

2021. 07

유망시장 Issue Report

AI 반도체

INNOPOLIS
연구개발특구진흥재단



목 차

I . 개요	01
II . 정책 동향	06
III . 기술 동향	18
IV . 시장 동향	31
V . 산업 동향	36

I. 개요

1. 아이템 개요

○ 인공지능 반도체란 인공지능 알고리즘을 효율적으로 계산할 수 있는 반도체를 의미

- 인공지능 기술의 핵심 기술 중 학습·추론 기술을 구현하기 위한 데이터 연산 처리를 저전력, 고속 처리하여 효율성에 특화된 반도체를 의미

[표] 기존 반도체와 인공지능 반도체 비교

구분	기존 반도체	인공지능 반도체
기능	범용 목적 (단순한 인지 수준으로 제약)	인공지능 최적화 (복잡한 상황 인식·판단 등 가능)
기술 특징	데이터를 프로그램대로 순차적 처리	대량의 데이터를 동시(병렬) 처리

※ 출처 : 인공지능 반도체 육성으로 인공지능과 종합 반도체 강국 시대 연다, KIPA, 2020

○ 인공지능 시스템의 구현 목적에 따라 크게 학습용(Training)과 추론용(Inference)으로 구분되며 두 가지 과정을 반복 실행해 최적의 답을 찾도록 성능 강화

[표] 인공지능 반도체의 분류

시스템 구현 목적	학습용		추론용
서비스 플랫폼	서버용		엣지 디바이스용
기술 구현 방식	기존 반도체 진화형	1세대 AI 반도체	2세대 AI 반도체
	CPU/GPU/FPGA (예 : 인텔, 엔비디아, 자이링스)	ASIC/ASSP (예 : 인텔, 구글)	뉴로모픽 프로세서 (예 : IBM)

※ 출처 : 인공지능 프로세서 반도체 기술 및 표준화 동향, ETRI, 2020

○ AI 반도체는 기술적으로 3가지 유형으로 분류 가능

- AI 반도체 유형마다 장단점이 존재하고 AI 반도체가 적용되는 분야 특성에 따라 선호도가 달라, 단지 인공지능 연산 성능이나 진화 단계가 높다고 하여 시장에서 경쟁력이 높은 것은 아님

- (기존 반도체 진화형) CPU · GPU · FPGA 등이 이에 해당되며, 인텔 · 엔비디아 · 자일링스 등의 업체가 대표적
 - 상대적으로 가격이 싸고 유연성(Flexibility)이 높다는 것이 장점이나, 인공지능 연산 성능(Performance)과 소비전력 효율(Power Efficiency)이 낮다는 것이 단점
- (1세대 AI 반도체) 인공지능 연산 고속화를 위해 반도체 구성을 최적화시킨 ASIC/ASSP가 이에 해당되며, 구글 · 인텔(모빌아이, 모비디우스) 등의 업체가 대표적
 - 기존 반도체 진화형의 반대 경우로 인공지능 연산 성능과 소비전력 효율이 높지만, 가격이 비싸고 유연성이 낮아 디자인된 알고리즘으로만 사용할 수 밖에 없는 것이 단점
- (2세대 AI 반도체) 인간의 뇌를 모방한 非폰노이만 방식의 뉴로모픽 반도체가 현재까지는 가장 진보된 형태의 AI 반도체로 평가
 - 인공신경망 연산 성능과 소비전력 효율은 AI 반도체 가운데 가장 뛰어나지만, 아직은 기술성숙도가 낮고 폰 노이만 구조를 사용하지 않기 때문에 범용성이 낮은 것이 단점
 - IBM TrueNorth의 경우, 초당 46억 회 실행되는 시냅스의 동작을 70~200mW의 보청기 수준의 소비전력으로 수행



[그림] AI 반도체 진화 단계 및 특성 비교

※ 출처 : 반도체 산업의 차세대 성장엔진 AI반도체 동향과 시사점, IITP, 2018

- 방대한 양의 데이터 연산 처리로 인한 발열 및 전력 소모로 인해 인공지능 반도체의 효율 개선 필요
 - 현재 인공지능 학습·추론은 대부분 데이터센터에서 실행됨
 - 인공지능 서비스에 요구되는 대규모 연산 처리 성능을 확보하기 위해 인공지능 반도체를 서버에 장착하여 활용
- 데이터센터 서버(클라우드)와의 연결 최소화 및 디바이스 자체에서 인공지능 연산이 가능하도록 소형화·저전력·고성능 중심의 인공지능 반도체 개발 가속화
 - 스마트폰, IoT 기기 등의 보급 확산 및 클라우드 기술의 발전에 따라 디바이스의 추론 기능에 대한 수요 증가

2. Value Chain

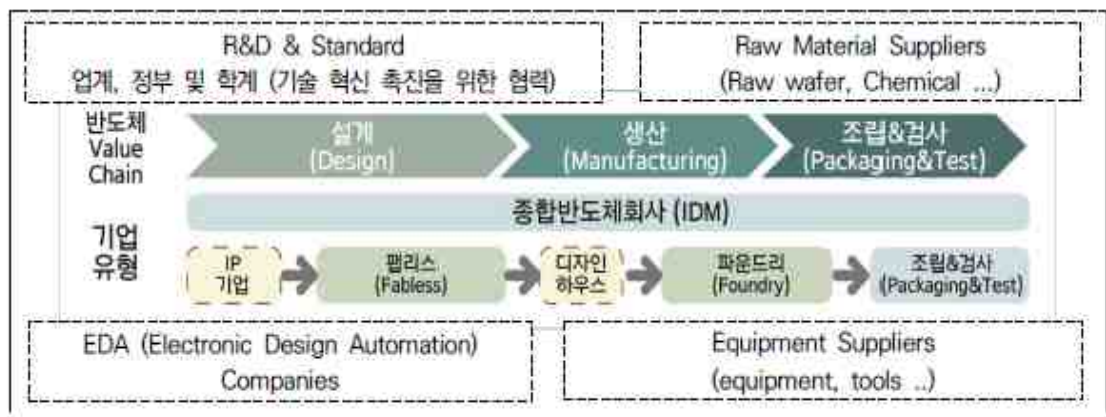
- AI 생태계와 반도체 생태계가 만나는 접점에 AI 하드웨어인 AI 반도체가 존재하며 역할의 중요성 증대

[표] 인공지능과 반도체 생태계 관계

AI 생태계	AI 서비스 및 앱			
	AI 플랫폼			
	AI frameworks Tools & Interfaces			
	AI 라이브러리			
AI 반도체	AI 하드웨어			
반도체 생태계	R&D	설계	생산	조립·패키징·테스트

※ 출처 : 인공지능 기술의 진화와 AI 반도체·컴퓨팅의 변화, 정보통신정책연구원, 2019

- 반도체 가치사슬에서 지원 활동을 수행하는 전문회사를 중심으로 AI 반도체 등장과 수요 증가에 따른 변화가 진행
 - 전문회사는 특히 기존 가치사슬에서 원가, 품질, 성능 및 수요의 선호도 등에 있어서 경쟁력 있는 반도체를 생산하는 데 기여



주: 점선으로 처리된 영역이 전문회사 영역

[그림] 반도체 Value Chain 및 기업 유형

※ 출처 : 인공지능 기술의 진화와 AI 반도체·컴퓨팅의 변화, 정보통신정책연구원, 2019

- 현재 반도체 생태계를 지원하는 전문회사는 아래의 유형의 기업들이 존재
 - 반도체를 설계하는 SW인 EDA를 제공하는 전자 설계 자동화 공급기업
 - 제조·조립·테스트 및 포장을 위한 특수 장비 및 공작 기계를 생산하는 장비 제조업체
 - 웨이퍼 제조, 각종 케미컬, packaging 재료 등을 제공하는 원재료 공급기업
 - 반도체 회사가 자체 칩을 설계할 때, 외부에서 사전에 디자인된 설계 블록을 사용할 수 있도록 이를 개발하고 라이선스를 부여하는 IP 기업

- (EDA 공급기업) 인공지능 반도체를 설계할 수 있는 솔루션을 제공하며
 현재 EDA 시장은 Synopsys, Cadence, Mentor 등의 기업이 시장 과점
 - AI 반도체를 구현할 수 있도록 새로운 설계 방법론과 설계 툴을 팹리스와 파운드리에게 제공

- (IP 기업) 전문적인 표준화된 설계자산(IP)을 개발하여 제공하는 대표적인 IP 기업인 ARM, Synopsys, Cadence의 경우 AI 반도체를 위한 IP를 제공
 - 주로 AI 작업만을 담당하는 단독정보다는 AI 처리기능이 시스템 반도체에 내장된 형태인 통합형을 위한 IP 코어 제공

- 설계자산(IP)을 공개하여, 기존의 상용 IP와는 달리 오픈소스처럼 무료로 사용할 수 있고, 수정할 수 있는 개방형 IP인 리스크파이브(RISC-V)가 기존 사업의 영역을 확장
 - 리스크파이브(RISC-V)를 활용하여, 시스템 반도체 또는 AI 반도체를 더 쉽고 빠르게 낮은 비용으로 만들 수 있도록 지원하여 수익을 창출하는 스타트업 탄생

- (클라우드 서비스 제공 기업) 대형 데이터센터를 보유하고 있는 구글, 마이크로소프트, 아마존 등의 인터넷 기업들을 중심으로 AI 서비스를 위한 빅데이터 및 연산 처리와 더불어 차별화를 위하여 직접 AI 반도체 개발 착수

II. 정책 동향

1. 국내 정책 동향

- 우리나라 정부는 인공지능 반도체 분야 핵심원천기술 개발 및 인프라 지원을 위한 범부처 차원의 계획을 마련하였으며, 관련 부처별 중장기 지원방안을 수립하여 기술경쟁력 확보를 위한 정책을 추진 중

가. 시스템 반도체 산업경쟁력 강화방안

- 2017년 산업통상자원부에서는 시스템 반도체 산업 선도국으로 도약하기 위해 3가지의 정책과제를 제시

- ① 저전력·초경량·초고속 반도체 설계기술 확보, ② 반도체 수요·공급 협력, ③ 반도체 설계·생산 컨소시엄 구성 등을 위한 정책과제를 제시

- (공급-기술) 3대 유망 기술개발 등에 민·관 합동으로 2,645억 원 투자

- (연구개발) 저(低)전력-초(超)경량-초(超)고속 반도체 개발을 위해 시스템 반도체 연구개발
- (인력양성) 차량용 반도체 석사과정 신설 등을 통하여 시스템 반도체 개발 전문 인력 4년간 총 2,880명 양성
- (소재·공정) 정부·기업 1:1 공동투자로 차세대 반도체 소재·공정 원천기술 개발
- (민간투자) 증가하는 낸드 수요 대응을 위해 민간주도로 낸드 기술 확보 및 적기에 투자가 이행될 수 있도록 정부 합동 투자 지원반 운영

- (수요-시장) 신수요·신시장 창출을 위한 협업프로젝트 추진

- 사물인터넷(IoT) 플랫폼 확산 및 자동차·가전·헬스케어 수요 창출

- (생태계) 설계·생산 협력체계 및 성장 인프라 구축

- (플랫폼) 다품종·소량생산 수요 대응을 위해 디자인하우스를 중심으로 설계·생산기업 컨소시엄(Virtual Fab) 구성

- (성장인프라) (창업) 반도체 설계 공용 랩(LAB) 구축 → (성장) 반도체 펀드 2,000억 원 투자 → (회수) 인수합병(M&A) 지원단 운영으로 반도체 설계 기업 성장의 연결고리 강화

나. 혁신성장동력 시행계획

- 2018년 05월, 관계부처 합동으로 4차 산업혁명 대응을 선도할 혁신 성장 동력 13대 분야를 선정하였으며, 이 중 인공지능 반도체 분야는 기술개발, 인프라 구축, 인력양성을 주축으로 지원하여 인공지능 시대의 핵심부품 경쟁력 확보를 위한 정책을 수립

핵심 의제	인공지능 시대의 핵심부품 경쟁력 확보	
목표	지능형반도체 핵심기술 개발 및 선순환 생태계를 구축하여 글로벌 경쟁력 확보	
변화상	As Is ('18) · 정보화사회(시스템반도체) · 세계시장 점유율 3% · 고숙실행, 고집적화	To Be ('22) · 지능정보사회(지능형반도체) · 세계시장 점유율 7%달성 · 초지능, 초저전력(초경량)

[그림] 혁신성장동력 시행계획

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

- (연구개발) 超지능, 超저전력 분야를 선정하여 기술적 차별화를 통해 新기술선도와 기업 육성이 가능한 기술 지원
 - 초지능, 초저전력 반도체를 위한 설계 원천기술 개발
 - 新산업 구현을 위해 미래 유망시장에 필요한 지능형 반도체 상용화
 - 국내 주력산업인 자동차, 바이오, 가전, 에너지, 첨단기계·로봇 분야와 협력체계를 구성하고 AI, IoT 기술을 활용한 상용화 지원
- (인프라구축) 팹리스-파운드리-학계-연구소 간 생산연계, 기술공유 등 기반 조성을 위한 지원 체계 구축 추진

