Draft of new V1 section

Sean Bittner, Agostina Palmigiano

October 6, 2020

1 EPI clarifies the implications of contrast-response flipping on [noise quenching, inhibition stabilization] in V1

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], surround suppression [3] and normalization [4]. 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 [5, 6, 7], and that these inhibitory cell types follow specific connectivity patterns (Fig. 1A) [8]. Recent theoretical advances [9, 10, 11], 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 analyze the posteriors of a stochastic nonlinear dynamical model of V1 conditioned on the emergent property of contrast-dependent S-V flipping. We then [use this info to make a statement about ...]

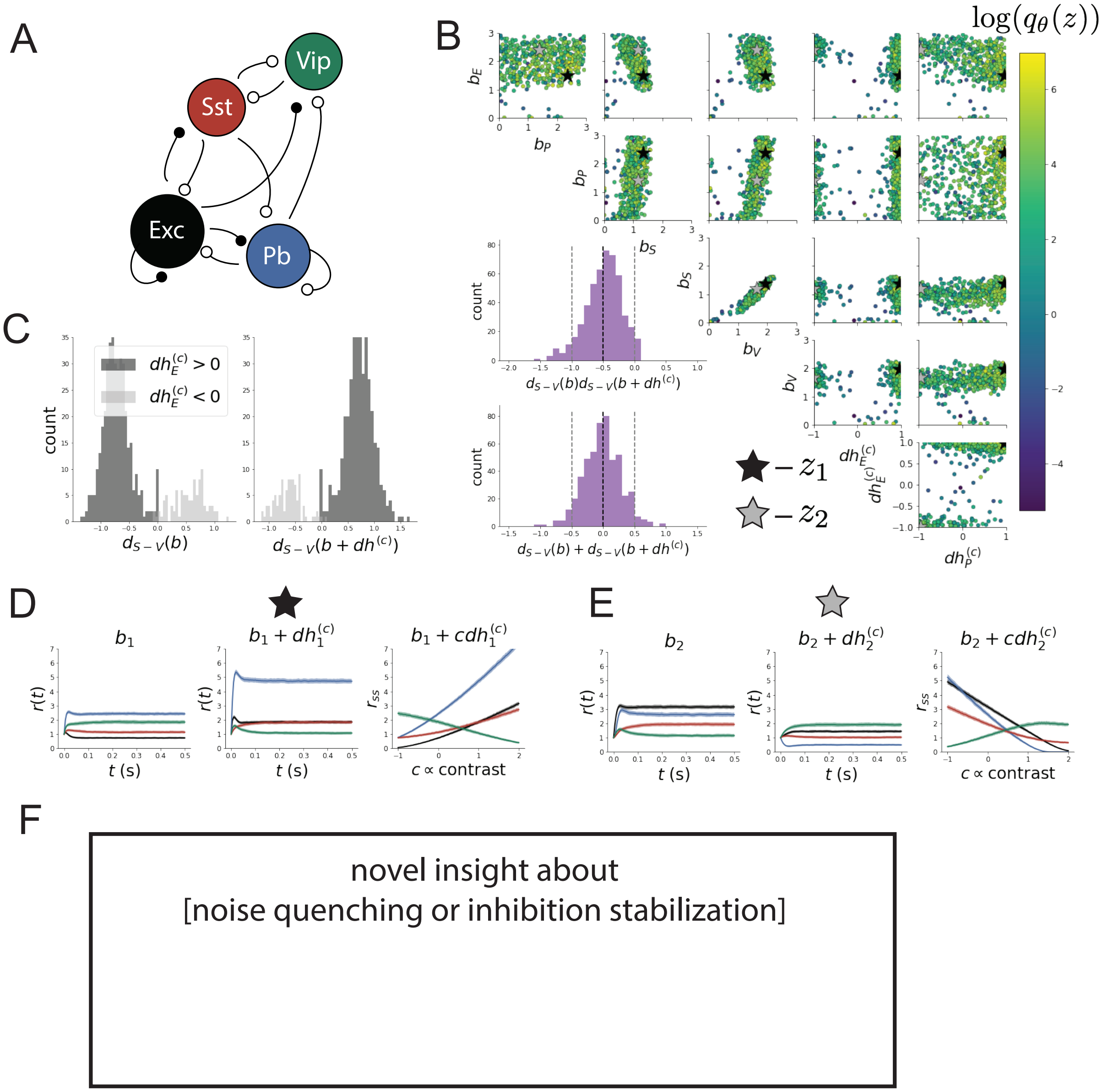
Specifically, we consider a four-dimensional circuit model with dynamical state given by the firing rate x of each neuron-type population $x = [x_E, x_P, x_S, x_V]^\top$. Given a time constant of $\tau = 20$ ms and a power $n = 2$, the dynamics are driven by the rectified and sum of recurrent inputs Wx, external inputs h, and external noise $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$:

