Ph.D. in Information Technology
Thesis Defenses

February 17th, 2022
at 9:00
online by Zoom

Davide CONFICCONI – XXXIV Cycle

On the Role of Reconfigurable Systems in Domain-Specific Computing

Supervisor: Prof. **Marco Santambrogio**

Abstract:

Computer architectures field is facing technological and architectural obstacles that are limiting the general-purpose processor scaling in the delivered performance at a reasonable energy cost. Domain-specialized solutions are a novel alternative to deliver extremely high performance at a relatively low energy profile and even more whenever combined with high-level abstractions for designing and programming it. Domain-Specific Architectures (DSAs) generally are the prominent exponent for domain specialization. Moreover, DSAs are programmable software architectures designed for few tasks, with the minimal amount of advanced CPU-based microarchitectural techniques, and to be efficiently implemented as Application-Specific Integrated Circuits (ASICs), or part of System on Chip (SoC). However, developing custom silicon devices is a time-consuming and costly process that is not always compatible with the time-to-market and fast evolution of the applications. Thus, adaptable computing platforms represent the most viable alternative for these scenarios. Field-Programmable Gate Arrays (FPGAs) are the candidate platforms for their on-field reconfigurable heterogeneous fabric. On top of the reconfigurability, FPGAs can implement large spatial computing designs and are publicly available on cloud computing platforms.

Therefore, this dissertation focus on Domain-Specific Reconfigurable Architectures (DSRAs): domain-specialized architectures with adaptable datapaths implemented on FPGAs. The design of such architectures demands a clear view of system-level trends and FPGAs' abstraction layers. On top of that, advanced design methodologies enable domain-tailored energy-efficient architectures, and design automation toolchains open the path to apply iterative development cycles and reproducibility of results. Moreover, usability layers that span from the hardware-software interfacing to ways of programming the architecture are necessary for a user base creation and deploying usable hardware. For these reasons, this dissertation explores these crucial issues and presents relevant takeaways of the domain specialization role of reconfigurable computing systems.

Alberto PARRAVICINI – XXXIV Cycle

Accelerating Graph and Sparse Information Retrieval through High-Performance Reconfigurable Architectures

Supervisor: Prof. **Marco Santambrogio**

Abstract:

Graph analytics, information retrieval, and recommender systems are a pervasive component of our society, helping billions of users in finding web pages, movies to watch, and products to buy.
Sparse linear algebra is now an essential building block of these workloads.
By not storing redundant information, it enables real-time processing of an increasingly large amount of data. In particular, Sparse Matrix-Vector Multiplication (SpMV) is the cornerstone of many complex applications, from graph ranking to approximate search.
SpMV does not perform well on general-purpose hardware architectures with traditional caching strategies, as it contains little data reuse and unpredictable memory access patterns.
Modern Field Programmable Gate Array (FPGA) accelerator cards have a few tricks up their sleeve. With abundant on-chip memory and HBM they enable novel data representations that maximize operational intensity and parallelism. Reduced-precision fixed-point arithmetic, a distinctive feature of FPGAs, opens the doors to trade-offs between performance and numerical accuracy in error-tolerant workloads such as recommender systems.
This thesis proposes a set of SpMV FPGA hardware designs targeted at the needs of graph analytics and recommender systems. These SpMV designs can quickly adapt to different workloads, such as graph ranking, sparse eigensolvers, and sparse embedding similarity search. In all cases, we provide state-of-the-art performance and energy efficiency and study the impact of reduced precision on accuracy. Overall, we establish that FPGAs are a fearsome contender in the race for high-performance sparse computations in graph analytics and recommender systems.

PhD Committee
Prof. **Manuel Roveri**, Politecnico di Milano
Prof. **Diana Goehringer**, TU Dresden
Prof. **Dionisios N. Pnevmatikatos**, National Technical University of Athens (NTUA)