Draft of new V1 section

Sean Bittner, Agostina Palmigiano

November 24, 2020

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. **??**A) [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 consider contrast responses of a nonlinear dynamical V1 circuit model (Fig. 1A) with a state comprised of each neuron-type population's rate $x = [x_E, x_P, x_S, x_V]^\top$. Each population of neuron-type α receives recurrent input $(W f_r(x))_\alpha$ from synaptic projections, where W is the connectivity matrix and $f_r = [\cdot]_+^2$ is the rate nonlinearity. Driven by a baseline input $\mathbf{b}$, the circuit evolves form an initial condition $\mathbf{x}(0)$ to a steady-state solution (Fig. 1B, solid). When slow noise ϵ is introduced, circuit activity fluctuates around this solution (Fig. 1B dashed), and the model is then a stochastic stabilized supralinear network (SSSN) [13] (see Section 2). As contrast is enhanced, input to the E- and P-populations via $h^{(c)}$ increases causing the steady state solution (and fluctuations thereabout) to change (Fig. 1C). In this analysis, we consider W,b, and $h^{(c)}$ that 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) with increasing contrast (Fig. 1D inset). While the ISN-regime and it's effects are well understood in E-I networks, it's role in the stability of this SSSN with inhibitory multiplicity has not been explored.

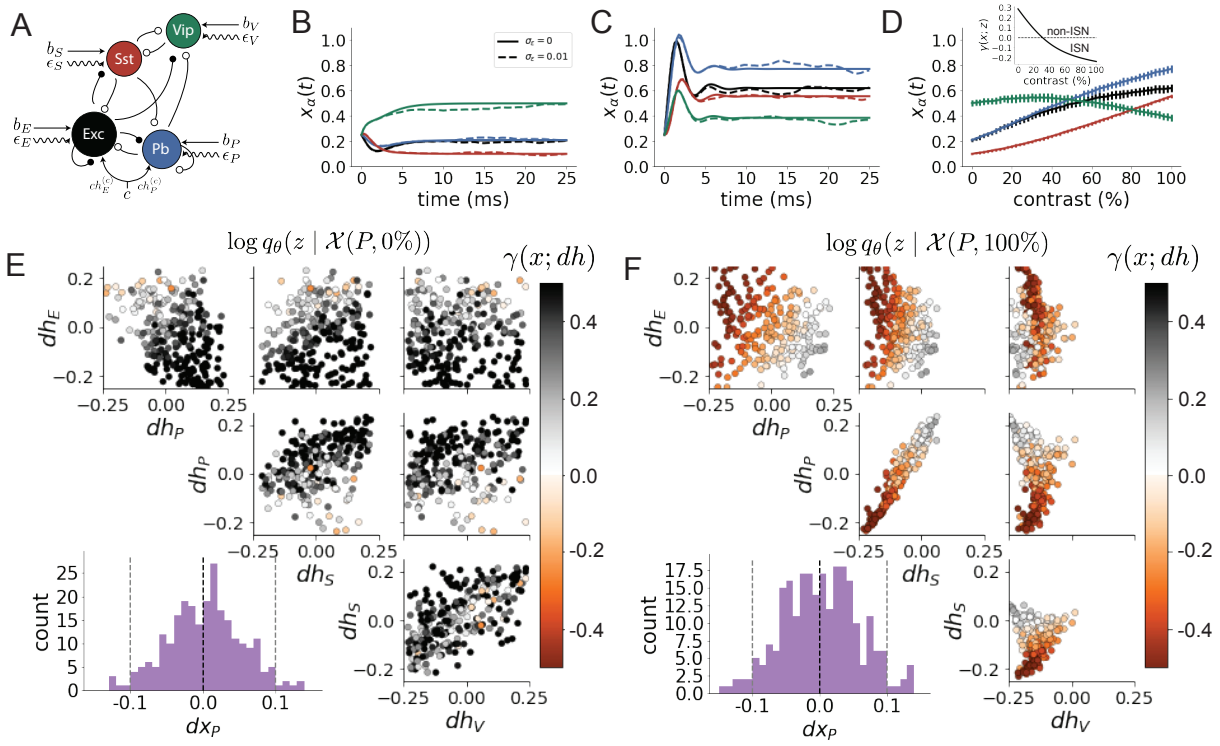
To study the role of inhibition stabilization in this network, we explore the changes in input

$$\mathbf{z} = \begin{bmatrix} dh_E & dh_P & dh_S & dh_V \end{bmatrix}^\top \tag{1}$$

resulting in each neuron-type's stability around it's mean rate. The emergent property of "neuron-type stability" for population α at contrast level c is defined as

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

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**. EPI posterior $q_\theta(z \mid \mathcal{B}_{S-V})$ for S-V flipping. The obtained posterior is visualized as 500 samples from the inferred distribution colored by $\log(q_\theta(z))$. This posterior is bimodal and concentrated in planes $h^{(c)} > 0$ and $h^{(c)} < 0$ at distinct modes z_1 (black star) and z_2 (gray, star), respectively. Bottom-left: Posterior predictive distribution of the emergent property statistics with respect to the constrained means (black, dashed line) and variances (gray, dashed lines at two standard deviations). **C**. Posterior predictive distribution of $d_{S,V}$ for each mode. The z_1-mode produces V-to-S flipping with increasing contrast. **D**. Model simulations at the mode z_1 at $c = 0$ (solid) and $c = 1$ (dashed). **E**. Mean system response at varying contrasts for z_1. Error bars show noise in rate. **F**. Plot of noise with contrast from E. Error bars are standard error measured across simulations.



