

## 1 Exploratory analysis of V1 with EPI produces a novel theory

Dynamical models of excitatory (E) and inhibitory (I) populations with supralinear input-output function have succeeded in explaining a host of experimentally documented phenomena. In a regime characterized by inhibitory stabilization of strong recurrent excitation, these models give rise to paradoxical responses [1], selective amplification [2, 3], surround suppression [4] and normalization [5]. Despite their strong predictive power, E-I circuit models rely on the assumption that inhibition can be studied as an indivisible unit. However, experimental evidence shows that inhibition is composed of distinct elements – parvalbumin (P), somatostatin (S), VIP (V) – composing 80% of GABAergic interneurons in V1 [6, 7, 8], and that these inhibitory cell types follow specific connectivity patterns (Fig. 1A) [9]. Recent theoretical advances [10, 11, 12], have only started to address the consequences of this multiplicity in the dynamics of V1, strongly relying on linear theoretical tools. Here, we use EPI to elucidate the mechanisms of neuron-type stability in V1 at different levels of contrast.

We considered the contrast responses of a nonlinear dynamical V1 circuit model (Fig. 1A) with a state comprised of each neuron-type population’s rate  $\mathbf{x} = [x_E, x_P, x_S, x_V]^\top$ . Each population receives recurrent input  $W\mathbf{x}$  from synaptic projections of effective connectivity  $W$  and an external input  $\mathbf{h}$ , which determine the population rate via nonlinearity  $f_r = \|\cdot\|_+^2$  (see Section 2). The deterministic circuit model evolves from an initial condition  $\mathbf{x}(0)$  to a steady-state solution (Fig. 1B, solid). When slow noise  $\epsilon$  is introduced, circuit activity fluctuates around this solution (Fig. 1B dashed), and the model is then a stochastic stabilized supralinear network (SSSN) [13]

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + f_r(W\mathbf{x} + \mathbf{h}) \quad (1)$$

At full contrast, input to the E- and P-populations increases

$$\mathbf{h} = \mathbf{h}_b + c\mathbf{h}_c, \quad (2)$$

changing the steady state solution (and fluctuations thereabout) (Fig. 1C). In this analysis, we fixed  $W, \mathbf{h}_b$ , and  $\mathbf{h}_c$ , which have been fit to contrast responses in mouse V1 using the deterministic model (TODO cite, see Section 2). As contrast changes, so does the degree of stochastic variability of each neuron-type (Fig. 1D). Moreover, the circuit transitions into an inhibition-stabilized network (ISN) ( $\gamma(\mathbf{x}; \mathbf{z}) < 0$ ) with increasing contrast (Fig. 1D inset). While the ISN-regime and its effects are understood well in E-I networks, its role in this SSSN with inhibitory multiplicity has not been explored.

To study the influence of circuit input on stabilization mechanisms in this stochastic model, we asked what changes in circuit input  $d\mathbf{h}$  have little effect on the responses of each neuron-type.