- 산·학·연 협동 지원을 통한 지능형반도체 기술개발 지원 및 창업생태계 활성화
- SiC 기반 파워반도체용 일괄공정라인 구축(연구개발 지원 플랫폼)을 통해 차세대 파워반도체 연구 지원

로드맵		'18년	'19년	'20년	'21년	'22년
	기술개발	초절전 하이퍼바이저 기반 지능정보 매니코어 프로세서 및 SW 기술개발				
		시냅스 소자기반 패턴인식 하드웨어 시스템 개발				
	인프라구축	파워반도체 연구개발 지원 플랫폼 구축				
			지능형반도체 지원 체계 구축			

[그림] 혁신성장동력 시행계획

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

- (해외 진출 지원) ‘한·중 시스템IC 협력연구원’ (중국 심천) 등 지역거점을 확대·활성화하고 마케팅 등 해외 진출 지원 확대
 - (지역거점 활성화) 한·중 시스템IC 협력연구원을 통해 해외수요 발굴, 현지 수요연계형 R&D, 현지 진출 컨설팅 등 지원 확대
 - (전문전시회) 반도체 전문전시회(SEDEX)를 통해 프로모션 기회 제공
 - (중국·대만) 현지 수요기업 대상 국내 설계기업 제품 공동 마케팅
 - (인도) 현지 모바일·자동차 기업과 국내 팹리스 간 네트워킹
- (기타 : 상생 협력) 반도체 상생발전위원회(2018년)를 통해 반도체 대기업과 중소 반도체 업계 간 협업체계 구축

다. 인공지능 반도체 산업발전 전략 발표

- 2020년 10월, 과학기술정보통신부는 세계적 수준의 인공지능 반도체 기술 및 인재 확보와 선제적·자생적 산업 생태계 구축을 위해 기술 개발, 인력양성, 인프라 지원을 위한 정책을 수립

- 글로벌시장 점유율 확대 및 인공지능 반도체 전문기업 및 고급 인재 육성을 목표로 전략 및 과제를 수립

○ (비전) 인공지능 반도체 선도국가 도약으로 인공지능·종합반도체 강국 실현

- 독자적 기술력 확보(2022년) → 세계적 산업경쟁력 확보(2026년) → 제2의 D램 신화 실현

○ (추진방향) 민간의 혁신 역량과 정부의 전략적 지원으로 국가적 역량 결집

- 인공지능 반도체의 퍼스트 무버 도약을 위한 혁신 기술인재 및 레퍼런스 확보
- 산업경쟁력 조기 확보를 위한 선제적·자생적 산업생태계 구축

구 분	2026년	2030년
글로벌 시장 점유율	10%	20%
인공지능 반도체 혁신기업	10개	20개
인공지능 반도체 고급인재	1,000명(24)	3,000명

[그림] 인공지능 반도체 산업 육성 추진 목표

※ 출처 : 「인공지능 강국」실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020

○ 퍼스트무버형 혁신 기술·인재와 혁신성장형 산업생태계로 2대 전략 분야를 선정하고 각 전략 분야별 6개의 세부과제를 발표

[전략1] 퍼스트무버형 혁신 기술인재	[전략2] 혁신성장형 산업 생태계
◆ 세계 최고 기술력 도전 ① 인공지능 반도체 플래그십 프로젝트 ② 메모리 기반 신개념 반도체	◆ 민간-공공 수요 마중물 창출 ① 수요-공급 연계 시장 조기 성숙 ② DNA 서비스 연계 공공 시장창출
◆ 기술-사업화 장벽 해소 ① 인공지능데이터 인프라 시범 도입확산 ② 맞춤형 Tech Jump-up 프로그램	◆ 연대·협력의 밸류체인 구축 ① 기업간 연대·협력 생태계 조성 ② 파운드리 소부장 강화
◆ 차세대 전문인재 양성 ① 인공지능 반도체 고급인재 확보 ② 실무융합인재 양성 및 저변 확산	◆ 기업 성장 인프라 강화 ① 혁신기업 Scale-up 자금지원 ② 인공지능 팹리스 성장 자원 인프라 확대

[그림] 인공지능 반도체 산업 발전 추진 전략 및 과제

※ 출처 : 「인공지능 강국」실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020

[표] 퍼스트무버형 혁신 기술·인재 확보 전략 세부내용

세계 최고의 기술력 도전	
인공지능 반도체 플래그십 프로젝트 추진	세계 최고 메모리 기반 신개념 반도체 개발
(개발목표) 선도국과의 기술격차 극복을 넘어 세계 최고 수준의 성능 전력효율을 갖는 인공지능 반도체 기술 확보로 글로벌 시장 선도 (개발전략) ①설계·소자·제조 분야별 원천 기술을 개발 상용화하고, ②핵심 기술의 연계 융합을 통해 혁신 기술 및 선도제품 확보 (추진체계) 범부처 '차세대 지능형 반도체 기술개발 사업단'을 구성(2020년), 민간 중심의 체계적인 사업 기획·관리 및 성과확산 지원	(PIM 반도체) ①상용 주력 공정 기반 가시적 성과 창출, ②차세대 메모리(신소자) 공정 기반 원천기술 확보를 위한 Two-Track 추진 (차세대 메모리) 신개념 PIM 반도체 개발·제조 등에 필요한 차세대 메모리(PRAM, MRAM, ReRAM 등) 제조 공정 기술 고도화 추진(2020년~)
기술사업화 장벽 해소	
국가 인공지능데이터 인프라를 통한 시범 도입확산	맞춤형 Tech Jump-up 프로그램 지원
(공공 컴퓨팅 인프라) 인공지능 융합 산업 생태계 조성, 인공지능·데이터 기반 연구 등을 위한 대규모 공공 인프라에 인공지능 반도체 도입·활용(2021년~) (민간 데이터센터) 클라우드 인프라 기업, 통신·포털사 등 민·관 협력을 통해, 국가 R&D 및 국내 팹리스를 통해 개발된 인공지능 반도체의 민간 인공지능 데이터센터 도입·검증 지원(2020년~) (실증 지원) 인공지능 반도체의 공공·민간 인공지능 데이터센터 도입·확산을 위한 실증 지원(2021~2025년)으로 인공지능 칩·서버 활용 촉진	(소프트웨어) 인공지능 반도체 상용화에 필수적이나 국내기업이 취약한 SW 역량 강화 및 생태계 조성을 위한 국가 R&D 본격 추진(2021년~) (맞춤형 기술지원) 인공지능 반도체 유망기업의 기술장벽 해소 및 빠른 글로벌 시장 변화 대응을 위한 맞춤형 지원 프로그램 신설(2021년) (산 학 연 협력) 산업 파급력이 높은 응용분야를 중심으로 기존 원천기술을 활용한 산·학·연 협력 응용기술 연구 지원(2020~2024년) (수요연계 기술실증) 자율주행차·인공지능IoT 등 민간수요와 연계한 중소 팹리스 및 스타트업의 선도제품 개발 및 현장실증 동시 지원
차세대 전문인재 양성	
인공지능 반도체 고급인재 확보	실무융합인력 양성 및 저변 확산
(인공지능 반도체 아카데미) 기업과 정부가 1:1 매칭을 통해 '핵심 기술 개발+고급인력 양성+채용 연계'의 1석 3조 프로젝트 지원 (선도대학 육성) 대학 내 인공지능 반도체 특화 인력양성 센터 확대, 우수 해외 인재 교류 등 석·박사급 고급인재 육성(2020년~) (산업맞춤형 인재 양성) 산업특화형 커리큘럼 운영, 기업 주관 국가 R&D 참여 등을 통해 인공지능 반도체 기업 수요에 맞는 핵심인재 육성(2020년~)	(실무역량 강화 지원) 교육·실습 인프라 강화, 재직자·학부생 대상 교육 프로그램 운영 등을 통해 인공지능 반도체 실무융합인력 양성 (설계 경연) '대한민국 반도체 설계대전'을 통해 대학(원)생들의 인공지능 반도체 설계능력 배양 및 산업계 협력(연구성과 상용화) 유도

※ 출처 : 「인공지능 강국」실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020

[표] 혁신성장형 산업 생태계 활성화 전략 세부내용

민간·공공분야 초기 시장창출	
시장 조기 성숙을 위한 수요-공급 연계 확대 (1社1Chip) 수요기업-팹리스의 수요기업 전용 인공지능 반도체 공동 개발을 지원하는 '1社1Chip 프로젝트'를 통해 2030년까지 인공지능 반도체 50개 개발 (인공지능 핫라인) 융합 얼라이언스 참여 없이도 누구든 수요-공급 연계를 지원받을 수 있는 '인공지능 반도체 핫라인(Hotline)' 구축 (인공지능 칩 성능평가) 인공지능 반도체 상용화를 위해 출연연 및 민간 연구소를 중심으로 시제품 테스트베드를 구축하고, 성능평가 지원 (성과홍보) '인공지능 반도체 포럼'을 개최 (과기정통부, 산업부 공동)하여 인공지능 팹리스의 성과를 홍보하고, 투자자·수요기업 매칭 유도	D.N.A 서비스 혁신과 연계한 공공분야 초기시장 창출 (공공 서비스) 디지털 뉴딜 프로젝트 (인공지능+X, 5G 융합서비스, 헬스케어 등) 추진을 통해 다양한 공공분야에서 인공지능 반도체 적용 확대 (공공 제품) 공공분야 제품 R&D 추진 시 인공지능 반도체 개발 연계 (SW융합 인프라) '인공지능 반도체+SW'를 통해 신제품·서비스 발굴 및 지역산업 혁신을 위한 '충북 SW융합 지능형반도체 클러스터' 조성(2020~2024년)
연대·협력의 밸류체인 구축	
설계 역량 강화를 위한 연대·협력 생태계 조성 (팹리스 ↔ IP기업) 인공지능 IP 기업과 팹리스가 공동으로 R&D에 참여하는 'SoC-IP 패키지형 과제' 추진 (디자인하우스 ↔ 팹리스) 디자인하우스가 인공지능 팹리스의 IP 설계~공정 최적화를 지원하는 '인공지능 IP 토탈 솔루션' 제공 (2021년~) (파운드리 ↔ 팹리스) 국내 파운드리와 인공지능 팹리스 대상 공정·IP 지원	첨단 파운드리 소·부·장 클러스터 조성 (인공지능 반도체 공정혁신 밸리) 10nm 이하 인공지능 반도체 제조 장비를 위한 인공지능 반도체 공정혁신 밸리 신규 구축(2022~2025년) (첨단 공정장비 개발) 첨단 파운드리와 공정 안정성, 생산성 향상을 위해 10nm 이하 초미세 (원자레벨) 공정제어 기술개발 추진 (신소재 개발) 10nm 이하 초미세 공정에 사용되는 소재 기술개발 확대
기업 성장 인프라 강화	
첨단 파운드리 소·부·장 클러스터 조성 (반도체 펀드) 인공지능 반도체 팹리스의 R&D, M&A 등에 총 700억원 투자 (뉴딜 펀드) 정책형 뉴딜펀드 투자 대상에 차세대 반도체(시스템 반도체, 인공지능 반도체)를 포함하여, 산업 생태계 활성화 및 디지털 뉴딜 가속화(2021년~)	인공지능 반도체 팹리스 집중 지원체계 구축 (설계지원센터 강화) 팹리스 종합 지원 플랫폼인 '시스템 반도체 설계지원센터'의 인공지능 반도체 팹리스 창업~성장 통합지원 강화 (인공지능 반도체 혁신설계센터 신설) 설계지원센터 제2캠퍼스로 시스템 반도체의 인공지능화에 대응하는 '인공지능 반도체 혁신설계센터' 조성 (~2022년) (창업지원체계 구축) 글로벌 기업이 보유한 역량을 국내 팹리스에 개방하여 설계부터 생산까지 책임지는 창업지원체계 구축 (지원체계 활용) 규제 완화, 특허 등 팹리스 성장 지원체계 구축

※ 출처 : 「인공지능 강국」 실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020

2. 해외 정책 동향

가. 미국

- 산학연 중심의 중장기 프로젝트를 통해 인공지능 반도체 기초·원천 기술 개발에 주력
- 2008년부터 방위고등연구계획국(DARPA)에서는 뉴로모픽 반도체 개발을 위해 시냅스(SyNAPSE) 프로젝트를 추진하였으며, 산업계(IBM) 주도의 연구기관 및 대학들이 참여하여 2014년 상용화 단계인 뉴로모픽 반도체 (트루노스) 개발함
- DARPA는 그간 마이크로 시스템 기술실(MTO)에서 반도체 R&D 프로젝트를 직접 수행하였으나, 2016년 산·학 컨소시엄 ‘JUMP(Joint University Microelectronics Programme)’ 공동 R&D 프로젝트를 추진
 - JUMP는 유망 반도체 기술들을 선정하여 정부 기관과 민간 기업이 공동 투자하는 SRC(Semiconductor Research Corporation)를 중심으로 응용 연구 중심의 컨소시엄 4개(인지 컴퓨팅, 지능형 메모리 및 스토리지, 분산 컴퓨팅 및 네트워킹, 센서 및 통신 시스템)와 학문 연구 중심의 컨소시엄 2개 (첨단 아키텍처 및 알고리즘, 디바이스 및 패키징, 첨단 소재)를 운영
- 2018년 DARPA는 산·학·연 차세대 반도체 R&D 프로그램인 ‘ERI(Electronics Resurgence Initiative)’를 발표하여 미국의 반도체 산업의 부활(Resurgence)을 위해 무어의 법칙을 극복하기 위한 장기 프로젝트 추진
 - ERI 프로그램은 반도체 설계자동화(EDA) 등 반도체 설계, 신소재, 시스템 설계구조(architecture) 등 3개 분야의 총 6개 프로젝트로 추진되며 2025~2030년 사이 R&D 기간이 끝나면 상용화 작업에 착수하겠다는 계획
 - 참여 기관으로는 미국 내 반도체 글로벌 기업인 Intel, 어플라이드머티어리얼즈, 퀄컴, IBM, 시놉시스 등과 MIT, 스탠포드, 프린스턴, 예일 대학 등 우수 대학 연구소들이 무어의 법칙을 지속하는 목표를 가지고 프로젝트에 참여

