POLITECNICO DI MILANO

PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 41st cycle

Research Area n. 1 - Computer Science and Engineering

THEMATIC Research Field: DESIGN OF ENERGY-EFFICIENT AI ACCELERATORS BASED ON IN-MEMORY COMPUTING

Monthly net income of PhDscholarship (max 36 months)
1400.0 In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	System-on-Chip (SoC) architectures for edge devices and low-end high-performance computing (HPC) integrate multicores and heterogeneity to handle a wide range of control and data processing tasks to meet the increasing computational demands of modern Deep Learning (DL) workloads efficiently. The primary design challenges include compute bottlenecks due to the high complexity of deep neural networks (DNNs), memory bandwidth limitations that dominate power consumption and reduce throughput, and scalability issues as integrating additional compute cores or dedicated AI accelerators increases silicon area and energy costs. Digital In-Memory Computing (D-IMC) is a promising paradigm that mitigates these limitations by enabling computation directly within memory macros, significantly reducing data movement and improving energy efficiency. Among different IMC technologies, SRAM-based D-IMC provides key advantages over non-volatile memory (NVM) implementations, including faster weight loading, reconfigurability, and compatibility with standard CMOS fabrication. However, integrating D-IMC into a scalable, modular, and workload-flexible SoC architecture for both edge AI and low-end HPC remains an open research challenge.

Methods and techniques that will be developed and used to carry out the research	This PhD research focuses on designing innovative heterogeneous architectures based on RISC-V that integrates D-IMC technology as a fundamental computing unit, enabling the design of new hardware accelerators for AI with a focus on energy efficiency, performance, scalability and workload flexibility. The research will explore the tradeoffs between performance, area, and power consumption while considering constraints unique to the Deep Learning domain, given its significant market volumes and computational complexity.A key aspect of this research is investigating various design space exploration (DSE) strategies for integrating D-IMC into a multicore RISC-V-based SoC. The research will evaluate different coupling approaches, such as loosely and tightly coupled integration, and assess their implications on instruction set extensions, programming tool support, achievable performance, and power efficiency. Furthermore, a critical step in the study will be integrating D-IMC within the microprocessor pipeline data-path leveraging RISC-V's flexibility for specialized AI acceleration techniques.Another critical aspect is scalability, ensuring that the proposed accelerator is adaptable for edge AI applications and low-end HPC workloads. The research will define a modular approach where multiple D-IMC tiles can be instantiated based on the computational demand, optimizing parallelism and workload distribution in heterogeneous multicore architectures.By the end of the PhD program, the research aims to design a novel AI accelerator architecture based on RISC-V, integrating D-IMC at both the architectural and instruction set levels. The ultimate objective is to develop a toolchain and a proof-of-concept prototype, based on an FPGA to assess design trade-offs in terms of energy consumption, computational efficiency, and system cost. - Develop design techniques to combine DIMC tiles either loosely- or tightly coupled to the central processing unit of a RISC-V processor. - Develop design techniques of a RISC-V based data-path and their function units to directly exploit the functionality provided by DIMC tiles. - System-level modeling and exploration techniques to evaluate design tradeoffs and tune System-on-Chip

	platforms with respect to various metrics associated with a system configuration such as performance, area and power consumption.
Educational objectives	Acquire and/or consolidate knowledge and/or practical skills around: • Modeling and simulation of computer architectures at the system-on-chip level; • Hardware architectural design, simulation and exploration; • Design of experiments and Machine-Learning models; • Analysis and design for power/performance tradeoffs.
Job opportunities	The research proposal addresses an output profile that responds to the needs of the edge-computing industry for technical experts in the design and the development of next generation digital system architectures to accelerate Deep Learning applications.
Composition of the research group	2 Full Professors 3 Associated Professors 3 Assistant Professors 6 PhD Students
Name of the research directors	Prof. Cristina Silvano

Contacts
cristina.silvano@polimi.it – 0223993692 -- https://silvano.faculty.polimi.it/

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents	--

Scholarship Increase for a period abroad	
Amount monthly	700.0 €
By number of months	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information
EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student. TEACHING ASSISTANTSHIP: availability of funding in recognition of supporting teaching activities by the PhD student. There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations. COMPUTER AVAILABILITY: 1st year: Yes 2nd year: Yes 3rd year: Yes **Economic awards** up to total euro **5029,50** (gross amount), meaning 1676,50 (gross amount) euro per year, will be recognized to the PhD candidate in case of significant contribution in the research project, subject to the evaluation of the research director.