Machine Learning for Approximate Computing

Basic information:
Applications are invited for 15 full-time PhD student positions (called Early Stage Researchers or ESR) for PhD degrees for 36 months. The positions are offered within the Approximate Computing for Power and Energy Optimisation (APROPOS) International Training Network action of the highly appreciated EU-funded Marie Skłodowska Curie grants. The positions are fully funded for a 36-month period. You can choose your position in one of 14 top-level universities and companies in Europe and complement your experience within periods from 3 to 9 months of cross-sector secondments by working with the highly committed industrial partners of the action. You are encouraged to apply for up to three positions within APROPOS network if you believe your profile is suitable for multiple topics. A separate application is needed for each position in case you apply for multiple positions.

APROPOS is an international, multidisciplinary and multi-sectorial training network programme on approximate computing. APROPOS is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No. 956090.

The ESR-specific project is described below. Please read the description carefully before applying. We recommend you choose the position you apply for based on your technical profile and desired technical expertise, rather than based on the hosting location, as all positions involve high-level research mobility.

Project information:
The main focus of the research will be the study of variable- and mixed-precision computing techniques to improve energy efficiency. The rationale behind this idea is that a significant decrease in processing accuracy at different parts of the computing chain can drastically reduce power consumption and computing time, which are the two factors composing energy consumption. The goal is to reduce the number of bits required by program variables while respecting bounds on the accuracy of the application output, measured in terms of error ratio w.r.t. the output generated with full precision. The current state-of-the-art is mainly based on iterative methods that repeatedly execute the application with different configurations, exploring the variable precision range through binary search; the process stops when a minimal configuration that respects the target error constraint is found. This process is extremely time-consuming, especially when the number of variables increases. Moreover, different target errors require to restart the solution search from scratch.

Approaches based on machine learning (ML) can greatly improve state-of-the-art by leveraging previously collected knowledge that can be integrated into an optimization model capable of making smarter decisions. ML models should accurately predict the error associated with a variable configuration; in conjunction with an optimization model, optimal precision levels respecting the target error could be computed. There are two main approaches to implement the ML models that will be investigated: 1) an off-line approach, where a data set comprising numerous previous applications executions with different configurations can be used to train an ML to predict the error given a configuration; 2) an active learning approach, which alternates learning and optimization phases by starting with

a simple ML model and refine it after having computed and tested a configuration. The methodologies deriving from this research must consider the issues related to their integration into a full software stack, including compilation toolchains (e.g., GCC and LLVM) and parallel runtime environments (e.g., OpenMP). The proposed activities also require extensive testing on the benchmark applications developed in the context of APROPOS on target hardware platforms, including (but not exclusively) the open-source RISC-V based platform PULP (www.pulp-platform.org).

ESR Host Unit: University of Bologna (UNIBO), Italy

ESR Objectives:
> • Understand the impact of variable precision on the application error in off-line and mixed off-line/on-line machine learning approaches
> • Support the benchmarking activities to evaluate efficiency and efficacies of the proposed strategies.

ESR Expected Results:
Proof-of-concept implementation of adaptive reconfigurable edge computing node exploiting approximate computing methods; completion of the PhD program at the hosting unit (within 48 months).

Planned secondment(s): IBM, supervisor Dr. Cristiano Malossi, M18, 6 months for experimenting with the methodology on real platforms in real working environments, and for gaining industrial experience.

Supervisory team:
- Prof. Luca Benini (UNIBO), main supervisor
- Prof. Michela Milano (UNIBO), academic co- supervisor
- Dr. Giuseppe Tagliavini (UNIBO), academic co- supervisor
- Dr. Cristiano Malossi (IBM), industrial co- supervisor
- Dr. Christoph Hagleitner (IBM), industrial co- supervisor

Enrolment in Doctoral degree: University of Bologna (UNIBO), Italy

Target degree: PhD in Data Science and Computation

Degree details: https://www.unibo.it/en/teaching/phd/2020-2021/data-science-and-computation

Expected start date: September 2021 (M11 from the beginning of the project)

Approximate gross salary: about 3250 EUR/month for researchers without family (and about 3650 EUR/month for researchers with family); Note: for different positions within the same APROPOS network, the salary is country- and unit-dependent.