[표] 미국 ERI 6개 세부 프로젝트 및 목표

세부 프로젝트 명	목표
소프트웨어 정의 하드웨어 (Software-defined Hardware, SDH)	특화 프로세서는 범용 프로세서보다 작업을 더 빨리 완료할 수 있으나, 가격이 비싸고, 제작 시간이 오래걸림 소프트웨어 정의 하드웨어는 처리 중인 데이터를 기반으로 실시간으로 재구성할 수 있는 하드웨어 및 소프트웨어를 개선하고 제작하는 방법을 연구
도메인 특화 시스템 온칩 (Domain-specific System on Chip, DSSoC)	도메인 특화 시스템 온칩은 개발자에게 ASIC, CPU 및 GPU와 같은 부품을 쉽게 조합할 수 있도록 도와주는 단일 프로그래밍 프레임워크를 개발하는 것을 목표
전자 칩 지능형 설계 (Intelligent Design of Electronic Assets, IDEA)	실제로 칩을 물리적으로 배치하고 레이아웃을 설계하는 것은 매우 복잡하고 고도의 전문성을 필요로 하는 작업으로 IDEA는 일반적으로 1년이 걸리는 칩 설계를 하루로 단축하여 24시간 설계 시대를 여는 것을 목표
Posh 오픈소스 하드웨어 (Posh Open Source Hardware, POSH)	POSH는 오픈 소스로 만들어지는 전문 프로그램이며, 라이선스 제한이 없이 최상의 시스템을 제작하고 테스트하는 것을 목표
3차원 단일 시스템 온어칩 (3-Dimensional Monolithic System-on-a-chip, 3DSoC)	현재 전자공학의 표준은 프로세서와 칩을 중앙 메모리에 연결하는 방식으로 3DSoC는 모든 것을 3차원 방식으로 결합하여 표준 방식보다 적은 전력을 사용하면서 계산 시간을 50배 이상으로 단축하는 것이 목표
FRANC (Foundations Required for Novel Compute)	컴퓨팅 이론의 창립자인 폰 노이만형 컴퓨팅 아키텍처를 뛰어넘는 혁신을 추구함 이 프로그램의 목표는 데이터 이동을 제한하거나 최소화하는 방식으로 데이터를 처리하기 위해 새로운 재료 및 통합 체계를 활용하는 회로를 설계하는 것임 이를 통해 기존 로직 프로세서와 근본적으로 다른 구조로 저장된 데이터를 처리할 수 있도록 함으로써 컴퓨팅 성능을 크게 향상하는 것을 목표로 함

※ 출처 : 「인공지능 강국」실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020

○ 최근 미국 정부는 반도체가 국가 안보에 영향을 주는 ‘전략 산업’ 이라고 인식하고 반도체 기술경쟁력 강화 및 보호 의지를 강하게 표명

나. EU

- 2012년 HBP(Human Brain Project)의 대형 프로젝트를 통해 인공지능 반도체 기술 개발 추진
 - 유럽에서는 독일의 하이델베르크대를 중심으로 ‘브레인 스케일스’ (Brain ScaleS)라는 프로젝트가 진행
 - 최신 뇌 과학 지식을 바탕으로 사람의 뇌의 동작과 구조 등을 슈퍼컴퓨터에 모델링 하여 시뮬레이션을 통해 뇌를 규명하겠다는 목표로 전 세계 100개 이상의 연구기관이 10년 목표로 참여 중
 - 스위스 로잔연방공과대학교 및 로잔의과대학교, 그리고 독일 하이델베르크 대학교가 주축을 이루고 있음
- 사람의 생각 및 행동 등의 상위 수준의 뇌기능이 아닌 개별 뉴런 세포와 같은 근본적 수준에서 뇌를 시뮬레이션 한다는 목표 발표(Henry Markram, HBP Project Leader)
- 인지형태를 프로그램으로 구현하여 인간처럼 지식처리가 가능한 인공지능 시스템을 개발할 계획이며, 세부 연구분야에는 Application, ICT Platform, Theory, Data로 구성

[표] 일본 원격 의료 정책 경과

시기	설명	연구내용	투자내용
Application	인간 지식 식별과 일처리 우선순위 결정하기 위한 실험데이터 수집 및 통합을 수행하는 신경정보학과 뇌 시뮬레이션 사용 영역	신경과학, 의학 및 컴퓨팅 기술에 의한 프로토타입 개발	2억 2,100만 유로
ICT Platform	인간이 질병으로 인한 손상을 입기 전에 진단법을 개발하거나 각 개별 환자의 맞춤형 치료 방법, 질병과 약물의 시뮬레이션을 통한 개선된 진단 방법 등 여러 가지 치료 방법을 개발하는 학문적 탐구 영역	신경과학 및 임상 연구의 가속을 위한 통합형 ICT 플랫폼 개발	4억 5,600만 유로
Theory	뇌 시뮬레이션에 활용할 수 있는 빠른 연산장치를 개발하며, 에너지 효율성 및 신뢰성을 제고하는 기술 개발 영역	뇌 활동영역 간 관계를 파악하여 수학적모델 개발	7,200만 유로

시기	설명	연구내용	투자내용
Data	주로 신경과학, 의학 및 컴퓨팅 기술에 의한 프로토타입 개발 영역	뇌지도에 필수적인 데이터를 생성하거나 분석에 필요한 데이터 확보	3억 1,600만 유로

※ 출처 : 인공지능 분야 국내외 동향분석 및 투자전략 수립, KISTEP, 2018

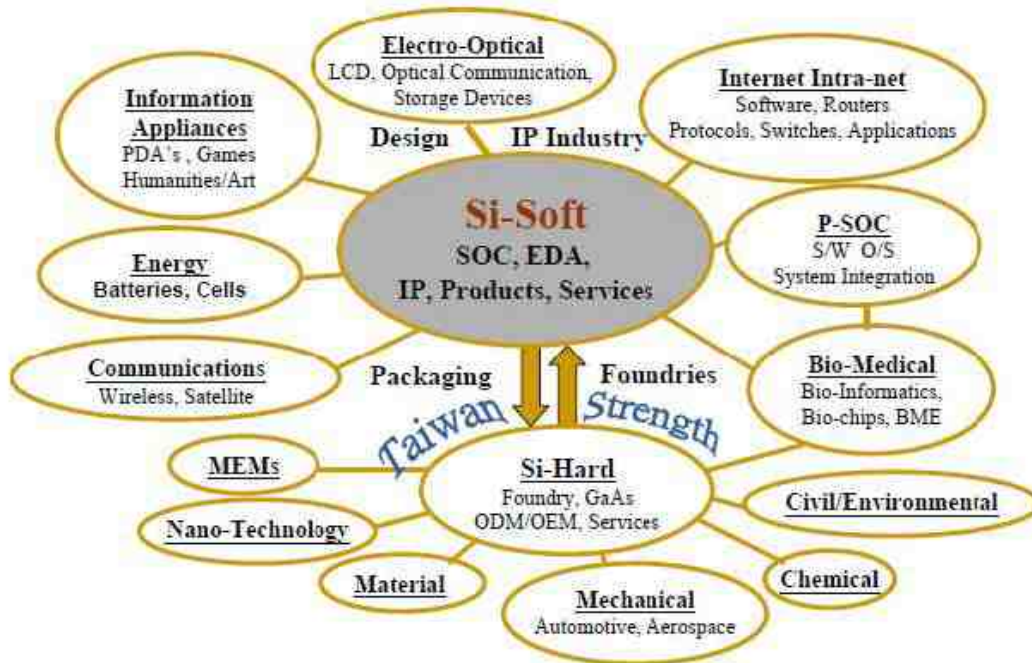
다. 일본

- 일본은 2001년부터 2007년까지 MIRAI, 2007년부터 2011년까지는 ASUKA 등의 프로젝트를 진행해 왔으며 최근 반도체 개발 거점 구축을 통한 산학연협력 기반 연구개발 지원 추진
- 일본 정부는 2018년 신형 반도체 개발 거점을 구축하여, 산학연 협력 기반으로 설계 작업부터 시제품까지 개발에 필요한 설비 지원 계획
 - 개발에 필요한 설비를 정부가 제공하면 반도체 기업이나 대학의 기술자들이 이 설비를 무료로 이용해 설계 작업부터 시제품까지 만드는 방식으로 추진
- PC나 스마트폰에 탑재되는 반도체에 비해 처리 속도가 10배 이상 빠르면서도 소비 전력은 100분의 1 이하인 자율주행 차량용과 인공지능용 반도체를 개발할 계획

라. 대만

- 대만의 SoC 산업은 대만 정부의 적극적인 국내 기업 육성정책에 따라 급성장
 - Si-Soft는 “Silicon Software”의 약자로, 기존의 전자제품 생산 수준 업그레이드를 위해 대만 정부가 대대적으로 시행한 시스템 반도체 투자 지원 프로젝트
 - Si-Soft 프로젝트의 목적은 시스템 반도체 인프라를 개발하는 것으로, 설계, IP, 생산, 집적회로 미세공정을 위한 국제적인 공급 체인을 형성
 - Si-Soft 프로젝트는 1) 혁신적인 반도체 제품 설계 2) SIP(Silicon

Intellectual Property) 개발 3) 전자설계 자동화(Electronic Design Automation, EDA) 통합 4) SIP Mall 5) 디자인 서비스 솔루션 5개 부문으로 구분



[그림] 대만 Si-Soft의 SoC 생태계 구축 목표

※ 출처 : The National Si-Soft Project, National Chiao-Tung University, 2010

- (CIC) 대만정부는 시스템IC 설계능력 확충을 위해 1992년부터 국가과학위원회에 CIC(Chip Implementation Center)를 설립
 - 국가응용연구랩(National Applied Research Laboratories) 소속인 CIC는 비영리연구기관으로서, 4개의 부서와 1개의 지사로 구성되어 있으며, 신주과학산업단지에 위치
 - 세계수준의 시스템 반도체 설계연구 및 서비스 연구센터로 도약하는 것을 목표로, 1) 시스템 반도체 설계 환경 및 서비스 향상, 2) 칩 생산 및 측정, 3) 시스템 반도체 설계 및 국제 협력 증진을 위한 사업 수행
 - CIC는 기업과 대학의 반도체칩 제작을 지원하여 시스템 반도체 산업의 경쟁력 강화에 기여한 것으로 평가

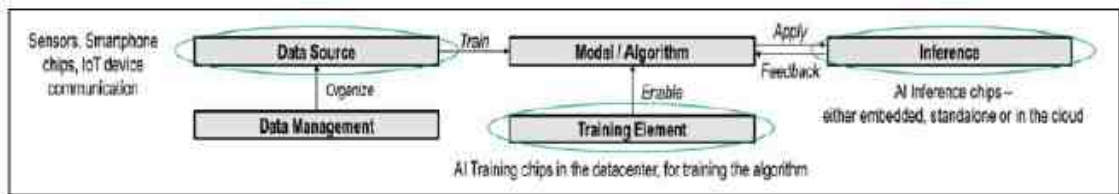
마. 중국

- 2017년 7월 인공지능 분야에서 세계를 이끌겠다는 ‘차세대 인공지능 개발계획’을 발표하였으며, 2020년까지 인공지능 반도체 기술 개발의 구체적 계획을 밝힘
- 차세대 인공지능 개발 계획((A Next Generation Artificial Intelligence Plan))은 연구개발, 산업화, 인재 개발, 교육 및 역량 습득, 표준 설정 및 규제, 윤리적 규범, 보안 등을 포함한 이니셔티브로서 국가 차원의 포괄적인 인공지능 전략 계획임
 - 2020년까지 중국의 인공지능 산업을 경쟁자와 동일 선상 수준까지 이끄는 데 주력하고 2025년까지 일부 인공지능 분야를 세계 최고 수준으로 육성, 2030년까지 인공지능 혁신을 위한 ‘중심센터’ 실현
- 인공지능 반도체에 대한 계획으로 2020년까지 다음과 같은 구체적인 계획을 밝힘
 - 차량, 서비스 로봇, 인증 시스템과 같은 지능화되고 네트워크화된 제품 개발
 - 지능형 센서 및 신경망 칩을 포함한 인공지능 지원 시스템 개발의 강조
 - 지능형 제조 시스템 개발 장려
 - 인공지능 개발 환경 향상-산업 훈련 자원 투자, 표준 테스트 및 사이버보안 등
- 중국 정부 연구기관 중 하나인 중국과학원은 인공지능 영역 R&D를 주도하는 조직자동화연구소의 역량을 강화하는 등 정부 차원에서 인공지능 반도체 지원을 본격적으로 추진
 - 2018년 중국과학원 산하 자동화연구소는 난징에 인공지능 연구개발과 성과 도출을 위한 핵심 기지로서 ‘인공지능혁신연구원’을 설립하고, 인공지능 반도체 설계 영역의 핵심인재 육성 등 계획 발표
 - 난징은 칭화유니그룹, TSMC를 포함한 중화권 반도체 업체의 첨단 제조 인프라가 집결하고 있는 곳인 만큼 산업계와 연계해 시너지를 낼 수 있는 인공지능 반도체 개발에 주력할 것으로 예상

