Internship Proposal

Design and Evaluation of Computing in Memory Solutions for AI Hardware Accelerators

Hosting Institution: ***TIMA Laboratory***
Supervisor: ***Mounir BENABDENBI and Ioana VATAJELU***
Duration of contract: ***6 months***
Contact: Mounir BENABDENBI (Mounir.Benabdenbi@univ-grenoble-alpes.fr)

Context:

Since the appearance of modern computers, the widely adopted architecture has been based on the separation between the computing unit (or processor) and the memory storing the program to be executed and its data, i.e., the Von Neumann architecture. The separation between processor and memory has become an issue in modern computers due to the uneven evolution of processing speed and memory access times (also known as memory wall). With the technology advancements, the memory wall became increasingly important. Therefore, there is an urgent need to explore alternative architectures in the light of emerging non-volatile technologies (resistive memories), not only to further increase the computing efficiency at lower cost, but also to further reduce the overall energy.

Among emerging device technologies, non-volatile memory technologies such as Ferroelectric [1] (FeFET) favor increasing system complexity and performance at lower power consumption; thus, providing the scientific community with opportunities for new computer architecture innovations being able to tackle today's limitation. For example, moving the computation to the memory (rather than doing it in the CPU) will significantly reduce the communication and therefore reduce the power consumption and increase the performance. This computing paradigm is also referred to as Computation-in-Memory (CIM), emerging concept based on the tight integration of traditionally separated memory elements and combinational circuits, that ensures the minimization of the time and the energy needed to move data across the processor. Computation-In-Memory (CIM) architecture, aims at eliminating the communication bottleneck while supporting massive parallelism. However, to achieve the ultimate objective of fully integrating the processing units and the memory in the same physical location, several technological challenges need to be overcome.

There is a very wide variety of CIM solutions proposed today that exploit existing technologies. They enable logic and/or arithmetic operations directly inside the memory boundaries. The operations are performed without the need of transferring data to/from the CPU, thus saving time and energy. This can be achieved exploiting the physical characteristics of the FeFET transistor.

The purpose of this project is to create an exhaustive library of logic gates implemented with FeFET transistors and evaluate their behavior for different FeFET devices, fabrication processes and operation conditions.

The project will be conducted in close collaboration (regular meetings) with renowned scientists from the INL laboratory at Lyon (France).

Internship Proposal

Objectives:

O1: Create a data-base of primitive gates for computing-in memory for different types of resistive-memory devices and arrays.

O2: Evaluate the FeFET-based logic gates behavior to voltage and other parameters' tuning, device-to-device and cycle-to-cycle variability.

O3: Create a library of HDL models of FeFET based gates, back annotated with extracted values from the simulations (voltage, speed, consumption, reliability level, …)
These models will allow the implementation and simulation of artificial neural networks based on FeFET logic in memory.

Tasks:
The work of the internship will consist in accomplishing the following tasks:

- Task 1: State-of-the-art – study the existing computing in memory solutions on FeFETs;
- Task 2: Implementation and spice simulation of basic logic gates in-FeFET transistors array;
- Task 3: Evaluate the behavior of the logic gates implemented at task 2 for different transistor parameters
- Task 4: The HDL models will be annotated with the duration of the logic operation, power consumption, etc.
- Task 5: Design and validation of a MAC operator based on the results of the previous tasks

We are looking for a highly-motivated Engineering School or M2 Masters student. Applicants must have a Master 1 (or equivalent) in a related field of microelectronics. Interpersonal skills, dynamism, rigor and teamwork abilities will be appreciated. Candidates should be fluent in English and/or in French and have good English writing skills.

Our team welcomes applicants with diverse backgrounds and experiences. We regard gender equality and diversity as strength and an asset.

Required skills: gate-level design, spice simulations, Cadence, scripting and programming languages

For highly motivated student, with a good CV, the internship may be followed by a PhD application.

About TIMA:
TIMA Laboratory is a public joint research laboratory located in Grenoble, France, and held jointly by Institut Polytechnique de Grenoble (Grenoble INP), University Grenoble-Alpes and French National Research Council (CNRS). TIMA is a multinational team of over 100 people, with members and interns from all over the world. The research topics of TIMA cover the specification, design, verification, test, CAD tools and design methods for integrated systems, from analog and digital components on one end of the spectrum, to multiprocessor Systems-on-Chip together with their basic operating system on the other end.
This call is from the AMfoRS team, and targets people motivated by hardware design and test. More information about the team is available at http://tima.imag.fr/tima/en/AMfoRS/AMfoRSoverview.html

Application:
Please send your CV and letter of motivation, together with 2 references to:
Mounir Benabdenbi (Mounir.Benabdenbi@univ-grenoble-alpes.fr)

[1] Y. Long et al., "A Ferroelectric FET-Based Processing-in-Memory Architecture for DNN Acceleration," in IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, vol. 5, no. 2, pp. 113-122, Dec. 2019, doi: 10.1109/JXCDC.2019.2923745.