

정책연구 2024-10

국내 AI 반도체 기업의 성장요인 분석을 통한 지원·육성 방안 연구

Fostering measures through analysis of the growth factors of
domestic AI semiconductor companies

김용기 · 김영환 · 박동운 · 이치호 · 이아람

내 부 저 자

연구책임자 | 김용기 | 과학기술정책연구원 부연구위원
연구참여자 | 김영환 | 과학기술정책연구원 연구위원
박동운 | 과학기술정책연구원 연구위원
이치호 | 과학기술정책연구원 부연구위원
이야람 | 과학기술정책연구원 연구원

기여자 및 자문위원

기여자 | 이형동 | 과학기술정책연구원 초빙연구위원
자문위원 | 강광욱 | Salisbury University 교수
전원 | 한국전자통신연구원 선임연구원
박종윤 | 고양시정연구원 부연구위원
채명식 | 한국과학기술기획평가원 부연구위원
김승범 | 흥의대학교 교수

정책연구 2024-10

국내 AI반도체 기업의 성장 요인분석을 통한 지원·육성 방안 연구

2024년 12월 31일 인쇄

2024년 12월 31일 발행

發行人 | 윤지웅

發行處 | 과학기술정책연구원

세종특별자치시 시청대로 370

세종국책연구단지 과학·인프라동 5~7F

Tel: 044)287-2000 Fax: 044)287-2068

登 錄 | 2003년 9월 8일 제2015-000014호

組版 및 印刷 | 미래미디어

Tel: 02)815-0407 Fax: 02)822-1269

ISBN 978-89-6112-967-1 93300

비매품

| 목 차 |

국문 요약 i

영문 요약 I

제1장 서론 1

제1절 관련 현안 및 연구 배경	1
1. 반도체 산업과 AI 반도체 산업의 개요	1
2. AI 산업의 발전과 AI반도체 산업의 부흥	3
제2절 연구 대상 및 주요 연구 내용	5

제2장 AI반도체 생태계 현황 및 문제점 7

제1절 AI 반도체 생태계 현황	7
1. 기술	7
2. 주요국 정책	28
3. 주요 기업	38
제2절 정부 R&D 투자 동향	65
1. 개관	65
2. 주요 정책	80
3. 지원 현황	87
4. 시사점	95

제3장 AI 반도체 생태계 마인드 맵 분석 98

제1절 개관 및 마인드맵 프레임워크	98
1. 기술 혁신/창업 생태계의 성장단계	99

2. 기술 스타트업의 생존 및 성장	102
3. 초기 기술 스타트업의 생존 및 성장 요인	103
4. 마인드맵 생태계 분석 프레임의 도출	105
제2절 프레임워크 요소별 분석	108
1. 기업 경쟁력 및 생태계 부문	108
2. 투자	126
3. 인재	137
4. 기술	145
제3절 기업의 성장 요인 도출	155
1. 서면 자문 및 분석	155
2. 주요 기업 인터뷰	161
3. 시사점	171
제4장 결론	175
제1절 결론	175
1. 시사점	175
2. 정책 제언	180
참고문헌	187

Contents

Summary (in Korean)	i
Summary (in English)	i
Chapter 1. Introduction	1
1. Current Issues and Research Background	1
2. Research Objectives and Main Research Contents	5
Chapter 2. Current Status and Problems of the AI Chip Ecosystem	7
1. Current Status of the AI Chip Ecosystem	7
2. Trends in Government R&D Policy Support	65
Chapter 3. Mindmap Analysis of the AI Chip Ecosystem	98
1. Overview and Mind Map Framework	98
2. Framework elements analysis	108
3. Derivation of growth factors for companies	155
Chapter 4. Conclusion	175
1. Conclusion	175
References	187

| 표 목 차 |

〈표 2-1〉 온라인 서비스별 월간 1억 사용자 달성을 걸린 시간	20
〈표 2-2〉 국내 인공지능 반도체 하드웨어 성능 현황	22
〈표 2-3〉 국외 인공지능 반도체 하드웨어 성능 현황	24
〈표 2-4〉 미국의 반도체 분야 수출·접근통제 주요 내용	31
〈표 2-5〉 중국 정부의 반도체 기금 운용 현황	32
〈표 2-6〉 TSMC의 신규 구축 중인 글로벌 제조시설	34
〈표 2-7〉 세레브라스 매출	50
〈표 2-8〉 Celestial AI의 투자유치	53
〈표 2-9〉 딥엑스 AI 주요 기술	60
〈표 2-10〉 반도체 분야 주관부처별 투자 현황(18~'23년)	66
〈표 2-11〉 반도체 주요 분야별 투자 현황(18~'23년)	67
〈표 2-12〉 반도체 분야 수행주체별 투자 현황	67
〈표 2-13〉 반도체 주요 연구개발 단계 별 투자 현황(18~'23년)	68
〈표 2-14〉 AI반도체 기술개발 분야 주요 사업	69
〈표 2-15〉 AI반도체 기업지원 분야 주요 사업	72
〈표 2-16〉 AI반도체 분야 영역별 지원 현황(18~'23년)	72
〈표 2-17〉 AI반도체 분야 정부R&D과제 심층 분석 결과	74
〈표 2-18〉 세부 분야의 구분과 정의	75
〈표 2-19〉 수행주체별 중점 추진 중인 세부기술(상위 3개 노출)	76
〈표 2-20〉 연구개발단계 별 중점 세부기술(상위 3개 노출)	78
〈표 2-21〉 시스템반도체 비전과 전략	81
〈표 2-22〉 인공지능(AI) 반도체 산업발전 전략 요약	82
〈표 2-23〉 인공지능(AI) 반도체 산업 성장 지원대책 요약	83
〈표 2-24〉 첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트	84
〈표 2-25〉 국산 AI반도체를 활용한 『K-클라우드』추진방안	85

〈표 2-26〉 AI-반도체 이니셔티브 (2024.4)	86
〈표 2-27〉 「차세대지능형반도체기술개발」사업 투자 현황('20~'24)	87
〈표 2-28〉 차세대지능형반도체기술개발(소자) 과제 지원 현황	88
〈표 2-29〉 차세대지능형반도체기술개발(소자) 지원과제 참여 기관	89
〈표 2-30〉 차세대지능형반도체기술개발(설계) 과제 지원 현황	90
〈표 2-31〉 차세대지능형반도체기술개발(설계) 지원과제 참여 기관	91
〈표 2-32〉 차세대지능형반도체기술개발(설계·제조) 과제 지원 현황	92
〈표 2-33〉 차세대지능형반도체기술개발(설계·제조) 지원과제 참여 기관	92
〈표 2-34〉 차세대지능형반도체기술개발(소자, 설계, 설계·제조) 과제 지원 현황 93	93
〈표 2-35〉 차세대지능형반도체기술개발(소자, 설계, 설계·제조) 과제 지원 현황 94	94
〈표 3-1〉 생태계 성장 단계별 주요 특징	101
〈표 3-2〉 국내외 AI반도체 대표기업의 스케일업 요인 비교	111
〈표 3-3〉 AI반도체 관련 주요 분야별 연구조직(대학·출연연) 현황	123
〈표 3-4〉 국내 반도체 계약학과 현황	139
〈표 3-5〉 주요국별 첨단 산업 관련 외국인 대상 비자 프로그램 현황	140
〈표 3-6〉 AI반도체 관련 부처별 주요 인재양성 사업	143

| 그 림 목 차 |

[그림 2-1] AI 반도체 관련 생태계	9
[그림 2-2] Large Language Model(LLM)들의 모델 매개변수 수	11
[그림 2-3] IP Provider를 활용한 설계 간소화와 그 비용	14
[그림 2-4] CPU, GPU, NPU의 비교	15
[그림 2-5] 머신러닝 연산을 위한 데이터의 분배와 이동	16
[그림 2-6] 머신러닝 연산을 위한 소프트웨어 스택	17
[그림 2-7] 트랜스포머 모델의 학습 파라미터 및 연산량 규모	19
[그림 2-8] 국내외 다양한 인공지능 반도체	21
[그림 2-9] 인공지능 반도체 단일 칩 FP16/BF16 성능 비교	25
[그림 2-10] 다양한 인공지능 반도체의 SDK 또는 소프트웨어 스택	27
[그림 2-11] DARPA의 ERI 2.0 지원 기술 분야	29
[그림 2-12] NSTC의 역할과 목표	30
[그림 2-13] TSMC의 글로벌 제조시설	34
[그림 2-14] Overview of the Chips for Europe Initiative	36
[그림 2-15] NVIDIA의 Full-Stack Platform	39
[그림 2-16] NVIDIA의 연도별 특허 출원 동향	40
[그림 2-17] NVIDIA의 주요 사업 모델	41
[그림 2-18] NVIDIA의 주요 사업 모델	42
[그림 2-19] 사업 구조별 반도체 기업 분류	43
[그림 2-20] ARM Architecture	43
[그림 2-21] ARM의 라이센스 및 로열티 수취 모델	44
[그림 2-22] ARM의 라이센스 및 로열티 수취 모델	45
[그림 2-23] ARM의 IoT 토탈 솔루션	46
[그림 2-24] ARM의 주요 사업	47
[그림 2-25] 하이퍼엑셀 오리온 서버 성능	64
[그림 2-26] 왼쪽부터 '19~'23년간 중점분야, 수행주체, 연구개발단계별 투자 비중	64

.....	73
[그림 2-27] (참고) 세부기술 별 연구개발 단계 비중	80
[그림 3-1] 기술순환 모형	100
[그림 3-2] 혁신 깔때기	102
[그림 3-3] 퓨리오사AI의 2세대 레니게이드 성능 비교(vs NVIDIA)	108
[그림 3-4] AI반도체 구분에 따른 제품전략의 다양성	109
[그림 3-5] AI반도체 PoC를 위한 MPW의 개념 및 필요성	115
[그림 3-6] AI반도체 설계기업 단계별 성장요건	117
[그림 3-7] AI반도체 산업의 가치사슬 및 참여기업	119
[그림 3-8] AI반도체 산업생태계 내 협력 구조	121
[그림 3-9] 삼성 파운드리 SAFE 프로그램의 개념	126
[그림 3-10] 주요 국내 AI반도체 팝리스 유치 실적	134
[그림 3-11] 아키텍처 논리흐름도	146
[그림 3-12] NVIDIA 인터커넥트 기술 개괄	149
[그림 4-1] 반도체 분야 글로벌 투자 현황	179
[그림 4-2] AI 바우처 운영 개념	184

| 요약 |

1. 서론

□ 연구 배경 및 목적

○ 주요 현안

- AI 반도체 시장의 급성장: AI 산업 발전과 함께 AI 반도체 수요 급증, 특히 생성형 AI의 등장으로 폭발적 성장 중
- 글로벌 경쟁 심화: 미국 기업 중심의 IP, EDA 및 첨단 반도체 팩리스 시장에서 한국 AI 반도체 설계 기업의 경쟁력 확보 필요
- 국내 AI 반도체 생태계 미성숙: 핵심기술 부족, 인재 부족, 투자 부족 등 다양한 문제점 존재
- 정부 지원 정책의 한계: 다양한 정책 및 지원에도 불구하고, 구체적인 성장 전략 및 생태계 지원 부족

○ 연구 대상

- NPU 설계 팩리스 스타트업 및 기업: NPU 설계에 집중하는 기업 (퓨리오사 AI, 리밸리온, 사피온, 모빌린트, 딥엑스 등) 및 AI 반도체 생태계

○ 연구 방법

- 문헌 연구: AI 반도체 산업 동향, 정부 정책, 해외 사례 등 관련 자료 분석
- 사례 분석: 국내 AI 반도체 기업의 사업 모델, 기술, 투자, 인재 현황 분석
- 마인드 맵 분석: 사업 모델, 기술, 투자, 인재 프레임워크를 기반으로 AI 반도체 생태계 문제점 분석
- 기업 및 주요 관계자 인터뷰: 대면 및 서면 인터뷰를 통해 기업의 성장과 생존에 영향을 미치는 요인 도출

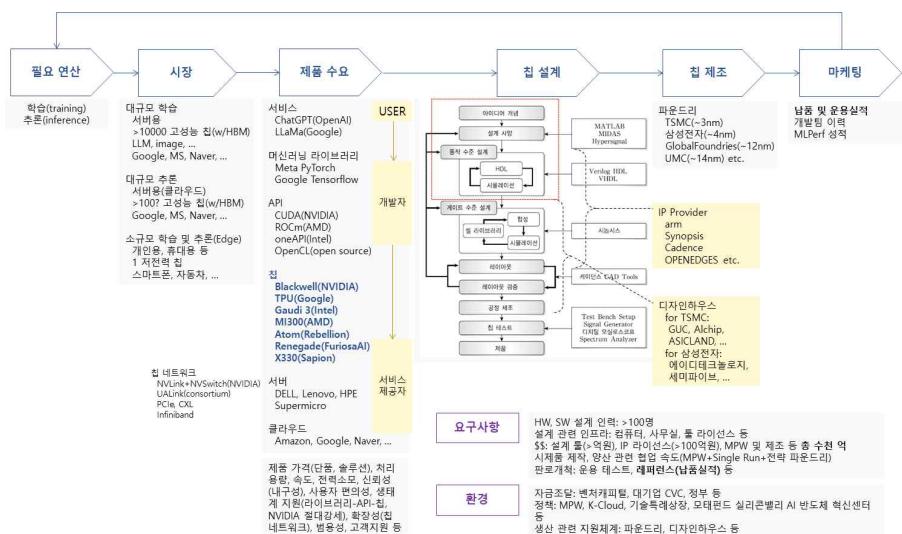
○ 주요 연구 내용

- 국내 AI 반도체 생태계 현황 및 문제점 분석 (사업 모델, 생태계, 기술, 인재, 투자)
- AI 반도체 기업의 성장과 생존 요인 도출
- AI 반도체 생태계 발전을 위한 정책 시사점 제시

2. AI반도체 생태계 현황 및 문제점

□ AI반도체 관련 생태계

[그림 1] AI반도체 관련 생태계



1. 인공지능 구현 단계:

- 학습 (Training):

- 인공지능 모델 생성 단계, 방대한 데이터를 인간의 뇌처럼 처리하여 모델 생성
(예: GPT-3는 753GB 데이터, 1750억개 매개변수 사용), 데이터 처리 방식:
인공신경망 (ANN), 딥러닝 (DL) 등 사용

- 목표: 인간의 언어, 그림 등을 이해하고 생성하는 능력 학습
 - 특징: 방대한 데이터, 큰 모델 크기, 높은 연산량 필요
- 추론 (Inference):
- 생성된 모델을 사용하여 질문에 답변하는 단계, 예: 영문 번역, 이미지 인식 등, 학습 단계보다 연산량 적음
 - 특징: 빠른 응답 속도, 정확도 중요

2. 시장:

- 대규모 학습:
 - LLM (Large Language Model) 등 거대 모델 학습
 - 막대한 컴퓨팅 자원 필요 (예: GPT-3 학습에 285,000개 CPU, 10,000개 GPU 사용)
 - 고성능 칩 (초미세공정), 대용량 메모리 (HBM) 필수
 - 주요 기업: NVIDIA, Google, Microsoft, Amazon, Meta 등
 - 최근 트렌드: 자체 NPU 개발 증가 (비용 절감 목표)
- 대규모 추론:
 - LLM 등 거대 모델 사용, 빠른 응답 속도 중요
 - 다수 사용자 요청 처리, TCO (Total Cost of Ownership) 중요
 - 주요 기업: NVIDIA, FuriosaAI, Rebellion, Saption 등
- 소규모 학습 및 추론:
 - 로봇, CCTV, 스마트폰 등 단말기 수준에서 연산 수행
 - 이미지 인식 등 특정 목적에 맞는 소규모 모델 사용
 - 저전력, 저가격 중요
 - 다양한 기업 참여 (대기업부터 중소기업까지)

3. 제품 수요:

- NPU 칩:

- 학습 및 추론 연산의 핵심 요소
- 가격, 처리용량, 속도, 신뢰성 중요
- 시장별 요구사항 상이 (대규모 학습 vs. 소규모 추론)

- 생태계 지원 필수:

- 라이브러리-API-칩-서버 연동 중요
- NVIDIA CUDA 생태계 우위 (머신러닝 라이브러리 지원, 개발자 친화적)
- 국내 기업, 소프트웨어 지원 확보 필요

- 연결성 중요:

- 대규모 연산 시 NPU 칩 간 고속 데이터 전송 중요
- NVIDIA NVLink, NVSwitch 기술 우위

4. 칩 설계:

- 칩 설계 과정: 아이디어 → 아키텍처 설계 → RTL 설계 → 게이트 레벨 설계
→ 레이아웃 설계 → 검증
- IP 활용: Synopsys, Arm 등 IP 제공 회사로부터 주변 기능 라이선스
- 디자인하우스 협업: 파운드리별 공정에 맞춰 레이아웃 설계, 오류 수정
- 핵심 역량: 아키텍처 설계 (연산, 메모리 효율성)

- 주요 기술:

- 데이터 타입 최적화 (32bit → 16bit, 8bit, 4bit)
- AI 개발툴 호환성 (PyTorch, TensorFlow 등)

- 칩 간, 노드 간, 시스템 간 데이터 이동 및 통신 기술

5. 칩 제조:

- 파운드리: TSMC, 삼성전자 (초미세공정)
- TSMC 우위: 칩 성능, 고객 지원 우수
- 삼성전자: TSMC 대안, MPW (Multi-Project Wafer) 기회 부족 등 개선 필요

6. 마케팅:

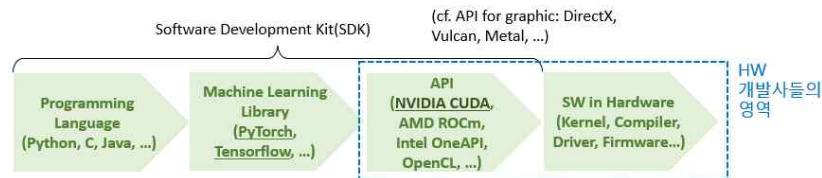
- 납품 실적, 개발팀 이력 중요: 고객 신뢰도 확보
- 벤치마크 참여: MLPerf 등 성능 평가 (실제 운용 환경 반영 한계)
- MLPerf: 학습, 추론 성능 측정, 다양한 모델, 데이터셋 사용

7. 거대 인공지능 시대:

- 트랜스포머 모델 등장: 수천억~조 단위 파라미터 (Llama-3.1: 4,000억개)
- 인공지능 서비스 사용자 폭증: ChatGPT, Gemini, CLOVA X 등 (ChatGPT: 출시 2달 만에 월간 1억명)
- 대규모 학습 시스템 필수: (Llama-3.1 학습에 NVIDIA H100 GPU 16,384개 사용)
- 인공지능 반도체 기술 발전 방향:
 - 단일 칩 성능 향상, 컴퓨팅 디바이스 간 데이터 전송 기술, 인공지능 분산 학습 기술
- 시장 현황: NVIDIA 시장 독점, 국내 기업 참여 필요

○ AI반도체 칩 설계 기술의 세 가지 특성

- Software Stack

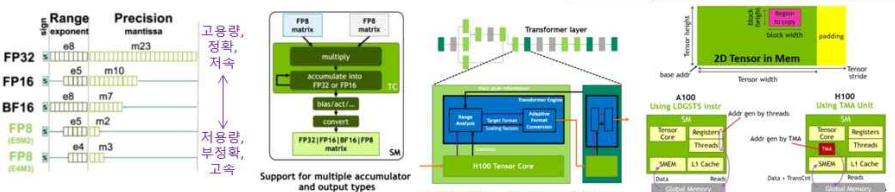


- 사용자(서비스 개발자)가 효율적으로 SW를 개발하기 위해서는 기존에 많은 연구자, 개발자들이 축적한 PyTorch, Tensorflow와 같은 machine learning library*를 활용할 수 있어야 하므로, HW 개발사는 SDK 혹은 API, SW in HW를 충실히 제공해야 함
 - C, C++등의 PGM 언어를 사용해 처음부터 서비스를 개발하는 것은 막대한 고급 개발자 자원이 필요하므로, 기업 규모가 작아지면 시도하기 어려움
 - *: 머신 러닝에 필요한 연산을 미리 프로그래밍해 모듈화 한 프로그램의 모음
- 폭넓은 연산 지원, 운영 실적, 참고할 만한 개발 문서의 양과 질, 커뮤니티 지원(Q&A) 등 서비스 개발 시 네트워크 효과가 작용
 - NVIDIA CUDA는 현존하는 대부분의 연산을 지원하며 등장이 검증된 API**로, PyTorch, Tensorflow 모두 고속연산을 위해 NVIDIA CUDA를 필수적으로 요구하므로***, NVIDIA의 절대적인 시장지배력의 근간이 되고 있음
 - **: HW를 원활히 구동하기 위한 SW 계층, 기업 고유 혹은 범용(OpenCL)이 있음
 - ***: 최근 AMD ROCm, Intel OneAPI 등의 지원이 확대되고 있으나, 완성도가 충분히 높지 않음

- Hardware Architecture

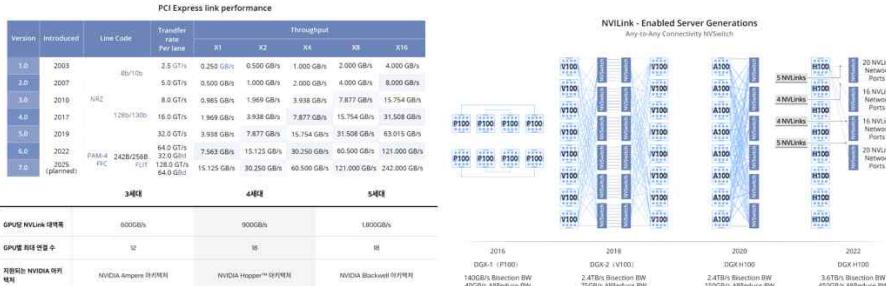
- ① Data type* 최적화: 높은 학습, 추론 정확도와 빠른 속도를 양립할 수 있도록 최적화
- ② 다양한 data type과 연산에 대응
 - *: 하나의 숫자를 표현할 때 사용하는 data의 양, 정밀도
 - **: ①, ②는 대체로 trade-off 관계에 있으며, 극복을 위해서는 정교한 architecture 구현이 필요
- ③ 메모리의 효율적 사용: 제한된 메모리 용량 하에서 데이터의 불필요한 이동을 최소화해야 함. 고대역폭 메모리(HBM) 도입 및 활용 기술 필요
 - **: 현재 머신 러닝 연산의 bottleneck은 메모리 용량 및 대역폭
- ④ 고속, 저전력, 저 발열: 최신 미세공정(3~5nm) 요구, 비용 상승의 주 원인

HW Spec.	FuriosaAI Warboy	NVIDIA H100 SXM	Rebellions ATOM
Tech node	2022, 14nm, 180mm ²	2022, 4nm, 814mm ²	2023, 5nm, -
Data types	INT8	FP64, TF32, FP16, BF16, FP8, INT8, INT4, INT2	
16-bit TOPS	-	989.4 FP16/BF16	32 FP16
8-bit TOPS	64 INT8	1978.9 FP8	128 INT8
On-chip memory size(MB)	32	116 (50+66)	64
Off-chip memory size(GB)	32, LPDDR4X	80, HBM3	16, GDDR6
Off-chip memory bandwidth(GB/s)	66	3352	256
16-bit TOPS/W	-	1.413	0.213
8-bit TOPS/W	1.067	2.827	0.853
Peak Power(W)	60	700	150
Major objective	AI vision inference	LLM training	AI vision, NLP inference



- Interconnect

- Chip간, 노드간, 시스템간 고속 데이터 이동 및 통신 구현 기술 필수
 - LLM과 같은 대규모 모델의 경우 단일 칩으로는 연산 불가능: 학습에는 10,000개 이상의 GPU와 TB 단위의 메모리가 사용되며(GPT-3.5)¹⁾, 추론의 경우에도 320~640GB의 메모리가 필요할 것으로 추산²⁾
- 범용 기술(chip간): PCIe gen5, 시스템간: Infiniband 등)이나 경쟁사 기술에 비해 NVIDIA NVLink, NVSwitch가 월등한 대역폭과 구성 유연성을 갖추고 있는 것으로 알려져 있음



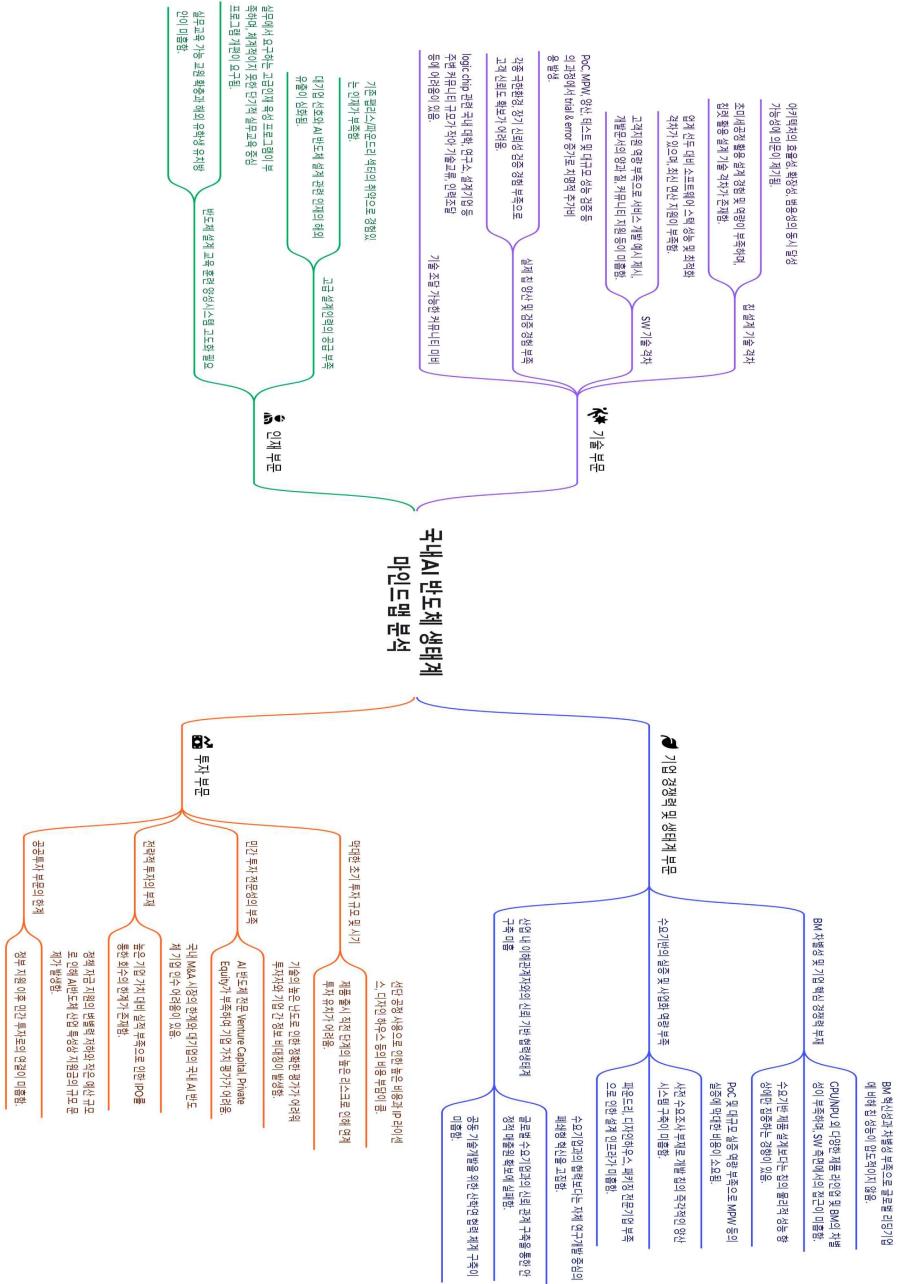
<표 1> 주요국 정책

국가	정책 목표	주요 내용	국가별 특징
미국	첨단기술 확보, 생산역량 확충, 기업 지원	<ul style="list-style-type: none"> 반도체과학법('22.8) 국가반도체기술센터(NSTC) 설립 - 경쟁국 견제 (중국, 러시아 등) 첨단 반도체 기술 연구 및 인력 양성 	<ul style="list-style-type: none"> 국가적 차원의 지원정책 추진 제조시설 리쇼어링 및 NSTC 설립을 통한 제조역량 강화 및 첨단기술 연구 경쟁국의 기술개발 견제 장치 마련
중국	핵심 기술 확보 및 장비 국산화	<ul style="list-style-type: none"> 차세대 인공지능 발전계획('17) - AI 반도체 기술 선도 ('25) 14차 5개년 계획 및 2035 중장기 목표 ('21) 	<ul style="list-style-type: none"> AI, 반도체 분야 국가 전략화, 핵심 기술 확보 및 장비 국산화 추진 미국의 제재로 자립화 및 기술 선도 담보 자국산 AI 반도체 사용 시 보조금 지원 및 수요 창출 지원
EU	첨단 반도체 생산능력 확대, 전문인력 양성	<ul style="list-style-type: none"> 유럽 반도체법(Chips Act, '23.9) 유럽 반도체 이니셔티브(CEI) 발족 2나노 미만 소자 설계 및 제조 기술 연구 지원 반도체기금(Chips Fund) 조성 	<ul style="list-style-type: none"> 유럽 반도체 생태계 활성화 및 공급 역량 확보 노력 반도체 칩 설계, 제조, 패키징 역량 강화 전문 인력 양성 및 교육 지원 강화
대만	첨단 제조 기술 확보, 핵심 소재 및 장비 경쟁력 강화, 고급 인재 육성	<ul style="list-style-type: none"> 산업혁신 조례 ('23.1), 대만형 반도체 법 인공지능 반도체 제조공정 및 칩 시스템 R&D 프로젝트('18~) 대만반도체연구소(TSRI) 설립 ('19) AI on chip Taiwan Alliance(AlTA) 출범 ('19) 	<ul style="list-style-type: none"> 첨단 공정 개발 촉진, 반도체 기업 활동 기반 강화 R&D 및 첨단 설비 투자에 대한 세액 공제 등 인센티브 제공 해외 투자 확대 중인 TSMC에 대한 자국 내 유인책 마련

<표 2> 주요 기업의 핵심 제품과 경쟁력

기업	개요	핵심 제품	핵심 경쟁력
NVIDIA	미국 캘리포니아주에 위치한 펩리스 기업. GPU, API, SoC 등을 개발하며 하드웨어 및 소프트웨어 분야에서 뛰어난 성과를 보임. 2024년 6월 시가총액 세계 1위 달성.	GeForce RTX 시리즈 (게이밍용 GPU), 데이터센터 제품 A100, H100 서버용 GPU, Grace-Hopper Superchip, 등	GPU 기술력, CUDA 기반 소프트웨어 생태계, AI 솔루션, 다양한 산업 분야 진출, 매출
ARM	1990년 영국에서 창립된 IP 코어 및 블록 제공 기업. 모바일 AP 부문 글로벌 선두 IP 기업.	Arm 아키텍처, Armv9, Arm AE IP	저전력 RISC 프로세서, 시장별 제품 다각화, 높은 점유율, IoT/임베디드 시장 경쟁력
세레브라스	2015년 설립된 AI 하드웨어 스타트업. 딥러닝 애플리케이션에 특화된 하드웨어 솔루션 제공.	Wafer-Scale Engine (WSE), CS-3 AI 슈퍼컴퓨터	WSE 칩, 대규모 AI 모델 훈련, 빠른 AI 칩, 의료 분야 협력
Celestial AI	2016년 설립된 AI 가속기 회사. 광학 상호 연결 기술 기반 컴퓨팅 시스템 제공.	Photonic Fabric™	광학 상호 연결 기술, 컴퓨팅 방식 전환, 대규모 AI 모델 지원
Hailo	이스라엘의 엣지 AI 반도체 스타트업. AI 프로세서 및 가속기 설계 및 제조.	Hailo-8 AI 프로세서, Hailo-10, Hailo-15	엣지 AI 칩, 저전력, 고성능, 다양한 분야 적용
퓨리오사 AI	2017년 설립된 NPU 및 소프트웨어 개발 기업. 서버용 NPU 시장 공략	WARBOY, RENEGADE	고성능 저전력 NPU, 소프트웨어 역량, 국내외 파트너십, 시장 경쟁력
리밸리온	GPU 대비 에너지 효율성이 높고 전력 소모가 낮은 AI 반도체 개발.	아이온(Ion), 아톰(Atom), 리벨(Rebel)	고성능 저전력 AI 반도체, 에너지 효율성, 뛰어난 인적 자원, 전략적 협력 관계
딥엑스	한국의 엣지 및 서버 AI 응용 분야에 맞춤형 AI 반도체를 개발	DX-V1, DX-V3, DX-M1, DX-H1	다양한 엣지 및 서버 AI 응용 분야에 맞춤형 AI 반도체, 독자적인 기술력, 컴퓨터 비전 및 언어 처리 성능
모빌린트	2019년 설립된 고성능 저전력 엣지 NPU 반도체를 개발하는 기업.	에리스(ARIES), 레귤러스(REGULUS)	고성능 저전력 엣지 NPU, 소프트웨어 개발 키트(SDK), 다양한 AI 모델 지원
하이파엑셀	2023년 설립된 저비용 저지연 LPU를 개발하는 기업.	LPU, 오리온 서버, 베르다 AI 반도체	저비용 저지연 LPU, FPGA 기반 AI 서버, 고성능

3. AI 반도체 생태계 마인드맵 분석



x 국내 AI반도체 기업의 성장 요인분석을 통한 지원·육성 방안 연구

가. 기업 경쟁력 및 생태계 부문

1. BM(Business Model) 차별성 및 기업 핵심 경쟁력 부재

1.1. BM 혁신성 및 차별성 부족

- 글로벌 리딩기업에 비해 입도적이지 못한 칩 성능,
- GPU/NPU 외 다양한 제품 라인업 및 BM의 차별성 부족,
- BM 설계에서의 SW 측면에서의 접근 부족(HW 측면만 강조)

1.2. 기업 핵심역량(Core Competence) 및 시장전략의 부재

- 수요기반 제품 설계보다는 칩 칩의 물리적 성능 형상에만 집중
- 지금조달, 협력제휴, 시장진출 등의 경영전략의 차별성 부족, 성능대비 가격경쟁력을 고려한 틈새시장(미국 외) 진입전략 부재

1.3. 창업팀 및 혁신인재 경쟁력 부족

- 창업자/CEO 외 AI 반도체 개발 실무경험 및 글로벌 선도기업 이력 부족
- 기술-비즈니스 간 균형 잡힌 창업팀/경영진 구성 한계
- 글로벌/국내 대기업 대비 우수인재 영입 및 유지 한계

2. 수요기반의 실증 및 사업화 역량 부족

2.1. PoC 및 대규모 실증 역량 부족

- PoC의 중요함에도 불구하고 MPW 등의 실증에 막대한 비용 소요
- AI 기술의 상용화 제한 등으로 대규모 실증 기회 부족(데이터 확보 문제)

2.2. 제품 양산 전략 및 레퍼런스 부족

- 사전 수요조사 부재로 개발 칩의 즉각적인 양산 시스템 구축 미흡
- 대기업 외 다양한 수요처 부재로 대규모 양산을 위한 사전 레퍼런스 부족

2.3. 칩 제조를 위한 전후방 밸류체인(공급망) 미흡

- 파운드리, 디자인하우스, 패키징 전문기업 부족으로 인한 설계 인프라 미흡
- 벤처스타트업에 친화적이지 않은 국내 파운드리 기업 (MPW 빈도 등)

3. 산업 내 이해관계자와의 신뢰 기반 협력생태계 구축 미흡

3.1. 다양한 수요기업과의 신뢰협력 관계 구축 노력 부족

- 수요기업과의 협력보다는 자체 연구개발 중심의 폐쇄형 혁신 고집
- 글로벌 수요기업과의 신뢰 관계 구축을 통한 안정적 매출원 확보 실패
- 유력 제품 수요자와 협력 부족으로 다양한 실사용 환경에서의 제품 피드백 어려움

3.2. 공동 기술개발을 위한 산학연 협력 체계 구축 미흡

- 대학 및 공공연구기관 등의 기술개발 선도 역할 및 공공기술 이전 성과 미흡
- 공동 기술개발을 위한 개발자 커뮤니티 및 협력 플랫폼 구축 애로

나. 투자 부문

1. 막대한 초기 투자 규모 및 시기

1.1. 막대한 초기 투자 필요성

- 선단 공정 사용으로 인한 높은 비용 (수백억 원 단위)
- IP 라이센스, 디자인 하우스 등 비용의 부담 (국산 반도체 설계에서도 케이던스, 시냅시스 등 해외 IP, EDA 툴의 높은 지배력)

1.2. 어려운 중후기 메가 투자 연계

- 제품 출시 직전 단계의 높은 리스크
- 가시적 성과 부족으로 인한 연계 투자 유치 어려움

2. 민간 투자 전문성의 부족

2.1. 투자자와 기업 간 정보 비대칭 (ROI 예측의 불확실성)

- 기술의 높은 난도로 인한 정확한 평가가 어려움 (기술에 대한 이해 부족)
- AI 반도체 전문 Venture Capital, Private Equity 부족

2.2. 기업 가치 평가의 어려움

- 지분희석 방어를 위한 기업 가치의 상승 고려 필요 (Founder의 지분율 문제를 고려할 필요, Poison Pill 문제)
- 높은 밸류에이션으로 인한 투자 리스크 증가

3. 전략적 투자의 부재

3.1. M&A 시장의 미성숙

- 국내 M&A 시장의 한계
- 대기업의 국내 AI 반도체 기업 인수 어려움 및 연계된 대기업-스타트업간 M&A에 대한 시선

3.2. IPO를 통한 회수의 한계

- 높은 기업 가치 대비 실적 부족으로 인한 리스크 (검증된 실적 부족, 투자 근거 확보 필요)
- 성공사례 부족 (장기적인 ROI 불확실성)
- IPO 이후 추가 자금 조달의 어려움

4. 공공투자 부문의 한계 (정부 정책으로 생성된 펀드 대상)

4.1. 정부 지원 펀드의 한계

- 정책 자금 지원의 변별력 저하 (선별적 지원 vs 광범위한 지원), AI반도체 산업 특성상 지원금의 규모 문제가 발생하므로, 선별적 지원에 대한 논의가 지속적으로 이뤄짐
- 작은 예산 규모 (AI반도체 투자 규모에 맞지 않음)
- 명확한 전략적 초점 부재
- 정부 지원 이후 민간 투자로의 연결 미흡

다. 인재 부문

1. 고급 설계인력의 공급 부족

1.1. 설계 및 개발(팹리스)에 필요한 석사급 이상의 고급설계 인력 요구

- 기존 팹리스/파운드리 섹터 취약으로 경험있는 인재 부족(국내 시스템반도체 세계시장 점유율 3% 수준, 파운드리는 TSMC 중심)
- 시스템 반도체 설계인력 양성이 미진
- 기존 팹리스/파운드리 기업의 폐쇄성(산학연 협력을 위한 정보공유 부족)

1.2. 인력확보 경쟁 심화

- 대기업 선호
- AI 반도체 설계 관련 인재의 해외 유출 심화
- 시스템 반도체 설계 관련 해외 인재 유치방안 미흡

1.3 고급 설계인력에 대한 수요 확대

- AI반도체 시장의 성장에 따른 인력수요 지속적 확대(학계, 출연연, 기업 등)
- 스타트업 및 기존기업의 시장진입 확대(AI반도체 스타트업, 기존반도체 기업의 AI반도체 진출, 대기업의 신규사업 진입 등)

2. 반도체 설계 교육 훈련 양성시스템 고도화 필요

2.1. 교육 훈련 양성 프로그램의 한계

- 실무에서 요구하는 고급인재 육성 프로그램 부족(체계적이지 못한 단기적 실무교육 중심 프로그램 개편 요구)
- 실무교육 가능 교원 확충 필요(산업계 출신 교원 확보)
- 해외 유학생 유치방안 미흡(유치 및 취업 연계방안 마련)

• 2.2. 인재육성 정책을 통합 관리하는 컨트롤타워 부재

라. 기술 부문

1. 칩 설계 기술 격차

1.1. 아키텍처 설계 기술 격차

- 아키텍처의 효율성(속도, 전력), 확장성(최소 수백 대 연결), 범용성의 동시 달성을 가능성에 의문

1.2. 최신 공정 활용 설계 격차 칩렛 활용 설계 기술 격차

- 초미세공정(FinFET 이후) 활용 설계 경험 및 역량 부족
- 칩렛(+HBM) 활용 설계 경험 및 역량 부족

1.3. 인터커넥트 기술 격차

- 업계 선두(NVIDIA NVLink & NVSwitch) 대비 인터커넥트 기술 격차: 범용 규격 의존, 속도, 구성의 유연성, 신뢰성 등

2. SW 기술 격차

2.1. SW 기술 격차

- 업계 선두(NVIDIA CUDA) 대비 소프트웨어 스택(SDK or API, SW in HW(kernel), driver, etc.) 성능 및 최적화 격차(network 효과 작용)

2.2. SW 관련 고객지원 역량 격차

- 최신 연산(알고리즘) 지원 부족
- 고객지원 역량 부족: 서비스 개발 예시 제시, 개발문서의 양과 질, 커뮤니티 지원(Q&A) 등

3. 실제 칩 양산 및 검증 경험 부족

3.1. PoC(proof of concept) 및 대규모 실증 경험 부족

- PoC, MPW, 양산, 테스트 및 대규모 성능 검증 등의 과정에서 trial & error 증가, 목표성능 달성을 실패, 불량 발생(낮은 수율), 개발 지연 등의 치명적 추가비용 발생
- 각종 극한환경, 장기 신뢰성 검증 경험 부족, 고객 신뢰도 확보 어려움

3.2. 호환성 검증 부족

- (Hardware) 칩 외 다양한 카드 형태, CPU 및 GPU, 서버, 데이터센터, 핸드셋(스마트폰 등) 등 다양한 주변기기와의 호환성 검증 부족
- (Software) 윈도, 리눅스 등 다양한 OS, 프로그래밍 언어(Python, C+, ...) 및 API(PyTorch, Tensorflow, ...) 등에 대한 호환성 검증 부족

4. 기술 조달 가능한 커뮤니티 미비

4.1. 기술 발전 및 피드백을 위한 커뮤니티 미비

- logic chip 관련 국내 대학(학계), 연구소, 설계기업 등 주변 커뮤니티 규모가 작아 기술교류, 인력조달 등에 어려움

마. 시사점

구분	주요 문제점	우선 고려 사항
BM과 생태계	<ul style="list-style-type: none"> • 수요 기반 한계 <ul style="list-style-type: none"> - 국내 시장 규모의 협소성, 대부분의 수요처가 퍼스트 무버 역할에 미온적, 빅테크 기업과의 교류/협력 부족으로 인한 수요 파악 지연, 보완(Complementary) 산업 생태계 미성숙, Value Chain 상 최종 고객 니즈 반영 지연 • 시장 진입 장벽 <ul style="list-style-type: none"> - 서버 시장: 높은 개발비용 대비 제한적 창업팀 역량, Edge 시장: 과다 경쟁 및 중국/대만 대비 열위, B2B 판로확보의 어려움과 IP 전략 미흡, 글로벌 마케팅 및 네트워크 역량 부족 	<ul style="list-style-type: none"> - 대기업 수요자 인센티브 제도 도입(세제혜택 등), 극초가 투자 지원체계 구축(CVC, Angels, 대학 / 출연연 등), 비영리 Incubator /Accelerator 통한 실증지원, 글로벌 협력생태계 구축을 통한 경쟁력 강화
기술	<ul style="list-style-type: none"> • PoC(Proof of Concept) 및 실증 문제 <ul style="list-style-type: none"> - 개발된 AI 가속기의 실질적 수요처 확보 어려움, 대규모 실증을 위한 데이터센터 테스트 환경 부족, 스타트업 제품 적용에 대한 기업들의 위험 회피 성향, 레퍼런스 확보 실패가 사업화의 주요 걸림돌 • 기술적 호환성 문제 <ul style="list-style-type: none"> - 기존 Training AI 칩과의 호환성/법용성 부족, 시스템 레벨에서 SW(PyTorch 등)와 HW 호환성 부족, 주변기기 와의 통합 및 검증 경험 부족, 	<ul style="list-style-type: none"> - PoC 및 대규모 실증 역량 부족으로 인하여 데이터센터 용 NPU 등 AI 가속기(HW) 설계/개발이 이루어져도 실질적으로 적용 가능한 수요처 확보에 어려움이 존재 → 글로벌 수요처 확보를 위한 노력, 해외 빅테크 데이터센터 수요에 부합하는 HW/SW 스펙을 제시하는 동시에 기술적 호환성 문제도 집중할 필요

구분	주요 문제점	우선 고려 사항
인재	<ul style="list-style-type: none"> • 인력 수급 현황 및 문제점 - 국내 반도체 생태계가 메모리 반도체 IDM 2개사 중심으로 형성되어 팝리스 기업의 고급인력 확보가 어려움, AI 반도체 분야의 석박사급 전문인력이 극히 부족하며, 특히 시스템 SW 개발 전문가가 매우 부족, 기존 인력의 고령화 및 전문 교수진 부족으로 인재양성에 어려움 • 핵심 인재 유출 문제 - 우수 인재들이 대기업이나 해외기업 취업을 선호, 팝리스 스타트업 재직자들의 대기업 이직이 지속해서 발생, 직업 안정성과 처우 문제로 인한 인재 유출이 심각 	<ul style="list-style-type: none"> - 대학원 중심의 연구환경 구축으로 중장기적 석박사 인력양성 필요, 글로벌 인재 유치를 위한 해외 Returnee 발굴 및 외국 전문가 영입 지원, 스타트업의 성과보상체계 개선(스톡옵션 등)과 처우 개선을 통한 인재 유치, 협회(학계, 산업계 Bridge 기능) 추진 맞춤형 교육시스템 구축으로 실무형 인재 양성
투자	<ul style="list-style-type: none"> • 대규모 초기 투자 필요성 - AI 반도체 개발은 10nm 이하 선단공정 사용시 IP, DSP, PKG & TEST 등에 400~500억원 규모의 막대한 초기 자금이 필요, 현재 정책펀드는 VC 수준인 10~20억원 규모로, 실제 필요 자금과 큰 격차가 있음. 기술개발과 시장 가치 창출까지 장기간의 지속적 투자가 필수적이나 공공투자만으로는 한계가 있음 • 중후기 투자 환경의 한계 - 해외 VC들은 대규모 후속 투자가 활발하나, 국내는 초기 투자 이후 대규모 추가투자가 어려운 구조, 매출이 없는 상황에서도 PE들이 과감한 투자를 할 수 있는 특수목적 펀드 등의 환경 구축이 필요, 국내 M&A 시장이 미흡하여 IPO가 유일한 Exit 전략이 되는 상황이며, 이는 투자 유치를 더욱 어렵게 만들 	<ul style="list-style-type: none"> - 팝리스 산업에 특화된 투자환경 혁신이 필요하며, 하이리스크-하이리턴 특성을 고려한 전폭적 지원 필요, 글로벌 경쟁력 있는 기업을 선별하여 정부 지원으로 개발자금 일부를 지원하는 방안 검토, 현재 주요 AI 반도체 스타트업의 성공이 향후 전체 시스템 반도체, 팝리스 부문의 벤처 투자 생태계에 큰 영향을 미칠 것으로 예상

4. 결론

가. 시사점

인공지능 반도체 기술 발전 전망

현황

- NVIDIA가 데이터센터 규모의 인공지능 학습 하드웨어 시장 90% 이상 독점, 국내 외 기업들의 NVIDIA GPU 성능 추격 및 소프트웨어 개발 환경 지원 노력, 단시 일에 거대 인공지능 모델 학습 과정에서 NVIDIA GPU를 대체하기 어려움 예상

○ 기회 요인:

- 자체 데이터센터 또는 클라우드 구축: 자체 거대 인공지능 학습 위한 데이터센터 활용 (Google TPU 사례), 인공지능 반도체 자체 개발 통한 비용 절감 및 전략 물자 확보 (Meta MTIA, Microsoft MAIA100 사례)
- 인공지능 프레임워크 발전: PyTorch, TensorFlow 등 프레임워크 고도화로 CUDA 라인 해제 가능성, 후발 주자 SDK 및 소프트웨어 스택과의 호환성 확보 중요
- 새로운 인공지능 모델 등장: 트랜스포머 이후 새로운 모델 등장 및 최적화된 하드웨어/소프트웨어 선점 기회 (얀 르쿤 주장)

○ 시사점:

- 국내 AI 반도체 기술 경쟁력 강화 필요
- 해외 인공지능 모델 의존 탈피, 자체 거대 인공지능 학습 위한 노력 필요
- 대규모 인공지능 데이터센터 구축 및 국내 AI 반도체 활용

□ AI 반도체 생태계 전반의 경쟁력

○ 글로벌 가치 사슬: AI 반도체 산업은 설계, 생산, 패키징/테스트 등 전문 분업화

- 스타트업은 초기 비용/전문성 낮은 팩리스 분야 집중, NVIDIA 독점 속 틈새시장 존재 (엣지, 추론, 학습 등), 산업 특화 AI 반도체, 각 분야별 S 곡선 성장 가능성 고려

○ 팩리스 세부 분야: IP, EDA, 디자인하우스 등 전문 공정으로 구성

- 스타트업, 모든 설계 프로세스 경쟁력 갖추기 어려움, 세부 분야별 전문 기업 육성 필요 (특히 IP, EDA), 국내 생태계 경쟁력 강화, 해외 의존도 감소 필요

○ 시사점:

- 팩리스 스타트업 뿐 아니라 IP, EDA 등 세부 분야 전문 기업 육성 통해 생태계 전반 경쟁력 강화.