$$\tau \frac{dx}{dt} = -x + [Wx + h + \epsilon]_+^n. \tag{1}$$

We considered fixed effective connectivity weights W approximated from experimental recordings of publicly available datasets of mouse V1 [12, 13] (see Section 2). The input $h = b + dh$ is comprised of a baseline input $b = [b_E, b_P, b_S, b_V]^\top$ and a differential input $dh = [dh_E, dh_P, dh_S, dh_V]^\top$ to each neuron-type population. Throughout subsequent analyses, the baseline input is $b = [1, 1, 1, 1]^\top$.

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.

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 $dh_E > 0$ and $dh_E < 0$ 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 stds). **C**. Posterior predictive distribution of $d_{S-V}(h)$ of each mode shows that the z_1-mode produces V-to-S flipping with increasing contrast and the z_2-mode produces S-to-V flipping. **D**. Model simulations at the mode z_1 at b_1 (left), $b_1 + dh_1^{(c)}$ (middle), and steady state solutions for varying levels of contrast (right). Shaded area is one standard deviation according to randomness of ϵ. **E**. Same as D. for z_2. **F**. ...



TODO Summarize S-V flipping phenomena and why it's a good EP TODO Explain that we can't really fit the whole data-set of responses yet because the model is quite constrained. TODO Motivate this input model, where $dh^{(c)}$ is the direction of increasing contrast.

$$\begin{bmatrix} h_E \\ h_P \\ h_S \\ h_V \end{bmatrix} = \begin{bmatrix} b_E \\ b_P \\ b_S \\ b_V \end{bmatrix} + \begin{bmatrix} dh_E^{(c)} \\ dh_P^{(c)} \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

Since we are unsure what either the baseline input b and contrast-dependent change in input $dh^{(c)}$ should be, we treat them as free parameters when running EPI.

$$z = \begin{bmatrix} b_E & b_P & b_S & b_V & dh_E^{(c)} & dh_P^{(c)} \end{bmatrix}^\top \tag{3}$$

We consider positive baseline inputs $b_\alpha \in [0, 3]$ and small contrast-dependent changes in input $|dh_\alpha| \leq 1$. To find parameters resulting in S-V flipping, we focus on models driven by nominal amounts of external noise $\sigma_\eta = 0.1$.

We formulate the emergent property of S-V flipping with a statistic measuring he difference in steady state (x_ss) between the S- and V-population at a given input:

$$d_{S-V}(h) = x_{ss,S}(h) - x_{ss,V}(h). \tag{4}$$

For S and V to flip their steady states, the difference between the S- and V-population rates at $h = b$ must have opposite sign from $h = b + dh$. Therefore, we stipulate the emergent property of S-V flipping to require the product between $d_{S-V}(b)$ and $d_{S-V}(b+dh)$ to be appreciably negative. Second we stipulate that the differences between S and V in each input condition cancel out on average ($d_{S-V}(b) + d_{S-V}(b + dh)$ is 0 on average). The means and variances of the emergent property statistics were sensibly chosen based on some inexpensive model simulations within the parameter bounds.

$$\mathcal{B}_{S-V} \triangleq \mathbb{E} \begin{bmatrix} d_{S-V}(b)d_{S-V}(b+dh) \\ (d_{S-V}(b)d_{S-V}(b+dh) - (-0.25))^2 \\ d_{S-V}(b) + d_{S-V}(b+dh) \\ (d_{S-V}(b) + d_{S-V}(b+dh))^2 \end{bmatrix} = \begin{bmatrix} -0.25 \\ 0.125^2 \\ 0 \\ 0.125^2 \end{bmatrix} \tag{5}$$

We ran EPI to inspect the structure of the posterior distribution of z conditioned on S-V flipping shown in Fig 2. It is clear from this visualization that S-V flipping in this V1-model is sensitive with respect to some parameter settings and robust with respect to others. Additionally, the posterior is bimodal: there is one mode in each of the hyperplanes $dh_E < 0$ and $dh_E > 0$.