2 V1 Supplemental section

The dynamics are driven by the rectified and exponentiated sum of recurrent inputs Wx, external inputs h, and external noise:

$$\tau \frac{dx}{dt} = -x + W f_r(x) + b + h + \epsilon \tag{3}$$

where $\tau = 1$ms, and $f_r(x) = [x]_+^2$.

The noise is modeled as an Ornstein-Uhlenbeck process:

$$\tau_{\text{noise}} d\epsilon_\alpha = -\epsilon_\alpha dt + \sqrt{2\tau_{\text{noise}}}\sigma_\alpha dB \tag{4}$$

where τ_{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 using connectivity (W) and input (b and $h^{(c)}$) parameters fit to the deterministic model (TODO cite)

$$W = \begin{bmatrix} W_{EE} & W_{EP} & W_{ES} & W_{EV} \\ W_{PE} & W_{PP} & W_{PS} & W_{PV} \\ W_{SE} & W_{SP} & W_{SS} & W_{SV} \\ W_{VE} & W_{VP} & W_{VS} & W_{VV} \end{bmatrix} = \begin{bmatrix} .785 & -.102 & -1.24 & -.306 \\ .809 & -.100 & -.620 & -.264 \\ .833 & -.000128 & -.0000483 & -.742 \\ .708 & -.306 & -.452 & -.0000605 \end{bmatrix}, \tag{5}$$

$$b = \begin{bmatrix} b_E \\ b_P \\ b_S \\ b_V \end{bmatrix} = \begin{bmatrix} .590 \\ .504 \\ .515 \\ .670 \end{bmatrix}, \tag{6}$$

and

$$h^{(c)} = \begin{bmatrix} h_E^{(c)} \\ h_P^{(c)} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} .595 \\ .396 \\ 0 \\ 0 \end{bmatrix}. \tag{7}$$

This ISN coefficient $\gamma(x; z)$ is calculated by simulating

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

$$= \mathbb{E}_{\mathbf{z}} \left[dx_\alpha(\mathbf{x}; \mathbf{z}, c) \right] = \mathbb{E}_{\mathbf{z} \sim q_\theta} \left[\mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})} \left[x_\alpha(\mathbf{x}; \mathbf{z}, c) - x_\alpha(\mathbf{x}; \mathbf{0}, c) \right] \right]. \tag{8}$$

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.

With this model, we are interested in the differential responses of each neuron-type population to changes in input dh. Initially, we studied the linearized response of the system to input $\frac{dx_{ss}}{dh}$ at the steady state response x_{ss}, i.e. a fixed point. All analyses of this model consider the steady state response, so we drop the notation ss from here on. While this linearization accurately predicts

differential responses $dx = [dx_E, dx_P, dx_S, dx_V]^\top$ for small differential inputs to each population $dh = [0.1, 0.1, 0.1, 0.1]^\top$ (Fig **??**B left), the linearization is a poor predictor in this nonlinear model more generally (Fig. **??**B right). Currently available approaches to deriving the steady state response of the system are limited.

The input $h = b + dh$ is comprised of a baseline input $b = [b_E, b_P, b_S, b_V]^\top = [1, 1, 1, 1.25]^\top$ and a differential input $dh = [dh_E, dh_P, dh_S, dh_V]^\top$ to each neuron-type population.

We want to know the differential inputs dh that maintain the steady state x_α for $\alpha \in \{E, P, S, V\}$. We see from Figure 1B that input to a single population in the recurrent circuit elicits a variety of responses across populations: E same, P up, S down, and V up. We define the differential steady state dx_α as the change in steady state x_α when receiving input $h = b + dh$ with respect to the baseline $h = b$. Maintaining the steady state of a neuron-type population amounts to the emergent property

$$\mathcal{B}(\alpha, \sigma) \triangleq \mathbb{E}\begin{bmatrix} dx_\alpha \\ dx_\alpha^2 \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}. \tag{9}$$

In the following analyses, we chose $\sigma = 0.25$.

3 EPI agrees with ABC

To get an idea of what distribution of parameters (dh) 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 σ_{ABC}

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

In other words, we ran ABC accepting parameters that generate differential responses within two standard deviations $\sigma_{ABC} = 0.25$ of $dx_\alpha = 0$. In Figure 2, we see that the distributions obtained via EPI (colored by $\log q_\theta(z)$) are visually similar to those obtained via ABC (colored by dx_α).

Figure 2: EPI (left) vs ABC (right). Arrows in EPI distribution indicate dimensions of maximal sensitivity at selected parameters dh. The importance of v_S and v_P are explained in section 4.2.

