SORS: AI Engines to Accelerate the Age of AI

Objectives

[Click here to download the slides](#)

Abstract: AI is rapidly changing. Models become larger and more compute and memory intensive. This requires specialized architectures to provide high performance while taking little power and occupying little space. AI is also becoming pervasive with inference needs from the edge to the cloud and HPC systems. The AMD AI Engine provides AI acceleration that scales from laptops to data center with optimized compute and data movement to maximize reuse and latency tolerance. This talk will cover the range of systems targeted by the AI Engine and provide an overview of its architecture and the major features providing high performance and energy efficiency.

Short Bio: Alex Rico is a Principal AI Processor Architect at AMD. He works defining the AI Engine architecture for future AMD devices. He was previously a Principal Research Engineer with Arm Research where he led a team researching processor technologies for HPC and data center. He got his PhD at UPC/BSC on simulation methodologies for HPC workloads while contributing to novel system architectures for HPC. He holds more than 10 patents and has published over 40 scientific papers.

Speakers

Speaker: Alex Rico, Principal AI Processor Architect at AMD
Host: Miquel Moretó, High Performance Domain-Specific Architectures Associated Researcher, CS, BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación