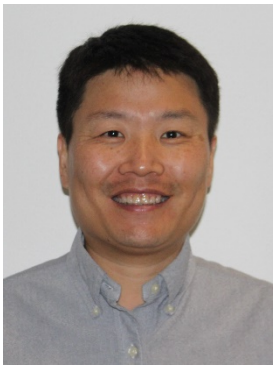
Key Challenges and Directional Path of Memory Technology for AI and High-Performance Computing

Keith Kim, NVIDIA



Bio

Keith Jangryul Kim is a Distinguished Engineer and memory architect at NVIDIA ASIC Memory Sub System group, leading future memory technology and early enabling memory technology for AI/HPC applications. Before Joining the NVIDIA, He worked for SK Hynix as memory engineer and technology lead. He had a lot of experience in various fields such as memory design, test, architecture and planning for HBM/DDR memory at SK Hynix. Especially, he was main member of HBM development and architect from 2013 and has been involved at beginning state of HBM development and contributed to design, test, PKG and eco-system enabling for SIP (System In Packaging). Based on this experience, he is architecting memory solution for AI/HPC and working with memory company to figure out optimum memory solutions. He received Engineering bachelor's degree in Ulsan University and He served as committee chairing for memory technology of Joint Electron Device Engineering Council (JEDEC).

Abstract: AI/ML continues to evolve beyond inferencing and training the image. Especially Generative AI unlocks new possibilities for all industries and larger language model is to accelerate computing resources, driving memory performance and density while focusing on energy efficiency. In addition, security and memory RAS cannot be ignored since datacenter data integrity is one of key technologies in datacenter. This short course will address the directional path of memory technology requested by AI/HPC applications, and technical challenges for bandwidth, power, density, RAS, and security. Both barriers and potential technology solutions will be reviewed. Towards resolving memory walls and meeting system requirements, initiatives to develop new memory technologies for AI/HPC are needed.