Michael James

Session 4

Title:

Revolutionizing Supercomputers: Post-Exascale Insights from AI Architectures

Abstract:
As growth in AI has commercial institutions poised to possess the world's largest computers, the scientific computing community should focus on what supercomputer platforms can achieve beyond AI. The future of supercomputing lies in architectures that can unlock new computational opportunities beyond exascale capabilities. AI-driven computer architectures offer insights into designing supercomputer systems that can address problems inaccessible to current exascale platforms, such as long time-scale simulations and data-intensive applications. These architectures not only extend the feasible space for physical simulations by orders of magnitude but also enable real-time high-performance computing (HPC).

In this talk, we will delve into the Cerebras wafer-scale platform, showcasing its capability to achieve hundred-fold accelerations in strong-scaling HPC workloads. We will explore how the physical structure of a computer dictates the feasible space of problems it can address. Additionally, we will discuss research directions and examine hardware architecture possibilities for future systems and converged AI + HPC workloads.

Bio:
Michael is the Founder and Chief Architect of Advanced Technologies at Cerebras, the company renowned for creating the world’s largest and most powerful computer processor. He leads the initiative to redefine the algorithmic foundations for the next generation of AI technologies. Before Cerebras, Michael was a Fellow at AMD, where he developed adaptive and self-healing circuits based on cellular automata, enhancing the reliability of distributed fault-tolerant machines. Throughout his career, Michael has focused on the intersection of natural phenomena, mathematics, and engineered systems. His degree from UC Berkeley is in Molecular Neurobiology, Computer Science, and Mathematics.