Ⅲ. 기술 동향

1. 기술범위 및 특징

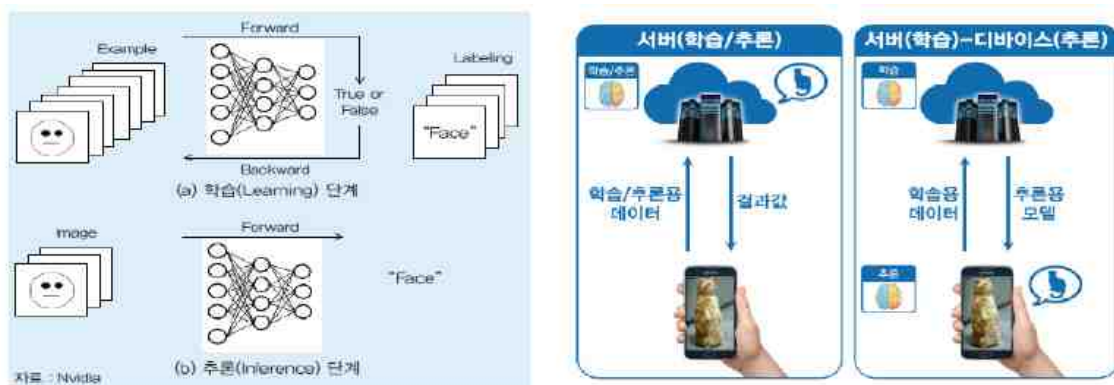
- 인공지능 반도체에 대한 정의는 활용 범위에 따라 다양하나 본 간행물에서는 인공지능 구현을 위해 요구되는 데이터 연산을 효율적으로 처리하는 반도체로 정의



[그림] 인공지능 구현을 위한 반도체 활용 범위

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

- (활용 목적) 인공지능 반도체는 인공지능 시스템의 구현 목적에 따라 크게 학습용과 추론용으로 구분할 수 있으며, 두 가지 과정을 반복 실행하여 최적의 답을 찾도록 성능을 강화하는 데 주로 사용
 - (학습용) 딥러닝 등 기계 학습의 특정 작업을 수행하기 위해 방대한 데이터를 통해 반복적으로 지식을 배우는 단계
 - (추론용) 학습을 거친 최적의 모델을 통해 외부 명령을 받거나 상황을 인식하면 학습한 내용을 토대로 가장 적합한 결과를 도출하는 단계



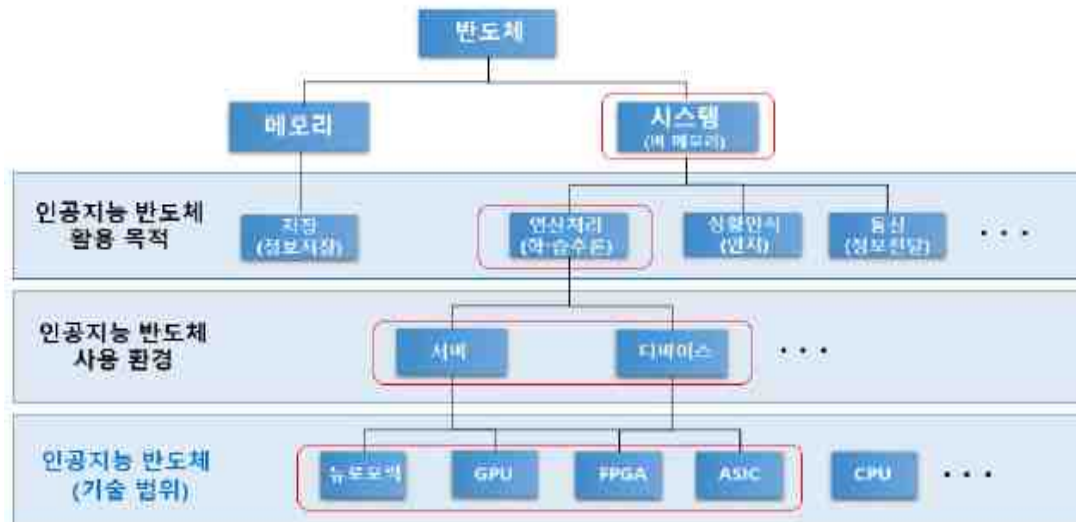
[그림] 인공지능 반도체의 주요 활용 목적(좌) 및 사용 환경(우)

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

- (사용 환경) 기존 인공지능 시스템은 주로 데이터센터에서 학습과 추론을 병행하여 사용되었으나, 스마트폰 및 IoT 등의 보급 확산, 클라우드 기술 발전과 동시에 디바이스의 추론 기능의 수요가 증가하면서 이를 수행하기 위한 반도체 기술 중심으로 발전
 - (데이터센터용) 현재 인공지능 학습/추론은 대부분 데이터센터에서 실행되며 일반적인 하드웨어로는 CPU가 담당하고 있지만, 인공지능 서비스에 요구되는 대규모 연산 처리 성능을 위해 인공지능 반도체를 서버에 장착하여 활용
 - 데이터센터 전용 반도체는 방대한 데이터를 처리하기 때문에 발열과 전력 소모로 인한 효율성 개선이 지속적으로 필요
 - (엣지 디바이스용) 데이터센터 서버(클라우드)와 연결을 최소화하고 디바이스 자체에서 인공지능 연산이 수행되는 경우가 점차 확대되면서 소형화·저전력·고성능 중심의 인공지능 반도체 기술 개발이 가속화

- 시스템 반도체 중 연산 처리에 최적화된 반도체를 중심으로 기술범위를 설정하였으며, 신기술로 주목받고 있는 차세대 인공지능 반도체인 뉴로모픽 반도체까지 포함

- 인공지능의 연산 성능 고속화 및 소비전력 효율(Power Efficiency)을 위해 최적화시킨 반도체이며, 아키텍처 구조 및 활용 범위에 따라 크게 GPU, FPGA, ASIC, 뉴로모픽 반도체로 구분
 - GPU(Graphical Processing Unit): 동시 계산 요구량이 많은 그래픽 영상 처리를 위해 고안된 병렬 처리 기반 반도체로 수천 개의 코어를 탑재하여 대규모 데이터 연산 시 CPU 대비 성능이 우수함
 - FPGA(Field-Programmable Gate Arrays): 회로 재프로그래밍을 통해 용도에 맞게 최적화하여 변경이 가능한 반도체로 활용 목적에 따라 높은 유연성이 특징
 - ASIC(Application Specific Integrated Circuits) : 특정 용도에 맞도록 제작된 주문형 반도체로 가장 빠른 속도와 높은 에너지 효율이 특징
 - 뉴로모픽 반도체(Neuromorphic Chips) : 기존 반도체 구조가 아닌 인간의 뇌를 모방한 非폰노이만 방식의 인공지능 전용 반도체로 연산처리, 저장, 통신 기능을 융합한 가장 진화한 반도체 기술



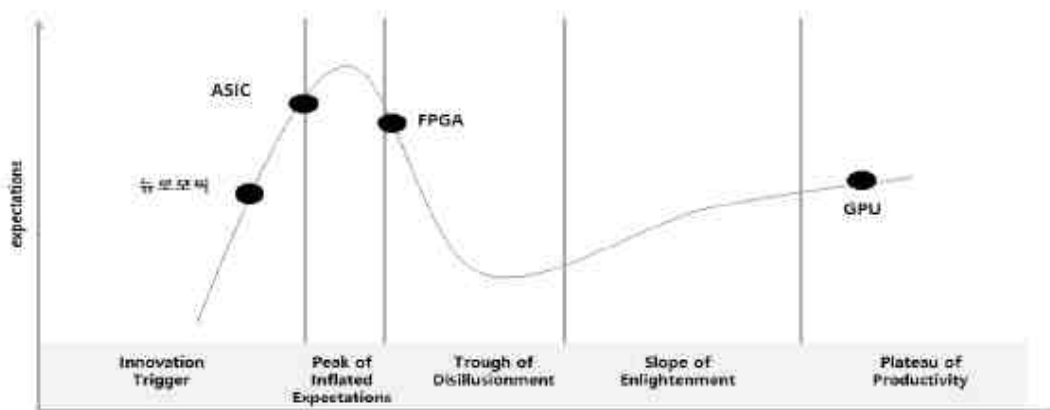
[그림] 인공지능 반도체의 기술 범위

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

2. 국내/외 기술 Trend

- 인공지능 알고리즘의 고도화 및 요구되는 데이터가 지속 증가함에 따라 신속한 데이터 연산 처리를 위한 반도체 기술의 발전이 가속화
 - 연산을 주목적으로 하는 가속기(AI Acceleration)를 탑재하여 CPU와의 분담을 통해 데이터 처리를 가속화하는 이기종 시스템 구조(HSA, Heterogeneous System Architecture) 기술이 등장하면서 인공지능 구현에 최적화된 반도체 개발의 움직임이 가속화되고 있음
 - 또한, 병목 현상 및 무어의 법칙 등 기존 반도체 구조의 한계를 돌파하고자 인간의 두뇌(뉴런-시냅스 구조)를 모방하여 구현하는 뉴로모픽 반도체 기술이 등장하면서 인공지능 반도체 기술의 새로운 변화 시기 도래

- 인공지능 반도체는 기술성숙도, 사용 환경 및 기능에 따라 기술 간 서로 대체되어 사용되기도 하나 공통적으로는 초고성능·초저전력 중심으로 진화할 전망
 - Gartner의 ‘Technology Hype Cycle’에 따르면 현재 인공지능 반도체 기술은 성숙도가 높은 순으로 ‘GPU’ > ‘FPGA’ > ‘ASIC’ > ‘뉴로모픽 반도체’로 나타남



[그림] 인공지능 반도체 Technology Hype Cycle

※ 출처 : Hype Cycle for Artificial Intelligence, Gartner, 2018

○ 인공지능 반도체는 활용 목적에 따라 용도가 다양하며 사용 환경 및 기술 발전에 따라 기술 간 서로 대체될 가능성이 큰 특성을 가짐

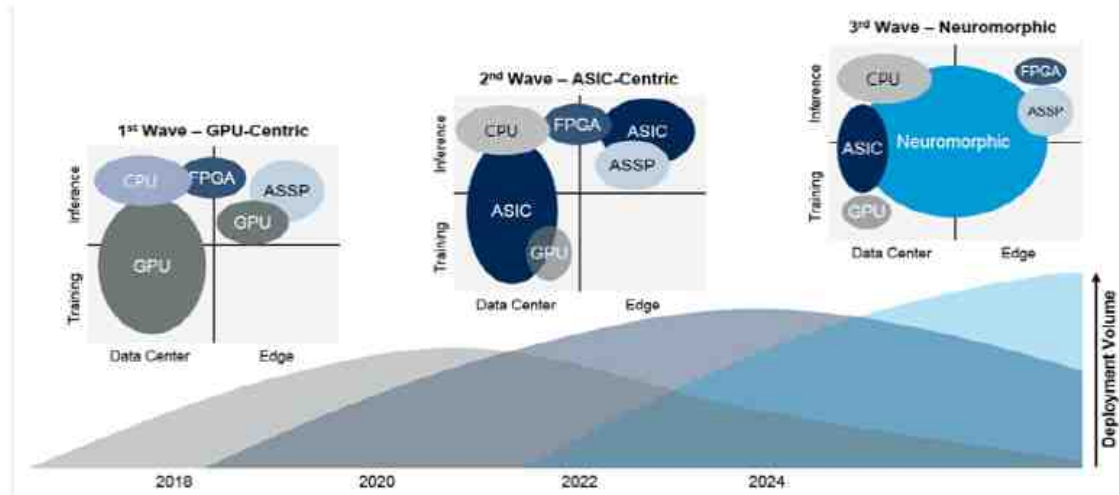
[표] 인공지능 반도체 유형별 특징

유형	GPU	FPGA	ASIC
장점	- 병렬 처리에 최적화된 프로세서로, CPU에 비해 빠른 가속 성능 - NVIDIA사의 CUDA 등 개발자 환경이 잘 갖춰져 있으며, 적용 사례가 많아 지원받기 용이	- ASIC보다 초기 개발 비용이 저렴 - CPU와 병렬 작동이 용이하여 전체 시스템 병목현상 발생 없음 - 회로 재구성이 가능, 개발 중인 AI 알고리즘을 유연하게 적용 가능 * (예) A라는 업무에 최적화하여 사용하다 반도체 회로 구성을 다시 설정(재프로그래밍)하여 B라는 업무에 맞춰 사용 가능	GPU, FPGA 대비 매우 빠른 속도와 우수한 전력효율
단점	- FPGA, ASIC 대비 낮은 전력 효율 - 기존 x86 시스템(CPU)에 추가 구축 시, 확장성과 호환성에 한계 * (예) 데이터 전송 병목 문제, 시스템 호환 문제 등	ASIC보다 연산속도가 느리고 CPU나 GPU 같은 범용 프로세서 대비 프로그래밍 전문성을 요함	- 매우 비싼 초기 제작비용, 장시간의 개발소요 시간 - 특정 연산에 최적화되었기 때문에 응용 분야가 한정

