COMPUTER AND CONTROL ENGINEERING

MUR DM 117/STMicroelectronics - Design techniques for low-area digital circuits in industrial and medical applications based on machine learning

Funded By	MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] STMICROELECTRONICS S.R.L. [P.iva/CF:00951900968] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	MACII ENRICO - enrico.macii@polito.it
Contact	MACII ENRICO - enrico.macii@polito.it JAHIER PAGLIARI DANIELE - daniele.jahier@polito.it
Context of the research activity	The goal of this thesis, co-funded by ST Microelectronics, is to develop innovative solutions for the integration of complex software algorithms on silicon. The work will focus in particular on deploying Machine Learning algorithms on industrial and medical smart sensors, whose digital components are ultra-tightly constrained in terms of area occupation (hence memory space) and power consumption, leveraging the flexible and extensible the RISC-V Instruction Set Architecture. Progetto finanziato nell'ambito del PNRR - DM 117/2023 - CUP E14D23002020004
	The main objective of this thesis is to investigate and develop innovative solutions for the integration of algorithms on silicon through interpreters, using hardware-software co-design techniques. The candidate will focus on developing solutions based on the RISC-V Instruction Set Architecture (ISA), enhanced with custom extensions. This will require a comprehensive understanding of the RISC-V ISA, including the design and implementation of custom instructions that can be integrated into an existing processor pipeline. The ultimate goal of this research is to reduce the area of specific smart sensor applications in the industrial and medical domains, currently on the market, such as step-counters, heart rate monitors, infrared vision sensors, etc. The main direction that will be pursued to achieve this objective will be the hardware acceleration of critical portions of the involved computation,

Objectives	thereby trading off the addition of new logic with a reduction in the required program memory. In fact, by replacing entire sections of code with a single accelerator invocation, the total area of the system can be reduced, while also possibly reducing peak power consumption and improving energy efficiency. The candidate will develop a full hardware-software co-design flow to identify the critical sections of an application, accelerate them, and then provide the required software infrastructure (ISA extension, compiler) to support the newly designed accelerators. Throughout this process, the candidate will work closely with ST Microelectronics, a world leader in the design of intelligent platforms. The company will direct the candidate towards the most relevant research problems, giving them a unique opportunity to integrate their work within state-of-the-art industrial design flows and possibly even contribute to the development of new products. This collaboration will provide the candidate with access to cutting-edge hardware and software tools, allowing them to explore the latest developments in the field of smart sensor technology. In summary, this thesis will provide a unique opportunity for the candidate to develop innovative solutions for the integration of algorithms on silicon, while also gaining hands-on experience in hardware-software co-design and microprocessor design. The work plan of the project will be structured as follows: Months 1-9: Literature review and background study, focusing on i) the RISC-V ISA and its peculiarities; ii) Hardware-software co-design and RISC-V ISA extensions; iii) Compilation toolchains, and iv) Techniques for Machine Learning deployment on constrained systems (quantization, pruning, NAS, etc). Furthermore, the candidate will also become familiar with the smart sensor applications identified by ST Microelectronics as main benchmarks of interest for the rest of the work. Months 9-24: Implementation of hardware extensions and software deployment toolchains to reduce area on the target smart sensor applications based on machine learning. In this phase, the candidate will intervene on the hardware and software sides of the system separately, first developing an accelerator/ISA extension (e.g., for low-precision or sparse computation) and then adding compiler support for it, to enable its usage starting from a high-level specification (e.g., in Python) of the target machine learning application, transparently to the developer. Months 24-36: Connection of the HW and SW sides of the work into a unique, automated, HW/SW co-design tool, in which the architecture and configuration of the developed extensions/accelerators are co-optimized with the characteristics of the deployed ML model “in the loop”. Possible publication venues for this thesis include: - IEEE Transactions on CAD - IEEE Transactions on Computers - IEEE Journal on Internet of Things - IEEE Transactions on Circuits and Systems (I and II) - IEEE Transactions on Parallel and Distributed Systems - ACM Transactions on Embedded Computing Systems - ACM Transactions of Design Automation of Electronic Systems - ACM Transactions on Architecture and Code Optimization - Conferences such as DAC, DATE, MLSys, and others.

	The EDA Group has many active industrial collaborations and funded projects on these topics, many of which involve ST Microelectronics, including: - TRISTAN (ECSEL-JU 2023) - ISOLDE (ECSEL-JU 2023) - StorAIge (ESCEL-JU 2021) - Etc.

Skills and competencies for the development of the activity	1. Good programming skills in Python and C. 2. Familiarity with embedded systems and computer architectures 3. Familiarity with compilers and compiler optimizations is a nice-to-have, but not a hard requirement, since these concepts will be studied during the first period of the thesis. 4. Basics of hardware design in Verilog/VHDL are nice-to-have since the candidate will develop accelerators and ISA extensions. However, the group internally has skills on this, and can support the candidate.