VC 투자 측면의 특징 (글로벌화)

○ 국가별 투자 현황:

- 중국: 투자자 수, 투자 기업 수, 투자 횟수 모두 압도적 1위, 미국: 2위, 투자 규모는 중국보다 큼, 한국: 투자자 수 7위, 스타트업 수 8위, 투자 건수 적음, 투자자 대비 스타트업 비율 낮음 (한국 15%, 중국 56%)

○ 국가 간 투자:

- AI 반도체, 국가 간 VC 투자 활발 (기술 독점, 규모의 경제), 한국, 투자 유출 규모 큼 (해외 투자 유치 노력 필요)

○ 시장 진출:

- 글로벌 시장 진출, 수요처 발굴 및 투자 유치 측면 중요, 혁신적 기술 기반 시제품 개발, 시장 수요 창출, 초기/후속 투자 연계, 칩 설계 검증 및 시제품 제작 지원

○ 시사점:

- 국내 투자 활성화, 해외 투자 유치, 글로벌 시장 진출 지원 통해 스타트업 성장 촉진

나. 정책 제언

1) 실증 및 수요처 발굴 문제

- AI 반도체 특징: 특정 목적 최적화 설계, 단독 소비자 효용 낮음, B2B 사업 모델 중심, 데이터 확보 및 수요처 확보 필수
- 수요처 확보: 스타트업, 네트워크/경험 부족으로 대규모 수요처 확보 어려움, 정부 출연 연구기관, 초기 수요처 역할 및 투자 지원
 - 정부 주도의 실증 프로젝트 (미국 NASA, DARPA 사례 참조 필요)
- 실증: 제품 설계, 수요처 발굴, 실증 (PoC), MPW까지 연계될 필요
 - de alio (기존 기업 연계) 스타트업, 수요처 확보 유리 (사피온 사례)

- de novo (신규) 스타트업, 창업자 경험/네트워크 활용 (리밸리온, Tenstorrent 사례)
- 반도체 NRE(Non-Recurring Engineering) 고려한 정부 지원 R&D, 수요 기업 초기 연계 지원, 실증 기회 제공(POC, MPW), de alio/de novo 특성 고려한 맞춤형 지원 필요.

2) AI 반도체 스타트업 Seed 투자 역할

- AI 반도체 설계: IP, 팹리스, EDA, 디자인하우스 등 다양한 과정 필요
- 대규모 초기 투자 필수, 딥테크 SoC 기반 개발 비용 높음
- Seed 투자 규모: 미국/중국, 한국보다 평균 Seed 투자 규모 큼
 - 한국 TIPS 프로그램은 유용하나 지원 규모 제한적, 높은 설계비 고려 시 민간 투자 유치 연계 어려움
- 정책 제언:
 - ASIC 반도체 수요처 지원, 바우처 형식 설계/PoC 지원 (수요 연계 목적)
 - 미국 SBIR 프로그램과 유사한, 단계별 지원 및 투자 매칭 제도 도입
 - 시장 매출/투자 유치 비율 기반 정부 지원금 매칭, 투자 유치 촉진
- Seed 투자 규모 확대, 정부 지원 프로그램 개선, 투자 유치 촉진 통해 스타트업 초기 성장 지원.

Summary

A study on support and fostering measures through the analysis of growth factors of domestic AI semiconductor companies

- Project Leader: Yonggi Kim
- Participants: Younghwan Kim · Dongun Park · Chiho Lee · Aram Lee

A comprehensive review of the prevailing issues in AI semiconductor industry reveals that there exists considerable challenges within the ecosystem, which can be categorised as follows: firstly, in terms of the business model, secondly, in relation to the investment environment, thirdly, with regard to technological capabilities, and finally, in regard to talent acquisition. Primarily, with regard to business models, the absence of PoC (Proof of Concept) and large-scale demonstration capabilities is regarded as the most significant issue. Specifically, the exorbitant costs associated with the development and production of server AI semiconductors for data centres severely hinder the capacity of nascent enterprises to devise competitive products in comparison to global leaders. The B2B market's inherent characteristics pose significant challenges for new entrants, particularly in the context of the edge-type AI semiconductor market, which is characterised by the presence of numerous domestic startups. These challenges stem from the complex relationship with existing customers and the intense competition that characterises this market. Furthermore, Chinese and Taiwanese companies have a comparative advantage over Korean companies with regard to manufacturing infrastructure and cooperation with demand companies.

In terms of the investment environment, there are structural limitations in the large-scale financing that is essential for the development of AI semiconductors.

The unique characteristics of the B2B market further complicate the entry of nascent companies, as they are often constrained by their relationships with existing clientele. Within the competitive landscape of the edge-type AI semiconductor market, characterised by numerous domestic start-ups, achieving a

distinct competitive edge becomes an arduous task due to the intense rivalry present. Furthermore, with regard to manufacturing infrastructure and collaboration with demand companies, Chinese and Taiwanese companies demonstrate a comparatively superior advantage over Korean companies. Specifically, in the context of developing chips utilising a front-end process of 10 nm or less, the financial outlay required for securing IP, DSP (design house), packaging, and testing can reach 40 to 50 billion won. However, the majority of current domestic policy funds are constrained to initial VC investments of 10 to 20 billion won. Concurrently, due to the nascent state of the domestic M&A market, the prevailing exit strategy is oriented towards IPO, which curtails the attainment of founder equity and the financing of substantial development costs. This is further compounded by the conservative investment tendencies of domestic investors, such as the avoidance of investment risk, which significantly hinders the attraction of the substantial investment required in the period preceding sales. The immaturity of the domestic M&A market is evidenced by its bias towards IPO, which limits the securing of founder equity and the procurement of large-scale development costs. This perpetuates a detrimental cycle, as the absence of companies with established sales hinders the facilitation of subsequent investments. From a technical standpoint, the predominant challenges pertain to the absence of software competitiveness and system-level compatibility. In contrast to memory semiconductors, AI semiconductors such as SoC and NPU demand a profound comprehension of programming languages. Furthermore, the necessity for comprehensive design that considers the HW-SW interface, API, and application is paramount, yet the chip SW competitiveness, such as SDK, exhibits significant deficiencies when benchmarked against global leading companies, consequently leading to inefficiency in the testing and performance verification process.

The AI semiconductor industry demands a convergence of talents encompassing both hardware and software domains, yet the domestic talent pool exhibits notable limitations in this regard. Specifically, there is a paucity of talent with both an understanding of AI algorithms and semiconductor design capabilities.

With regard to technological challenges, the primary concerns pertain to the absence of software competitiveness and system-level compatibility. In contrast to memory semiconductors, AI semiconductors, such as SoCs and NPUs, demand a profound comprehension of programming languages. Furthermore, the necessity for comprehensive design that considers the HW-SW interface, API, and application is paramount, yet the chip SW competitiveness, exemplified by the SDK, exhibits significant deficiencies when benchmarked against global leading companies, consequently leading to inefficiencies in the testing and performance verification process. Furthermore, the aggressive recruitment of talent by global companies is making it more difficult for domestic companies to secure talent, and start-ups are struggling to attract and retain talented people due to the limitations of providing competitive compensation packages.

Confronted with these challenges, the growth trajectories of major companies in the AI semiconductor ecosystem are exhibiting a stark contrast to those of existing analogue IP companies and fabless companies. The existing initial, growth, and mature stages of the semiconductor industry have been shown to differ from those of analogue IP companies and fabless companies. The MPW (Multi-Project Wafer), Single-run, and strategic foundry (using legacy semiconductor processes to produce semiconductors for various industries, such as 28nm and above, 40nm, etc.) are also evident in the mature process. However, with the rise of the AI industry and the emergence of generative AI, the need for AI computing power is increasing, making it difficult to follow the existing growth path. The requirement for advanced low-power designs, such as GPUs and NPUs, in the AI industry has led to an increase in the complexity of chip design. Consequently, foundries are compelled to utilise leading-edge processes, such as 10, 7, 3, and 2 nm, in the production of designed chips. In addition, as highlighted in the characteristics of AI semiconductors, the architecture of the chip itself is becoming more sophisticated, necessitating the use of expensive Cadence, Synopsys, Siemens IP, and EDA in the chip design process. When utilising Samsung Foundry, the production of semiconductors can be accomplished through the utilisation of IP from Qualitas Semiconductors and OpenEdge Technology, both domestic companies, and DSP (Design Solution

Partners) companies such as Semifive. It should be noted that, in the case of IP, all IP companies contracted with Samsung Foundry can be used in addition to domestic IP. However, it should be noted that the cost of MPW may be subject to further escalation in the context of small-volume production.

The initial investment cost for AI semiconductors is estimated to be in the range of 50 to 100 billion won. This is primarily due to the substantial costs associated with the licensing of IP and EDA tools, in addition to the need for a significant number of new employees. This is consistent with the preparation of the software stack, which is a feature of AI semiconductors. The complexity of the hardware design necessitates a multifaceted approach, incorporating not only the design of the chip itself but also the software that will be utilised. In the market for data centres, which is a substantial learning market, NVIDIA has a 95% share of the global market, and in the market for server-based AI systems-on-chips that participate in the area of large-scale inference, there are two aspects of challenges: the advancement of the design sector to improve the performance of semiconductors and the provision of software for the use of chips by customers who use them. In terms of the growth trajectory, the AI semiconductor sector requires a sufficient workforce of hardware and software engineers at the start-up stage, including professionals who can design in this field.

In the market for data centres, which is a substantial learning market, NVIDIA has a 95% share of the global market. In the market for AI systems on chips for servers that participate in the field of large-scale inference, there are two challenges: the design of the semiconductor to improve performance and the provision of software for the chip utilisation of the customer. This phenomenon is further accentuated when examining the backgrounds of prominent Korean startup founders, who often accumulate experience in prominent corporations such as Samsung, Apple, AMD, NVIDIA, and Microsoft, subsequently challenging themselves to establish their own ventures. Prior to embarking on their entrepreneurial endeavours, these individuals engage in the domain of AI semiconductor design, contributing to the development of the initial workforce and the identification of a distinct niche market. However, it is imperative to recognise the significance of investment in IP and EDA, which is instrumental

in the production of prototypes.

To address this challenge, the following priorities must be considered. In terms of the establishment of an industrial ecosystem, it is necessary to establish a one-stop support system that spans from the discovery of cutting-edge technologies to large-scale demonstrations, and to introduce institutional incentives to promote cooperation between large companies and startups. Of particular note is the necessity for government support in the establishment of strategic partnerships with globally prominent technology companies. The provision of opportunities, such as demonstrations, for establishing initial references or creating a platform for cooperation in the emerging technology sector on major national platforms, including the Science and Technology Joint Committee, is recommended.

In terms of enhancing technological competitiveness, it is imperative to institute a systematic programme for the training and security of software development personnel, and to establish a public infrastructure for system-level compatibility verification. There is a need for continuous monitoring and response to global standards and the latest technology trends in a form that can be disclosed at the government level. Furthermore, companies must establish a differentiated positioning strategy in the edge AI semiconductor market and target niche markets by developing solutions specialised for each application area. In this regard, it is imperative for companies to explore avenues for overcoming barriers to market entry through collaborative development and verification initiatives with companies that are in demand. Even if a company has a technological advantage, its competitiveness in the market is maximising sales through collaboration with demand companies, and it can promote strategies such as product diversification through market share expansion. In terms of talent cultivation and retention, there is an imperative for the augmentation of talent development programmes that emphasise convergence, facilitated through collaborative endeavours between industry, academia and research entities. These programmes should be designed to produce talent that can be seamlessly integrated into the industry. It is imperative to prioritise the development of an education programme that focuses on the training of professionals with practical

experience in domains such as chip manufacturing, design, and design-manufacturing. To this end, it is essential to consider a practical-oriented talent training plan through the enhancement of industry-academia cooperation programmes between companies and universities. Furthermore, institutional support is required to attract outstanding talent from overseas. The government should consider introducing an incentive system for recruiting talent at the government level to secure talent for startups, which are the driving force of innovation in this field. In terms of innovation in the investment ecosystem, the creation of a large-scale project fund (worth 40-50 billion won) specialising in AI semiconductors is imperative, and the introduction of a risk-sharing mechanism is recommended to encourage active participation by private equity firms. The establishment of an institutional foundation to foster patient capital, capable of making long-term investments, is imperative. In terms of policy support, it is necessary to expand public investment for the construction of empirical infrastructure after the design of chips, and to consider improving the public procurement system to create demand. In terms of innovation, the introduction of an incentive system for government-level talent recruitment is recommended in order to secure talent for startups, who are the main innovators in this field. In terms of innovation in the investment ecosystem, the creation of a large-scale project fund (worth 40-50 billion won) specialising in AI semiconductors is imperative, as is the introduction of a risk-sharing mechanism to encourage active participation by private equity firms. Furthermore, in order to facilitate marketing and network expansion in order to penetrate the global market, it is essential to adopt a strategy that will augment the networks that companies have already established.

The combination of technology, Investment, ecosystem, and talent is imperative for the successful development of the AI semiconductor industry, and the establishment of a collaborative ecosystem among the government, companies, investors, and educational institutions is a pressing task. A long-term strategic approach is imperative for nurturing and securing talent, necessitating systematic efforts to establish a talent pool that boasts global competitiveness.

| 제1장 | 서론

제1절 관련 현안 및 연구 배경

1. 반도체 산업과 AI 반도체 산업의 개요

반도체 산업은 크게 메모리 반도체와 비메모리 반도체로 구분하며, 메모리 반도체의 경우 휘발성 메모리를 저장하는 반도체로 한국 기업들이 세계시장에서 높은 시장 점유율을 보이고 있다. 반면에 비메모리 반도체는 특정 연산에 특화된 반도체로 ‘시스템 반도체’라고도 불린다. 전체 반도체 시장을 살펴보면, 메모리 반도체가 약 30% 정도이며 비메모리 반도체가 약 70% 정도를 차지한다. 최근 들어 다양한 형태의 정보를 수집하여 처리하는 AI 산업의 성장과 더불어, AI의 학습과 추론에 활용하는 첨단 반도체 (AI 반도체)의 수요가 급증함에 따라서 비메모리 반도체 시장은 더욱 폭발적으로 성장하고 있는 상황이다(이선재, 2023).

AI 반도체는 Generative AI(생성형 AI) 성장에 따른 AI 파운데이션 모델의 학습과 추론을 위하여 이용된다. 기존의 신경망 알고리즘에서 트랜스포머(Transformer)로의 변화 및 Flash Attention¹⁾의 활용 등 생성형 AI의 발전과 AI반도체 산업의 발전은 깊은 관련이 있다. 인공지능 반도체 산업과 시스템 반도체 산업은 기술적 특성과 시장 요구에 따라 뚜렷한 차이를 보인다. 이 두 산업은 각각의 응용 분야와 기술적 접근 방식에서 다르며, 이는 기업의 전략, 연구 개발(R&D) 투자, 그리고 시장 진입 방식 및 사업 모델에도 영향을 미친다.

첫째, 인공지능 반도체 산업은 데이터의 학습과 추론을 위한 AI 알고리즘을 실행하기 위한 반도체를 설계하고 제조하는 데 중점을 두고 있다. 예를 들어, AI 반도체는 신경망 처리 장치(NPU)와 같은 특수화된 아키텍처를 포함하여, 머신러닝 및 딥러닝 알고리즘을 효율적으로 실행할 수 있도록 설계된다(Kumar, 2024). 이러한 반도체는

1) 파이토치 한국 사용자 모임(24.07.11), “FlashAttention-3: 비동기 및 저정밀도에서의 빠르고 정확한 어텐션 제공”, <https://pytorch.kr/blog/2024/flashattention-3/> (검색일 24.10.15)

대량의 데이터와 연산을 필요로 하며, 이는 고속 처리와 낮은 전력 소비를 요구한다 (Prytkova & Vannuccini, 2022). AI 반도체는 AI의 발전과 함께 빠르게 성장하고 있으며, 다양한 AI-powered 첨단 산업에서 ASIC(주문형 반도체, Application Specific Integrated Circuit) 형태로 수요가 증가하고 있다. 반면, 시스템 반도체 산업은 보다 넓은 범위의 산업을 위한 전자 기기와 시스템을 위한 반도체를 대상으로 하고 있다. 시스템 반도체는 통신, 컴퓨팅, 소비자 전자 제품 등 다양한 분야에서 사용되며, 일반적으로 시스템 온 칩(SoC) 형태로 제공된다. 이러한 반도체는 다양한 기능을 통합하여, 여러 전자기기에서의 성능을 극대화하는 데 중점을 두고 있다(Saha, 2018). 시스템 반도체는 다양한 프로세서, 메모리, 그리고 주변 장치를 포함하여, 복잡한 시스템을 구성하는 데 필수적인 요소이다. 이들은 일반적으로 대량 생산을 통해 비용 효율성을 높이며, 다양한 응용 분야에 맞춰 설계되고 있다(Ren et al., 2023).

둘째, 두 산업의 기술적 접근 방식에서도 차이가 존재한다. AI반도체는 진화하는 AI산업에 활용되는 주요한 매개체로서, 설계의 난이도 및 회로의 집적도가 높다. 즉, 이를 위하여 3nm 이하의 선단 공정이 주로 사용되며, HBM(High Band Width Memory)과 같은 고도화되고 집적화된 DRAM이 설계된 반도체에 필수적으로 사용되어야 한다. 이러한 선단 공정 및 기술들은 고도화된 설계의 AI반도체 생산에 필수적이며, 이를 통해 AI 연계 산업의 응용 범위를 넓히고, 새로운 시장 기회를 창출하는 데 중요한 역할을 한다. 이에 비하여, 시스템 반도체 산업은 보다 전통적인 제조 공정과 기술을 기반으로 하며(Legacy 공정, 28nm, 14nm), 대량 생산과 비용 절감에 중점을 두고 있다. 시스템 반도체는 일반적으로 표준화된 프로세스를 통해 생산되며, 이는 생산 효율성을 높이고, 시장의 다양한 요구에 신속하게 대응할 수 있도록 한다 (Bahinipati & Deshmukh, 2012). 이러한 접근 방식은 시스템 반도체가 다양한 응용 분야에서 경쟁력을 유지하는 데 중요한 요소로 작용한다.

셋째, 시장의 요구와 트렌드에 대한 반응에서도 두 산업간 차이가 존재한다. AI 반도체 산업은 AI가 적용되는 산업의 성장과 맞물려 폭발적으로 성장하고 있으며, 이는 AI의 도입이라는 측면에서 향후에도 지속될 것으로 전망된다. AI의 발전은 첨단 반도체에 대한 수요를 증가시키는 동시에, 기업들이 AI 반도체 개발에 더 많은 자원을 투

자하도록 유도하고 있다(Prytkova & Vannuccini, 2022). 이러한 변화는 인공지능 반도체 산업의 성장을 가속화하고 있으며, 기업들은 AI 기술을 활용하여 동 분야의 경쟁력을 강화하고 있다. 반면, 시스템 반도체 산업은 안정적인 수요와 시장의 성숙에 따라 성장하고 있다. 이 산업은 전통적인 전자 기기와 시스템의 수요에 기반하여 안정적인 성장을 이루고 있으며, 이는 기업들이 지속적으로 효율성을 높이고, 비용을 절감하는 데 중점을 두도록 하고 있다(Hwang & Lu, 2013). 즉, 성숙한(Mature) 하이테크 제조 산업과 비슷한 형태의 Business model을 사용하고 있다. 시스템 반도체는 다양한 산업에서 활용되는 전력 반도체, 디스플레이 구동칩 등을 포함하고 있으며, 이는 동 분야 기업들의 장기적인 시장 전략이 AI반도체 기업과는 다르다는 것을 보여준다.

정리하면 인공지능 반도체 산업과 시스템 반도체 산업은 기술적 특성, 접근 방식, 그리고 시장 요구에 따라 뚜렷한 차이를 보이고 있다. 인공지능 반도체 산업은 AI 기술의 발전에 따라 빠르게 변화하고 있으며, 시스템 반도체 산업은 안정적인 수요와 효율성을 중시하는 전통적인 접근 방식을 유지하고 있다. 이러한 차이는 각 산업의 성장전략과 미래 방향성을 결정짓는 중요한 요소로 작용하며, 이에 따른 정부의 지원 정책도 서로 달라져야 함을 시사한다.

2. AI 산업의 발전과 AI반도체 산업의 부흥

AI 산업의 발전은 Google의 Transformer 발표 및 Open AI의 Chat-GPT 3, 3.5, 4와 같은 Generative AI (생성형 인공지능)의 활용의 증가에 따라 발전하고 있다. 특히 NVIDIA의 GPU와 같이 병렬 연산을 효율적으로 처리할 수 있는 반도체의 활용은 AI 파운데이션 모델을 통한 다양한 서비스의 제공의 측면에서 주요한 역할을 한다. 즉, 첨단 반도체 산업은 AI 산업이 주도하는 혁신적 변화를 체감하고 있으며, 글로벌 AI 산업을 위한 첨단 반도체 칩 시장은 더욱 성장할 것으로 예상된다. 이는 AI 서비스의 일상화로 AI 반도체가 핵심 수요처인 데이터베이스에서 스마트폰, 커넥티드카, IoT 등 엣지 디바이스로 확대됨에 따라 첨단 시스템 반도체 시장 뿐만 아니라 메모리 시장 성장도 견인할 전망이다(임현 외 2024). 특히, 생성형 AI로 촉발된 수요 증가는 AI 반도체 시장 성장의 원동력이 될 것이다. 글로벌 AI 반도체 시장은 2023년

537억 달러 규모에서 2028년 1,975억 달러 규모로 향후 5년간 연평균 29.7% 성장이 전망²⁾되고 있다.

앞선 AI 산업을 보면, 연산 목적에 따라 훈련(training)과 추론(inference) 시장으로 구분할 수 있으며 데이터 처리 위치에 따라 클라우드형(Cloud)과 엣지형(Edge)로 구분할 수 있다. 훈련은 AI를 학습시키는 과정이며 주로 고성능 GPU와 HBM을 사용하는 클라우드 데이터 센터에서 수행하고 있으며, 추론은 학습된 AI 모델을 실제로 적용하는 과정으로 추론 전용 AI 반도체(가속기)가 활용된다. 클라우드는 데이터 센터에서 대규모 연산을 처리하는 방식이고 (데이터 센터의 위치에 따라 클라우드와 온프레미스로 구분), 엣지는 엣지 플랫폼을 통해 기기에서 실시간으로 데이터를 처리하는 방식으로 스마트폰, IoT 기기, 자율주행 차량 등 엣지 환경에서 추론을 작업 수행하는 방식이다.

폭발적으로 증가하는 AI-powered industry의 수요를 고려할 때, AI 반도체의 역할은 기존보다 증가할 것으로 예상된다. 하지만, 데이터 센터의 학습용 목적의 AI반도체 시장에서의 NVIDIA의 절대적 영향력 아래 Magnificent 7으로 설명되는 미국의 구글 (TPU), 아마존 (イン페런시아), 테슬라 (도조) 등 기업은 수직적 통합을 통해 자사의 반도체를 설계하여 사용하는 상황이다. Application Processor(스마트폰 용 반도체) 시장에서 절대적 강자인 Qualcomm의 경우에는 ARM과의 소송 분쟁으로 인하여, 향후에 관련된 칩들의 판매에 대한 로열티가 배로 부과되거나 IP공급이 끊길 가능성도 높후하다. 이는 AI반도체의 설계에 연계된 Chipless, Fabless기업의 중요성이 글로벌 하게 부각되고 있으며, 설계의 각 공정에서의 글로벌 의존도 등 고려해야하는 조건이 많아졌음을 의미한다.

2) Gartner(24.07.04.), “Forecast: AI Semiconductors, Worldwide, 2022–2028, 2Q24 Update”, <https://www.gartner.com/en/documents/5558995> (검색일 24.11.13.)

제2절 연구 대상 및 주요 연구 내용

본 연구의 주요 연구대상은 인공지능 서비스 구현에 필요한 대규모의 데이터 및 알고리즘을 효율적으로 처리하는데 특화된 반도체인 AI반도체를 설계하는 팹리스 스타트업 및 기업을 대상으로 하고 있다. 여기서 AI반도체 설계는 NVIDIA가 글로벌 시장을 주도하고 있는 GPU(Graphics Processing Unit)에 대한 부분이 아닌, 인공 신경망 연산에 최적화된 NPU(Neural Processing Unit, 가속기 등 총칭) 설계를 하는 기업을 의미한다. NPU는 기존의 CPU, GPU보다 저전력, 고효율로 대규모 데이터 처리에 특화된 칩을 의미한다. 하지만, 앞서 언급한바와 같이, 글로벌 NPU 및 AI 반도체 시장에서의 한국의 위상은 높지 않다. 삼성전자와 하이닉스의 HBM과 같은 GPU 및 NPU의 메모리 부문의 글로벌 경쟁력은 최고 수준이지만, 비메모리 영역의 딥테크 SOC(System on Chip)부문에서는 NVIDIA와의 비교는 아직 이른 수준이다. 물론 퓨리오사 AI, 리밸리온, 사피온(리밸리온과 24.8월 합병 발표), 모빌린트, 딥엑스 등 차별화된 기술로 NPU시장에 뛰어드는 스타트업과 신생기업은 있으나, 아직 시장에서의 매출 또는 글로벌 표준 제정에서의 영향력 등 부문에서는 미비한 실정이다.

정부에서는 2019년부터 다양한 정책, 전략, 이니셔티브를 활용하여 AI 반도체 및 시스템 반도체 부문의 경쟁력을 도모하고 있다. 과학기술정보통신부, 산업통상자원부, 중소벤처기업부 등 다양한 전략과 이니셔티브를 발표하여 AI반도체 산업 발전과 기업의 경쟁력 향상을 위한 지원을 하고 있으나, 다수의 이니셔티브와 전략의 정책대안은 다양한 정책제언을 동시에 제공하고 있는 한계점이 존재한다. 하지만, AI반도체 부문의 혁신을 선도하는 기업의 성장과 생존을 위한 정책 제언은 보다 더 전략적으로 이뤄져야 할 필요가 있으며, 성장과 생존의 과정에서 상호 영향을 주고받는 AI 반도체 생태계 부문으로 확장이 될 필요가 있다.

본 연구에서는 국내 AI 반도체 생태계의 현황과 당면한 문제점을 사업 모델 및 생태계, 기술, 인재, 투자 부문의 프레임워크를 통하여 분석한 결과를 기반으로, AI 반도체 기업의 성장과 생존에 주요하게 작용하는 요인을 도출할 예정이다. 이를 기반으로 향후 AI반도체 부문의 정책에 의미있는 시사점을 제공하고자 한다. 아직은 미성숙한

6 국내 AI반도체 기업의 성장 요인분석을 통한 지원·육성 방안 연구

국내 AI반도체 생태계 특성으로 인하여, 주요 기업의 대면·서면 인터뷰 및 사례 분석을 기반으로 사업모델, 기술, 투자, 인재의 프레임워크를 반영하여 인공지능 반도체 생태계의 문제점을 마인드맵을 통하여 분석하였다. 앞선 프레임워크의 요소별로 기업의 성장과 생존에 미치는 주요 요인을 도출하였다.

| 제2장 | AI반도체 생태계 현황 및 문제점

제1절 AI 반도체 생태계 현황

1. 기술

가. 기술 측면에서 본 AI 반도체 혁신 생태계

인공지능(Artificial Intelligence, 이하 AI)이라는 용어는 학계에서는 이미 1956년부터 사용되었을 정도로 오래된 개념이나³⁾, 일반인들에게도 널리 받아들여지게 된 것은 최근 기계학습(machine learning)의 급속한 발전에 따라 생성형 AI 등 현실에서 활용할 수 있는 사례가 증가한 후이다. Google Cloud는 인공지능을 다음과 같이 정의하고 있다.

“인공지능은 일반적으로 인간의 지능이 필요하거나 인간이 분석할 수 있는 것보다 규모가 큰 데이터를 포함하는 방식으로 추론, 학습 및 행동할 수 있는 컴퓨터 및 기계를 구축하는 것과 관련된 과학 분야입니다. AI는 컴퓨터 공학, 데이터 분석 및 통계, 하드웨어 및 소프트웨어 엔지니어링, 언어학, 신경 과학은 물론 철학과 심리학을 포함하여 여러 학문을 포괄하는 광범위한 분야입니다. 비즈니스의 운영 수준에서 AI는 주로 머신러닝과 딥 러닝을 기반으로 하는 기술 모음으로, 데이터 분석, 예상 및 예측, 객체 분류, 자연어 처리, 추천, 지능형 데이터 가져오기 등을 수행할 수 있습니다.”⁴⁾

이를 참고하여, 본 연구에서 AI는 컴퓨터나 기계가 수행하는 인간의 지적 활동, 보다 구체적으로는 컴퓨터와 기계가 인간의 뇌로는 감당할 수 없는 대규모의 데이터를 가지고 기계 학습과 딥 러닝 등의 기법을 통해 인간의 추론, 학습 및 행동 등 인간의 지적 활동 행태를 학습하여 그것을 모방하는 것으로 정의하고자 한다. 그리고 위의

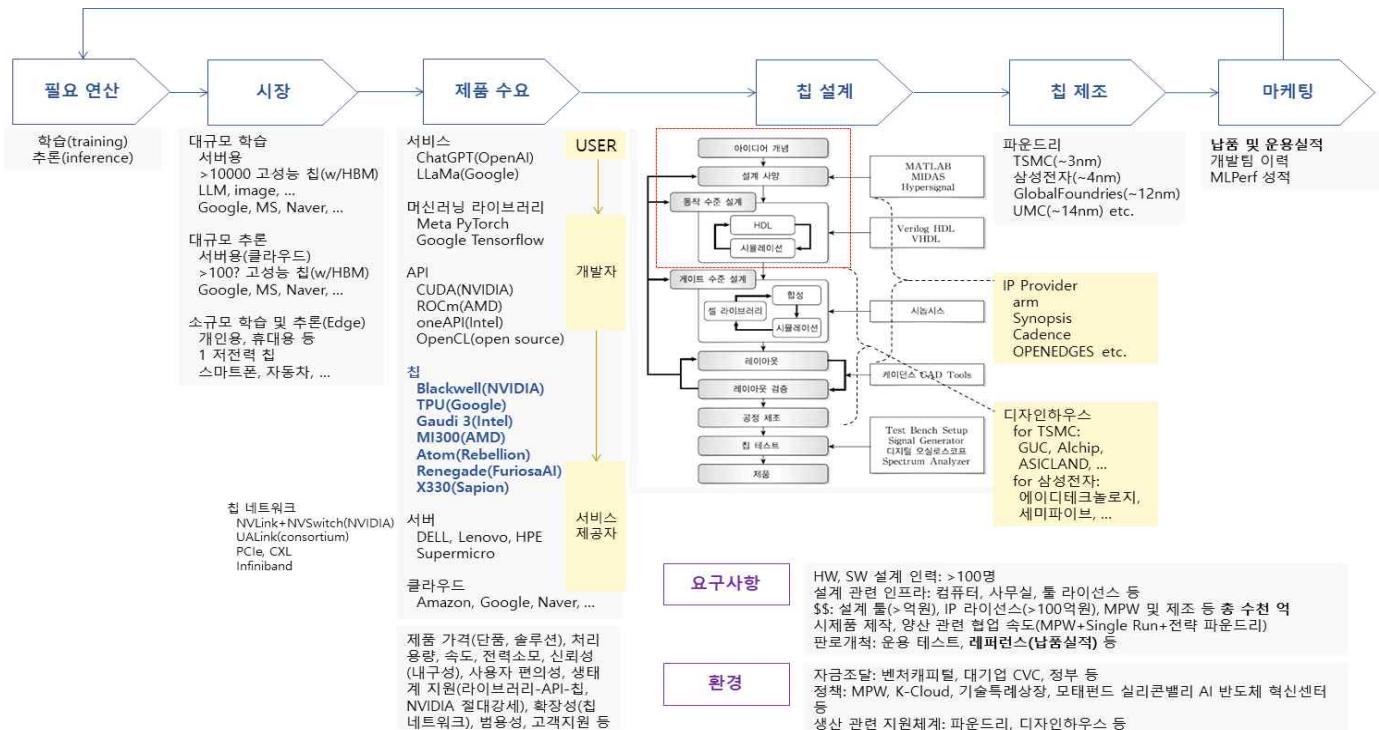
3) 출처: IBM, <https://www.ibm.com/kr-ko/topics/artificial-intelligence> (검색일: 2024.6.2.)

4) 출처: Google Cloud, <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=ko> (검색일: 2024.6.2.)

정의는 그 길이가 짧지만, AI를 구현하는 데 필요한 요소들을 잘 정리하고 있다. 인간의 지적 활동을 흉내 내기 위해서는 데이터가 필요하며, ‘컴퓨터 및 기계’가 대변하듯 해당 데이터를 처리하기 위한 입력 및 연산, 출력장치가 필요하다. 여기서 ‘AI 반도체’는 인간의 지적 활동을 모방하고자 하는 연산장치를 의미하며, 그것은 컴퓨터 내 반도체 공정을 통해 제작되는 칩의 형태로 구현된다.

따라서 AI 반도체 기업은 인간의 지적 활동을 모방하기 위한 연산에 사용되는 칩인 AI 반도체, NPU(Neural Processing Unit)를 만드는 기업이 된다. 그렇다면 구체적으로 AI 반도체는 어떤 과정을 통해 만들어지며, 어떤 요소가 그에 관여하고 있는가? 여기서는 AI 반도체를 만들고자 하는 기업의 입장에서 논리적인 순서에 따라 우선 아래 그림과 같이 정리해 보고자 한다. 그리고 단계별 세부사항을 뒤에서 차례대로 다룰 것이다.

[그림 2-1] AI 반도체 관련 생태계



자료: 전문가 인터뷰 등을 참조하여 연구팀 작성.

1) 필요 연산

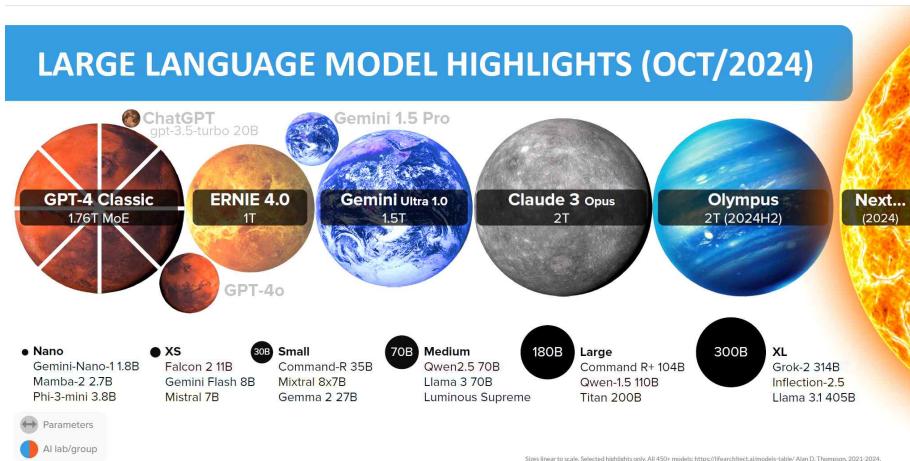
가) 학습(Training)

인공지능을 구현하기 위해서는 우선 인간이 지금까지 생성한 여러 데이터를 인간이 데이터를 처리하는 방식대로 처리함으로써 모델, 즉 경험 내지는 사고방식을 만들어 낼 필요가 있다. 엄밀한 설명은 아니지만 예를 들자면, 어린아이가 어릴 때부터 무수한 말을 듣고 말하면서 언어를 익히듯이 컴퓨터에게 대량의 데이터를 주고 인간의 뇌를 모방한 연산장치를 가지고 인간의 언어나 그림 등을 학습하게 하여 모델을 생성하게 하는 것이다.

학습 연산의 특징은 입력하는 데이터의 양이 방대하고 모델의 크기(매개변수 수)가 매우 크다는 것이다. 컴퓨터에게 입력하는 데이터의 양이 인간 한 명의 지식과 경험을 아득히 웃돌 정도로 많다면, 인간 한 명이 도저히 감당할 수 없는 다양한 분야의 지식들에 대해 알고 질문에 대해 답을 할 수 있는 쓸모 있는 인공지능이 생성될 수 있을 것이기 때문이다. 현재 인공지능 개발사들은 막대한 양의 데이터를 학습에 사용하고 있는데, 2020년에 나온 OpenAI의 GPT-3는 753GB의 데이터를 학습에 투입하였고 모델을 구성하는 매개변수(parameter)가 1,750억 개에 달하는 것으로 알려져 있으며⁵⁾, [그림 2-2]와 같이 최근에는 더 큰 모델도 많이 있다(물론 LLM과 같은 초대형 모델들만 있는 것은 아니며, 목적에 따라서는 작은 모델들도 많이 있다).

5) Thompson(2023.5.), 「Overview of GPT-3 (May/2020)」, <https://lifearchitect.ai/chatgpt/> (검색일: 2024.6.2.)

[그림 2-2] Large Language Model(LLM)들의 모델 매개변수 수



자료: Thompson(2024.10.), 「Language model sizes」, <https://lifearchitect.ai/models/> (검색일: 2024.11.6.)

나) 추론(Inference)

일단 학습을 통해 모델이 만들어지고 나면, 이를테면 영문 문장을 주고 한국어로 번역해 보라고 하거나 동물 이미지를 주고 무슨 동물인지 알려달라는 식으로(입력) 인공지능에게 정보처리를 요구해서 원하는 답(출력)을 얻어낼 수 있다. 이 때는 학습했던 모든 데이터를 다시 읽어 들여 연산을 할 필요 없이 이미 만들어진 모델을 사용해 답을 하므로 연산의 복잡도는 학습에 비해 크게 낮아진다.

2) 시장

가) 대규모 학습

LLM과 같은 대규모 모델은 학습에 막대한 자원이 필요하다. GPT-3의 경우 약 285,000대의 CPU와 약 10,000대의 NVIDIA GPU가 사용된 것으로 알려져 있다.⁶⁾ 이러한 연산은 서버나 데이터센터에서 며칠에서 수개월에 걸쳐 실행되므로, 적시 서비스 제공을 위해서는 속도와 처리용량이 매우 중요하다. 따라서 고성능 칩(~3nm 초

6) <https://lifearchitect.ai/chatgpt/> (검색일: 2024.6.2.)

미세공정)과 대용량 메모리(특히 광대역폭의 HBM)을 다수 연결해서 사용하는 경우가 대부분이다. 연산장치의 가격이나 기반시설 비용, 전기요금 등 금전적 비용은 상대적으로 중요도가 낮았으나, 최근 모델의 거대화 및 NVIDIA NPU의 가격 상승 및 품귀 현상에 따라 비용이 치솟고 있어 Google, Microsoft, Amazon, Meta 등 자체 NPU 개발을 시도하는 서비스 기업이 늘어나고 있다.

나) 대규모 추론

LLM과 같은 대규모 모델은 모델 사이즈 자체가 크며, Google, Microsoft 등의 서비스 기업은 고객의 요청(입력)에 즉각적으로 답을 내놓아야 하므로 추론에도 상당한 자원이 필요하다. 다수의 고객의 요청에 응답하기 위해서는 TCO(Total Cost of Ownership) 역시 중요하다. NVIDIA의 고성능 혹은 보급형 NPU들이 서버 및 클라우드 환경에서 사용되고 있으며, 국내 기업들인 FuriosaAI, Rebellion, Sapeon 등 의 제품들도 이 시장의 진출을 목표로 하고 있다.

다) 소규모 학습 및 추론

간단한 이미지 인식 등 특정 목적에 부합되는 소규모 모델을 로봇, CCTV, 스마트폰 등 단말기 수준에서 처리하는 경우이다. 성능보다는 낮은 소비전력과 가격이 중요해 미세공정에 대한 요구가 상대적으로 적다. 대기업에서 중소기업에 이르기까지 다양한 규모의 기업들이 저마다 다른 종류의 연산의 수요자가 된다. 따라서 범용 제품 외에도 특정 작업에 특화된 다양한 제품들이 요구된다. NVIDIA 등 대기업 외에도 DeepX 등 국내 기업들이 진입을 시도하고 있다.

3) 제품 수요

학습 및 추론 연산을 위해 필요한 핵심 요소는 NPU 칩이라고 할 수 있다. NPU 칩은 각 시장의 요구에 따라 제품 가격(단품, 솔루션), 처리용량, 속도, 신뢰성(내구성) 등의 특성을 요구받는다.

그러나 NPU 칩은 그 자체 성능으로만 성립되는 제품이 아니라, 라이브러리-API-칩-서버 등으로 이어지는 주변 생태계 지원이 필수적이라는 점에 유의해야 한다. ChatGPT, LLaMa를 비롯한 각종 서비스를 구현하기 위해서는 서비스 기업의 개발자들이 Python 등의 프로그래밍 언어를 가지고 PyTorch, TensorFlow 등의 머신러닝 라이브러리를 활용해 필요한 연산을 프로그래밍해야 한다. 그리고 이들 머신러닝 라이브러리는 API라는 소프트웨어를 통해 하드웨어로서의 GPU 혹은 NPU를 구동하게 되는데, 2006년경 머신러닝에 GPU를 처음 활용할 때부터 지원에 힘써 온 NVIDIA의 CUDA(compute unified device architecture)와 타사 API와의 격차가 크다.

한편 서버나 데이터센터에서 NPU 칩이 사용되기 위해서는 카드 단독 및 다수 카드의 결합 솔루션이 중요해지게 되며 DELL, Supermicro 등 서버 시장의 기업들에게 납품할 수 있어야 하는데 여기에도 여러 CPU나 OS와의 호환성, 확장성 등 여러 조건이 필요하다. 특히 대규모 학습이나 추론 시장의 경우 많은 수의 NPU 칩을 결합해서 사용해야 하는데, 이때 NPU 칩 간 고속광대역 데이터 전송기술이 중요하게 작용한다. 여기서도 NVIDIA의 NVLink와 NVSwitch가 타사 기술들에 비해 절대적인 우위를 갖고 있어 강력한 시장지배력을 형성하고 있다.

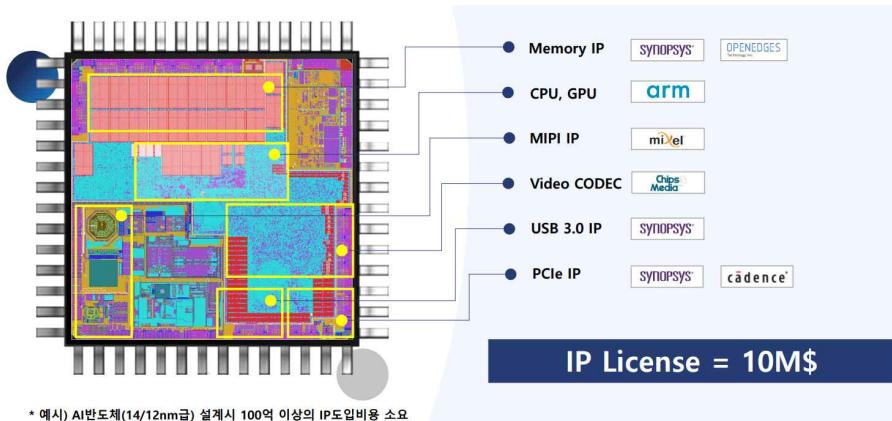
이렇듯 주변 생태계 지원이 매우 중요하게 작용한 결과, 칩의 연산 성능만 놓고 보면 차이가 크지 않은 AMD나 Intel의 제품들이 시장에서 NVIDIA와 거의 경쟁이 되지 않고 있다. 따라서 우리나라 기업들도 적어도 진출하고자 하는 시장 내에서는 충분한 소프트웨어 지원이 뒷받침되어야만 NVIDIA와 경쟁할 수 있을 것임을 짐작할 수 있다.

4) 칩 설계

칩 설계는 위 [그림 2-1]과 같은 과정으로 이루어진다. 최근의 AI 반도체 회사들(소규모 팹리스)의 경우, 핵심 연산 기능 외에 칩에 필요한 메모리, USB, 비디오 코덱 등 주변 기능들은 Synopsys, ARM 등 IP(지적재산권) 제공 회사들로부터 라이선스 비용을 주고 설계를 사 움으로써 설계 속도를 높이는 경우가 많다. 또한 게이트 수준, 레이아웃 등 실제 제작공정에 사용될 도면을 제작하는 후반 설계 부분은 파운드리별 공정에 맞춰 최적화 및 오류 수정을 돋는 디자인하우스와 협업해서 진행한다. 설계에

있어서도 이러한 주변 생태계를 잘 활용하는 것이 중요하다. TSMC 공정의 경우 GUC, Alchip, ASICLAND 등의 업체가 유명하며, 삼성전자 공정의 경우 에이디테크 놀로지, 세미파이브 등의 업체가 있으나 TSMC 공정에 비해 디자인하우스의 역량이 상대적으로 부족한 것으로 알려져 있다.

