

초점

July 2024 No.1

새로운 기회의 창으로 Si반도체 시장 현황과 전망

김민식 부연구위원

오정숙 부연구위원

정보통신정책연구원 디지털플랫폼경제연구실

새로운 기회의 창으로 AI반도체 시장 현황과 전망

김민식 부연구위원

정보통신정책연구원 디지털플랫폼경제연구실, kimmin@kisdi.re.kr

오정숙 부연구위원

정보통신정책연구원 디지털플랫폼경제연구실, redrock5@kisdi.re.kr

요약

- AI반도체는 AI서비스 인프라의 기본·핵심 기술 계층으로 반도체 시장의 성장을 이끌고 있으며, 시장 참여자에게 수요·기술·제도의 측면에서 새로운 기회의 창을 제공
- 반도체 시장에서 AI반도체의 부상은 기존 가치사슬을 변화시키고, 이는 반도체 산업 경쟁 구도와 가치 창출 방식의 변동으로 이어져, 나아가 생태계에 혁신을 가져옴
 - EDA : AI반도체 설계를 지원하는 EDA툴의 중요성 증대 및 제품 포트폴리오 확대
 - SIP : Fabless, EDA, SIP 공급기업들은 AI반도체 SIP 제공 사업을 전략적 수행
 - Fabless·IDM : 다양한 수요처에 활용되는 AI반도체 고도화로 점유율 확대
 - Design House : AI반도체 수요기업에 대한 종합적인 지원역량 강화
 - Foundry : AI반도체 생산 활동에 필요한 부가적·보완적 활동을 통합·제공하는 플랫폼 전략 강화
 - OSAT : AI반도체 성능향상의 핵심기술 영역으로 반도체 가치사슬에서 중요성 증대
- AI반도체 시장 변화 동인으로는 ① (수요) On-device AI 활성화에 따른 AI반도체의 다양화 ② (시장 참여자) 빅테크 기업의 AI반도체 시장진입 활성화 ③ (요소기술) AI반도체 성능의 최적화에 필수적인 시스템SW의 전략화 ④ (생태계) 주요 선도기업의 플랫폼화 등이 작용

- 주요국은 적극적으로 AI 반도체 주도권 확보를 위한 반도체 정책 발표, 우리나라도 AI 반도체 이니셔티브 전략(2024.4)을 통해 수요처별 지원정책, 수평적인 생태계 정책, 기술 계층별 지원정책이라는 3가지 관점에서 정책목표를 이루기 위한 구조화된 정책 수립
- 향후 AI반도체 정책은 기존 AI반도체 정책의 효과를 제고하고 가치사슬의 변화 및 시장 변화의 동인을 반영한 방향으로 제도적인 기회의 창을 적극 활용해야 함
 - AI반도체 수평적 가치사슬 영역별 참여 기업의 경쟁력 열위 부문 지원
 - AI반도체의 다양한 수요처 및 다품종·소량생산 방식에 대응하는 세밀한 지원
 - AI반도체를 지원하는 다양한 기술 계층 변화에 대응하는 기술 계층별 지원
 - AI반도체 플랫폼 생태계의 참여 및 국내 선도기업의 플랫폼 구축·강화 지원

이 개요

- 본 고는 AI반도체의 부상이라는 새로운 기회의 창에 대해 분석하기 위하여, AI반도체 정의 및 중요성, AI반도체 시장·전망 및 관련 반도체 시장 등에 대하여 살펴봄
- 아울러 AI반도체의 부상에 따른 주요 영역별 가치사슬 변화 및 반도체 시장 변화 동인을 분석하고, 최근 우리나라의 AI반도체 정책도 분석하여 기회의 창을 활용할 수 있는 정책적인 시사점을 제공하고자 함

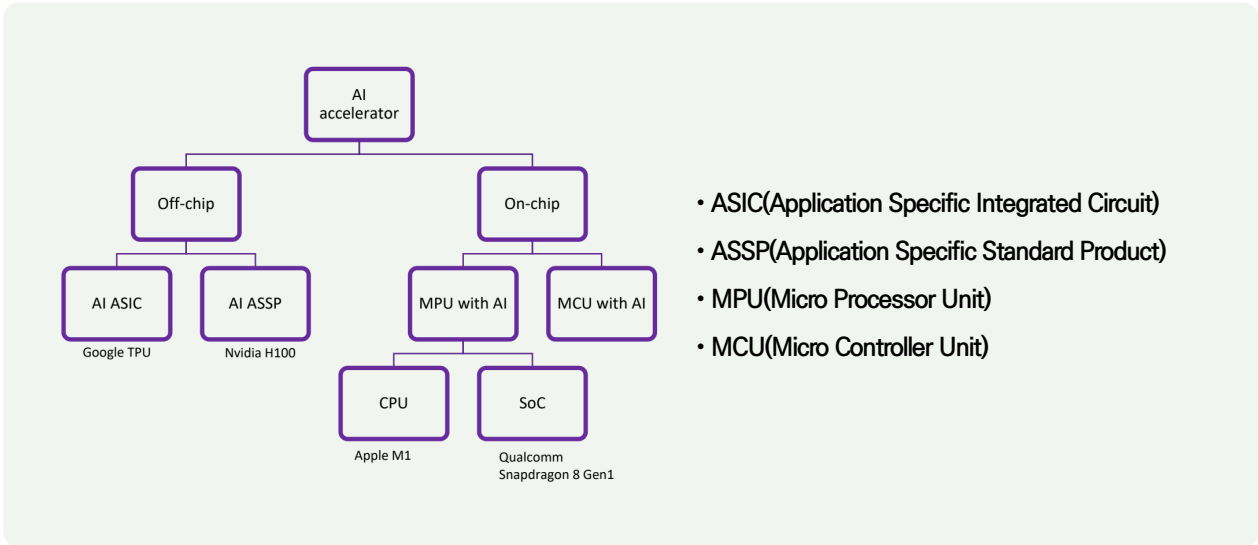
02 새로운 기회의 창(Windows of opportunity)으로 AI반도체

I 기존 반도체 산업에서 새로운 기회의 창(Windows of opportunity)¹으로 AI 반도체 등장

- 주요 시장조사업체 따라 차이가 존재하지만, 공통으로 AI 연산을 지원하는 훈련 및 추론에 사용되는 AI Processor 또는 AI accelerators를 AI반도체로 규정
 - 독립적인(Off-chip) AI반도체는 일반적으로 범용 또는 애플리케이션별 프로세서(시스템 반도체)와 연계되어 보완적으로 활용
- 또한, 반도체 칩 내부에 AI 기능을 통합하여 제공하는 예도 AI반도체 정의에 포함
 - AI반도체 설계 시 기존 시스템 반도체에 AI 기능을 담당하는 IP(intellectual property)를 통합하여, 하나의 반도체 칩(On-chip)으로 생산

1 "기회의 창"이라는 개념은 Perez와 Soete(1988)가 새로운 패러다임을 이용하여 후발주자가 기존기업을 추월하는 데 있어, 새로운 기술 경제 패러다임의 부상이 미치는 역할을 지칭하기 위해 처음 사용

1 그림 1 | AI반도체 분류



자료: OMDIA(2024)

- Gartner에 따르면 AI반도체의 수요처 측면에서는 5개 분야, AI반도체의 기술 측면에서는 8개 분야로 분류
 - * (수요처) 소비자 가전, 유무선 통신, 컴퓨팅, 산업, 자동차 등 5개로 분류
 - * (기술) AP(Application Processor), DSP(Digital Signal Processor), GPU(Graphics Processing Unit), FPGA, Microcontroller 32-bit, Microprocessor-Compute, Microprocessor-Embedded, Other Application Specific 등 8개로 분류
- AI 기술의 지속적인 발전과 이를 지원하는 핵심 인프라로써 AI반도체의 중요성 증가
 - AI 서비스를 구현하는 데 필요한 Hardware와 Software에는 필수적인 기술 요소로 ①AI 프로세서(반도체) ②AI 메모리 및 스토리지 ③AI 학습 및 모델 최적화 ④AI 응용 등 4개 모듈이 존재
 - AI 데이터 학습 및 추론 연산을 수행하는 i) AI 프로세서 기술과 ii) AI 빅데이터를 저장하고 처리하는 메모리 및 스토리지 기술 등 관련 반도체 기술이 AI 인프라의 기본·핵심(basic·core) 기술 계층임