$$\mathbf{z} = [dh_E \ dh_P \ dh_S \ dh_V]^\top. \quad (3)$$

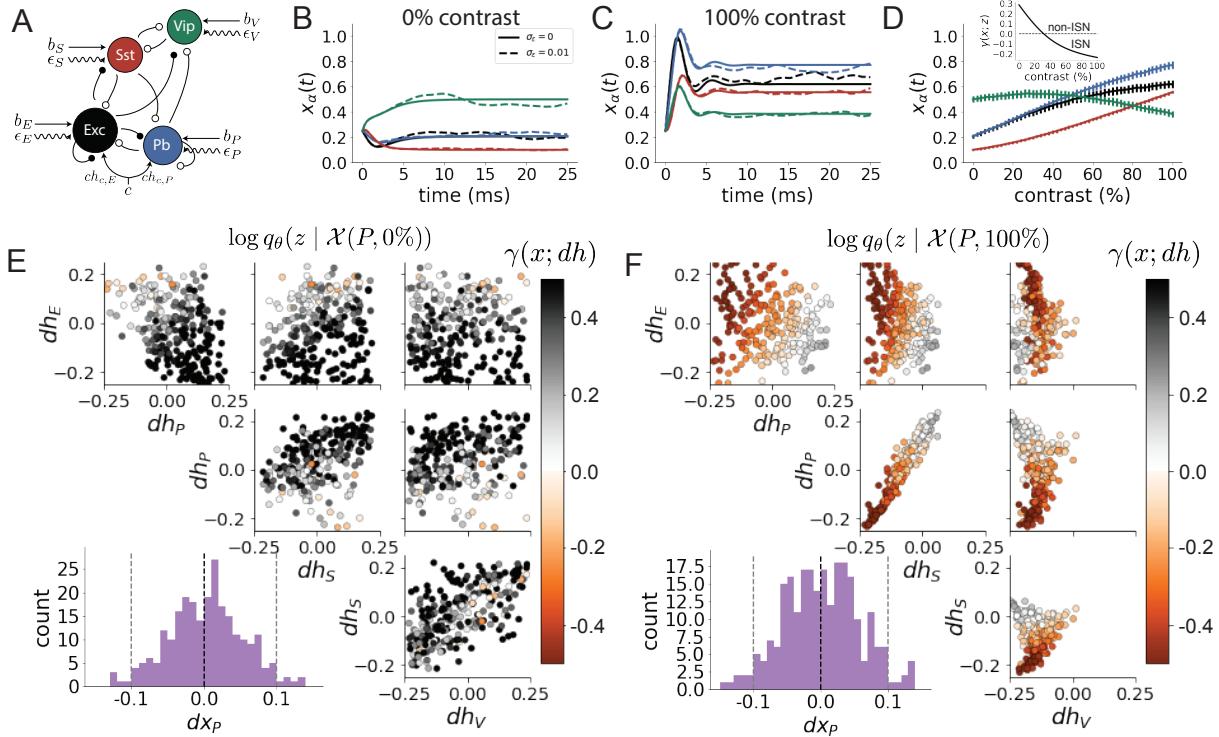
To formulate “fluctuation constraints” for population  $\alpha$  at contrast level  $c$  as an emergent property in this stochastic model, we prescribed that the mean change in activity with  $dh_\alpha$  is 0 and the variance over both EPI posterior and model stochasticity is a small amount

$$\begin{aligned} \mathcal{X}(\alpha, c) : \mathbb{E}_{\mathbf{z}} [dx_\alpha(\mathbf{x}; \mathbf{z}, c)] &= 0 \triangleq \mu \\ \text{Var}_{\mathbf{z}} [dx_\alpha(\mathbf{x}; \mathbf{z}, c)] &= [0.05^2] \triangleq \sigma^2. \end{aligned} \quad (4)$$

By conditioning on this emergent property, we constrained stochastic fluctuations of circuit models sampled from the EPI posterior to be unbiased and vary to a fixed extent from the steady state solution at  $d\mathbf{h} = \mathbf{0}$ .

Visualization of the resulting  $q_{\theta}(\mathbf{z} | \mathcal{X}(\alpha, c))$  shows that the EPI posteriors largely agree with sets of parameters obtained using ABC (Fig. 2). Focusing our attention on  $P$ -fluctuation constraints at 0% and 100% contrast ( $\mathcal{X}(P, 0\%)$  and  $\mathcal{X}(P, 100\%)$ , respectively), we gain insight about how  $d\mathbf{h}$  governs mechanisms of stability. At 0% contrast, the SSSN is not inhibition-stabilized, and only local perturbations of positive  $dh_E$  and negative  $dh_P$  place the circuit in the ISN regime, while  $P$ -population fluctuations are kept near the 0% contrast level (Fig. 1E). At 100% contrast, where  $h_E$  and  $h_P$  are much greater, the SSSN is inhibition-stabilized, and  $q_{\theta}(z | \mathcal{X}(P, 100\%))$  reveals distinct patterns of  $d\mathbf{h}$  that obey the  $P$ -fluctuation constraint (Fig. 1F). As  $dh_V$  decreases, either  $dh_P$  and  $dh_S$  must increase or decrease together. When  $dh_P$  and  $dh_S$  increase together, the circuit leaves the ISN regime, and when they are decreased, the strength of inhibition stabilization grows. The pattern in  $d\mathbf{h}$  producing transitions out of the ISN regime is unapparent when running traditional ABC, even when SSSN noise is reduced to a negligible order of magnitude (Fig. 3).