The structure of the V1 model S-V flipping posterior yields the following insights:

1. The marginal distributions of $q_\theta(h_P \mid \mathcal{B}_{S-V})$ and $q_\theta(dh_P \mid \mathcal{B}_{S-V})$ are approximately uniform along their allowed range. The approximate uniformity of the $P - population$ parameter marginal distributions shows that the P-population plays little role in S-V flipping.

2. The pairwise marginal distribution of $q_\theta(h_S, h_V \mid \mathcal{B}_{S-V})$ shows strong correlation between these parameters. The strong correlation between h_S and h_V in the posterior reveals that S-V flipping is sensitive with respect to the baseline inputs to the S- and V-populations. **Augment this to show that multidim info from Hessian is useful.** The Hessian

provided by EPI at the modes indicates that coordinated increases of h_S and h_V by a ratio of $\frac{h_S}{h_V} = 0.701$ will preserve S-V flipping, while changes in an orthogonal dimension will disrupt it.

3. This distribution is multimodal.

2 V1 Supplemental section

We considered fixed effective connectivity weights W approximated from experimental recordings of publicly available datasets of mouse V1. Specifically, Billeh et al. [13] produce estimates of the synaptic strength (in mV)

$$M = \begin{bmatrix} 0.36 & -0.48 & -0.31 & -0.28 \\ 1.49 & -0.68 & -0.50 & -0.18 \\ 0.86 & -0.42 & -0.15 & -0.32 \\ 1.31 & -0.41 & -0.52 & -0.37 \end{bmatrix} \tag{6}$$

and connection probability

$$C = \begin{bmatrix} 0.16 & 0.411 & 0.424 & 0.087 \\ 0.395 & .451 & 0.857 & 0.02 \\ 0.182 & 0.03 & 0.082 & 0.625 \\ 0.105 & 0.22 & 0.77 & 0.028 \end{bmatrix}. \tag{7}$$

Multiplying these connection probabilities and synaptic efficacies gives us an effective connectivity matrix:

$$W = C \odot M = \begin{bmatrix} 0.0576 & -0.197 & -0.131 & -0.0244 \\ 0.589 & -0.307 & -0.429 & -0.00360 \\ 0.157 & -0.0126 & -0.0123 & -0.200 \\ 0.138 & -0.0902 & -0.400 & -0.0104 \end{bmatrix}. \tag{8}$$

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.

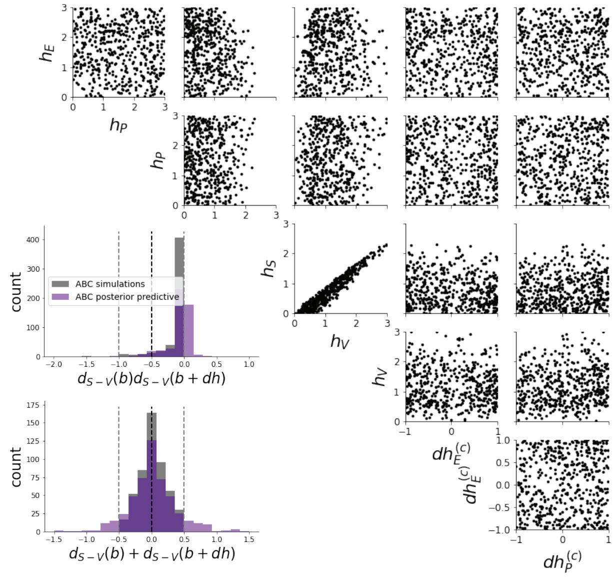
[2] Brendan K Murphy and Kenneth D Miller. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.

[3] 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.

[4] 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.

[5] 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.

Figure 2: **A**. ...



[6] 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.

[7] Robin Tremblay, Soohyun Lee, and Bernardo Rudy. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*, 91(2):260–292, 2016.

[8] 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.

