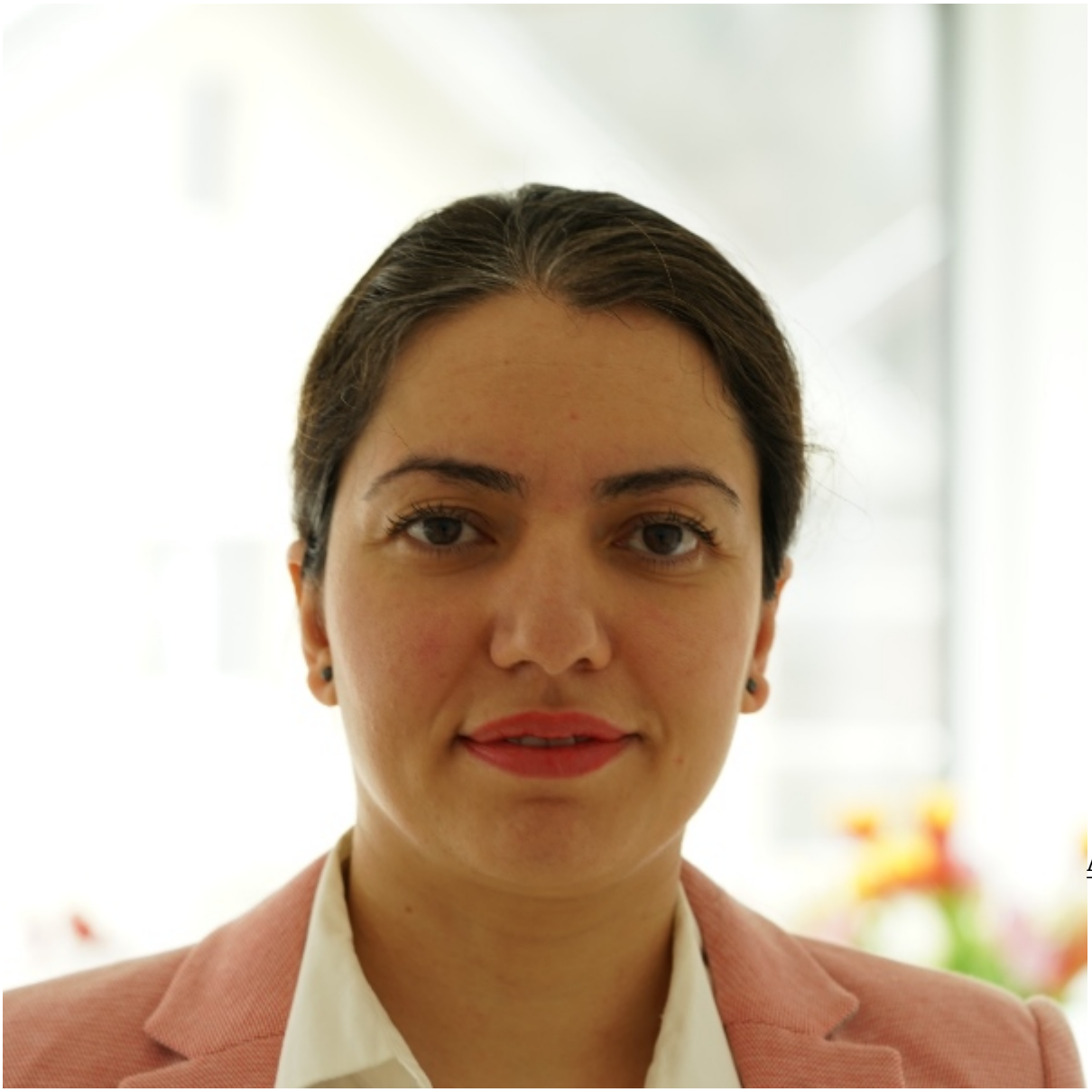
SORS/WomenInBSC: "A dynamic task scheduling method based on run-time information from the Miss Status Holding Register (MSHR) tables"

Objectives

Download the Presentation slides here

Abstract: The rise of new applications like data mining and graph analysis has heightened the demand for enhanced processing power at the hardware level. Traditional static task scheduling struggles to meet the intricate requirements of such applications, particularly when executed on Graphics Processing Units (GPUs). The challenge lies in distributing millions of instructions among a limited number of processing cores efficiently. In our recent work, we introduced a dynamic task scheduling method based on run-time information from the Miss Status Holding Register (MSHR) tables, aiming to rectify the inefficiencies of static scheduling. Additionally, we investigated the near LLC processing structures, and compared it to the near main memory processing alternative, and proposed a solution to dynamically offload instructions to different memory hierarchy levels, equipped with processing cores.

Further, our research extended to investigating load balancing across processing cores, interconnection networks, and memory controllers. This exploration led to the development of a method capable of balancing the load on memory controllers, thereby reducing memory request round-trip latency.



Azin Ebrahimi

Short Bio: Masoumeh (Azin) Ebrahimi is currently an Associate Professor at KTH Royal Institute of Technology, Sweden. She is also Adjunct professor at University of Turku, Finland. She received her PhD from University of Turku, Finland with honors in 2013. She also received MBA jointly from Turku School of Economics and European Institute of Innovation & Technology (EIT) ICT Labs. She joined KTH Royal Institute of Technology, Sweden as a Marie-Curie Prostdoctoral fellow in 2014. Currently she is leading several projects at KTH and (co)supervising several M.Sc. and PhD students. Her main areas of interests include interconnection networks, neural network accelerators, and GPU power/performance optimization.

Speakers

Speaker: Azin Ebrahimi, Associate Professor at KTH Royal Institute of Technology, Sweden
Host: Miquel Moretó, Associate Researcher-High Performance Domain-Specific Architectures, Computer Sciences, BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación