Ph.D. in Information Technology: Thesis Defense

February 11th, 2021
online by Zoom – at 9.30

Luca STORNAIUOLO – XXXIII Cycle

Hardware/Software Co-design for Scientific Computing and Convolutional Neural Networks on FPGA-based Embedded Architectures

Supervisor: Prof. **Donatella Sciuto**

Abstract:

FPGAs are increasingly used in embedded heterogeneous architectures as they provide high computational power, low power consumption, flexibility, and adaptability thanks to their reconfiguration property. The rapid evolution of fields such as Machine Learning, Statistical Computing, and Biomedical Computing, along with the ending of Moore's era, is moving the attention of industry and academia towards less traditional computer architectures that can satisfy the ever-increasing need for high computational power and low power consumption, with FPGA-based Heterogeneous System Architectures (HSAs) being a promising solution. However, owing to the learning curve required to implement FPGA-based accelerators, the number of applications that benefit from their use is quite small. Although the technology of these devices continues to evolve, their difficulty in use is still preventing them from spreading widely.

To solve the FPGA-based HSAs programmability and usability challenges, this thesis focuses on the formalization and implementation of techniques and tools to develop and exploit accelerators that take advantage of reconfigurable embedded architectures and integrate them within more complex heterogeneous infrastructures. The goal is twofold. First, the thesis aims at providing new workflows and innovative techniques to hardware developers in order to optimize specific domain functions and to distribute their designs ready to be integrated by end users. In particular, we focus on accelerating applications for Scientific Computing by taking advantage of the programmable logic of the FPGA. We propose innovative FPGA-based hardware/software designs for Audio Signal Alignment, Template Matching and Biomedical applications capable of outperforming their software versions on embedded devices. Moreover, for applications based on Convolutional Neural Networks, which are generally difficult to implement on embedded systems due to their hardware resource constraints, we propose methodologies for their mapping to a distributed embedded heterogeneous system. Second, in parallel to this, the thesis aims at providing software developers with new methodologies to take advantage of FPGA-based hardware accelerators transparently directly from High Level Languages. Starting from the efficient management of the on-chip memory of FPGAs up to the automatic runtime management of the heterogeneous system directly from Python.

PhD Committee

Prof. **Marco Santambrogio**, DEIB
Prof. **David Atienza**, Ecole Polytechnique Federale de Lausanne – EPFL
Prof. **Theocharis Theocarides**, University of Cyprus