I 그림 2 | AI Full stack의 구조와 연동



자료: SKT(2023.10.22)

- AI반도체의 등장은 반도체 시장에 새로운 진입자가 기존 기업들을 추격·대체하여 선도적인 기업이 될 수 있는 기회의 창(Windows of opportunity)을 제공하고 있음
 - AI반도체의 등장은 반도체 시장 참여자에 새로운 기회의 창(기술·수요·제도)²으로 작용
 - * 기술적인 기회의 창 : 서버 및 PC에 탑재되는 CPU 분야의 선두 주자였던 기존 Intel, AMD에 비해, AI 연산 처리를 위한 병렬 데이터 처리 중심의 GPU를 지속해 자가 발전시킨 Nvidia가 AI반도체의 선두 주자로 자리매김
 - * 수요적인 기회의 창 : 새로운 다양한 산업에 AI 모델 활용됨에 따라, 다양한 영역에서 활용되는 디바이스에 탑재되는 저전력·고성능 AI반도체 수요가 크게 성장할 전망으로 신규기업이 반도체 시장에 진입할 기회가 발생
 - * 제도적인 기회의 창 : AI반도체에 대한 공적인 정부의 개입(보조금 지급)이나 제도적 조건(실증사업, 검증)으로 시장 진입 기회가 발생

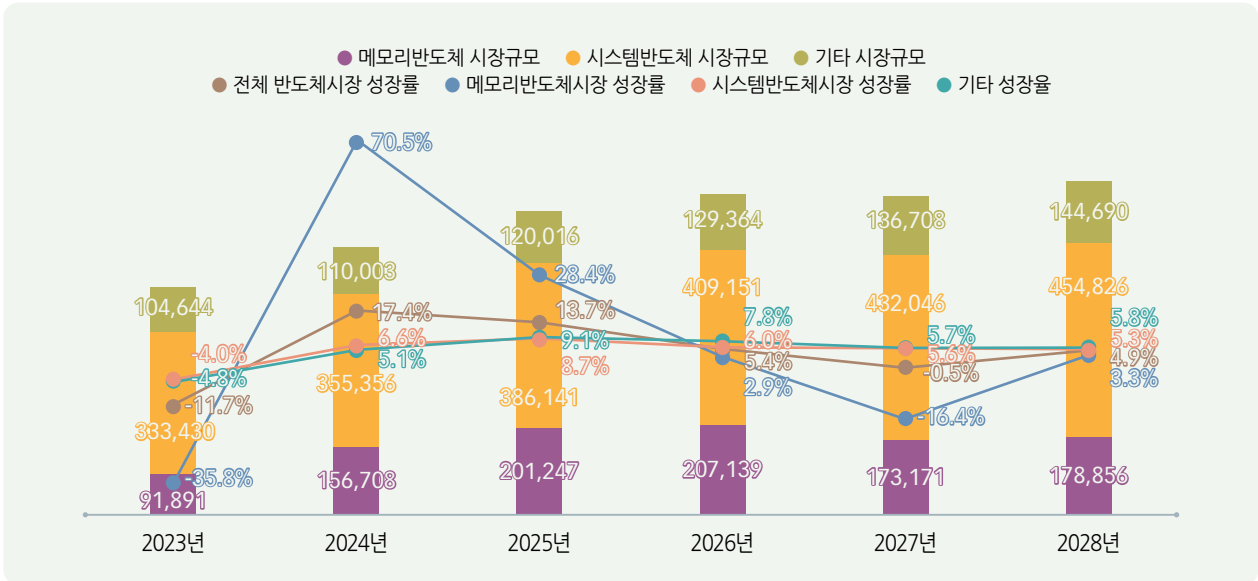
I 글로벌 AI반도체 시장 성장 추이·방향

- '23년 글로벌 반도체 시장은 부진했으나, AI반도체 시장은 생성형 AI로 촉발된 수요 증가로 전년 대비 27.2% 성장, 이후 '28년까지 연평균 24.3% 증가하며 글로벌 반도체 시장 성장 견인
 - '23년 글로벌 메모리반도체 시장 및 시스템반도체 시장 모두 수요 부족으로 전년 대비 각각 -35.8%, -4.0% 역성장하여 글로벌 반도체 시장도 -11.7% 위축
 - 반면, 글로벌 AI반도체 시장은 두 자릿수 성장하며, 전체 반도체 시장 및 시스템반도체 시장에서 차지하는 비중은 각각 '22년 7.0%, 12.1%에서 '23년 10.1%, 16.1%로 증가
 - ※ '28년 글로벌 AI반도체시장의 전체 반도체시장 및 시스템반도체시장 기여도는 각각 20.4%, 35.0%까지 대폭 확대

2 Lee, K., & Malerba, F. (2017)

그림 3 | 글로벌 반도체 시장규모 및 성장률 전망(2023년-2028년)

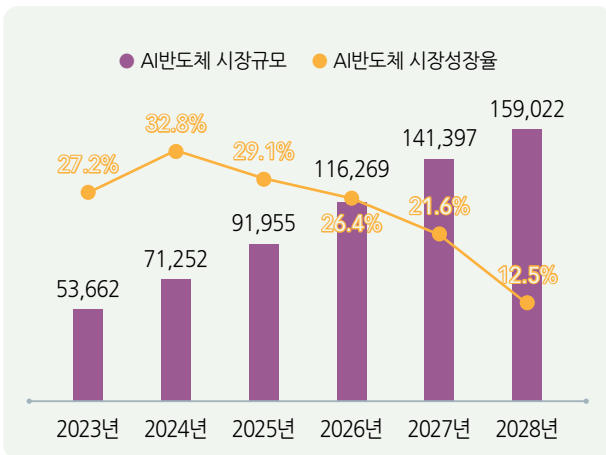
(단위: 백만 달러)



주: 1) 매출액 기준임 2) 기타에는 Analog, Discrete, Optoelectronics 반도체 포함
 자료: Gartner(2024.4a), 저자 재구성

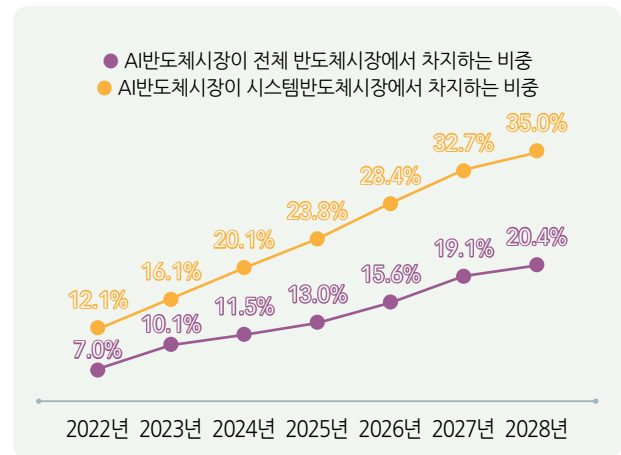
그림 4 | 글로벌 AI반도체 시장규모 및 성장률 전망 (2023년-2028년)

(단위: 백만 달러)



주: 매출액 기준임
 자료: Gartner(2024.4a), Gartner(2024.4b), 저자 재구성

그림 5 | 글로벌 AI반도체 시장이 전체 반도체 시장 및 시스템반도체 시장에서 차지하는 비중 추이 및 전망(2022년-2028년)



주: 매출액 기준임
 자료: Gartner(2024.4a), Gartner(2024.4b), 저자 재구성

- 한편, 메모리시장 부진에도 AI서비스 구현을 위해 서버에서 대규모 데이터 처리에 필요한 고대역폭메모리(HBM)의 글로벌 매출은 '23년 전년 대비 82.1% 상승할 것으로 추정³