[그림 2-3] IP Provider를 활용한 설계 간소화와 그 비용

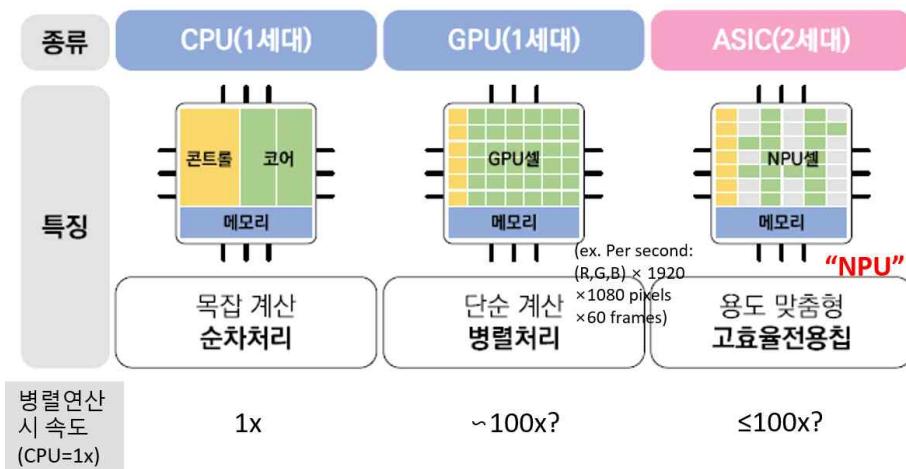


자료: 유병두(2024.3.25.).

따라서 AI 반도체 회사의 핵심 역량은 필요한 연산과 시장, 제품 수요에 맞춰 아이디어 및 목표 사양을 정하고 동작 수준 설계(아키텍처 설계)를 수행하는 쪽에 집중된다. NPU는 기본적으로 복잡한 계산을 순차처리 하는데 적합한 CPU와는 달리 주로 머신러닝의 학습, 추론에 필요한 대량의 행렬 및 벡터에 대한 선형대수 연산을 처리하는 데 최적화되어 있다. 한편 NPU는 100만 개 이상의 다수의 화소에 대한 동시 병렬 처리를 위해 개발된 GPU와 비교하면, 병렬화에 의해 특정 단계의 연산 결과가 이후 연산에 참조되는 경우가 많아 연산유닛 간 데이터 이동이 보다 빈번하게 일어나는 등 머신러닝 고유의 학습 및 추론 연산 특성에 맞게 최적화되어야 한다. 그리고 필요한 연산의 종류, 데이터의 크기 등이 늘어날수록 병렬연산의 효율적 수행이 어려워진다. 따라서 NPU의 경우 연산을 적절한 크기로 쪼개고, 각 unit에 적시에 할당하여, 결과를 효율적으로 메모리에 저장하고 합치고 불러오는 등 연산을 얼마나 효율적으로 병

결합할 수 있는지, 그리고 데이터 크기에 비해 메모리 용량과 대역폭이 한정된 상황에서 CPU, NPU 내 다른 블록, 다른 NPU 칩 등의 메모리 간 불필요한 데이터 이동을 줄이고 얼마나 효율적으로 메모리를 활용할 수 있는지가 기본적으로 중요한 사항이 된다. 현재 머신러닝 연산의 bottleneck은 메모리 용량 및 대역폭에 있는 것으로 알려져 있다.

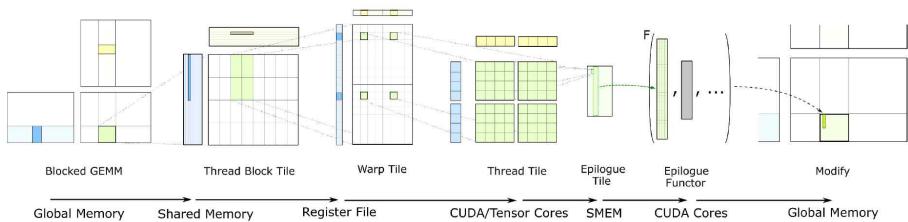
[그림 2-4] CPU, GPU, NPU의 비교



자료: 이선재(2023)를 연구진 가공.

이와 관련해서 AI 반도체 회사의 중요한 역량을 세 가지로 생각할 수 있다. 첫째, 데이터 타입의 최적화와 적절한 소형화이다. 하나의 숫자를 32bit를 사용해서 표현하던 것을 필요한 경우 16bit, 8bit, 4bit까지 줄이면 연산 속도를 2배, 4배, 8배 빠르게 할 수 있는데, 여러 연산 단계 중에는 사용하는 숫자의 정확도(bit수)를 낮추어도 모델의 정확성에는 거의 차이가 없는 경우가 있으므로 해당 단계에 대해서는 유연하게 데이터 타입을 변경해 최적화하면 높은 연산 속도를 얻을 수 있다. 그리고 다양한 데이터 타입과 연산에 대응해야 제품이 범용적으로 활용될 수 있을 것이다. 이러한 효율성과 범용성은 대체로 trade-off 관계에 있으며, 이를 극복하기 위해서는 정교한 아키텍처 구현이 필요하다.

[그림 2-5] 머신러닝 연산을 위한 데이터의 분배와 이동



자료: 전원(2024.5.28.).

둘째, 기존 머신러닝 라이브러리 등 AI 개발툴과의 높은 호환성이다. 기존 AI 개발 툴과의 호환성이 낮을 경우, 연산 속도가 느려지거나 아예 연산을 할 수 없는 등 칩의 사용성을 크게 떨어뜨릴 수 있으며, 결과적으로 고객사 소프트웨어 개발자의 생산성을 저해하게 된다.

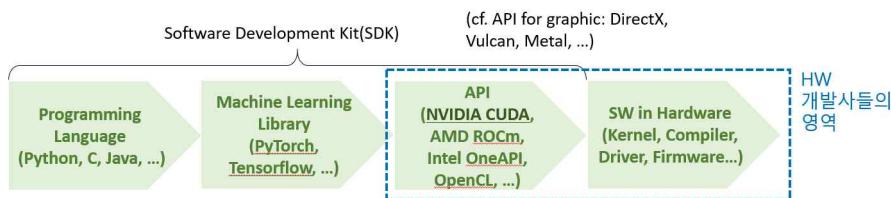
셋째, 대규모 연산을 수행하기 위한 칩간, 노드간, 시스템간 원활한 데이터 이동 및 통신을 지원하기 위한 네트워크 기술이다. LLM과 같은 대규모 모델의 경우 단일 칩으로는 연산이 불가능하기 때문에, 이러한 네트워크 기술은 필수적이다. 그리고 범용 기술(chip 간: PCIe gen5, 시스템간: Infiniband 등)이나 경쟁사 기술에 비해 NVIDIA NVLink, NVSwitch가 월등한 대역폭과 구성 유연성을 갖추고 있는 것으로 알려져 있어 NVIDIA의 시장지배력의 근원이 되고 있다.

이상과 같은 AI 반도체의 성능 향상을 위해서는 우선 반도체 칩(하드웨어, HW)의 아키텍처가 효율적으로 설계되어야 하며, 소프트웨어 스택(Software Stack)을 충실히 갖추는 것이 그 못지않게 중요하다. 사용자(서비스 개발자)가 효율적으로 SW를 개발하기 위해서는 기존의 많은 연구자, 개발자들이 축적한 PyTorch, Tensorflow, ONNX와 같은 머신러닝 라이브러리⁷⁾를 쉽게 활용할 수 있어야 한다. 다시 말해 머신러닝 서비스 개발자들이 프로그램을 개발할 때 CUDA가 지원되는 NVIDIA NPU를 사용할 경우와 사용하지 않을 경우의 생산성 격차가 크다(기존 서비스나 논문 등에서 제안된 프로그램을 활용함으로써 작업시간을 줄이고, 예기치 않은 오류 발생 가능성

7) 머신러닝에 필요한 연산을 미리 프로그래밍해 모듈화한 프로그램의 모음.

을 낮추며, 높은 연산 속도를 얻는 데 크게 유리하다). 그리고 이러한 머신러닝 라이브러리를 쉽게 활용함과 동시에 위에서 언급한 효율적인 연산 성능을 얻기 위해서는 사용자와 HW를 잇는 SW 스택, 즉 API8)를 포함한 Software Development Kit(SDK)과 정교하게 HW를 제어할 수 있는 SW in HW(kernel, compiler, driver, firmware...)를 충실히 갖춰야 한다.

[그림 2-6] 머신러닝 연산을 위한 소프트웨어 스택



출처: 연구진 작성.

그리고 이러한 SW 스택의 개발은 폭넓은 연산 지원, 운영 실적, 참고할 만한 개발 문서의 양과 질, 커뮤니티 지원(Q&A) 등과 직결되므로 기존 사용자가 많을수록 신규 서비스 개발이 간편해지는 네트워크 효과가 작용하게 된다. NVIDIA의 CUDA는 2006년경부터 시장을 선점한 결과 현존하는 대부분의 연산을 지원하며 동작이 검증된 API로 자리매김하고 있으며, PyTorch, Tensorflow 모두 NVIDIA CUDA를 필수적으로 요구하고 있다⁹⁾. 현재 NVIDIA의 시장지배력이 절대적인 것은 칩 성능뿐만 아니라 CUDA로 인한 lock-in 효과가 강력하게 작용한 결과로 볼 수 있다.

한편 AI 반도체 회사의 운영에는 타 분야 팜리스들과는 달리 막대한 자원이 요구된다. 우선 숙련도가 높은 하드웨어 설계, 소프트웨어 개발자가 다수 필요한데, 국내 업체들도 수십 명에서 100명 이상의 인원이 재직하고 있는 경우가 많아 인건비 부담이 크다. 또한 Edge 시장을 제외하면 요구성능을 충족시키기 위해 3nm, 5nm 등 초미세 공정을 사용해야 하는데, 그럴 경우 설계 툴, IP 라이선스, 파운드리 비용 등을 고려하면 수천 억원의 비용이 필요한 것으로 알려져 있다.

8) HW를 원활히 구동하기 위한 SW 계층, 기업 고유(NVIDIA CUDA, AMD ROCm 등) 혹은 범용(OpenCL)이 있다.

9) 최근 AMD ROCm이 공식적으로 지원되기 시작하였으나, 완성도는 NVIDIA CUDA만큼 높지 않다.

5) 칩 제조

칩 설계를 마치면 파운드리에 위탁해 칩을 생산해야 한다. 초미세공정은 현재 TSMC(~3nm)와 Intel(~3nm), 삼성전자(~4nm)의 3사가 생산하고 있으며, 그 외 GlobalFoundries, UMC 등이 12nm 혹은 14nm까지의 칩을 생산하고 있다.¹⁰⁾ 이들 중 TSMC는 제작된 칩의 성능뿐만 아니라 고객사(팹리스) 대응 및 지원체계 측면에서도 독보적인 선두 업체로 인식되고 있다. 삼성전자는 초미세공정에서 TSMC의 유일한 대안이지만, 칩 성능과 고객사 지원 모두 뒤지는 측면이 있다. 이를테면 여러 고객사를 모아 시제품을 제작하는 기회(MPW)를 TSMC가 연간 120회 정도, 삼성전자의 3~4 배 제공하는 것으로 알려져 있다.¹¹⁾

6) 마케팅

최종적으로는 제작된 칩을 보드에 장착해 고객사를 발굴, 판매해야 한다. 여기서 가장 중요하게 작용하는 것은 이전의 납품 및 운용 실적이라고 할 수 있다. 거대 비즈니스를 운영하는 고객사 입장에서는 검증된 솔루션을 채택하는 것이 안전하기 때문이다. 그 외에 개발팀의 이력이 뛰어난 경우 마케팅에 긍정적으로 작용할 수 있다.

마지막으로 신생 회사 입장에서 가장 현실적으로 접근 가능한 마케팅 수단은 MLPerf와 같은 벤치마크에 참여하는 것이다. 그러나 전문가 인터뷰에 따르면 이러한 벤치마크의 몇 가지 항목에서 뛰어난 성능을 거두었다고 해서 바로 납품 성공 가능성 이 높아지는 것은 아닌데, MLPerf의 테스트는 성능 평가로서 의미가 없지는 않지만

10) TSMC, 「Logic Technology」, <https://www.tsmc.com/english/dedicatedFoundry/technology/logic> (검색일: 2025.2.3.)

Intel, 「Semiconductor Manufacturing Process Technologies」, <https://www.intel.com/content/www/us/en/foundry/process.html> (검색일: 2025.2.3.)

경향신문(2024. 10. 9.), 「초미세공정 낮은 수율에 불접혀…파운드리 '찬바람', 삼성은 춥다」, <https://www.khan.co.kr/article/202410090600015> (검색일: 2025.2.3.)

UMC, 「Logic」, <https://www.umc.com/en/Product/technologies/Index/logic> (검색일: 2025.2.3.)

GlobalFoundries, 「FinFET 12LP+」, <https://gf.com/technology-platforms/finfet/> (검색일: 2025.2.3.)

삼성전자의 경우 3nm 공정 수율이 낮아 갤럭시 S25 탑재 모바일 AP 등 의미 있는 양산을 하고 있지 못하므로 4nm로 표기하였음.

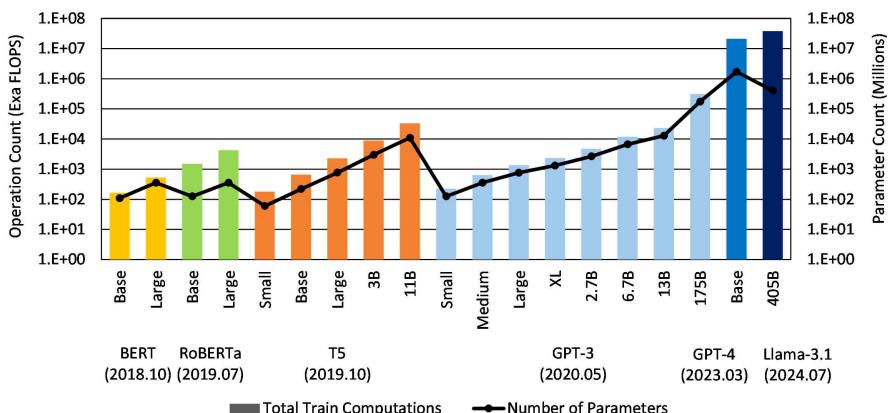
11) 헤럴드경제, 「TSMC 'MPW' 연간 120회 테스트, 삼성은 4분의1 수준」, <https://biz.heraldcorp.com/view.php?ud=20231004000304> (검색일: 2024.6.2.)

비교 조건이 엄격하게 통제되지 않고 실제 운용상 필요한 다양한 연산이나 볼륨상황을 모두 반영하고 있지 않기 때문이다.

나. 거대 인공지능 시대의 인공지능 반도체 기술 발전 양상¹²⁾

2010년대 초 AlexNet을 시점으로 등장하기 시작한 합성곱 인공신경망(convolutional neural network, CNN) 기반의 시각 인공지능은 주로 소규모의 그래픽 처리장치(graphics processing unit, GPU)로 구성된 학습 환경에서도 인공신경망 모델의 학습이 가능했다. 따라서, 초기 인공지능 전용 반도체 기술은 슈퍼컴퓨터 규모의 인공지능 학습을 위한 대규모 기술보다는 소규모의 인공신경망 추론 전용 반도체 등의 기술이 주목받았다. 하지만, 2017년 등장한 트랜스포머 기반의 언어 모델이 주목받기 시작하고 거대 언어 모델이 높은 성능을 보이자, 인공지능 반도체 기술도 거대 언어 모델을 위한 형태로 바뀌어 나가기 시작했다.

[그림 2-7] 트랜스포머 모델의 학습 파라미터 및 연산량 규모



자료: 저자 작성

12) 본 항목은 전원, 여준기(2023), 전원(2024.5.28.)을 주로 참고하고 전문가(전원) 자문을 통해 최신 동향에 맞게 작성하였다.

인공지능을 학습하고 사용하는 데에 필요한 연산량 및 컴퓨팅 파워는 해당 인공신경망 모델의 파라미터 수를 기준으로 파악할 수 있다. 시각 인공지능은 수천만에서 수억 개의 상대적으로 적은 수의 파라미터로도 학습될 수 있었다. 반면, 트랜스포머를 기반으로 설계된 거대 언어 모델은 수천억에서 조 단위의 파라미터를 필요로 한다. OpenAI의 ChatGPT에 적용되었던 GPT-4 모델에는 약 1.7조 개의 파라미터가 사용된 것으로 알려져 있으며, 최근 Meta에서 공개한 Llama-3.1에는 약 4,000억 개의 파라미터가 쓰인 모델이 존재한다.

<표 2-1> 온라인 서비스별 월간 1억 사용자 달성을 걸린 시간

(단위: 월)

서비스	출시 월	개월 수
ChatGPT	2022.11.	2
TikTok	2016.09.	9
Instagram	2010.10.	30
Pinterest	2009.12.	41
Spotify	2008.10.	55
Telegram	2013.08.	61
Uber	2010.06.	70
Google Translate	2006.04.	78

자료: Yahoo!finance(2023). 「ChatGPT on track to surpass 100 million users faster than TikTok or Instagram: UBS」, <https://finance.yahoo.com/news/chatgpt-on-track-to-surpass-100-million-users-faster-than-tiktok-or-instagram-ubs-214423357.html> (검색일: 2024.10.25.)

거대 인공지능의 등장은 각 인공지능 모델의 규모가 커지는 것뿐만 아니라, 인공지능 서비스의 사용자를 폭발적으로 증가시켰다. 초기 시각 인공지능은 이미지 속 사물 인식, 물체 테두리 인식 등 다양한 이미지 처리 기능을 제공했지만, 자율 주행 등 매우 제한된 사용처만 있었을 뿐, 일반 사용자를 대상으로 하는 대규모 서비스가 주를 이루지는 못했다. 반면, ChatGPT 등의 거대 언어 모델은 일반 사용자도 쉽게 접근할 수 있는 편의성과 작문, 번역, 교정 등 매우 다양한 언어 기능을 제공하여 매우 짧은 시간 안에 대규모 사용자를 지닌 서비스로 발전했다. ChatGPT는 출시 후 두 달 만에 월간 1억 사용자를 확보했는데, 이는 당시까지 출시된 온라인 서비스 중 가장 빠른 달성

속도였다. 특히, TikTok, Instagram과 같은 소셜 미디어와 같은 일상 서비스가 아닌 인공지능 서비스가 폭발적인 사용자를 확보한 것은 의미가 컸으며, ChatGPT는 2024년 중순에 주간 2억 사용자를 돌파했다(Reuters, 2024). ChatGPT, Gemini, CLOVA X, Copilot 등의 인공지능 챗봇 서비스뿐만 아니라, DALL-E, Stable diffusion, Sora와 같은 생성형 인공지능이 일반 사용자에게 높은 관심을 받으면서, 전반적인 인공지능 서비스 모두에 대한 수요가 높아지고 있다.

개별 인공지능 모델의 규모가 거대해지고, 그를 필요로 하는 사용자 수가 폭증하면서, 인공지능 반도체 기술의 발전은 초기 인공지능 시대와 완전히 다른 양상을 띠고 있다. 2012년 AlexNet은 NVIDIA GTX580 GPU 2개를 사용해서 학습되었다. 반면, 2024년 Meta의 Llama-3.1 405B 모델은 NVIDIA H100 GPU 16,384개로 구성된 대규모 컴퓨팅 시스템을 사용하여 학습되었다(Meta, 2024). 이처럼 인공지능 학습을 위한 컴퓨팅 시스템이 단일 노드를 넘어 슈퍼컴퓨터 수준으로 거대해짐에 따라, 단일 인공지능 반도체의 성능을 향상시키는 것 뿐만 아니라, 수많은 컴퓨팅 디바이스 간 데이터를 전달하고, 인공지능 분산 학습을 수행할 수 있는 하드웨어 및 소프트웨어 기술이 필수 요소가 되었다.

[그림 2-8] 국내외 다양한 인공지능 반도체



자료: NVIDIA H100, Tesla D1, Google TPU, Samsung HBM-PIM, SK Hynix AiM, SAPEON X330, rebellion_ ATOM, Mobilint Aries, Graphcore IPU, Intel Gaudi, AMD MI300, FuriosaAI RNDI, DeepX DX-M1, Cerebras WSE-3, ETRI AB21

국내 다양한 반도체 설계 기업에서도 인공지능 반도체를 개발하고 있지만, 거대 인공지능을 위한 대규모 인공지능 학습 시스템에는 아직 활용되지 못하고 있다. 국내뿐만 아니라 해외에서도 다양한 그룹에서 인공지능 반도체를 설계하고 있지만, GPT-4, Llama-3.1과 같은 거대 인공지능 학습을 위한 인공지능 반도체 시장은 NVIDIA가 사실상의 독점하고 있다.

다. 국내외 인공지능 반도체 기술 비교 분석¹³⁾

초기 인공지능 연산은 병렬 처리에 유리한 GPU를 활용해 수행되어 왔다. 하지만, 그래픽 연산 처리에 특화된 GPU를 그대로 활용하는 것이 아닌, 인공지능 연산에 특화된 전용 가속기 구조와 반도체에 대한 수요가 증가함에 따라 다양한 형태의 인공지능 반도체가 제안되고 있다. 특히, Google은 2015년부터 빠르게 자체 설계 인공지능 가속기인 텐서 처리장치(tensor processing unit, TPU)를 제작하여 사용하고 있으며, NVIDIA도 2017년부터 V100 GPU 내부에 인공지능 연산을 더욱 효율적으로 처리할 수 있는 텐서코어(tensor core) 구조를 도입했다.

<표 2-2> 국내 인공지능 반도체 하드웨어 성능 현황

HW Spec.	Mobilint ARIES	SAPEON X330	Rebellions ATOM	FuriousAI RNGD
Tech node	2022, 14nm	2023, 7nm	2023, 5nm	2024, 5nm
Data types	INT8	FP16, FP8, INT8	FP16, INT8, INT4, INT2	BF16, FP8, INT8, INT4
16-bit TOPS	-	368 FP16	32 FP16	256 BF16
8-bit TOPS	80 INT8	734 FP8	128 INT8	512 INT8/FP8 1024 INT4
On-chip memory size(MB)	-	-	64	256
Off-chip memory size(GB)	32, LPDDR4	32, GDDR6	16, GDDR6	48, HBM3
Off-chip memory BW(GB/s)	66.7	512	256	1500

13) 본 항목은 전원, 여준기(2023)와 전원(2024.5.28.)을 주로 참조하고 전문가 자문(전원)을 통해 최신 동향에 맞게 작성하였다.

HW Spec.	Mobilint ARIES	SAPEON X330	Rebellions ATOM	FuriosaAI RNGD
16-bit TOPS/W	-	1.472	0.213	1.707
8-bit TOPS/W	3.2	2.936	0.853	3.413
Peak Power(W)	25	250	150	150
Major objective	AI vision	LLM	AI vision, NLP inference	LLM

자료: 전원, 여준기(2023), Oh, Y.(2023)

2023년 이전에 국내에서 설계된 Mobilint ARIES, SAPEON X220, Rebellions ION, FuriosaAI Warboy 등의 인공지능 반도체는 일반적으로 저전력 연산기와 메모리 모듈을 지니는 저전력 고효율 가속기 형태를 지녀, 시각 인공지능의 추론 연산을 엣지, 모바일 등의 환경에서 효율적으로 처리할 수 있는 것을 목표로 했다. 반면, 2023년 이후 새롭게 발표되고 있는 SAPEON X330, Rebellions ATOM 및 REBEL, FuriosaAI RNGD 등의 인공지능 프로세서들은 시각 인공지능보다 거대 언어 모델(large language model, LLM)을, 모델 추론보다 학습을 목표로 설계되고 있다. 이처럼 인공지능 연산의 주요 목표가 시각 인공지능 추론에서 거대 언어 모델 학습으로 변경됨에 따라 발생하게 된 인공지능 반도체의 주요 변경점은 다음과 같다.

- ① (다양한 데이터 탑입 연산 지원) 인공신경망의 추론 연산은 양자화 등의 기술을 통해 정수형 데이터로 처리할 수 있었다. 반면, 거대 언어 모델의 학습은 부동소수점 데이터 연산이 필수적이며, 32-bit, 16-bit, 8-bit, 4-bit 등 다양한 크기를 갖는 부동소수점 연산을 통해 성능을 올릴 수 있다.
- ② (칩당 성능 대폭 향상) 단일 디바이스에서도 처리될 수 있는 시각 인공지능 추론 연산 대비, 거대 언어 모델의 학습 연산은 수천~수만의 디바이스를 필요로 한다. 이때, 디바이스 간 비효율적인 데이터 이동을 최소화하기 위해서는 최대한 높은 단일 디바이스 연산 성능을 확보하는 것이 필요하다.
- ③ (고대역폭, 고용량의 메모리) 거대 언어 모델은 파라미터 수가 수천억~조 단위에 이를 만큼 규모가 크기 때문에, 이를 담기 위한 고용량의 메모리가 필수적이며, 대폭 향상된 연산기에 데이터를 지속적으로 공급하기 위한 HBM 등의 고대역폭 메모리가 필수적이다. FuriosaAI RNGD는 고대역폭 메모리 HBM3를 탑재했으

며, 발표 예정인 Rebellions REBEL 역시 HBM3E를 탑재할 예정이다.

- ④ (칩 간 데이터 통신) 수천~수만의 인공지능 반도체를 하나로 묶어 하나의 거대 언어 모델을 학습하기 위해 칩/노드/랙 간 고성능의 데이터 통신 기술이 필요하다. 예를 들어, Rebellions는 서버, 랙 단위의 scalable system solution을 제공하며, 기가비트 이더넷 등의 기술을 제공한다.

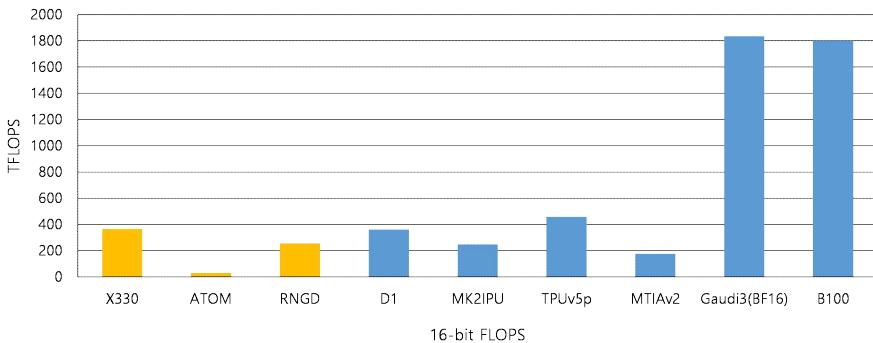
<표 2-3> 국외 인공지능 반도체 하드웨어 성능 현황

HW Spec.	Tesla D1	Graphcore MK2IPU	Google TPUs v5p	Meta MTIAv2	Intel Gaudi3	NVIDIA B100
Tech node	2021, 7nm, 645mm ²	2021, 7nm, 832mm ²	2023, -, -	2024, 5nm, 421mm ²	2024, 5nm, -	2024, 4nm, -
Data types	FP32, BF16, FP8, INT8	FP32, FP16	BF16, INT8	FP16, BF16, INT8	FP32, TF32, FP16, BF16, FP8, INT8, FP6, FP4	FP64, FP32, TF32, FP16, BF16, FP8, INT8, FP6, FP4
16-bit TOPS	362 BF16	250 FP16	459 BF16	177 FP16/BF16	459 FP16, 1835 BF16	1800 FP16/BF16
8-bit TOPS	362 FP8	-	918 INT8	354 INT8	1835 FP8	3500 FP8, 7000 FP4
On-chip memory size(MB)	442.5	897	-	256	96	-
Off-chip memory size(GB)	-	-	95, HBM2E	128, LPDDR5	128, HBM2E	192, HBM3E
Off-chip memory BW(GB/s)	-	-	2765	204	3700	8000
16-bit TOPS/W	0.905	0.833	-	1.967	3.058	2.571
8-bit TOPS/W	0.905	-	-	3.933	3.058	5.000
TDP(W)	400	300	-	90	600	700
Major objective	FSD training	LLM training	LLM training	DLRM	LLM training	LLM training

자료: 전원, 여준기(2023), Vahdat, A.(2023), Tal, E.(2024), NVIDIA(2024)

범위를 국내에서 해외로 넓혀보면 더 많은 그룹에서 인공지능 반도체를 설계하고 있다. 특히, Google과 NVIDIA는 인공지능 반도체를 초기부터 이끄는 선두 주자로, 각각 2015년과 2017년에 1세대의 TPU와 V100 텐서코어 GPU를 선보인 이후 지금 까지 다른 인공지능 반도체 대비 기술적인 우위를 점하고 있다. 특히, 두 반도체 모두 매우 초기부터 HBM을 채택하여 사용하고 있으며, Google은 TPU Pod, NVIDIA는 NVLink, NVSwitch 등의 고성능 칩 간 통신 기술을 보유하고 있어 거대 인공지능 학습에 필요한 기술을 고루 갖추고 있다. 실제로 GPT-4, Llama-3.1 등 다양한 거대 언어 모델이 NVIDIA GPU 위에서 학습되었으며, Google의 자체 거대 언어 모델인 Gemini는 TPU를 활용하여 학습되었다.

[그림 2-9] 인공지능 반도체 단일 칩 FP16/BF16 성능 비교



자료: 연구진 작성

개별 칩의 성능으로 보았을 때도 NVIDIA의 GPU, Intel Gaudi3는 특히 압도적인 성능을 보이고 있는데, B100의 3500TFLOPS(1012 floating-point operations per second)의 FP8 연산 성능은 국내에서 개발된 SAPEON X330의 734TFLOPS, FuriosaAI RNGD의 512TFLOPS 대비 4.77배, 6.84배 높은 성능이다.

표에 소개된 인공지능 반도체 외에도, Microsoft MAIA100, AMD MI300, Cerebras WSE-3, 삼성전자 MACH-1 등 수 많은 빅테크에서 인공지능 반도체를 발표, 개발하고 있다. 하지만, NVIDIA는 수많은 경쟁 칩을 제치고 데이터센터 규모의 인공

지능 반도체 시장의 90%를 독점하고 있다.¹⁴⁾ 물론 NVIDIA GPU 연산 칩의 성능, 고대역폭 메모리칩의 성능, 칩 간 데이터 통신 성능이 높은 것은 사실이지만, Gaudi3, MAIA100, MI300 등 일부 칩은 단일 칩 기준 NVIDIA GPU와 유사한 성능을 제공한다.

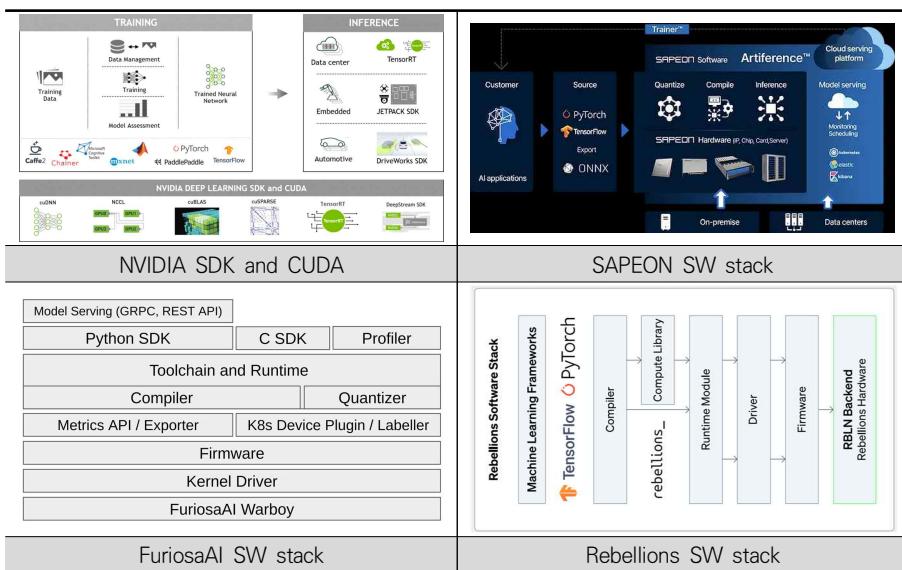
NVIDIA가 위와 같이 성능 경쟁에 위협이 되는 칩들 사이에서 높은 점유율을 유지할 수 있는 이유는 오랫동안 NVIDIA가 투자해 온 CUDA 기술 덕분이라고 볼 수 있다. NVIDIA GPU를 사용하기 위해서 프로그래머는 반드시 CUDA를 사용해야 한다. CUDA는 약 2007년 부터 병렬프로그램을 작성하는 데에 사용됐으며, 2010년대 초, GPU를 활용한 인공지능 기술이 등장하기 시작했을 때부터 인공지능 병렬 처리의 표준 격으로 자리 잡았다. 이후, 전 세계 연구자들이 연구, 개발해 온 각종 인공지능 기술과 병렬 가속 알고리즘은 CUDA를 기반으로 작성되었으며, 이는 인공지능 연구자가 NVIDIA GPU에 락인(lock-in)되는 효과를 가져왔다. 즉, 인공지능 반도체의 사용자는 그간 CUDA로 개발되어 온 소프트웨어 자산을 모두 버리고 NVIDIA GPU에서 다른 인공지능 반도체로 환경을 옮겨갈 이유가 부족하기 때문에 NVIDIA의 독점이 지속된다고 볼 수 있다. 이러한 상황은 NVIDIA가 아닌 후발 주자에게, 앞서 소개된 1~4번의 요구사항 외에, 아래와 같은 추가 조건이 필요하게 만들었다.

- ① (과거 인공지능 연산 호환) 2012년 GPU로 가속되었던 AlexNet의 등장 이후 10년 이상의 기간 동안 제안된 대부분의 인공지능 기술은 NVIDIA GPU에서 자연스럽게 실행될 수 있다. 후발 주자는 CUDA 없이도 해당 인공지능 모델 또는 알고리즘들이 새롭게 제안하는 인공지능 반도체에서 구동될 수 있도록 지원해야 한다.
- ② (편리한 인공지능 개발을 위한 SDK 제공) CUDA로 개발된 각종 인공지능 라이브러리는 PyTorch, Tensorflow 등 각종 인공지능 개발 프레임워크와 자연스럽게 연결되어 있다. 후발 주자의 인공지능 반도체 위에서 새로운 인공지능 모델을 연구하고 개발하기 위해서는 해당 개발자가 프레임워크 안에서 새롭게 구현하는 인공지능 모델이 자연스럽게 새로운 인공지능 반도체로 전달될 수 있도록 SDK(software development kit)를 제공하는 것이 중요하다.
- ③ (소프트웨어 성능 최적화) 동일한 FLOPS의 성능을 지니는 서로 다른 인공지능 반도체에

14) 머니투데이(2024.2.27.). AI 칩 90% 독점한 엔비디아, 그 중심엔 ‘쿠다’…장기 집권의 무기. <https://news.mt.co.kr/mtview.php?no=2024022712195070004> (검색일: 2024.10.25.)

서 같은 인공지능 모델을 실행할 때, 소프트웨어 단계에서 병렬 연산 및 온칩 메모리 최적화 등이 잘 이루어져 있는지에 따라 실제 성능에 매우 큰 차이가 발생할 수 있다. NVIDIA는 오랜 기간 CUDA에서 실행되는 병렬 연산 알고리즘에 대하여, low level(assembly 등)에서 고강도의 최적화를 수행해 왔다. 후발 주자는 이미 사용되고 있는 병렬 연산이 제안하는 인공지능 반도체 위에서 최적의 성능을 달성할 수 있게 만들어야 하며, 이후에도 새로운 연산 알고리즘이 나올 때마다 지속적인 최적화 지원을 할 수 있어야 한다.

[그림 2-10] 다양한 인공지능 반도체의 SDK 또는 소프트웨어 스택



자료: NVIDIA, SAPEON, FuriosaAI, Rebellions 각사 홈페이지(검색일: 2024. 10. 25.)

SAPEON, FuriosaAI, Rebellions, Mobilint 등 국내 인공지능 반도체 설계 그룹도 개발된 칩을 효과적으로 사용할 수 있게 도와주는 SDK 및 소프트웨어 스택을 함께 개발하여, PyTorch, Tensorflow와 같은 인공지능 프레임워크와 연동되어 개발된 인공지능 반도체를 활용할 수 있는 환경을 제공한다. 이처럼 인공지능 반도체를 개발하는 그룹은 하드웨어에 관한 연구, 개발 투자뿐만 아니라, 인공지능 소프트웨어에 대한 넓은 개발 지원을 제공해야 한다.

2. 주요국 정책

전 세계적으로 전 사회·산업 분야의 디지털전환/인공지능전환(DX/AX)의 가속화에 따른 AI반도체를 비롯한 시스템반도체의 수요가 확대되며, 반도체는 전통적인 무역·통상 품목에서 전략자산 관점으로 인식이 급변, 주요국 간 기술·통상 경쟁이 심화하고 있다. 지난 수년간 반도체 산업을 중심으로 미국-중국 간 패권 경쟁이 이어지고 있으며, 최근 미국은 일부 그래픽카드에서 설계 SW 및 제조 장비까지 제재 품목을 확대하는 등 경쟁국의 첨단 AI 및 AI반도체 기술 확보를 저지하고 있다. 또한, 미국은 우방국 중심의 공급망 재편을 통해 대한민국, 네덜란드, 일본 등 대중국 반도체 장비 수출통제에 직·간접적인 참여를 요구하고 있다.

또한, 미국을 필두로 유럽, 대만 등 주요 선도국은 자체적인 「반도체법」을 마련함으로써 국가적 차원에서의 첨단기술 확보, 생산역량 확충, 기업지원 등을 위한 대대적인 지원에 나서고 있다. 미국의 「반도체과학법(22.8.)」을 필두로 대만 「산업혁신 조례 수정안(22.11.)」, EU 「반도체법(22.12.)」 등을 통해 역내 반도체 생산시설의 확보를 위한 보조금 지원, 첨단 설계·제조 기술 강화를 위한 연구개발 및 인력양성 등 국가적 지원을 위한 상위 근거를 마련하였다.

AI 반도체를 중심으로 미국, 중국, EU, 대만 등 주요 선도국의 반도체 기술·산업 지원 정책을 살펴보고자 한다.

가. 미국

먼저, 미국은 과거 차세대반도체 선행 연구개발 프로젝트 중심의 투자에서 '22년 이후 본격적인 역내 생산역량 강화 및 글로벌 기술·산업 주도를 위한 국가적 차원의 지원 정책을 추진하고 있다. '20년까지 국방부를 비롯한 연방정부는 차세대반도체 분야의 기술·산업 주도권 확보를 위해 원천 기술에 대한 중장기적인 투자를 추진한 바 있다. 대표적으로 국방부는 DARPA를 통해 뉴로모피 반도체 등 차세대 AI반도체 개발을 위한 'SyNAPSE(Systems of Neuromorphic Artificial Intelligence Components)' 프로젝트를 '08년부터 추진한 바 있으며, 일환으로 '14년 개발한 뉴로모피 반도체

'TureNorth'를 '17년 IBM社가 제한적으로나마 상용화를 추진한 바 있다. 한편, DARPA는 '18년 산·학·연 차세대 반도체 R&D 프로그램인 'ERI(Resurgence Initiative)'를 발표하였으며, 인텔, 퀄컴, IBM, Applied Materials, Synopsys 등 반도체 생태계 전반에 걸친 자국 대기업과 대학 등의 참여를 유도하여 '22년까지 15억 달러를 지원한 바 있다. 한편, '22년 후속 프로그램인 'ERI 2.0'을 출범하고 5년간 30억 달러 지원을 통해 연구개발 지원의 연속성을 확보하고 있다.

[그림 2-11] DARPA의 ERI 2.0 지원 기술 분야

DARPA ERI 2.0 may expand investment in dual use research

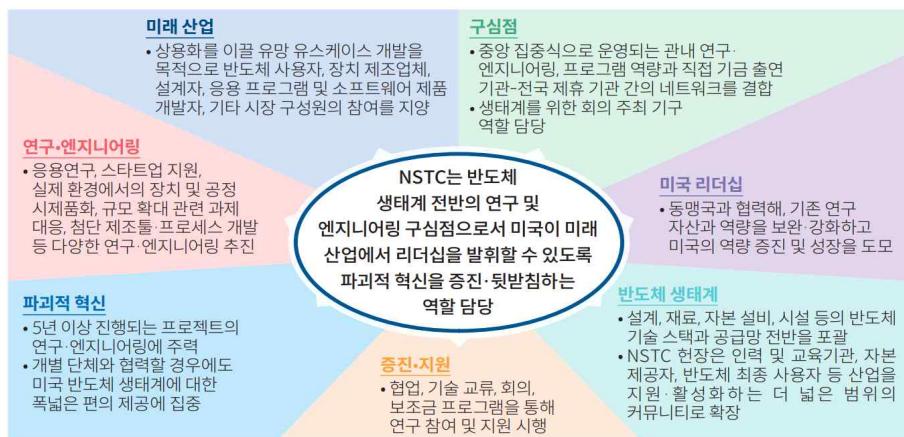
Current Areas of Emphasis				Potential New Areas
Overcoming security threats across the entire hardware lifecycle	Mitigating the skyrocketing costs of electronics design	Revolutionizing communications (5G and beyond)	Manufacturing complex 3D microsystems	
				
<ul style="list-style-type: none"> EDA based technology Inspection and supply chain based technology 	<ul style="list-style-type: none"> Foundry-portable IP Design tools with ML capability 	<ul style="list-style-type: none"> Power efficient digital arrays Techniques for secure communications 	<ul style="list-style-type: none"> ML / AI automated tools Desktop assembly 	
Overcoming the inherent throughput limits of 2D electronics	Accelerating innovation in AI hardware to make decisions at the edge faster	Increasing information processing density and efficiency	Developing electronics for harsh environments	
				
<ul style="list-style-type: none"> Heterogeneous electronics with Si-like back-end Integration of photonics / optics 	<ul style="list-style-type: none"> Alternative / approximate computing AI / ML integrated with HW 	<ul style="list-style-type: none"> New computing architectures, devices, and materials Quantum-inspired algorithms 	<ul style="list-style-type: none"> Radiation-hardened electronics High-temperature electronics 	

출처: "Evolving the Electronics Resurgence Initiative (ERI 2.0)" – Briefing prepared for NDIA, Mark Rosker, MTO, 2021.04.21. DARPA 11p

한편, COVID-19 팬데믹 중 전통적인 글로벌 공급망에 대한 불확실성 확대에 따라 미국은 반도체 산업을 국가적 안보 관점으로 접근하며 「반도체과학법((Chips and Science Act, 이하 美반도체법)」을 통해 반도체 제조시설의 리쇼어링과 「국가반도체 기술센터(National Semiconductor Technology Center, 이하 NSTC)」 설립 등을 통해 자국 제조역량 강화 및 첨단반도체 기술 연구에 앞장서고 있다. 「美반도체법」에서 제시한 520억 달러의 투자액 중 다수는 첨단 제조시설 유치를 위한 글로벌 기업에의 보조금과 세액공제 혜택을 위해 지원되고 있다. 이에 따라 삼성전자, TSMC, 인텔 등 주요

글로벌 반도체 기업은 애리조나·뉴욕·텍사스주 등에 신규 제조사설을 구축을 계획하고 있다. 이처럼 자국 중심의 공급망을 형성하는 한편, 보조금 수혜기업에는 우려 대상 국가(중국, 러시아, 이란, 북한 등)에 향후 10년간 반도체 시설에 대한 투자를 금지하는 '가드레일 조항('반도체법 지원금 오용 방지 규칙')'을 적용함으로써 경쟁국의 기술개발을 견제하는 장치를 마련하였다¹⁵⁾. 또한, '美반도체법'에 의거, 미국 반도체 경쟁력·안보 강화를 위해 'NSTC'를 설립하고, 첨단반도체 기술 연구·개발 및 인력양성 추진하고 있다. NSTC는 NSTC는 민간부문, 에너지부, 국립과학재단이 참여하는 민관 컨소시엄으로 운영, 정부, 국립연구소, 산업체, 공급업체, 교육기관, 기업가, 노동계, 투자자 등이 플랫폼을 구축하여 미국 내 반도체 제조부문 활성화 및 첨단 시제품화 역량을 지원을 계획하는 한편, 자국의 반도체 생태계 기술·산업·연구의 구심점으로써 동맹국과의 협력을 통한 리더십을 발휘하고 있다.

[그림 2-12] NSTC의 역할과 목표



출처: NIST. (2023). A Vision and Strategy for the National Semiconductor Technology Center, 3-12. 국회도서관, 「현안 외국에선? 2024-8호」, 2024 4p “NSTC의 미션” 재인용

15) [기획]“AI반도체 주도권 잡아라” 글로벌 대전, 2023.07.03., 매일일보, <https://www.m-i.kr/news/articleView.html?idno=1027363>

또한, 미국은 '22년 이후 AI 및 AI반도체 기술에 대한 중국, 러시아 등 경쟁국의 접근을 차단하고, 미국 기업의 대중국 투자를 제한하는 행정명령(23.08.)을 발표하는 등 제재 수단을 자속해서 강화해 나가고 있다. 중국의 AI·반도체 기술·산업 발전 억제를 위해 지난 '22년 두 차례(9월, 10월)에 걸쳐 글로벌 공급망 길목(Check Point) 통제 조치를 강화하는 한편 (S&T GPS, 2022), 美상무부 산업안보국은 수출관리규정(EAR) 개정안을 통해 대중국 수출 통제 대상 반도체 칩과 장비의 범위를 확대하고, 우회 수출 경로 차단(23.10.)한 바 있다 (S&T GPS, 2023). 이와 같은 수출통제에 따라 중국 기업은 막대한 타격을 입게 되었으며, 주요 반도체 업체는 미국 내 시설 투자를 확대하는 등 기술 경쟁 장기화에 대비하고 있다¹⁶⁾. 더불어 AI 산업에 활용가능한 반도체 수출 통제를 통하여 중국의 AI 부문의 발전에 제동을 걸 수 있는 장치로 활용하고 있다.

<표 2-4> 미국의 반도체 분야 수출·접근통제 주요 내용

제한 품목	주요 내용
첨단 인공지능 반도체	<ul style="list-style-type: none"> 대규모 인공지능 모델을 위한 데이터센터 및 슈퍼컴퓨터 구현 등을 위한 일정 성능 이상의 첨단반도체*는 상무부의 수출 허가가 필요 * 강력한 연산성능(300TFLOPS 이상)과 빠른 데이터 입출력 속도(600GB/s 이상) ※ 미국의 NVIDIA와 AMD는 이러한 첨단 반도체 설계·생산이 가능한 대표적인 업체로, 이들 제품을 사용하지 않는 경우 인공지능 구현을 위한 범용SW 생태계에서 제외될 수 있음
설계SW	<ul style="list-style-type: none"> 해외직접생산품구칙*을 통해 중국 팝리스 기업이 자국 SW를 활용하여 설계한 반도체는 중국 외에서 생산할 수 없도록 제한 * Foreign Direct Product Rule : 제3국 생산 제품이더라도, 미국의 기술이나 SW를 활용 하여 생산한 경우 미국(상무부) 허가 없이 수출금지 ※ 반도체 설계 SW(EDA) 선도 기업은 대부분 미국에 본사를 두고 있어 이와 같은 조치를 통해 중국 등 해외 반도체 설계 산업 성장 억제 가능
반도체 제조장비	<ul style="list-style-type: none"> 반도체 종류(로직, DRAM, NAND 등)에 따라 수출제한 조치를 차별화*하여 적용하고, 기존 제조 장비 뿐 아니라 검사·측정장비 등 응용 도구의 공급 제한 * (로직) FinFET 구조나 16/14nm 이하, (DRAM) 18nm 이하, (NAND) 128-layer 이상
자국 제조 부품	<ul style="list-style-type: none"> 중국의 자체적인 생산 장비 개발을 막기위해 관련 부품의 수출을 금지 ※ 반도체 제조 기기·부품 분야는 미국·일본·네덜란드 등 소수의 주요국이 자속적인 경쟁 우위를 가짐

출처: KISTEP INI Vol.45 (2023) 53p

16) 제229호 과학기술&ICT동향 2022.12.30. 4p 재인용

나. 중국

중국은 AI·반도체 분야 등 핵심기술의 국가 전략화를 강조, 핵심기술 확보 및 장비 국산화를 추진 중이었으나, 미국의 제재로 난관에 봉착한 것으로 보인다. 중국은 '00년 부터 반도체 산업 발전정책을 발표, 제조기업에 대한 세수혜택 부여 및 핵심기술 R&D에 대한 투자를 시작하였으며, '14년부터는 반도체 국산화를 지원하기 위해 기금을 조성·운영하고 있으며, 최근에는 소재·장비·부품 기업을 중심으로 투자가 이루어진 바 있다.

