of the diversity of STP patterns on the operation of balanced networks with multiple inhibitory populations. In particular, one should investigate how the number of fixed points and their stability, changes depending on a given pattern of STP in the connections. Specifically, it should focus on the possibility of multi-stability between states with approximately the same activity level in the excitatory population but different activity levels in the inhibitory populations.

Recent studies report strong evidence supporting the view that PV neurons are a key player in the generation and maintenance of gamma, theta, and ripple oscillations in mouse cortex (Cardin et al., 2009; Sohal et al., 2009; Stark et al., 2014, 2013). These rhythms, and the circuits underlying their generation, illustrate the importance of the fast signaling properties of PV cells (Bartos et al., 2007). The mechanisms by which PV cells control the oscillatory components of neural activity are still to be elucidated.

Cortical circuits are known to generate rhythmic activity in different frequency ranges. However, in two-population balanced models, due to the linearity of synaptic interactions, no balanced oscillatory regime is possible. The non-linearity provided by STP could endow the network with the ability to produce balanced oscillations. Future work must investigate the mechanism underlying the generation of rhythm in multi-inhibitory population network models with STP as a function of the levels of feedforward inputs to the different neuronal populations. In this case, it will be of extreme interest to characterize the statistical features stimulus-evoked activity depending on the global state of the network (*i.e.*, asynchronous *v.s.* balanced oscillations).

Another interesting question is how the state of the network affects the statistics of stimulus-evoked activity (*e.g.*, average activity, selectivity in different neuronal populations). Besides providing a source of static non-linearity, STP also can transiently modify the response properties of the network on relatively long time-scales (*i.e.*, 100-1000 ms). This can in principle enhanced the network computational abilities since the response to an afferent stimulation will depend on previously encountered stimuli. Further work should characterize the statistical features of stimulus-evoked activity depending on the previous history of stimulation. In particular, it would be interesting to measure the fraction of variability of evoked response that can be explained by history-dependent effects.

References

Atallah BV, Bruns W, Carandini M, Scanziani M. 2012. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**:159–170.

Bartos M, Vida I, Jonas P. 2007. Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat Rev Neurosci* **8**:45–56.

Bower JM. 2010. Model-founded explorations of the roles of molecular layer inhibition in regulating Purkinje cell responses in cerebellar cortex: more trouble for the beam hypothesis. *Frontiers in Cellullar Neuroscience*. doi:10.3389/fncel.2010.00027

Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, Moore CI. 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**:663–667.

Dizon MJ, Khodakhah K. 2011. The role of interneurons in shaping Purkinje cell responses in the cerebellar cortex. *J Neurosci* **31**:10463–10473.

El-Boustani S, Sur M. 2014. Response-dependent dynamics of cell-specific inhibition in cortical networks in vivo. *Nat Commun* **5**:5689.

Gupta A, Wang Y, Markram H. 2000. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**:273–278.

Haider B, Duque A, Hasenstaub AR, McCormick DA. 2006. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* **26**:4535–4545.

Hansel D, Mato G. 2013. Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. *Journal of Neuroscience*. doi:10.1523/jneurosci.3455-12.2013

Hijazi S, Heistek TS, Scheltens P, Neumann U, Shimshek DR, Mansvelder HD, Smit AB, van Kesteren RE. 2019. Early restoration of parvalbumin interneuron activity prevents memory loss and network hyperexcitability in a mouse model of Alzheimer’s disease. *Mol Psychiatry*. doi:10.1038/s41380-019-0483-4

Karnani MM, Agetsuma M, Yuste R. 2014. A blanket of inhibition: functional inferences from dense inhibitory connectivity. *Curr Opin Neurobiol* **26**:96–102.

Kato HK, Asinof SK, Isaacson JS. 2017. Network-Level Control of Frequency Tuning in Auditory Cortex. *Neuron* **95**:412–423.e4.

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

Lee S-H, Kwan AC, Zhang S, Phoumthipphavong V, Flannery JG, Masmanidis SC, Taniguchi H, Huang ZJ, Zhang F, Boyden ES, Deisseroth K, Dan Y. 2012. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**:379–383.

Litwin-Kumar A, Rosenbaum R, Doiron B. 2016. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *J Neurophysiol* **115**:1399–1409.

Mainen ZF, Sejnowski TJ. 1995. Reliability of spike timing in neocortical neurons. *Science* **268**:1503–1506.

Mastro KJ, Zitelli KT, Willard AM, Leblanc KH, Kravitz AV, Gittis AH. 2017. Cell-specific pallidal intervention induces long-lasting motor recovery in dopamine-depleted mice. *Nat Neurosci* **20**:815–823.

Mongillo G, Hansel D, van Vreeswijk C. 2012. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys Rev Lett* **108**:158101.

Nowak LG, Sanchez-Vives MV, McCormick DA. 1997. Influence of low and high frequency inputs on spike timing in visual cortical neurons. *Cereb Cortex* **7**:487–501.

Runyan CA, Schummers J, Van Wart A, Kuhlman SJ, Wilson NR, Huang ZJ, Sur M. 2010. Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron* **67**:847–857.

Shu Y, Hasenstaub A, McCormick DA. 2003. Turning on and off recurrent balanced cortical activity. *Nature*. doi:10.1038/nature01616

Sohal VS, Zhang F, Yizhar O, Deisseroth K. 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459**:698–702.

Stark E, Eichler R, Roux L, Fujisawa S, Rotstein HG, Buzsáki G. 2013. Inhibition-induced theta resonance in cortical circuits. *Neuron* **80**:1263–1276.

Stark E, Roux L, Eichler R, Senzai Y, Royer S, Buzsáki G. 2014. Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron* **83**:467–480.

van Vreeswijk C, Sompolinsky H. 1998. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*. doi:10.1162/089976698300017214

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

Verret L, Mann EO, Hang GB, Barth AMI, Cobos I, Ho K, Devidze N, Masliah E, Kreitzer AC, Mody I, Mucke L, Palop JJ. 2012. Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* **149**:708–721.

Wichmann T, DeLong MR. 1996. Functional and pathophysiological models of the basal ganglia. *Curr Opin Neurobiol* **6**:751–758.

Wilson NR, Runyan CA, Wang FL, Sur M. 2012. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature* **488**:343–348.