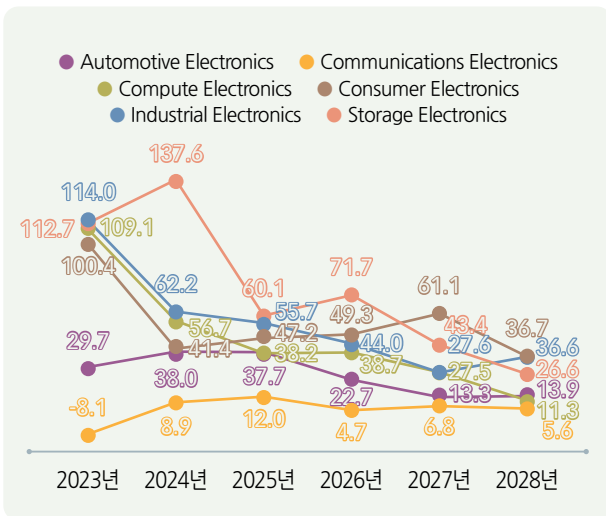
※ 매출액 기준 글로벌 HBM시장의 '22년부터 '27년까지 연평균 성장률은 36.3%에 달할 전망

3 Gartner(2023.8)

- 또한, 향후 AI서비스 일상화로 Si반도체가 기존 주요 수요처인 데이터 센터에서 스마트폰, PC, 자동차, 소비자가전, IoT 등 엣지 디바이스로 확대됨에 따라 저전력 D램(LPDDR) 수요도 증가하여 Si반도체는 시스템 반도체시장 뿐만 아니라 메모리 반도체시장 성장을 촉진할 것으로 기대

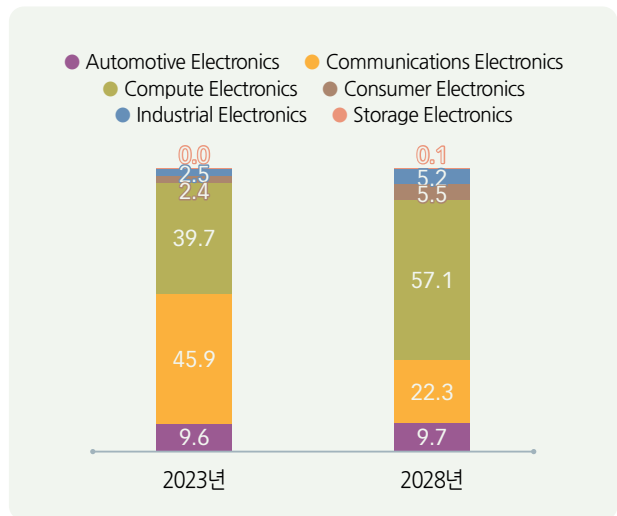
※ 매출액 기준 글로벌 LPDDR시장은 '23년부터 '28년까지 연평균 12.6%씩 성장할 전망⁴

그림 6 | 글로벌 Si반도체 기기별 성장률 전망 (2023년~2028년) (단위: %)



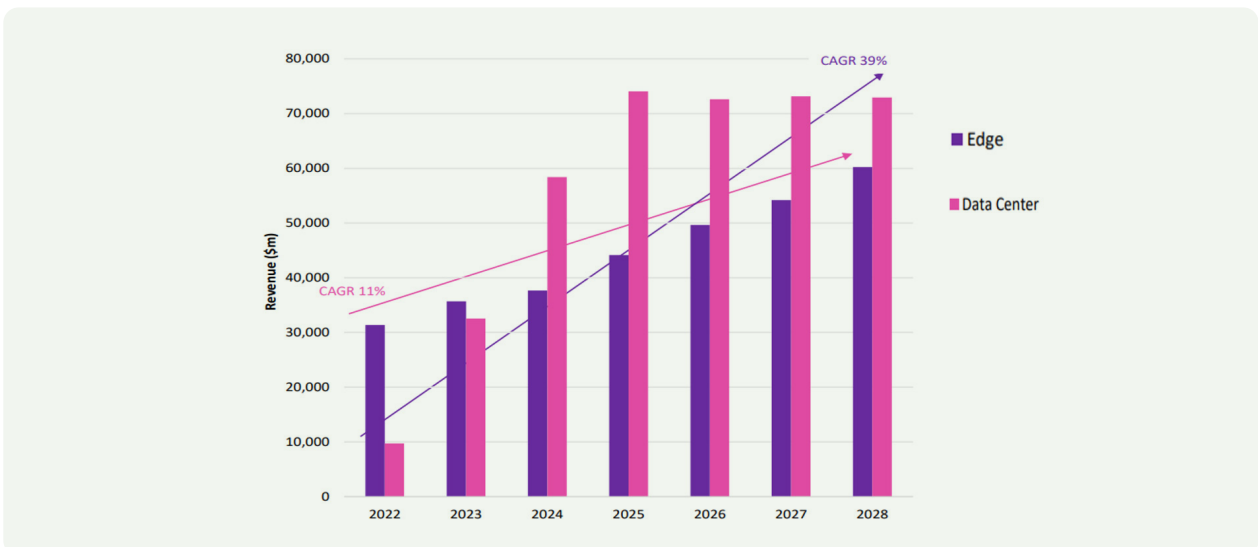
주: 매출액 기준임
자료: Gartner(2024.4b), 저자 재구성

그림 7 | 글로벌 Si반도체 기기별 비중 전망 (2023년, 2028년) (단위: %)



주: 매출액 기준임
자료: Gartner(2024.4b), 저자 재구성

그림 8 | 글로벌 엣지 디바이스 · 데이터센터용 Si반도체시장 매출액 규모 추이·전망 (2022년~2028년)



자료: OMDIA(2023.12), OMDIA(2024) 재인용

4 Gartner(2024.4c)

- 엣지 디바이스로 AI반도체 수요가 확대됨에 따라 데이터센터 수요 중심인 GPU 비중은 축소되는 반면, 다양한 수요처별 맞춤형 AI반도체(Other Application Specific)는 확대

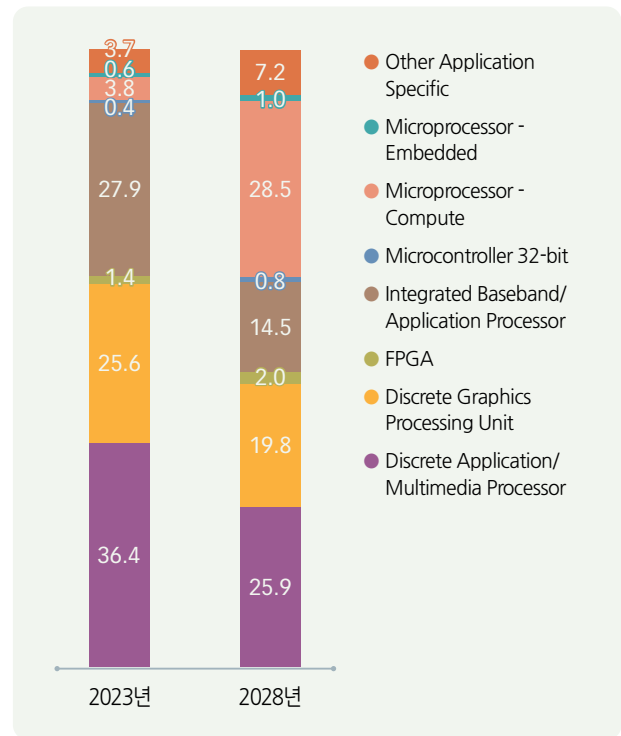
※ 매출액 기준 GPU가 전체 AI반도체에서 차지하는 비중은 '23년 25.6%에서 '28년 19.8%로 축소되는 반면, 수요처별 맞춤형 기타 AI반도체는 '23년 3.7%에서 '28년 7.2%로 늘어날 전망

그림 9 | 글로벌 AI반도체 제품별 전망(2023년, 2028년) (단위: 백만 달러)

	2023년	2028년	2023-2028 CAGR
Digital Signal Processor	74	493	46.2%
Discrete Application/Multimedia Processor	19,516	41,261	16.2%
Discrete Graphics Processing Unit	13,738	31,542	18.1%
FPGA	749	3,107	32.9%
Integrated Baseband/Application Processor	14,984	23,044	9.0%
Microcontroller 32-bit	232	1,293	41.0%
Microprocessor - Compute	2,061	45,266	85.5%
Microprocessor - Embedded	297	1,521	38.6%
Other Application Specific	2,012	11,495	41.7%

