Self-Testable Silicon Neural Networks

Mots clés :

- Directeur de thèse : Haralampos STRATIGOPOULOS
- Co-encadrant(s) :
- Unité de recherche : Laboratoire d'informatique de Paris 6
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Sciences et technologies de l'information et de la communication

Résumé du projet de recherche (Langue 1)

Nowadays, AI algorithms typically run in a general-purpose CPU or in the cloud in a cluster of CPUs. However, there are many applications that demand dedicated AI hardware, i.e. building neurons and synapses and control circuitry in silicon. Examples include: (a) building an artificial brain in silicon to serve as a platform for neuroscientists to perform brain activity simulations; (b) performing AI onto the IoT devices instead of transferring the data to the cloud which poses security, bandwidth, latency, and energy issues; and (c) performing AI in real-time on AI hardware accelerators that offer the minimum possibly latency required for example in autonomous driving technology. Nowadays, there exist a number of approaches for performing neuromorphic computing. Neural networks can be emulated in an FPGA or GPU. Still the speedup can be up to 10x and power consumption is too big, for example for fitting these devices onto IoT nodes. Another big bottleneck is the memory wall problem of the data movement between the processor and off-chip memory storage in the conventional von-Neumann computer architecture. For these reasons, there is a great interest in building efficient VLSI implementations of neural networks. Today there exist CMOS ASIC designs and there is a lot of activity on emerging non-volatile memory devices and that efficiently implement synaptic arrays. With the foreseen industrialization and high-volume production of hardware neural networks in the coming years, testing strategies specific to hardware neural networks is an emerging topic that is largely unexplored. In general, post-manufacturing testing aims at detecting manufacturing errors and is done per manufactured chip. Moreover, for safety- and mission-critical applications, testing also needs to be performed in the field concurrently with the operation or in idle times. For this purpose, BIST capabilities need to be added into the design, which will allow stand-alone evaluation of the health status of the chip. BIST is a critical block for putting in place a self-healing methodology towards a fault-tolerant design. In this thesis, we will focus on self-testing and fault-tolerance specifically for spiking neural networks that are more biologically plausible and offer the greatest energy efficiency since they are event-driven. We plan to demonstrate BIST for spiking neurons and synapses and show how the BIST can detect failures due to process variations and physical defects.