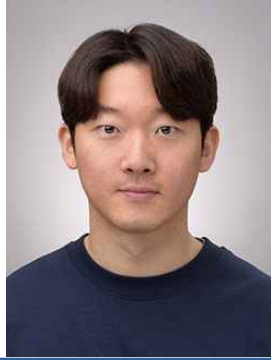


2024년도 대한전자공학회
하계종합학술대회 특별세션 강연 발표정보

■ 발표자 정보

성명	김희탁	사진	
소속(학교)	한국전자기술연구원(KETI)		
부서(학과)	SoC플랫폼연구센터		
직위	선임 연구원		
발표분야	디지털 하드웨어 설계		
약력	학력 및 경력 2011~2017: 경희대학교 전자공학과 학사 2017~2019: 고려대학교 전기전자공학과 석사 2019 ~: 한국전자기술연구원 SoC플랫폼연구센터 선임연구원 2023 ~: 고려대학교 전기전자공학과 박사 과정 연구실적 Heetak Kim, et al “An Energy-Quality Scalable STDP Based Sparse Coding Processor with On-Chip Learning Capability”, IEEE Transactions on Biomedical Circuits and Systems, vol.14, no. 1, pp. 125-137, Feb. 2020. Heetak Kim, et al “An Energy-efficient On-chip Learning Architecture for STDP based Sparse Coding”, ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED), July. 2019. 주요 연구 및 관심 분야 - 인공지능 하드웨어 가속기 - 경량 뉴로모픽 알고리즘 설계 - 저전력 DSP 설계		

■ 강연 정보

제 목	엣지 디바이스에서 인공지능 서비스를 위한 경량 인공지능 가속기 및 SoC 구조
Abstract	최근 급증하는 인공지능 서비스를 처리하기 위하여 클라우드 서버에서는 수십억 개의 변수를 가지는 Large Language Model (LLM)과 같은 거대 인공지능 서비스를 구동하는 반면, 엣지 디바이스에서는 경량화된 인공지능 서비스를 구동하는 엣지-클라우드 모델 기반 기술이 발전하고 있다. 본 논문에서는 엣지 디바이스에서 사용할 수 있는 인공지능 가속기 및 SoC 구조를 제안한다. 제안하는 인공지능 가속기는 Network on chip (NoC) 구조로 설계 되었으며, Zero 입력에 대한 연산을 건너 뛰어 하드웨어 에너지 소모를 줄였다. 또한, matmul 연산과 depthwise convolution 연산에서 요구하는 높은 입력 NoC bandwidth 요구를 만족하기 위해 입력 NoC bandwidth를 조절할 수 있게하였다. 제안하는 인공지능 가속기는 AXI-4 64bit 인터페이스로 설계되어 AMBA를 포함하는 SoC 구조에 연결되었으며, Xilinx의 XCKU115 FPGA를 통해 검증하였다.