Device-Level Modeling of Photonic Integrated Circuits for Neuromorphic Computing

Andrea Marchisio

Under the supervision of

Professor Andrea Carena
Professor Paolo Bardella
Professor Vittorio Curri

Abstract

In recent years, the world experienced the Artificial Intelligence (AI) revolution: since the launch of ChatGPT in 2022, AI has suddenly become an extremely popular topic, not only at the academic research level, but also for the broader general public, on one hand, sparking countless moral discussions about its ethical implications, on the other, attracting billions of dollars in investments from the major tech companies (e.g., Google, Meta, NVidia, etc.).

This AI surge has been powered both by the computational capabilities of modern computers and by the capillary availability of data. Indeed, AI, in particular Machine Learning (ML), is a "data-driven" application, meaning that it requires vast and diverse datasets in order to be effectively trained. However, the required dataset size has been doubling every 3.5 months and, at some point, traditional computers based on the Von Neumann architecture will no longer be able to effectively handle the input data.

In this context, a new paradigm could be required to maintain the projected AI trends. Photonic Integrated Circuits (PICs) appear to be an interesting candidate for the creation of the next-generation hardware accelerators for AI. Indeed, despite being a relatively new and less mature technology with respect to standard CMOS, PICs offer several beneficial advantages. For instance, they show very high operational speed (information is carried by light), low power consumption (most of the commonly used components are fully passive), and high parallelism (various techniques could be employed, such as wavelength, polarization, and mode division multiplexing). Moreover, photonic platforms can implement matrix multiplications very naturally, whereas traditional Von Neumann architectures perform them inefficiently due to the high number of

memory accesses required.

To leverage these advantages, it is possible to design PICs for neuromorphic computing applications. Among the possible strategies, it is possible to create Photonic Neural Networks (PNNs) made with a linear section performing the weighted sum of the inputs and a nonlinear section applying the nonlinearity to the output, thus mimicking the operation of a traditional artificial neuron.

In this context, the work of this thesis tackles the modeling of PICs for photonic neuromorphic computing applications, dealing with both the linear and nonlinear sections of the PNNs.

In particular, the linear section of the PNN is typically implemented with a mesh of interconnected thermally-controlled Mach-Zehnder Interferometers (MZIs): by controlling the propagation of light through the MZIs via thermal tuning, the weighted sum can be performed. However, without proper insulation, neighboring devices will influence each other with spurious thermal interactions (i.e. thermal crosstalk), which can strongly reduce the final accuracy of the trained PNN. For this reason, we have developed a comprehensive model for MZI meshes able to predict both the light propagation through the circuit and the thermal crosstalk between the devices. The model is validated with the measurements of a 3X3 MZI mesh with incomplete interconnections and it is shown to accurately predict the crosstalk in this reference PIC.

Furthermore, in this thesis, an all-optical Nonlinear Activation Function (NLAF) device able to implement nonlinearities for neuromorphic applications has been analyzed. This device is realized on the Indium phosphide Membrane on Silicon (IMOS) technological platform and it is made with an MZI cascaded with a ring-assisted MZI. A model for predicting the nonlinearities of the IMOS Microring Resonator (MRR) has been constructed and validated against experimental data and it has been used to create a simulative model for the complete NLAF device.