Attojoule Modulators for Photonic Neuromorphic Algorithms

R. Amin\textsuperscript{1}, Z. Ma\textsuperscript{1}, J. B. Khurgin\textsuperscript{2}, T. El-Ghazawi\textsuperscript{1}, P. R. Prucnal\textsuperscript{3}, V. J. Sorger\textsuperscript{1}

\textsuperscript{1}Department of Electrical and Computer Engineering, George Washington University
\textsuperscript{2}Department of Electrical and Computer Engineering, Johns Hopkins University
\textsuperscript{3}Department of Electrical Engineering, Princeton University

Neuromorphic networks are computational algorithms and network models inspired by signal processing in the brain with societal-relevant applications in machine-learning such as speech and image recognition, non-linear optimization, and real-time simulation. Recent progress by both industry and academia has demonstrated compute efficiencies surpassing the digital efficiency wall set by von-Neuman compute architectures (Fig. 1a). An artificial neural network consists of a set of input artificial neurons connected to both the hidden- and the output layer. Within each layer, information propagates by a linear combination such as matrix multiplication followed by the application of a nonlinear activation function (Fig. 1b). Optical neural network implementations offer unique advantages over microelectronics because the linear transformation (e.g. matrix multiplication) can be executed in the optically with temporal response times <10ps.

Here we show scaling vectors for (nano)photonic-based neurons in optical neural networks (ONN). The systems energy efficiency depends on the sum of all optical components’ quantum efficiencies, $\eta$, the fan-in-out ratio, and the energy efficiency of the electrooptic modulator ($E$/bit)\textsubscript{EOM} providing the non-linear activation function through its transfer function.

In this talk I will discuss modulator scaling laws and share our recent demonstrations of plasmonic attojoule per bit efficient modulators on a silicon photonics platform (Fig. 1b). The modulator’s efficiency scales inversely with the physical device volume, $V$, and underlying cavity quality $Q$-factor, thus inversely with the Purcell factor, $F_P$. Thus, increasing the weak light-matter-interaction and using Graphene’s Pauli blocking mechanism, we demonstrated a modulator with a record-high performance (1 dB/V-µm) while being micrometer compact.

References

Figure 1. a, Performance vectors for compute systems. Neuromorphic (= brain inspired) systems break the digital efficiency wall set by CMOS and von-Neuman architectures [1]. b, Model of an optical perceptron neuron inside a artificial neural network. The nonlinear activation function is provided by the transfer function of the plasmonic electrooptic modulator. The computational system efficiency scales with the modulator efficiency, which in turn scales inversely with device the physical volume, $V$, and material broadening [1,2]. Our plasmon modulators are micrometer-compact consuming 100’s aJ [3,4].