※ 출처 : 인공지능 반도체 산업동향 및 이슈 분석(ETRI Insight), 한국전자통신연구원, 2017

○ 일부 시장조사기관에서는 인공지능 반도체가 병렬연산 처리에 최적화된 GPU 중심에서 초고성능·초저전력 중심의 뉴로모픽 반도체로 기술이 진화할 것으로 전망함

- (1st Wave) 현재의 인공지능 반도체는 GPU 중심으로 데이터센터 및 엣지 디바이스에 탑재되어 주로 활용
- (2nd Wave) 인공지능을 지원하는 저전력·고성능의 특화된 반도체인 ASIC 방식으로 변화될 전망
- (3rd Wave) 최종적으로는 추론·학습, 데이터 센터·디바이스 등 인공지능 시스템의 다기능을 지원하는 초저전력·초고성능의 뉴로모픽 반도체 중심으로 진화 전망



[그림] 인공지능 반도체의 기술 진화 방향

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

가. GPU (Graphic Processing Unit) 관련 동향

○ 게임 산업의 3D 그래픽 등을 처리하기 위해 개발되었으나 대규모 데이터의 연산을 효율적으로 처리하는 병렬 처리 기반 반도체로 각광을 받으면서 인공지능 구현을 위한 핵심 반도체로 부상

○ 직렬(순차) 처리 방식인 CPU는 한 개의 명령어에 의해 입력된 순서대로 데이터를 처리하는 반면, GPU는 여러 명령어를 동시에 처리할 수 있는 병렬 처리(Parallel Processing) 구조로 방대한 데이터의 연산을 지원하기 위한 반도체로 각광 받음

- CPU는 연산을 담당하는 ALU의 개수가 최적화된 소규모 코어로 구성되어 있으며, GPU는 수천 개 코어로 구성됨



[그림] CPU(좌)와 GPU(우)의 회로도 비교

※ 출처 : 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019

- GPU를 통해 기계학습(딥러닝 등)에 요구되는 복잡·고도화된 데이터 연산 처리의 성능 향상을 보여줌으로서 인공지능 구현의 발전을 견인
 - 2010년 인공지능 분야 전문가인 스탠퍼드 대학의 앤드루 응 교수는 12개의 GPU가 무려 2,000개의 CPU에 맞먹는 딥러닝 성능을 발휘한다는 사실을 발견
 - 2012년 이미지넷 대회에서 GPU 기반 딥러닝 기술을 활용해 우승을 차지하면서 인공지능 분야에서 GPU의 비약적인 발전 가능성을 보임

- GPU 기술은 NVIDIA社가 시장 선두 기업으로 데이터센터, 자율주행차 등 적용 분야에서 인공지능 반도체로 활용

- (데이터 센터용) 2016년 GTC(GPU Technology Conference)에서 데이터 센터에 최적화된 ‘테슬라(Tesla) P100’ GPU를 발표하였으며, 2017년에는 GPU 컴퓨팅 아키텍처인 ‘볼타(Volta)’에 기반한 최초의 프로세서인 ‘테슬라(Tesla) V100’ GPU를 공개함
 - 볼타의 성능은 120 TFLOPS로 기존 버전인 파스칼 대비 5배, 2015년 출시된 맥스웰 대비 15배 향상됐으며, CPU 100대와 같은 수준의 성능으로 딥러닝을 구현

- (자율주행차용) 2017년 GTC에서는 자율주행 차를 위한 AI Supercomputer chip, Xavier를 발표하면서 엣지 디바이스에 최적화된 반도체 기술을 선보임
 - Xavier의 구성을 보면, 512개의 GPU, 8개의 Core로 구성된 ARM64 CPU, CVA(computer vision accelerator), Video Processor로 구성되어, 30W의 전력으로 30 TFLOPS 연산을 구현
 - 자율주행 차량에서 수행해야 하는 주변 환경 감지, 차량 스스로 위치 파악, 주변 사물의 행동과 위치 예측 등의 연산 등을 구현

나. FPGA (Field-programmable Gate Array) 관련 동향

- 활용 목적에 따라 재프로그래밍이 가능한 반도체로서 개발에 투입되는 시간이 짧고 원하는 작업에 맞춰 연산 처리가 가능해 유연성이 높은 반도체로 각광
- 기존 반도체들은 한번 생산되면 수정할 수 없지만 FPGA는 하드웨어를 재설계하지 않고 프로그래밍을 통해 사용 목적에 맞게 변경할 수 있음
 - 예를 들어 FPGA는 번역 작업에 최적화해 사용하다가 반도체 회로 구성을 다시 설정해 가상비서 서비스에 맞춰 쓸 수 있음
 - 특정 목적에 맞춘 하드웨어 프로그래밍이 가능해 기계 학습을 통해 최적화된 학습모델의 출력 값을 빠르게 계산하는 인공지능 추론 서비스 구현에 적합하다는 평가를 받음
 - 알고리즘을 수정하거나 연구개발 중일 때는 다른 인공지능 반도체보다 FPGA가 효율적
- 또한, CPU와 병렬로 작동하므로 시스템 혼란이나 병목 현상 없이 사용 가능하여 최근 인공지능 구현을 위한 기술로 주목받고 있음
- FPGA 기술은 데이터센터 및 엣지 디바이스 등의 저전력화·고성능화를 위해 활용되고 있으며 글로벌 기업들의 기술개발이 활발히 진행 중임
 - 자일링스社は FPGA 시장에서 약 56% 점유율을 보유한 선두 기업으로 데이터센터 전용 반도체인 ‘reVISION’ 을 통해 머신러닝, 데이터 분석 등과 같이 고성능 작업에 최적화된 제품으로 경쟁사 대비 2~6배의 컴퓨팅 성능을 발휘한다고 발표
- Microsoft社は 2015년 FPGA를 탑재한 ‘캐터펄트(Catapult)’ 서버 메인보드를 통해 자사의 검색엔진 ‘빙(Bing)’ 의 성능을 일반 CPU를 이용한 것보다 30% 비용 절감과 10% 전력 감소에 성공
 - 데이터센터에 있는 반도체를 FPGA로 통합하면 용도 재구성이 가능한 구조인 만큼 단일 서비스에 수천 개에 이르는 FPGA를 이용할 수도 있어

유연성이 뛰어남

- Intel은 최근 인공지능 특화 기술을 확보하기 위해 독립형 FPGA와 통합형 FPGA(FPGA-CPU)을 제공하고 있으며, 추론용 가속기인 ‘intel FPGA’와 엣지 디바이스에서 머신 러닝을 구현하기 위한 저전력 비전 기술 ‘모비디우스’ 등을 발표

다. ASIC (Application Specific Integrated Circuit) 관련 동향

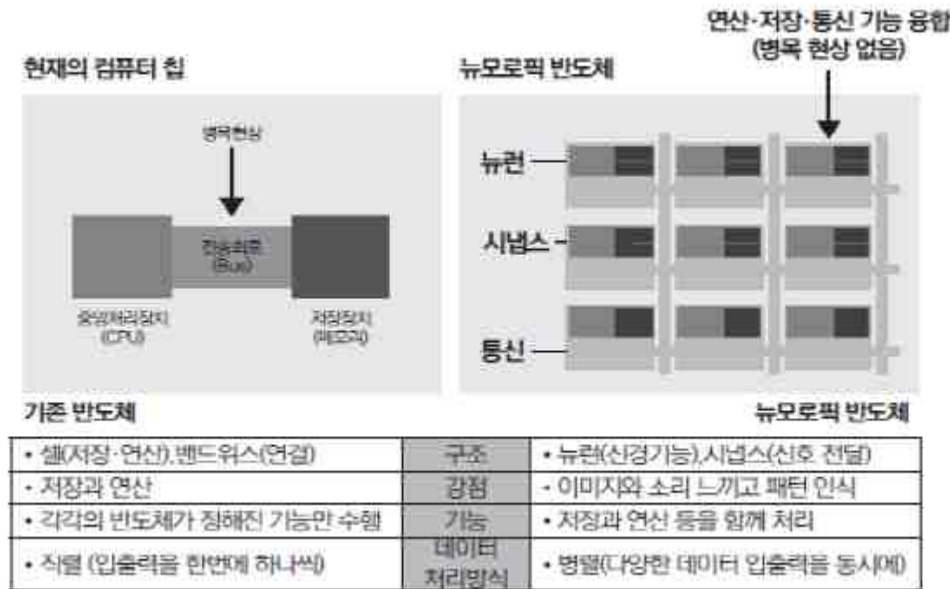
- 어플리케이션의 특성에 적합한 인공지능 시스템을 구현하기 위해 범용 프로세서를 사용하는 대신 특정 목적으로 제작되는 주문형 반도체 기술이 부상
 - 특정한 용도에 맞도록 제작된 주문형 반도체인 ASIC는 빠른 속도와 높은 에너지 효율의 특성을 지니고 있어 인공지능 전용 반도체로 각광받고 있음
- ASIC는 다른 인공지능 전용 반도체보다 비용이 많이 들고 개발 기간이 길며, 한번 제품을 만들고 나면 기능을 바꿀 수 없지만 범용 프로세서(CPU, GPU)나 FPGA보다 성능 개선에 유리함
 - ASIC는 개발이 끝난 알고리즘의 전력 소모량을 줄여야 하는 기기에 도입할 때 효율적임
- 전통적인 반도체 업체 외에도 다양한 산업체들은 인공지능 알고리즘이 내장된 ASIC 반도체를 자체 개발하는 추세
- Google은 SW 및 빅데이터 센터 기반 인터넷 전문기업으로, 인공지능 기술을 활용한 서비스의 수요가 증대되면서 자사의 데이터센터에 사용하기 위한 병렬 프로세서 반도체를 직접 개발하기 시작
 - 인공지능 딥러닝을 구현하기 위해 가장 널리 이용되는 SW 플랫폼인 오픈소스 기반 ‘텐서플로’를 구현하는데 최적화된 하드웨어를 자체 개발
 - 2016년 5월, 데이터 분석과 딥러닝을 위해 개발한 TPU(Tensor Processing Unit) 1세대를 공개하였으며, GPU에 비해 병렬처리에 특화되어 있으며 추론 등의 연산을 수행하는데 92 TLOPS의 성능을 보임

- 2017년에는 TPU 1세대를 개선하여 기계 학습에도 사용할 수 있는 TPU 2세대를 개발
 - TPU1 세대는 학습된 모델을 사용한 추론, 이미지나 언어 등의 인식에 특화됐다면, 2세대 TPU는 머신러닝 연산 과정에서 추론뿐 아니라 학습 연산에서 성능 향상에 특화됨
 - 데이터센터용으로는 TPU 2세대 4개를 탑재한 Cloud TPU와 64개를 연결한 TPU Pod을 발표함
- 바이두는 최근 AI 개발자 컨퍼런스 행사에서 AI 연산용 ASIC ‘쿤룬(Kunlun)’을 공개하였으며, 260 TFLOPS 수준으로 초당 512GB 데이터를 주고받으며 무인 자동차부터 데이터센터까지 모든 곳에 사용 가능
- 삼성전자는 초고속 모뎀을 탑재하고 인공지능(AI) 연산 기능을 강화한 고성능 모바일 애플리케이션 프로세서(AP) ‘엑시노스9(9810)’를 개발

라. 뉴모로픽 반도체 관련 동향

- ‘폰 노이만(Von Neumann)’ 구조의 기존 반도체 한계를 극복하기 위해 인간처럼 저 전력으로도 고성능의 인공지능을 수행하는 반도체 기술이 각광을 받기 시작
- 기억장치(메모리)·중앙처리장치(CPU)·입출력장치(IO) 등 3단계 구조로 명령을 순차적으로 수행하는 방식의 폰 노이만 구조는 복잡한 작업일수록 시간이 오래 걸리고 에너지 소모가 크게 증가하는 구조
- 산업연구원에 따르면 폰 노이만 구조는 인간의 뇌와 같은 정보처리를 위해 필요한 소비 전력을 감당하기 위해선 원자력발전소 1기가 필요하다고 언급
- 뉴로모픽 반도체 안에는 여러 개의 ‘코어(Core)’들이 존재하며, 코어에는 트랜지스터를 포함한 몇 가지 전자 소자들과 메모리 등이 탑재
- 코어의 일부 소자는 뇌의 신경세포인 뉴런의 역할을 담당하며, 메모리 반도체는 뉴런과 뉴런 사이를 이어주는 시냅스 기능을 담당