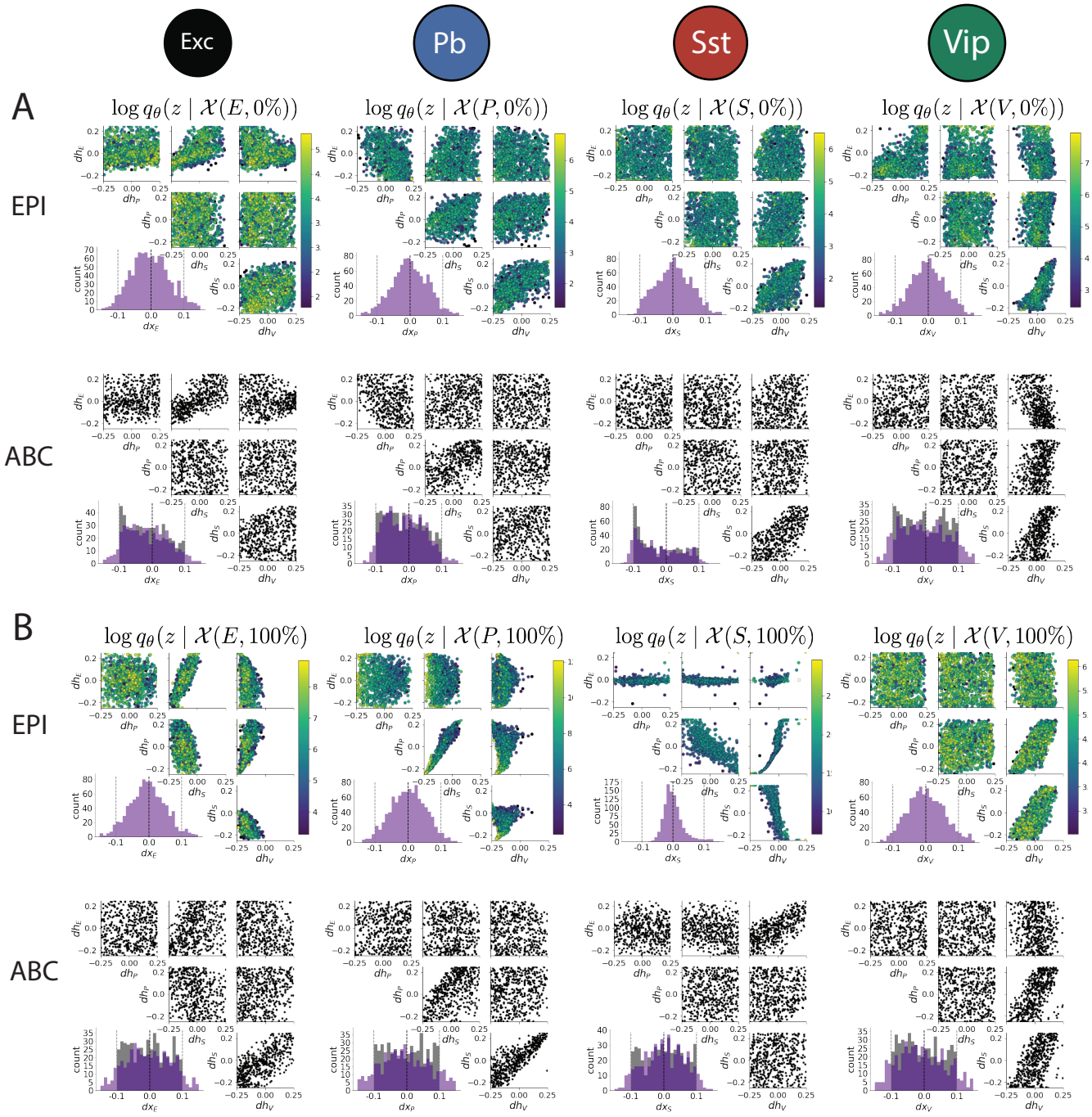
There are some subtle differences between the EPI and ABC distributions, but some difference is to be expected. Rather than simply attributing these differences to imperfection of the EPI optimization routine, we consider the the effect of bias in dx_α from ABC (and lack thereof from EPI).

When $\epsilon > 0$ in ABC, the samples are *not* from the posterior distribution. For $\epsilon > 0$, the "posterior" predictive means of ABC $\mathbb{E}_{ABC(\alpha,\sigma_{ABC})}[dx_\alpha]$ may be far from zero. In Figure 3, we see that with ABC, there is an increasingly negative bias in $\mathbb{E}_{ABC(\alpha,\sigma_{ABC})}[dx_\alpha]$ for greater error tolerances across all neuron-types. Additionally with ABC, there is no precise control of the variance $\mathbb{E}_{ABC(\alpha,\sigma_{ABC})}\left[dx_\alpha^2\right]$, which may be undesireable.

In contrast, the first and second moments of dx_α are controlled to a specified degree of accuracy (Figure 4). The variances of each distribution of dx_α are close to their target value $0.25^2 = 0.0625$ (see variances next to histograms of Figure 4).

Sean Bittner
srb2201@columbia.edu

Figure 2: SX 1



Sean Bittner
srb2201@columbia.edu

Figure 3: SX 2