Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Electrical, Electronics and Communications Engineering
(34^{th}cycle)

Memristor-based hardware accelerators: from device modeling to AI applications

Francesco Marrone

******

Supervisor:
Prof. Fernando Corinto

Doctoral Examination Committee:
Dr. Alon Ascoli, Referee, Technische Universität Dresden
Prof. Mauro Di Marco, Referee, Università degli Studi di Siena
Prof. Michele Bonnin, Politecnico di Torino
Prof. Rodrigo Picos Gayá, Universitat de les Illes Balears
Prof. Sung-Mo "Steve" Kang, University of California Santa Cruz

Politecnico di Torino
2022

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Francesco Marrone
2022

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Memristor-based hardware accelerators: from device modeling to AI applications

Francesco Marrone

In recent times, the interest in artificial intelligence algorithms has skyrocketed well out of academia marking a substantial shift in most, if not all, the industrial sectors. This revolution started in the early 2010s and was made possible by the combined availability of very powerful parallel processing units (e.g. GPUs) and more and more abundant "big data". Although the GPU has served well as platform to kick-start this revolution, currently this conventional digital hardware accelerator struggles to satisfy the increasingly high requirements of machine learning algorithms within a reasonable power consumption range. Most of the energy inefficiency which characterizes conventional digital hardware accelerators (e.g. GPU, TPU, FPGA etc) is to blame on the traditional von Neumann architecture they implement which involves data being sent back and forth between memory and the processor. In this context, neuromorphic computing has emerged as an alternative to contemporary processing units' architectural design choices which may rise to the challenges posed by modern AI training and inference tasks. Neuromorphic computing takes inspiration from the biology of the animal brain by putting the storage of information and its processing in very close spatial proximity if not encoded and performed by the same processing element. It is this the case of memristive technologies which have attracted the interest of lots of researchers and companies as good candidates to implement the synaptic function in neuromorphic hardware accelerators. The overall goal of this doctoral thesis is the investigation of memristive based computing architecture exploiting an holistic approach ranging from the modeling of memristive devices to the mapping of advanced machine learning algorithm on crossbar arrays. The first part of the thesis is devoted to memristive technologies modeling techniques while its second part deals with the design and study of meristor-based architectures for machine learning tasks. In particular the thesis contains novel results on the use of DRMs for Phase Change Memory devices, the study of Spiking Neural Networks and the design of advanced memristive crossbar based accelerators for linear algebra problems (Pagerank).