(Samuel Eckmann, Yashar Ahmadian, Máté Lengyel, Cosyne 2023, Montreal, Canada)

Input-dominated Hebbian learning enables image-computable E-I networks

Recurrent network models of excitatory (E) and inhibitory (I) neurons with supralinear activation functions have successfully explained several cortical computations, including response normalization and surround suppression.[1,2] Unlike more abstract approaches,[3,4] such networks allow direct comparisons with experimentally measured neural activities and synaptic strengths. However, the scope of these networks remained limited as their connectivity needed to be designed by hand or fitted by complex machine learning algorithms to yield stable activity and computations.[5,6] Here we present a method to efficiently construct stable recurrent E-I networks with connectivity that reflect the statistical regularities of high-dimensional natural images using a diverse set of feedforward receptive fields. We build on recent work that demonstrated the emergence of functional E-I networks via online Hebbian learning from an input population with homogeneous tuning curves,[7] and employ a simple covariance plasticity rule at all recurrent synapses without constraints on input tuning. When the network's activity is dominated by feedforward inputs, we can solve for the steady-state weight matrix analytically. This allows us to construct stable networks with recurrent weights adapted to complex input statistics without hand-tuning or numerical optimization. We demonstrate our approach by constructing two fully connected networks of 6,500 neurons and >40 million synapses each, encoding natural image datasets resembling the upper and lower visual field of mice, respectively.[8] We found that correlations between cross-oriented receptive fields in the upper visual field were weaker than those in the lower visual field, while iso-oriented receptive fields were more strongly correlated in the upper visual field. These statistics became reflected in the networks' synaptic connectivity and predicted weaker cross-orientation, but stronger iso-orientation surround suppression in the lower compared to the upper visual field. In summary, our method enables image-computable models of stable, supralinear E-I networks that allow for detailed comparison with heterogeneous cortical circuits.

We consider E-I rate networks with dynamics

$$\dot{\mathbf{y}} \propto -\mathbf{y} + [\mathbf{W}\mathbf{y} + \mathbf{H}\mathbf{z}]_+^{\circ n}, \quad [x]_+ = \max(x, 0), \tag{1}$$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_E \\ \mathbf{y}_I \end{pmatrix}, \ \mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & -\mathbf{W}_{EI} \\ \mathbf{W}_{IE} & -\mathbf{W}_{II} \end{pmatrix}, \ \mathbf{H} = \begin{pmatrix} \mathbf{H}_E \\ \mathbf{H}_I \end{pmatrix}, \tag{2}$$

where bold symbols denote matrices and vectors, and $[\cdot]^{\circ n}$ denotes the element-wise power. The network receives external input $\mathbf{z}$ via fixed feedforward receptive fields (RFs) $\mathbf{H}$. We assume a competitive Hebbian learning rule at synapses between neurons i, j of type A, B

$$\dot{w}_{ij}^{AB} \propto y_j^B \left(y_i^A - \bar{y}_i^A\right) - \gamma_i^A w_{ij}^{AB}, \quad A, B \in \{E, I\}, \tag{3}$$

where y are firing rates with means $\bar{y}$. The scalar γ maintains the total synaptic weight of all recurrent excitatory or inhibitory inputs[7,9] such that

$$\sum_j w_{ij}^{AB} = W_{AB}, \quad \mathcal{W} \equiv \begin{pmatrix} W_{EE} & W_{EI} \\ W_{IE} & W_{II} \end{pmatrix}, \tag{4}$$

while we set negative weights to zero, adhering to Dale's law. Following previous work,[10] we make the assumption that during learning, the network is input-dominated and we further ignore the neuronal nonlinearity, such that $\mathbf{y} \propto [\mathbf{H}\mathbf{z}]_+ \equiv \mathbf{p}$. The expected synaptic weight change becomes (cf. Eq. 3):

$$\langle \dot{\mathbf{W}} \rangle \propto \mathbf{C} - \mathbf{\Gamma}\mathbf{W}, \quad \mathbf{C} = \langle \mathbf{p}\mathbf{p}^\mathsf{T} \rangle - \langle \mathbf{p} \rangle \langle \mathbf{p}^\mathsf{T} \rangle, \tag{5}$$

