PhD positions in TIMA Laboratory - SLS Team

October 13, 2023

Introduction

The TIMA Laboratory (website) is an academic research entity situated in Grenoble city, in France. 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.

The SLS team (System Level Synthesis - website) focuses on highly efficient architectures for general purpose computing or AI-dedicated algorithms, and system-level modeling and design methodology : specification, simulation and verification of hardware/software systems on chip; design exploration and synthesis of hardware.

This document presents several PhD positions that are available in the SLS team, linked to AI (Artificial Intelligence), in particular acceleration of AI and usage of AI for EDA (Electronic Design Automation).

About TIMA: TIMA is located within historical building of the Grenoble Polytechnic Institute, just near the city train station. It is fully accessible for people with disabilities.

TIMA is a multinational team, with members and interns from all over the world. The Laboratory and its parent institutions ensure that everyone is treated equally, respectfully, regardless of gender, religion, disabilities, etc.

Requirements: New applicants should send a CV and an overview of their work experience and full coursework and grades, by email and preferably in PDF format, to the contacts below. For students, Masters degree or engineer diploma or other degree with formal equivalence is mandatory. Please present clear, exhaustive and verifiable curriculum, as there is a standard inquiry before taking position (especially for non-European applicants).

Contacts:

Adrien Prost-Boucle, adrien.prost-boucle@univ-grenoble-alpes.fr
Frédéric Pétrot, frederic.petrot@univ-grenoble-alpes.fr

1.2 Mapping of neural networks on distributed low-quantization neuron blocks

Context: Hardware-accelerated learning and/or inference solutions can involve optimized hardware blocks that model the behavior of a certain number of neurons. This is the case in existing AI accelerators found in certain FPGAs, and this will be the case with emerging technologies of analog-inspired models of neurons (e.g., based on ReRAM, memristors, etc.). Mapping of neural networks on distributed computing blocks is not a new topic. But it has always been assumed that the underlying hardware natively supports the quantization required by the neural network model.

The energy savings brought by binary or ternary quantifications are also very high, at the neuron hardware unit level. So in the near future, efficient hardware targets will probably consist of neuron units with quantization support actually lower what the neural network model requires. This may need revised requirements about memory bandwidth, multiplexing and general data distribution and routing circuitry between clusters of neuron units.

Objectives: Assuming the nature of the blocks of dedicated neuron units is given, the challenge involves optimization of the cost of routing between blocks, the merging of block results, the remaining glue logic, FIFOs, etc. Conversely, there are opportunities to identify which properties the optimized blocks should have, such that the glue logic stays relatively minimal while keeping a high level of genericity. Indeed, the increased efficiency brought by the optimized blocks should not be cancelled by the potential overhead of extra glue logic and memory operations.

Potential heterogeneous layer quantizations will be taken into account, inter-layer and intra-layer. The targeted hardware blocks may also need extra time and/or energy to load new weights and/or activations, which impacts design of mapping and inference algorithms to keep high efficiency.

Proof of concepts can involve transaction-level simulation and/or monolithic or distributed FPGA implementations.

Involvement in the design of the low-level optimized ASIC neuron blocks will be possible and be recommended.

Requirements:

- Experience with AI, neural network architectures and tools
- Good programming skills
- Good knowledge of digital circuit architectures
- Knowledge of usual HDL languages would be a plus
- Knowledge of ASIC design would be a plus
- Knowledge of analog circuits would be a plus
- Interest in technologies of microelectronics
- Interest in problem solving