Duration: 36 months

Researcher Profile: First Stage Researcher (R1)

Research Field: Engineering; Computer science; Technology

Type of contract: Temporary

Job Status: Full-time

Hours per week: 40

Trial period: 2 months

Working and living conditions in the country: **Italy** is one of the most culturally and economically advanced countries in the world, with the eighth-largest economy by nominal GDP (third in the European Union), sixth-largest national wealth and third-largest central bank gold reserve. It ranks very highly in life expectancy, quality of life, healthcare, and education. Location of the oldest University in the world, **Bologna** is an important industrial center in Northern Italy, where many mechanical, electronic and food companies have set their headquarters**.** The city is also a "UNESCO creative City of Music" and has one of the largest and most well-preserved medieval historical centres, full of restaurants, taverns, theatres, and shops.

Read more about Italy and Bologna:
- https://www.enit.it/wwwenit/en/
- https://emiliaromagnaturismo.it/en/towns/bologna

Unit details: UNIBO is the second largest university in Italy and active in research and technology transfer. It is one of the most important institutions of higher education in EU with 87,000 enrolled students, 2,857 Academic staff, 1,198 post-docs, 1,606 PhDs, 3,014 administrative and technical staff units. UNIBO is the top ranked university in Italy according to the Times Higher Education Rankings, 2021. The activity of the University of Bologna will be conducted within the Department of Electrical, Electronic and Information Engineering (DEI) and the Department of Computer Science and Engineering (DISI).

DEI is one of the largest departments in UNIBO with an excellent research profile. Within DEI, the research activity in APROPOS are carried out by the Energy-Efficient Embedded Systems Lab, led by Prof. Luca Benini.

DISI is one of the strongest centers for computer science and engineering research in Italy, Within DISI, the research activity in APROPOS is carried out by the Artificial Intelligence Laboratory, led by Prof. Michela Milano.

Read More about UNIBO at: https://www.unibo.it/en

Seconding Unit description: IBM Research GmbH, (IBM Research - Zurich), with approximately 300 employees, is a wholly owned subsidiary of the IBM Research division with headquarters at the T.J. Watson Research Center in Yorktown Heights, NY, USA. IBM Research - Zurich, which was established in 1956, represents the European branch of IBM Research. At the lab, scientific and industrial research is conducted in three scientific and technical departments: Cloud & Computing Infrastructure, Cognitive Computing & Industry Solutions,

Science and Technology. Main research topics are cognitive computing, AI, machine learning, blockchain, IoT, nanotechnology, supercomputing, advanced server technology, RAID systems & tape storage, security and privacy. IBM Research - Zurich employs a steady stream of post-docs, PhD candidates, and summer students who pass through the laboratory. More than 30 nationalities are represented among the research staff members, including such specialists as computer scientists, mathematicians, electrical engineers, physicists, and chemists. Read more about IBM Research at: https://www.research.ibm.com

Eligibility criteria: Please check carefully that you are eligible; all conditions below are compulsory

Transnational mobility: The researcher must not have resided or carried out their main activity (work, studies, etc.) in the country of the host organization for more than 12 months in the 3 years immediately prior to the start date. Note: the mobility rule applies to the (main) beneficiary where the researcher is recruited, and not to partners to which the researcher is sent or seconded. It is also only determined at one point in time: at the time when the ESR starts working for APROPOS.

Early-Stage Researchers (ESRs)/fresh MSc graduates: All researchers recruited in APROPOS must have less than 4 years since the completion of their first MSc degree and have not been awarded any doctoral degree at the date of the employment.

Background requirements: The applicant must be in possession of Master of Science (MSc) diploma in a relevant field, such as: electrical engineering, computer engineering, computer science. Strong programming skills and exposure to fundamental concepts and techniques in computer architecture, compilers and run-time systems is required. Exposure to machine learning theory and practice is a plus.

English language requirements: The candidate must be in possession of an English certificate with good level according to the regulations of the university to which the student will be enrolled for the PhD degree – check the degree requirements of the corresponding doctoral program for details.