

글로벌 ICT 표준 컨퍼런스 2022

Global ICT Standards Conference 2022

2022. 11.9.(수)~11.(금)
서울 양재 엘타워 오르체홀(5F)

부대행사명

글로벌 기술패권 대응을 위한 AI 반도체 표준화 현황 및 전망

안성덕, 책임연구원, 한국전자통신연구원

INDEX

01 AI 반도체의 개요

02 AI 반도체의 정책동향

03 AI 반도체의 산업현황

04 AI 반도체의 기술현황

05 AI 반도체의 표준화동향

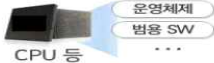
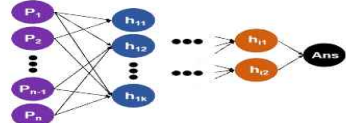
06 맺음말

01. AI 반도체의 개요

AI 반도체의 정의

AI 서비스를 구현하기 위해 요구되는 데이터 및 알고리즘을 효율적으로 처리할 수 있는 반도체

- 학습·추론 기술을 구현하기 위한 데이터 연산 처리를 저전력으로 고속 처리하여 효율성에 특화된 반도체

구분	기존 반도체	인공지능 반도체
기능	 운영체제 범용 SW ... 범용 목적 (단순한 인지 수준으로 제약)	 AI SW(학습모델) 학습 AI반도체 추론 빅데이터 AI 서비스 인공지능 최적화 (복잡한 상황 인식·판단 등 가능)
기술 특징	데이터를 프로그램대로 순차적 처리 	대량의 데이터를 동시(병렬) 처리 

AI 반도체의 구분

(시스템 구현목적 측면) AI 학습용(Learning)과 추론용(Inference)으로 구분

- 학습용의 경우, 방대한 데이터를 통해 딥러닝 등의 알고리즘을 이용하여 지식을 습득하는 학습 단계에 적합
- 추론용의 경우는 외부 명령을 받거나 상황을 인식하면 학습한 내용을 토대로 가장 적합한 결과를 도출하는 추론 단계에 적합

(서비스 플랫폼 측면) 데이터센터 서버용과 엣지 디바이스용으로 구분

- 데이터센터 서버용 AI 반도체는 무엇보다도 병렬연산 처리 능력과 전력효율이 중요하며, 운영측면에서 확장성과 유연성도 고려
- 엣지 디바이스용 AI 반도체는 자율주행차, 드론, IoT 등 개별 AI 서비스에 특화되어 초저전력, 경량화, 제조원가 등이 핵심 경쟁력

(기술구현 방식 측면) FPGA, ASIC/ASSP로 구분

- FPGA는 HW를 재설계하지 않고 프로그래밍을 통해 사용 목적에 맞게 프로그래밍을 통해 변경이 가능한 반도체
- ASIC은 어플리케이션의 특성에 적합한 AI 시스템을 구현하기 위해 범용 프로세서를 사용하는 대신 특정 목적으로 제작되는 주문형 반도체
- ASSP는 범용 ASIC으로서, 특정한 사용자의 요구나 주문에 의해서 만들어진 ASIC이 평판 등에 따라 특정용도 표준으로 되어 여러 사용자가 사용하는 되는 주문형 반도체

01. AI 반도체의 개요

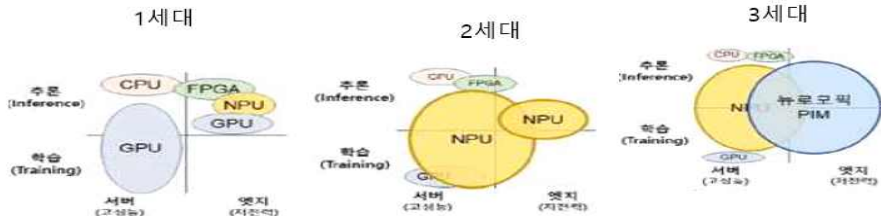
AI 반도체의 발전전망

AI 반도체는 1세대(CPU, GPU), 2세대(NPU), 3세대 (뉴로모픽) 반도체로 진화

(1세대 AI 반도체) GPU는 딥러닝 등 AI 알고리즘 수행에 요구되는 데이터 연산 처리에 있어서 CPU 대비 높은 성능 향상을 제공해 AI 발전을 견인