Figure 1: **A.** Four-population model of primary visual cortex with excitatory (black), parvalbumin (blue), somatostatin (red), and VIP (green) neurons. Some neuron-types largely do not form synaptic projections to others (excitatory and inhibitory projections filled and unfilled, respectively). **B.** V1 circuit model simulations at 0% contrast without (SSN, solid) and with (SSSN, solid) slow noise. **C.** Simulations at 100% contrast. **D.** Average responses of the SSSN for all contrasts. Error bars show the standard deviation of stochastic fluctuations. **E.** EPI posterior conditioned on P-fluctuation constraints at 0% contrast. Parameter samples are colored by their ISN coefficient. Bottom-left inset is the posterior predictive distribution of  $dx_P$ . **F.** Same as E. at 100% contrast.



## 2 V1 Supplemental section

The dynamics are driven by the rectified and squared sum of recurrent inputs  $W\mathbf{x}$ , external inputs  $\mathbf{h}$ , and external noise:

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + f_r(W\mathbf{x} + \mathbf{h} + \epsilon) \quad (5)$$

where the input is comprised of a baseline and contrast-dependent input

$$\mathbf{h} = \mathbf{h}_b + c\mathbf{h}_c \quad (6)$$

$\tau = 1\text{ms}$ ,  $f_r(\mathbf{x}) = [\mathbf{x}]_+^2$ , and  $c \in [0, 1]$ .

The noise is modeled as an Ornstein-Uhlenbeck process:

$$\tau_{\text{noise}} d\epsilon_\alpha = -\epsilon_\alpha dt + \sqrt{2\tau_{\text{noise}}} \sigma_\alpha dB \quad (7)$$

where  $\tau_{\text{noise}}$  is slow at 5ms,  $\sigma_\alpha = 0.01$ , and  $dB$  is a standard Wiener process.

We considered the stability of this stochastic stabilized supralinear network with respect to an additive differential input  $d\mathbf{h}$

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + f_r(W\mathbf{x} + \mathbf{h} + d\mathbf{h} + \epsilon). \quad (8)$$

We used connectivity ( $W$ ) and input ( $\mathbf{h}_b$  and  $\mathbf{h}_c$ ) parameters fit to the deterministic model (TODO cite)

$$W = \begin{bmatrix} W_{EE} & W_{EP} & W_{ES} & W_{EV} \\ W_{PE} & W_{PP} & W_{PS} & W_{PV} \\ W_{SE} & W_{SP} & W_{SS} & W_{SV} \\ W_{VE} & W_{VP} & W_{VS} & W_{VV} \end{bmatrix} = \begin{bmatrix} .785 & -.102 & -1.24 & -.306 \\ .809 & -.100 & -.620 & -.264 \\ .833 & -.000128 & -.0000483 & -.742 \\ .708 & -.306 & -.452 & -.0000605 \end{bmatrix}, \quad (9)$$

$$\mathbf{h}_b = \begin{bmatrix} h_{b,E} \\ h_{b,P} \\ h_{b,S} \\ h_{b,V} \end{bmatrix} = \begin{bmatrix} .590 \\ .504 \\ .515 \\ .670 \end{bmatrix}, \quad (10)$$

and

$$\mathbf{h}_c = \begin{bmatrix} h_{c,E} \\ h_{c,P} \\ h_{c,S} \\ h_{c,V} \end{bmatrix} = \begin{bmatrix} .595 \\ .396 \\ 0 \\ 0 \end{bmatrix}. \quad (11)$$

The expectation and variance in the emergent property are taken over both the posterior distribution and the stochasticity of the SSSN:

$$= \mathbb{E}_{\mathbf{z}} [dx_\alpha(\mathbf{x}; \mathbf{z}, c)] = \mathbb{E}_{\mathbf{z} \sim q_\theta(\mathbf{z})} [\mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} [x_\alpha(\mathbf{x}; \mathbf{z}, c) - x_\alpha(\mathbf{x}; \mathbf{0}, c)]] . \quad (12)$$

