Trainable structural plasticity with radio frequency spintronic neural networks

Théophile Rageau(1), Julie Grollier(1), Dédalo Sanz-Hernandez(1) and Frank Mizrahi(1)

(1) Laboratoire Albert Fert, CNRS, Thales, Université Paris-Saclay Palaiseau, France

One critical limitation for edge learning is that it requires large, dense networks with more parameters than edge devices can support. While we can remove as high as 99% of weights in pre-trained networks for inference, training sparse networks with a small number of parameters is challenging. Both neuroscience and machine learning research point to the solution of structural plasticity (i.e. tuning the strengths and also the topology of the connections) to train a network with a smaller number of parameters. However, a hardware implementation is challenging, as in-memory computing platforms (typically crossbar arrays of non-volatile memories) implement hard-wired network topologies.

Here we propose to use radio-frequency (RF) spintronic networks [1], which perform in-memory computing, and where we can dynamically train the network topology (we refer to this system as "RF spintronic network"). In RF spintronic networks, neurons are implemented as oscillators emitting RF signals at distinct frequencies, while synapses act as resonators functioning as tunable frequency-selective filters. As described in [1], fine-tuning the frequency of a synapse modifies its synaptic weight. By applying large shifts to the resonance frequency of synapses, we change to which neurons they are attached, thus reconfiguring the network topology.

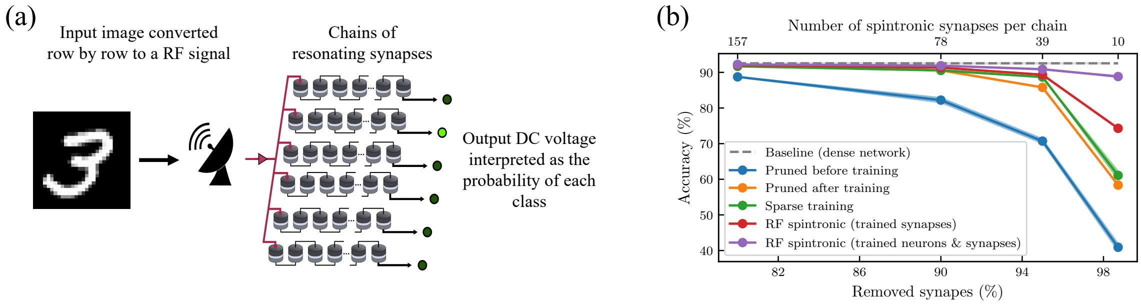

Figure 1: (a) Diagram of a RF spintronic perceptron fed with an image taken from the MNIST database. (b) Accuracy as a function of the percentage of removed synapses (or equivalent number of spintronic parameters) for different sparse feed-forward networks (blue, orange and green) and RF spintronic networks (red and purple).

Using numerical simulations (based on Pytorch with physical models of RF spintronic networks), we show that we can train the topology and weights of a sparse RF spintronic perceptron (as shown in Figure 1a) to recognize images from the MNIST database. Figure 1b provides a comparison between the RF spintronic network (red and purple) and feed-forward networks (blue, orange, green) in which a large amount of synapses have been removed. For a given number of trainable parameters, we train the different networks and compare their test accuracy. For extreme sparsity level (1.28% of remaining synapses, equivalent to a RF network with 10 spintronic synapses per chain), the RF network (red) outperforms the sparse training perceptron (green) by 13.21%. By also training input frequencies (purple), method which does not require additional nano-devices, the RF spintronic network reaches 88.83% accuracy, which is only 3.71% less than a dense feed-forward network (dashed grey line).

This work paves the way toward a hardware implementation of a spintronic neural network with on-chip training for machine-learning tasks.

[1] A. Ross et al., Nature Nanotechnology, 18(11), 1273-1280, 2023.