**Abstract – Bernstein**

**Noise and load asymmetry enforce an optimal heterogeneous mixing in cerebellar-like circuits**

Across various cerebellar-like architectures in the brain, a recent flurry of results from electrophysiology point towards the presence of neurons that are mixedly selective to various input modalities[[1]](#footnote-1)[[2]](#footnote-2)[[3]](#footnote-3)[[4]](#footnote-4). Moreover, **many of these neurons display a heterogeneous degree of mixing** across a specific anatomical region - i.e some neurons are found to be fully selective to all modalities, others purely selective to only one modality, and many that lie in between those two limits. This begs the question of what the computational benefits are of displaying such a heterogeneous mixing degree.

Studying a bimodal model of associative learning with a Hebbian readout[[5]](#footnote-5), without loss of generality, we find that the dependence of the generalization on the mixing heterogeneity is largely controlled by the symmetry of two input modalities, specifically the input noise ratio, - – and the sample number ratio - - for a given sparseness and expansion ratio. In the fully symmetric case where both and equal to unity, the optimal network parameters – i.e those that minimize the generalization error - are those that fully mix all modalities across all neurons on the cortical layer – an intuitive result given that no modality is preferred (see red region in Figure). However, upon deviations from the fully symmetric case, we find that **a heterogeneous mixing degree is preferred**. Indeed, in the limit of a large load imbalance, the network prefers to be purely selective towards the modality with the lesser load, a result we are able to rationalize analytically via simple scaling arguments.

Our results show that in the multi-modal paradigm – on a canonical model of associative learning with Hebbian readout weights - full mixed selectivity across cortical neurons on feed-forward cerebellar-like circuits i**s not necessarily more computationally beneficial** than other selectivity schemes, in stark contrast to the unimodal case [[6]](#footnote-6). In reality, the optimal degree of mixing largely depends on a rich interplay between input parameters of the respective modalities that are used to train the network, providing a theoretical underpinning for widely observed results from electrophysiology.



Caption: (a) – (b): Phase diagrams are shown, denoting the optimal mixing probabilities, to , for the cerebellar-like model studied, with the parameters shown in the above. N and M denote the number of input neurons to each modality (denoted ξ and φ respectively), is the number of neurons on the intermediate cortical layer, P is the number of independent stimuli ξ to be learned by the network in the associative learning task, with K the respective one for φ, is the sparseness of the cortical layer, and the ’s represent the input noise parameters in each modality, providing a distance measure from the training and test datum. (a) and (b), we note that when sparseness is decreased the system exhibits a much stronger preference to being fully mixed (red region in the center grows larger). Figures (c) and (e) display two examples of realizations of the model studied for separate choices of to , an extension of Babadi and Sompolinksy5 . In (c) we display an example of a network with parameters , , whereas in (d) we have . Note that σ denotes the labels (valences) for the associative learning task, containing two indices from both input modalities (μ and γ respectively). Figures (d) and (f) are colour coded by the region highlighted in Fig (a). Ternary plots show the generalization error for all possible combinations of to on a grid spacing of 10. In (f) we see that for the fully symmetric modalities, is preferred (blue region in bottom left), whereas the purely selective limits are the least optimal (yellow shaded region in other two corners). In (d) we have instead a vast asymmetry in the input noise parameters, and we have instead as the preferred configuration (top corner), thus the network prefers to be purely selective to the less noisy modality.

1. Ishikawa, Taro, Misa Shimuta, and Michael Häusser. "Multimodal sensory integration in single cerebellar granule cells in vivo." *Elife* 4 (2015): e12916. [↑](#footnote-ref-1)
2. Huang, Cheng-Chiu, et al. "Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells." *Elife* 2 (2013): e00400. [↑](#footnote-ref-2)
3. Yagi, Ryosuke, et al. "Convergence of multimodal sensory pathways to the mushroom body calyx in Drosophila melanogaster." *Scientific reports* 6.1 (2016): 1-8. [↑](#footnote-ref-3)
4. Poo, Cindy, et al. "Spatial maps in olfactory cortex during olfactory navigation." *bioRxiv* (2020). [↑](#footnote-ref-4)
5. Babadi, Baktash, and Haim Sompolinsky. "Sparseness and expansion in sensory representations." *Neuron* 83.5 (2014): 1213-1226. [↑](#footnote-ref-5)
6. Fusi, Stefano, Earl K. Miller, and Mattia Rigotti. "Why neurons mix: high dimensionality for higher cognition." *Current opinion in neurobiology* 37 (2016): 66-74. [↑](#footnote-ref-6)