

Semiconductor

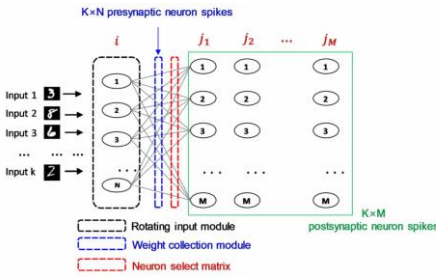
시냅스의 집적도를 높이는 인공지능 반도체 기술

서울대학교 공과대학 박병국 교수

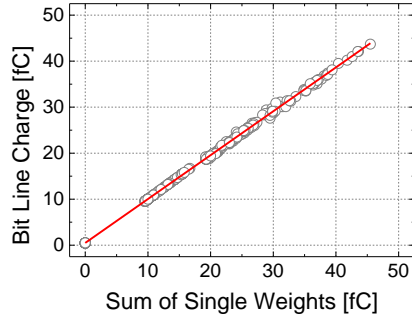
기술내용

- Spiking neural network (SNN) inference system, multi input processing을 하나의 synapse array를 공유하여 가능하게 하는 기술임
- 고집적 대용량 시냅스 어레이를 안정적으로 구동하기 위한 시냅스 어레이 구조를 제한함
- 정전용량을 변화시킬 수 있는 커패시터 기반의 시냅스 소자를 활용하여 구성한 어레이에서의 벡터-행렬곱 연산 방법에 관련한 기술임

주요도면 및 사진



[뉴럴 네트워크 구조도]



[실시에에 따른 벡터-행렬곱 연산 결과 그래프]

기술 개발 배경

- 인공지능 반도체 시장은 반도체 시장보다 가파른 추세로 확장되고 있어 4차 산업 혁명 시대 대응을 위해 연구가 필수적임
- 복잡한 뉴럴 네트워크의 경우, 많은 수의 시냅스와 뉴런으로 인해 고집적화의 필요성이 증가, 이에 따라 short channel effect, 가중치의 multi-level 수 감소 등의 문제 발생
- SNN을 비롯한 뉴럴 네트워크를 하드웨어로 구현할 경우, 온도 변화나 시간의 경과에 따라 시냅스 어레이의 가중치가 변화할 수 있으며, 이는 필연적으로 추론의 정확도를 저하시킴

특장점(효과)

- 유효 채널 길이가 증가함에 따라 시냅스의 전류를 줄일 수 있으며, short channel effect 완화 및 multi-level을 구현하기 용이함
- 기존의 4-terminal NOR-type array에서 게이트 및 드레인 단자를 공유함으로써 cell size를 반으로 줄일 수 있음
- 기존의 시스템보다 input processing 평균 시간 감소로 throughput 증가 가능
- 하나의 공유된 synapse array만 구동하기 때문에 energy efficient system 구현 가능
- 시냅스 소자의 구현 방법의 제약 없이 정전 용량의 변화가 가능한 모든 커패시터 어레이에 적용이 가능함
- 저항이 큰 커패시터를 활용하여 sneak path의 억제가 가능해 백터 · 행렬곱 연산의 에러를 줄일 수 있음

기술활용분야

- 인간의 두뇌 내부에서 뉴런이 수행하는 것과 유사한 방식으로 정보를 처리하는 기술이 필요한 모든 산업 분야

응용분야 및 적용제품	관련 업체
<ul style="list-style-type: none"> • 응용분야 - 뉴로모픽 컴퓨터, 뉴로모픽 소프트웨어, 모바일 프로세서 등 뉴럴 네트워크 기술이 적용되는 산업 분야에 활용 • 적용제품 - CUDA 기반 플랫폼, OpenCL 기반 플랫폼 등 소프트웨어 플랫폼과 이를 운영하기 위한 운영 체제 기술 개발 분야에 적용 - Tensorflow, Caffe, PyTorch 등 	<ul style="list-style-type: none"> • 뉴럴 네트워크 활용 기업 (데이터 마이닝, 네트워크 관리, 언어 인식, 금융서비스, 머신 비전 등)

기술개발단계



지식재산권 현황

No.	기술명	출원번호	등록번호	국가
1	스파이크 신호를 병렬처리하는 뉴로모픽 장치	10-2021-0159710	-	KR
2	하드웨어 기반 인공 신경망 제공 장치	10-2021-0154893	-	KR
3	오버패스형 채널을 포함하는 반도체 소자	10-2021-0159121	-	KR

기술이전 상담 및 문의: 서울대학교 산학협력단 신양일 변리사 ✉ youmei21@snu.ac.kr ☎ 02-880-2026