

2024년도 대한전자공학회
하계종합학술대회 초청강연 발표정보

■ 발표자 정보

성명	김동균	사진	
소속(학교)	SK하이닉스		
부서(학과)	Next Generation Memory Architecture		
직위	연구위원		
발표분야	SK하이닉스 세션		
약력	<p>Dongkyun Kim is a SK hynix fellow and the head of next-generation memory architecture for DRAM design. He received his B.S and M.S degrees in information and communication engineering from Chungbuk National University, Korea, in 1993 and 1995, respectively. He joined Hynix in 1995 and has been instrumental in the development of various DRAM technologies including EDO, SDR, DDR1/2/3/4/5, and LPDDR4/5. Since 2013, he has been leading the design team and achieved significant milestones such as the development of the world's first LPDDR4 and DDR5 DRAMs. In addition, he has been an active member of the JEDEC task group for current and future DRAM SPEC definitions since 2012. And His current focus is on exploring and developing the next generation of memory and sub-systems, particularly in relation to AI."</p>		

■ 초청강연 정보

제 목	Memory Devices for on device AI
-----	---------------------------------

Abstract

In the AI era, the LLM model size of Generative AI is increasing exponentially, getting closer to human ability, and accordingly, the demand for server's computing power is also rapidly increasing. However, in preparation for the explosion of demand for AI applications, it is predicted that the server system alone may lack response capabilities, so it is necessary to allocate roles for the AI inference operation of the client system. For this reason, interest in on-device AI is growing, and hardware and software research are actively being conducted to cover "AI inference" by PCs and mobile systems. In this presentation, I would like to introduce the advantages, disadvantages, and development directions of wide IO type high bandwidth memory, processing in memory, and Compute in memory, which are on device-oriented memory solutions.