주: 매출액 기준임
자료: Gartner(2024.4b), 저자 재구성

그림 10 | 글로벌 AI반도체 제품별 비중 전망 (2023년, 2028년) (단위: %)



주: 매출액 기준임
자료: Gartner(2024.4b), 저자 재구성

I 글로벌 AI반도체 시장 경쟁 구조

- AI반도체 시장은 GPU와 기타(AP, Microprocessor, FPGA, ASIC) 구분하여 분석
- GPU의 경우, 기존 GPU 공급업체인 엔비디아, 인텔, AMD 등이 시장을 주도하고 있음
 - 특히 AI 연산을 위해 데이터센터에서 컴퓨팅용으로 사용하는 GPU의 경우, NVIDIA가 데이터센터 GPU 시장을 압도적으로 선도하며 시장점유율 92% 차지
 - ※ '23년 데이터센터 GPU 시장규모는 전년 대비 182% 증가한 490억 달러 수준이며, NVIDIA가 주도 (시장점유율 Nvidia 92%, AMD 3%, 기타 5%)⁵

5 IoT Analytics(2023.12.14)

- 반면, AP, FPGA, ASIC, Microprocessor 등의 경우, 아마존, 삼성, 애플, 테슬라 등 다양한 빅테크, 기존 반도체 업체들과 더불어 다수의 스타트업 기업들이 진출하고 있음

표 1 | AI반도체 공급업체 현황 I (비Mobile 분야)

반도체 기업	분류	AI chip Model	출시 년월
NVIDIA	Leading producer	GH200	March 2024
AMD	Leading producer	MI350	June 2024
Intel	Leading producer	Gaudi 3	April 2024
Alphabet	Public cloud & chip producer	Trillium	May 2024
AWS	Public cloud & chip producer	Trainum2	November 2023
IBM	Public cloud & chip producer	NorthPole	October 2023
Alibaba	Public cloud & chip producer	ACCEL	November 2023
SambaNova Systems	AI startup	SN40L	September 2023
Cerebras Systems	AI startup	WFE-3	March 2024
Groq	AI startup	LPU Inference Engine	November 2023
Apple	Upcoming producer	M4	May 2024
Meta	Upcoming producer	Artemis	April 2024
Microsoft Azure	Upcoming producer	Maia 100	November 2023
Rebellions	Upcoming producer	Rebel	January 2024
Graphcore	Other producers	Bow IPU	March 2022
Mythic	Other producers	M2000	March 2023
_etched	Other producers	Sohu	January 2024

주) 소개된 모델은 최신 발표를 중심으로 구성. 2024년 이전에 발표된 모델은 대부분 시장에 출시
 자료: aimultiple(2024.6.12.)

표 2 | AI반도체 공급업체 현황 II (Mobile 분야)

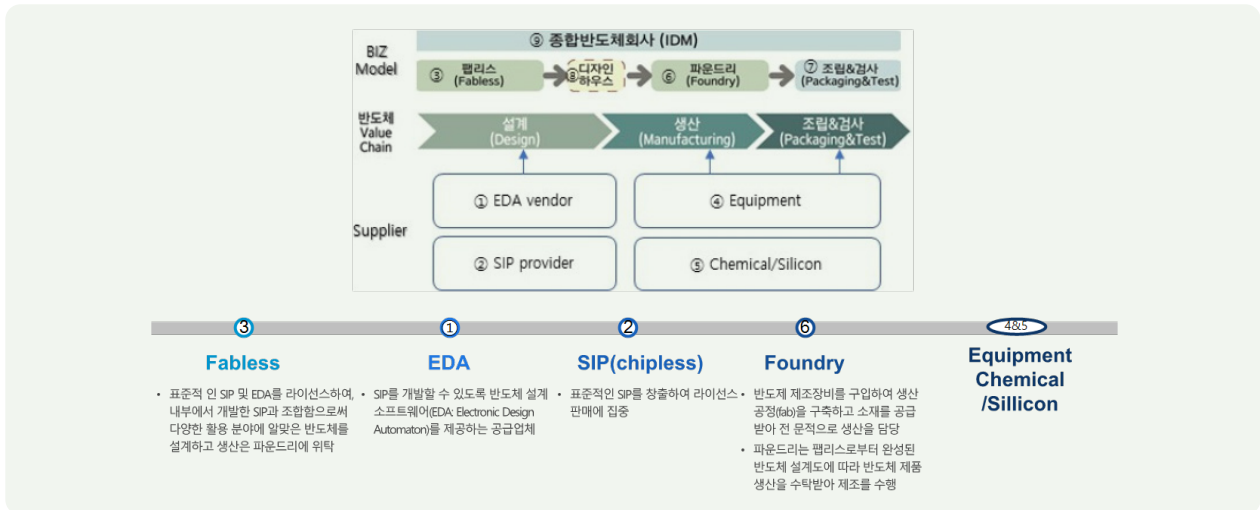
반도체 기업	AI chip Model	활용 기기 및 수요처
Apple	A 15 Bionic, A-16 Bionic	iPhones & iPads
Huawei	Kirin 9000 Huawei	smartphones
MediaTek	Dimensity 1200, Dimensity 9000	Oppo, Vivo, Realme, Xiaomi, POCO, Samsung
Qualcomm	Snapdragon 888, Snapdragon 8 gen 1 (2, 3)	Samsung Galaxy Series, Motorola, OnePlus, Oppo and Vivo, Pixel, Xiaomi
Samsung	Exynos 2100, Exynos 2200 (2400)	Samsung Galaxy devices

자료: aimultiple(2024.6.12.)

03 Si반도체 부상에 따른 반도체 산업의 가치사슬의 변화

- 반도체 시장에서 Si반도체의 부상은 기존 가치사슬을 변화시키고, 이는 반도체 산업 경쟁 구도와 가치 창출 방식의 변동으로 이어져, 나아가 생태계에 혁신을 가져옴
 - AI 모델 성능 강화에 대한 지속적인 경쟁과 급격한 AI 서비스 확산은 다양한 Si반도체에 대한 수요를 견인
 - * 현재까지 데이터센터에서 빅데이터를 학습·추론하는 대규모 AI 연산을 위해 동시에 병렬처리가 가능한 GPU 중심으로 수요 증가
 - * 최근 AI가 다양한 산업에서 활용됨에 따라, 기기 자체에서 AI 서비스를 제공하는 On Device AI가 관심을 받으며 고성능·저전력 Si반도체(NPU 중심)의 신시장이 창출되고 있음
 - Si반도체의 등장과 확산으로 기존 반도체 가치사슬의 주요 영역인 SIP·EDA⁶, Fabless, Foundry, Design House, OSAT⁷ 부문에서 변화가 발생하고 있음
 - * 반도체 설계 단계(Design): 다양한 Fabless 및 IDM 기업이 반도체 설계 SW(EDA)와 반도체 지식자산(SIP)을 외부 전문기업으로부터 공급받아 반도체 설계와 개발 진행
 - * 반도체 생산 단계(Manufacturing): Foundry가 Fabless·IDM에서 디자인한 반도체 설계를 위탁받아 생산하는 전(前)공정을 담당하고, 이때 외부 전문기업이 반도체 생산에 필요한 장비와 소재를 Foundry에 공급, 그리고 최종적으로 OSAT 업체가 패키징·테스트 등의 후(後)공정을 담당

그림 11 | 반도체 수평적 가치사슬의 범위



자료: 저자 작성

- 6 • SIP(Semiconductor Intellectual Property): 반도체 회로를 설계하는 데 적용·활용할 수 있는 표준 반도체 설계자산(IP)(해외기업으로 ARM, Synopsys 등이 존재)
- EDA(Electronic Design Automation): 복잡하고 미세한 반도체 회로를 설계(디자인)할 수 있는 필수적인 SW Tool(해외기업으로 Synopsys, Cadence, Mentor 등이 존재)
- 7 OSAT(Outsourced Semiconductor Assembly and Test): 반도체 조립과 패키지 및 테스트 공정