where $\mathbf{C}$ is a covariance matrix, and the diagonal matrix $\mathbf{\Gamma}$ holds the appropriate γ_i^{AB} normalization factors. We make the simplifying assumption that for each excitatory neuron there is an inhibitory neuron with the same receptive field, i.e., $\mathbf{H}_E = \mathbf{H}_I \equiv \mathcal{H}$, resembling a cortical microcolumn. Then $\mathbf{C}$ becomes a 2×2 block matrix with four identical submatrices, so that the learning fixed point, for which $\langle \dot{\mathbf{W}}^* \rangle = 0$, becomes a Kronecker product

$$\mathbf{W}^* = \mathcal{W} \otimes \bar{\mathbf{C}}_+, \tag{6}$$

where $\bar{\mathbf{C}}_+$ is the positive part of the $\mathbf{C}$-submatrix, normalized such that each row sums to one and the total synaptic weights are set by the 2×2 matrix $\mathcal{W}$. Synaptic weights that connect two neurons with negative covariance decay to zero due to Dale's law. The recurrent connectivity within each E-I microcolumn becomes proportional to $\mathcal{W}$, and the activity of the full network remains bounded if activity in each microcolumn remains bounded, which can be analysed analytically.[1,11] To load a set of natural images into a network, we define a fixed set of receptive fields $\mathcal{H}^\mathsf{T} = \left(\mathbf{h}_1, \mathbf{h}_2, \ldots\right)$, compute response patterns $\mathbf{p}_i = [\mathcal{H}\mathbf{z}_i]_+$ for each input image $\mathbf{z}_i$, and set the recurrent weight matrix according to Eq. 6 (Fig. 1a, b).

The statistics of visual stimuli are not homogeneous across the visual field. For example, tree branches in the upper visual field of wild mice induce spatial correlations across large distances, while grass blades in the lower visual field imply more short-ranged correlations (Fig. 1b) (we expect similar differences in correlations for lab-grown mice). These statistics should be reflected in the connectivity between neurons in the visual cortex and lead to distinct response patterns when a stimulus is presented in either the upper or the lower visual field.

To test this hypothesis, we defined a set of oriented receptive fields that varied in their location, orien-

