Title: Memory Management for Scalable Many-Core Architectures
Author: Nikil D. Dutt

Abstract

Technology scaling along with the growing demand for media-rich software stacks have motivated the need for many-core platforms. With the increase in compute power and its inherent demand for high memory bandwidth comes the need for vast amounts of on-chip memory space. Thus, designers must carefully provision the limited on-chip memory budget to meet their application's needs. Embedded systems often use both software controlled memories (e.g., scratchpad memories) and hardware-controlled memories (e.g., caches), with each having their pros and cons. Efficient on-chip memory management is extremely critical as it has a great impact on the system's power consumption and throughput. Traditional memory hierarchies primarily consist of SRAM-based on-chip caches, however, with the emergence of non-volatile memories (NVMs) and mixed-criticality systems, we expect to see heterogeneous on-chip memory hierarchies, not only in type (cache vs. scratchpad) but also in technology (e.g., SRAM vs. NVM). This talk will survey the state of the art in memory subsystems for many-core platforms, and presents strategies for efficiently managing software-controlled memories in the many-core domain, while addressing emerging challenges faced by designers, and will also propose a holistic software/hardware solution to managing the memory subsystem for scalable many-core architectures.