I Si반도체와 EDA의 변화

- EDA 공급업체들은 AI 기능이 적용된 제품 포트폴리오 확대
 - 기존 물리적 설계 관련 EDA뿐만 아니라 반도체 회로 설계의 검증, 테스트에 사용되는 EDA에도 AI가 적용
 - * 나아가 반도체 설계인력의 고유 영역으로 여겨지던 논리적 설계(Logical Design)에도 생성형 AI가 적용
 - 향후 3D, Chiplet 등 첨단 패키징 시장은 빠르게 성장할 것으로 전망됨에 따라, 반도체 설계뿐만 아니라 시스템(패키지) 레벨에서 사용되는 EDA의 중요성이 증가
 - * 3D, Chiplet 등 패키징 과정에서 시뮬레이션 SW가 패키징 관련 EDA 기능으로 제공
- Si반도체 수요가 증가하면서 EDA Tool의 중요성 증대
 - EDA를 이용한 반도체 설계 검증 과정의 정확성과 효율성이 AI 반도체 개발의 중요한 역량
 - EDA를 활용하면 Si반도체 설계에 걸리는 시간이 줄뿐만 아니라 오류를 감지해 불량률 감소
 - 파운드리업체는 공정에서 지원하는 EDA 툴을 늘려 고객사들의 선택 폭을 넓혀 사업 경쟁력을 강화

I Si반도체와 SIP의 변화

- Fabless 기업이 SIP 제공 사업을 동시에 수행하거나, 반도체 설계/검증용 소프트웨어를 공급하는 EDA 전문기업이 SIP 제공 사업을 동시에 진행함
 - * 대표적인 기업(Fabless+SIP)은 AMD와 퀄컴이며, ① 퀄컴의 경우 OTL(Qualcomm Technology Licensing)이라는 SIP 라이선스 사업을 유지 중이며 ② AMD의 경우 Semi-Custom 사업으로, 자체적으로 개발해서 보유하고 있는 Core IP를 활용해서 특정 고객사를 위한 맞춤형 반도체 설계
 - * 대표적인 기업(EDA+SIP)은 Synopsys, Cadence, Siemens로 이들은 EDA를 중심으로 SIP을 보완적으로 제공하는 플랫폼 방식으로 변화
- 소프트뱅크의 ARM Holdings와 같이 순수하게 반도체용 SIP 사업만을 영위하는 기업의 경우, 보완적인 SIP에 대해 인수합병이나 협업을 통해 통합 솔루션을 제공하는 플랫폼 전략 수행
 - Si반도체 SIP 개발의 복잡도 증가, Si반도체 설계시 필요한 검증된 SIP 확보, 연구개발 및 서비스인력의 추가 확보 등의 요구를 협업과 M&A로 해결
- Si반도체를 개발하는 주체가 구글, 아마존, 페이스북과 같은 최종 서비스 기업, 애플과 같은 최종 수요기업으로 다변화되어, SIP 회사의 고객도 기존 Fabless 외에 다양한 수요기업과 파운드리, 디자인하우스 등으로 확장

- 최종서비스 및 수요기업은 자사가 원하는 성능 및 가격을 담보할 수 있는 Si반도체를 직접 개발하기 위한 내부 인적 역량을 확보하고, 외부에서 조달할 수 있는 EDA 및 SIP을 활용

I Si반도체와 Fables·IDM의 변화

- 엔비디아·인텔·퀄컴·삼성 등은 데이터센터·모바일·자동차 등에 다양한 수요처에 활용되는 Si반도체 제품을 고도화하여, 자사의 시장점유율을 확대하려 함
 - * 삼성전자가 현대자동차와 차량용 인포테인먼트 분야에서 협력을 시작하여, 현대자동차의 차량에 프리미엄 인포테인먼트(IVI, In-Vehicle Infotainment)용 프로세서인 '엑시노스 오토 (Exynos Auto) V920'을 '25년에 공급할 예정⁸
 - * 엔비디아는 산업용 로봇과 같은 새로운 AI 시장을 창출하기 위해 자사 GPU 신제품인 블랙웰 등을 활용해 50조 달러 규모로 성장할 AI 자동화 시장을 공략한다는 계획⁹

I Si반도체와 Design House의 변화

- 디자인하우스의 수요기업에 대한 종합적인 지원역량 제고 필요성 증대
 - Si반도체의 다양화와 고기능화로 기존 반도체 제조 서비스의 부가적인 형태에서 반도체 제조 전반 서비스를 제공하는 토털 솔루션 서비스 형태로 진화
 - 반도체 설계 능력은 물론 공정 최적화, 패키징, 테스트, 유통 외에도 EDA 및 SIP 협업 능력까지 반도체가 설계되는 과정에서 판매되는 과정까지 전방위적으로 다양하게 요구되는 역할 수행이 요구
- 중소규모 디자인하우스의 지원역량 확보를 위해, 중소규모 디자인하우스 간 M&A를 통한 규모 확대, 업무영역 확장 등 추진 중

I Si반도체와 Foundry의 플랫폼 전략

- Foundry에서는 다양한 수요기업이(일반기업, Fables, IDM 등) 자사의 Si반도체 칩을 만들 수 있도록 자유도와 유연성을 제공하는 것이 중요해짐
 - 파운드리가 Si반도체의 성능향상을 위한 새로운 첨단 공정을 도입하면, 다양한 수요 업체가 이에 맞춰 원활히 반도체 설계를 할 수 있도록 파운드리와 협력하는 IP, EDA 등을 지원

8 삼성전자(2023.6.7)

9 뉴스1(2024.6.30)

- 대표적인 Foundry 기업은 AI반도체 생산 활동에 필요한 부가적·보완적 활동을 통합하여 제공하는 플랫폼 전략을 강화
 - 물리적인 반도체 생산 역량을 핵심기반에 두고, 다양한 IP·EDA와 함께 부가적인 서비스와의 협업을 통해 다양한 수요자 요구를 충족시킴

I AI반도체와 OSAT의 중요성

- 반도체 공정의 미세화가 한계에 이르면서 AI반도체 성능향상의 핵심기술 영역으로 OSAT가 발전하면서, OSAT의 가치활동 범위가 확대되어 협업이라는 요소가 필수적으로 등장
 - 첨단 미세공정 경쟁이 치열한 가운데, 기술적 한계와 비용 상승이라는 한계를 극복하기 위해 칩렛(Chiplet) 기술의 역할이 강조됨
 - * 칩렛기술은 여러 종류의 반도체를 하나의 패키지로 조립하는 기술로, 이를 통해 각각의 반도체가 역할을 나눠 수행할 수 있도록 해 전체적인 성능을 끌어올릴 수 있음
 - 칩렛 기술은 반도체 기업들이 미세공정의 한계를 극복하고, 개발 리스크를 감소시키며, 고객 중심 또는 애플리케이션별 반도체 양산을 더욱 신속하게 창출할 수 있는 새로운 방식 제시
- AI반도체의 설계 난이도의 어려움, 다양한 반도체와의 결합으로 인한 비용 최적화를 위해 3D 혹은 2.5D 방식으로 반도체 칩을 쌓아 올리는 적층 기술과 서로 다른 반도체칩을 연결하는 이중 집적 기술 대두

