POLITECNICO DI MILANO

PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 40th cycle

Research Area n. 2 - Electronics

PNRR 630 Research Field: EDGE AI BY ANALOG IN-MEMORY COMPUTING: FROM DEVICE TECHNOLOGY TO ALGORITHM MAPPING

Monthly net income of PhDscholarship (max 36 months)
€ 1500.0 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	In the emerging field of edge AI, the imperative is to maximize energy efficiency and cost. As the capabilities of neural processing units (NPUs) evolve, the demand escalates for enhancements in computational efficiency, performance, and memory footprint minimization. To transcend the constraints of conventional Von Neumann architectures, the industry and academic sectors are pioneering novel designs that leverage computational memories, incorporating both analog and digital methodologies.In this scenario, one of the most promising methodologies is to merge low-precision, low-power in-memory computing (IMC) modules with high-precision dataflow inference engines to facilitate the acceleration of deep neural networks (DNNs). Despite the recognition of IMC in both industry and academia, the substantial storage costs required to accommodate advanced AI algorithms have yet to align with the stringent constraints of TinyML and embedded devices.To explore the large design space from the single bit cell to the analog/digital circuits required to create larger computational memory macros, it is essential to model the effects of different schemes via an advanced modeling framework focused specifically on DNN algorithms. Pruning, quantization and sparsity appear as a pivotal techniques in optimizing computational graphs for standard NN algorithms. While

	these compression methods have proven beneficial in software (SW) and, to a lesser extent, digital hardware (HW) accelerators, their application in in-memory computing, particularly in analog formats utilizing resistive memories like RRAM and PCM, is less defined.In this context, the PhD thesis will aim at the design of novel DNN accelerators based on IMC in the presence of massive quantization and sparsity. The thesis will first address the problem from a high level of abstraction to identify the main figures of merits and the theoretical correlation linking sparsity, quantization, accuracy, memory/area occupation and energy consumption. In the second phase, integrated circuits blocks will be developed to allow a realistic simulation of the various figures of merit of the in-memory computing accelerator. The system level architecture will then be developed by including converters and digital processors. Finally, real-life applications in the automotive, industrial and biomedical sectors will be explored to identify a path to commercial exploitation.
Methods and techniques that will be developed and used to carry out the research	- Integrated circuit design to develop the analog/digital in-memory computing circuits - Device characterization and engineering to control the number of conductance levels, their precision and readout time - Neural processing to tailor quantization and sparsity in the deep network - Circuit simulation to assess the impact of sparsity and quantization on performance, efficiency and accuracy
Educational objectives	Acquire and/or consolidate knowledge and/or practical skills around: - Circuit simulation and design for application specific integrated circuits - Neural modeling for the deep learning - Device, circuit, algorithm codesign for optimized neural networks
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 computing systems to accelerate Deep Learning applications.
Composition of the research group	1 Full Professors 2 Associated Professors 3 Assistant Professors 6 PhD Students
Name of the research directors	Prof. Daniele Ielmini

Contacts
daniele.ielmini@polimi.it - 0223996120 - https://ielmini.faculty.polimi.it/

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	750.0 €
By number of months	6

National Operational Program for Research and Innovation	
Company where the candidate will attend the stage (name and brief description)	STMICROELECTRONICS S.R.L., Agrate Brianza (Monza e della Brianza)
By number of months at the company	6
Institution or company where the candidate will spend the period abroad (name and brief description)	STMICROELECTRONICS S.R.L., Francia
By number of months abroad	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