(2세대 AI 반도체) NPU는 CPU와 연계하여 병렬로 작동되어 시스템 혼란이나 병목현상 없이 사용 가능하며, 최근 AI 구현을 위한 기술로 주목받고 있다. DNN ASIC은 비용이 높고 개발 기간이 길며 한번 제품을 만들고 나면 기능을 바꿀 수 없지만, 이미 개발이 끝난 알고리즘의 전력소모량을 줄이고자 할 때 효율적

(3세대 AI 반도체) 뉴로모픽은 뉴런 역할을 하는 코어를 사람의 뇌처럼 병렬로 구성하여 저전력으로 인간의 뇌처럼 학습하기 때문에 연산 성능이 대폭 향상



02. AI 반도체의 정책동향

글로벌 정책동향

각국의 AI 반도체 정책은 크게 ① 기술개발, ② 인프라 확보, ③ 인재양성

(기술개발) AI 반도체 기술개발은 다양한 정보를 빠르고 정확하게 판단하기 위한 HW와 SW의 기술개발임
HW적으로 AI 기술을 구현하기 위해 NPU나 뉴로모픽과 같은 전용 AI 반도체를 개발하는데 정책 과제를 포함하여 대규모의 투자를 진행하고 있으며, AI 반도체를 구동하기 위한 전용 SW도 대규모의 지원 정책이 진행됨

(인프라확보) AI 반도체가 고속·고성능·저전력 반도체 특성이 요구되는 만큼 이를 제조하기 위한 인프라(제조 시설)에 대한 투자 유인책이 진행 중임

미국이 자국 내 반도체 기업을 유치하기 위해 약 100조원을 지원하는 입법을 추진한 것이 대표적인 지원책임
AI 반도체를 자국 내에서 안정적으로 생산하기 위한 기업의 반도체 제조시설(Fab, 팹) 건설(투자)에 대한 여러 가지 지원 정책을 추진하고 있는데, 반도체 제조시설을 건설하기 위해서는 대규모의 전기, 용수, 도로, 부지 등이 필요하고 이것들이 대표적인 인프라임

(인재양성) AI 반도체 시장에서의 경쟁력을 확보하기 위해서는 AI 반도체 시장에서 필요한 인력을 미리 양성하고 산업계에 유입하는 정책이 필수적임

AI 반도체는 다양한 융합 전공을 습득한 인재가 요구되고, 고급 전문 인력이 필요한 만큼 각국의 정부에서는 '핵심 인력 = 국가기술경쟁력'이라는 비전하에 대규모 정책적 지원을 함

02. AI 반도체의 정책동향

국내의 AI 반도체 산업 성장 지원대책



02. AI 반도체의 정책동향

AI 반도체 국내 R&D 현황

반도체 R&D

과제 건수 (연구비금액 : 단위 백만원)

AI 반도체 R&D

구분	2018년	2019년	2020년	2021년	2022년	계
국공립연구소	2 (180)	1 (150)	1 (15)	2 (383)	67 (22,355)	73 (23,083)
출연연구소	286 (225,990)	311 (241,743)	400 (349,093)	429 (353,847)	331 (354,240)	1,757 (1,542,913)
대학	1623 (308,320)	2,089 (334,925)	2,162 (351,749)	2,334 (446,942)	2,064 (411,703)	10,272 (1,853,639)
대기업	9 (32,600)	9 (23,231)	8 (17,029)	15 (36,657)	18 (34,007)	59 (143,524)
중견기업	38 (36,778)	43 (42,838)	69 (25,144)	137 (139,455)	107 (143,933)	394 (388,148)
중소기업	952 (338,008)	899 (350,828)	1,180 (451,954)	1,279 (582,266)	1,206 (615,622)	5,516 (2,338,679)
정부부처	-	-	-	-	2 (2,470)	2 (2,470)
병원	-	-	-	-	5 (747)	5 (747)
기타	76 (68,687)	79 (85,594)	109 (101,275)	133 (125,318)	101 (85,358)	498 (466,231)