04 AI반도체 시장 변화 동인 분석

I (수요) On-device AI 활성화에 따른 AI반도체의 다양화

- 학습하고 추론하는 AI에 관한 관심이 높아짐에 따라, 앞으로는 스마트폰, 로봇, 가전, 자동차, 방산 등 엣지 디바이스(edge device)에서 광범위하게 AI 기술 적용이 가능
 - 기존 AI 기술은 스마트기기에서 수집한 정보(데이터)를 중앙 클라우드 서버로 전송해 분석하고 다시 기기에 보내는 방식으로 진행
 - 이와는 대조적으로 On-device AI는 네트워크로 연결된 클라우드 서버를 거치지 않고 엣지 디바이스에서 자체적으로 정보를 수집하고 학습 및 추론
 - On-device AI는 엣지 디바이스 내부에서 데이터를 처리(학습 및 추론)하기 때문에 지연이 없는 빠른 작업 수행이 가능
 - 중앙 서버를 통하지 않기 때문에 클라우드 기반 AI의 문제점으로 대두되었던 보안 문제(개인정보 보호)도 해결
 - 별도의 네트워크가 필요 없으므로 인터넷 연결이 어려운 상황에서도 실시간 학습 및 추론이 가능
- On-device AI Model, On-device AI SW, On-device AI HW 등 각 기술 계층에서 On-device AI 서비스를 실현하기 위하여 노력¹⁰
 - (Lightweight AI Model) 단말기라는 제한된 성능과 공간에서 AI 연산의 매개변수(parameter)를 줄이면서 성능의 저하를 최소화한 AI 모델 경량화 추진¹¹
 - * Meta의 Llama는 GPT-3보다 적은 매개변수(parameter)로도 벤치마크에서 우수한 성능을 보였으며, Llama 2는 parameter 규모 70억개, 130억, 700억개 등 총 3개의 모델로 구분
 - (AI Inference SW) On-device AI를 위해 모델개발 혹은 최적화를 지원하는 프레임워크 및 SW 기술이 지속해 연구¹²
 - * (Google) TensorFlow Lite 프레임워크를 통해 마이크로컨트롤러, 모바일 등에서 AI 모델변환, 실행, 최적화를 지원하는 모바일라이브러리 제공

10 이석준(2024)

11 Meta 웹페이지, <https://llama.meta.com/llama2>

12 김병수(2023)

- (On-device AI반도체) AI 모델이 가벼워지고 있는 동시에, 이를 지원하는 저전력·고성능의 AI반도체(AI Neuromorphic Chip & NPU) 성능은 강화

* (삼성전자) NPU를 독자적으로 설계하고 개발에 투자하고 있는데, '18년 엑시노스 9820에 처음으로 NPU를 탑재, 엑시노스 10시리즈(엑시노스 1080, 엑시노스 2100)에 모두 NPU가 탑재

● On-device AI의 활성화는 다양한 분야에서 활용되는 기기에 탑재되는 AI반도체의 세분화를 가져올 전망

- AI반도체는 사용처와 특성이 서로 다른 다품종·소량생산인 맞춤형 제품 생산으로 변화될 전망

표 3 | AI반도체의 수요처 및 기술별 세분화 현황

기술 수요처	AP	DSP	GPU	FPGA	Microcontroller 32-bit	Microprocessor- Compute	Microprocessor- Embedded	Other Application Specific
소비자 가전	Appliances, Digital Camcorder, Digital Set-Top Box, Digital Still Camera, Portable Media Players Smart Speakers TV, Smart, Video Game Consoles, Video Game Handhelds, Wearable ...							
유무선 통신	Smartphone, Mobile Infrastructure - Base Station Equipment, Wireless LAN Infrastructure Enterprise WAN, Optical Transport, Enterprise LAN, Service Provider Routers and Switches Fiber Broadband Access, DSL Broadband Access, Enterprise Telephony Equipment ...							
컴퓨팅	Server, Desktops, Graphics Cards, Laptops, Accelerator Cards ...							
자동차	ADAS-ECU, ADAS-Sensors & Connectivity, Automotive HPC, Body, Chassis, Powertrain, Infotainment, Safety ...							
산업	Surveillance, Other Security, Automation, Medical/Healthcare, Energy Management ,Transportation, Agriculture, Military/Civil Aerospace Electronics ...							

자료: 저자 작성

I (시장 참여자) 빅테크 기업의 AI반도체 시장진입 활성화

● Google, Amazon, Tesla, Apple 등 전통적인 반도체 기업이 아닌 인터넷 기업, 자동차 기업 등이 자사의 제품에 탑재하기 위한 AI반도체를 자체적으로 설계 및 개발

- MS·Google·Amazon·Tesla·Apple·Meta 등 글로벌 빅테크 기업은 전문성, 호환성, 안정성 측면에서 기존 반도체 기업의 AI반도체를 사용했으나, 최근 자사 기기·서비스에 특화된 맞춤형 AI반도체를 자체 개발 중

* 오픈AI는 엔비디아 GPU의 고비용·품귀현상 문제 해결할 목적으로 전방위적(IP·메모리·파운드리·투자사 등) 동맹을 모색하기 위하여 최대 7조달러(약 9,300조원) 규모 펀딩을 추진 (2024.2)¹³

13 관계부처합동(2024.4.25)

- 자사 AI 알고리즘 수행에 최적화된 AI반도체의 필요성으로 인해, 자체 AI반도체를 설계하는 단계를 지나서 최종적으로는 기존 반도체 생태계를 적극 활용하여 자사 AI반도체를 양산할 전망
 - * 기존 반도체 기업의 AI반도체로는 성능 강화에 한계가 있다고 판단, 자사 시스템에 맞춤형된 AI반도체를 사용해 제품 및 서비스를 개선하고자 하는 의지가 작용, 자사의 AI반도체를 직·간접으로 설계하고자 함
 - * AI반도체를 설계·생산하는 데에는 큰 비용이 소요되기 때문에, 자체 AI반도체를 개발하기 위해서는 충분한 자본력과 자체 수요가 보장되어야 함
- 현재 생성형 AI 기기·서비스 분야에 사용되는 AI반도체(GPU)의 약 80%가 엔비디아에 의존하는 독점구조로 고착되어 있어, 빅테크 기업에 중장기적인 공급망 리스크 회피 필요
 - * 엔비디아의 AI반도체(GPU)가 상당한 고가로 판매되고 있으며, 공급량이 수요를 따라가서 못하는 공급난으로 인해 빅테크 기업의 비용 부담이 커지고 있음
 - * 빅테크 기업의 AI 제품 및 서비스가 구현하는 성능이 다른데도, 동일하게 엔비디아 AI반도체를 사용하고 있으며, AI반도체 활용 따른 소비전력도 커다란 문제로 작용
 - * 최신 엔비디아 AI반도체를 주문해서 서비스 활용하는 데까지 소요되는 리드 타임이 평균 50주 이상
 - * AMD와 인텔은 엔비디아와 동등한 성능을 지닌 GPU를 생산하기 위해 노력하고 있지만, 경쟁 업체가 엔비디아와 경합하는데 필요한 '소프트웨어 생태계' 개발에 큰 비용과 시간이 소요될 전망

● AI반도체 시장에 빅테크의 참여에 따른 변화 전망

- Google, Amazon, Tesla, Apple, Meta 등 전통적인 반도체 기업이 아닌 인터넷, 자동차, 기기, 컴퓨팅 등 수요기업이 반도체 설계 조직을 내부적으로 구축하여, AI반도체 설계역량을 본격적으로 갖추기 시작
 - * AI반도체의 경우, 빅테크 수요기업이 직접 설계하고 Foundry(전문생산업체)에 위탁생산
 - * 빅테크 수요기업이 원하는 AI반도체는 ① 합리적인 가격, ② 사용처별 성능(사업적 필수조건: 전력소모, 수행능력, 사용 애플리케이션...) 등을 충족할 수 있는 맞춤형 AI반도체
 - * AI반도체의 주도권(설계능력: EDA/SIP에 대한 기본적인 이용역량)도 기존 전통적 팹리스를 넘어서 빅테크 기업 전체로 확대
- 최근 생성형 AI 시장 선점을 위해 인프라 영역의 AI반도체에 대한 국내외 빅테크 기업의 수요가 급증함에 따라, AI 서비스 시장 선점을 위해 AI반도체 기술을 보유한 전문업체에 대한 투자와 인수가 진행될 전망
- 반도체 생태계와 AI 생태계가 융합된 특성으로 인해 주요 기존 반도체 기업은 기존의 사업 영역을 확대할 전망

I (요소기술) AI반도체 성능의 최적화에 필수적인 시스템SW의 전략화

- AI서비스는 하드웨어·시스템SW·응용SW의 최적화로 결정되기 때문에 하드웨어뿐만 아니라 이를 운영하고 응용SW를 지원하는 시스템SW, AI모델, 서비스까지 모두 중요