<표 2-5> 중국 정부의 반도체 기금 운용 현황

반도체 산업기금 1기, '14~'19	<ul style="list-style-type: none"> 반도체 제조 능력 확대에 중점을 두고 총 23개 기업의 70개 프로젝트에 투자(총 투자규모 1,387억 위안)
반도체 산업기금 2기, '19~	<ul style="list-style-type: none"> 1기 대비 약 68% 증가하였으며, 소재·부품·장비의 국산화와 함께 EDA 분야, 통신, AI반도체, 화합물반도체 등 지원 범위를 확대(총 투자규모는 2,041억 위안)

출처: 중국의 반도체 국산화 추진 현황과 시사점, 세계경제포커스 Vol. 6 No. 20, 대외경제정책연구원, 오종혁, 2023.06.27

중국 정부는 '17년 '차세대 인공지능 발전계획'에서 '20년까지는 AI 기술·응용 수준을 선도국 수준으로 향상, '25년까지 일부 AI 기술·응용 분야에서 세계 최고 수준으로 도약함으로써 '30년까지 세계 1위 AI 중심 국가로의 도약을 천명한 바 있다. 후속 방안으로는 '차세대 AI사업 발전 3개년 행동계획('18~'20)을 통해 AI 관련 차세대 SW·HW 등 발전을 위한 기반 구축 등을 제안하였으며, 동 정책의 일환으로 '18년, 중국과학원 산하 자동화연구소는 인공지능혁신연구원을 설립하고 중국 반도체기업 Spreadtrum(現 UniSOC) 등과 협력해 AI반도체 개발 등을 추진하였다.

한편, '21년 '14차 5개년 계획 및 2035 중장기 목표'의 7대 전략육성 분야로 AI, 반도체 등을 강조, 대대적인 지원을 추진 중이나 자립화와 기술 선도는 담보 중인 실정이다. 차세대 AI 분야 적용을 위한 전용 칩 개발, AI 알고리즘 플랫폼 구축 등, 반도체 분야는 설계도구(EDA), 소재, 화합물 반도체 등 기술 확보를 위한 국가적 지원을 제시한 바 있다. 다만, 대규모 펀드와 각종 세제지원 등 전방위적인 자국 AI·반도체 기술·산업 육성정책을 추진 중이나, 여전히 자립화율은 10% 수준으로 담보하고 있는

것으로 알려져 있다¹⁷⁾. 또한, 미국의 전방위적인 대중국 압박으로 우방국 중심의 EUV 등 장비 수출 및 신규 파운드리 구축이 사실상 금지됨에 따라 자국 내 7나노 미만 첨단 반도체 생산이 사실상 불가능한 실정으로, AI반도체를 비롯한 차세대 반도체 분야에 대한 자체적인 기술·생산 역량 확보에 난항을 겪고 있다.

중국 정부는 이와 같은 기술·산업 제재를 타개하고자, ‘컴퓨팅 인프라 고품질 발전을 위한 행동계획(‘23)’ 등 주요 정책을 통한 자국산 AI반도체 사용 시 보조금 지급 및 수요 창출을 위한 데이터센터 구축 프로젝트 등을 추진하고 있다. 또한, 자국의 보안 강화 등 이유로 정부기관 및 공기업에서는 ’27년까지 현재 전자기기를 자국 반도체를 포함한 제품으로 교체하는 지시가 이루어지는 등 기술·산업 내재화를 위한 움직임을 보이고 있다.

다. 대만

대만은 첨단 반도체 제조 분야 기술우위와 법·정책 등 기반을 바탕으로 전세계 기술·산업 선도를 위한 첨단 제조 기술 확보, 핵심 소재·장비 경쟁력 강화, 고급인재 육성·유치 등을 지속 지원하고 있다. 특히, ’23년 1월 ‘대만형 반도체법’으로 일컫는 「산업혁신 조례」를 통해 첨단 공정개발 촉진을 통한 경쟁우위와 반도체기업의 활동 기반을 강화한 것으로 평가된다(김동수 외 2024). 주요 내용으로는 세계 시장에서 기술혁신 및 글로벌 공급망을 주도하는 기업이 R&D 및 EUV 등 선진 생산 공정 설비에 투자하는 경우 각각 투자비의 25%, 5% 세액을 공제하는 인센티브를 포함하고 있으며, 대만 내 주요 반도체기업이 신청서를 제출한 가운데 사실상 해외 투자를 확대 중인 TSMC의 자국 내 유인책이라는 해석이 우세하다. 앞서 언급한 바와 같이 전세계 1위 파운드리 기업인 TSMC는 자국 외에 미국, 중국, 일본, 독일 등에 신규 제조시설을 구축하는 등 첨단 반도체 제조 분야의 지정학적 영향력을 지속해서 확대하고 있다. 자국 내 신주, 타이중, 타이난, 가오슝 지역에 신규 반도체 공장을 건설·계획 중이며, 글로벌 주요국의 수요 및 특수성을 고려하여 첨단·레거시 제조 공장을 대대적으로 구축하고 있다. 대표적으로 일본 구마모토현에 ’27년 가동을 목표로 제2공장 추가 건설을 발표하였으며, 미국 애리조나주에는 현재 건설 중인 1, 2공장 외에도 첨단 패키징

17) 정형/비정형데이터 기반 반도체 시장 이슈 분석 및 예측 모형 개발, 정책자료 23-07, 정보통신 정책 연구원, 고동환 외 39p

시설 중심의 3, 4공장 구축 계획을 발표한 바 있다.

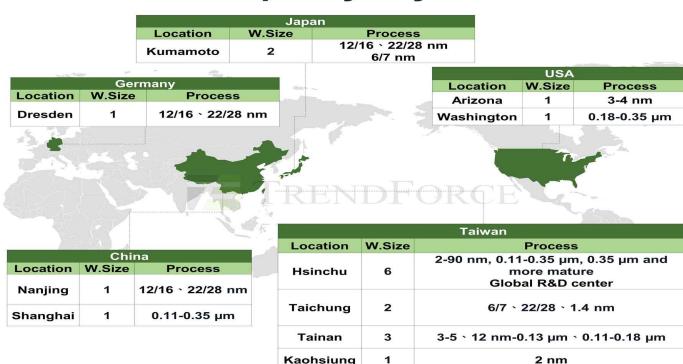
<표 2-6> TSMC의 신규 구축 중인 글로벌 제조시설

국가	위치	주요 공정 노드	TSMC 해외 제조시설
대만	신주	2~90나노, 110~350나노, 350나노 이상, 글로벌R&D센터	드레스덴* (12~28nm) 난징 (12~28nm) 워싱턴 (0.18~0.35μm) 애리조나 1~3공장* (2~4nm)
	타이중	6/7나노, 22/28나노, 첨단 제조 계획	중국
	타이난	3~5나노, 12~130나노, 110~180나노	상하이 (0.11~0.35μm)
	가오슝	2나노	일본 구마모토 1공장 (12~28nm) 구마모토 2공장* (6, 40nm)
미국	애리조나	4나노	
	워싱턴	180~350나노	
중국	난징	12/16나노, 22/28나노	
	상하이	110~350나노	
독일	ドレス덴	12/16나노, 22/28나노	
일본	구마모토	12/16나노, 22/28나노	

출처: (그림) TSMC 홈페이지 “Fab Location”, 중앙일보 (2024.04.30.), “미중 쌔움에 ‘한대’ 맞았다…삼성-TSMC 서로의 구원자 될까”, <https://www.joongang.co.kr/article/25246035>, Semicon Electronics(24.11.18), “TSMC’s Global Expansion Blueprint: Ten New Factories by 2025, with an Investment of up to 270 Billion Yuan” <https://www.semicon.com/article-99.html>

[그림 2-13] TSMC의 글로벌 제조시설

TSMC's Latest Global Production Capacity Layout



TRENDFORCE

출처: Trendforce(24.02.24.), “Latest Overview on TSMC’s Global Expansion Initiatives”, <https://www.trendforce.com/news/2024/02/24/news-latest-overview-on-tsmcs-global-expansion-initiatives/>

한편, 대만 정부는 '18년부터 「인공지능 반도체 제조공정 및 칩 시스템 R&D 프로젝트(Semiconductor Moonshot Project)」를 추진, 4년간 약 1,526억원의 예산을 투입하였다. 특히, 동 정책의 일환으로 '19년 「대만반도체연구소(TSRI)」를 설립하여 인공지능 반도체 개발 역량을 강화하기 위한 설계·공정·검증·인력양성 등의 통합 서비스를 제공하고 있다¹⁸⁾. 또한, '19년부터 'AI on chip Taiwan Alliance(AITA)'을 출범, 초기 ITRI(대만산업기술연구소), UMC社 등 56개 기관·기업의 참여로 4개의 협의체(Special Interest Groups, SIG)를 구성하며 자국 AI반도체 분야의 산업생태계 강화를 지원하고 있으며, '23년 현재 참여사는 152개 국내외 주요 반도체 업체·기관으로 확대되었다¹⁹⁾.

라. EU

EU회원국 역시 '23년 9월, '유럽 반도체법'으로 일컫는 「반도체 법안(Chips Act, 이하 '유럽 반도체법')」을 제정하였으며, 반도체 연구 및 기술 분야에서의 리더십 강화, 지속 가능한 첨단반도체 개발 능력 확충, 유럽의 첨단반도체 생산능력을 4배 확대, 전문인력 양성 등의 5대 목표를 제시하였다²⁰⁾. 그간 유럽은 전 세계 반도체 수요의 약 20%를 차지하였으나, 공급능력은 10%대 수준으로 알려져 있어 해외 의존도가 높은 실정이었다. 이를 극복하기 위해 '유럽 반도체법'은 '30년까지 만·관 투자를 통해 430억 유로 이상의 펀드를 조성하고, 반도체 생산(매출액 기준)에서 EU 비중을 최소 20%(현재 약 9% 수준)으로 확대할 것을 목표로 제시²¹⁾하였다.

먼저, 역내 첨단 칩 설계·제조·패키징 혁신 역량 구축·강화하기 위해 '유럽 반도체 이니셔티브(Chips for Europe Initiative, CEI)'를 발족하였으며, '30년까지 110억

18) National Applied Research Laboratories(2019.01.30.), "Taiwan Semiconductor Research Institute (TSRI) was officially launched on January 30, 2019", <https://www.narlabs.org.tw/en/xm/doc/cont?sid=0J049400520435607039&xsmid=0I160457997407279810> (검색일: 24.11.25.)

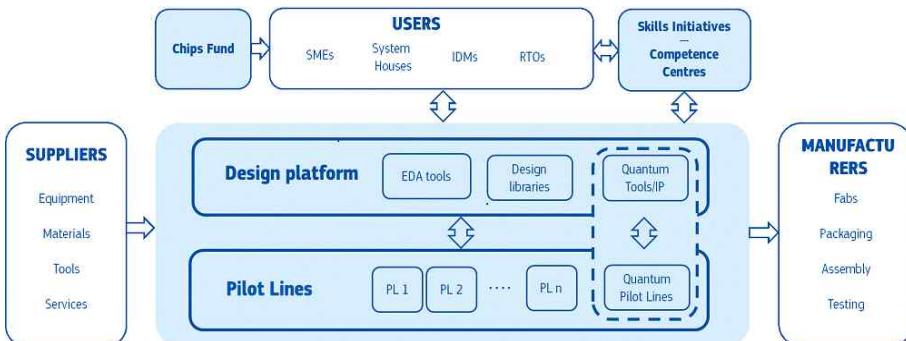
19) Nestrends Asia(2022.02.07.), "The golden age of IC design in Taiwan", <https://nxtrendsasia.org/the-golden-age-of-ic-design-in-taiwan/> (검색일: 24.11.25.)

20) [이슈분석 250호] 글로벌 공급망 재편에 관한 EU의 주요 정책방향, https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=48745&nPage=3

21) 20) 상동

유로 규모의 공공투자를 바탕으로 ▲기준 연구·개발·혁신 강화 ▲첨단 반도체 설계 툴, 차세대 소자 생산라인 ▲최신 반도체 기술 응용을 위한 테스트 시설 보급 ▲인력 교육 등을 추진할 계획이다²²⁾. 특히, CEI를 바탕으로 마이크로전자공학 부문 대학원 프로그램, 단기 훈련 과정, 취업 알선·연수·견습, 첨단 실험실 내 훈련 등에 대한 접근성을 제공하고, 유럽 전역 역량센터 네트워크 지원을 통해 중소기업을 비롯한 기업의 설계 역량 제고, 스킬 개발을 뒷받침하는 등 교육·훈련·재교육을 통한 전문인력 양성에 집중하고 있다. 또한, 2나노 미만의 소자 설계·제조 기술, AI 혁신 기술, 초저전력 에너지효율 프로세서, RISC-V 컴퓨팅 아키텍쳐 등에 기반한 새로운 설계 솔루션, 신소재 통합 등의 연구 활동 지원 「Horizon Europe」 및 「Digital Europe Programme」 등 EU의 연구 프로그램에서 총 33억 유로의 예산을 투입할 예정이다. 이와 더불어 프랑스·벨기에·독일 등 주요 EU 회원국별로 자국 내 반도체 분야 연구시설(클러스터)를 보유 중으로, “30년까지 역내 생산역량 20% 확보”를 천명한 「유럽 CHIPS법」의 목표와 같이 사실상 유럽 전역이 거대한 반도체 클러스터로써 협력 추진하고 있다.

[그림 2-14] Overview of the Chips for Europe Initiative



출처: European Commission(24.11.04.), “European Chips Act: The Chips for Europe Initiative”, <https://digital-strategy.ec.europa.eu/en/factpages/european-chips-act-chips-europe-initiative>

22) European Commission(24.11.04.), “European Chips Act: The Chips for Europe Initiative”, <https://digital-strategy.ec.europa.eu/en/factpages/european-chips-act-chips-europe-initiative>

또한, 유럽 역내 반도체 생태계 활성화 및 공급역량 확보를 위해 반도체기금(Chips Fund)으로 통칭되는 투자 활동 실시함으로써 금융 접근성 촉진 및 역동적이고 복원력 있는 반도체 생태계 발전 지원을 추진하고 있다²³⁾. 또한, 민간투자 촉진을 위해 유럽 반도체 공급 확보에 기여할 수 있는 ▲개방형 EU 파운드리²⁴⁾, ▲통합 생산 설비²⁵⁾ 등 두 가지 유형의 시설에 대해 건설·운영 관련 회원국의 허가·승인 절차 관련 패스트 트랙을 제공, 유럽 반도체 이니셔티브에 따라 구축된 시범생산라인의 우선적 이용권 부여 등 실효성 있는 지원 정책을 제시하고 있다. 마지막으로 글로벌 공급망에 대한 이해 제고와 위기 예측을 위해 공급망 부족 위험에 대한 조기 경고 지표를 파악, 주요 위험을 예측할 수 있는 ‘조기경보지표’를 개발하고 모니터링 시스템을 구축할 예정이다. 공급망 위기 발생 시 집행위는 위기의 심각성을 파악하고, 위기 단계 활성화 여부 파악 및 주요국·기업간 협력체계 마련 등의 조치를 취하게 된다.

한편, EU는 ’18년 역내 필수 제품의 대외 무역의존성에 따라 수직적 산업 정책 필요성이 증대되면서 전략적 자율성 강화 방안으로 「공통관심 주요 프로젝트(Important Project of Common European Interest, IPCEI)」를 통해 AI 프로세서, 엣지컴퓨팅, 모빌리티, 5G·6G 등 차세대 핵심기술에 대한 연구개발 프로젝트를 지원하고 있다(권지연 외 2022). 동 프레임워크는 EU의 경쟁력 강화, 지속가능성, 역내 가치창출, 복수국 참여 등을 조건으로 회원국의 산업보조금 지원을 허용하고 있으며. 정부-민간 공동 투자 원칙, R&D 집중도 및 혁신 사업 등에 진행되며 EU 집행위원회는 직접적인 보조금 지급에 참여하지 않으나, 프로젝트의 진행 과정을 모니터링하고, 관련 업계 간의 포럼 등을 독려²⁶⁾하고 있다.

23) 미·중 기술패권 전쟁 2.0 [전자자료] : 반도체 전쟁(2022), 한국산업기술진흥원 산업기술정책센터 동향조사연구팀(KlaT)

24) (개방형 EU 파운드리, Open EU Foundries) 주로 다른 산업 행위자를 위해 칩을 설계·생산하는 시설

25) (통합 생산 설비, Integrated Production Facilities) 자체 시장에 서비스를 제공하는 부품의 설계·생산 공장이 해당되나, 유럽 최초 시설로서 운영자가 EU 반도체 부문 혁신을 위한 지속적인 투자를 약속하는 경우로 한정

26) [이슈분석 250호] 글로벌 공급망 재편에 관한 EU의 주요 정책방향, https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=48745&nPage=3

3. 주요 기업

가. NVIDIA(NVIDIA Corporation)

□ 기업 개요

- NVIDIA는 미국 캘리포니아주에 위치한 반도체 설계 (팹리스) 기업.
 - 주요 제품으로 GUI(그래픽처리장치), API(애플리케이션 프로그래밍 인터페이스), SoC(시스템 온 칩 유닛) 등이 있으며, 하드웨어 뿐 아니라 소프트웨어 영역에서도 돋보이는 성과를 내고 있다²⁷⁾. 2024년 6월에는 시가총액 3조 3350억달러로 전세계 시가총액 1위 기업으로 올라서서²⁸⁾, 구글과 MS가 양분하던 글로벌 빅테크 시장의 주요한 기업으로 올라섰다.

□ 기술

- NVIDIA는 1999년도 최초의 GPU 개발이 동 기업 성공의 단초로 볼 수 있다. 현재 NVIDIA의 주요 기술은 아키텍처(Architectures), 엔터프라이즈 및 개발자(Enterprise & Developer), 게이밍(Gaming), 산업 기술(Industry Technologies)의 4가지로 분류할 수 있으며,²⁹⁾ 자세한 내용은 다음과 같다.
 - 아키텍처 부문: GeForce와 같은 게이밍 GPU, Tesla와 같은 AI 및 데이터 센터 전용 GPU, Quadro와 같은 전문 그래픽 솔루션, 관련 응용 프로그램에 필요한 성능 제공
 - 엔터프라이즈 및 개발자 : CUDA(Computed Unified Device Architecture) 가 핵심 기술로 CUDA는 GPU의 병렬 처리 능력을 활용, 대규모 데이터 처리 와 복잡한 계산을 수행 지원. AI 및 머신러닝 모델 최적화, 이미지 및 영상

27) 한겨레 (24.06.28.), “NVIDIA는 어떻게 ‘대세’가 됐나 [The 5]”, <https://www.hani.co.kr/arti/economy/it/1146889.html> (검색일 24.01.11.)

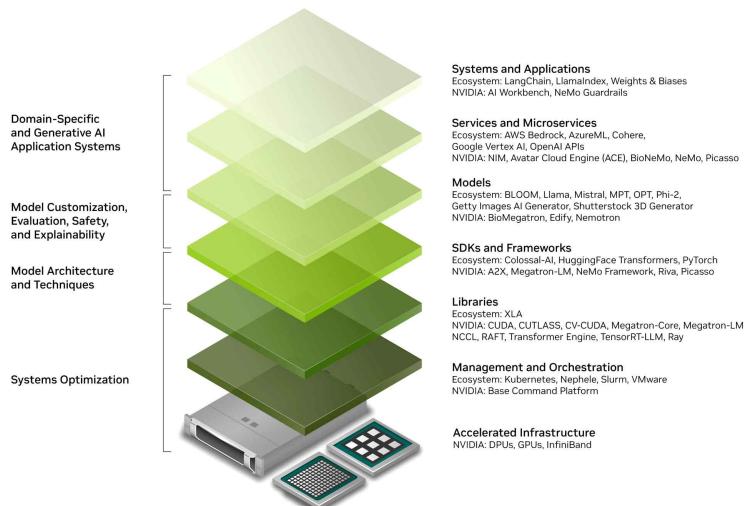
28) 한겨레 (24.06.19.), “MS·애플 제친 NVIDIA 시총 1위…‘인공지능 3파전’ 최후의 승자는?” <https://www.hani.co.kr/arti/economy/global/1145589.html> (검색일 24.01.11.)

29) NVIDIA 홈페이지 “Technologies”, <https://www.NVIDIA.com/en-us/technologies/> (검색일 24.01.11.)

처리, 과학적 시뮬레이션 등의 성능을 극대화 지원.

- 산업기술 부문: 자율주행, 헬스케어, 제조업 등에 NVIDIA DRIVE 및 Omniverse 플랫폼을 통해 실시간 모델링과 분석 솔루션을 제공.
- 특히 Generative AI를 위한 풀 스택 플랫폼을 제고할 수 있는 기술 역량을 갖춘 점이³⁰⁾ AI 하드웨어의 최강자 위치를 공고히 하는데 제일 큰 역할을 하고 있다고 볼 수 있는 점이다.

[그림 2-15] NVIDIA의 Full-Stack Platform

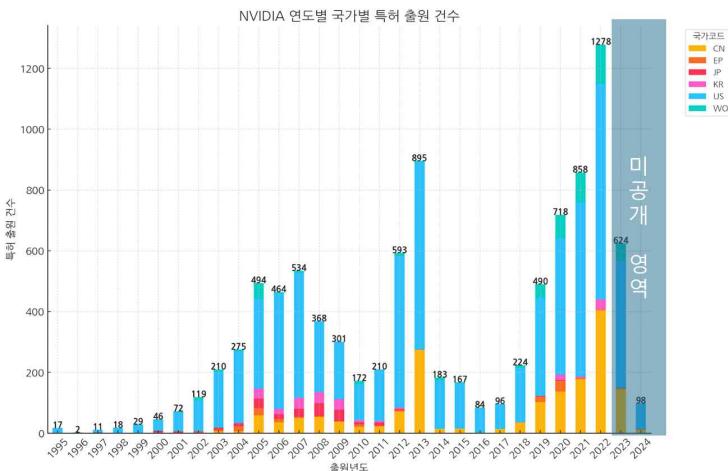


자료: <https://developer.NVIDIA.com/generative-ai> (검색일 25.1.11)

- 인공지능, 딥러닝, 병렬처리 등 첨단 기술 분야에서 다량의 특허를 보유하고 있으며, 2024년 기준 인공지능 및 신경망 관련 특허 1,010건, GPU 및 병렬 처리 분야 관련 특허 850건 등 인공지능 관련 주요 핵심 기술에 다량의 특허를 출원하였다.

30) “NVIDIA Full-Stack Generative AI Software Ecosystem”, <https://developer.NVIDIA.com/generative-ai> (검색일 25.1.11.)

[그림 2-16] NVIDIA의 연도별 특허 출원 동향



자료: 파인특허법률사무소(24.09.30), “2024년 엔비디아(NVIDIA)의 특허 출원 동향 분석”, <https://www.pinepat.com/ko/insights/columns/envidia-nvidia-yi-teugheo-culweon-donghyang-bunseog> (검색일 25.02.06.)

□ 핵심경쟁력과 생태계

- 사업모델을 보면, NVIDIA는 고성능 그래픽 처리 장치(GPU) 개발을 통해 데이터 센터, 게임, 자율주행차 등 다양한 산업에 관련 솔루션 제공하고 있다. 초창기에 게임 부문의 병렬처리에 특화된 GPU 장치의 제공에서 시작하여, 생성형 AI 학습을 위한 데이터 센터 및 클라우드 컴퓨팅 기술 부문을 통해 더욱 고성장 하였다. 특히, 고성능 GPU 기술과 강력한 소프트웨어 생태계³¹⁾(CUDA, cuDNN 등)를 통해 AI 및 딥러닝 작업에 최적화된 솔루션을 제공하여, AI 중심의 전략으로 데이터센터와 클라우드 컴퓨팅 시장에서 강력한 입지를 확보하였다. 최근에는 기존의 게임, 자율주행차 지원 뿐 아니라 ‘바이오니모’(BioNeMo)³²⁾의 출시 등을 통한 AI-powered service (바이오 분야 생성형 AI 활용) 등 다양한 산업에 AI 기술을 응용하는 영역으로도 사업을 확장하고 있다.

31) NVIDIA, “NVIDIA cuDNN”, <https://developer.nvidia.com/cudnn> (검색일 25.01.13.)

32) NVIDIA, “바이오파마용 Clara”, <https://www.nvidia.com/ko-kr/clara/biopharma/> (검색일 25.01.17.)

[그림 2-17] NVIDIA의 주요 사업 모델



자료: NVIDIA 2024 annual report 4p

○ 주요 거래 기업

- 최대 고객은 마이크로소프트(MS)로, NVIDIA 매출의 약 15%를 차지, 메타(Meta)가 약 13%로 그 뒤를 잇고 있으며, 아마존(Amazon) 약 6%, 구글(Google) 약 6%로 이들 네 개의 기업이 NVIDIA 전체 매출의 40% 차지³³⁾하고 있다. 팹리스 제조 전략을 채택하여, 웨이퍼 제작, 조립, 테스트 및 포장 과정에서 TSMC와 삼성전자와 같은 주요 공급업체와 협력하고 있다. 이들 공급업체 보유한 ISO 9001 및 IATF 16949 인증은 반도체 생산 전 과정에 걸친 표준화된 품질 관리를 보장하는 것으로, NVIDIA는 이들을 통해 제조비용과 위험을 줄이고, 제품 설계 및 품질 보증에 집중할 수 있다.³⁴⁾ 메모리는 마이크론, SK 하이닉스, 삼성으로부터 구매하며, CoWoS 기술을 활용한 패키징을 수행. Hon Hai Precision, Wistron, Fabrinet과 같은 계약 제조업체를 통해 최종 제품의 조립 및 테스트를 진행하고 있다.

33) AI 타임즈 (24.06.10), “NVIDIA 최대 고객은 MS...상위 4곳이 GPU 40% 구매” , <https://www.aitimes.com/news/articleView.html?idxno=160226> (검색일 25.01.17.)

34) NVIDIA(25.01.06), “NVIDIA DRIVE Hyperion Platform Achieves Critical Automotive Safety and Cybersecurity Milestones for AV Development”, <https://nvidianews.nvidia.com/news/nvidia-drive-hyperion-platform-achieves-critical-automotive-safety-and-cybersecurity-milestones-for-av-development> (검색일 25.01.11.)

○ 투자

- NVIDIA는 설립 초기 세콰이어 캐피털(Sequoia Capital) 등 주요 벤처캐피털로부터 Series A 및 B 투자를 유치하여 초기 자금을 조달하며 여러 라운드의 투자를 통해 성장해왔다. 최근에는 외부 투자 보다는 자사주 매입 등의 방식으로 재투자를 진행하고 있으며, DARPA, ARPA-E등의 미국의 혁신 부문 정부 지원 (GRANT)을 받아 연구를 수행하며, 사업을 진행하는 특징도 보이고 있다.

[그림 2-18] NVIDIA의 주요 사업 모델

funding-rounds-2024-10-27

Transaction Name	Funding Type	Money Raised	Money Raised Currency	Money Raised (in USD)	Announced Date	Investor Names
Seed Round - NVIDIA	Seed				1993-01-01	Sequoia Capital
Post-IPO Debt - NVIDIA	Post-IPO Debt				2009-01-01	Jean Abrial, TriplePoint Capital
Grant - NVIDIA	Grant	25000000	USD	25000000	2010-08-09	DARPA
Post-IPO Equity - NVIDIA	Post-IPO Equity	400000000	USD	400000000	2017-05-24	SoftBank Vision Fund
Post-IPO Equity - NVIDIA	Post-IPO Equity				2021-02-18	Grand 365 Capital
Post-IPO Equity - NVIDIA	Post-IPO Equity	65000000	USD	65000000	2022-08-09	ARK Investment Management
Grant - NVIDIA	Grant	5000000	USD	5000000	2023-05-09	ARPA-E

자료: crunchbase (NVIDIA_funding rounds)

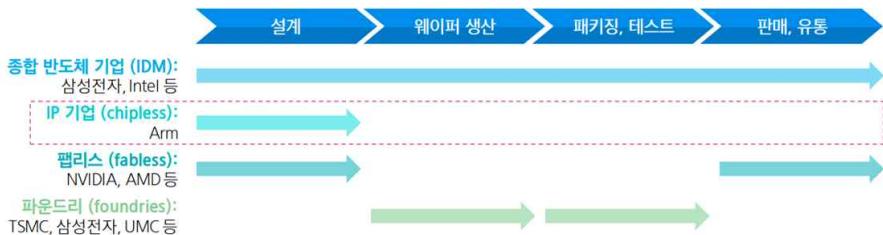
나. ARM

□ 기업 개요

- ARM은 1990년 영국에서 창립되었으며 반도체 설계에 필요한 IP 코어 및 블록 제공 기업이며, 현재 AP(Application Processor) 부문 글로벌 선도 IP 기업이다. 1990년 ARM(Advanced RISC Machines)은 Acorn Computer, Apple, VLSI Technology 세 회사의 조인트 벤처(협작회사)로 탄생하였다. ARM 홀딩스의 전신 기업인 영국 컴퓨터 제조사 에이콘(Acorn)은 1978년에 설립되었고, 1990년 저 전력 CPU 개발과 함께 RISC의 가능성을 눈여겨보던 Apple이 VLSI를 포함하여 에이콘과 조인트 벤처를 만들면서 ARM(Acorn RISC Machines)가 탄생하였다. 1998년, 회사명에서 'Acorn'을 'Advanced'로 변경하고 런던과 나스닥에 동시 상장하면서 독립적인 기업으로 활동을 시작하였다. 반도체를 디자인하고 설계하는 팜리스에 속한다고 볼 수 있지만, ARM 홀딩스는 기업 자체의 반도체 칩을 개발하는 기존의 팜리스 기업과 달리, 코어 등 주요 IP 기술 개발만을 전문으로 하고 있다.

[그림 2-19] 사업 구조별 반도체 기업 분류

사업 구조별 반도체 기업 분류



자료: 삼성반도체이야기, 삼성증권

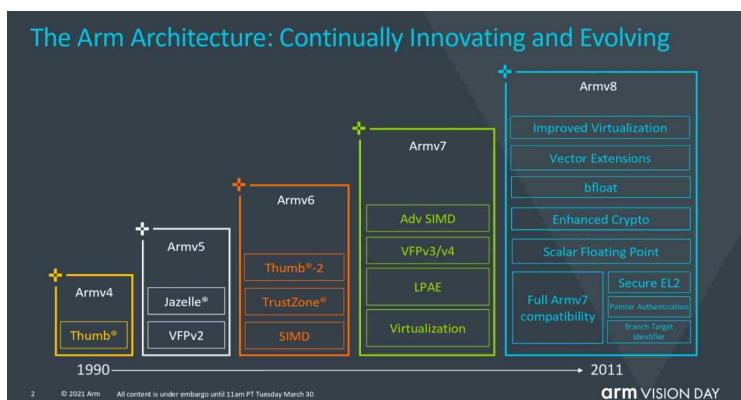
자료: 삼성증권, “Arm holdings, 무한한 공간 저 너머로?” 2023.09.12

 기술

○ 주력 제품

- Arm 아키텍처: ARM 프로세서를 설계하는 디자인으로, 프로그래머 모델을 의미, 임베디드 기기에서 많이 사용되는 RISC 프로세서로 저전력으로 사용할 수 있다는 장점이 있음

[그림 2-20] ARM Architecture

자료: Cadence (21.04.28.), “Arm V9A”, https://community.cadence.com/cadence_blogs_8/b/breakfast-bytes/posts/arm-v9a (검색일 25.01.11.)

- Armv9: ·AI 수요 증가로 등장한 Armv9는 기존의 Armv8 대비 높은 CPU 연산 능력을 보유하고 있으며, 슈퍼컴퓨터용 명령어 세트인 SVE2를 지원하는 Armv9는 기존보다 빠른 머신러닝(ML)과 DSP 성능을 구현³⁵⁾
- 유망 제품으로는 Arm AE IP: 2024년 3월, Arm은 AI 지원 차량을 위한 Arm AE(Automotive Enhanced) IP 5종을 출시하였고, 자율주행, SDV, ADAS 등 개별 임무 별 연산량 고려하여 IP를 다각화 하여 출시
- 라이센스 비용 및 로열티를 받는 형태가 주요한 사업 모델임
 - 반도체 업체에게 IP를 제공 후 라이센스 비용으로 수익 창출, Arm IP 기반으로 개발된 반도체의 생산과 패키징은 각각 파운드리와 반도체 패키징 및 테스트 (OSAT) 수탁기업에 위탁. 이후 완성된 최종 제품을 모바일 OEM 업체들에게 판매하며 수익을 창출하고 있으며, 제품에 대한 1~2% 가량의 로열티를 제공받음.

[그림 2-21] ARM의 라이센스 및 로열티 수취 모델

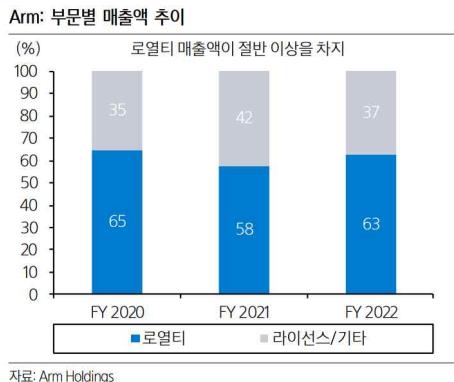


자료: KIPOST(2019.07.22.), “Arm에게 오픈소스 아키텍처란?”, <https://www.kipost.net/news/articleView.htm?idxno=201469> (검색일 25.01.11.)

- 특징점으로는 시장별, 제품별 수익률에 차별이 존재하고 있으며, PC/서버 시장은 모바일보다 높은 가격대 형성으로 인해 로열티 매출이 높음

35) Zdnet(21.03.31.), “ARM, 차세대 아키텍처 ‘Armv9’ 공개”, <https://zdnet.co.kr/view/?no=20210331194029> (검색일 25.01.11.)

[그림 2-22] ARM의 라이센스 및 로열티 수취 모델



자료: 인포스탁데일리 “Arm홀딩스(ARM), 나스닥 상장 임박...업황 회복의 조짐 확인한 이후 접근 고려”, <https://www.infostockdaily.co.kr/news/articleView.html?idxno=194512>

□ 시장

- 모바일 AP 시장에서 ARM은 선도적 위치를 지니고 있음. 모바일 AP 시장 CPU 점유율 99% 이상. 전력 효율성이 높은 RISC 방식으로 경쟁 우위를 창출하고 있음. 하지만, 성장률 관점으로는 숙제가 존재함. 향후 3년간 연평균 4% 성장에 그칠 것 예상되어 미래의 유망 시장으로 판단하기에는 어려우나, Qualcomm등 주요 기업의 ARM 의존도는 매우 높으므로, 단기간에 ARM을 대체할 ARM IP 사용기업이 없음.
- PC/서버 시장에서는 선도적 우위를 가지고 있지는 않음. 전통적인 강자인 인텔과 그 뒤를 추격하는 AMD가 주도권을 쥐고 있음. 이는 AMD가 ARM 아키텍처가 RISC 방식인 데 비하여 PC 시장에서는 높은 성능과 호환성을 수반하고 있는 CISC 방식이 주도권을 가지고 있기 때문임. 애플의 MAC이 기존의 인텔 칩에서 ARM 기반의 자체 프로세서로 대체하였고, 이를 바탕으로 PC 시장의 점유율을 10% 내외로 끌어올릴 기회를 얻을 수 있었음³⁶⁾. 여기서 중요한 점은 제품의 판

36) TECHWIRE Asia, “Apple vs Microsoft as new M2 chips look to dent Windows market share”, <https://techwireasia.com/2022/06/apples-m2-chips-may-just-dent-microsofts-windows-market-share/> (검색일: 25.01.17.)

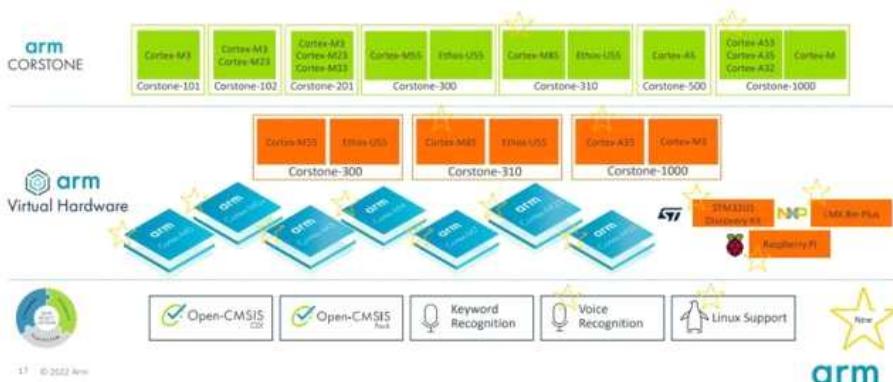
매 이후 마진을 고려할 때, 최신 스마트폰 AP 가격대가 40~160달러인데 비하여, 데스크톱 및 서버 CPU 가격대는 400~10,000달러까지로 높은 가격대를 형성하고 있으므로, ARM도 수익률이 높은 PC/서버시장의 점유율 향상을 위하여 노력³⁷⁾.

- IoT/임베디드 시장을 보면, 가정 및 산업 현장의 각종 기기들이 인터넷 접속 기능 제공 및 데이터를 생성하면서 전송하고 이를 연결하여 활용하는 IoT 시장이 기존보다 빠른 속도로 성장하고 있음. 높은 성능을 요구하지 않고 전력 효율성 및 빠른 처리를 요구하고 있기 때문에 ARM 아키텍처의 장점이 작용할 수 있는 시장으로 볼 수 있음. 이에 따라 IoT 시장에서 64.5% 점유율 가지고 있으며, 2021년에는 IoT 생태계 활성화를 위해 ARM IoT 토탈 솔루션 발표.³⁸⁾

[그림 2-23] ARM의 IoT 토탈 솔루션

An Expanding Library of Technology for Building IoT Solutions

Addressing developers' needs for more choice, security, performance and flexibility



자료: 동아일보(22.05.13.), “IoT 시장의 ‘실세’ Arm, 개발 생태계 강화까지 이끈다”, <https://www.donga.com/news/lt/article/all/20220513/113387175/1> (검색일: 25.01.17.)

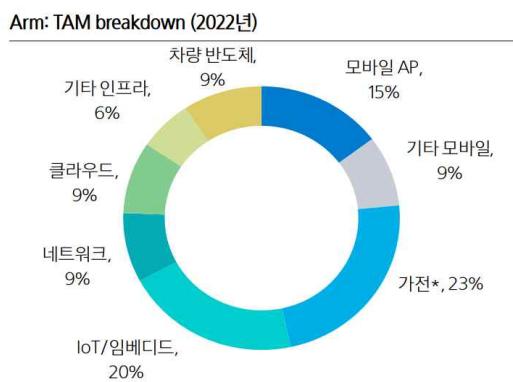
37) Forbes(24.10.15.). “AMD And Intel Unite Against Existential Threat Of Arm”, <https://www.forbes.com/sites/stevemcdowell/2024/10/15/amd-and-intel-unite-against-common-existential-threat/> (검색일: 25.01.17.)

38) 동아일보(22.05.13.), “IoT 시장의 ‘실세’ Arm, 개발 생태계 강화까지 이끈다”, <https://www.donga.com/news/lt/article/all/20220513/113387175/1> (검색일: 25.01.17.)

□ 핵심경쟁력과 생태계

- 모바일 시장의 강자로서, ARM은 RISC 방식 기반의 ARM 아키텍처를 통하여 CISC 방식보다 뛰어난 전력 효율성 구축³⁹⁾하였음. 스마트폰의 발달로, RISC 방식이 더욱 선호되고 있으며, ARM은 모바일 시장에서 독보적인 위치를 만들었고, RISC 방식에 대한 선점과 이어지는 기술 개발로 후발 주자에게 시장의 진입장벽을 형성하였음⁴⁰⁾.
- 사업 다각화 노력 측면에서, ARM은 모바일 AP 시장을 벗어나 미래 성장 가능성이 높은 여러 시장을 노린 제품 다각화, 차량 반도체, 클라우드, 네트워크, AR/VR 영향력을 확장하고 있음⁴¹⁾. ARM의 주력 시장은 모바일(모바일 AP, 기타 모바일, 가전 포함) 47%, IoT/임베디드 20%. 로열티 매출액이 중요하기 때문에 고객사의 반도체 출하량에 영향을 많이 받는 구조를 지니는 한계가 있음.

[그림 2-24] ARM의 주요 사업



참고: * 태블릿 등 스마트폰 외의 모바일 제품들도 포함

자료: Arm Holdings

자료: 인포스탁데일리 “Arm홀딩스(ARM), 나스닥 상장 임박...업황 회복의 조짐 확인한 이후 접근 고려”, <https://www.infostockdaily.co.kr/news/articleView.html?idxno=194512>

39) ARM, “What Is RISC?”, <https://www.arm.com/glossary/risc>

40) CNBC(23.11.09.), “How Arm is gaining chip dominance with its architecture in Apple, Nvidia, AMD, Amazon, Qualcomm and more”, <https://www.cnbc.com/2023/11/09/how-arm-gained-chip-dominance-with-apple-nvidia-amazon-and-qualcomm.html> (검색일 24.11.28.)

41) Futurum,(24.11.13.), “Research Note-Arm Q2 FY 2025 Earnings: AI & Data Center Growth Drive Revenue” <https://futurumgroup.com/insights/arm-q2-fy-2025-earnings-ai-data-center-growth-drive-revenue/> (검색일 24.12.13.)

다. Cerebras (세레브라스)

□ 기업 개요

- 2015년에 설립된 AI 하드웨어 스타트업. 딥러닝 애플리케이션에 맞춘 혁신적인 하드웨어 솔루션 통해 AI 처리에 혁명을 일으키려는 비전으로 설립⁴²⁾. Cerebras는 AI 서비스에 특화된 반도체를 개발 중. 2024년 8월, CS-3 칩을 기반으로 하는 최신 AI 프로세서를 공개. NVIDIA 호퍼 칩보다 AI 추론 속도가 20배 빠른데 가격은 낮다고 설명⁴³⁾하고 있으며, 2024년 9월, Cerebras는 나스닥 상장을 위해 미국 증권거래위원회에 증권신고서(S-1) 제출.⁴⁴⁾

□ 기술

- 세계 최대 규모의 웨이퍼 스케일 엔진(Wafer-Scale Engine; WSE)을 주력으로 하여 GPU 클러스터의 복잡성을 감소하고 대규모 AI 모델에 대한 훈련을 단일 칩으로 수행할 수 있도록 함.
- CS-3은 초기대(3세대) 웨이퍼 스케일 AI 가속기(Wafer-scale AI accelerator)로 AI 모델을 트레이닝할 수 있도록 설계되어있음. 4조개의 트랜지스터는 동급 최대 기종의 57배이며, 이전 모델에 비해 2배 빠른 성능을 제공. 복잡성이 높고 다량의 연산의 매우 효율적으로 수행할 수 있음.
- CS-3를 기반으로 한 AI 슈퍼컴퓨터⁴⁵⁾는 최대 1.2PB의 거대한 메모리 시스템을 갖추고 있으며, CS-3는 GPT-4 및 Gemini보다 10배 더 큰 차세대 프론티어

42) Myscale(24.04.17.), “Revolutionizing AI: A Peek into Cerebras Systems’ Wafer-Scale Technology”, <https://myscale.com/blog/revolutionizing-ai-cerebras-wafer-scale-technology/?t> (검색일 25.01.11.)

43) 한경(24.10.01.), “AI칩 샛별’ 세레브라스, 나스닥 도전장”, <https://www.hankyung.com/article/2024100190231>(검색일 25.01.11.)

44) AI 타임즈 (24.10.02.), “세레브라스, 엔비디아 아성에 본격 도전장?… 세계에서 가장 큰 AI 칩셋으로 나스닥 상장 한다!” <https://www.aitimes.kr/news/articleView.html?idxno=32349> (검색일 24.11.11.)

45) Cerebras (24.03.11.), “Cerebras Systems Unveils World’s Fastest AI Chip with Whopping 4 Trillion Transistors” <https://cerebras.ai/press-release/cerebras-announces-third-generation-wafer-scale-engine?t> (검색일 24.11.11.)

모델 훈련하도록 설계됨. 24조 개의 파라미터 모델을 하나의 로직에 파티션이나 리팩토링 없이 저장할 수 있어, 모델 훈련 과정을 극적으로 단순화할 수 있으며 개발자의 생산성 향상 가능. CS-3은 기업과 대규모 사용 모두 적합하도록 설계. 4개의 시스템 설정으로 70B 모델들을 하루에 미세하게 조정할 수 있으며, 최대 2048 시스템을 모두 이용할 경우 Llama 70B를 하루 만에 처음부터 훈련 가능 함. CS-3는 사용의 용이성을 제공하며, LLM에 비해 97% 적은 코드가 필요하고, 데이터 병렬 모드에서는 1B에서 24T개의 변수 모델 학습 가능. Cerebras에서 GPT-3 크기의 모델 또한 565줄의 코드로 구현 가능

□ 핵심경쟁력과 생태계

○ 사업모델

- Wafer-Scale Engine(WSE) 칩 설계와 제조가 사업모델이며 최근 뛰어난 성능을 제공하는 CS-3칩 생산 및 AI 슈퍼컴퓨터 생산.

○ 주요 협력 기업

- Mayo Clinic과 Cerebras Systems는 대규모 언어 모델(LLM)을 개발해 의료 분야에서 환자 진단 및 치료 계획 개선 위한 다년간의 전략적 협력 체결하였고, 환자 기록, 유전자 데이터, 약물 분자를 활용한 류마티스 관절염 진단 모델을 포함한 다양한 질병 상태에 대한 모델 개발 예정. 이는 의료진이 환자 데이터를 보다 효율적으로 분석하고 최적의 치료 방법을 제안하는 데 도움 줄 것이라 기대되는 부분임⁴⁶⁾⁴⁷⁾
- GSK와 협력하여 신약 개발 및 생명과학 연구를 지원, 인간의 에피게놈을 모델링할 수 있는 EBERT 모델을 개발.⁴⁸⁾

46) Cerebras(24.01.15.) “Cerebras Collaborates with Mayo Clinic to Advance AI in Healthcare” <https://cerebras.ai/press-release/cerebras-collaborates-with-mayo-clinic-to-advance-ai-in-healthcare> (검색일 24.11.11.)

47) Data Centre Dynamics(24.01.09.), “Data Centre Dynamics”, <https://www.datacenterdynamics.com/en/news/mayo-clinic-to-use-cerebras-chips-for-medical-ai-models> (검색일 24.11.11.)

48) HPCwire(22.01.28.), “Cerebras, GSK Team for Unprecedented Epigenomic Models”. <https://www.hpcwire.com/2022/01/28/cerebras-gsk-team-for-unprecedented-epigenomic-models> (검색일 24.11.11.)