Thus, over the EPI posterior predictive distribution, the mean and variance of the slow fluctuations about the steady state solution are constrained to  $\mu = 0$  and  $\sigma^2 = 0.05^2$

This ISN coefficient of the network at a given input is calculated by simulating the system to a steady state in the absence of noise

$$\gamma(\mathbf{x}; \mathbf{z}) = 1 - 2 [u_E]_+ W_{EE} \quad (13)$$

where

$$\mathbf{u} = W\mathbf{x} + \mathbf{h} + d\mathbf{h}. \quad (14)$$

When the ISN coefficient is negative, the network is inhibition-stabilized. To get an idea of what distribution of parameters ( $d\mathbf{h}$ ) we should expect from EPI, we can use ABC to obtain a set of parameters related to the emergent property. We compare EPI to ABC with a rejection heuristic defined by the standard deviation of the differential responses  $\sigma_{ABC}$

$$f_{ABC}(dx_\alpha; \sigma_{ABC}) = |dx_\alpha| > 2\sigma_{ABC}. \quad (15)$$

In other words, we ran ABC accepting parameters that generate differential responses within two standard deviations  $\sigma_{ABC} = 0.05$  of  $dx_\alpha = 0$ .

## References

- [1] Misha V Tsodyks, William E Skaggs, Terrence J Sejnowski, and Bruce L McNaughton. Paradoxical effects of external modulation of inhibitory interneurons. *Journal of neuroscience*, 17(11):4382–4388, 1997.
- [2] Mark S Goldman. Memory without feedback in a neural network. *Neuron*, 61(4):621–634, 2009.
- [3] Brendan K Murphy and Kenneth D Miller. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.
- [4] Hirofumi Ozeki, Ian M Finn, Evan S Schaffer, Kenneth D Miller, and David Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009.
- [5] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015.
- [6] Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*, 5(10):793, 2004.
- [7] Bernardo Rudy, Gordon Fishell, SooHyun Lee, and Jens Hjerling-Leffler. Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neurobiology*, 71(1):45–61, 2011.
- [8] Robin Tremblay, Soohyun Lee, and Bernardo Rudy. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*, 91(2):260–292, 2016.
- [9] Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068, 2013.

- [10] Ashok Litwin-Kumar, Robert Rosenbaum, and Brent Doiron. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *Journal of neurophysiology*, 115(3):1399–1409, 2016.
- [11] Luis Carlos Garcia Del Molino, Guangyu Robert Yang, Jorge F. Mejias, and Xiao Jing Wang. Paradoxical response reversal of top- down modulation in cortical circuits with three interneuron types. *Elife*, 6:1–15, 2017.
- [12] Guang Chen, Carl Van Vreeswijk, David Hansel, and David Hansel. Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. 2019.
- [13] Guillaume Hennequin, Yashar Ahmadian, Daniel B Rubin, Máté Lengyel, and Kenneth D Miller. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860, 2018.
- [14] (2018) Allen Institute for Brain Science. Layer 4 model of v1. available from: <https://portal.brain-map.org/explore/models/l4-mv1>.
- [15] Yazan N Billeh, Binghuang Cai, Sergey L Gratiy, Kael Dai, Ramakrishnan Iyer, Nathan W Gouwens, Reza Abbasi-Asl, Xiaoxuan Jia, Joshua H Siegle, Shawn R Olsen, et al. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *bioRxiv*, page 662189, 2019.

Figure 2: Supplemental Figure: A. Top row: EPI posteriors  $q_\theta(z | \mathcal{X}(\alpha, c))$  for neuron-type  $\alpha$ , 0% contrast,  $\epsilon = 0.01$  visualized as samples colored by their log probability. Bottom-left of pairplots indicate that the posterior predictive distribution obeys the emergent property constraints. Bottom row: Sets of parameters produced using the ABC approach. Bottom-left of pairplots shows statistics of accepted parameters (black) and the ABC posterior predictive distribution given that set (purple). B. Same as A at 100% contrast.