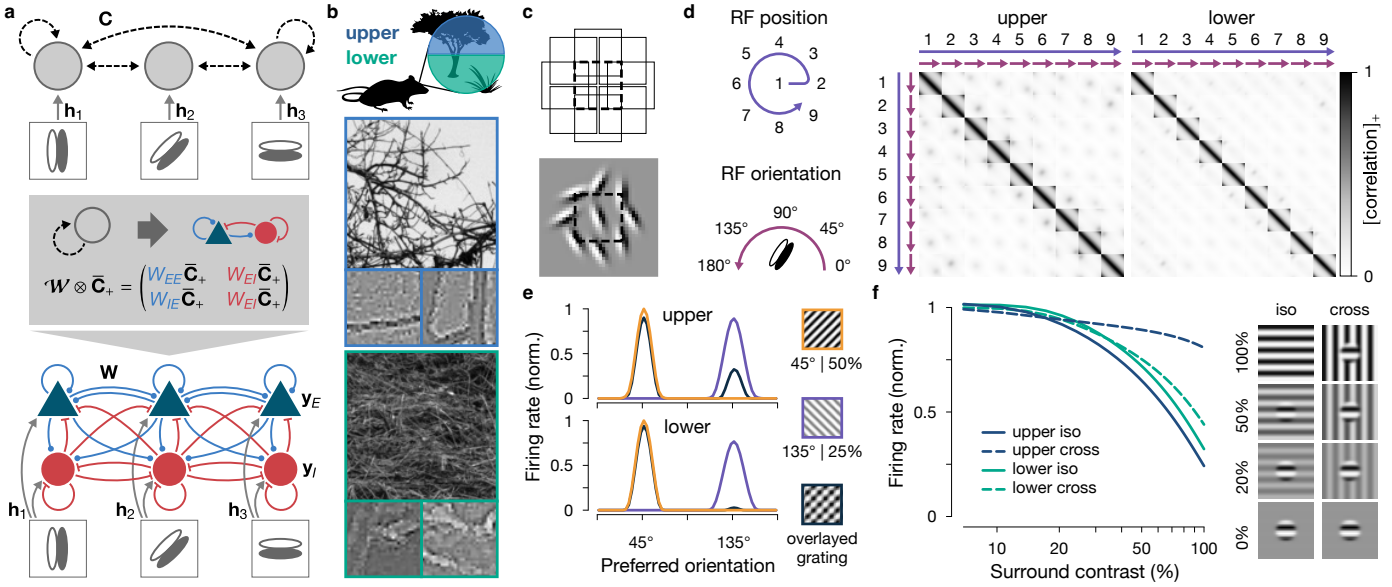

Figure 1. **a**, Network construction. Top: computation of response covariances in the input-dominated learning phase. Center: evolving network units to E-I column via Kronecker product of weight norm matrix and normalized covariance matrix. Bottom: full E-I network (some synapses not drawn for clarity). **b**, Upper (blue) and lower (green) regions of the visual field differ in their image statistics (tree branches compared to grass). Bottom: example natural images (500 × 500 pixels) characterising the upper and lower visual field in mice,[8] below – whitened example patches $\mathbf{z}_i$ (28 × 28 pixels) used for network construction. **c**, Overlapping Gabor receptive fields arranged in a centre (dashed square) surround pattern. **d**, Rectified response correlation between different receptive fields in the upper (centre) and lower (right) visual field. RFs are sorted according to their orientation and location tuning, indicated by coloured arrows (left). Correlations were averaged and renormalized over RFs with different phases. **e**, Response normalization in the upper (top) and lower (bottom) visual field. Two different test gratings of different orientations and contrasts (right) were presented to the network either separately (orange and purple) or overlayed (dark blue). For the overlayed grating mostly the higher contrast orientation is encoded in the network's population response. **f**, Surround suppression in the upper (blue) and lower (green) visual field. Different orientations and contrasts in the surround result in different suppression levels of activity in a neuron tuned to the centre orientation. Firing rates are normalized to responses at zero surround contrast).

tation, and phase tuning (Fig. 1c). We loaded two image sets corresponding to the upper and lower visual field[8] into two separate networks. We observed that neurons with RFs in the upper visual field were strongly connected to neurons with iso-oriented receptive fields in the surrounding regions, while neurons with RFs in the lower visual field showed weaker orientation preference in their recurrent connectivity (Fig. 1d).

This difference in connectivity between networks encoding the upper and lower visual fields is reflected in a difference in circuit dynamics. After initiation, we observed the networks' responses to test stimuli while activities progressed *outside* the input-dominated regime (according to Eq. 1). While both networks showed response normalization (Fig. 1e) and surround suppression (Fig. 1f), in accordance with experimental results,[4] the network trained on input from the upper visual field showed stronger iso-oriented surround suppression but weaker response normalization compared to the network trained on input from the lower visual field.

While these results demonstrate the usefulness of our approach to model heterogeneous circuits in the visual cortex, our method generalizes naturally to arbitrary input statistics and thus paves the way for a new generation of functional recurrent E-I network models of unprecedented scale and biological realism.

References

1. Y. Ahmadian *et al.*, *Neural Comp.* (2013).
2. D. B. Rubin *et al.*, *Neuron* (2015).
3. O. Schwartz, E. P. Simoncelli, *Nature neuroscience* (2001).
4. M. Carandini, D. J. Heeger, *Nature Reviews Neuroscience* (2012).
5. S. Di Santo *et al.*, *bioRxiv* (2022).
6. W. W. M. Soo, M. Lengyel, *bioRxiv* (2022).
7. S. Eckmann, J. Gjorgjieva, *bioRxiv* (2022).
8. L. Abballe, H. Asari, *PloS one* (2022).
9. K. D. Miller, D. J. MacKay, *Neural computation* (1994).
10. J. J. Hopfield, *Proceedings of the national academy of sciences* (1982).
11. N. Kraynyukova, T. Tchumatchenko, *Proceedings of the National Academy of Sciences* (2018).