- GSK 연구팀은 CS-1 시스템을 사용하여 EBERT 모델을 2.5일 만에 훈련시켰으며, 이는 16노드 GPU 클러스터로 24일이 걸릴 것으로 추정되는 작업. EBERT 모델은 ENCODE-DREAM이라는 업계 벤치마크의 13개 데이터셋 중 4개에서 가장 높은 예측 정확도를 달성한 바 있음.⁴⁹⁾
- Cerebras의 주요 고객 중 하나는 아부다비의 AI 기업인 G42. G42는 대규모 언어 모델, 컴퓨터 비전, 헬스케어 등 다양한 분야에서 AI 기술 개발 중이며, Cerebras의 AI 칩 활용해 연구 개발 가속화 중. 또한 Cerebras는 중동 지역에서 AI 인프라를 확장하고 Condor Galaxy AI 슈퍼컴퓨터 네트워크 구축 위해 G42와 파트너십 체결. G42는 2024년 상반기에만 세레브라스 매출의 87%를 차지했으며, 현재 세레브라스 제품을 대규모로 구매 중.⁵⁰⁾

□ 투자

○ 매출

<표 2-7> 세레브라스 매출

년도	매출
2022	\$25M
2023	\$79M
2024	\$206M

자료: SACRA. <https://sacra.com/c/cerebras-systems/> (검색일 25.01.15)

- 2024년 상반기 동안 Cerebras는 1억 3640만 달러의 매출 기록. 이는 2023년 상반기 매출 870만 달러에 비해 크게 증가한 수치이며, Cerebras의 성장 가능성 보여줌.⁵¹⁾

49) Cerebras(22.01.26.), “GlaxoSmithKline and Cerebras are Advancing the State of the Art in AI for Drug Discovery”, <https://cerebras.ai/chip/glaxosmithkline-and-cerebras-are-advancing-the-state-of-the-art-in-ai-for-drug-discovery/> (검색일 24.11.11.)

50) Cerebras(24.09.11), “Aramco and Cerebras Sign a MoU to Accelerate the Deployment of AI”, <https://cerebras.ai/press-release/aramco-and-cerebras-sign-a-mou-to-accelerate-the-deployment-of-ai> (검색일 24.12.18.)

51) AI 타임즈 (24.10.02.), “세레브라스, 엔비디아 이성에 본격 도전장?… 세계에서 가장 큰 AI 칩셋으로 나스닥 상장한다!” <https://www.aitimes.kr/news/articleView.html?idxno=32349> (검색일 24.11.11.)

○ 투자

- 총 투자 유치 금액: \$1.12B (2024년 기준), 24년 9월 Series F로 \$254.38M 유치⁵²⁾
- 주요 투자자: Abu Dhabi Growth Fund, Altimeter Capital Management, Alpha Wave Ventures, Moore Strategic Ventures 등

라. Celestial AI

□ 기업 개요

- Celestial AI는 2020년 설립된 AI 기술 회사로서 David Lazovsky가 설립. Celestial AI는 AI 컴퓨팅 인프라를 위한 광학 상호 연결 기술 플랫폼을 개발 중⁵³⁾. 2022년 56백만 달러 규모의 시리즈 A 투자 유치에 성공했으며 2024년 175백만 달러의 시리즈 C 투자 유치⁵⁴⁾.

□ 기술

○ Photonic Fabric™ 기술⁵⁵⁾

- 칩 내부와 칩 간의 데이터 전송에 빛을 사용하는 독자적인 포토닉 패브릭 (Photonic Fabric™) 기술 플랫폼을 기반으로 한 새로운 프로세싱 시스템으로 컴퓨팅 방식의 근본적 전환을 도모하고 있는 혁신적 기술을 제시하고 있으며, 이는 패키지 비치프론트의 제약 없이 컴퓨팅 다이의 어느 위치로든 직접 데이터 전달 가능. 기존 광학 상호 연결 대비 25배 높은 대역폭 제공, 지연 시간과 전력 소비는 10배 이상 저감. 대규모 AI 모델에 필요한 메모리 대역폭과 용량을 저 지연

52) Forge Investment Outlook Q1 2025, https://forgeglobal.com/cerebras_stock/#financingsAndValuations

53) businesswire(23.06.28.), “Celestial AI, the Creator of the Photonic Fabric Optical Interconnect Technology Platform, Raises \$100 Million in Series B Funding”, <https://www.businesswire.com/news/home/20230627995436/en/Celestial-AI-the-Creator-of-the-Photonic-Fabric-Optical-Interconnect-Technology-Platform-Raises-100-Million-in-Series-B-Funding> (검색일 24.12.18.)

54) Semiconductor Today(24.04.02.), “Celestial AI closes \$175m Series C funding round”, https://www.semiconductor-today.com/news_items/2024/apr/celestialai-020424.shtml (검색일 24.12.18.)

55) Laser Focus World(23.07.24.), “Q&A with Celestial AI's Dave Lazovsky”, <https://www.laserfocusworld.com/optics/article/14296728/qa-with-celestial-ais-dave-lazovsky> (검색일 24.12.18.)

으로 제공, 데이터를 컴퓨팅 지점으로 직접 전달하여 “Memory Wall” 문제 해결⁵⁶⁾

□ 핵심경쟁력과 생태계

○ 사업모델

- 대형 대규모데이터센터 운영 기업들과 협력하여 시스템 인프라의 병목 현상을 개선하기 위해 노력하였으며, 그 결과 빛을 기반으로 데이터 전송을 수행하는 Photonic Fabric 기술 개발⁵⁷⁾.

○ 주요 거래 기업

- 삼성전자와 HBM 메모리, DDR, 고급 패키징 솔루션 분야 등에서 협력 중. 특히 Photonic Fabric 기술을 통해 AI 학습 및 추론을 위한 나노 초 수준의 지연 시간을 제공하기 위해 노력⁵⁸⁾ 중.
- Broadcom 등 주요 기업들과 파트너십을 구축하여 Photonic Fabric 생태계 육성을 도모⁵⁹⁾

□ 투자

○ 투자사⁶⁰⁾

- 시리즈 C 라운드 주도 투자자: Thomas Tull의 US Innovative Technology

56) 뉴스와이어(22.02.04.), “셀레스티얼 AI, 5600만달러 조달… 새로운 광전자 기술 플랫폼으로 AI 칩셋 산업 혁신”, <https://www.newswire.co.kr/newsRead.php?no=938919> (검색일 24.12.18.)

57) Design&Reuse(23.06.29.), “Celestial AI, the Creator of the Photonic Fabric Optical Interconnect Technology Platform, Raises \$100 Million in Series B Funding”, <https://www.design-reuse.com/news/54347/celestial-ai-series-b-funding-photonic-fabric-optical-interconnect.html> (검색일 24.08.11.)

58) BusinessWire(24.03.25.), “Celestial AI Announces Photonic Fabric™ Adoption by Lead Hyperscaler Customers”, <https://www.businesswire.com/news/home/20240325010932/en/Celestial-AI-Announces-Photonic-Fabric%E2%84%A2-Adoption-by-Lead-Hyperscaler-Customers> (검색일 24.08.11.)

59) theNESTPLATFORM(23.06.28.), “Photonics To Make Celestial HBM3 Memory Fabric”, <https://www.nextplatform.com/2023/06/28/photonics-to-make-celestial-hbm3-memory-fabric/> (검색일 24.08.11.)

60) Celestial AI(24.03.27.), “Celestial AI Closes \$175 Million Series C Funding Round Led by U.S. Innovative Technology Fund”, <https://www.celestial.ai/blog/celestial-ai-closes-175-million-series-c-funding-round-led-by-us-innovative-technology-fund> (검색일 24.08.11.)

Fund (USIT), 시리즈 C 라운드 참여 투자자: AMD Ventures, Koch Disruptive Technologies (KDT), Temasek, Samsung Catalyst, Smart Global Holdings (SGH), Porsche Automobil Holding SE 등

- 다양한 산업 분야의 주요 기업들과 벤처 캐피털 펀드들이 투자 참여. 특히 AMD, 삼성, Porsche와 같은 글로벌 기업들의 참여는 Celestial AI의 기술에 대한 업계의 높은 관심을 반영하고 있음. Koch Industries와 Temasek 같은 대형 투자사의 지속적인 참여는 회사의 성장 잠재력에 대한 신뢰라고 해석 가능. Celestial AI의 총 투자 유치 금액은 \$331 million (약 4,270억 원). 특히 시리즈 C 라운드가 “highly oversubscribed”(초과 청약)됨을 통해 Celestial AI에 대한 시장의 관심도 유추 가능.

<표 2-8> Celestial AI의 투자유치

년도	라운드	금액
2022	시리즈 A 라운드	\$56 million (약 560억원)
2023.06	시리즈 B 라운드	\$100 million (약 1,350억원)
2024.03	시리즈 C 라운드	\$175 million (약 2,360억원)

자료: "Celestial AI closes \$175m Series C funding round", semiconductor-today, (24.04.02.), https://www.semiconductor-today.com/news_items/2024/apr/celestialai-020424.shtml , "Celestial AI Raises \$56 Million Series A to Disrupt the Artificial Intelligence Chipset Industry with Novel Photonic-Electronic Technology Platform", (22.02.04.), <https://financialpost.com/pmn/press-releases-pmn/business-wire-news-releases-pmn/celestial-ai-raises-56-million-series-a-to-disrupt-the-artificial-intelligence-chipset-industry-with-novel-photonic-electronic-technology-platform> (검색일 24.08.11.)

마. Hailo

□ 기업 개요

- Hailo는 이스라엘의 엣지 인공지능(AI) 반도체 스타트업
 - AI 프로세서와 AI 가속기 설계 및 제조, 주요 적용 분야는 자율주행 차량, 보안 카메라, 자율 이동 로봇 분야임⁶¹⁾. 북미, 유럽, 아시아 등 여러 곳에 지사를 두고 있으며 2021년 시리즈 C 투자 유치로 기업 가치 10억 달러의 유니콘 기업 등극⁶²⁾

61) CBInsight, <https://www.cbinsights.com/company/hailo-1> 기업소개

□ 기술

○ 주요 기술

- Hailo-8 AI 프로세서⁶²⁾: 딥 러닝 워크로드에 최적화된 엣지 AI 칩, 초당 최대 26테라 연산(26 TOPS) 처리 가능. 2.5W의 낮은 전력 소비. 1페니 동전보다 작은 크기를 가지고 있음. GPU 대비 성능 최대 2배 향상, 가격 대비 성능 19배 향상. ResNet-50 실행 시 NVIDIA Xavier AGX 대비 20배 낮은 전력 소비의 장점을 지니고 있음. 구조 중심 데이터 흐름 아키텍처(structure-defined data flow architecture), 분산된 온 칩 메모리 패브릭, 효율적인 상호 연결 및 제어 방식 구현하여, 데이터를 로컬에서 처리하는 장점이 있고, 이를 통한 데이터 프라이버시 강화의 장점
- 생성 AI 워크로드를 엣지에서 직접 실행 가능한 Hailo-10도 개발, 5W 미만 전력으로 초당 최대 10개 토큰으로 LLaMA2-7B 실행 가능. 최대 40 TOPS 성능 제공.
- Hailo-15는 스마트 카메라를 위한 AI 비전 프로세서 제품으로 다각화의 일환으로 판단 가능

□ 핵심경쟁력과 생태계

- 사업모델로서 AI 프로세서의 성능과 효율성을 높이면서도 비용을 낮춰 엣지 AI 시장 경쟁력 확보 전략을 활용. 최근 출시한 Hailo-10을 통해 생성 AI 워크로드 까지 엣지 디바이스에서 처리할 수 있게 함으로써 AI 애플리케이션 경쟁력 확보
- 주요 거래 기업

- J-Squared Technologies(북미)와 2024년 3월, 북미 지역 유통 파트너십 체결. 미국과 캐나다에서 헤일로의 엣지 AI 기술 유통을 담당, 스마트 리테일, 스마트

62) TheJerusalempost(25.01.04.), “Hailo: The Tel Aviv unicorn company set to conquer Vegas”, <https://www.jpost.com/consumerism/consumer-news/article-834421> (검색일 25.01.11.)

63) Things Embedded Limited, “Benefits of HAILO-8 AI Accelerators for Edge Computing”, <https://things-embedded.com/uk/white-paper/benefits-of-hailo-8-ai-accelerators-for-edge-computing/> (검색일 25.01.11.)

시티, 스마트 홈, 군사, 항공우주, 산업 4.0 등 다양한 분야에 제품 공급 목표⁶⁴⁾

- AITg(WPG Holdings의 자회사, 대만 및 중국)과 2022년 3월 대만 및 중국 지역 유통 파트너십 체결, 스마트 리테일, 스마트 시티, 스마트 홈, 산업 4.0 등의 분야에 제품 공급 목표⁶⁵⁾
- IMDT(이스라엘)과 AI 비전 솔루션 개발을 위한 파트너십 체결. MDT의 단일 보드 컴퓨터(SBC)에 Hailo-8 AI 가속 모듈 통합하며 로봇공학, 자율 주행 차량, 스마트 시티, 보안 시스템 등의 분야에 적용 예정.⁶⁶⁾

투자

투자내역

투자 라운드	연도	금액
시드	2017.06	\$3.5M
시리즈 A	2018.06-2019.01	\$21M
시리즈 B	2020.03	\$60M
시리즈 C	2021.1	\$136M
시리즈 C 확장	2024.04	\$120M

출처: Crunchbase, https://www.crunchbase.com/organization/hailo-technologies/company_financials

- 이스라엘 국내 투자자와 글로벌 투자자가 균형있게 참여⁶⁷⁾

64) HAILO AI,(24.03.04.) “Hailo Strengthens Global Reach with Addition of Two New Distribution Partners in America” <https://hailo.ai/company-overview/newsroom/news/hailo-strengthens-global-reach-with-addition-of-two-new-distribution-partners-in-america/?t> (검색일 25.01.11.)

65) PR Newswire(22.03.02.), “Leading AI Chipmaker Hailo Partners with AITg of WPG Holdings to Support Growing Customer Base in Asia Pacific”<https://www.prnewswire.com/news-releases/leading-ai-chipmaker-hailo-partners-with-aitg-of-wpg-holdings-to-support-growing-customer-base-in-asia-pacific-301493756.html?t> (검색일 25.01.11.)

66) IMDT(24.04.17.), “IMDT and Hailo Join Forces to Introduce an Edge AI Solution for Ultimate Performance in Real Time” <https://www.imd-tec.com/news-media/imdt-and-hailo-join-forces-to-introduce-an-edge-ai-solution-for-ultimate-performance-in-real-time/?t> (검색일 25.01.11.)

67) 디지털투데이(24.04.03.), “이스라엘 반도체 기업 헤일로, 1조원 투자 유치 ‘성공적’” <https://www.digitaltoday.co.kr/news/articleView.html?idxno=512355> (검색일 25.01.11.)

바. 퓨리오사

□ 기업 개요

- 퓨리오사AI는 2017년 서울에서 설립된 기업으로 대표는 백준호이다. NPU (Neural Processing Unit) 및 소프트웨어 개발하고 있으며 서버용 NPU를 주 타겟 시장으로 삼고 있음⁶⁸⁾
 - 주요 파트너: 삼성전자, 카카오, Hugging Face, SK hynix, TSMC, ASUS, LG AI Research 등 글로벌 파트너십
 - 제품: AI 연산 최적화된 NPU 칩(WARBOY, RENEGADE 등)

□ 기술

○ 주요 기술

- NPU(Neural Processing Unit): AI 연산을 효율적으로 수행하는 반도체 칩으로, 복잡한 인공지능 모델의 추론을 빠르고 저전력으로 처리⁶⁹⁾.
 - WARBOY (1세대 제품): 삼성 14nm 공정으로 설계된 서버용 AI 추론 가속기, MLPerf 추론에서 NVIDIA T4를 뛰어넘는 성능을 기록.
 - RENEGADE (2세대 제품): TSMC 5nm 공정으로 개발된 고성능 LLM 및 다중 모드 배포를 위한 데이터센터 가속기^{70).}

□ 핵심경쟁력과 생태계

○ 핵심경쟁력

-
- 68) 월간인물(24.03.04.), “[주)퓨리오사에이아이 백준호 대표 – AI의 등장과 함께 몸집 키우는 AI반도체…딥테크 파워로 글로벌 빅테크의 아성에 도전장 내미는 [주)퓨리오사에이아이” <https://monthlypeople.com/news/articleView.html?idxno=700326> (검색일 24.06.28.)
 - 69) IT비즈뉴스(23.02.28.), “[단독] 퓨리오사AI 백준호, “2세대 NPU 내년 상용화, GPU와 전 영역서 경쟁””, <https://www.itbiznews.com/news/articleView.html?idxno=91500> (검색일 24.06.28.)
 - 70) 시사저널e(24.05.07.), “퓨리오사AI, 7월부터 2세대 칩 시제품 공급…대기업 품질 평가”, <https://www.sisajournal-e.com/news/articleView.html?idxno=402602> (검색일 24.06.28.)

- 기술력 부분에서는 1세대 AI 반도체 ‘워보이’가 NVIDIA 대비 이미지 분류, 객체 검출 처리속도에서 1.5배 우수한 성능을 보임⁷¹⁾. 2세대 AI 반도체 ‘RNGD’는 150W의 낮은 전력으로 고성능 LLM 및 멀티모달 모델 구현이 가능⁷²⁾함.
- 소프트웨어 경쟁력 측면에서는 하드웨어와 소프트웨어를 동시에 개발하여 최적화된 성능 구현⁷³⁾하는 동시에, 컴파일러를 통해 다양한 AI 모델에 대응 가능한 소프트웨어 스택을 보유⁷⁴⁾하고 있음.
- 생태계 측면에서 업스테이지⁷⁵⁾, LG AI 연구원⁷⁶⁾ 등 AI 소프트웨어 기업들과의 파트너십을 통한 생태계 확장 및 수요처의 적극적 발굴이라는 강점이 있음

사. 리밸리온

□ 기업 개요

- GPU 대비 에너지 효율성이 높고 전력 소모가 낮은 제품을 통해 데이터센터 및 클라우드 기업의 자속 가능성을 지원하고 있음. 2023년 포브스 선정 ‘가장 빠르게 성장하는 한국 스타트업’, CB인사이트 선정 ‘글로벌 100대 AI 스타트업’에 한국 유일 포함⁷⁷⁾되었음. 2024년 8월 SK텔레콤과 사피온코리아와의 합병 본

71) ZDnet(22.12.22.), “데이터센터향 AI반도체 글로벌 리더 꿈꾼다”, <https://zdnet.co.kr/view/?no=20221222084802> (검색일 24.07.08.)

72) 퓨리오사 AI 레나게이드 소개, <https://furiosa.ai/rngd> (검색일 24.07.08.)

73) IT조선(23.03.23.), “[AI 2023] 퓨리오사AI 김한준 CTO “AI반도체 경쟁력은 SW에 달렸다”” <https://it.chosun.com/news/articleView.html?idxno=2023032201124> (검색일 24.07.08.)

74) 디일렉(23.07.12.), “[딥테크포럼] 퓨리오사AI, “국내 AI 반도체 경쟁력 기우려면 자체 소프트웨어 개발 필요”, <https://www.thelec.kr/news/articleView.html?idxno=21993> (검색일 24.07.08.)

75) 업스테이지 (22.05.30.), “업스테이지-퓨리오사AI, AI 생태계 확장 위해 맞손” <https://www.upstage.ai/news/furiosa-ai-ecosystem> (검색일 24.07.08.)

76) 뉴스핌(23.06.07.), “LG AI 연구원, 퓨리오사AI와 AI 반도체 연구 생태계 구축”, <https://www.newspim.com/news/view/20230607000133> (검색일 24.07.19.)

77) Zdnet(24.08.18.), “기업가치 1조 AI반도체”…사피온-리밸리온 합병 본계약 체결” <https://zdnet.co.kr/view/?no=20240818090428#:~:text=SK%ED%85%94%EB%A0%88%EC%BD%A4%EA%B3%BC%20AI%EB%B0%98%EB%8F%84%EC%B2%B4%20%EC%8A%A4%ED%83%80%ED%8A%B8%EC%97%85%20%EB%A6%AC%EB%B2%A8%EB%A6%AC%EC%98%A8%EC%9D%B4%20SK%ED%85%94%EB%A0%88%EC%BD%A4%20%EA%B3%84%EC%97%4B%EC%82%AC%20%EC%82%AC%ED%94%BC%EC%98%A8%EC%BD%94%EB%A6%AC%EC%95%84%EC%99%80%20%EB%A6%AC%EB%B2%A8%EB%A6%AC%EC%98%A8,%ED%95%A9%EB%B3%91%EB%B2%95%EC%9D%B8%EC%9D%84%>

약 체결, 리밸리온이 통합법인 경영 담당, SK텔레콤이 전략적 투자자로 글로벌 시장 지원하여 국내에서 규모 부문으로 제1의 AI반도체 팩리스 업체가 되었음.

□ 기술

- 주요 기술제품으로 ‘아이온(Ion)’, ‘아톰(Atom)’, ‘리벨(Rebel)’
 - 아이온: 고주파 거래 최적화, 금융 시장 대상, 듀얼 뉴럴 코어 적용⁷⁸⁾
 - 아톰: 데이터센터용 SoC, 삼성 5nm 공정, 효율적 연산과 에너지 절감
 - 리벨: LLM 가속화용, 4nm 공정과 HBM3e 메모리 탑재, 2024년 하반기 출시 목표

□ 핵심경쟁력과 생태계

- 기술적 우수성으로 AI 반도체 ‘ATOM’은 NVIDIA GPU T4 대비 3.4배, 퀄컴 제품 대비 1.4배 빠른 처리 속도를 보유⁷⁹⁾하고 있으며, 컴퓨터 비전과 언어 처리 두 부문 모두에서 해외 대기업을 능가하는 성능 달성⁸⁰⁾. 기존 GPU 대비 20-30% 수준의 전력 소모로 AI 연산이 가능한 에너지 효율성 확보⁸¹⁾
- 인적 자원 측면에서도 ARM, 삼성전자, SK하이닉스 출신의 박사급 베테랑 개발자들로 구성되어 있는 남다른 기업의 인재 구성이 강점. 더불어, 삼성전자, ARM 등 글로벌 기업들과의 전략적 협력 관계 구축하여 글로벌 시장 진출을 도모하고 있는 점은 강점⁸²⁾으로 작용.

20%EC%9D%B4%EB%81%8C%EA%B8%B0%EB%A1%9C%20%ED%95%98%EB%A9%84%EC%84%9C%20%EC%83%88%20%ED%9A%8C%EC%82%AC%EC%9D%98%20%EC%82%AC%EB%AA%85%EC%9D%80%20%E2%80%98%EB%A6%AC%EB%B2%A8%EB%A6%AC%EC%98%A8%E2%80%99%EC%9C%BC%EB%A1%9C%20%EA%B2%80%EC%A0%95%EB%90%90%EB%8B%A4. (검색일 24.09.08.)

78) Zdnet(24.08.22.), “리밸리온, 차세대 AI칩 ‘리밸-쿼드’ 양산 앞당긴다…‘삼성 지원 덕분’” <https://zdnet.co.kr/view/?no=20240821105745> (검색일 24.09.08.)

79) 한경(23.04.06.), “3년차 리밸리온, 세계 최고 수준 AI반도체 개발…퀄컴·엔비디아 꺾었다”, <https://www.hankung.com/article/202304061064i> (검색일 24.07.08.)

80) 상동

81) 매경(24.04.19.), “‘K팹리스’ 라이징 스타…‘유니콘’ 눈앞”, <https://www.mk.co.kr/economy/view.php?no=291473&sc=50000001&year=2024> (검색일 24.11.15.)

아. 딥엑스

□ 기업 개요

- 딥엑스(DEEPX)는 AI 반도체를 개발하는 한국의 스타트업 기업임⁸³⁾.
 - 2024년 4월 시리즈 C 펀딩으로 1,100억 원 투자를 유치 받았으며, 이는 국내 스타트업 중 상반기 최대 규모⁸⁴⁾
 - 딥엑스의 투자 유치는 상반기 콘텐츠 분야 스타트업 기업에 대한 투자액이 96% 감소하고, AI 기술 분야에 대한 총 투자액이 5,230억 7,000만원이라는 점을 고려할 때 상당히 큰 규모라고 할 수 있음
- 딥엑스는 현재 국내 대표적인 AI 반도체 팝리스 기업으로 평가받고 있음⁸⁵⁾.
 - 특히 보유 개수가 252개. 특히 진흥 유공 단체 최고상인 ‘대통령상’ 수상
 - EE타임스 ‘실리콘 100 기업’ 선정, VSD 저널 ‘혁신가상’ 수상
 - 2024년 CES에서 3개 부문 ‘혁신상’ 수상 (전 세계 AI 반도체 팝리스 최초)

□ 기술

○ 주요 기술

- 현재 다양한 엣지 및 서버 AI응용분야에 맞춰 성능과 기능이 최적화된 AI 반도체 4종 보유

82) 서울신문(24.10.15.), “리밸리온, Arm·삼성과 손잡고 AI CPU 칩셋 플랫폼 개발”, <https://www.seoul.co.kr/news/economy/industry/2024/10/15/20241015500228> (검색일 24.11.15.)

83) 시사저널e(24.07.23.), “투자 유치 1위 스타트업 ‘딥엑스’, 특허협업으로 시장 공략” <https://www.sisajournal-e.com/news/articleView.html?idxno=404477> (검색일 24.11.15.)

84) Edge AI and Vision Alliance(24.05.10.), “Korean Semiconductor Industry Titans Back DEEPX in Series C Funding Round”, <https://www.edge-ai-vision.com/2024/05/korean-semiconductor-industry-titans-back-deepx-in-series-c-funding-round/> (검색일 24.11.15.)

85) 보안뉴스(24.05.13.), “AI 반도체 팝리스 ‘딥엑스’, 국내 반도체 산업 주목 이끌며 1,100억 원 규모 신규 투자 유치” <https://www.boanews.com/media/view.asp?idx=129726&direct=mobile> (검색일 24.11.15.)

<표 2-9> 딥엑스 AI 주요 기술

기 술	내 용
DX-V1	- 비전 시스템 특화 - 한 카메라로 최신 AI 알고리즘 연산 처리 가능
DX-V3	- 카메라 3~4대 동시 연산처리 가능 - 자율 주행, 로봇 비전 등 특화
DX-M1	- IPC(산업용 PC) 수준 성능 - 공장 적용 예정 - 1개 칩으로 다채널(16채널 이상) 영상 대상 30FPS/SEC 이상 실시간 AI 연산 처리 지원
DX-H1	- 대규모 AI 연산용 서버 솔루션 - 성능, 전력, 비용 효율성을 극대화

자료: DeepX 홈페이지, Edge AI and Vision Alliance(24.01.11.), “DEEPX Unveils ‘All-in-4 AI Total Solution’ at CES 2024: Four AI Chips Transforming the On-device AI Market”, <https://www.edge-ai-vision.com/2024/01/deepx-unveils-all-in-4-ai-total-solution-at-ces-2024-four-ai-chips-transforming-the-on-device-ai-market/> (검색일 24.11.18.)

○ 독자적인 기술

- 독자적인 기술력 구축이 강점이며, 다양한 칩을 제작하여 수요기업의 니즈에 대응을 기민하게 할 수 있는 강점이 존재. 엣지 컴퓨팅용 신경망 프로세싱 유닛 (NPU) 분야 공개특허는 34건, 269건의 특허를 출원 상태. 자체 확보 메모리 저감 기술 통해 거대언어모델 (LLM)을 구동 가능 AI 반도체 개발에 착수⁸⁶⁾ 및 금년 하반기 1세대 반도체 제품 양산을 시작, 내년 LLM AI 반도체 프로토타입 제품을 출시 목표⁸⁷⁾

□ 핵심경쟁력과 생태계

- 사업 모델은 AI 반도체 설계, 개발, AI 컴퓨팅 솔루션 제공, 온디바이스 AI 반도체 제품군 공급이며, LG 유플러스, 포스코 DX등 주요 기업들과 협력

86) 매경(24.07.09.), “김녹원 딥엑스 대표 “가전, 로봇 등에서 생성형 AI 구동하는 반도체 만들 것”, <https://www.mk.co.kr/news/it/11062973> (검색일 24.11.15.)

87) 시사저널e(24.07.23.), “투자 유치 1위 스타트업 ‘딥엑스’, 특허협업으로 시장 공략” <https://www.sisajournal-e.com/news/articleView.html?idxno=404477> (검색일 24.11.15.)

○ 주요 거래 기업

- LG 유플러스의 소형언어모델(sLLM) ‘익시젠(ixi-GEN)’을 적용한 AI반도체 연구 개발 목표 온디바이스 AI는 통신장비, AI컨택센터(AICC), 소호(SOHO), 로봇, 모빌리티 등 LG유플러스 자체 사업 도입 예정⁸⁸⁾
- 포스코 DX와 기술 개발 협력 계획, AI 반도체 기반 공장 자동화 및 지능화 물류 시스템 상용화 목적. 산업현장 설비시스템 운영을 자동화 가능한 대단위 AI 솔루션 공동 개발 계획⁸⁹⁾.

자. 모빌린트

□ 기업 개요

- 모빌린트는 2019년 설립된 AI반도체 설계 전문 팝리스 기업으로 AI 반도체 (ARIES; 에리스)⁹⁰⁾ 양산 및 차세대 칩(REGULUS; 레귤러스)⁹¹⁾ 개발에 집중
 - 2022년 AI반도체 에리스 개발
 - 모빌린트의 MLA100의 높은 경쟁력: 경쟁사 제품 대비 AI 성능 4배, 에너지 사용 수준 1/5배 이하, 가격 1/2 수준
 - 모빌린트의 차세대 칩 레귤러스: 독립형 AI 반도체. 저전력(5W 이내) 고성능 AI 기능 수행. 소형 로봇, 드론 등 다양한 분야에 활용가능성 多.

88) 디일렉(24.07.02.), “LG유플러스, AI 반도체 직접 만든다” https://www.thelec.kr/news/articleView.html?id_no=28867 (검색일 24.11.15)

89) zdnet(23.04.06.), “딥엑스-포스코DX, AI 반도체 기반 공장 자동화 개발 협력”, <https://zdnet.co.kr/view/?no=20230406143842> (검색일 24.11.15.)

90) Aving.net (22.12.23.), “Mobilint Set out to Introduce ‘Aries’, an AI Semiconductor for Edge Computing, at CES 2023… “To Dominate the Global NPU Market”, <https://us.aving.net/news/articleView.html?idxno=49994> (검색일 24.11.15.)

91) PR newswire(25.01.08.), “Mobilint Unveils LLM Demo on High-Efficiency Edge AI Accelerator at CES 2025”, <https://www.prnewswire.com/news-releases/mobilint-unveils-llm-demo-on-high-efficiency-edge-ai-accelerator-at-ces-2025-302345545.html> (검색일 25.01.15.)

□ 제품

○ 에리스(ARIES)

- 에리스는 초당 최대 80조 번(80 TOPS) 연산 가능한 AI 추론 칩
- 200개 이상의 오픈 소스 딥러닝 모델로 테스트를, 첨단 비전 애플리케이션, 엣지 서버, 하이퍼스케일 데이터 센터에 활용 가능⁹²⁾
- 2025년 초부터 기존 합성곱신경망(CNN), 순환신경망(RNN), LSTM(Long short-term memory)에 더하여 트랜스포머 계열 모델 연산까지 지원 계획⁹³⁾

○ 레귤러스(REGULUS)

- 레귤러스는 온디바이스 AI용 고효율 시스템 온 칩(SoC)
- 레귤러스는 고성능 CPU, 코덱, ISP등을 내장하고 3 와트(w) 이하의 전력으로 10TOPS(1초당 1조 번 연산)이상의 AI 연산 성능을 낼 수 있는 온디바이스 AI 용 고효율 시스템 온 칩 제품이며 로봇, 드론, 가전, 블랙박스, CCTV 등에 활용돼 AI 기능 수행⁹⁴⁾
- CES2025 혁신상 수상으로 기술력 검증⁹⁵⁾

○ 주요 거래 기업

- 모빌린트는 세미파이브와 협력해 AI 반도체(ARIES) 양산 돌입⁹⁶⁾
- 모빌린트는 어드밴텍과 에지 인공지능(Edge AI) 사업 확대를 위해 전략적 업무 협약을 체결함. 고성능 에지 AI PC 솔루션을 포함, 다양한 산업현장에서 사용 되는 에지 AI 수요에 대응하며 사업 확대를 위해 협력할 계획. 어드밴텍의 에지

92) 디일렉(24.03.28.), “세미파이브, 모빌린트 AI 반도체 ‘에리스’ 양산 돌입” <https://www.thelec.kr/news/articleView.html?idxno=26866> (검색일 24.11.15.)

93) AI타임즈(24.09.09.), “모빌린트, 실리콘밸리 ‘AI 서밋’서 엣지 AI 반도체·테스트 플랫폼 공개” <https://www.aitimes.com/news/articleView.html?idxno=163218> (검색일 24.11.15.)

94) <https://www.aitimes.com/news/articleView.html?idxno=163218> (검색일 24.11.15.)

95) AI타임즈(24.12.11.), “모빌린트, 온디바이스 AI 저전력 반도체로 ‘CES 혁신상’ 수상”, <https://www.aitimes.com/news/articleView.html?idxno=166129> (검색일 24.12.29.)

96) 디일렉(24.03.28.), “세미파이브, 모빌린트 AI 반도체 ‘에리스’ 양산 돌입” <https://www.thelec.kr/news/articleView.html?idxno=26866> (검색일 24.11.15.)

AI 제품과 모빌린트의 NPU(Neural Processing Unit)를 통해 갈수록 늘어나는 시장 수요에 대응해 시너지를 발휘, 국내 AI 비즈니스에 영향력을 확대해 나간다는 전략 수립⁹⁷⁾

차. 하이퍼엑셀

□ 기업 개요

- 인공지능(AI) 반도체 개발 기업 (거대언어모델(LLM)에 최적화)
 - 2023년 1월 김주영 대표(KAIST 전기및전자공학부 교수)가 창업
 - 550억원 규모의 시리즈 A투자 유치⁹⁸⁾
 - 창업 이전부터 KAIST 연구소에서 LLM 특화 반도체 연구개발(R&D) 수행
 - AI 반도체 공급 및 비용문제 해결 목표, 생성형 AI 특화 AI 반도체와 솔루션 개발 목표⁹⁹⁾

□ 기술

○ 주요 기술

- 2023년 7월, LPU(Latency Processing Unit: 자연어 처리 특화용 반도체) 개발: NVIDIA A100 기반 슈퍼컴퓨터 대비 성능 최대 50%, 가격 대비 성능 2.4배 가량 향상¹⁰⁰⁾.

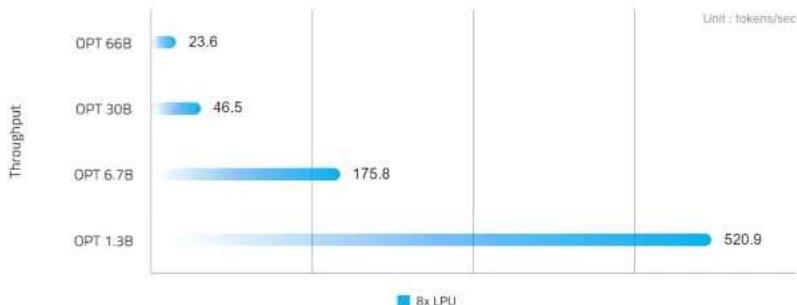
97) IT데일리(24.01.02.), “어드밴텍, 예지 AI 사업 확대 위해 모빌린트와 MOU 체결” <http://www.itdaily.kr/news/articleView.html?idxno=219276> (검색일 24.11.29.)

98) Wowtale(24.12.26.), “LLM 특화 AI 반도체 개발 ‘하이퍼엑셀’, 550억원 규모 시리즈A 투자유치”, <https://www.wowtale.net/2024/12/26/235287/> (검색일 25.1.15.)

99) 전자신문(24.07.07.), “[우리가 AX 주역]〈24〉하이퍼엑셀, LLM에 최적화된 AI반도체 기업”, <https://www.eetnews.com/20240707000060> (검색일 24.11.29.)

100) KAIST 뉴스(23.08.04.), “2.4배 가격 효율적인 챗GPT 핵심 AI반도체 개발”, https://kaist.ac.kr/news/html/news/?mode=V&mng_no=30750&skey=keyword&sval=2&GotoPage=5 (검색일 24.11.29.)

[그림 2-25] 하이퍼엑셀 오리온 서버 성능



자료: 파이낸셜포스트(23.11.03.), “하이퍼엑셀, 챗GPT 고성능 GPU 대체 기술개발…비용·고효율 가속 서버 ‘오리온’ 출시”, <https://www.financialpost.co.kr/news/articleView.html?idxno=172970> (검색일 24.11.16.)

- 2024년 11월, FPGA(프로그래머블반도체) 기반 생성형 AI 전용 서버 ‘오리온’ 발매. 메타 자체 서비스 대비 서비스 답변 속도 30~50% 향상¹⁰¹⁾. LLM을 여러 개의 LPU에 효율적으로 분산하는 모델 병렬화 기술과 LPU 간 데이터 동기화를 위한 자체 네트워킹 기술로 성능과 효율을 높임. 베르다(Bertha) AI 반도체를 개발 중. 4나노 공정 기술 적용 예정. 2026년 1분기 양산 목표.¹⁰²⁾

○ 주요 거래 기업

- 정원엔시스(IT 인프라 공급 기업)와 총판 계약(2024년 5월)¹⁰³⁾, 세미파이브(반도체 설계 솔루션 기업)와 생성형 AI 반도체 ‘베르다(Bertha)’ 양산 계약(2024년 10월)¹⁰⁴⁾

101) 카포스트(24.01.10.), “세미파이브, 하이퍼엑셀과 AI 반도체 개발 협력”, <https://www.kipost.net/news/articleView.html?idxno=315709> (검색일 24.11.29.)

102) Zdnet(24.10.11.), “세미파이브, 하이퍼엑셀과 4나노 AI칩 양산 계약 체결”, <https://zdnet.co.kr/view/?no=20241011082022> (검색일 24.11.29.)

103) 헬로티(24.05.21.), “하이퍼엑셀, 정원엔시스와 AI 반도체 서버 총판 계약 체결”, <https://www.hellot.net/news/article.html?no=89965> (검색일 24.12.16.)

104) 전자신문(24.10.11.), “세미파이브 “하이퍼엑셀 AI 칩 양산 계약””, <https://www.etnews.com/2024101100046> (검색일 24.12.16.)

제2절 정부 R&D 투자 동향

1. 개관

가. 반도체 분야 정부R&D 투자 동향¹⁰⁵⁾

과거, 민간 주도 분야라는 인식에 따라 반도체 분야의 정부 지원은 다소 미미하였으나, 전 세계적인 디지털전환 흐름에 따른 시스템반도체 산업의 수요 및 중요도 확대로 '19년 이후 국내 반도체 산업 생태계 강화를 위한 정부 투자가 급격히 증가하였다. 특히, 우리 정부는 '19년 이후 AI반도체를 중심으로 한 시스템반도체 산업 육성을 통한 종합 반도체 강국 도약을 위해 매년 주요 정책을 발표¹⁰⁶⁾하고 있으며, '20년 이후로는 「차세대지능형반도체기술개발사업」, 「PIM인공지능반도체기술개발사업」 등 1조 원 규모의 대형 예타사업을 추진하는 등 대대적인 지원을 이어나가고 있다.

지난 6년('18~'23년) 간 반도체 분야 정부R&D 전용 사업의 총 투자액은 1조 4,308억 원 규모로 '18년 256억 원 규모에서 '23년 4,454억 원을 달성하며, 연평균 77.4% 증가한 것으로 나타났다. '18년 과학기술정보통신부(이하 과기정통부) 주관 사업을 중심(96.6%)으로 투자가 이루어졌으나, '20년 이후 앞서 언급한 다부처 예타 사업을 시작으로 산업통상자원부(이하 산업부)의 지원 규모는 연평균 약 200% 수준으로 급격히 확대가 이루어졌다. 과기정통부는 세부사업을 중심으로 AI반도체 분야 기술개발, 기업지원, 인프라 등을 지원 중이며, 산업부는 그간 프로그램형 사업을 통해 화합물·전력 반도체를 비롯한 산업 수요 연계형 기술개발에 집중한 것으로 나타났다. 두 부처는 개발·응용 분야에 각각 70.6%, 89.1%로 상용화 지향형 기술개발 사업을 지원하였으며, 과기정통부는 출연연·대학, 산업부는 중소·중견기업을 중심으로 지원¹⁰⁷⁾ 중이다.

105) 본 연구에서는 주요R&D사업 중 사업의 목적·내용·범위가 반도체 기술·산업·인력양성·기업지원 등에 집중된 세부 /내역사업 단위를 '전용사업'으로 구분하여 분석 범위로 설정하였으며, 일부 내역/과제단위에서 지원중인 내역은 제외

106) ('19) 시스템반도체 비전과 전략 → ('20) 「인공지능반도체산업발전전략」 → ('21) 「K-반도체 전략」→ ('22) 「인공지능반도체산업성장지원대책」→ ('23) 「반도체 미래기술 로드맵」 → ('24) 「AI-반도체 이니셔티브」 등

107) (과기정통부) 출연연·대학 72.3%, (산업부) 중소·중견기업 95.9%

<표 2-10> 반도체 분야 주관부처별 투자 현황('18~'23년)

단위: 백만원, %

구 분	2018	2019	2020	2021	2022	2023	총계
과기정통부	24,751	64,226	141,391	159,885	223,411	246,847	860,511
산업부	865	18,953	88,701	105,730	153,411	202,597	570,257
합계	25,616	83,179	230,092	265,615	376,822	449,444	1,430,768

그간 정부의 반도체 분야 전용 R&D사업은 크게 ①기술개발 ②기업지원 ③인력양성·국제협력 ④인프라 등으로 구분할 수 있다. '20년까지는 기술개발 및 인프라 구축을 중심으로 지원이 이루어졌다. '20년 이후 급격한 투자 규모 확대에 따라 기업지원, 인력양성·국제협력을 포함한 전방위적인 분야로 지원 범위가 넓어졌다. 지난 6년간 기술개발 분야의 투자 규모는 총 9,736억 원 수준으로 68.4%의 가장 높은 비중(연평균 79.7% 확대)을 차지하였다. 주요 기술개발 분야로는 과기부·산업부 다부처 예타 사업을 통해 차세대 반도체 소자·설계·제조 기술 및 화합물·전력반도체, 센서 소자 등 개발 내용을 포함하고 있다. 산업부의 본격적인 투자 이후 기업지원 분야는 '20년 179억 원 규모에서 '23년 612억 원 규모로 연평균 50% 수준 투자 확대가 이루어졌으며, 인력양성·국제협력 분야는 '23년 민·관협력 예타 사업¹⁰⁸⁾ 추진에 따라 지원 규모가 큰 폭으로 증가한 것으로 나타났다. 기업지원 분야는 국내 팹리스의 역량 강화를 통한 설계 생태계 조성을 통한 AI반도체, 치량용 반도체 등 핵심 수요산업과의 연계를 지원하며, 그간 과기정통부·과기부의 인재양성 사업 내 내역사업 및 과제 단위¹⁰⁹⁾에서 지속 추진되던 인력양성 분야는 산업 수요형 인력양성을 위한 민·관협력 사업 추진에 따라 세부사업 단위에서 '23년 이후 신규 투자가 추진 중이다. 마지막으로, 인프라 구축 분야는 그간 한국과학기술원 부설 나노종합기술원¹¹⁰⁾을 중심으로 공공 인프라 구축·운영 지원이 이루어졌으며, '22년 이후 차량용 반도체, 통신용 반도체, 전력·화합물반도체 등 수요 산업과의 연계 및 제조역량 강화를 위한 신규 지원 확대 중인 것으로 나타났다.

108) (산업부) 민관공동투자반도체고급인력양성사업, ('23년)100억 원

109) (과기정통부) 과학기술혁신인재양성사업 (내내역: 시스템반도체융합분야, '반도체설계구현인재양성'), (산업부) 산업 혁신인재성장지원사업(내역: 교육훈련' 내 과제 단위) 등을 포함하나 내내역사업/과제 등은 본 통계에서는 제외

110) (과기정통부) 나노종합기술원운영비지원(주요사업비), ('23년)365억 원

<표 2-11> 반도체 주요 분야별 투자 현황('18~'23년)

단위: 백만원, %

구 분	2018	2019	2020	2021	2022	2023	총계
①기술개발	16,194	56,109	139,191	184,726	274,109	303,240	973,569
②기업지원	-	-	17,884	31,089	48,253	61,153	158,379
③인력양성·국제협력	-	-	3,600	7,200	8,600	26,021	45,421
④인프라구축	9,422	27,070	69,417	42,600	45,860	59,030	253,399
합계	25,616	83,179	230,092	265,615	376,822	449,444	1,430,768

반도체 분야 전용 R&D사업의 수행 주체 및 연구개발 단계는 과거 대학·출연연 중심의 응용연구에서 점차 산업계 중심의 개발·응용연구 중심으로 무게중심이 이동한 것으로 나타났다. 대학·출연연은 '18년 79.3%로 절대적인 수행비중이 나타났으나, 중소·중견·대기업 등 산업계 참여 확대로 지난 6년간 연평균 263% 확대하여 전체 비중의 45.0%를 차지하여 산·학·연 중 가장 높은 수행 비중을 차지하고 있다. 특히, 중견·대기업의 참여 비중은 연평균 133.2%, 95.7% 확대하며 5년간 1,166억 원, 292억 원 규모의 연구를 수행한 것으로 나타났다. 산업계는 상용화 중심의 기술개발, 기업지원 분야의 사업을 수행하였으며, 대학은 기초·원천 중심의 기술개발 및 인력양성·국제협력 분야를 중점적으로 수행한 것으로 나타났다. 국공립·출연연구소는 상대적으로 4개 분야 사업에 유사 비중으로 참여하였다.

한편, 그간 반도체 분야 응용·개발 단계에 77.9%를 집중함으로써 제품·상용화 성과 창출을 위한 정부 투자가 이루어진 것으로 나타났다. 그간 개발연구는 산업계가 74.7%의 비중으로 주도하였으며, 응용연구는 국공립·출연연 및 산업계가 38.7%, 36.1%를 차지, 기초연구는 대학을 중심(71.7%)으로 지원한 것으로 나타났다.

