

Neocortical Keys and Locks: A Neural Model of Associative Learning by Coherence Induction Between Spike Patterns and Ongoing Membrane Potentials

René Doursat and Philip H. Goodman

Brain Computation Laboratory, ARF/400, University of Nevada, Reno, NV 89557 – doursat, goodman@unr.edu

In vivo recordings suggest precise and reproducible temporal order in neural signals that underlie behavior. In particular, neurons engaged in a perceptual or associative task seem to exhibit quick onsets of transitory correlations among firing patterns. Moreover, documentation of pervasive feedback connectivity suggests an alternative paradigm to the traditional feed-forward model, in which lower areas activate higher areas. Thus, interest is turning to the synchronous interplay of active subnetworks, which we will refer to as *coherence induction*. Here, local groups of neurons possess spontaneous modes of “ongoing activity” [1] that are influenced by stimuli in various ways. These preferred transient modes may be overt patterns of spikes or covert fluctuations of subthreshold potentials. In this context, stimulus-behavior association tasks can be reformulated as processes of *selection* among several transients. The stimulation of a subnetwork L (as in “lock”) by another subnetwork K (as in “key”) *engages*—but does not create—one of L ’s intrinsic modes. Learning (e.g., by STDP) then becomes a question of *reinforcing successful matches* rather than creating new matches. Our model of coherence induction contains N neuronal membrane potentials $\{V_i(t)\}_{i=1\dots N}$ that fluctuate quasi-periodically, representing simplified *in vivo* subthreshold recordings (Fig. a). Through recurrent connections, the potentials are pulled into relative coherence at a characteristic frequency f and each neuron is described by a phase ϕ_i . The set of phase shifts $\{\phi_i - \phi_1\}_{i=2\dots N}$ is an attractor mode of L : when an external stimulus is applied and removed, L consistently relaxes back to this mode. (Several factors could support this slow dynamic attractor, including apical dendritic background activity [2], transmission delays, and inhibitory pathways.) We define the order parameter as the interference sum $V_L(t) = \sum_i V_i(t)$ and look at its patterning and peak magnitudes when perturbing the neurons with specific spiking signals (Fig. b). We propose that the amplitude ΔV_L may represent the real propensity of a subnetwork to respond instantaneously as a population code to external perturbation. When L is relaxed, ΔV_L is near 0 due to mutual cancellation among scattered phases. When L is perturbed, phases are pulled toward each other and form transient, history-dependent coherent clusters, increasing ΔV_L in an irregular pattern. The main point is that network L ’s response will depend on the relative match between the temporal structure of the input stimulus K and the spectral composition and instantaneous phase distribution of V_i . Let this input be a spike train $V_K(t) = \sum_j \delta[\sin(2\pi ft + \psi_j(t))]$, where δ is a spike at 0. The core influence of K on L is then the following: if cell j sends a spike shortly after (before) V_i reaches its peak, then V_i peaks slightly later (earlier) and ϕ_i decreases (increases). Thus, j always attracts ϕ_i towards its own phase ψ_j . We model the displacement $\Delta\phi_i$ as a decreasing function of $\phi_i - \psi_j$ (e.g., a cosine). This mirrors the physiological nonlinearity that a cell is less likely to be brought to fire by incoming post-synaptic potentials the further it lies below firing threshold. Viewed on the phase circle, L is a pattern of scattered dots $\{\phi_i\}_{i=1\dots N}$. The net effect of one spike j on this circle is a sudden jerk of all the dots, to varying degrees, towards j ’s phase ψ_j (Fig. c). One repeated spike with constant ψ_j eventually coalesces the dots and raises the response amplitude. Two spikes with opposite phases interfere to form two opposite clumps of dots and perpetuate the cancellation. More spikes elicit more complex responses. When the stimulus is removed, the dots relax toward their original phase distribution constrained by the ongoing activity. Fig. c presents snapshots of L ’s phases: before stimulation, at different spike times (stars), and after stimulation. Our numerical experiments show (i) the remarkable uniqueness of the transient response of a specific phase distribution L to a specific incoming spike pattern K , despite identical mean rates, (ii) the reproducibility of this unique response, and (iii) its sensitivity to variations in either pattern, K or L . Thus, there is evidence for distinct “key and lock” engagement, provided a sufficient diversity of lock combinations (analogous to the complexity of the tumbler of a safe). These preliminary observations based on phase-space dynamics offer a promising approach to future models of real-time pattern recognition and stimulus-response learning based on spiking neural networks, and may also explain, in part, rhythmic components of EEG recordings. [1] Kenet *et al.*, *Nature* 425, 2003. [2] Destexhe & Paré, *Neurocomputing* 32-33, 2000.