- 인공 뉴런 역할을 하는 코어를 사람의 뇌처럼 병렬로 구성하였기 때문에 적은 전력만으로 많은 양의 데이터 처리가 가능하며, 인간의 뇌처럼 학습하기 때문에 연산 성능이 대폭 향상



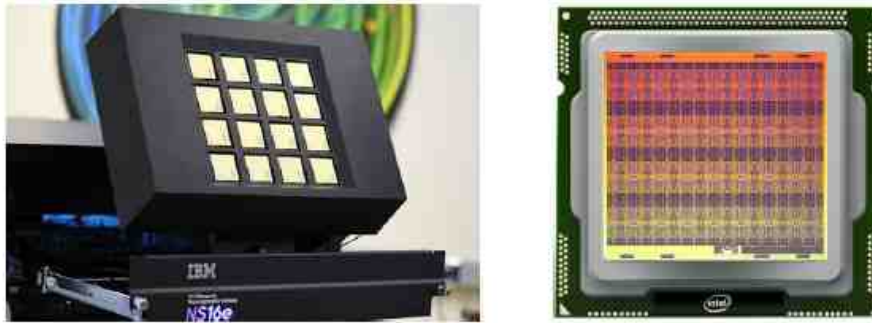
[그림] 기존 반도체와 뉴로모픽 반도체의 구조도 비교

※ 출처 : 두뇌 신경회로 모방 뉴로모픽 칩(융합 Weekly TIP), 융합연구정책센터, 2018

- 전 세계적으로 뉴로모픽 반도체 기술에 대한 연구가 활발히 진행되는 가운데 산·학·연 협력 중심으로 추진되는 것이 특징
- (IBM) 미국 국방부 산하 방위고등연구계획국(DARPA)이 주도하는 ‘인공두뇌 만들기 프로젝트’에 참여하여 ‘트루노스’(TrueNorth)라는 뉴로모픽 반도체를 만드는 데 성공
 - 트루노스 반도체는 무려 54억 개 트랜지스터를 내장한 4,096개의 프로세서로 이루어져 전자 회로 소자들을 인간의 신경망처럼 연결해 인간 두뇌 활동을 모방
 - 100만 개의 뉴런과 2억 5000만 개의 시냅스로 얹혀 초당 1,200프레임에서 2,600프레임으로 이미지를 분류하는데, 이때 25mW에서 275mW 수준의 적은 전력을 소비하며 이는 기존 마이크로프로세서의 1만 분의 1의 전력을 소요
- (Intel) 사람이 입력 데이터와 함께 정답을 알려 주는 지도학습 대신 실시간으로 유입되는 정보를 받아들여 스스로 학습하는 뉴로모픽 반도체 ‘로이히(Loihi)’를

개발 (2017년)

- 0.47mm² 크기의 코어 128개로 이뤄진 로이히에는 총 13만 개의 뉴런과 1억 3000만 개 시냅스로 구성
- 반도체 집적도는 14nm(나노미터 · 1nm는 10억분의 1m) 수준으로 현재까지 개발된 뉴로모픽 반도체 중 최고 수준임



[그림] 뉴로모픽 반도체 (좌) IBM의 트루노스 (우) intel의 로이히

※ 출처 : IBM, intel

- (퀄컴) 세계 최초로 뇌에서 영감을 얻어 신경세포처럼 스파이크 형태의 신호를 주고받고 시냅스 연결 강도를 조절해 정보를 처리하는 프로세서인 ‘제로스’ (Zeroth) 개발(2013년)
 - 제로스는 회로 구조뿐만 아니라 기능도 인간의 학습 방법을 모방하여 강화학습(reinforcement learning)을 활용해 로봇을 제어하는 데모 영상 시연
- (SK 하이닉스) 미국 스탠퍼드대학과 강유전체(Ferroelectrics) 물질을 활용한 ‘인공신경망 반도체 소자 공동 연구개발’ 협약 체결(2016년)
 - SK하이닉스는 전하 유입 여부를 통해 0과 1을 구분하는 기존의 반도체 입력 방법 대신 전압 크기에 따라 다양한 신호를 저장할 수 있는 유기 물질 강유전체를 사용해 뉴로모픽 반도체 개발 추진 중
- KIST, KAIST, 서울대, 포스텍, 울산과학기술원(UNIST), 국민대, 어바인 캘리포니아대 등 7개 기관이 연구단을 구성하여 2021년까지 자가 학습이 가능한 뉴로모픽 반도체 ‘네오(NeO)2C’ 개발을 목표로 연구 추진
 - ‘네오(NeO)2C’의 반도체 집적도는 55nm, 소모전력은 56mW 수준이며

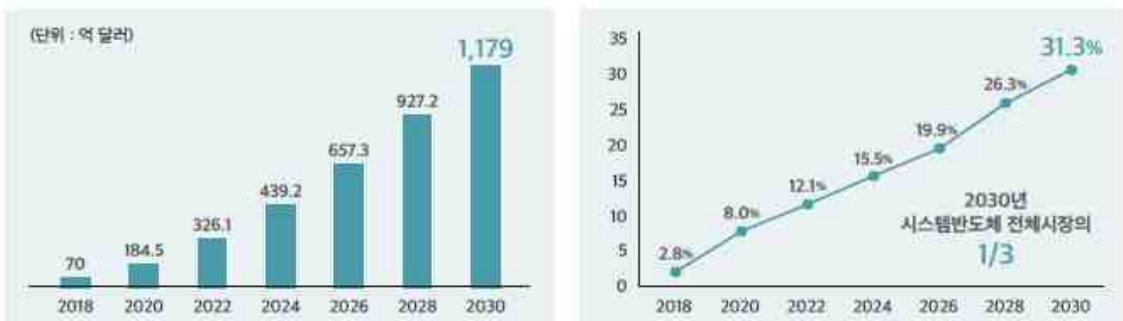
코어당 뉴런 수는 1,024개로 ‘로이히’와 유사하나, ‘로이히’는 128개 코어이며 ‘네오(NeO)2C’는 단일코어

- (네패스) 반도체 패키징·테스팅(후공정) 기업인 네패스는 미국의 반도체 설계업체인 제너럴 비전社와의 협업을 통해 엣지 디바이스용 인공지능 반도체인 NM500을 상용화 추진(2017년)
 - NM500은 0.4mm 반도체에 576개의 인공뉴런을 집적하여 고속·병렬연산 처리를 수행
- 차세대 유망 기술인 뉴로모픽 반도체는 일부 상용화가 진행되고 있으나 산업적 적용 확대를 위해서는 다소 시간이 소요될 것으로 전망
 - Gartner의 ‘Technology Hype Cycle’에 따르면 뉴로모픽 반도체는 10년 후 시장이 형성될 것으로 예측하였으며, 10년 이후에는 지능형 로봇, 무인기, 자율주행 자동차, AI 비서 등에서 폭넓게 사용할 것으로 전망
 - 뉴로모픽 반도체는 다양한 분야의 학문이 함께 연구되어야 하는 대표적인 융합 기술이며, 생물학에서 연구되는 뇌의 학습, 기억, 그리고 인지 기능 등에 대한 이해와 더불어 이를 공학적으로 구현하기 위한 뉴로모픽 시스템, 알고리즘, 소자 등 다양한 공학 분야에서의 기술 발전이 필요

IV. 시장 동향

1. 글로벌 시장

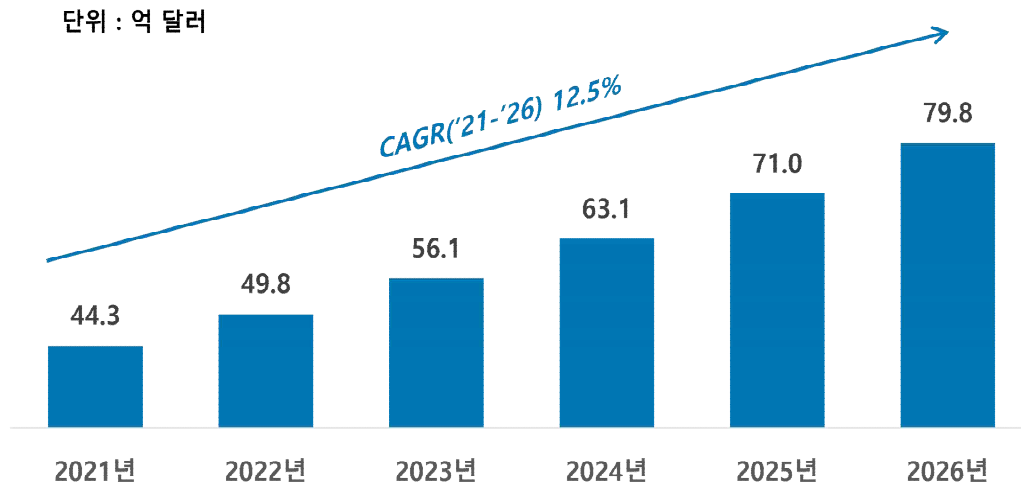
- 세계 AI 반도체 시장은 성장이 둔화되고 있는 반도체 산업에서 새로운 성장 동력이 될 것으로 기대
- 2030년이면 인공지능반도체 글로벌시장 규모가 1,179억 달러에 이를 전망이며 시스템 반도체 시장 규모 3,769억 달러에서 차지하는 비중도 30%를 웃돌 것으로 예상
- AI 반도체 시장이 3년 후인 2024년 낙관적 전망대로 400억 달러 내외의 시장을 형성하게 된다면, 시스템 반도체 시장에서 약 15%의 비중을 차지할 전망
 - 다른 반도체와 비교하여 당장은 큰 규모의 시장은 아닐지 모르지만, PC · 스마트폰의 성장 둔화로 정체되어 있는 기존 반도체 시장 대비 성장 속도가 빨라, 앞으로 5~10년 후에는 AI 반도체 시장이 메모리 반도체나 마이크로프로세서에 버금가는 시장으로 성장할 전망



[그림] 인공지능 반도체 시장 전망(좌) 및 시스템 반도체 시장 중 인공지능 반도체 비중(우)

※ 출처: 인공지능 반도체 육성으로 인공지능과 종합 반도체 강국 시대 연다 KIPA, 2020

- 글로벌 리서치회사 Technavio에 따르면 글로벌 AI반도체 시장은 2020년 44.3억 달러에서 연평균 성장률 12.5%로 증가하여, 2025년에는 79.8억 달러에 이를 것으로 전망됨



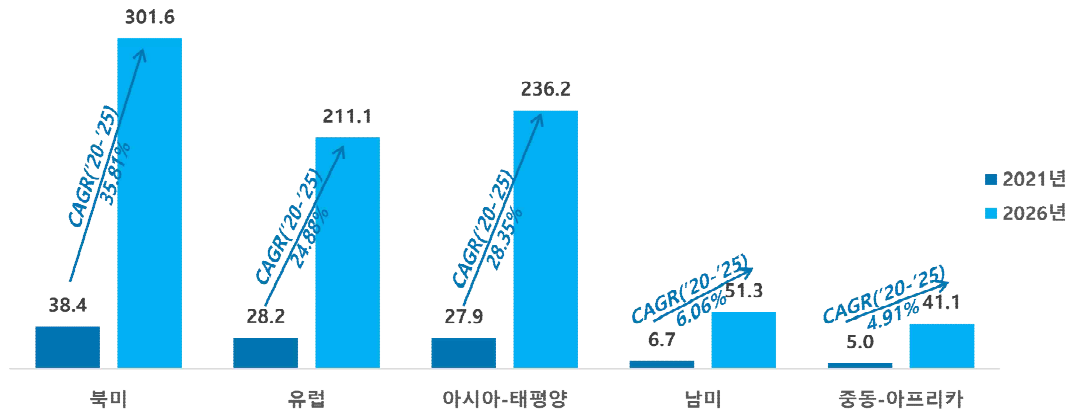
[그림] 글로벌 AI반도체 시장 규모 및 전망

※ 출처 : Global Artificial Intelligence (AI) Chips Market, Technavio, 2020

○ **(지역별 현황)** 글로벌 AI반도체 시장을 지역별로 살펴보면, 2020년을 기준으로 북미 지역이 38.4억 달러로 가장 큰 시장을 형성하고 있음

- 북미 지역은 2020년 38.4억 달러에서 연평균 성장률 35.81%로 증가하여, 2025년에는 301.6억 달러에 이를 것으로 전망됨
- 유럽 지역은 2020년 28.2억 달러에서 연평균 성장률 24.88%로 증가하여, 2025년에는 211.1억 달러에 이를 것으로 전망됨
- 아시아-태평양 지역은 2020년 27.9억 달러에서 연평균 성장률 28.35%로 증가하여, 2025년에는 236.2억 달러에 이를 것으로 전망됨
- 남미 지역은 2020년 6.7억 달러에서 연평균 성장률 6.06%로 증가하여, 2025년에는 51.3억 달러에 이를 것으로 전망됨
- 중동-아프리카 지역은 2020년 5.0억 달러에서 연평균 성장률 4.91%로 증가하여, 2025년에는 41.1억 달러에 이를 것으로 전망됨