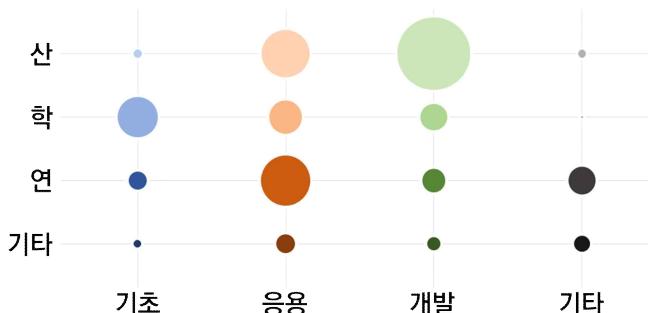
<표 2-12> 반도체 분야 수행주체별 투자 현황

구 分	2018	2019	2020	2021	2022	2023	총계
산업계	2,615	60,374	91,616	143,352	155,877	453,833	645,142
중소기업	2,615	50,324	65,482	106,539	120,089	345,049	499,349
중견기업	-	6,450	20,534	28,824	25,999	81,806	116,598
대기업	-	3,600	5,600	7,989	9,789	26,978	29,195
대학	5,818	16,871	40,437	71,617	53,303	188,045	314,743
국공립·출연연구소	4,200	11,749	18,412	29,158	39,164	102,683	372,086
기타	-	4,365	4,320	7,505	9,695	25,885	98,798
합계	25,616	83,179	230,092	265,615	376,822	449,444	1,430,768

<표 2-13> 반도체 주요 연구개발 단계 별 투자 현황('18~'23년)

단위: 백만원

구 분	2018	2019	2020	2021	2022	2023	총계
기초연구	7,494	14,929	24,088	38,831	54,278	60,675	200,295
개발연구	5,603	23,242	70,758	112,838	147,207	214,140	573,788
응용연구	10,807	36,773	120,473	102,449	138,972	131,860	541,334
기타	1,713	8,235	14,774	11,497	36,365	42,769	115,353
합계	25,616	83,179	230,092	265,615	376,822	449,444	1,430,768



단위: 백만원

	기초연구	응용연구	개발연구	기타	합계
산	11,178	195,169	428,449	10,346	645,142
학	143,679	99,018	69,470	2,576	314,743
연	35,495	209,760	54,684	72,147	372,086
기타	9,943	37,386	21,185	30,284	98,798
합계	200,294	541,333	573,788	115,353	1,430,768

나. AI반도체 분야 정부R&D 투자 동향¹¹¹⁾

그간 정부에서 지원한 반도체 전용 사업 중 53.8%는 AI반도체 분야에 집중, 중장기적인 투자가 이루어지고 있다. AI반도체 전용 사업으로는 지난 '19년 내역사업 1건¹¹²⁾에서부터 시작하여, '20년 다부처 협력 추진 등 대형사업 중심의 지원을 추진하며 연간 112.6% 확대, 지난 6년간 총 7,704억 원 투자¹¹³⁾가 이루어졌다. 그간 AI반도체의 중점 투자 분야는 90% 이상 기술개발에 집중되어 있으며, 수행 주체로는 대

111) 본 연구에서는 반도체 전용 사업 내 AI반도체 특화 세부/내역사업 및 지원 과제(총 843개)를 중심으로 분석

112) 과기정통부, 「ICT융합산업혁신기술개발」 내 '지능형반도체' 내역 사업

113) '23년 기준 15개의 전용사업을 지원 중

학·중소기업이 각각 38.8%, 37.8%로 주도하고 있다. 기술개발 측면에서는 2개의 예타사업¹¹⁴⁾을 중심으로 NPU, PIM반도체 등 차세대 핵심기술 확보를 위한 중장기적인 투자 중이며, 과기정통부는 핵심 SW 기술 확보¹¹⁵⁾ 및 도메인별 AI반도체 활용¹¹⁶⁾을 위한 다수의 사업을 추진 중이다. 한편, 산업부는 다부처 예타 사업 구조 내에서 차세대 공정·제조·장비 분야의 핵심기술 개발을 추진 중이다.

<표 2-14> AI반도체 기술개발 분야 주요 사업

부처	사업명	주요 내용	사업기간	총사업비
과기 정통 부	차세대지 능형반도 체기술개 발(소자)	<ul style="list-style-type: none"> • (목적) 대량의 데이터를 처리하기 위한 고성능, 초저전력의 반도체 핵심기술 개발을 통해 반도체 생태계의 균형 발전 및 미래 산업 경쟁력 확보 - (신소자 원천기술) 기존 반도체 한계를 넘어서는 초저전력·고성능 신소자 개발 - (신소자집적·검증기술) 조기상용화 가능한 신소자 개발 및 IP 확보를 위한 웨이퍼 레벨의 신소자 집적/검증기술 개발 - (신개념 소자 기초기술) 반도체 패러다임을 바꿀 신소재, 신공정 등 창의적·도전적 아이디어 기반의 혁신적 소자 기초기술 개발 	'20~'29 (예타)	2,405억 원 (국비 2,405억 원)
	차세대지 능형반도 체기술개 발(설계)	<ul style="list-style-type: none"> • (목적) 인공지능 강국 실현을 위한 기반을 구축하고, 글로벌 반도체 패러다임 전환을 선도하기 위한 AI반도체(NPU) 핵심기술 확보 - (인공지능 프로세서) 서버, 모바일, 엣지(IoT 등) 분야에 활용 가능한 세계 수준의 인공지능 반도체(Neural Processing Unit) 개발 - (초고속 인터페이스) AI반도체 성능향상을 위한 프로세서 간, 프로세서와 메모리간 대용량 데이터를 고속으로 전송하기 위한 인터페이스 기술개발 - (지능형반도체SW) 다양한 인공지능 응용 분야에 대하여 인공지능 반도체의 고성능·저전력 특성의 최적화 및 활용성 증대를 위한 SW 개발 	'20~'29 (예타)	2,475억 원 (국비 2,045억 원)

114) (다부처(과기정통부·산업부)) 「차세대지능형반도체기술개발사업」, 「PIM인공지능반도체핵심기술개발」

115) (과기정통부) SW컴퓨팅산업원천기술개발(내역: AI반도체SW핵심기술)(‘21~’24), 과기정통부거대인공 신경망인공지능반도체SW기술개발(‘23~’27), 과기정통부인공지능반도체SW통합플랫폼기술개발(‘23~’27) 등

116) 과기정통부스마트엣지디바이스기술개발(‘22~’26), 과기정통부자율주행용인공지능반도체핵심기술개발(‘22~’25) 등

부처	사업명	주요 내용	사업기간	총사업비
산업부	차세대지능형반도체기술개발(설계·제조)	<ul style="list-style-type: none"> (목적) 반도체 산업 경쟁력 강화를 위한 주력산업 경쟁력 강화를 위한 차세대 시스템반도체 및 미세화 제조 공정기술 지원 (시스템반도체상용화 설계) 주력산업에 필요한 반도체 설계 핵심기술 R&D 지원을 통해 시스템반도체 분야 성장 촉진 (반도체제조공정장비) 차세대 반도체 제조에 필요한 공정/장비 기술개발을 통해 반도체 장비 분야 국산화율 제고 	'22~'26 (예타)	5,216억 원 (국비 4,277억 원)
과기정통부	PIM인공지능반도체핵심기술개발(소자)	<ul style="list-style-type: none"> (목적) 미래 반도체 시장의 새로운 패러다임인 기억(메모리)과 연산(프로세서)을 통합한 신개념 반도체(PIM) 등 핵심기술 확보 신구조·신재료 PIM 소자 및 어레이 기술, 신개념 PIM 기초 기술 및 소자·단위 셀 공정 및 검증기술 개발 추진 	'22~'28 (예타)	630억원 (국비 630억원)
	PIM인공지능반도체핵심기술개발(설계)	<ul style="list-style-type: none"> (목적) 저장(메모리)과 연산(프로세서) 기능을 통합한 PIM반도체 기술개발 및 산업 생태계 구축을 통해 AI반도체 글로벌 기술 선점과 신시장 주도권 확보 (PIM 설계기술) 프로세서로직과 메모리(DRAM, SRAM, MRAM, PRAM, RRAM 등)를 융합한 PIM 인공지능반도체 설계기술 개발 (PIM 혁신기반기술) 시스템 아키텍처 및 SW, 인터페이스, 설계툴, 인력양성 등 PIM 기반 구축 기술개발 	'22~'28 (예타)	2,262억 원 (국비 1,912억 원)
산업부	PIM인공지능반도체핵심기술개발(제조)	<ul style="list-style-type: none"> (목적) DRAM 제조 공정 기술 고도화, PIM용 차세대 상용 메모리(MRAM, PRAM) 아키텍처 및 공정·소재·장비 상용화 기술 개발을 통한 PIM용 메모리 기술 고도화 PIM 컴퓨팅구조 적용위한 메모리 인터페이스·상용플랫폼 기술개발 PIM 기반 컴퓨팅 시스템용 SCM 제어 반도체 기술개발 PRAM, MRAM 기반 PIM 제조를 위한 장비 상용 기술개발 	'22~'28 (예타)	1,118억 원 (국비 957억원)
과기정통부	SW컴퓨팅산업원천기술개발	<ul style="list-style-type: none"> (목적) 전 세계적인 디지털 경제 확산과 기술패권 경쟁 속에서 국가·사회 디지털 혁신의 근간인 SW·AI·데이터·클라우드 등 핵심기반을 강화하여 전 산업에 응용하는 SW 원천기술 개발 지원 (AI반도체SW핵심기술) 국내 인공지능 반도체 성능 최적화를 위한 인공지능 반도체 SW 원천기술 개발 	'21~'24	345억원 (국비 345억원)
	거대인공지능신경망인공지능반도체SW기술개발	<ul style="list-style-type: none"> (목적) 국산 AI반도체(PIM, NPU) 기반 초거대 인공지능 서비스 구현을 위한 AI반도체 SW 기술 및 거대인공신경망 처리 SW 플랫폼 개발 거대인공신경망 모델의 학습·추론에 최적화된 연산 처리가 가능한 국산 AI반도체(PIM, NPU) 기반 시스템 SW 개발 국산 NPU, PIM 반도체(2종 이상)를 활용한 컴퓨팅 서버를 구축하고, 동 사업에서 개발한 시스템 SW을 탑재하여 성능 검증 및 실증 	'23~'27	337억원 (국비 253억원)

부처	사업명	주요 내용	사업기간	총사업비
	인공지능 반도체S W통합 플랫폼기 술개발	<ul style="list-style-type: none"> (목적) 다양한 스마트 디바이스(엣지)에 국내 기술의 인공지능반도체(NPU)의 적용을 위한 인공지능반도체 SW 통합플랫폼 기술개발 - 공개 가능한 AI반도체 모듈(핵심 IP, 오픈 라이브러리 등)을 조합하여 엣지용 AI반도체 설계할 수 있는 공개 NPU 아키텍처 및 최적화 컴파일러 개발 및 성능시험평가(BMT) 지원 툴 개발 	'23~'27	447억원 (국비 393억원)
	스마트엣 지디바이 스기술개 발	<ul style="list-style-type: none"> (목적) 디지털(데이터·AI 등) 생태계 조성 및 국산 인공지능 반도체 활성화를 위한 맞춤형 스마트 엣지 디바이스 기술 개발 	'22~'26	263억원 (국비 263억원)
	자율주행 용인공지 능반도체 핵심기술 개발	<ul style="list-style-type: none"> (목적) 레벨4 이상 자율주행 실현을 위해 국내 자체기술 기반으로 자율주행 인공지능 반도체 및 자율주행 인공지능 컴퓨팅 모듈 기술 개발 - 자율주행 레벨4 이상 다중센서 인지 인공지능 SW와 기능안정성이 내재된 인공지능 반도체(NPU) 및 반도체 SW 기술 개발 - 국산 인공지능 반도체를 탑재하여 고화질 영상·레이더 등 센서 제어 SW 등 자율주행 기능안전성 기술을 통합 연계 지원하는 컴퓨팅 모듈 개발 	'22~'25	497억원 (국비 390억원)

출처: (다부처(과기정통부·산업부)) 「차세대지능형반도체기술개발사업」, 「PIM인공지능반도체핵심기술개발」, (과기정통부) SW 컴퓨팅산업원천기술개발(내역: AI반도체SW핵심기술)(`21~`24), 과기정통부거대인공 신경망인공지능반도체SW기술개발(`23~`27), 과기정통부인공지능반도체SW통합플랫폼기술개발(`23~`27), 과기정통부스마트엣지디바이스기술개발(`22~`26), 과기정통부자율주행용인공지능반도체핵심기술개발(`22~`25)

'20년 이후 과기정통부 주도로 AI반도체 분야 조기 상용화 촉진, 시장경쟁력 강화 등 국내 영세 팹리스를 대상으로 한 기업지원 사업을 추진 중이다. 기업지원의 특성상 중소·중견기업(팹리스) 주관으로 산·학 컨소시엄을 구성하여 도메인 별 AI반도체 상용화를 위한 과제를 추진중이다. 대표적으로 '인공지능반도체응용기술개발' 사업에서는 카메라용 SoC(NPU), C-ITS 및 차량통신용 반도체 등 상용화 반도체 중심으로 개발 중이며, '인공지능반도체혁신기업집중육성' 사업은 AI반도체 관련 SW 및 IP 개발에 집중하고 있다. 다만, 현재 진행 중인 사업의 종료 시점(`24년) 도래에 따라 지원의 연속성을 위한 후속 사업 추진이 요구되는 실정이다.

<표 2-15> AI반도체 기업지원 분야 주요 사업

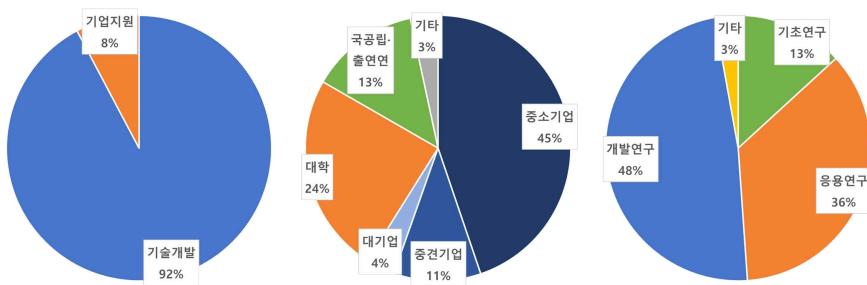
부처	사업명	주요 내용	사업기간	총사업비
과기 정통 부	인공지능 반도체 응용기술 개발	<ul style="list-style-type: none"> (목적) 팝리스 기업과 학연 간의 기술 및 인력 교류와 독자 적 혁신적 기술의 제품화 지원을 통해 AI반도체의 조기 상용화와 산업 생태계 조성 (내용) 학연의 경쟁력 있는 AI 반도체 R&D 결과물(IP, 특허)을 바탕으로 기술이전, 사업화연구인력 지원을 통한 기업의 제품화 추진 	'20~'24	489억 원 (국비 398억 원)
	인공지능 반도체 혁신기업 집중육성	<ul style="list-style-type: none"> (목적) AI·데이터 생태계의 핵심기반인 인공지능 반도체 시장에 도전하는 유망기업 발굴 및 집중지원을 통해 '인공지능 반도체 혁신기업' 육성 (내용) 인공지능 반도체 시장에서 성과 창출이 유망한 기업을 선정하여, 첨단 인공지능 반도체 기술개발 과정에서 직면하는 미세공정 전환, 신규 IP 개발·활용, SW 최적화 등 기술애로 특화 지원 	'21~'24	386억 원 (국비 298억 원)

<표 2-16> AI반도체 분야 영역별 지원 현황('18-'23년)

단위: 백만원

구분		2019	2020	2021	2022	2023	총계
주관부처	과기정통부	12,633	46,659	91,796	164,328	165,042	480,458
	산업부	-	46,700	62,988	87,303	92,997	289,988
중점분야	기술개발	12,633	91,512	142,514	228,184	235,879	710,722
	기업지원	-	1,847	12,270	23,447	22,160	59,724
수행주체	중소기업	2,615	50,324	65,482	106,539	120,089	345,049
	중견기업	-	6,450	20,534	28,824	25,999	81,806
	대기업	-	3,600	5,600	7,989	9,789	26,978
	대학	5,818	16,871	40,437	71,617	53,303	188,045
	국공립·출연연	4,200	11,749	18,412	29,158	39,164	102,683
	기타	-	4,365	4,320	7,505	9,695	25,885
	기초연구	800	8,334	24,492	39,553	28,446	101,624
연구개발 단계	응용연구	2,400	34,281	45,170	100,252	92,904	275,006
	개발연구	7,718	44,714	81,289	107,780	130,442	371,943
	기타	1,715	6,031	3,834	4,046	6,247	21,873
	합계	12,633	93,359	154,784	251,631	258,039	770,446

[그림 2-26] 원쪽부터 '19~'23년간 중점분야, 수행주체, 연구개발단계별 투자 비중



한편, '19~'23년 간 AI반도체 분야 전용 R&D사업 내 연구개발과제의 지원 동향을 파악하기 위해 총 843개를 심층 분석¹¹⁷⁾하였다. 과제 정보는 NTIS 내 사업명을 기준으로 산출한 '19~'23년 정보를 활용하였다. 분석 결과 그간 지원 과제는 크게 ▲소자기술 ▲설계기술 ▲제조기술 ▲공통기술 ▲기타 등 중점분야와 12개의 세부기술¹¹⁸⁾로 구분할 수 있었다. 이중 소자기술의 투자액은 총 3,444억 원으로 그간 투자액의 44.7%를 차지하였다.

'소자기술' 중점분야는 NPU, PIM, GPU, SoC, 메모리반도체 분야의 기술개발 과제를 포함하고 있으며, 산업 도메인 맞춤형 반도체인 SoC가 총 1,788억 원(169개 과제)으로 가장 높은 비중을 차지하였다. '19년에는 GPU와 SoC를 중심으로 투자가 이루어졌으나, '20년 예타 및 신규사업 추진 이후 NPU, PIM, GPU 등 개발 범위의 다양성이 증가하였다. 대표적인 산업 도메인으로는 바이오헬스, 자율주행 등 자동차·모빌리티, 디스플레이 구동 반도체, 각종 센서 등을 포함하고 있다. NPU와 PIM반도체 맞춤형 과제의 투자규모는 각각 1,124억 원(66개, 14.6%), 348억 원(35개, 4.5%)으로, 엣지컴퓨팅 프로세서 모듈, 인공지능 프로세서, 모빌리티 분야에 특화된 NPU, DRAM 기반 PIM 반도체, 인메모리 오류 정정 디바이스 등의 기술 분야를 포함하고 있다. 메모리 분야는 ReRAM, MRAM, eMRAM 등 AI반도체向 차세대 메모리반도체 개발에 집중하고 있으며, GPU는 모바일에 최적화된 GPU 소자 개발 과제를 포함하고

117) 본 과제 분석은 필자의 주관에 따라 분류한 기준에 따르므로, 주관부처 및 수행기관의 의도와 상이할 수 있음

118) 구분과 정의는 <표 2-18> 참고

있다. SoC 분야는 소자기술 내 가장 높은 비중을 차지하고 있으나, 산술평균에 따른 평균 과제비는 NPU 1.7억 원, SoC 1.0억 원, PIM 9.9억 원 규모로 나타나 단가 측면에서는 차이가 존재했다.

반도체 설계, SW, IP개발을 포함하는 ‘설계기술’ 중점분야는 그간 투자 총액에 22.0%를 차지하였다. 세부 분야로써 설계 분야는 인터페이스, 시스템 아키텍처, 소자 배선 설계, 로직 회로 설계, PIM 구조 설계 등 소자 구현을 위해 필요한 세부 컴포넌트별 설계 분야를 포괄하고 있다. 제작한 AI반도체의 AI모델 구현을 위한 시스템SW와 반도체 설계에 필요한 SW 등을 포함하는 SW 분야는 지속적으로 투자 규모가 확대 중인 것으로 나타났으며, 저전력 구동 반도체 및 영상 콘텐츠 서비스를 위한 AI 반도체 IP 기술개발 과제의 투자규모는 ’21년 이후 담보상태인 것으로 나타났다.

‘제조기술’ 중점분야는 차세대 공정 기술 및 이에 부합하는 소재·장비 관련 과제를 포함하고 있다. 공정기술 분야 대비 소재·장비 분야의 투자 규모는 약 3배 수준으로, 차세대 공정 구현을 위한 전/후공정 장비 개발에 집중 지원이 이루어지고 있다. 공정 기술은 증착(ALD, PECVD), CMP 등 전공정과 함께 WLP, 차세대 배선 구현 등 후공정(패키징) 분야를 집중적으로 지원 중인 것으로 나타났으며, 소재·장비는 이러한 첨단 공정 구현을 위한 다양한 증착, 식각, 세정 및 패키징, 테스트 장비 개발에 투자가 이루어지고 있다.

<표 2-17> AI반도체 분야 정부R&D과제 심층 분석 결과

구분	2019		2020		2021		2022		2023		합계	
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제
소자기술	6,215	5	47,790	45	75,891	70	106,004	97	108,463	89	344,363	306
NPU	-	-	21,100	11	23,952	14	32,430	20	34,893	21	112,375	66
PIM	-	-	-	-	9,000	5	12,967	13	12,793	17	34,760	35
GPU	500	1	480	1	870	2	490	1	-	-	2,340	5
SoC	5,715	4	24,656	28	38,685	42	53,044	52	53,646	43	175,746	169
메모리	-	-	1,554	5	3,384	7	7,073	11	7,131	8	19,142	31
공통기술	-	-	8,313	30	26,272	60	36,818	71	22,148	49	93,550	210
설계기술	6,118	3	13,935	13	23,898	27	56,736	52	68,710	53	169,398	148
설계	318	1	8,371	11	11,761	16	36,670	36	43,070	37	100,190	101
SW	5,800	2	5,564	2	10,545	8	17,402	12	24,012	14	63,323	38
IP	-	-	-	-	1,592	3	2,664	4	1,629	2	5,885	9
제조기술	-	-	18,537	22	24,489	28	46,473	51	52,973	54	142,472	155
공정	-	-	5,037	9	7,889	12	10,663	15	13,323	17	36,912	53
소재·장비	-	-	13,500	13	16,600	16	35,810	36	39,650	37	105,560	102
기타	300	1	4,784	6	4,234	5	5,601	6	5,745	6	20,664	24
총합계	12,633	9	93,359	116	154,784	190	251,631	277	258,039	251	770,446	843

<표 2-18> 세부 분야의 구분과 정의

중점분야	세부기술	정의
소자기술	NPU	AI 구동을 위한 소자(프로세싱 유닛)
	PIM	PIM반도체 특화 소자·설계
	GPU	모바일 GPU
	SoC	AI 적용(상용화) 산업/목적형 반도체 (자율주행, 통신, 헬스케어 등)
	메모리	SRAM, ReRAM, MRAM 등 차세대 메모리
	공통기술	차세대 소자·소재 등 원천·공통기술
설계기술	설계	인터페이스 기술, 배선 및 전기적 설계, 로직 회로 설계 등
	SW	매니코어 프로세서 등 시스템SW, 설계 및 SW 플랫폼
	IP	반도체 IP 개발
제조기술	공정	차세대 전/후공정 기술
	소재·장비	차세대 AI반도체 공정에 부합하는 공정 소재·장비 기술
기타	상용화 지원, 사업단·연구실 운영비, 기획평가관리비 등	

수행주체·연구개발단계 별 중점적으로 추진중인 세부기술 분야를 심층 분석한 결과는 아래 <표 2-19>, <표 2-20>와 같았다. 그간 수행한 세부기술 분야 과제 투자액 기준으로 지원 경향성을 분석한 결과, 수행주체 및 연구개발단계의 성격에 따라 집중적으로 지원중인 세부 기술이 상이함을 확인할 수 있었다.

중소기업은 SoC 분야 과제 수행 비중이 41.8%로 타 분야 대비 상대적으로 높았으며, 다음으로 소재·장비 및 NPU 분야가 각각 19.8%, 14.3%를 차지하였다. 중소기업이 중점적으로 추진중인 세 분야는 개발연구 단계에서 높은 비중을 차지하는 세부기술 분야로써 AI반도체 분야 과제 수행은 대체로 제품화·사업화를 염두하고 있는 것으로 해석할 수 있었다(<표 2-19>). 중견기업은 소재·장비 분야가 45.3%, 뒤를 이어 NPU, SoC 분야가 각각 26.8%, 17.5%의 비중을 차지하였으며, 대기업은 NPU, SW 분야 과제 총 8개만 수행한 것으로 나타났다. 중견기업에서 중점적으로 추진 중인 세부 분야는 대체로 개발 및 응용분야에 집중되어 있었으며, 상용화 연계 지원이 이루어지고 있는 것으로 파악되었다. 대기업은 인공지능 딥러닝 프로세서·모듈, 서버용 SW 프레임워크 개발, 시스템SW 프레임워크 개발 등 시스템 단위의 상위 분야 기술개발을 수행중이다. 중소·중견을 포함한 산업계 수행 과제 중 SoC 분야의 투자비중이 가장 높긴 하였으나 '22년 이후 그 규모가 담보 상태인 반면, 소재·장비 분야는 지속적으로 투자 규모가 확대중인 점을 고려하였을 때 산업계 중심의 AI반도체 분야 핵심 소재·장

비의 내재화를 위한 지속 투자 경향성을 확인할 수 있었다.

한편, 대학의 중점 수행 분야는 소자공통기술로 총 투자액의 38.6%를 차지하였으며, 뒤를 이어 설계 및 SW가 각각 18.8%, 14.2%를 차지하였다. 소자공통기술 및 설계 분야는 원천성에 따라 기초연구의 비중이 높은 세부기술로써 대학의 인적·물적 특수성을 기반으로 한 연구를 집중적으로 수행중인 점을 확인할 수 있었다. 또한, 대학은 주관기관으로써 수행주체 중 가장 많은 수의 과제(총 327개)를 수행한 것으로 나타나, 기술개발의 다양성 및 참여 수준은 높은 것으로 판단할 수 있었다. 다만, 대학의 수행 규모는 '19년 이후 '22년까지 증가 추세였으나, '23년 이후 답보·축소 중인 실정으로, AI반도체 분야의 전반적인 투자 비중 측면에서 감소하는 경향으로 나타났다.

국공립·출연연은 비중에 차이가 있지만, 대학과 유사하게 설계(22.3%), 소자공통기술(18.3%), SW(11.0%)을 중심의 연구개발을 수행하고 있다. 이 외에도 PIM(10.9%), NPU(10.3%), SoC(8.5%) 등 소자 기술 분야 과제를 전반적으로 참여하고 있는 것으로 나타났다. 국공립·출연연의 투자 규모는 '19년 이후 지속적으로 확대중이나, 주관 기관으로써 과제 수행 수는 총 100개 미만으로 산업계와 대학 대비 절대량은 적은 것으로 확인하였다.

<표 2-19> 수행주체별 중점 추진 중인 세부기술(상위 3개 노출)

구분	2019		2020		2021		2022		2023		합계	
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제
중소기업	2,615	2	50,324	46	65,482	68	106,539	103	120,089	100	345,049	319
SoC	2,615	2	23,531	25	31,400	34	43,927	43	42,877	36	144,350	140
소재·장비	-		9,700	10	10,650	11	20,620	22	27,320	24	68,290	67
NPU	-		10,200	6	8,152	7	13,838	12	17,093	13	49,283	38
기타	-		6,893	5	15,280	16	28,154	26	32,799	27	83,126	74
중견기업	-		6,450	5	20,534	16	28,824	23	25,999	21	81,806	65
소재·장비	-		3,800	3	5,950	5	15,140	13	12,180	11	37,070	32
NPU	-		1,650	1	6,550	3	6,850	3	6,850	3	21,900	10
SoC	-				5,437	5	4,707	5	4,169	4	14,313	14
기타	-		1,000	1	2,597	3	2,127	2	2,800	3	8,523	9
대기업	-		3,600	1	5,600	2	7,989	2	9,789	3	26,978	8
NPU	-		3,600	1	3,600	1	5,300	1	7,100	2	19,600	5
SW	-				2,000	1	2,689	1	2,689	1	7,378	3

구분	2019		2020		2021		2022		2023		합계	
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제
대학	5,818	5	16,871	46	40,437	78	71,617	112	53,303	86	188,045	327
공통기술	-		7,063	26	20,196	48	29,317	59	17,945	40	74,521	173
설계	318	1	2,826	8	4,293	11	13,740	22	14,165	18	35,342	60
SW	4,000	1	3,837	1	5,497	4	6,958	4	6,423	4	26,714	14
기타	-		3,145	11	10,451	15	21,602	27	14,770	24	51,468	80
국공립·출연연	4,200	2	11,749	15	18,412	23	29,158	30	39,164	28	102,683	98
설계	-		1,052	1	1,052	1	7,552	3	13,262	7	22,918	12
공통기술	-		1,250	4	6,076	12	7,500	12	4,003	8	18,829	36
SW	1,800	1	1,727	1			1,570	1	6,234	3	11,331	6
기타	-		7,720	9	11,284	10	12,536	14	15,665	10	49,605	44
기타	-		4,365	3	4,320	3	7,505	7	9,695	13	25,885	26
합계	12,633	9	93,359	116	154,784	190	251,631	277	258,039	251	770,446	843

연구개발단계별 중점 지원중인 세부기술에도 차이가 존재하였다. 대학 및 출연연의 참여 비중이 높은 기초연구¹¹⁹⁾단계는 AI반도체 공통기술의 비중이 58.1%로 과반을, PIM반도체와 설계가 각각 15.7%, 14.4%를 차지하는 등 원천성이 높은 분야를 중심으로 기초연구가 추진중인 점을 확인할 수 있었다. 다만, 기초연구의 연간 투자액은 '22년까지 지속적으로 증가하였으나, '23년 이후 감소하는 경향으로, 응용·개발연구 단계 대비 투자 비중이 지속적으로 하향 중인 것을 확인할 수 있었다. 한편, 응용연구¹²⁰⁾는 설계(26.0%), NPU(18.8%), SoC(15.9%) 순으로 지원중이며, 중장기적인 관점에서 상용화에 요구되는 소자 및 설계기술 중심으로 연구가 추진 중인 것으로 나타났다. 마지막으로 개발연구 단계는 SoC 33.4%, 소재·장비 20.2%, NPU 15.1%를 차지하는 등 기술·산업 수요를 고려, 산업계 주도¹²¹⁾의 단·중기적인 제품·상용화를 고려한 과제 수행이 이루어지고 있는 것을 확인할 수 있었다. 대표적인 SoC·NPU 품목으로는 비전 센서, 통신 모뎀, 바이오헬스 응용을 위한 반도체 등, 소재·장비 중에는 증착, CMP, 식각 및 테스트 장비 등을 포함하고 있다.

119) AI반도체 분야 기초연구의 참여비중은 대학(77.6%) > 출연연(16.8%) > 중소·중견기업(3.1%) 순

120) AI반도체 분야 응용연구의 참여비중은 중소기업(42.4%) > 대학(21.8%) > 출연연(19.0%) 순

121) AI반도체 분야 개발연구의 참여비중은 중소기업(59.3%) > 중견기업(16.9%) > 대학(13.1%) 순

<표 2-20> 연구개발단계 별 중점 세부기술(상위 3개 노출)

구분	2019		2020		2021		2022		2023		합계	
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제
기초연구	800	2	8,334	25	24,492	51	39,553	71	28,446	61	101,624	210
공통기술			4,034	16	15,442	37	23,356	46	16,177	38	59,008	137
PIM					4,000	2	6,625	8	5,293	11	15,918	21
설계			1,152	2	1,873	5	6,375	10	5,273	8	14,673	25
기타			3,148	7	3,177	7	3,197	7	1,703	4	12,025	27
응용연구	2,400	1	34,281	41	45,170	45	100,252	91	92,904	66	275,006	244
설계			5,364	5	6,655	6	26,663	21	32,869	24	71,551	56
NPU			8,250	5	11,850	6	15,789	8	15,729	8	51,618	27
SoC	2,400	1	7,200	9	4,500	6	14,543	11	15,200	7	43,843	34
기타			13,467	22	22,165	27	43,257	51	29,106	27	107,994	127
개발연구	7,718	5	44,714	44	81,289	90	107,780	109	130,442	117	371,943	365
SoC	1,600	2	14,105	16	32,767	34	37,464	40	38,446	36	124,382	128
소재·장비			6,900	7	10,700	11	20,750	18	36,890	35	75,240	71
NPU			12,850	6	12,102	8	15,141	11	16,014	11	56,107	36
기타			10,859	15	25,720	37	34,425	40	39,092	35	116,214	130
기타	1,715	1	6,031	6	3,834	4	4,046	6	6,247	7	21,873	24
합계	12,633	9	93,359	116	154,784	190	251,631	277	258,039	251	770,446	843

다. 정부R&D 투자 동향 및 현황

본 장에서는 AI 반도체 분야 정부R&D 투자동향을 살펴보기 위해 과기정통부, 산업부 등 주요 주관부처의 반도체 분야 전용 R&D사업을 식별하고, 해당 과제에서 지난 5~6년간 지원한 R&D과제를 바탕으로 세부적인 분석을 추진하였다. AI반도체에 앞서 반도체 분야 전반에 대한 투자 경향을 살펴본 결과, 지난 '18~'23년간 반도체 분야 총 투자액의 60%를 과기정통부에서, 40%는 산업부에서 지원한 것으로 확인할 수 있었다. 과거(18년) 과기정통부 주도로 추진하였으나, 다부처 대형사업 추진 및 지원 분야 다양화, 기업지원 활성화 등에 따라 산업부의 참여 비중이 점차 증가하여 '23년 대략 과기정통부 55%, 산업부 45% 비중을 나타내고 있다. 한편, 정부 반도체 R&D는 대표적으로 기술개발, 기업지원, 인력양성·국제협력, 인프라 구축 등 분야로 구분할 수 있으며, 그간 과기정통부는 인력양성 및 인프라 구축에 집중, 산업부는 기업지원 분야에 집중해 온 것으로 확인할 수 있었다. 그간 반도체 분야는 총 투자액의 30%를 산업계 주도의 개발연구가 주도하였으며, 대학의 기초연구는 10%, 국공립·출

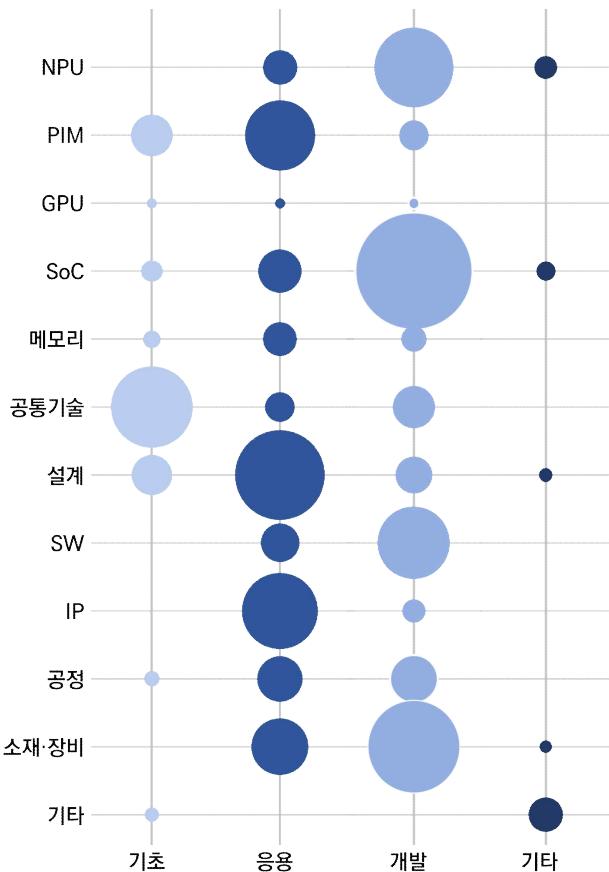
연연의 개발연구가 14.7%를 차지하는 등 수행주체별 연구개발 단계가 명확히 구분되는 특징을 보여주었다.

한편, AI반도체는 '19년부터 본격적으로 전용사업 기반 투자가 시작되었으며, 전체 반도체 분야와 달리 기술개발 및 기업지원 분야를 중심으로 지원이 이루어지고 있다. 대표적인 다부처 예타 사업인 「차세대지능형반도체기술개발」 및 「PIM인공지능반도체핵심기술개발」 사업이 그간 AI반도체 전체 투자액의 약 76.5%¹²²⁾를 차지하는 등 절대적으로 영향력을 보여주었다. 반도체 전반의 투자 동향과 달리 AI반도체는 과기정통부 주도(62.3% 비중)로 추진 중이며, 기술개발 분야가 92.2%를 차지하고 있다. 그간 AI반도체 분야의 중소·중견·대기업 참여 비중은 약 58.9%¹²³⁾이며, 산업계 주도의 개발연구 비중은 38.2%로 전체 반도체 분야 내 비중(30.0%)를 다소 상회하는 것으로 나타났다. 또한, 과기정통부에서 추진 중인 AI반도체 분야 내 기업지원 사업의 수행주체가 100% 기업인 점을 고려하였을 때, 단기 및 중·장기 제품·상용화 중심의 정부 투자가 이루어지고 있는 점을 확인할 수 있었다. 또한, AI 반도체 분야를 다시 3개의 중점분야와 11개의 세부기술(표9)로 구분하여 심층분석 한 결과, 산업계는 도메인 별 SoC, 소재·장비, NPU 등 제품화 중심의 응용·개발 과제를 수행중인 것으로 나타났으며, 대학과 국공립·출연연은 소자공통기술, 설계 등 원천성이 높은 분야 과제를 지속적으로 수행 중인 것으로 나타났다.

122) '19~'23년 AI반도체 분야 정부R&D 사업 기준 총 투자액 7,704억원 중 5,893억원을 차지

123) 중소기업 44.8%, 중견기업 10.6%, 대기업 3.5%

[그림 2-27] (참고) 세부기술 별 연구개발 단계 비중



2. 주요 정책

대한민국 정부의 산업정책에서 AI 반도체 산업에 대한 전략은 2019년 「시스템반도체 비전과 전략」에서 처음 등장했으며, 2020년 「인공지능 반도체 산업 발전전략」부터 본격화되었다. 정부에서 발표된 주요 전략과 이니셔티브를 통해 AI 반도체 산업정책의 주무 부처는 주로 과학기술정보통신부와 산업통상자원부임을 알수 있다. 2022년에는 「첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트」를 통해서 새롭게 중소벤처기업부가 시스템 반도체(AI 반도체) 펩리스 기업을 정책적으로 지

원하기 시작하였다. 각 부처는 주로 연구개발, 인프라 지원 등의 기술 부문, 인력양성과 공급을 중심으로 한 인재 부문, 수요연계, 협력체계 구축 등의 기업 경쟁력 및 생태계 부문, 정부 펀드 등의 투자 부문 정책을 통해서 AI 반도체 산업을 성장을 지원하고 있다. 본 연구에서는 주요 전략과 이니셔티브의 하위 전략과 과제를 간략하게 기술(R&D), 기업·생태계(인프라), 투자, 인재로 구분하여 살펴보았다. 이를 통해, 정부의 주요 전략과 이니셔티브 전략적 고려사항에 대한 이해도를 높일 수 있다.

가. 시스템반도체 비전과 전략(관계부처 합동, 2019.5.1)¹²⁴⁾

2019년 관계부처 합동으로 ‘메모리반도체 강국에서 종합 반도체 강국으로 도약’을 비전으로 하여 ‘시스템반도체 비전과 전략’ 발표하였다. 이 산업전략은 대한민국의 시스템반도체 산업의 한계를 기업 성장기반수요확보 부족, 팹리스-파운드리 연계 미흡, 기술생산 역량 취약 등 산업생태계 전반의 경쟁력 부족의 측면에서 지적하였고, 이를 극복하기 위한 세부전략을 팹리스, 파운드리, 상생협력, 인력 그리고 기술 측면에서 제시하였다.

<표 2-21> 시스템반도체 비전과 전략

주무부처	관계부처 합동	
비전	메모리반도체 강국에서 종합 반도체 강국으로 도약	
전략	산업의 패러다임을 바꾸는 차세대반도체 핵심기술 확보	
과제	AI 반도체 등 미래 반도체시장을 좌우할 Next-Generation 반도체 개발	기술
	산·학·연 공동의 혁신적 연구거점 구축 통합연구 추진	기술
	특히 빅데이터 활용, 미래 유망 반도체기술 발굴	기술

출처: 관계부처합동(2019.5.1.)

나. 인공지능(AI) 반도체 산업 발전전략(관계부처 합동, 2020.10.12.)¹²⁵⁾

2020년 10월 12일 정부는 관계부처합동(과학기술정보통신부, 산업통상 자원부)으로 ‘인공지능(AI) 반도체 선도국가 도약으로 인공지능·종합반도체 강국 실현’을 비전

124) 시스템반도체 비전과 전략(관계부처 합동, 2019.5.1.)의 포괄적 인용 및 요약

125) 인공지능(AI) 반도체 산업 발전전략(관계부처 합동, 2020.10.12.)의 포괄적 인용 및 요약

으로 하여 인공지능(AI) 반도체 산업 발전전략을 발표하였다. 이 전략의 가장 큰 특징은 AI 반도체에 대한 본격적인 산업 전략이라는 점이다. 반도체 부문의 기존 정부 산업 전략은 시스템 반도체 전략의 하위로 AI 반도체 정책을 설명하였지만, 본 산업전략은 AI 반도체를 중심 대상으로 하고있다. 이 전략은 기존의 기술(R&D)부문만이 아니라 인재, 기업경쟁력 및 생태계, 투자 부문에 대한 정책을 제안하여, 연구개발 뿐만 아니라 기술사업화 전주기의 관점을 통해 산업 전반의 발전을 도모한다는 점에서 의미가 있다.

<표 2-22> 인공지능(AI) 반도체 산업발전 전략 요약

주무부처		관계부처 합동	
비전		AI반도체 선도국가 도약으로 AI·종합반도체 강국 실현	
전략		선도자형 혁신기술·인재 확보	
과제	인공지능(AI) 반도체 플래그십 프로젝트 추진으로 글로벌 최고 기술 리더십 확보	기술	
	세계 1위 메모리 역량으로 신개념 PIM 반도체 초격차 기술에 도전	기술	
	국가 인공지능·데이터댐 기반에 인공지능 반도체 시범 도입·실증	기업·생태계	
	민·관 공동투자, 선도대학 육성으로 '30년 고급인재 3,000명 양성	인재	
전략		혁신성장형 산업 생태계 활성화	
과제	1사 1칩(Chip) 프로젝트를 통해 '30년까지 수요 맞춤형 인공지능(AI)칩 50개 출시	기업·생태계	
	기업간 연대·협력으로 인공지능(AI) 반도체 설계 역량 강화 + 공정혁신밸리 조성	기업·생태계	
	AI반도체 혁신기업 스케일업(Scale-up)을 촉진하는 대규모 뉴딜펀드 지원	투자	
	30년 혁신기업 20개사 육성을 위한 “인공지능 반도체 혁신설계센터” 신규 구축	기술, 기업·생태계	

출처: 관계부처 합동 (2020.10.12.)

다. 인공지능(AI) 반도체 산업 성장 지원대책(과학기술정통부, 2022.6.27.)¹²⁶⁾

2022년에 과학기술정보통신부에서 발표한 인공지능(AI) “반도체 산업 성장 지원대책”은 “AI반도체 시장선점 및 강국 도약”을 비전으로 하고 있으며 AI 반도체 산업을 정책대상으로 하고 있다. 이 전략의 특징으로는 기술 부문에서는 기존의 HW 뿐만 아니라 SW 연구개발에 대한 과제가 논의되며 시작하였다는 점이다. 기업경쟁력 및 생태계 부문 측면에선 정부 수요측면의 공공조달을 통해서 수요를 창출하고 기업의 경쟁력을 강화시키는 과제들이 본격적으로 등장했다는 점을 역시 들 수 있다. 인재부문이

126) 인공지능(AI) 반도체 산업 성장 지원대책(과학기술정통부, 2022.6.27.)의 포괄적 인용 및 요약

가장 특징적인데 하나의 세부전략으로서 인재부문이 대두되었고 관련된 정책과제도 3개가 제시되어, 기존의 전략에 비해 인재 부문 정책을 구체화하였다는 점 역시 주목할 점이다.

<표 2-23> 인공지능(AI) 반도체 산업 성장 지원대책 요약

주무부처	과학기술정보통신부	
비전	AI반도체 시장선점 및 강국 도약	
전략	인공지능 반도체 초격차 기술력 확보	
과제	인공지능 반도체 초격차 기술 선점을 위한 원천기술 개발 강화	기술, (공동 기반 조성을 통한 연구자 생태계)
	국산 인공지능 반도체 성능을 극대화 하는 SW 기술력 강화	
	인공지능 반도체 기술 선도국과의 협력 기반 마련	
전략	인공지능 반도체 초기 시장수요 창출	
	국산 인공지능 반도체 대형 테스트베드 구축	기업·생태계
과제	인공지능 반도체 적용 제품서비스 실증	기업·생태계
	공공분야 전반으로 국산 인공지능 반도체 도입 확산	기업·생태계
전략	인공지능 반도체 산학연 협력 생태계 조성	
과제	반도체 대기업과 산학연 간 협력 강화	기업·생태계
	유망기업 경쟁력 강화 및 상용화 지원	기술, 기업·생태계
	산학연 협력체널 구성	기업·생태계
전략	인공지능 반도체 전문인력 양성	
과제	공공 연구 인프라를 활용한 현장중심 교육 강화	인재
	연구 중심의 석박사 고급인력 양성	인재
	실무형 학부 인재 양성	인재

출처: 과학기술정보통신부(2022.6.27.)

라. 첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트 (중소벤처기업부, 2022.11)¹²⁷⁾

2022년 11월 중소벤처기업부는 “첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트”를 발표했다. “독보적인 기술우위를 바탕으로 글로벌 시장을 선도하는 초격차 스타트업 육성”을 비전으로 하였고 이를 위한 전략으로 “5년간, 초격차에 도전하는 1,000+개의 스타트업 선발·육성”, ①분야·②스타트업선정 → ③핵

127) 첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트(관계부처 합동, 2022.11)의 포괄적 인용 및 요약

심수단투입 → ④전문기관매칭 → ⑤스케일업·글로벌화+⑥부처협업·연계“를 통한 초격차달성”을 내세웠다. 전략을 달성하기 위한 과제로 먼저, “10대 초격차 분야 선정”을 과제로 하였고 10대 초격차 분야 중 하나가 시스템 반도체이다. 단, 본 전략은 AI 반도체 및 시스템 반도체만을 정책대상으로 한다고 보기는 어렵다는 한계가 있다. 기존의 AI 반도체에 대한 산업 전략과 정책들이 주로 과학기술정보통신부와 산업통상자원부 중심으로 만들어지고 집행되어 왔지만, 이 전략을 통해 중소벤처기업부도 시스템 반도체 스타트업을 정책대상으로 삼았다는 특징이 있다. 투자 부문, 기업경쟁력 및 생태계 부문, 기술 부문의 정책이 주요하게 나타나 기업의 성장과 스케일업을 도모하는 전략인 것도 특징점이다.

<표 2-24> 첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트

주무부처	중소벤처기업부	
비전	독보적인 기술우위를 바탕으로 글로벌 시장을 선도하는 초격차 스타트업 육성	
전략	<ul style="list-style-type: none"> - 5년간, 초격차에 도전하는 1,000+개의 스타트업 선발·육성 - ①분야·②스타트업 선정 → ③핵심수단투입 → ④전문기관매칭 → ⑤스케일업·글로벌화+⑥부처협업·연계“를 통한 초격차 달성 	
과제	10대 초격차 분야 선정	기업·생태계,
	1,000개 이상 유망 스타트업 선별	기업·생태계,
	5년간 2조원의 과감한 민·관 공동자금 투입	투자
	정부출연연구소 등 전문기관의 특화지원	기업·생태계, 기술
	스케일업 및 글로벌화	기업·생태계, 기술, 투자
	부처 협업 및 연계	기업·생태계

출처: 중소벤처기업부(2022.11.)