구분	2018년	2019년	2020년	2021년	2022년	계
국공립연구소	-	-	-	1 (140)	3 (845)	4 (985)
출연연구소	18 (18,111)	29 (30,236)	43 (37,070)	56 (49,581)	69 (67,641)	215 (202,640)
대학	95 (24,437)	167 (33,442)	250 (60,997)	353 (103,836)	353 (107,488)	1,218 (330,200)
대기업	1 (10,400)	1 (13,790)	2 (10,106)	2 (7,350)	2 (10,980)	8 (52,626)
중견기업	4 (2,592)	5 (4,888)	5 (1,755)	10 (20,735)	18 (25,561)	42 (55,531)
중소기업	30 (12,442)	51 (24,782)	91 (51,951)	128 (88,569)	142 (116,307)	442 (294,050)
정부부처	-	-	-	-	-	-
병원	-	-	-	-	2 (494)	2 (494)
기타	3 (7,161)	3 (8,237)	5 (9,858)	5 (14,198)	3 (3,448)	19 (42,903)

03. AI 반도체의 산업현황

응용 분야별 AI 반도체 시장전망

- 시장 조사 전문기관인 Gartner에 따르면 '21년 AI 반도체 시장규모는 약 347억 달러 규모이며, '26년에는 860억 달러 규모로 연평균 20% 가까이 성장할 것으로 전망
- AI반도체는 스마트폰과 컴퓨터 및 서버를 중심으로 한 응용 분야가 전체 시장의 약 80%를 차지

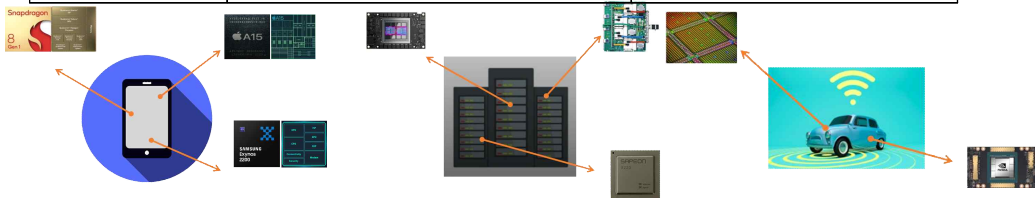
AI 반도체 응용 분야	2020	2021	2022	2023	2024	2025	2026	연평균 성장률
자동차	1,740	2,746	3,422	4,460	6,171	7,811	9,534	28.3%
통신	14,994	21,597	25,630	28,750	31,491	31,948	32,645	8.6%
컴퓨터 및 서버(CPU)	5,119	9,943	14,124	19,261	24,747	30,089	35,353	28.9%
소비가전	87	262	728	1,554	2,131	2,932	3,922	71.7%
산업용	84	221	519	1,244	2,035	3,133	4,431	82.1%
저장장치	3	7	17	44	86	130	194	95.8%
합 계	22,026	34,777	44,440	55,314	66,662	76,043	86,079	19.9%

03. AI 반도체의 산업현황

AI 반도체의 활용분야

- AI 반도체는 스마트폰, 데이터센터, 자동차, IoT 가전, 산업용 등 다양한 분야에서 활용

AI반도체 응용분야	해외기업	국내기업
통신	애플(美), 퀄컴(美), 미디어텍(臺), 화웨이(中) 등	삼성전자 등
컴퓨터 및 서버	Google(美), MS(美), IBM(美), 엔비디아(美), 그래픽코어(英) 등	SAPEON, 퓨리오사ai, 리벨리온, 딥엑스 등
자동차	엔비디아(美), 모빌아이(인텔, 美), 테슬라(美) 등	삼성전자, 텔레칩스, 넥스트칩, 딥엑스, 모빌린트 등



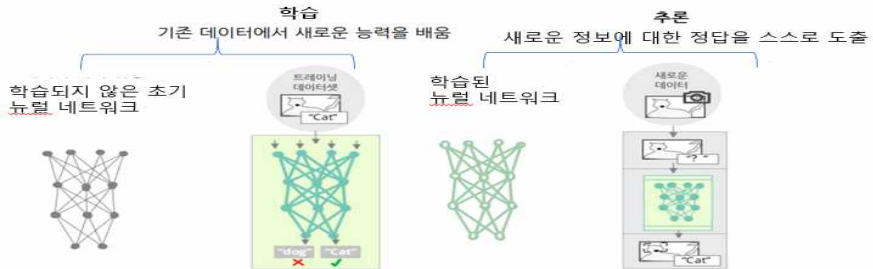
스마트폰과 AI 반도체

컴퓨터 및 서버와 AI 반도체

자율자동차와 AI 반도체

AI의 두 가지 핵심 기능인 학습과 추론

차세대 컴퓨터가 주어진 이미지를 인지하고 분류하는 추론 작업과 사람과 같이 스스로 배우고 판단할 수 있는 학습 작업을 수행할 수 있도록 AI 반도체 기술이 발전



AI의 반도체 발전방향

(1세대 AI 반도체 기술)

CPU 외에도 그래픽 영상 처리 및 고사양 3D 게임 등에 활용되는 그래픽 처리 장치 GPU를 활용함.
연산을 수행할 수 있는 코어 부분을 세분화시켜 명령어를 동시에 병렬적으로 처리할 수 있도록 설계됨.

(2세대 AI 반도체 기술)

동일한 실리콘 반도체 소자를 이용하여 AI 기술에 특화된 칩을 만들기 위해 반도체 회로 구성과 배치를 최적화함.
AI 알고리즘을 빠르게 가속화 시킬 수 있도록 특정 목적을 위해 개발된 ASIC을 뉴럴 프로세싱 유닛(NPU)이라 지칭함.

(미국 Google 회사가 개발한 텐서 프로세싱 유닛(TPU), 미국 테슬라의 자율주행 칩, 미국 애플의 노트북 발열을 획기적으로 제어한 M1 칩)

(3세대 AI 반도체 기술)

인간 두뇌에서의 정보 처리 과정에서 영감을 받은 뉴로모픽 방식임.
신경계에서 정보가 저장된 생물학적 시냅스에서 연산이 동시에 수행되기 때문에 인간은 고도의 작업을 수행하면서도 약 20와트(Watt) 정도의 전력만을 사용함.
에너지 효율적인 구조와 기능을 전자소자로 모사하기 위하여 저장과 연산을 동시에 수행하는 메모리 반도체 소자 기술이 중요하고, 두뇌 모사 구조에 적합한 신개념 비휘발성 메모리 소자가 도입되어 가장 진보된 형태로 AI 반도체 기술이 발전하고 있음



04. AI 반도체의 기술현황

1세대 AI 반도체

(CPU 기술)

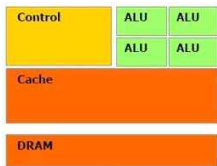
CPU는 외부 명령을 실행하고 제어하는 Control Unit, 덧셈과 뺄셈과 같은 연산을 담당하는 ALU와 계산을 위해 정보를 저장할 수 있는 메모리(Cache 혹은 DRAM)로 구성

(GPU 기술)

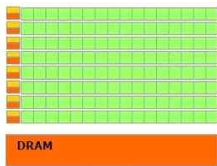
GPU는 Graphic 처리를 위한 병렬처리 성능을 높일 수 있는 SIMT 구조를 가지고 있으며, 이는 인공지능망의 Matrix, Convolution 연산을 병렬처리를 가속할 수 있게 인공지능망 처리에 활용

(SRAM 기반 온칩 및 Near 메모리 기술)

거대 인공지능망 뿐만 아니라, 빅데이터 등 다양한 HPC 어플리케이션에서 높은 용량의 온칩 메모리를 요구함에 따라, 데이터센터 급 프로세서에서 온칩 메모리 용량을 높이기 위한 시도



CPU



GPU

04. AI 반도체의 기술현황

2세대 AI 반도체

(NPU 기술)

광범위한 응용 분야 중 특정 목적만을 위해 맞춤형으로 설계된 ASIC을 AI 알고리즘에 최적화한 AI 알고리즘을 빠르게 가속시킬 수 있는 전용 반도체

(Processing-In-Memory 기술)

메모리 기반 로직 칩 (로직이 반영된 1세대 PIM, 메모리와 로직이 융합된 2세대 PIM)

(HBM 기술)

삼성전자, AMD, 하이닉스의 3D 스택 방식으로 DRAM을 적층하고 2.5D Advanced Package를 통해 고성능 RAM 인터페이스를 가지는 메모리로 고대역폭 데이터 처리를 요구하는 반도체와 결합하기 위해 사용