- 특히, 핵심 인프라인 AI반도체를 효과적으로 활용하면서, AI 모델개발을 지원하는 핵심 인프라로서 시스템SW의 전략적 중요성 증가
 - 엔비디아의 경우, GPU 성능의 우월함도 상당한 강점이 있지만, 엔비디아 GPU 전용 AI 프로그래밍 개발을 지원하는 시스템SW인 CUDA 제공이 중요
 - * 엔비디아의 AI 모델 개발용 소프트웨어 인프라 쿠다는 AI 개발자에게 표준적인 지위를 확보 중
 - 국내외 AI반도체를 개발하는 팹리스 및 IDM 기업들의 경우 이러한 시스템SW도 함께 개발하여 제공하는 것이 요구
- On-device AI 환경에서는 제한된 자원을 활용하기에 효율성이 중요하고, 특히 특정 애플리케이션에 전문화된 시스템SW 개발이 요구
 - 다양하고 세부적인 활용 분야가 매우 많아 환경에 따른 하드웨어 및 소프트웨어 스펙이 서로 달라 범용적인 솔루션을 만들기 어렵다는 문제가 대두
 - On-device AI에서는 범용적인 솔루션이 대응할 수 없는 다양한 애플리케이션별 특성에 대응하기 위해 개별화된 AI반도체와 시스템SW가 요구

I (생태계) 주요 선도 기업의 플랫폼화

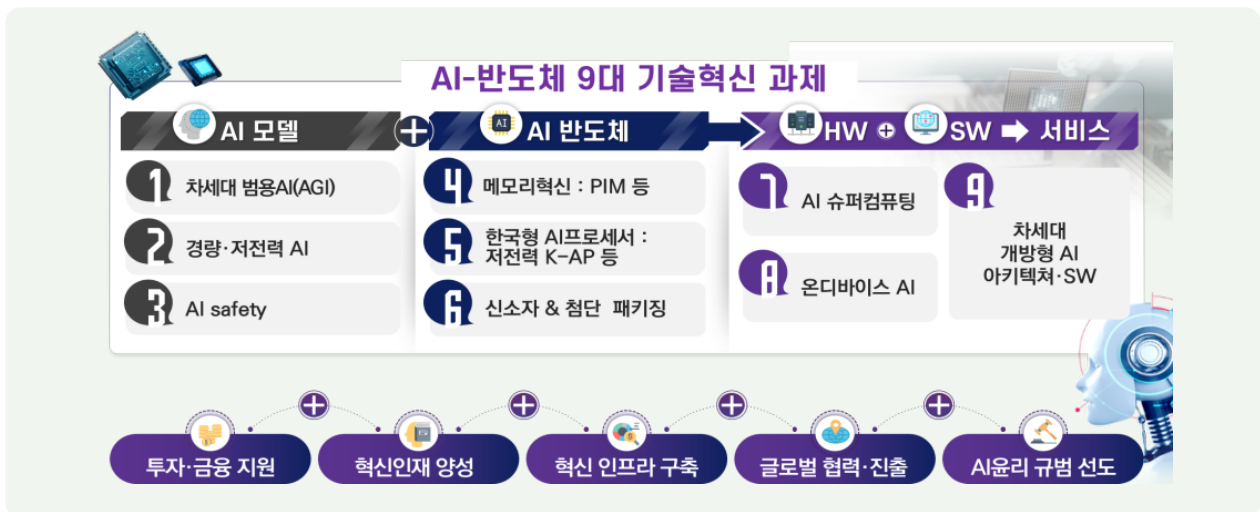
- AI반도체의 부상으로 주요 가치사슬의 영역의 선도 업체들은 플랫폼 방식으로 보완적인 자원들을 모아서 제공하는 전략을 강화함
 - SIP 선도기업(ARM), EDA 선도기업(Synopsys), Foundry 선도기업(TSMC, Samsung) 등은 자사 중심의 생태계를 구축하여 영향력을 확대하고 있음
 - 특히, 파운드리리는 주요 빅테크 및 팹리스 고객에게 첨단 미세공정 기반의 AI반도체 생산뿐만 아니라, AI반도체의 IP제공, 설계서비스, 패키징과 시스템 구성 최적화에 이르기까지 폭넓은 영역을 지원
 - * AI반도체 설계 및 제조는 단순히 하드웨어 설계·생산에서 그치는 것이 아니라, 사용처별의 요구를 깊이 이해하고, 이러한 요구에 맞는 AI반도체를 생산할 수 있어야 함

05 최근 AI반도체 정책 분석

I 정부의 제도적/공공 정책 시행으로 AI반도체 시장의 진입 계기가 발생하는 기회의 창으로써 최근의 종합적인 AI반도체 정책 분석

- 미국, 중국, EU, 일본 등 주요 반도체 기술 선도국가는 반도체 육성 및 보호 관련 정책을 통해 지속해서 정부의 공적 개입이나 제도적 조건을 변화시킴으로써 제도적인 기회의 창을 적극 활용
- 이에 대응하여 우리 정부도 최근 AI-반도체 이니셔티브 전략(2024.4.25)을 발표하며, AI반도체 중심의 종합적인 지원정책 수립
 - 우리나라 AI 가치사슬 분야별 강점과 요소기술을 분석하여 도출한 9대 기술혁신 과제와 이를 뒷받침하기 위한 중점 추진과제로 구성

그림 12 | AI-반도체 이니셔티브(안)

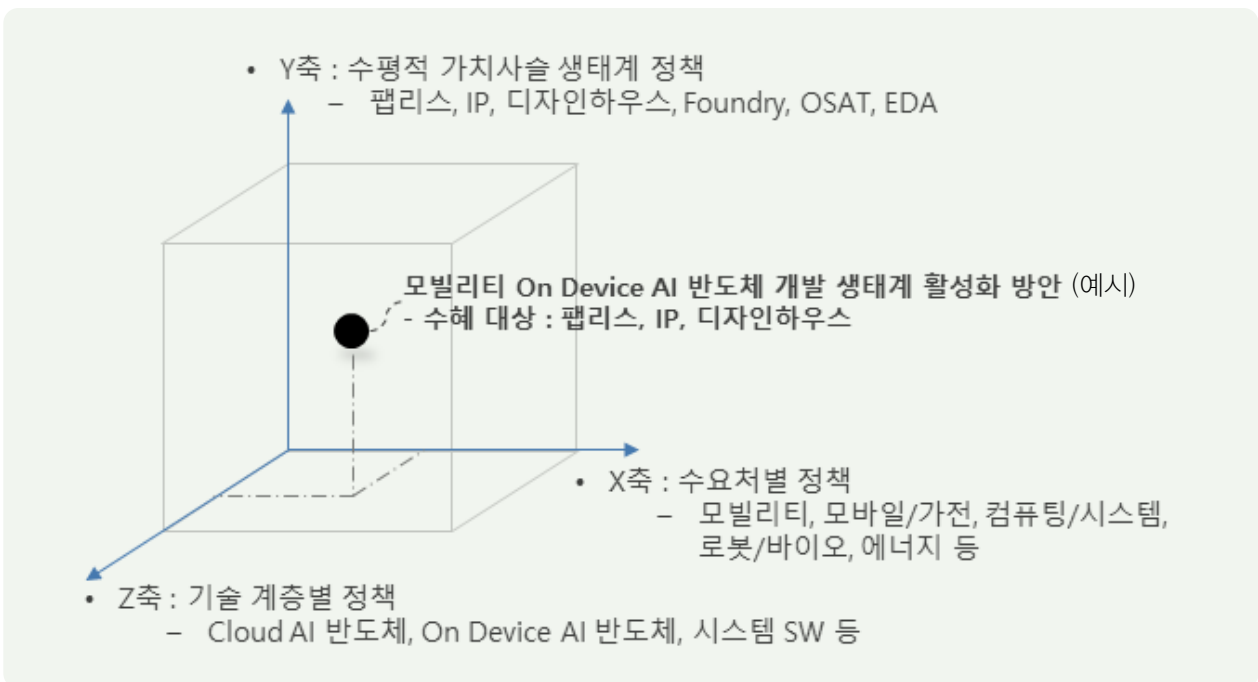


자료: 관계부처 합동

- AI-반도체 이니셔티브 전략의 경우 수요처별 지원정책, 수평적인 생태계 정책, 기술 계층별 지원정책이라는 3가지 관점에서 정책목표를 이루기 위한 정책의 조합이 상호 배타적이면서 전체 구조를 이루고 있음
 - AI-반도체 전략 이니셔티브를 수요처별 지원정책, 수평적인 생태계 정책, 기술 계층별 지원정책이라는 3가지 관점에서 입체적으로 분석하는 것이 필요

