An efficient coding theory for cortical connectivity

Recurrent synaptic connections strongly shape representations of sensory stimuli in cortical circuits, but their functional role in sensory processing is not well understood. Theoretical work has highlighted several possible functions for recurrent cortical connectivity. First, local excitation and global inhibition may generate attractor dynamics that amplifies, sharpens, or temporally integrates tuned sensory input to enhance information transmission. Second, tuned inhibition may stabilise network dynamics and generate a sparse and efficient code for complex natural sensory statistics. Third, cortical connectivity may generate dynamics close to criticality or chaos, characterised by long network time constants and high capacity for discrimination of sensory inputs. Common to each of these theories is the hypothesis that cortical circuits are optimised to efficiently encode the sensory input they receive. However, the optimal circuit architecture for efficient encoding of a particular sensory input ensemble can rarely be determined.

To address this, we develop a novel analytical approach to maximise the Fisher information of a network via gradient descent with respect to its recurrent weights. We find that networks optimised to encode stimulus orientation have long integration timescales and ultimately become unstable, consistent with the temporal integration and edge of chaos hypotheses. Upon introducing a stability penalty, the network optimally trades off response information against dynamical stability, and exhibits local excitation and local inhibition with a Mexican hat profile, consistent with previous theories of efficient coding and redundancy reduction. Thus, we show that several classic theories for the functional role of cortical connectivity can be reproduced and unified under an efficient coding principle.