TPU v4

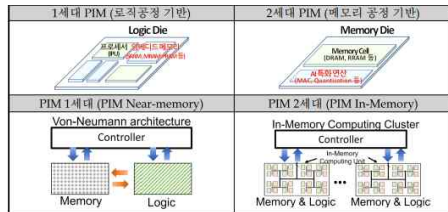


D1 칩

타일의 Seamless한 연결



- 7nm 공정, 645 mm²
- 500억개 트랜지스터
- BF16 연산에 대해 362 TFLOPs



04. AI 반도체의 기술현황

3세대 AI 반도체

시냅스 소자	장점	단점	응용 분야 (플레이어)	
			메모리	AI 반도체
SRAM	공정 미세화/ 신뢰성	0과 1 이진법/ 낮은 메모리 밀도	3 나노미터 이하 트랜지스터로 구성 (대만 TSMC, 한국 삼성전자, 미국 인텔)	인텔 4 나노미터 공정으로 제작, 1억 2천개의 시냅스 구성된 뉴로모픽 칩 (미국 인텔) 삼성전자 28 나노미터 공정으로 제작, 약 2억 6천개의 시냅스로 구성된 뉴로모픽 칩 (미국 IBM)
MRAM	고속 동작 /균일도	0과 1 이진법 /미세화 공정	Mbit 용량 수준의 임베디드 메모리 (한국 삼성전자)	64×64 크기의 뉴로모픽 칩 (한국 삼성전자)
PRAM	이진법 이상의 멀티레벨	높은 동작 전류	Gbit 용량 수준의 고집적 메모리 (미국 인텔)	약 16만개의 시냅스로 구성된 뉴로모픽 칩 (미국 IBM)
RRAM	이진법 이상의 멀티레벨/ 미세화 공정	특성 불균일도	Mbit 용량 수준의 임베디드 메모리 (대만 TSMC, 일본 파나소닉/ 대만 UMC)	256×256 크기의 뉴로모픽 칩 개발 (미국 스탠포드 대학/중국 칭화대)

05. AI 반도체의 표준화 현황

TTA 지능형반도체 프로젝트 그룹(PG417)

AI 반도체 분야에 대한 표준은 2019년도부터 스파이킹 신경망 기반의 인터페이스 표준, AI 프로세서 메모리 인터페이스 표준 및 차량용 반도체 관련 표준 개발

표준번호	표준명	완료년도
TTAK.KO-10.1144	지능형 반도체 적용을 위한 스파이킹 신경망 하드웨어 인터페이스	2019
TTAR-10.0111	인공지능 프로세서의 보안성 강화를 위한 안전성 플랜(기술보고서)	2019
TTAR-10.0112	인공지능 프로세서의 기능안전 인터페이스(기술보고서)	2019
TTAR-10.0113	다중 칩 구성이 가능한 인공지능 프로세서 구조(기술보고서)	2019
TTAR-10.0114	다중 칩 기반 인공지능 프로세서 소프트웨어 인터페이스(기술보고서)	2019
TTAK.KO-10.1229	차량용 반도체의 직렬 주변기기 인터페이스 버스를 위한 순환 중복 검사 방법	2020
TTAK.KO-10.1301	차량용 반도체의 집적회로간 버스를 위한 순환 중복 검사	2020

05. AI 반도체의 표준화 현황

인공지능 반도체 포럼

인공신경망, 딥러닝, 머신러닝에 관련된 단체 표준을 개발

표준번호	표준명	완료년도
ISFS-2020-0001	인공 신경망 압축기술에서 균등 양자화 기법	2020
ISFS-2020-0002	인공 신경망 압축기술에서 낮은 계수 근사화 기법	2020
ISFS-2020-0003	딥러닝 네트워크 양자화를 위한 비트스트림 포맷	2020
ISFS-2020-0004	딥러닝 커널 및 활성화 데이터의 압축을 위한 엔트로피 코딩 및 비트스트림 포맷	2020
AIFS-2021-0001	컴퓨터 비전을 위한 비디오 코딩 요구사항	2021
AIFS-2021-0002	인공 신경망 데이터 압축에서 특징맵 재배열 방법	2021
AIFS-2021-0003	머신러닝 부호화기술에서 물체 검출/분할 부호화 방법	2021