- * AI반도체 경쟁력을 강화하기 위해서는 i)특정한 수요처를 목표로, ii)수평적인 가치사슬 중 어떠한 영역에 정책적인 자원을 할당할 것인지, 그리고 iii)활용 생태계에서 특정 기술의 정책 지원 등을 분석
- 수요처별 지원정책은 ⑦AI 슈퍼컴퓨팅(K-클라우드 2.0 등) 및 ⑧온디바이스 AI 과제를 중심으로 데이터센터 및 특정한 애플리케이션에 활용되는 다양한 AI반도체 개발 지원
- 수평적 생태계 지원정책은 ④메모리혁신(PIM 등), ⑤한국형 AI프로세서(저전력 K-AP 등), ⑥신소자&첨단 패키징 과제를 중심으로 첨단 메모리, Fabless, OSAT 부분에 대한 투자 강화
- 기술 계층별 지원정책은 ⑨차세대 개방형 AI 아키텍처·SW 과제를 중심으로 AI반도체를 데이터센터와 온디바이스AI에 효과적으로 적용하기 위한 한국판 차세대 개방형 AI 아키텍처·SW에 대한 투자 강화

그림 13 | AI반도체 생태계 활성화를 위한 정책 구상의 틀(Framework)



자료: 저자 작성

06 정책적 시사점

- 최근 각국 정부는 반도체 산업에서 주도권 확보를 위한 양자 간 협정, 자국의 이익 추구, 다양한 가치사슬 영역에서의 영향력 강화 등을 목표로 정책을 수립
- 향후 Si반도체 지원정책은 기존 Si반도체 정책효과를 제고하기 위해 다음과 같은 방향으로 제도적인 기회의 창을 좀 더 적극 활용하는 구조화 필요
- 첫째, Si로 인한 환경변화에 따라 Si반도체 수평적 가치사슬 영역별 참여 기업의 경쟁력 부족한 부문에 지원하는 것이 필요
 - Si로 인한 환경변화에 기술적으로 대응하기 위해, 기존 반도체 산업에서 Si반도체 산출과 관련된 다양한 자원과 역량을 기업 내·외부에서 조달하거나, 협업 및 인수·합병해야만 하는 복잡한 기술환경으로 변화
 - Si반도체 설계의 복잡도가 극도로 고도화됨에 따라 기존 반도체 참여 기업은 M&A를 통하여 새로운 기술을 흡수하는 동시에, Si반도체 분야의 주도권을 조기 확보하는 것이 필요
- 둘째, Si반도체의 수요처가 다양해지고 다품종·소량생산으로 변화될 전망이어서 이를 지원하는 좀 더 세밀한 수요처별 정책이 요구
 - Si반도체 설계 능력 발전을 위해서는 결국 최종 수요자의 종합 시스템(최종 제품 Set + 최종 서비스)에 따른 설계역량이 중요해짐
 - Si반도체의 주도권이 기존 전통적 팹리스를 넘어선 중소규모의 최종 수요기업 전체로 확대될 전망
 - 기존 전통적 중소 팹리스 전문업체는 디자인하우스/SIP로 변화하거나 인수합병될 것으로 전망
 - 설계 최적화 및 테스트, 기타 지원 서비스를 제공하는 디자인하우스의 역할 중요
- 셋째, Si반도체의 수요처가 다양해지고 이를 지원하는 다양한 기술 계층 변화에 따라 발생하는 기회를 지원하는 기술 계층별 지원정책이 요구
 - AI 서비스가 갈수록 세분화, 특화하는 경향은 Si반도체 개발에 있어 기회로 작용
 - 특히 전력 효율성이 좋으면서 특화 연산 능력을 갖춘 Si반도체를 지원하는 풀스택에 관한 기술을 내부에서 개발하거나 다양한 파트너와 협력하여 제공하는 것이 필요

* 시스템SW, SDK, 최적화 솔루션, 보안 시스템 반도체, Si반도체 기판, 냉각 솔루션, 서버 공급 등 다양한 기술 풀스택에 대한 맞춤형 지원정책 제공

- 넷째, AI반도체 설계 및 생산 지원을 위한 반도체 선도기업의 플랫폼 전략이 강화되고 있어, 주요 플랫폼 생태계의 참여에 대한 지원과 더불어 국내 선도기업의 플랫폼 구축·강화를 위한 유인책 제공
 - 사업 범위와 역할, 보유역량이 주요 플레이어 사이에 겹치고 더욱더 포괄적인 영역에서 경쟁하는 동시에, 생태계 구축을 통해 협업하는 입체적인 가치사슬로 변화
 - 국내 파운드리 중심의 생태계 경쟁력을 강화할 수 있는 다양한 지원책이 필요
 - 플랫폼 선도기업과 참여 기업에 대한 인프라 투자비 지원, 참여기업에 대한 검증 및 상용화 지원, 생태계 참여에 따른 인센티브 제공

참고문헌

국내

관계부처합동(2024.4.25.), AI-반도체 이니셔티브(안).

김민식(2022), 반도체 분야의 기술주권 강화에 대한 시사점, ICT 통계 분석 22-02호, KISDI.

김민식, 김봉선, & 김언수(2024). 보완자 상호작용에 기반한 지식재산 생태계 분석 프레임. 전략경영연구, 27(1), 45-72.

김병수(2023), AI 반도체 개발 현황과 방향.

뉴스1(2024.6.30.), “[엔비디아의 교훈②]AI 넘어 산업 자동화 시장…항상 그다음을 본다.”

<https://www.news1.kr/industry/electronics/5462839>.

삼성전자(2023.6.7), “삼성전자, 현대자동차와 차량용 인포테인먼트 분야 첫 협력”,

<https://semiconductor.samsung.com/kr/news-events/news/samsungs-exynos-auto-v920-to-power-hyundai-motors-next-generation-in-vehicle-infotainment-systems/>

이석준(2024), 온디바이스 AI 기반 임베디드 최신 기술 동향, 세미나발표자료.

SKT(2023.10.22), “SKT AI 풀스택 글로벌 표준 최종 승인”, 보도자료.

해외

aimultiple(2024.6.12). “Top 20+ AI Chip Makers of 2024: In-depth Guide.”

<https://research.aimultiple.com/ai-chip-makers/#easy-footnote-bottom-1-36025>.

Gartner(2023.8). “Forecast Analysis: High-Bandwidth Memory, Worldwide, 2022-2027.”

_____ (2024.4a). “Semiconductors and Electronics Forecast Database, Worldwide, 1Q24 Update.”

_____ (2024.4b). “Forecast: AI Semiconductors, Worldwide, 2022–2028, 1Q24 Update.”

_____ (2024.4c). “Forecast: DRAM Market Statistics, Supply and Demand, Worldwide, 2021–2028, 1Q24 Update.”

IoT Analytics(2023.12.14). “The leading generative AI companies”,
<https://iot-analytics.com/leading-generative-ai-companies>.

Lee, K., & Malerba, F.(2017). Catch-up cycles and changes in industrial leadership: Windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Research Policy*, 46(2), 338–351.

OMDIA(2023.12). “Processors for Graphics and AI Market Tracker – 2J23 Analysis.”

_____ (2024.4). AI MOMENT, 세미나발표자료.

Soete, L.(1988). 21 Catching up in technology: entry barriers and windows of opportunity.

KISDI Perspectives 발간 내역

KISDI PERSPECTIVES는 국내 외 정보통신방송 관련 주요 정책 및 시장 동향을 분석한 리포트입니다.

문의 : 노희윤 전문연구원 (정보통신정책연구원 방송미디어연구본부, hyooooon@kisdi.re.kr, 043-531-4042)