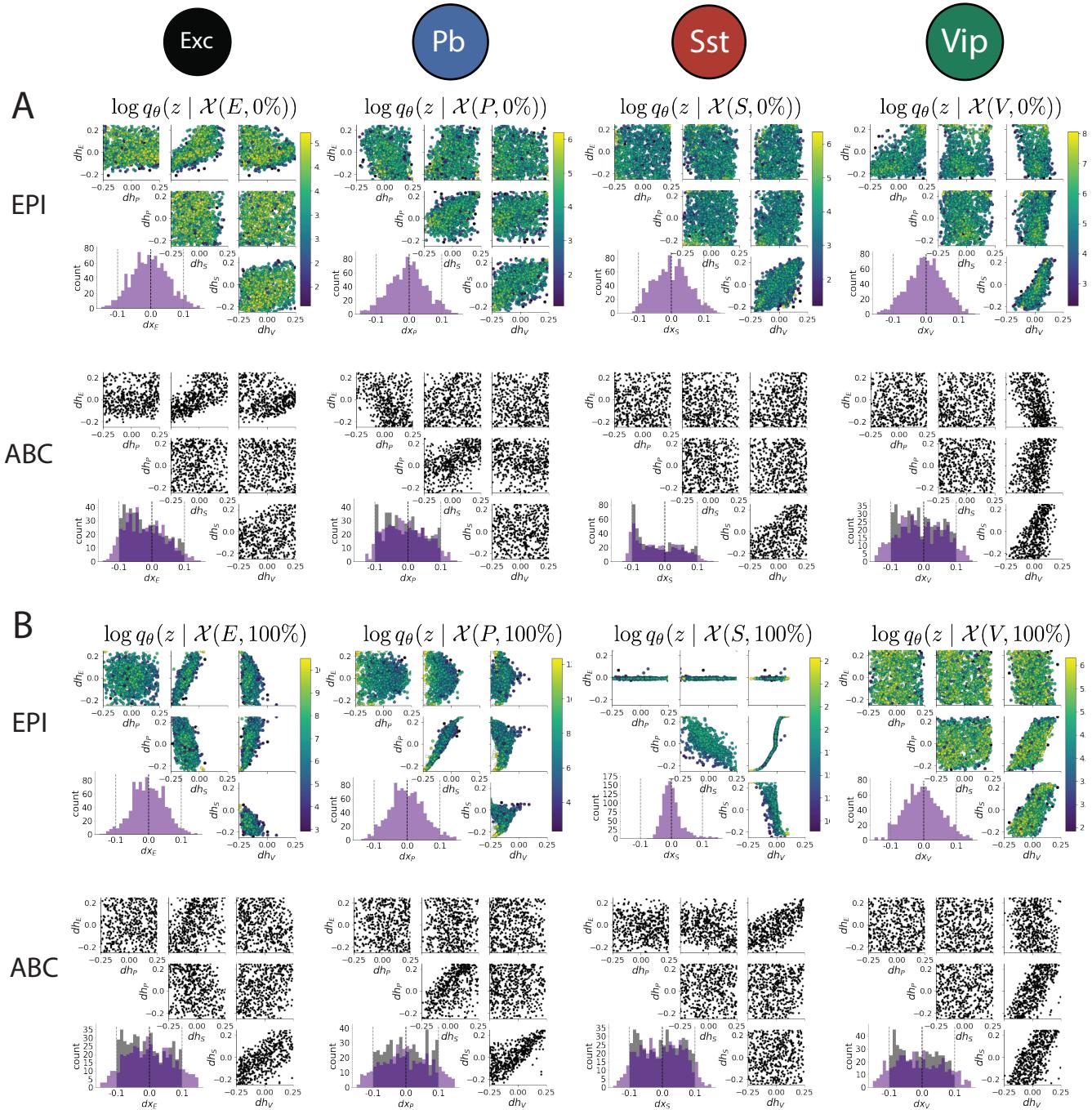


Figure 3: Supplemental Figure: Same as Figure 2 (will be supp fig) with  $\sigma_\epsilon = 0.001$ .