마. 국산 AI반도체를 활용한 『K-클라우드』추진방안(과학기술정보통신부. 2022.12.12.)¹²⁸⁾

과학기술정보통신부는 2022년 12월 22일 “메모리 반도체 기술력을 바탕으로 글로벌 최고 수준의 초고속 저전력 AI 반도체 개발에 승부를 걸어 국내 클라우드 경쟁력을

128) 국산 AI반도체를 활용한 『K-클라우드』추진방안(과학기술 정통부. 2022.12.12.)의 포괄적 인용 및 요약

혁신적으로 개선”을 비전으로 하는 “국산 AI반도체를 활용한『K-클라우드』추진방안”을 발표했다. “K-클라우드 프로젝트”는 “세계 최고 수준의 초고속·저전력 국산 AI반도체 개발과 데이터센터 적용을 통해 국내 클라우드 경쟁력을 강화하고 국민들에게 향상된 AI서비스를 제공하는 프로젝트”로서 AI 반도체 산업에 관련된 정책이다. 이 전략의 특징은 기술 부문에서 새로운 기술개발의 목표를 명확화 했다는 점이다. 3단계에 걸쳐서 기존의 NPU와 PIM 반도체의 고도화를 정책 과제로 추진한다는 점에서 AI 반도체 기술 부문을 특화한 전략으로 볼수 있다. 더불어, 실증과 수요 연계 측면에서 국가가 AI 반도체 팝리스 스타트업을 지원한다는 점에서 기존의 전략과 차별점이 있다.

<표 2-25> 국산 AI반도체를 활용한『K-클라우드』추진방안

추진체계		과학기술정보통신부
비전		메모리 반도체 기술력을 바탕으로 글로벌 최고 수준의 초고속·저전력 AI 반도체 개발에 중점을 걸어 국내 클라우드 경쟁력을 혁신적으로 개선
전략		국산 AI 반도체 고도화
과제	1단계 NPU 자속고도화 및 데이터 센터 적용 2단계: 저전력 PIM 개발 3단계: 극저전력 PIM 및 Smart SSD 개발	기술
전략		AI반도체용 SW 개발
과제	AI반도체를 데이터센터에 적용 하기 위한 SW 추가개발을 위한 신규 예타 사업 추진	기술
전략		데이터센터 실증 및 AI서비스 제공
과제	구축사업과 기존의 AI·클라우드 서비스 개발 사업을 연계	기업·생태계
	파급력·수요가 높은 4대 분야에 서비스 선도 적용, 주요분야로 확대	
전략		산·학·연 협력 강화를 위한 추진체계를 마련
과제	K-클라우드 얼라이언스 구축	기업·생태계, 기술
	PIM-HUB 역할 강화	
	R&D 및 지원조직 강화	
	AI 반도체 대학원	인재

출처: 과학기술정보통신부(2022.12)

바. AI 반도체 이니셔티브(관계부처 합동, 2024.4.25)¹²⁹⁾

2024년 4월 35일 관계부처 합동으로 발표한 AI 반도체 이니셔티브는 “인공지능

129) AI 반도체 이니셔티브(관계부처 합동, 2024.4.25.)의 포괄적 인용 및 요약

G3 도약, K-반도체 새로운 신화 창조”를 비전으로 하여 “AI 기술패권 선도”, “AI 반도체 초격차·신격차”, “AI-반도체 HW·SW 기술생태계”라는 전략 하에 AI-반도체 9대 기술혁신 과제를 제시하였다. “AI 기술패권 선도” 하 과제들은 주로 AI 알고리즘 및 모델과 관련되어 있기 때문에 본 연구에서는 다루지 않는다. 이 전략의 특징은 기술 부문에서 온디바이스 AI등의 동 산업의 주요 기술을 지원하는 정책을 새롭게 추진한다는 점이다. 투자 부문에서도 반도체 생태계 펀드 등의 펀드 구성을 계획했다는 점에서 특징이 있다. 더불어, 글로벌 협력에서 기존의 공동연구 및 기술협력 강화와 같은 정책에 덧붙여 국내기업과 글로벌 빅테크간의 협력관계 구축지원 및 주요기업 제품의 글로벌 경쟁력 강화를 위한 수입국(국산 AI반도체) 현지화 실증과 같은 수요자 맞춤형의 정책을 추진한다는 점도 주목할 점이다.

<표 2-26> AI-반도체 이니셔티브 (2024.4)

주무부처	관계부처 합동	
비전	인공지능 G3 도약, K-반도체 새로운 신화 창조	
전략	AI 반도체 초격차신격차	
과제	메모리혁신 : Processing in Memory (PIM) 등	기술
	한국형 AI프로세서 : 저전력 K-AP 등	기술, 기업·생태계
	신소자&첨단 패키징	기술
전략	AI-반도체 HW·SW 기술생태계	
과제	AI 슈퍼 컴퓨팅 : 「K-클라우드 2.0」 등	기술, 기업·생태계
	온디바이스 AI	기술, 기업·생태계
	차세대 개방형 AI 아키텍처SW	기술
전략	중점 추진과제	
과제	AI-반도체 분야 전방위적인 투자·금융 지원 추진	투자, 기업·생태계
	AI-반도체 산업을 이끌 혁신인재 양성	인재
	AI-반도체 산업·연구 혁신 인프라 구축	기술, 기업·생태계
	AI-반도체 글로벌 경쟁력 강화를 위한 글로벌 협력·진출 확대	기술, 기업·생태계

출처: 관계부처 합동, (2024.4.25)

3. 지원 현황

과기정통부와 산업부는 인공지능 시대 반도체 패러다임 전환기에 세계 반도체 시장 경쟁 우위를 위한 소자, 설계, 공정, 장비 기술 개발을 위한 대형 사업으로 ‘차세대지능형반도체기술개발’을 추진하고, 사업 기간 10년간(‘20년~’29년) 약 1조원의 투자를 추진하고 있다. ’19년 4월 ‘산업경쟁력 향상을 위한 기술개발 필요성이 인정되며, 중장기 기술경쟁력 확보라는 전략적 관점에서 추진당위성이 존재’의 결과에 따라 예비 타당성 조사가 통과하였으며, ’20년 7월부터 본격적인 과제 선정 및 사업을 개시한 바 있다. ‘차세대지능형반도체기술개발’ 사업은 목표·역할에 따라 ‘소자’, ‘설계’, ‘설계·제조’ 총 3개의 세부사업으로 구성되었으며, 공동으로 사업단을 운영하고 있다.

<표 2-27> 「차세대지능형반도체기술개발」사업 투자 현황(‘20~’24)

구분	2020		2021		2022		2023		2024		합계
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	
(과기정통부) 차세대지능형 반도체기술개발(소자)	12,000	47	33,977	81	47,077	95	28,257	60	21,168	-	142,479
(과기정통부) 차세대지능형 반도체기술개발(설계)	24,448	14	23,616	13	29,131	16	32,261	19	19,328	-	128,784
(산업부)차세대지능형 반도체기술개발 (설계·제조)	46,700	46	62,988	64	67,603	70	73,297	71	62,907	-	313,495
합계	83,148	107	120,581	158	143,811	181	133,815	150	112,002	-	593,357

※ ‘20~’23년 투자액은 NTIS 내 지원 과제의 총액, ’24년은 과제 협약 시점 등을 고려 국회안 예산 반영

먼저, 과학기술정보통신부 1차관에서 추진하는 ‘차세대지능형반도체기술개발(소자)’ 사업은 고성능, 초저전력 연산처리를 위한 신소자 원천기술과 신소자를 구현하여 성능을 검증 및 시스템을 설계하는 집적/검증기술 개발을 위한 ①신소자 원천기술개발 ②집적/검증기술개발 ③신개념 소자 기초기술개발 등 내역사업으로 이루어져 있다. 세부 지원 내용으로는 ①신소자원천기술개발 저전압, 정보밀도 개선, 미세 전류제어, 3차원 집적, 두뇌모사 등 초저전력에서 고성능으로 작동하는 반도체 소자 기술, ②신소자집적·검증

기술개발) 신소자의 조기 상용화 및 원천 설계 IP 확보가 가능하도록 웨이퍼레벨에서 신소자 성능을 검증하는 집적소자·검증 기술, (③신개념소자기초기술개발) 미래 반도체 패러다임 선도를 위해 새로운 원리, 신소재, 신공정 등 창의적 도전적 아이디어에 기반한 혁신적 소자 기초기술 등을 집중적으로 지원하고 있다.

지난 4년간 동 사업을 통해 총 1,213억 원의 정부 투자가 이루어졌으며, 연간 최대 90개의 과제(322억 원)를 지원한 바 있다. 내역사업 별로는 ①신소자원천기술개발이 전체의 67.2%를 차지하였으며, 내역사업②,③은 각각 20.0%, 10.7%를 차지하였다. 그간 투자액을 기준으로 55.7%가 기초연구에 집중되어 있었으며, 수행주체로는 대학의 비중이 79.3%(961억 원)로 대다수를 차지하는 등 차세대 AI반도체 소자의 요소기술 개발을 지향하는 동 내역사업의 중점 투자분야를 확인할 수 있었다. 이와 같은 특성에 따라 전체 예타 사업 구성에서 동 세부사업의 산업계 참여 비중은 중소기업 1개소 수준으로 가장 낮은 것으로 나타났다. 한편, 그간 수행주체의 지역으로는 서울특별시와 대전광역시가 각각 45.0%, 25.8%를 차지하는 등 전국대·경희대·고려대·서울대 및 한국과학기술원 서울·대전 권역의 주요 사립·국립대 중심으로 투자가 이루어진 것을 확인할 수 있었다.

<표 2-28> 차세대지능형반도체기술개발(소자) 과제 지원 현황

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
내역사업	내역사업①	7,750	33	21,568	53	32,246	62	19,942	34	81,506	182
	내역사업②	3,000	6	8,250	10	10,000	10	3,000	3	24,250	29
	내역사업③	700	7	3,400	17	4,301	22	4,570	22	12,971	68
	기타	550	1	759	1	530	1	745	1	2,584	4
연구개발 단계	기초연구	5,418	21	17,622	45	26,471	55	18,044	43	67,554	164
	응용연구	3,738	15	8,134	17	9,523	19	6,244	10	27,638	61
	개발연구	2,295	10	7,463	18	10,553	20	3,224	6	23,535	54
	기타	550	1	759	1	530	1	745	1	2,584	4
수행주체	대학	9,865	40	26,472	66	37,508	79	22,297	49	96,142	234
	출연연구소	2,135	7	7,505	15	8,811	15	4,003	8	22,454	45
	중소기업	-	-	-	-	758	1	1,012	1	1,770	2
	기타	-	-	-	-	-	-	945	2	945	2
지역	서울특별시	3,972	19	14,069	38	20,701	45	15,807	32	54,550	134
	대전광역시	3,980	13	9,605	18	12,607	20	5,115	9	31,306	60
	경상북도	1,607	4	3,944	5	4,780	6	1,367	4	11,698	19

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
경기도	801	3	1,582	3	2,482	4	2,507	3	7,372	13	6.1
울산광역시	890	3	1,984	5	2,777	6	1,198	3	6,849	17	5.6
대구광역시	150	1	1,107	5	1,266	5	1,246	4	3,769	15	3.1
인천광역시	436	2	1,006	3	1,312	3	200	1	2,953	9	2.4
충청북도	-	-	200	1	367	2	417	2	984	5	0.8
부산광역시	65	1	280	2	352	2	200	1	897	6	0.7
경상남도	100	1	200	1	233	1	-	-	533	3	0.4
광주광역시	-	-	-	-	200	1	200	1	400	2	0.3
합 계	12,000	47	33,977	81	47,077	95	28,257	60	121,311	283	-

<표 2-29> 차세대지능형반도체기술개발(소자) 지원과제 참여 기관

구분	참여 기관
산	• (중소기업) (주)알세미
학	• 건국대학교, 경북대학교, 경희대학교, 고려대학교, 국민대학교, 대구경북과학기술원, 동국대학교, 부산대학교, 삼육대학교, 서강대학교, 서울대학교, 성균관대학교, 아주대학교, 연세대학교, 울산과학기술원, 이화여자대학교, 인하대학교, 전남대학교, 전북대학교, 중앙대학교, 충남대학교, 포항공과대학교, 한국과학기술원, 한국교원대학교, 한국교통대학교, 한양대학교
연	• 나노종합기술원, 한국과학기술연구원, 한국세라믹연구원, 한국전자통신연구원, 한국표준과학연구원, 한국화학연구원 등
기타	• 차세대지능형반도체사업단, 한국연구재단 등

다음으로 과학기술정보통신부 2차관에서 추진하는 ‘차세대지능형반도체기술개발(설계)’는 반도체 산업의 새로운 성장동력인 인공지능 반도체 핵심기술 개발을 통한 글로벌 수준의 기술경쟁력 확보 및 팝리스 육성을 위해 ①인공지능프로세서, ②지능형 반도체SW, ③초고속인터페이스 등 3개의 내역사업을 지원하고 있다. 세부 내용으로는 ①인공지능프로세서 주요 응용 분야인 서버·모바일·엣지 분야에 활용 가능한 고성능 저전력의 인공지능 프로세서(NPU), ②지능형반도체SW인공지능 반도체 구동 및 성능 최적화를 위한 시스템 소프트웨어(OS, 컴파일러, 라이브러리 등), ③초고속인터페이스 프로세서 간프로세서와 메모리 사이에서 대용량 데이터를 고속으로 전송하기 위한 초고속 인터페이스 기술개발 지원을 주요 골자로 하고 있다.

지난 4년간 동 사업에 총 1,095억 원의 정부 투자가 이루어졌으며, 연간 20개 이내의 과제에 대해 집중적으로 지원하고 있다. ’23년까지 총 사업비의 79.1%를 내역사업

①을 지원하였으며, HW 설계 → 시스템 SW 등 동 사업의 기술개발 체계를 고려하여 내역사업②는 '23년부터 2개 과제의 신규 지원을 추진하고 있다. 연구개발 단계로는 향후 상용화를 위한 응용·개발 연구의 비중이 각각 54.1%(592억 원), 35.7%(390억 원) 이루어졌으며, 수행주체로는 중소·중견·대기업 전반을 아우르는 산업계의 비중이 74.4%를 차지하고 있다. 특히, 참여 기업으로는 넥스트칩, 텔레칩스, 퓨리오사AI, 네페스, 사피온코리아 등 국내 주요 팝리스 및 패키징 업체를 포함하고 있어, 당초 동 사업이 제시한 상용화 중심의 국내 산업생태계 육성 등 목적으로 부합한 지원 현황을 확인할 수 있었다. 한편, 투자 지역으로는 서울특별시·경기도에 72.5% 비중은 나타난 점은 민간 참여 수준(74.4%) 및 소재지 특성을 반영한 것으로 나타났다.

<표 2-30> 차세대지능형반도체기술개발(설계) 과제 지원 현황

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
내역사업	내역사업①	18,893	9	18,893	9	24,408	12	24,408	12	86,602	42
	내역사업②	-	-	-	-	-	-	2,520	2	2,520	2
	내역사업③	4,552	4	4,723	4	4,723	4	4,723	4	18,721	16
	기타	1,003	1					610	1	1,613	2
연구개발 단계	응용연구	10,243	7	13,843	8	17,343	9	17,764	9	59,193	33
	개발연구	12,150	5	8,721	4	9,236	5	8,985	5	39,092	19
	기타	1,003	1	-	-	1,500	1	5,512	5	8,015	7
	기초연구	1,052	1	1,052	1	1,052	1	-	-	3,156	3
수행주체	중소기업	12,493	6	8,664	5	11,279	7	12,599	8	45,035	26
	중견기업	1,650	1	5,650	2	5,650	2	5,650	2	18,600	7
	대기업	3,600	1	3,600	1	5,300	1	5,300	1	17,800	4
	출연연구소	4,705	4	3,702	3	3,702	3	4,312	4	16,421	14
	대학	2,000	2	2,000	2	3,200	3	4,400	4	11,600	11
지역	서울특별시	13,600	7	13,771	7	11,771	7	13,091	8	52,233	29
	경기도	2,993	1	2,993	1	9,993	3	11,193	4	27,172	9
	대전광역시	4,705	4	3,702	3	4,217	4	4,827	5	17,451	16
	충청북도	1,650	1	1,650	1	1,650	1	1,650	1	6,600	4
	충청남도	1,500	1	1,500	1	1,500	1	1,500	1	6,000	4
합 계		24,448	14	23,616	13	29,131	16	32,261	19	109,456	62

<표 2-31> 차세대지능형반도체기술개발(설계) 지원과제 참여 기관

구분	참여 기관
산	<ul style="list-style-type: none"> (중소기업) (주)아이닉스, (주)넥스트칩, (주)딥엑스, (주)텔레칩스, (주)퓨리오사에이아이, 오픈엣지테크놀로지(주), (주)모빌린트, (주)퓨처디자인시스템 (중견기업) (주)네페스, (주)텔레칩스 (대기업) SK텔레콤(주), (주)사피온코리아
학	서울대학교, 성균관대학교, 연세대학교
연	한국전자통신연구원
기타	정보통신기획평가원

마지막으로 산업부에서 추진하는 ‘차세대지능형반도체기술개발(설계·제조)’는 시스템반도체 범용기술을 국내 주력산업(미래차, 바이오, 스마트가전, 첨단기계로봇)과 연계한 시스템반도체 상용화 및 차세대 제조 공정·장비 개발을 위해 ’20년~’26년 간 ①시스템반도체상용화설계, ②반도체제조공정장비 등 내역사업을 지원하고 있다. 세부적으로는 ①시스템반도체상용화설계(미래차, 바이오, 스마트가전, 첨단기계·로봇 분야 응용 시스템반도체 개발을 위한 SoC 제작 및 검증, ②반도체제조공정장비) 원자레벨전공정장비, 첨단 패키징 등의 첨단 반도체 제조에 필요한 차세대 공정장비·부품 개발 및 내재화를 중심으로 지원하고 있다.

지난 4년간 동 사업의 총 투자액은 2,506억 원 규모로, ’20년 46개에서 ’23년까지 총 71개의 과제를 지원하고 있다. 그간 내역사업 ①과 ②의 지원 비중은 각각 56.1%, 42.0%로 시스템반도체 상용화를 위한 설계 분야에 대한 투자 비중이 공정·장비 개발 대비 다소 높은 것으로 나타났다. 또한, 상용화를 지향하는 동 사업의 목적을 고려하여 개발, 응용연구에 각각 71.4%(1,790억 원), 24.7%(619억 원)의 집중 지원과 함께, 중소·중견기업을 포함한 산업계의 과제 수행 비중이 89.7%로 대다수를 차지하고 있다. 참여 기업으로는 ‘차세대지능형반도체기술개발(설계)’ 사업과 같이 가온칩스, 넥스트칩, 어보브반도체 등 국내 팹리스와 함께 메카로, 테스, 피에스케이 등 주요 장비사 역시 다수 포함하고 있다.

<표 2-32> 차세대지능형반도체기술개발(설계·제조) 과제 지원 현황

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
내역사업	내역사업①	27,000	27	38,100	39	38,503	43	37,097	38	140,700	147 56.1
	내역사업②	17,365	18	22,570	24	29,100	27	36,200	33	105,235	102 42.0
	기타	2,335	1	2,318	1	-	-	-	-	4,653	2 1.9
연구개발 단계	개발연구	22,600	24	40,650	44	52,993	55	62,757	59	179,000	182 71.4
	응용연구	20,300	19	18,500	17	13,163	13	9,940	10	61,903	59 24.7
	기타	2,800	2	2,738	2	410	1	600	2	6,548	7 2.6
	기초연구	1,000	1	1,100	1	1,037	1			3,137	3 1.3
수행주체	중소기업	34,400	37	45,200	49	47,684	53	54,177	53	181,461	192 72.4
	중견기업	4,800	4	10,350	10	15,448	13	12,720	11	43,318	38 17.3
	기타	4,365	3	4,320	3	3,780	3	4,700	5	17,165	14 6.8
	출연연구소	3,135	2	3,118	2	691	1	1,700	2	8,644	7 3.4
지역	경기도	2,465	3	2,320	3	3,290	4	53,937	52	62,012	62 24.7
	서울특별시	-	-	-	-	-	-	7,270	7	7,270	7 2.9
	대구광역시	2,335	1	2,318	1	-	-	1,300	1	5,953	3 2.4
	대전광역시	-	-	-	-	-	-	3,650	4	3,650	4 1.5
	경상북도	-	-	-	-	-	-	2,200	2	2,200	2 0.9
	전라북도	-	-	-	-	-	-	1,300	1	1,300	1 0.5
	충청남도	-	-	-	-	-	-	1,100	1	1,100	1 0.4
	기타*	41,900	42	58,350	60	64,313	66	2,540	3	167,103	171 66.7
	합 계	46,700	46	62,988	64	67,603	70	73,297	71	250,588	251 -

*NTIS 조사 결과 '20~'21년간 대다수의 수행주체 지역 구분을 '기타'로 정의

<표 2-33> 차세대지능형반도체기술개발(설계·제조) 지원과제 참여 기관

구분	참여 기관
산	<ul style="list-style-type: none"> (중소기업) (주)가온칩스, (주)네스랩, (주)넥스트칩, (주)노바칩스, (주)로그프레트코리아, (주)메카로, (주)바코솔루션, (주)베라세미콘, (주)비아트론, (주)사피엔반도체, (주)서보산전, (주)솔리드부, (주)스카이칩스, (주)씨자인, (주)아르고, (주)아이닉스, (주)아이브이티, (주)아이씨디, (주)아이엔씨테크놀로지, (주)아이에스피, (주)에어포인트, (주)에이디테크놀로지, (주)에이투유정보통신, (주)에이투테크, (주)에코에이블, (주)에프와이디, (주)엑시콘, (주)에스티, (주)옵토레이인, (주)유니젯, (주)유케어트론, (주)이미지스테크놀로지, (주)이에스티, (주)제4기한국, (주)지니ックス, (주)지오엘리먼트, (주)지투이, (주)케이씨텍, (주)포인트텍, (주)한백교역, (주)헬스리안, (주)호진플라텍, (주)힉스컴퍼니, 넥스트랩(주), 넷솔(주), 노바칩스, 대양전기공업(주), 디에이치(주), 브레이인소프트(주), 빌리브마이크론(주), 사피엔반도체, 세인플렉스(주), 수퍼게이트(주), 스카이칩스, 씨자인, 아스텔(주), 아이브이티, 아이언디바이스, 알에프코어(주), 어보브반도체(주), 에너지마이닝, 에어포인트, 에이투테크, 에이피에스리서치(주), 에프와이디, 엠클리미칼, 윤텔코퍼레이션(주), (주)뉴라텍, (주)라운에스, (주)램쉽, (주)메카로, (주)아르고, (주)에스디옵틱스, (주)이에스티, (주)헬스리안, (주)미코세라믹스, 지투이정보기술, 큐알티(주), 크라이오애이치앤아이, 피에스케이(주), 현대엠시스템즈(주), 황성공업(주) (중견기업) (주)AVACO, (주)서플러스글로벌, (주)아이언디바이스, (주)에프에스티, (주)케이씨텍, (주)테스, (주)톱텍, (주)프로텍, 대양전기공업(주), 대원산업(주), 어보브반도체(주), 에이피에스홀딩스(주), 와이아이케이(주), 와이아이케이(주), (주)오토실리콘, 주식회사티에스이, 톱텍, 피에스케이(주)
학	• 없음
연	• 한국전자통신연구원
기타	• 한국반도체연구조합, 한국전자기술연구원

그간 ‘차세대지능형반도체기술개발’의 과제 지원현황을 바탕으로 종합 분석한 결과, 개발 및 응용연구의 비중이 50.2%, 30.9%로 전체 정부R&D 투자 평균¹³⁰⁾을 상회하며, 사업·과제 종료시 산업 도메인별 상용화 또는 이를 실현할 수 있는 지원기술을 중심으로 투자가 이루어진 것을 확인할 수 있었다. 다만, NPU, 뉴로모픽 등 데이터센터·엣지컴퓨팅향 차세대 AI반도체의 원천기술 개발은 ‘차세대지능형반도체기술개발(소자)’을 통해 대학이 중점적으로 추진¹³¹⁾하는 것으로 나타났다. 특히, 동 사업의 수행 주체 비중은 중소(47.4%)·중견(12.9%)·대기업(3.7%)을 포함한 기업이 전체의 64.0%를 차지하는 등 전체 정부R&D 투자 평균¹³²⁾을 크게 상회하며 산업계의 높은 참여율을 확인할 수 있었다.

<표 2-34> 차세대지능형반도체기술개발(소자, 설계, 설계·제조) 과제 지원 현황

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
연구개발 단계	기초연구	7,470	23	19,774	47	28,560	57	18,044	43	73,847	170
	응용연구	34,281	41	40,477	42	40,029	41	33,948	29	148,734	153
	개발연구	37,045	39	56,834	66	72,782	80	74,966	70	241,627	255
	기타	4,353	4	3,497	3	2,440	3	6,857	8	17,147	18
수행주체	중소기업	46,893	43	53,864	54	59,721	61	67,788	62	228,266	220
	중견기업	6,450	5	16,000	12	21,098	15	18,370	13	61,918	45
	대기업	3,600	1	3,600	1	5,300	1	5,300	1	17,800	4
	대학	11,865	42	28,472	68	40,708	82	26,697	53	107,742	245
	출연연구소	9,975	13	14,325	20	13,204	19	10,015	14	47,519	66
	기타	4,365	3	4,320	3	3,780	3	5,645	7	18,110	16
	서울특별시	17,572	26	27,840	45	32,472	52	36,168	47	114,053	170
지역	경기도	6,259	7	6,895	7	15,765	11	67,637	59	96,556	84
	대전광역시	8,685	17	13,307	21	16,824	24	13,592	18	52,407	80
	경상북도	1,607	4	3,944	5	4,780	6	3,567	6	13,898	21
	대구광역시	2,485	2	3,425	6	1,266	5	2,546	5	9,722	18
	충청북도	1,650	1	1,850	2	2,017	3	2,067	3	7,584	9
	충청남도	1,500	1	1,500	1	1,500	1	2,600	2	7,100	5
	울산광역시	890	3	1,984	5	2,777	6	1,198	3	6,849	17
	인천광역시	436	2	1,006	3	1,312	3	200	1	2,953	9
	전라북도	1,000	1	1,000	1	1,000	1	1,000	1	4,000	4

130) '22년 기준 정부R&D 전체 과제의 연구개발 단계는 개발연구(46.8%) > 응용연구(25.2%) > 기초연구(28.0%) 순으로 나타남 ('22년 국가연구개발사업 조사분석 보고서, 2023)

131) 지난 4년간 기초연구 총 투자액 738억원 중 대학의 비중은 74.6%(551억 원)으로 나타나 원천연구를 주도

132) '22년 기준 정부R&D 전체 과제의 산업계 참여 비중은 중소기업(19.2%), 중견기업(6.5%), 대기업(2.1%) 수준이며, 대학(24.3%), 국공립·출연연(40.7%)를 차지 ('22년 국가연구개발사업 조사분석 보고서, 2023)

구분	2020		2021		2022		2023		합계			
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중	
전라북도	-	-	-	-	-	-	1,300	1	1,300	1	0.3	
	부산광역시	65	1	280	2	352	2	200	1	897	6	0.2
	경상남도	100	1	200	1	233	1	-	-	533	3	0.1
	광주광역시	-	-	-	-	200	1	200	1	400	2	0.1
	기타	41,900	42	58,350	60	64,313	66	2,540	3	167,103	171	34.7
합 계	83,148	107	120,581	158	143,811	181	133,815	150	481,355	596	-	

<표 2-35> 차세대지능형반도체기술개발(소자, 설계, 설계·제조) 과제 지원 현황

구분	2020		2021		2022		2023		합계		
	투자액	과제	투자액	과제	투자액	과제	투자액	과제	투자액	과제	비중
기초연구	7,470	23	19,774	47	28,560	57	18,044	43	73,847	170	15.3
대학	4,426	18	13,784	37	21,750	47	15,141	36	55,100	138	11.4
출연연구소	2,043	4	4,890	9	5,773	9	2,703	6	15,410	28	3.2
중견기업	-	-	1,100	1	1,037	1	-	-	2,137	2	0.4
중소기업	1,000	1	-	-	-	-	-	-	1,000	1	0.2
기타	-	-	-	-	-	-	200	1	200	1	0.0
응용연구	34,281	41	40,477	42	40,029	41	33,948	29	148,734	153	30.9
중소기업	17,293	16	16,293	15	15,801	14	16,336	13	65,723	58	13.7
대학	5,197	15	9,110	17	10,762	19	8,932	12	34,000	63	7.1
대기업			3,600	1	5,300	1	5,300	1	14,200	3	3.0
출연연구소	3,991	5	4,474	5	4,544	5	1,000	1	14,009	16	2.9
중견기업	4,800	4	4,000	3	1,030	1	1,380	1	11,210	9	2.3
기타	3,000	1	3,000	1	2,592	1	1,000	1	9,592	4	2.0
개발연구	37,045	39	56,834	66	72,782	80	74,966	70	241,627	255	50.2
중소기업	28,600	26	37,571	39	42,420	46	49,952	48	158,543	159	32.9
중견기업	1,650	1	10,900	8	19,031	13	16,990	12	48,571	34	10.1
대학	2,242	9	5,579	14	8,196	16	1,924	4	17,942	43	3.7
출연연구소	53	1	1,883	4	2,357	4	3,000	4	7,293	13	1.5
기타	900	1	900	1	778	1	3,100	2	5,678	5	1.2
대기업	3,600	1	-	-	-	-	-	-	3,600	1	0.7
기타	4,353	4	3,497	3	2,440	3	6,857	8	17,147	18	3.6
출연연구소	3,888	3	3,077	2	530	1	3,312	3	10,807	9	2.2
중소기업	-	-	-	-	1,500	1	1,500	1	3,000	2	0.6
기타	465	1	420	1	410	1	1,345	3	2,640	6	0.5
대학	-	-	-	-	-	-	700	1	700	1	0.1
합 계	83,148	107	120,581	158	143,811	181	133,815	150	481,355	596	-

4. 시사점

AI 반도체 산업을 정책 및 예산 측면에서 분석한 결과, 전반적으로 기술 부분의 정책이 기업 경쟁력 및 생태계 부문의 정책으로 전환되고 있음을 알 수 있다. 이는 기술-사업-성장이라는 혁신 생애주기에서 기술에서 사업 구간으로 넘어가는 과정에 대응하여 전반적인 정부 정책의 성격이 연구개발정책에서 사업화 정책으로 전환되는 것으로 볼 수 있다. 단, 이 과정에서 사업화 정책 대응이 아직은 부족한 것으로 보인다. 문제는 사업화 정책을 어느 부처에서 주도하는가에 대한 문제이다. AI 반도체 산업이 막대한 재원이 필요한 만큼 정부정책과 예산의 집중이 필요하다. 기술 부문의 연구개발정책과 인프라지원 정책의 경우 과학기술정보통신부와 산업통상자원부로 나뉘어 정책이 형성 및 집행되고 있다. 특히, 기업의 기술개발과 육성에 대한 정책의 경우, 부처별로 과학기술정보통신부의 “인공지능 반도체 혁신기업 집중육성 (R&D)”, 산업통상자원부의 “글로벌스타트업30기술개발지원”, 중소벤처기업부 “초격차스타트업육성사업 (DIPS)”와 사업들이 존재하며, 부처별 고유한 사업 목적이 존재하지만 사업의 지원 대상은 일부 겹칠 수 있는 단점이 있다. 이러한 문제를 극복하기 위해서는 각 부처의 특성에 맞는 정책조정이 필요하다. 사업 간 유사·중복성은 분야를 막론하고 제기되는 이슈로 볼 수 있다. 하지만, AI반도체는 타 분야 대비 지원 목적·내용의 차별성이 존재하는 편으로 판단할 수 있다. 다만, 향후 R&D 지원의 효과를 보다 극대화하기 위하여, AI연계산업 및 데이터 센터 등 특정 영역에서 활용되는 AI반도체 산업을 고려하여, 주관부처에서는 최종 산출물 및 적용 도메인, 경제·사회적 파급효과 등을 종합적으로 고려하여 사업간 차별성을 가지고 사업을 진행할 필요가 있다. 더불어, 유사사업 간 ‘프로그램 사업화’를 추진하여 중장기적인 지원 체계 고도화 방안 마련이 요구된다.

인력분야의 경우 장기적 차원에서의 적시적 인재 공급을 고려할 필요가 있다. 전략 차원에서는 2020년에 이미 전략차원에서 고급인재 인력양성 대한 고려가 있었으나, AI 반도체 대학원이 개설과 인재양성 사업을 통한 인재 배출까지의 소요 시간을 고려하면 인력난을 당장 해결하기는 힘들다. 해외로의 인력유출¹³³⁾을 해결할 수 있는 장

133) 조선비즈(25.01.02.), “[단독] 韓, AI 인재 유출국 됐다… 日은 순유입국 유지”, <https://biz.chosun.com/it-science/ict/2025/01/02/3WZDNB5JCNDNKHUDHC6L6P4W2I/> (검색일 25.01.18.)

기적인 전문 인력 매칭 정책의 확충이 필요하다. 더불어, 2000년대 중반 이후부터 반도체 R&D에 대한 지원이 줄어들어 대학교육의 한계가 발생했다는 지적¹³⁴⁾이 있는데, 향후 현재 진행되는 정책들의 귀추를 주목할 필요가 있다.

투자 부문의 경우, 반도체 생태계 펀드와 같은 특화 펀드가 만들어지고 규모가 커지는 것은 매우 고무적이나, 이를 운용할 수 있는 전문성 있는 민간 운용사의 선정이 주요한 쟁점이 된다. 또한, AI 반도체 이니셔티브에 따르면 과학기술정보통신부의 출자를 통해 'AI 특화 펀드'의 '모펀드' 신설을 중점 추진과제로 발표했는데, 이미 2023년에 조성된 반도체 생태계 펀드가 존재하는 만큼, 정책 조정을 통해 연구개발과 사업화 정책을 단계적으로 지원하는 방향을 모색할 필요가 있다. AI 반도체 산업 생태계 특성을 고려할 때, 정부 투자는 민간투자와의 연계를 높이기 위한 마중물로서 하이리스크 첨단기술 부문의 혁신적 기업의 초기 투자에 집중하는 것이 효율성과 형평성을 동시에 잡는 방안으로 보인다. 단, 리스크가 높더라도 기술력 측면, 시장에서의 수요에 대응하는 혁신적인 Business Model의 유무 등 기업의 선별 등 구조화된 평가의 체계를 통한 투자가 강화되어야 한다.

마지막으로 세부적 지원 분야를 기준으로 분석한 결과, 반도체 분야 정부R&D 투자 방향은 대체로 실효성 있는 사화경제적 성과 창출을 위해 산업계 주도의 AI반도체 제품화를 중심으로 지원이 이루어지고 있다.¹³⁵⁾ 22년도 국가연구개발사업 조사분석보고서」에서 집계한 산업계 참여 비중이 중소기업(19.2%), 중견기업(6.5%), 대기업(2.1%)인 점과 비교하였을 때, AI반도체를 포함한 반도체 분야 전반에 있어, 산업계의 활성화된 참여 및 정부의 집중 수준을 명확히 확인할 수 있었다. 또한 정부 연구개발 사업의 기초·응용·개발 연구 비중(28.0%, 25.2%, 46.8%)을 고려하였을 때, 반도체 및 AI반도체 분야의 기초연구 비중이 10%인 점을 감안, 정부 투자의 방향성을 확인할 수 있었다.

그간 국내 취약한 시스템반도체 및 AI반도체 분야의 경쟁력 강화를 위한 집중 지원의 취지는 바람직하나, 실효성 있는 단기, 중·장기 경제·사회적 성과 창출을 위한 사업간 연계 및 활용 방안 마련은 필요할 것으로 판단된다. 특히, 기업지원 과제는 중소·

134) 조선일보(24.10.17.), ““반도체 인력 키워야할 대학 교육 시스템 붕괴… 삼성도 흔들릴 수밖에””, <https://www.chosun.com/economy/industry-company/2024/10/17/TRRPVGT3FBCJP4HGGXMUUGDMQ/> (검색일 24.12.28.)

중견 팹리스의 단기 상용화가 주요 사업 목적이었던 만큼, 종료 시점(‘24)을 고려하여 성과 추적에 따른 효과성 검토 등 근거를 바탕으로 후속 사업의 기획·추진이 필요하다. 한편, 인적·물적 구성에 따른 대학의 참여 비중이 제한적인 점을 고려, 원천·공통 기술의 산업 활용 연계를 위한 방안 및 추진체계 마련이 필요하다. 또한, 최근 정부에서 발표한 「AI-반도체 이니셔티브」와 같이 두 기술 간 연계성 마련을 위한 전 부처·전 사업을 아우를 수 있는 총괄적인 거버넌스의 확립과 로드맵을 고려한 사업(과제) 추진이 필요하다.

| 제3장 | AI 반도체 생태계 마인드 맵 분석

제1절 개관 및 마인드맵 프레임워크

생태계의 문제점 분석을 위하여 학계, 협회, 정부 사업 관계자, 기업 관계자의 인터뷰를 진행하였고, 이와 문헌 연구를 더불어 국내 AI반도체 생태계 현황을 분석하는 작업을 수행하였다. AI반도체 기업의 생존과 성장에 대한 요인의 분석은 AI 반도체 생태계가 가지는 고유한 특성과 이로인한 문제점의 파악이 최우선적 과제이며, 각 문제점 역시 다른 문제점과 상호 연결되어 톱니바퀴와같이 인과적으로 작용하는 경우가 있으므로, 이에 대한 문제점을 구체적으로 인식하기 위하여 문제점을 구체화하고 분석 영역별로 확장할 수 있는 마인드맵 분석을 활용하고자한다.

김정호 (2023)는 지원 및 역량, 제품 및 사업 특성, 시장 환경 및 산업 생태계, 정책 및 제도의 프레임워크를 통해 스케일업의 주요 요소를 파악하였다. 하지만, 본 연구는 시스템 반도체 중에서도 AI 반도체 기업 (팹리스 스타트업 및 초기 신생기업 중 가속기 설계 및 제작, 디자인 하우스, IP 제공 업체를 대상으로 진행)의 성장과 생존에 주요한 영향을 미치는 요인을 파악하는 연구로 앞선 스케일업의 대상이 되는 제품의 매출에 대한 부분이 아직 성숙하지 않았고, 규제나 제도적 프레임워크를 판단할 수 있는 사례가 발견되지 않았다. 주로, 육성과 지원에 대한 부처별 분산된 정책이 주를 이루고 있으며, 국내 시장을 선도하는 대표적인 기업도 아직은 단정 지을 수 없는 것이 현실이다.

AI반도체 생태계가 당면한 문제점을 분석하기 위하여, Tech 스타트업의 성장에 미치는 주요한 요인들을 살펴볼 필요가 있다. 우선, 지속적인 R&D투자는 기술 스타트업의 사회적 경제적 영향을 극대화 시키는 동시에 성장과 시장의 경쟁력 강화에 도움을 줄 수 있다(Triono et al, 2022). Eliakis(2020)의 지적과 같이 지속적인 기술 혁신과 기술 혁신을 위한 인재 확보의 측면은 주요한 기술 스타트업의 성장 요인으로 볼 수 있다. 더불어, 초기 고객 및 수요처 확보 등의 기업의 마케팅 및 사업 전략과 운영 자금조달 부문도 성장의 주요한 요인으로 작용한다 (Singh & Subrahmanyam,

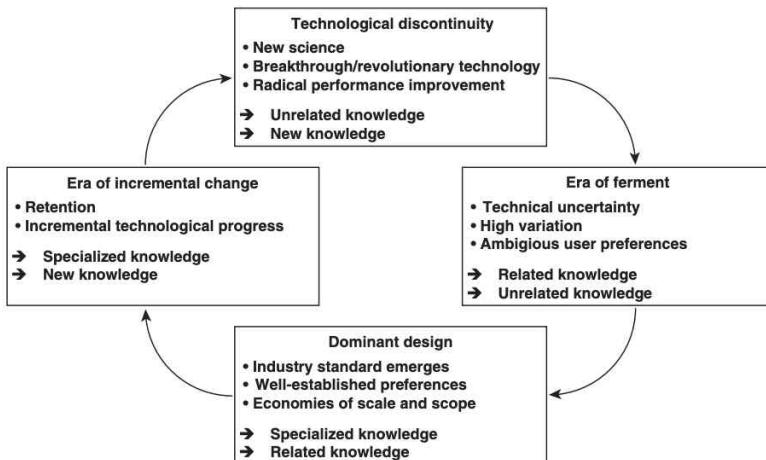
2020; Paik & Woo, 2013; Qoriawan & Apriliyanti, 2022). 기술 스타트업의 성장에 미치는 영향은 기업의 사업전략, 인재, 자금(투자 유치), 기술로 구분하여 볼 수 있다. 위의 선행 연구를 기반으로, AI 반도체 생태계 분석을 위한 마인드맵의 주요 요소 도출을 위하여 기술 창업 및 동 생태계의 발전과 생존의 주요 요인에 대한 문헌 연구를 실시하였다.

1. 기술 혁신/창업 생태계의 성장단계

기술 산업 생태계의 발전은 앞서 논의한 일반적 성장 단계를 따른다고 할 수 있다. 즉, 큰 틀에서 도입기 → 성장기 → 성숙기를 거쳐 쇠퇴기를 겪게 된다. 다만, 생태계의 일반적 수명주기를 따르는 기술수명주기와 기술순환모형 기술의 lifecycle과 기술 순환모형 (Tushman and Rosenkopf, 1992)를 결합하여 보면, 조금 더 기술 중심의 생태계 성장모형을 만들어 볼 수 있다. 앞선 모형을 발전시킨, Kalthaus (2020)에 따르면, 기술은 순환적 특성을 띠는데, 이는 ‘완전히 새로운 발명(Technological discontinuity)’이 등장을 하고 난 후, 새로운 기술과 관련된 다양한 제품 및 기술이 개발되는 ‘발효의 시대 (Era of ferment)’가 등장을 하게 된다. 발효의 시대는 Chasm¹³⁵⁾에 빠지지 않는 이상 De facto standard(사실상의 표준)이라고 불리는 ‘Dominant design(지배적 디자인)’이 시장에 소개될 때까지 지속되고, 지배적 디자인 이후는 ‘점진적 발전의 시대 (Era of incremental change)’가 등장하게 되고, 사실상의 표준인 지배적 디자인을 개선하는 기술 및 제품이 지속되게 되고 이는 다음의 완전히 새로운 기술이 발명될 때까지 지속된다.

135) Chasm은 Geoffrey Moore의 책 “Crossing the Chasm”에서 제시한 개념으로 기술확산에 있어서 초기 수용자 (Early Adopters)와 초기 다수(Early Majority) 수용자 간에 존재하는 간격을 의미한다. 이 Chasm(틈)이 중요한 이유는 이 두 수용자 간극이 기술 확산에 있어서 가장 중요하며, 기술확산 및 산업의 성장에 있어서 초기 다수 (Early Majority) 수용자로 넘어가는 것은 아주 중요하기 때문이다.

[그림 3-1] 기술순환 모형



자료: Kalthaus (2020)

위의 기술순환모형과 과거의 생태계 성장이론을 고려하면, 기술산업 분야의 성장 단계는 다음과 같은 단계로 정리하여 볼 수 있다.

1. 기술 아이디어 및 연구개발 단계 (Technology Idea and Research & Development Stage): 이 단계에서는 초기 기술개발을 위한 아이디어 및 연구개발을 시작한다.
2. 시드 및 초기 단계 (Seed and Early Stage): 이 단계에서는 개발된 기술의 시장 성(Feasibility)을 검증하고, 초기 prototype 제품을 만들어 초기 시장 기술 수용 모델 상의 혁신자(Innovators) 수용그룹에 진출하며, 다양한 기술제품을 시장에 소개하는 ‘발효의 시대(Era of ferment)’의 특성을 보일 가능성이 높다.
3. 시장 진입 및 성장 단계 (Market entry and Growth Stage): 이 단계에서는 Chasm을 지나서 초기 다수(Early Majority) 수용자 그룹까지 기술/혁신이 확산되어 시장이 본격적인 성장단계를 맞이하게 된다. 이 단계에서는 ‘사실상의 표준(De facto standard) 혹은 지배적 디자인(Dominant design)’이 등장할 가능성이 매우 높다.
4. 확장 단계(Expansion Stage): 이 단계에서는 사업을 대규모로 확장하며, 시장에서는 조정기(Shakeout)를 맞이할 확률이 높다. 각 기업의 M&A가 활발히 일

어나게 되고, 시장의 전체적 규모는 커지나 실제 활동하는 기업의 총 수는 줄어들게 된다.

5. 성숙 및 안정화 단계 (Mature and Stabilization Stage): 이 단계에서는 시장에서의 입지를 확고히 하고, 지속 가능한 성장을 추구하게 되는데, 이 단계에서는 ‘점진적 변화의 시대(Era of incremental change)’의 특성을 보이게 되며, 시장 크기 자체의 변화는 둔화되고 몇몇 기업이 전체 시장을 과점하는 특성을 보이게 된다. 이 시기에 새로운 스타트업들이 완전히 새로운 대체할 만한 기술을 소개하는 활동이 시작되게 된다.

각 단계별 주요 활동/요인을 정리해보면 다음과 같다.

<표 3-1> 생태계 성장 단계별 주요 특징

성장 단계	단계별 특징
기술 아이디어 및 연구개발 단계 ¹³⁶⁾	<p>기술아이디어 발굴 및 평가: 완전히 새로운 창의적인 아이디어를 발굴하고, 이에 대한 초기 시장 가능성과 기술적 타당성 평가 연구개발: 개인 혹은 조직적으로 핵심 기술을 개발하고, 프로토타입 개발 지식 재산권 확보: 특히 및 기타 지식 재산권을 확보 기존의 기술과는 그 접근법이 완전히 다른(Technology discontinuity) 기술을 목표로 함 정부의 연구개발 및 중소기업 지원 제도의 역할이 큼</p>
시드 및 초기 단계	<p>초기 자금 조달: 엔젤 투자자, 시드 펀딩, 정부 자금 등을 통해 초기 자금을 확보 프로토타입 개발 및 검증: 시제품을 개발하고, 시장에서 기술 검증(Proof of concept, PoC) 수행 초기 고객 확보: 기존 네트워크 혹은 정부 사업 등을 통해 초기 고객 확보를 위해 노력하며, 시장 반응을 분석 인큐베이터 및 액셀러레이터 프로그램 참여: 개발된 기술에 대한 비즈니스모델 개발(Business development, BD)을 위해 전문 프로그램 혹은 전문가와의 연계를 추진</p>
시장 진입 및 성장 단계	<p>제품 출시: 제품 또는 서비스를 공식적으로 시장에 출시 마케팅 및 세일즈 전략: 효과적인 마케팅 및 세일즈 전략을 수립하고 실행 벤처 캐피탈 유치: 추가 자금을 확보하여 제품 양산 능력을 포함한 제품의 본격적 시장 진출 도모 고객 기반 확장: 더 많은 고객을 확보하기 위해 제품의 다각화 등을 검토하며, 시장 점유율을 늘려감 기술 및 제품 개선: 시장의 초기 피드백을 반영하여 기술과 제품을 지속적으로 개선하여 ‘사실상의 표준 혹은 지배적 디자인’ 제품 양산을 위해 노력</p>

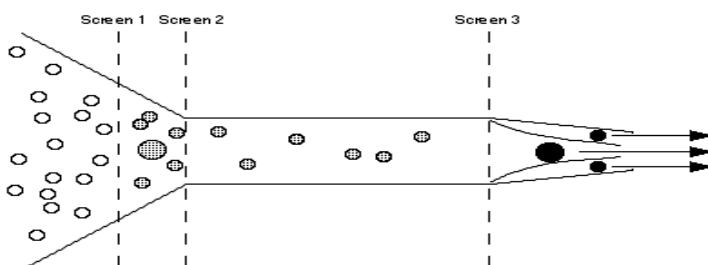
성장 단계	단계별 특징
확장 단계	국내 및 국제 시장 확장 : 국내 시장은 물론 해외 시장 진출을 도모함 파트너십 및 협력: 타 기업과의 전략적 협력을 추구하여 자원배분 및 활용을 극대화 함 조직 확장: 인력 및 조직 구조를 확장하게 됨 지속적인 혁신: 지배적 디자인을 중심으로 개선된 제품을 지속적으로 개발
성수 및 안정화 단계	지속 가능한 비즈니스 모델: 장기적인 성장을 위한 지속 가능한 비즈니스 모델을 구축 기업공개(IPO): 필요시 기업공개를 통해 대규모 자금 유치하고 기업의 가치를 극대화 재투자 및 인수합병: 본격적 시장 및 사업 다각화 추진 글로벌 시장에서의 위치 확보: 글로벌 시장에서 브랜드 가치를 높이기 위해 노력

자료: 연구진 작성

2. 기술 스타트업의 생존 및 성장

기술 스타트업의 성장에서는 당연하게도 기술의 역할이 중요하다. 기술의 발명(Invention)은 상대적으로 쉬운 편이지만, 시장으로의 성공적인 상업화까지를 포함하는 기술의 혁신(Innovation)은 쉽지 않다. 실제로 매일 아주 많은 특허가 등록되고 소개가 되지만, 이들이 시장에서 성공적으로 받아들여지는 상업화와는 거리가 있다. 이를 단적으로 표현하는 것이 혁신의 깔때기(Innovation funnel)로 표현할 수 있는데 이는 발명과 성공적인 상용화는 전혀 다른 문제임을 알려 준다 (Wheelwright & Clark, 1992).