단위 : 억 달러



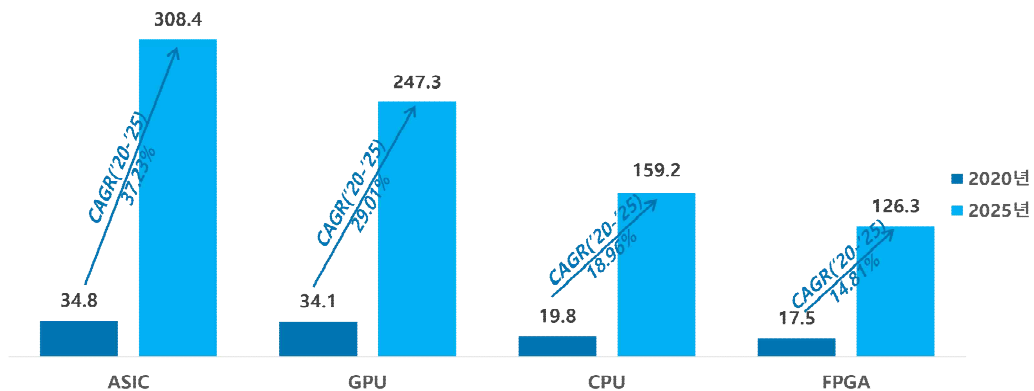
[그림] 지역별 AI반도체 시장 규모 및 전망

※ 출처 : Global Artificial Intelligence (AI) Chips Market, Technavio, 2020

○ (종류별 현황) 글로벌 AI반도체 시장을 아키텍처 구조 및 활용 범위에 따라 분류하면 크게 ASIC, GPU, CPU, FPGA 반도체로 구분

- ASIC 시장은 2020년 34.8억 달러에서 연평균 성장률 37.23%로 증가하여, 2025년에는 308.4억 달러에 이를 것으로 전망됨
- GPU 시장은 2020년 34.1억 달러에서 연평균 성장률 29.01%로 증가하여, 2025년에는 247.3억 달러에 이를 것으로 전망됨
- CPU 시장은 2020년 19.8억 달러에서 연평균 성장률 18.96%로 증가하여, 2025년에는 159.2억 달러에 이를 것으로 전망됨
- FPGA 시장은 2020년 17.5억 달러에서 연평균 성장률 14.81%로 증가하여, 2025년에는 126.3억 달러에 이를 것으로 전망됨

단위 : 억 달러



[그림] 종류별 AI반도체 시장 규모 및 전망

※ 출처 : Global Artificial Intelligence (AI) Chips Market, Technavio, 2020

2. 국내 시장

○ 한국은 메모리반도체 강국이나, 시스템 반도체 경쟁력은 상대적으로 취약

○ 우리나라의 반도체 세계시장 점유율(2019년)은 21%이나 시스템 반도체 세계시장 점유율(2019년)은 3.2% 시스템 반도체 부분은 지난 10년간 비슷한 수준의 점유율을 유지하며 정체됨

- 시스템 반도체 산업은 종합반도체기업 삼성전자와 다수의 중소 팹리스로 구성되며, 대기업 제외시 시스템 반도체의 세계시장 점유율은 1% 미만
- 특히 팹리스 점유율(2019년)은 1%로 대만(17%), 중국(15%) 대비 현저하게 낮은 상황

○ 시스템 반도체 세부 분야별 한국의 점유율(2019년)은 로직 IC는 6.2%, 아날로그 IC는 1.5%, 마이크로컴포넌트는 0.5% (반도체산업협회)

- 로직 IC의 점유율은 2015년 7.0%에서 2019년 6.2%로 하락, 대표 품목인 디스플레이 구동칩(DDI)*과 AP가 한국기업의 디스플레이, 스마트폰 출하량 둔화 등에 영향을 받음

* DDI는 삼성전자가 삼성디스플레이, 실리콘웍스(LG계열)가 LG디스플레이를 중심으로 공급

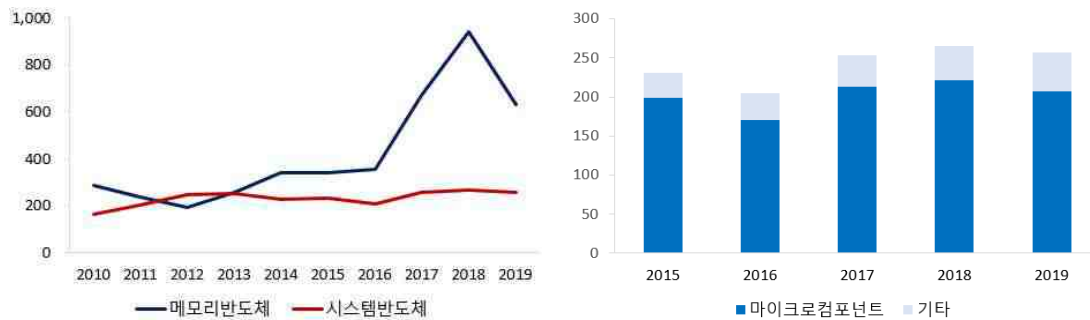
- 아날로그 IC의 점유율은 2015년 1.2%에서 2019년 1.5%로 소폭 상승
- 마이크로컴포넌트의 점유율은 2015년 0.9%에서 2019년 0.5%로 하락, 자동차용 마이크로컨트롤러 수요 증가 등에 대한 대응 미흡 등이 영향을 미침
- 마이크로컨트롤러의 수요처별 비중(2019년)은 자동차용 30%, 가전 20%이며, 한국은 국내 가전기업의 경쟁력 등에 힘입어 가전용 마이크로컨트롤러에서 선전

○ 시스템 반도체 수출은 수요증가, 파운드리 경쟁력 향상 등으로 2019년에 257억 달러 기록

- 시스템 반도체 수출은 2017년 이후 전자제품 등의 스마트화로 시스템 반도체 수요가 증가했으며, 수요증가와 경쟁력 제고로 파운드리들의 가동률도 상승

○ 시스템 반도체 수출의 주력 품목은 마이크로컴포넌트

- 마이크로컴포넌트의 시스템 반도체 수출 비중은 2015년 86.0%에서 2019년 80.5%로 낮아지는 추세이나 여전히 최대 주력 품목
- 우리나라의 시스템 반도체 대표 품목은 디스플레이 구동칩(DDI)과 AP이나 자사 Captive 수요 중심, 해외생산 등으로 수출기여도는 세계시장 점유율 대비 낮음



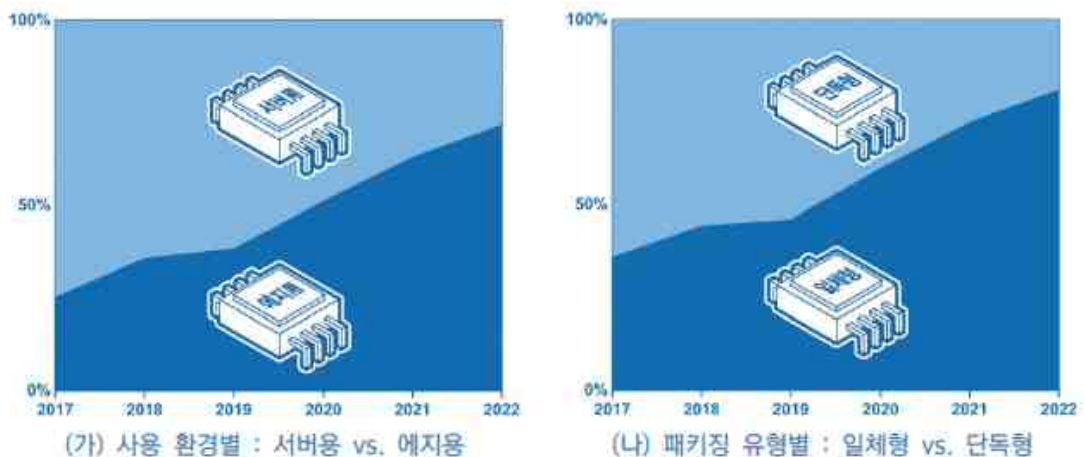
[그림] 메모리반도체 및 시스템 반도체 수출 추이(좌), 시스템 반도체 품목별 수출 추이(우)

※ 출처 : 뉴딜산업 분석보고서 : 시스템 반도체산업 현황 및 전망, 한국수출입은행 해외경제연구소, 2020

V. 산업 동향

1. 글로벌 산업 동향

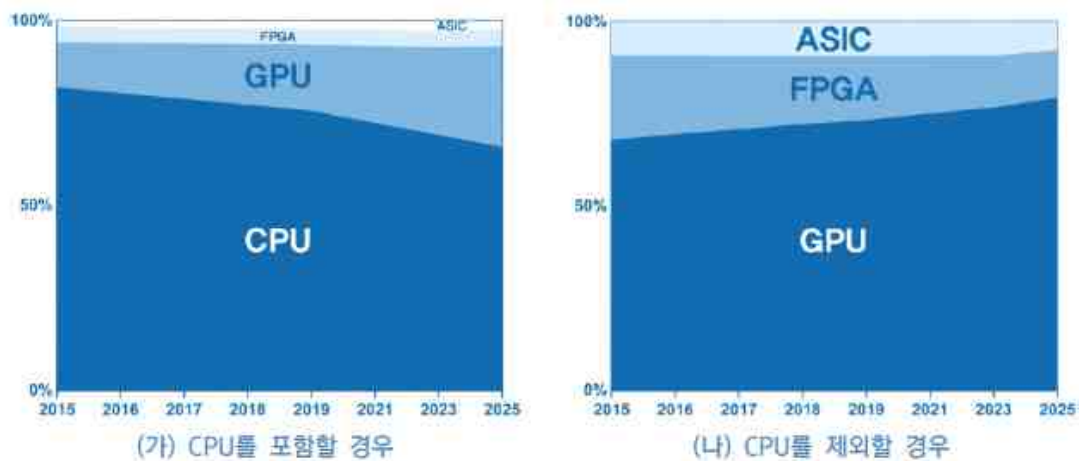
- 반도체 초기 시장에서 학습용/서버용 AI 시장의 비중이 상대적으로 높겠지만, 시간이 지남에 따라 추론용/엣지용 시장 비중이 증가하는 추세를 보일 전망
- 시장조사회사 가트너는 AI 반도체가 시장 초기 데이터센터 서버에서 주로 사용되다가 점차 엣지 디바이스용으로 무게 중심이 옮겨가고, 단독형보다 기존 반도체에 AI 프로세싱 기능이 내장되는 일체형이 대세가 될 것으로 전망
 - 가트너는 2021년 모든 프리미엄 스마트폰에 AI 연산이 가능한 AP가 탑재되고, 2022년 IoT 디바이스용 ASIC/ASSP의 1/5 이상이 AI 연산 기능을 갖출 것으로 전망함으로써, 엣지 디바이스용 AI 반도체 시장의 성장성을 매우 긍정적으로 평가
 - 이로 인해 전체 AI 반도체 시장에서 엣지 디바이스용 AI 반도체 시장 비중은 25%에서 72%까지 급격하게 높아질 전망
 - 엣지 디바이스용 AI 반도체 시장의 성장은 자연스럽게 일체형 AI 반도체 시장 성장과 궤를 같이하게 되며, 이에 일체형 AI 반도체 비중은 36%에서 81%까지 늘어날 것으로 기대



[그림] 가트너의 세계 AI 반도체 시장 전망

※ 출처: AI Neural Network Processing Semiconductor Revenue, Worldwide, Gartner, 2018

- 시장조사회사 마켓스앤마켓스에 따르면, CPU 기술이 AI 반도체 시장의 대부분을 차지하고 있으나, 시간이 지나면서 GPU 비중이 확대되고 CPU 비중은 점차 축소될 것으로 예상
- 전망에서 CPU를 제외하고 보면, 전체 시장에서 GPU 비중은 더욱 확대되어 2025년에는 80%까지 확대될 것으로 예상되며, FPGA/ASIC 기술 비중은 상대적으로 낮은 수준을 유지할 것으로 전망