[9] 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.

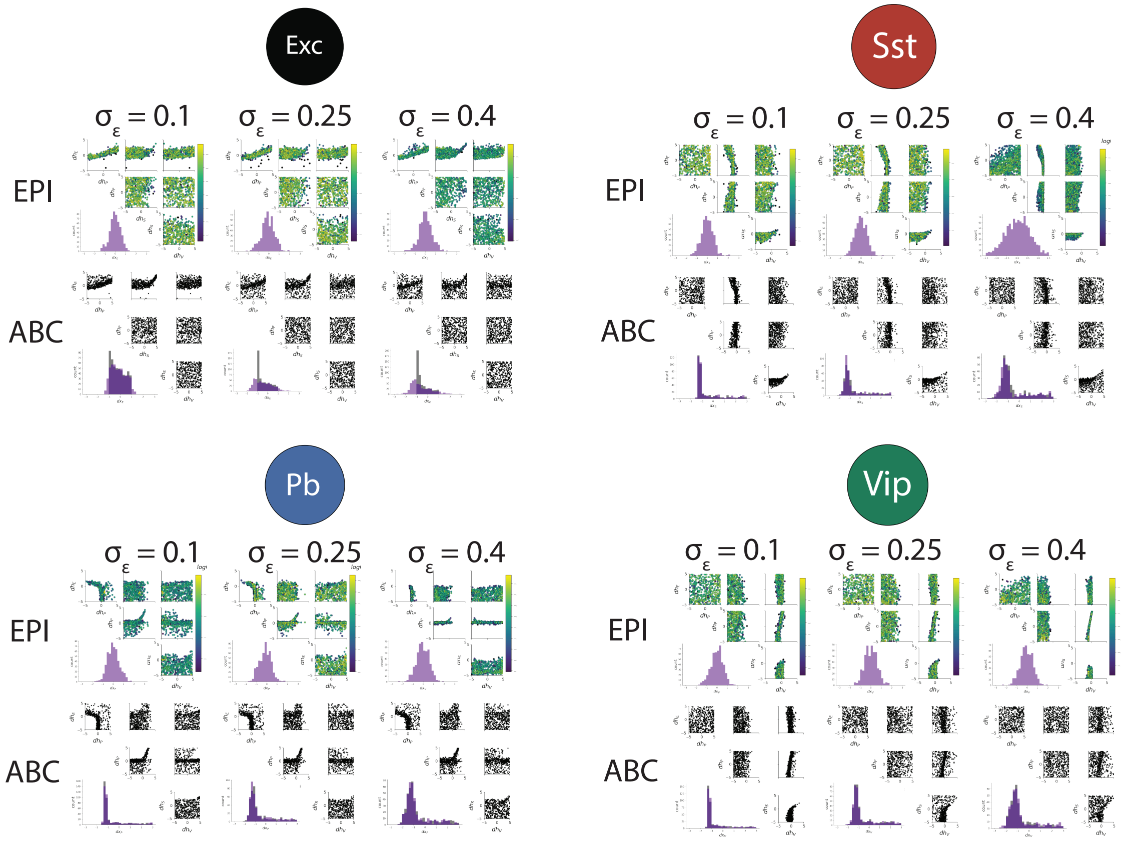
[10] Luis Carlos Garcia Del Molino, Guangyu Robert Yang, Jorge F. Mejias, and Xiao Jing Wang. Paradoxical response reversal of top- down modulation in cortical circuits with three interneuron types. *Elife*, 6:1–15, 2017.

[11] Guang Chen, Carl Van Vreeswijk, David Hansel, and David Hansel. Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. 2019.

[12] (2018) Allen Institute for Brain Science. Layer 4 model of v1. available from: https://portal.brain-map.org/explore/models/l4-mv1.

[13] 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 3: **A**. ...


Exc
σ_ε = 0.1
σ_ε = 0.25
σ_ε = 0.4
EPI
ABC
Sst
σ_ε = 0.1
σ_ε = 0.25
σ_ε = 0.4
EPI
ABC
Pb
σ_ε = 0.1
σ_ε = 0.25
σ_ε = 0.4
EPI
ABC
Vip
σ_ε = 0.1
σ_ε = 0.25
σ_ε = 0.4
EPI
ABC