05. AI 반도체의 표준화 현황

IEC TC 47

AI 반도체 관련 표준은 WG6에서 저항 메모리 관련 성능 측정 표준 1개와 자동차용 신뢰성 테스트의 표준화 로드맵 보고서 1건이 진행

표준번호	표준명	완료년도
IEC 62951-9 ED1	Semiconductor devices - Flexible and stretchable semiconductor devices - Part 9: Performance testing methods of one transistor and one resistor (1T1R) resistive memory cells	2023.02
IEC TR 63357 ED1	Semiconductor devices - Standardization roadmap of fault test method for automotive vehicles	2022.10

05. AI 반도체의 표준화 현황

ISO TC22 SC32 WG8(Electrical and electronic components and general system aspects)

ISO TC22 SC32 WG8은 자동차 기능안전 WG으로 자동차 전자제어장치의 고장으로 인한 위험을 제거/방지/완화하기 위한 설계/분석/시험 방법 등 최신 기술을 표준화

표준번호	표준명	완료년도
ISO 26262-11	Road vehicles — Functional safety — Part 11: Guidelines on application of ISO 26262 to semiconductors	2018.12
ISO 21448:2022	Road vehicles — Safety of the intended functionality	2022.06

05. AI 반도체의 표준화 현황

JEDEC JC-42(Solid State Memories)

JC-42에서는 제조 기술이나 응용 프로그램에 관계 없이 정적이든 동적이든 모든 memory integrated circuit과 programmable logic device(예를 들면 SRAM, DRAM, ROM, EEPROM 및 PLD)의 핀 아웃, 읽기 및 쓰기 알고리즘을 포함한 작동 특성, 테스트 매개변수, 특성화 및 등록 형식과 관련된 기술 정보 및 표준 개발

표준번호	표준명	완료년도
JESD235D	High Bandwidth Memory DRAM(HBM1, HBM2)	2021
JESD238	High Bandwidth Memory DRAM(HBM3)	2022

Item	Description	Standardization Areas
Execution Model	Both SIMD and MIMD	PIM Configuration and programming
Execution Management	Starting/Stopping/Status etc.	Enhancements to existing R/W commands or new commands
Controller-PIM Arbitration	For accessed to shared memory	New interface protocol
PIM Addressing	Linear address range per PIM DRAM	Define single DRAM channel
Media Errors	PIM DRAM on-die ECC	Specify error detection/correction requirements
Channel Errors	R/W CRC	Tweak existing DDR CRC if needed
Media Mgmt	Refresh for retention/row hammer etc.	PIM DRAM based scheme. Look for common solution with non-PIM DDR

- AI 반도체는 현재의 세계의 반도체 시장을 새롭게 재편할 수 있는 기회를 제공하고 있으며, 체계적이고 전략적인 기술 및 산업육성을 통해 우리나라의 미래 핵심 성장동력으로 예상
- AI 반도체는 현재 스마트폰과 일부 가전에 주로 도입되고 있지만, 조만간 자율주행차, 지능형 로봇 등 다양한 산업 분야로 빠르게 확산되어 우리나라 반도체 산업의 새로운 성장 기회를 제공할 것으로 예상
- (기술적인 측면) AI 반도체 후발국인 우리나라는 혁신적인 창의적이고 선도적인 AI 반도체 기술을 개발하여 미래 신시장을 선점하는 전략이 필요
- (정책적인 측면) AI 반도체 시장 선점을 위해서는 국가 주도의 체계적이고 전략적인 산·학·연 협력 R&D와 수요기업-팹리스-파운드리 가치사슬 구축 및 AI 반도체의 핵심 경쟁력인 설계 역량 확보를 위해 고급인재의 유치 및 육성이 필요
- (표준적인 측면) 차세대 AI 반도체의 R&D, 제품 및 서비스에 필요한 표준 구축에 대한 장기적인 로드맵 확보 및 이에 따른 표준 제정 전략이 필요
단기적으로는 2세대 AI 반도체인 AI 프로세스를 위한 인메모리 관련 선도 표준 개발 및 로드맵
중장기적으로 3세대 AI 반도체인 뉴로모픽 소자 관련 선행 표준 개발 전략 및 로드맵

Thank you

안성덕, 책임연구원, 한국전자통신연구원
lovesky@etri.re.kr