Synchronization and Semantization in Deep Spiking Networks

J. Oberste-Frielinghaus[1,2], A. C. Kurth[1,3], J. Göltz[4,5], L. Kriener[6,5], J. Ito[1], M. A. Petrovici[5], and S. Grün[1,7]

[1] Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany
[2] RWTH Aachen University, Aachen, Germany
[3] RIKEN Center for Brain Science, Wako, Japan
[4] Kirchhoff-Institute for Physics, Heidelberg University, Heidelberg, Germany
[5] Department of Physiology, University of Bern, Bern, Switzerland
[6] Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland
[7] Theoretical Systems Neurobiology, RWTH Aachen University, Germany

Recent advances in training spiking neural networks (SNNs) via exact spike-time gradients enable a more biologically faithful reassessment of the deep learning hypothesis in the brain [1]. Unlike conventional artificial neural networks, which rely on continuous rate-based transfer functions, SNNs communicate through discrete spikes, much like cortical neurons. We analyze deep, feedforward SNNs trained for visual image classification (MNIST) using a spike latency code, in which information is carried by the relative timing of spikes rather than their count. Adhering to Dale's law, each hidden neuron is either purely excitatory or purely inhibitory, approximating the ratio found in cortex with 300 excitatory and 100 inhibitory neurons per layer.

After training, we approach these networks as we would electrophysiological recordings and characterize how activity propagates through the network layers. Input patterns, synchronous by construction, first spread out over time in early layers before re-converging into tight pulse packets in deeper layers, reminiscent of pulse packet propagation in synfire chains [2]. In addition, we observe an increasing class-specific activation of the neurons in the deeper layers of the network, reminiscent of the concept cells observed in the human medial temporal lobe [3]. To quantify this observation, we measure the information gain (IG), i.e., the information about the input class obtained by observing a neuron as active. Strikingly, excitatory and inhibitory neurons show a different pattern: the IG distribution of excitatory neurons shifts progressively towards full specificity (IG $= 1$) in deeper layers, while inhibitory neurons maintain low IG throughout the hierarchy, indicating that they remain largely class-unspecific. This asymmetry reveals that this semantization is mainly a property of the excitatory neurons while the inhibitory neurons appear to serve a more global, class-independent role, in accordance with experimental observation in biological networks [4]. Based on the network connectivity, we identify class-specific feedforward pathways of excitatory neurons that preferentially route activity towards the correct output neuron, becoming increasingly pronounced with depth.

Together, these findings demonstrate that deep spiking networks trained by gradient descent naturally acquire activity patterns characterized by synchronicity and separability, without any explicit constraint enforcing either. This provides a computational hypothesis for the synchrony and feature selectivity observed experimentally in the visual cortex as a natural consequence of hierarchical learning with spikes, establishing a bridge between modern spike-based learning theory and in-vivo electrophysiology. Our results further provide an analysis framework applicable to both artificial and biological neuronal networks.

[1] J. Göltz et al., Nat. Mach. Intell., 3, 823–835, 2021.

[2] M. Diesmann et al., Nature, 402, 529–533, 1999.

[3] R. Q. Quiroga et al., Nature, 435, 1102–1107, 2005.

[4] K. Sohya et al., J Neurosci, 27.8, 2145–2149, 2007.