Dendrites and Rewiring Endow Recurrent Spiking Neural Networks with Compute- and Parameter-Efficient Time Series Regression

M. Lefebvre and C. Frenkel

Department of Microelectronics, Delft University of Technology, Delft, The Netherlands

Despite significant advances in edge artificial intelligence algorithms and hardware, efficiently regressing and classifying time series at the edge remains an open challenge. On the one hand, conventional approaches such as recurrent neural networks, state-space models, and edge transformers, reach high performance but with significant compute and memory requirements. On the other hand, neuromorphic systems inspired by the brain appear as a promising alternative given their explicit temporal dynamics and potential to improve energy efficiency by leveraging local and temporally-sparse computations. Yet, they lack clarity as to which brain-inspired mechanisms need to be modeled and which level of bioplausibility is right, to deliver performance while reducing model requirements.

In this work, we propose rewired dendritic recurrent spiking neural networks (rdRSNNs), which augment neurons with dendrites and rewire the network connections to enable high-performance yet compute- and parameter-efficient regression of time series. First, dendrites with different time constants endow neurons with a memory at multiple timescales, complexifying the temporal relationships they can capture and produce. Second, rewiring evolves the sparse connectivity of the network and, thereby, the receptive field of the dendritic compartments, adding a notion of space. These two features effectively create rich spatio-temporal representations within both the neurons and network.

Table 1: Comparison to state-of-the-art models trained on Neurobench NHP dataset.

	[1] tinyRSNN	[1] bigRSNN	[2] AEGRU	[3] GRU-t1	[4] SNN_3D	**This work**
Test R^2 (Indy + Loco)	**0.6604 ± 0.0021**	0.6978 ± 0.0027	0.6982 ± 0.0089	0.707 ± 0.012	0.7062	**0.7224 ± 0.0047**
Footprint [bytes]	27,144	4,833,360	45,520	352,904	33,219	706,128†
Connection sparsity	0.4549 ± 0.0111	**0**	0.5039 ± 0.0054	**0**	**0**	**0.8818**
Non-zero weights	**7326 ± 149**	**1,206,272**	26,930 ± 294	22,342	39,063	**16,476**
Activation sparsity	0.9836 ± 0.0001	0.9683 ± 0.0005	0	0	0	0.9647 ± 0.0021
Dense*	13,440	1,206,272	54,283 ± 2	22,342	39,063	139,392
Multiply-accumulate (MACs)	**66**$^\diamond$	1,026$^\diamond$	25,316 ± 322	8,518	**32,261**	**512**
Accumulate (ACs)	**304 ± 9**	**42,003 ± 550**	0	793	0	**855 ± 8**

* Dense reflects the number of operations on a hardware that does not support sparsity.
† Footprint includes non-sparse weight matrix used during training. $^\diamond$ MACs in neurons' synaptic compartment.

We demonstrate the capabilities of the proposed rdRSNNs on the neural decoding task of the Neurobench non-human primate (NHP) dataset [5]. It consists in predicting velocity components of a monkey's finger from spiking neural activity. The results are obtained for an rdRSNN with 64 hidden neurons with eight dendrites per soma and two readout neurons. We compare them to previous works in Table 1. Our rdRSNN reaches a record test R^2 of 0.7224 that surpasses RSNNs and gated recurrent units (GRUs) alike. It is both compute-efficient, with only 855 accumulate (AC) operations in the dendrites and 512 multiply-accumulate (MAC) operations in the somas, as well as parameter-efficient, with only 16.5k non-zero weights. The sparse connectivity mask entails a moderate footprint overhead of 19.7%, which could be cut by 8× with a binary mask representation instead of a boolean one (1 byte). This work thus paves the way for efficient time series regression at the edge.

[1] T. Liu et al., Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS), 1–5, 2024.

[2] Y. Wang et al., Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS), 1–5, 2024.

[3] A. Vasilache et al., Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS), 1–5, 2024.

[4] B. Zhou et al., Neuromorph. Comput. Eng., 5, 014013, 2025.

[5] J. Yik et al., Nat. Commun., 16 (1), 1545, 2025.