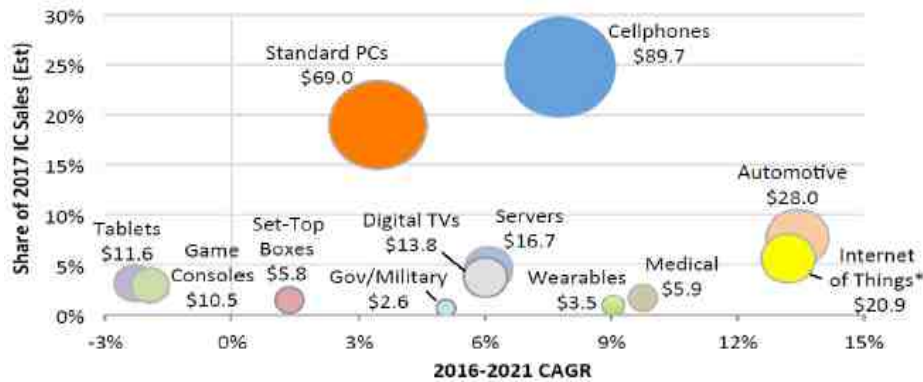


[그림] 마켓스앤마켓스의 세계 AI 반도체 시장 전망 : 기술별

※ 출처 : Artificial Intelligence Chipsets Market, Markets and Markets, 2018

- 인공지능 반도체는 융복합화·스마트화를 주도하는 지능형 산업의 핵심 기반 기술로서 수요가 지속적으로 증가할 것으로 전망
- 인공지능을 활용한 주요 산업의 변화 추세는 2016년을 기점으로 급성장하였으며, 금융 및 소비, 헬스케어, 자동차 등 인공지능의 산업적 적용이 확대되고 있는 상황에서 인공지능 구현을 위한 핵심 기반기술인 인공지능 반도체의 시장도 점차 확대될 전망
- 산업별 반도체 IC 시장에서 스마트폰 산업의 반도체 시장 점유율이 가장 높게 나타났으며, 향후 자동차, IoT, 의료, 웨어러블 산업에서 반도체 IC 시장 수요가 급증할 것이라고 전망
- 반도체 IC 전체 시장의 25% 비중을 차지하는 스마트 폰 산업은 2017년 기준 매출 규모가 89.7억 달러이며 향후, 연간 7.8% 성장하여 2021년에는 1,056억 달러에 달할 것으로 전망

- 자동차 및 IoT 산업의 반도체 IC 시장은 전 산업의 7.9% 비중을 차지한데 비해 향후, 연평균 13.4%, 13.2% 수준으로 타 산업의 비해 성장률이 가장 높게 나타나며 시장 수요가 급증할 것으로 전망
- 의료 분야는 연평균 9.7% 성장하여 2021년까지 78억 달러, 웨어러블 시스템은 연평균 9.0% 성장하여 2021년에는 49억 달러 규모로 시장이 확대될 전망



[그림] 주요 산업별 반도체 IC 시장 전망(2016~2021) (단위: 억달러)

※ 출처 : Game Changers 2018, CBInsights, 2018

○ (미국) AI 반도체의 본고장 미국에서는 AI 반도체 개발 춘추전국시대가 전개

- (대기업) CPU · GPU 시장을 과점 중인 인텔과 엔비디아가 AI 반도체 초기시장을 이끄는 가운데, 반도체 사업을 하지 않는 기업도 자사 제품/서비스 경쟁력 강화를 위해 독자 AI 반도체를 개발 추진

[표] 미국 주요 대기업의 AI 반도체 개발 현황

업체명	데이터센터 서버용				일체형	엣지 디바이스용			
	GPU	FPGA	ASIC/ASSP	뉴모로픽		GPU	FPGA	ASIC/ASSP	뉴모로픽
인텔		○	△	△				◎	
엔비디아	◎					◎			
AMD	○								
XILINK		◎							
퀄컴					○				
구글			◎					○	
마이크로소프트								○	
아마존								△	

업체명	데이터센터 서버용				일체형	엣지 디바이스용			
	GPU	FPGA	ASIC/ASSP	뉴모로픽		GPU	FPGA	ASIC/ASSP	뉴모로픽
IBM									◎
애플					◎				
Hewlett Packard				△					
TESLA								△	

※ ◎ : 시장선도자, ○ : 시장 추격자, △ : 제품 개발 중

※ 출처 : 반도체 산업의 차세대 성장엔진 AI반도체 동향과 시사점, IITP, 2018

○ (중국) 우수한 스타트업들이 중국 AI 반도체 기술 혁신을 주도

- 화웨이 자회사 하이실리콘을 제외하고는 나머지 기업들이 전부 스타트업일 정도로, 중국의 AI 반도체 산업은 스타트업이 중심
- 딥파이/비트메인/썬크포스를 제외한 나머지 기업들은 모두 엣지 디바이스용 AI 반도체를 개발하고 있으며, 그중에서도 ASIC/ASSP 형태가 대부분
- 최근 이들 중국 AI 반도체 스타트업에 대한 투자가 증가하고 있는데, 이는 AI 반도체 시대에는 국산 반도체를 사용하겠다는 중국 정부의 산업 육성 의지와 무관하지 않으며, 대학과 연구소에서도 사업화를 적극적으로 지원

[표] 중국 AI 반도체 개발 현황

업체명	데이터센터 서버용				일체형	엣지 디바이스용			
	GPU	FPGA	ASIC/ASSP	뉴모로픽		GPU	FPGA	ASIC/ASSP	뉴모로픽
하이실리콘					○				
캠브리콘								◎	
호라이즌 로보틱스								○	
디파이테크			○					○	
비트메인			○						
Think Force			△						
Unisound								△	
Intel If Usion								○	

※ ◎ : 시장선도자, ○ : 시장 추격자, △ : 제품 개발 중

※ 출처 : 반도체 산업의 차세대 성장엔진 AI반도체 동향과 시사점, IITP, 2018

2. 국내 산업 동향

- 우리나라의 인공지능 반도체 산업은 미국·중국 등과 비교하여 산업 저변이 열악한 상황
 - 우리나라 반도체 산업은 삼성전자, SK 하이닉스 등 메모리 반도체 시장 및 역량이 높은 반면, 인공지능 반도체 기술 역량 및 산업 저변은 미국·중국에 비해 열악한 상황
 - 특히, 우리나라 인공지능 반도체 설계·원천기술의 경우 대부분 수입에 의존하는 실정이며 대기업 중심의 ASIC 개발과 산학연 중심의 뉴로모픽 반도체 연구가 활발하다는 것이 특징

[표] 국내 AI 반도체 개발 현황

업체명	데이터센터 서버용				일체형	엣지 디바이스용			
	GPU	FPGA	ASIC/ASSP	뉴로로픽		단독형			
						GPU	FPGA	ASIC/ASSP	뉴로로픽
삼성전자					○				△
네페스									○
넥셀			△						
SK하이닉스			○						
ETRI								○	
서울대학교									△
KAIST								△	

※ ◎ : 시장선도자, ○ : 시장 추격자, △ : 제품 개발 중

※ 출처 : 반도체 산업의 차세대 성장엔진 AI반도체 동향과 시사점, IITP, 2018

- (삼성전자) 스마트폰 전용의 인공지능 반도체를 상용화하였으며, 뉴로모픽 반도체 선행 연구를 추진 중
 - 갤럭시S9 스마트폰에 탑재되는 ASIC 기술인 엑시노스 9810은 빠른 이미지 처리를 위해 뉴럴 프로세싱 엔진을 탑재
 - AI와 5세대(5G) 이동통신, 전장부품, 바이오 등 4개 신산업을 선정하여 2020년까지 25조원의 투자 계획을 발표하였으며, 이를 위해 실리콘밸리, 토론토 등 6곳에 글로벌 AI 연구센터를 구축하여 기술경쟁력 강화 추진

- 삼성종합기술원 산하 두뇌컴퓨팅 연구실을 중심으로 뉴로모픽 반도체 개발을 추진하고 있으며, 최근 서울대학교 · KAIST · UNIST과 함께 뉴로모픽 반도체 산학협력 모색

- (네패스) 반도체 패키징 업체 네패스는 美 뉴로모픽 반도체 업체 제너럴비전과 2016년 생산-판매에 관한 글로벌 독점 판매권을 갖는 계약을 체결하고, 2018년 1월부터 뉴로모픽 반도체 NM500 양산을 시작
- 다른 인공지능 반도체와 달리 엣지 디바이스 상에서 학습 및 추론이 가능하며, 2018년 100만 개 생산 · 판매를 목표

- (넥셀) AP 전문업체로, 국내에서 유일하게 인공지능 반도체 설계원천 기술 개발 추진
- 셀(XELL)로 명명한 인공지능 반도체 설계기술을 개발 중이며, 테슬라 K80이나 Google TPU2보다 우수한 수준이라고 평가

- (SK텔레콤) 글로벌 기업과의 협업을 통해 데이터센터 및 엣지 디바이스 전용 인공지능 반도체 공동연구 추진
- 2018년 SK텔레콤은 美 자일링스와 협업해 국내 대규모 데이터센터로는 처음으로 자사 데이터센터 인공지능 반도체로 자일링스 FPGA를 채택하였으며, 인공지능 가속 솔루션을 개발해 AI 서비스 ‘누구’에 적용
- 2019년 1분기에는 자일링스 ‘버텍스 VU9P’ FPGA를 적용하여 기존보다 전력 소모량은 3배 늘어나지만 컴퓨팅 능력은 10배 이상 향상된 새로운 인공지능 반도체를 발표할 계획

- (퓨리오사AI) 사람 뇌의 신경망을 모방한 NPU(신경망 처리 장치) 기술을 활용해 데이터센터와 기업용 서버에서 AI 연산 성능을 극대화할 수 있는 반도체 개발
- 2019년 글로벌 AI 반도체 테스트 대회인 ‘엠엘퍼프’에서 아시아 스타트업 가운데 유일하게 AI반도체 성능 지표를 공식으로 인정받음
- 2021년 네이버 D2SF · DSC인베스트먼트 · KDB산업은행 · IMM인베스트먼트 등으로부터 800억원 규모 투자를 유치

- (딥엑스) 딥러닝 알고리즘 경량화 기술을 보유한 벤처기업으로 방대한 딥러닝 연산처리를 돕는 NPU(차세대 신경처리장치) 개발 분야에서 뛰어난 기술력 보유
 - 마그나인베스트먼트와 캡스톤파트너, 패스파인더H 등의 기존 투자사와 한국산업은행과 SJ투자파트너스, 타임폴리오, 신한캐피탈 DS자산운용 등의 신규 투자사에게 210억 원 규모의 시리즈B 투자 유치

- (오픈엠티테크놀로지) 반도체 IP(핵심기능 블록 설계도면)를 공급하는 스타트업으로 직접 반도체를 제작하지 않고, 반도체 기업에 IP를 제공하고 로열티를 받아 수익을 얻는 비즈니스 모델 구성
 - NPU와 차세대 고성능 메모리 시스템 등의 핵심 기술을 보유하였으며 2021년 300억 원 규모의 시리즈C 투자유치 성공

- 핵심 인재 영입 및 활발한 인수합병을 통해 자사의 기술경쟁력 강화 및 글로벌 초기 시장 확보에 주력
 - 삼성전자는 빅데이터를 빠른 속도로 처리하는데 활용되는 GPU 기술을 자체 개발하기 위해 ‘SGPU(Samsung’ s GPU)’ 프로젝트를 가동

- 인공지능의 산업적 활용이 확대됨에 따라 인공지능 반도체는 데이터센터용과 엣지 디바이스용으로 시장이 세분화되는 추세이나 상호보완적 관계로 유지될 전망
 - (데이터센터용) 딥러닝 학습을 위한 병렬연산 처리 능력과 수요변화에 대응하기 위한 컴퓨팅 파워 구성의 확장성과 유연성, 운용비용 관점에서의 전력효율 등을 중심으로 시장 형성
 - (엣지 디바이스용) 자율주행차, 드론, IoT 등 개별서비스에 특화된 연산 가속, 배터리, 폼팩터 등의 제한 환경 극복을 위한 초저전력, 경량화, 제조원가 절감 등을 중심으로 시장 형성

[참고문헌]

- 인공지능 반도체 육성으로 인공지능과 종합 반도체 강국 시대 연다, KIPA, 2020
- 인공지능 프로세서 반도체 기술 및 표준화 동향, ETRI, 2020
- 「인공지능 강국」 실현을 위한 인공지능 반도체 산업 발전전략, 관계부처 합동, 2020
- 행정포커스: 인공지능 반도체 육성으로 인공지능과 종합 반도체 강국 시대 연다, KIPA, 2020
- 뉴딜산업 분석보고서 : 시스템반도체산업 현황 및 전망, 한국수출입은행 해외경제연구소, 2020
- Global Artificial Intelligence (AI) Chips Market, Technavio, 2020
- 인공지능 기술의 진화와 AI 반도체·컴퓨팅의 변화, 정보통신정책연구원, 2019
- 기술동향브리프 인공지능(반도체), 한국과학기술기획평가원, 2019
- 반도체 산업의 차세대 성장엔진 AI반도체 동향과 시사점, IITP, 2018
- 인공지능 분야 국내외 동향분석 및 투자전략 수립, KISTEP, 2018
- 두뇌 신경회로 모방 뉴로모픽 칩(융합 Weekly TIP), 융합연구정책센터, 2018
- 혁신성장동력 시행계획, 관계부처 합동, 2018
- 인공지능 반도체의 특허 동향 및 시장 트렌드, 정보통신기술진흥센터, 2018
- AI Neural Network Processing Semiconductor Revenue, Worldwide, Gartner, 2018
- Artificial Intelligence Chipsets Market, Markets and Markets, 2018
- Game Changers 2018, CBInsights, 2018
- 인공지능 반도체 산업동향 및 이슈 분석(ETRI Insight), 한국전자통신연구원, 2017
- 대만의 시스템반도체 육성 정책 현황, 정보통신산업진흥원, 2011
- The National Si-Soft Project, National Chiao-Tung University, 2010