[그림 3-2] 혁신 깔때기



자료: Wheelwright & Clark, (1992)

136) 기술의 난이도와 불확실성이 높을수록 정부지원 사업의 역할이 중요해 진다. 기술 아이디어 및 연구개발, 그리고 초기 시장 진출을 위해서 각 국에서는 다양한 지원 프로그램을 운영하고 있다. 미국에서는 SBIR (Small Business Innovation Research) 프로그램을 도입하여 운영 중에 있다.

구체적으로 초기 기술개발에 있어서 많은 경우 스타트업이 기술개발에 초점을 두고 시장을 도외시 하는 경우가 발생한다. 이는 위에서 설명한 혁신의 깔때기에서 왜 많은 기업들이 걸리는지를 잘 설명하는 이유 중에 하나이다. 존재하지 않는 문제를 해결해 려고 하거나, 너무 기술 중심의 접근법은 시장과의 괴리를 가져오게 만든다. 이는 많은 스타트업이 공통적으로 가지는 문제이긴 하지만, 기술 스타트업의 경우 그 가능성 이 높다고 할 수 있다. 이러한 문제를 Eric Reis는 그의 책 “The Lean Startup”에서 시장을 알아야 한다고 지적하고 있다.

기술과 시장과의 관계는 기술수용모형(Technology Acceptance Model)으로 설명해볼 수 있는데, 기술수용모형에서는 사용자(시장)가 새로운 기술을 받아들이기 위해서는 크게 두 가지 요인을 고려해야한다고 주장한다. 첫 번째는 PU(Perceived usefulness)이고, 두 번째는 PEOU(Perceived ease-of-use)이다. 즉, 새로운 기술은 소비자에게 유용성을 가져다 줄 수 있어야 하며, 또한 쉽게 사용할 수 있어야 한다는 것이다 (Davis, 1989).

성공적으로 초기 기술개발에 성장을 하더라도, Geoffrey Moore가 주장한 The Chasm을 넘어야 본격적인 성장에 접할 수 있고 (Moore, 1991), 이를 위해서는 스타트업이 가진 기술뿐만이 아니라 보완이 되는 기술(Complementary technology)과 그 기술이 사용이 될 수 있는 관련 산업이 함께 성장하여야 하며, 그 산업이 신생 산업 일수록 그 중요성은 커진다 (Kapoor and Furr, 2014).

성공적으로 기술개발에 성공하여 상용화를 진행한다고 하더라도 여전히 문제는 발생한다. 기술 스타트업이 성장하면서 꾸준히 기술혁신을 진행하여야 한다. 그렇지 않으면, 앞에서 살펴본 S자 곡선의 발전에 있어서 성장기 이후 산업은 조정기 (Shakeout)에서 퇴출될 가능성이 커지기 때문이다. Klepper and Simons (2005)에 따르면 조정기에 기술혁신이 부족한 기업이 퇴출된다고 설명한다.

3. 초기 기술 스타트업의 생존 및 성장 요인

기술 스타트업의 생존 및 성공적인 성장을 위해서는 시작 시점에서의 스타트업 특성을 파악할 필요가 있다. 이를 크게 두 가지 측면으로 나누어 볼 수 있는데 첫 번째는

스타트업 기업의 특징과 두 번째는 창업자의 특징으로 나누어 볼 수 있다 (Kalyanasundaram et al., 2021).

기술 스타트업의 특징은 여러 가지가 있을 수 있지만, 창업의 형태를 기준으로 de alio와 de novo 두 가지로 구분할 수 있다. de alio는 라틴어로 “다른 곳에서” 또는 “다른 출처로부터”를 의미하며, 즉 de alio 기업은 기존 기업이나 조직에서 분사되거나 파생된 기업으로 볼 수 있다. 반면에, de novo는 라틴어로 “처음부터” 또는 “새로”를 의미하는데 de novo 기업은 완전히 새로운 창업을 의미한다. 즉, de alio의 경우 기존 기업에서 Corporate venturing 등의 방법으로 spinoff 혹은 분사되는 것을 의미하며, de novo의 경우는 우리가 이해하는 과거 경험이 없는 startup이라고 이해하면 된다. 이 둘은 기본적으로 다음과 같은 차이점이 존재한다.

	de alio 기업	de novo 기업
특징	기존 자원의 활용 위험 감소 빠른 시장진입 속도 혁신의 한계 (기존 기업에 얹매임)	혁신의 자유도가 높음 높은 위험 유연성 자원의 부족

특히, 기술 산업분야의 경우 이 차이점이 두드러지게 된다. 여러 연구에서 이를 뒷받침하고 있는데 예를 들어 Gang (2018)에 따르면, 새로운 하부시장이 나타났을 때 기존 하부시장에 오래 있을수록 새로운 시장의 진입시점이 늦지만, 그 생존확률은 훨씬 더 큰 것으로 나타났다. 유사하게 Klepper & Simons (2000)의 연구에 따르면 라디오를 출시했던 기업들이 TV 시장에서 퇴출될 확률이 낮은 것으로 나타났다. 또한 Benner and Tripsas (2012)의 연구에서도 Digital camera 산업에 진출한 기업의 특성을 살펴보았다. Khessina & Carroll (2008)은 de novo 와 de alio 회사를 직접적으로 비교했는데, de novo 회사일수록 보다 혁신적인 제품을 생산해 내지만 생존확률은 낮음을 발견하였다. 즉, 이러한 연구의 공통점은 기술 스타트업 기업의 특징이 그만큼 중요하다는 의미이다.

창업자의 특성도 스타트업의 특성 만큼 중요한데, Hsu(2007)의 연구에서는 이를 잘 보여주고 있다. 이 연구에서는 창업자의 특성 - 학위 수준, 사회적자본(네트워크),

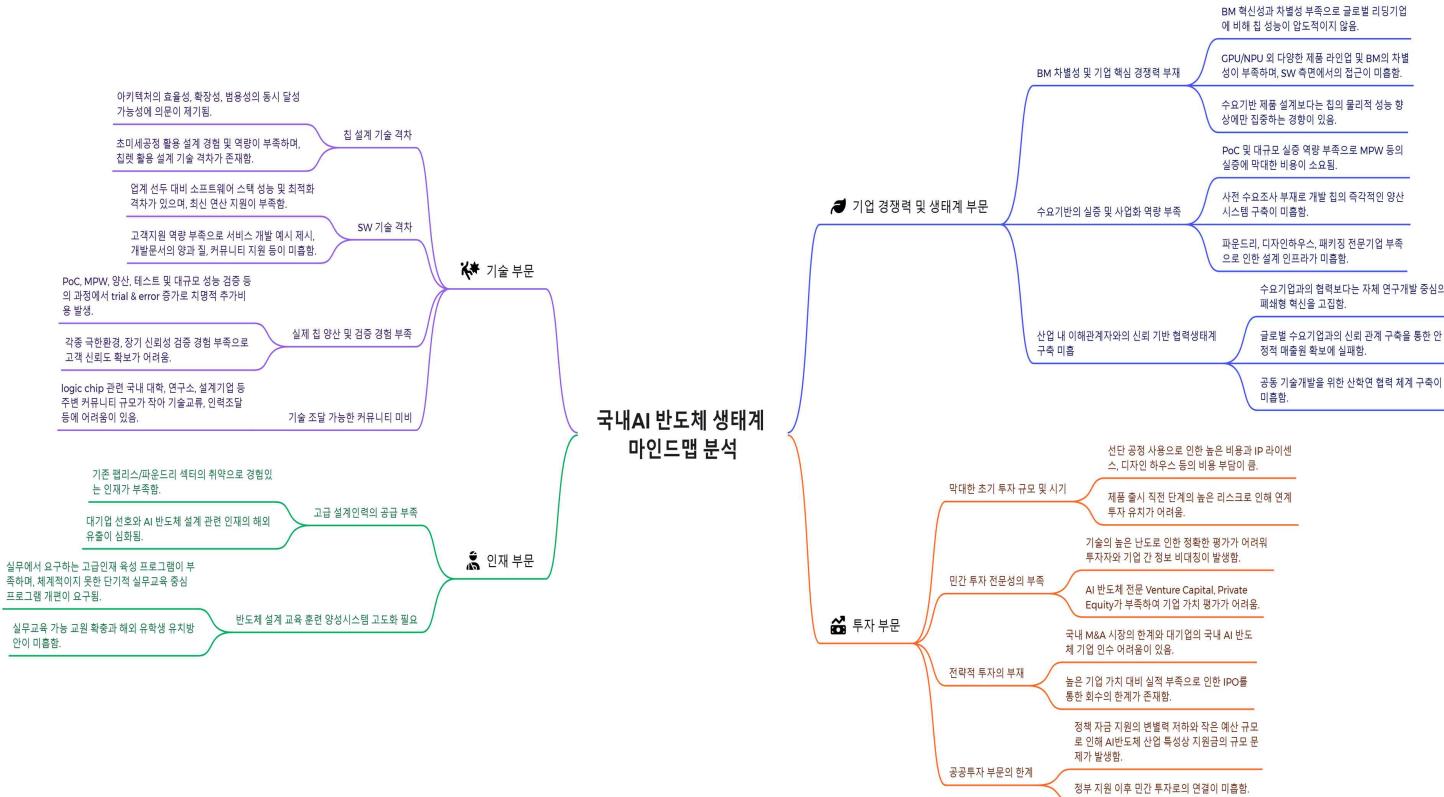
과거 창업경험이 기술 스타트업의 VC 투자 및 기업가치 평가가 높은 것으로 나타났다. 다만 학위가 높을수록 기술에 중심을 둘 가능성이 높으므로, 과거 창업경험은 물론 다양한 정부 프로그램(예, I-Corps 프로그램)으로 기술과 시장의 갭을 줄이는 프로그램이 필요하다. 스타트업의 특성, 창업자의 특성 외에도 생태계 자체의 특성(예를 들어, 대학 및 연구센터, 그 도시의 교육 수준, 인큐베이터의 수)도 중요한 역할을 한다(Bosco, et al., 2019).

더불어, AI 산업에 활용되는 반도체로서 학습 부문 및 연계 시장의 90% 이상의 시장 점유율을 가지는 NVIDIA의 성장 경로를 분석하는 연구들(Wang, 2024; Wang, 2024; Yin, 2023; Li, 2024)은 NVIDIA의 성장이 지속적인 기술의 혁신, 급변하는 시장에 대한 대응 전략, 재무적 성과를 활용한 투자, 지속적 혁신 추구를 위한 경영 전략(M&A, HPC Interconnect 업체 Mellanox Technologies의 인수 사례)에 기인한다고 설명하고 있다.

4. 마인드맵 생태계 분석 프레임의 도출

AI반도체 산업 생태계의 특징을 살펴보면, 기존의 산업에서 특화된 영역으로 발전되는 구조(시스템반도체 산업 → AI반도체 산업), 기존의 반도체 산업보다 기반 인프라 산업에 대한 의존도가 높은 특성(5nm 이하의 선단공정 사용 및 HBM의 탑재, IP, EDA 사용, 첨단 패키징), 국내의 경우 아직 양산된 제품을 통해 시장에서의 매출을 확인하기 어려우므로, 기술과 발전 가능성에 대한 가치를 산정하여 투자가 진행되는 특성을 가진 산업으로 볼 수 있다. 동 업계 주요 기업 창업자들의 경우, 반도체 산업의 글로벌 기업 경력, 다수의 박사 학위자(고학력), 교수 창업 등 다양한 형태로 구성되어 있다. 특히 반도체 산업 경력(칩설계, 소재 등 경험)이 창업에 미치는 영향이 큰 것을 알 수 있었다. 기업의 성장과 생존에 필요한 요소를 고려할 때, 앞선 선행연구에서 주요하게 논의된 산업에 대한 투자, 창업자의 사업 모델과 시장 전략, 동 산업 및 첨단 기술 분야 스타트업이 마주하는 문제점, 보완 기술 및 전반적 기술 수준에 대한 검토, 동 산업의 혁신적 인재 등 요소가 AI 반도체 기업의 생존과 성장을 위한 동 산업 생태계의 분석에 활용되었다.

생태계가 직면하고 있는 문제점을 분석하기 위하여 다음의 프레임워크를 구성하였다. 우선 기업의 사업 모델 및 전략을 고려하여 ‘1. 기업의 경쟁력과 생태계’, 민간, 공공 부문의 투자와 연계된 투자 부문에서의 문제점 파악을 위한 ‘2. 투자’, 동 생태계에 공급되는 인재와 인력 문제를 고려하기 위한 ‘3. 인재’, 다변하고 급변하는 기술 부문과 이에 대한 경쟁력과 구성요소를 파악하기 위한 ‘4. 기술’로 나누어 AI 반도체 생태계의 문제점을 파악하였다.



제2절 프레임워크 요소별 분석

1. 기업 경쟁력 및 생태계 부문

가. BM(Business Model) 차별성 및 기업 핵심 경쟁력 부재

1) BM 혁신성 및 차별성 부족

국내 AI반도체 선도 펫리스 기업들은 각자 뛰어난 성능을 강조하고 있다. 특히 NVIDIA의 GPU 대비 가성비, 전성비(전력 대비 성능)이 우수한 것으로 여겨진다. 예를 들어 퓨리오사AI의 2세대 모델 레니게이드의 경우 자사의 1세대 모델 워보이에 비해 메모리 성능의 경우 4세대 고대역폭 메모리(HBM3)가 탑재돼 25배 이상 좋아졌고, 정수 계산 능력은 8배 좋아져서, 종합적인 성능은 최대 20배까지 좋아졌다. 하지만 이러한 물리적인 성능은 자체 기술개발 결과로서 외부의 객관적 시각에서 완벽히 입증된 성능으로 보기엔 한계가 있다. 더불어 실제 해당 칩이 활용될 다양한 수요처의 환경 조건에 따라 성능에는 큰 편차가 있기 마련이다. 즉, 시장에서 검증된 성능으로서 아직 국내 AI반도체 기업들의 제품이 경쟁우위를 확보하였다고 보기는 어렵다.

[그림 3-3] 퓨리오사AI의 2세대 레니게이드 성능 비교(vs NVIDIA)



주: Llama 2 7B로 성능 비교한 결과 수치

자료: 퓨리오사AI 홈페이지(<https://furiosa.ai/rngd>); 이코노미조선(2024.3.25.) 「“차세대 반도체, 3분기 공개…잠재력 증명 후 IPO 도전”」, <https://shorturl.at/sIY1T>(검색일: 2024.10.25.) 재인용

한편 국내 AI반도체 생태계에서 각광을 받고 있는 선도기업(예: 퓨리오사AI, 리밸리온, 사피온)들은 주로 칩 설계에 초점을 맞춘 팝리스 기업들이다. 하지만 해당 시장은 이미 NVIDIA가 80~90%의 압도적인 점유율을 기록하고 있어 해당 시장에서의 경쟁력을 확보하기엔 한계가 있다. 따라서 제조업 경쟁력이 뛰어난 국내 시장에서는 학습이 끝난 AI모델을 실제 제조 현장에 적용하기 위한 장비(로봇, CCTV, 드론 등 각종 단말기)에 탑재되어 실행되는 ‘엣지 디바이스(Edge-Device)’ 형 반도체의 수요도 향후 폭발적으로 증가할 것으로 예상되는 바, 보다 명확한 목표 시장과 차별화된 제품 전략이 필요하다([그림 3-4] 참조).

[그림 3-4] AI반도체 구분에 따른 제품전략의 다양성



자료: 이동수(2022), 「지능형 반도체의 새로운 패러다임」, 2023 ICT 산업전망컨퍼런스, 네이버; 윤종혁(2024.4.22.), 「엣지 디바이스에서의 AI 이용에 관한 산업분석」,『ICT Standard Weekly』, 제1183호, 한국정보통신기술협회 재인용

또한 AI반도체는 BM 설계에 있어서 단순히 하드웨어(H/W)만이 아닌 소프트웨어(S/W)적인 특성도 동시에 고려해야 한다. 반도체라는 제품을 생산하기 위해서는 반도체를 구성하는 회로 설계가 매우 중요한데 이러한 설계 과정에서 고성능의 S/W 특성을 가진다. 같은 AI반도체라도 기업들마다 세부 특화 및 우위 분야가 다르고, 세부 논리구조도 상이하며, 설계에 필요한 인력 수준에 따라 제품 구현 및 성능에 있어서의

격차가 크게 벌어질 수 있다. 이렇게 H/W와 S/W의 양면적 특성 때문에 칩 설계 및 제품 실증에 많은 비용이 소요되어 초기 단계부터 대형 자본 투자가 필요하다.

데이터센터용 AI반도체 시장에서의 NVIDIA의 독보적인 입지 이러한 S/W적인 특성을 잘 살린 ‘쿠다(CUDA: Compute Unified Device Architecture)’라는 플랫폼 생태계를 구축한 데에 있다. 2006년 출시된 쿠다는 GPU로 수행할 수 있는 병렬처리 알고리즘을 기준 산업 표준 프로그래밍 언어를 사용해 작성할 수 있도록 돋는 소프트웨어 도구(API)로서, 개발자는 쿠다를 이용해 GPU의 병렬처리 성능을 쉽고 간편하게 활용할 수 있으며, 대규모 데이터처리, 이미지 처리, AI 개발 및 학습에 있어서의 성능과 효율을 크게 향상할 수 있다.¹³⁷⁾ 쿠다 플랫폼의 가장 중요한 특성은 NVIDIA의 GPU에서만 구동되는 폐쇄형 시스템(Closed System)이라는 것이다. 2023년 기준 450만 명에 달하는 쿠다 플랫폼 내 개발자들은 다양한 개발 코드를 만들어냈고, 이러한 코드들은 후임 개발자들이 학습할 거대 이론처럼 여겨져 ‘락인(Rock-in) 효과’를 발생 시킨다.¹³⁸⁾ 하지만 국내 AI반도체 기업들은 고성능의 칩 설계·제조에만 초점을 맞추고 있어, 효과적인 개발자 커뮤니티 및 생태계 구축에 대한 비전을 제시하지 못하고 있다.

2) 기업 핵심역량(Core Competence) 및 시장전략의 부재

국내 AI반도체 기업들의 칩 설계역량이나 BM의 차별성보다 더 큰 문제는 핵심 경쟁 우위 및 시장전략의 부재에 있다. 현장의 많은 기술 전문가들은 국내 AI반도체 스타트업들의 기술역량이나 인적수준은 상대적으로 높게 평가하고 있으나, 결국 NVIDIA와 같은 글로벌 선도기업과의 경쟁을 위해서는 수요기업이 원하는 제품을 기획하는 역량이 무엇보다 중요하다고 강조한다. 하지만 국내 AI반도체 기업들은 수요기반 제품 설계보다는 칩의 물리적 성능 향상에만 치중하는 경향이 있다. AI반도체 관련 국제 컨퍼런스 등에서도 많은 외국 기업들이 자사 제품에 대한 서비스 수요에 초점을 맞추어 홍보하는 반면, 국내 기업들은 제품의 기술적 성능을 보여주는데 집중하는 것이다. 따라서 AI반도체

137) 글로벌이코노믹(2023.12.20.), 「“추론 기술 시대 온다” 인텔 CEO, NVIDIA의 CUDA 종말 주장」,
<https://bulky.kr/ESxCDvx>(검색일: 2024.10.24.)

138) 아시아경제(2024.5.23.), 「[Why&Next]‘천비디아’ 독주에…빅테크·反NVIDIA 연대까지 추격」,
<https://view.asiae.co.kr/article/2024052307265909462>(검색일: 2024.10.24.)

기업들의 경쟁우위 확보를 위해서는 탁월한 기술력보다는 세밀하고 차별화된 시장 차원의 경쟁전략 수립에 초점을 맞추어야 한다.¹³⁹⁾

영국의 글로벌 선도 AI반도체 스타트업인 그래프코어(Graphcore)는 이러한 경쟁 전략에 있어서 대표적인 벤치마킹 사례이다. 그래프코어는 초기 단계부터 Dell, Bosch, 삼성전자, 이후 Microsoft와 BMW에 이르기까지 글로벌 수요기업으로부터의 전략적 대형투자 유치를 이끌어내면서 자금조달에 있어 독보적인 경쟁력을 확보하였다. 더불어 초기부터 글로벌 수요시장을 타겟으로 설정하여, 미국, 유럽 내 연구기관이나 대학과는 공동 R&D, TSMC와는 양산 협력, 앞서 언급한 수요기업들과는 실증 및 생산을 위한 협력활동을 수행하면서 가치사슬 내 다양한 이해관계자들과의 협력·제휴 관계를 강화하였다.¹⁴⁰⁾

<표 3-2> 국내외 AI반도체 대표기업의 스케일업 요인 비교

요인	해외 사례(Graphcore)	국내 사례(퓨리오사AI)	시사점
기술역량	세계 최고수준 (자체 핵심설계가 가능)	세계 상위권 수준 (자체 핵심설계가 가능)	해외 사례기업에 비해 낮지만 글로벌 수준임
BM/매출구조	제품 설계(파운드리 기업에서 생산), B2B 매출	제품 설계(파운드리 기업에서 생산), B2B 매출	큰 차이 없음
사업화 소요기간	약 3년 후부터 본격적 생산 (매출 급속히 증가)	약 4년부터 시제품 검증 (실증) 중	본격 사업화(대규모 매출 발생)까지 오래 걸림
해외진출	글로벌 수요에 대응 (글로벌 기업들에 공급 중, 해외기관과 실증 협력)	현재 주로 국내 수요에 대응 (장기적으로 해외진출 목표)	국내 사례기업의 해외진출 규모가 낮은 편
투자유치	초기단계(Seed, Series A)부터 대형투자 유치 (대기업 또는 CVC, 대형 VC, 전문·특화VC 참여)	초기단계에서 대형투자 유치 어려움 (Series B 이후부터 국내 대형 VC가 본격 투자)	대형투자 미흡 (초기단계에서 수요기업과 대형 VC의 투자부족)
수요-공급기업 간 협력	초기단계부터 글로벌 수요기업, 파운드리(생산) 기업과 적극 협력	초기단계에서 글로벌 수요 기업, 국내 파운드리(생산) 기업과 협력 어려움	국내수요 활성화 부족, 수요기업과 협력 부족

자료: 김정호(2023), 「하드웨어 유니콘의 스케일업 요인에 관한 연구: 해외 유니콘과 국내 예비유니콘의 사례 비교」, 『한국창업학회지』, 18(3), p. 340 표를 연구진이 요약·발췌.

139) 전문가 인터뷰 내용 중 발췌

140) 김정호(2023), 「하드웨어 유니콘의 스케일업 요인에 관한 연구: 해외 유니콘과 국내 예비유니콘의 사례 비교」, 『한국창업학회지』, 18(3), pp. 325~352.

하지만 국내 AI반도체 기업의 경우 자금조달, 협력·제휴(수요기업과의 파트너십), 시장진출, 특히 해외시장 진출에 대한 구체적인 전략수립 역량이 부족하다. 상대적으로 우수한 가격 대비 성능(아직은 검증되지는 않은)만을 강점으로 이미 NVIDIA가 선점하고 있는 미국 중심의 AI연산을 위한 반도체 시장에서 경쟁하겠다는 야심찬 포부 만으로는 시장에서의 성공을 확신하기 어렵다. 따라서 활용 수요(AI연산/데이터센터 외 다양한 산업 내 AI반도체 적용 수요)와 진출국가(미국, EU 외 중동, 동남아시아 등 신흥시장) 등에 대해 보다 다양한 선택지를 놓고, 제품의 특성에 적합한 틈새시장 진입을 위한 전략을 마련할 필요가 있다.

3) 창업팀 및 혁신인재 경쟁력 부족

반도체 분야는 자본집약적 첨단기술의 특성을 가지고 있고, 제품 설계, 실증, 양산에 이르기까지의 공정이 복잡하기 때문에 창업에 있어서의 인적 자본의 접근성이 크게 제한된다. 특히 AI반도체의 경우 기존 반도체 산업에서 분화된 기술이자 제품으로 볼 수 있기 때문에 반도체 전반의 기술 및 사업에 대한 지식과 경험이 풍부한 상태에서 사업을 시작할 수밖에 없다. 따라서 국내외 AI반도체 선도 스타트업의 경우 대부분 반도체 엔지니어링 경험을 보유한 인력이 창업을 실행하는 경우가 많다. 퓨리오사AI의 백준호 대표는 AMD의 GPU 설계팀과 삼성전자 메모리사업부에서 근무한바 있으며, 리밸리온 박성현 대표는 인텔, 삼성전자 모바일사업부를 거쳤다. 사피온코리아의 류수정 대표도 삼성전자에서 GPU 개발 담당 임원으로 근무했으며, 딥엑스 김녹원 대표는 SK하이닉스, 브로드컴, IBM(딥러닝 하드웨어 프로세서 개발), Cisco, 애플(AP 설계 개발)을 거치며 다양한 반도체 설계 프로젝트에 참여했다. 또한 세미파이브 조명현 대표도 미국 MIT에서 반도체 설계 분야로 박사학위를 취득하고 반도체 팹리스 SiFive에서 Country Manager로 일한 바 있다. 이와 같이 AI반도체 스타트업을 창업하기 위해서는 반도체 칩 설계에 대한 기술적·산업적 이해도가 무엇보다 중요하게 여겨지기 때문에 반도체 설계 프로젝트에 참여한 이력이 풍부한 엔지니어 출신의 창업자가 많다. 따라서 AI반도체 등 시스템반도체 분야 전문 투자사(VC)들도 이러한 글로벌 선도 반도체 기업에서의 엔지니어링 경력자들이 주축이 된 스타트업에 투자하

는 것을 선호하는 경향이 뚜렷하다.¹⁴¹⁾

하지만 단순히 글로벌 반도체 기업의 엔지니어링 경험만으로 AI반도체 스타트업의 성공을 이끌 수는 없다. 일단 해당 기업에서의 경력이 얼마나 깊고, 또 CPU, GPU 등의 차세대 반도체 설계에 있어서 얼마나 핵심적인 역할을 담당했는지가 상대적으로 더욱 중요하다. 더불어 중요한 것은 엔지니어링 경험 뿐만 아니라 시장을 보는 눈이다. 앞서 강조한 바와 같이 AI반도체의 새로운 시장기회를 포착하고 BM 및 제품의 차별화된 경쟁력을 만드는 능력, 그리고 무엇보다도 스타트업의 성장과정에서 겪게 될 자금투자, 자원조달, 경영전략, 조직관리, 위험관리 등의 경영역량을 갖추어야 한다. 리벨리온의 박성현 대표는 반도체 기업의 엔지니어링 뿐 아니라 스페이스X와 같은 거대 수요기업, 모건스탠리와 같은 투자컨설팅 기업의 다양한 경력을 보유하였고, 사피온코리아의 류수정 대표도 SKT의 AI 액셀러레이터를 맡으며 AI반도체 시장의 성장가능성을 탐색할 기회를 가졌으며, 세미파이브의 조명현 대표도 전략컨설팅 기업 BCG에서 반도체 기업 전략 프로젝트를 수행한 바 있다.¹⁴²⁾

이렇게 대표 창업자의 AI반도체 엔지니어링 경력과 산업전문가로서의 경험이 조화를 이루는 것이 기업 성장에는 도움이 될 것으로 생각되나, 대표 창업자만큼이나 중요한 것은 바로 창업팀이다. AI반도체는 설계 및 양산 과정이 매우 복잡하고 전문화되어 있으며, 연구개발 외 수요기업과의 파트너십 등 사업운영 관점에서도 다양한 전문성이 요구되기에 창업팀이 어떻게 구성되느냐에 따라 기업의 생존에도 큰 영향을 미칠 수 있다. 이런 측면에서는 퀄리타스반도체의 사례와 같이 기존 재직기업에서의 팀워크가 도움이 될 수 있다. 퀄리타스반도체는 삼성전자 인터페이스 IP 개발 업무를 담당한 김두호 대표를 비롯해 이재철 CIP 팀장, 성창경 SIP 팀장, 최광천 AE 팀장 등 주요 임원들이 삼성전자에서 일한 경력이 있다.¹⁴³⁾ 이러한 인연으로 삼성 파운드리의 공식 파트너로서 MIPI를 독점적으로 점유할 수 있었다. 한편 사피온코리아와 같이 창업팀이 기존 대기업의 사내벤처 형태로 만들어질 수도 있다. 사피온코리아는 2016년 SK

141)전문가 인터뷰 내용 중 발췌

142) 혁신의숲 기업별 대표자 정보 참조, <https://www.innoforest.co.kr>

143) 한경비즈니스(2023.10.17.), 「삼성전자 출신이 창업한 퀄리타스반도체, 제2의 파두될까 [전예진의 마켓 인사이트]」, <https://magazine.hankyung.com/business/article/202310111417b>(검색일: 2025.1.1.)

텔레콤 내부 연구개발(R&D) 조직으로 출발해 2022년 분사된 AI반도체 팹리스(설계) 기업으로, SKT 출신의 엔지니어 및 사업전략 임원들이 모기업의 대규모 투자에 힘입어 초기 사업운영에 안정성을 꾀할 수 있었다.

하지만 국내에서 아무리 우수한 창업자 및 창업팀을 보유한다고 하더라도, NVIDIA의 시장 독점 현상에서 보듯 우수한 인재의 해외 유출로 인해 국내 혁신인재의 경쟁력은 상대적으로 떨어지기 마련이다. 메타(페이스북), 알파벳(구글), 아마존과 같은 핵심 수요 대기업은 모두 미국에 있으며, AI 서비스가 활성화되고 있는 거대 수요기반을 가진 대기업은 중국에 모여있는 현실에서, 국내 AI반도체의 창업 및 핵심 기술인력의 수준 향상을 기대하기는 쉽지 않다. 더욱이 관련 인력의 공급에 비해 국제적인 수요가 훨씬 큰 현 상황에서 기존 인력의 해외 유출 등으로 인해 기업의 혁신경쟁력 및 생존 자체가 위협받을 가능성이 높다. 따라서 AI반도체 기업의 생존 및 성장을 위해서는 무엇보다도 기술과 사업 영역에서 균형을 가진 창업팀과 핵심 개발인력의 유출을 막을 수 있는 인센티브가 필요하다.

나. 수요기반의 실증 및 사업화 역량 부족

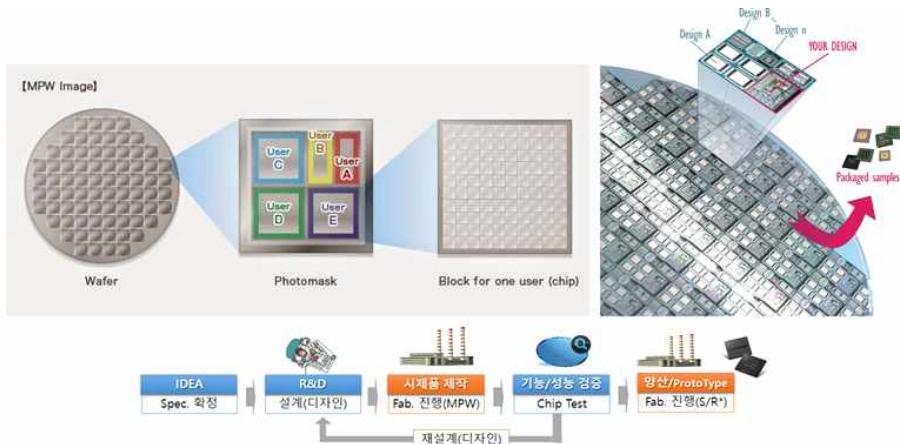
1) PoC(개념증명) 및 대규모 실증 역량 부족

AI반도체는 메모리반도체와 달리 명확한 스펙 상의 성능보다는 수요처가 원하는 제품의 특성에 맞추어 성능에 대한 인식 수준이 달라지기 때문에 개념증명(PoC: Proof of Concept)이 매우 중요하다. 특히 AI반도체 팹리스 회사가 설계한 칩의 테스트를 위해 시험적으로 칩을 제작할 필요가 있는데, 이때 파운드리에서 여러 회사에서 의뢰한 여러 개의 테스트 칩을 하나의 웨이퍼를 공유하여 제작하는 MPW(Multi-Project Wafer) 서비스를 활용하게 된다. MPW를 통해 상대적으로 낮은 비용으로 설계한 칩의 결과를 테스트할 수 있는 장점이 있으나, 여전히 초기 팹리스 스타트업이나 중소기업 입장에서는 초미세 공정의 경우 수십억 원 규모의 막대한 비용 부담이 존재한다.

이처럼 AI반도체 팹리스 스타트업들에게는 MPW가 자사의 시제품을 테스트하기 위한 중요한 수단이지만, 파운드리는 거대 수요기업들의 칩 생산을 위해 팝 가동률을

높이느라 MPW 기회를 충분히 제공하지 못하는 실정이다. 다만, 선도 혁신 팹리스 기업들의 우수한 시제품을 테스트할 기회를 제공함으로써 파운드리 기업들도 유망 AI 반도체의 양산을 위한 미래 핵심고객을 확보할 수 있다. 이런 측면에서 대만의 TSMC는 다양한 팹리스 스타트업들에게 연간 130여회의 MPW를 제공하는 반면, 한국의 삼성전자는 29회만의 MPW를 제공한 것으로 나타나 AI 반도체의 다양한 제품 사업화 및 생태계 구축에 있어 국내 시장의 한계를 드러내고 있다(2023년 기준).¹⁴⁴⁾

[그림 3-5] AI반도체 PoC를 위한 MPW의 개념 및 필요성



자료: 유병우(2024.3.25.), 「시스템반도체 산업동향」(발표자료), 시스템반도체설계지원센터.

한편 PoC 단계를 통과하더라도 AI 반도체 제품의 시장 활용을 위한 대규모의 실증 기회가 부족하다는 것도 역시 문제이다. AI 기술의 수요가 큰 자율주행, 클라우드 인프라, B2B AI 솔루션 등의 신산업 분야에서 국내 시장의 규모는 매우 제한적이다.¹⁴⁵⁾ AI 기술의 높은 잠재력을 사전에 예측하지 못한 정부에서는 관련 산업 분야에 대한 선도적인 기술투자 및 성장전략을 수립하지 못했고, 관련 시장에서의 규제 및 제약 조건들이 실증을 비롯한 본격적인 사업화 지원의 시작을 지연시켰다. 더 큰 문제는

144) 브릿지경제(2023.12.28.), 「‘脫삼성’ 가속화하는 K-팹리스…뇌관은 ‘MPW’ 횟수」,

<https://www.viva100.com/20231227010008242>(검색일: 2024.10.21.)

145) 김정호(2024.5.17.), 「AI반도체 기업의 스케일업 요인: 국내외 사례, 최근 연구 결과와 시사점」(발표자료).

정부 차원에서 AI 기술을 활용한 다양한 응용 분야의 성장 전략을 선제적으로 마련하지 못하면서, 실증을 통해 생성되는 막대한 데이터들을 활용할 기회가 제한된다는 것이다. 다양한 응용 분야에서의 한국 내 이용자·소비자들의 고유한 데이터를 확보하지 못하면서, 데이터 구입 및 활용에 더 많은 비용이 소요되고, 해당 산업의 경쟁력도 낮아지는 악순환이 이어지고 있다.

현재는 과학기술정보통신부가 정보통신산업진흥원(NIPA)을 통해 AI반도체(서버·엣지용) 기반 서버·제품(드론, 로봇, CCTV 등)에 다양한 AI 응용을 적용, 비공개 시범 테스트(성능 등) 등을 통해 AI반도체 실증 레퍼런스를 확보하기 위한 ‘AI반도체 응용 실증지원’ 사업을 2021년부터 수행하고 있는데, ① 드론, ② 로봇, ③ 국방, ④ 차량·모바일, ⑤ 자유 분야에 연간 88억여 원을 지원하는데 그치고 있다(2024년 기준).¹⁴⁶⁾ 따라서 상대적으로 열악한 AI반도체 기업에 대한 국내 민간투자 시장의 기능을 보완하기 위해서라도 정부 주도의 다양한 응용 분야의 실증사업이 확대될 필요가 있다.

2) 제품 양산 전략 및 레퍼런스 부족

AI반도체 설계기업이 PoC 및 실증 단계를 성공적으로 통과하더라도 결국 최종적으로 출시될 제품을 양산하기 위해서는 구체적인 양산 전략이 필요하다. 초기 단계인 기업의 경우 MPW를 통해 시제품을 만들 수 있지만, 양산을 위해서는 Single Run,¹⁴⁷⁾ 더 나아서는 전략 Foundry와의 협력 관계를 만들어야 한다. 물론 이러한 AI반도체 기업의 성장단계에 따른 양산 전략에는 보다 막대한 규모의 자본과 인력이 투입되며, 이는 스타트업에게는 큰 한계로 작용한다. 실제로 퓨리오사AI, 리밸리온, 사피온코리아와 같은 국내 선도 AI반도체 스타트업조차도 매출 성과가 미미한 것은 테스트를 위한 납품 수준으로 본격적인 양산 단계에 진입하지 못하였음을 의미한다.¹⁴⁸⁾ 결국 해외 빅테크 기업들을 고객사로 유치하기에 필요한 레퍼런스를 충분히 축적하지 못한 상황이라고 볼 수 있다.

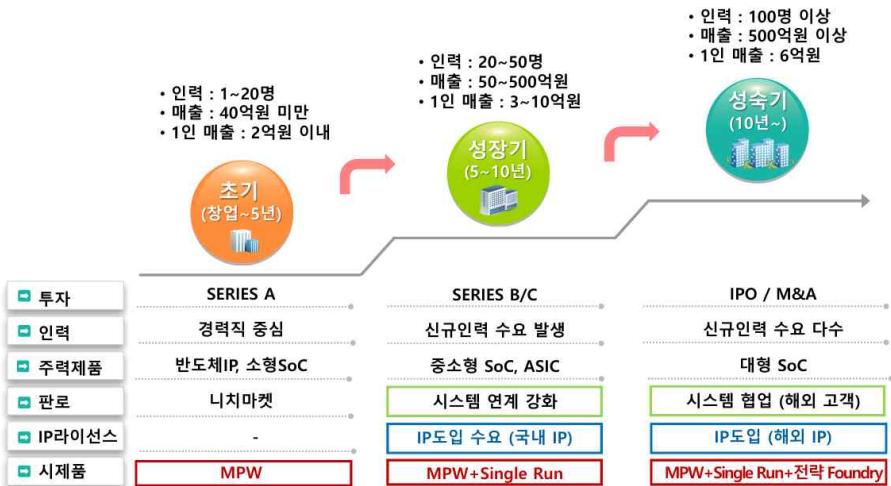
146) 정보통신산업진흥원(NIPA) AI반도체 응용실증지원 소개 홈페이지,

<https://www.nipa.kr/home/bsnsAll/3/detail?bsnsDtlItemNo=638&tab=1>(검색일: 2024.10.21.)

147) 하나의 웨이파에 여러 종류의 테스트 칩을 제작하는 MPW와 달리 하나의 반도체만을 생산하는 것을 의미

148) 2023년 기준 매출실적은 퓨리오사AI 36.2억 원, 리밸리온은 27.3억 원이며, 사피온코리아는 매출 실적이 없는 것으로 나타남(혁신의숲, <https://www.innoforest.co.kr/> 참조)

[그림 3-6] AI반도체 설계기업 단계별 성장요건



자료: 유병우 (2024.3.25.), 「시스템반도체 산업동향」(발표자료), 시스템반도체설계지원센터.

한편 설계된 AI반도체의 시험 결과 성능이 아무리 우수하다고 하더라도, 수요자의 반응을 이끌어내기 위해서는 결국 제품의 활용성과 안전성 등의 실질적인 검증이 필요하다. 앞서 언급한 정부 주도의 실증사업들이 상대적으로 판로 개척과 테스트베드 확보가 어려운 AI반도체 스타트업들에게 큰 도움이 되겠으나, 궁극적으로는 시장 수요자가 만족할 만한 성능의 제품을 제공하기 위해서는 수요자들에게 직접 제품을 판매한 실적이 있어야 한다.

NVIDIA, AMD 등이 AI반도체의 최강자로 군림하게 된 것은 아마존, MS, 구글 등 자국 빅테크 기업 영향이 크다. 역시 사피온코리아, 리밸리온, 퓨리오사AI 등이 상대적으로 고평가를 받은 건 각각 SK텔레콤, KT, 네이버, 카카오 등 동맹군이 있었던 덕분이다. 결국 우리나라에서도 AI반도체 수요를 초기에는 내수시장에서 확보해야 한다는 의미이다.¹⁴⁹⁾ 하지만 정책적 차원에서는 삼성전자, 현대자동차, KT와 같은 대기업을 정부가 직접적으로 도와야 하는지에 대한 딜레마가 존재하며, 대기업 입장에서도 전체 비용의 일부에 불과한 AI 반도체 구매 수요에 대해 스타트업을 우대하는 전향

149) 더밸(2024.8.14.), 「AI 캐즘 우려? 자체 생태계 조성 기회 ‘뭉치면 산다’」, <https://bully.kr/BTOJY32>(검색일: 2024.10.24.)

적 정책을 채택할 유인이 없는 것이 현실이다.

따라서 이러한 시장에서의 레퍼런스 확보를 위해서는 국내 AI반도체 기업들이 자사의 신제품을 글로벌 대기업에 바로 납품하려는 계획보다는 국내의 다양한 수요기업들을 대상으로 자사의 제품의 가능성을 먼저 탐진하여 의미있는 매출을 만들어내려는 노력이 필요하다. 즉, 대기업 수요를 한번에 만족하기보다는 소규모, 중견 규모의 업체에 여러 납품 경력을 가지며 레퍼런스를 만드는 것이 현실적인 방법이라는 것이다. 처음부터 제품을 개발할 때 기술 특성(Specification) 협의부터 고객사와 상의해야 하지만 이는 너무 이상적인 것으로, 다양한 수요기업과의 제품개발 이력을 통해 제품 성능과 활용성에 대한 평판을 축적할 수 있다. 그리고 이러한 국내 시장에서의 레퍼런스가 축적된다면 이에 대한 평판을 통해 해외 빅테크 수요기업들로의 판로 확장 기회를 모색할 수 있다.¹⁵⁰⁾

3) 칩 제조를 위한 전후방 밸류체인(공급망) 미흡

반도체 산업은 일반적으로 IDM(종합반도체회사) - Foundry(제조) - Fabless(설계) - Chipless(IP Provider) 등으로 분업화되고 공생·협력하는 생태계로 발전하고 있다. 특히 AI반도체는 데이터 저장이 주 용도인 메모리반도체와 달리 여러 기능을 단일 칩에 통합하여 경제성 및 편의성을 극대화하고, 소프트웨어와 융합하여 시스템의 고성능화, 소형화, 저전력화 및 스마트화를 추구하기 때문에, 설계·제조·패키징·테스트 등 의 산업생태계가 분화되어 있어 다품종 생산에 특화되어 있는 기술집약적 산업 특성을 가지며, 공정별로 특화된 기업에 의해 분업화가 가능하다.¹⁵¹⁾

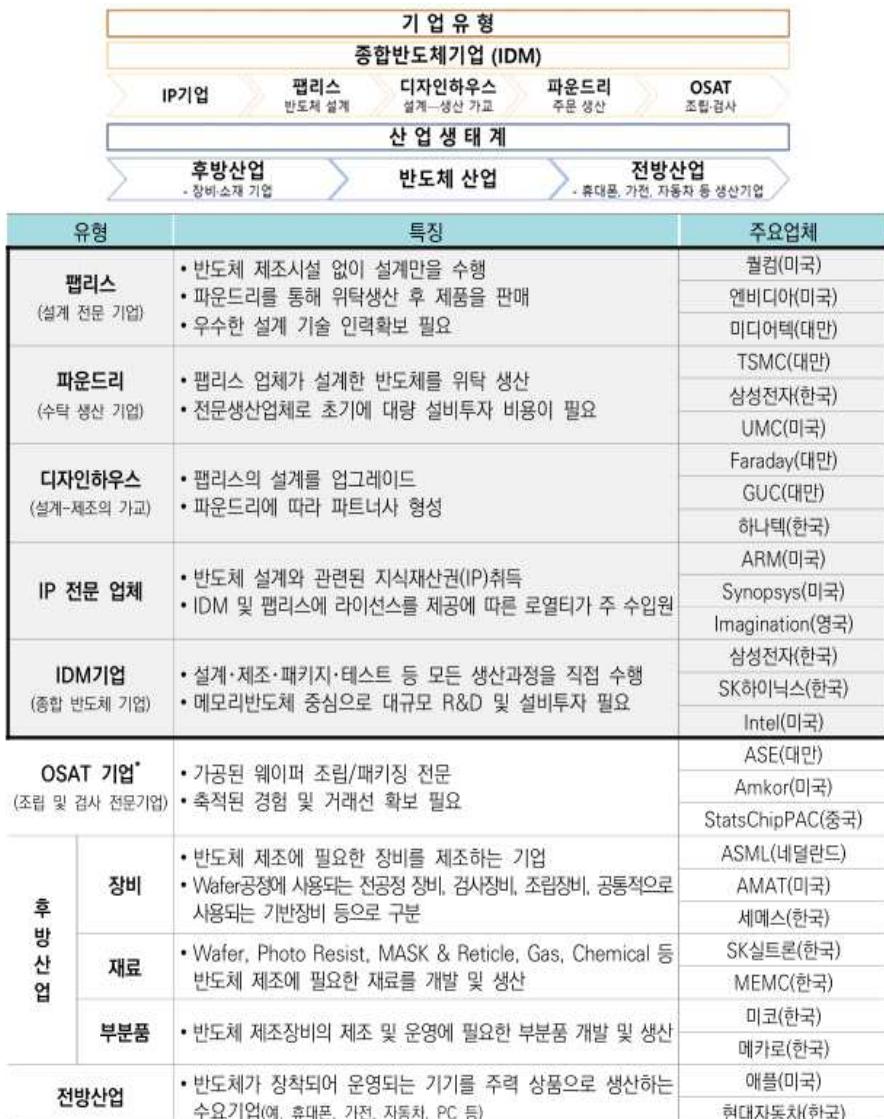
특히 최근 설계(팹리스)와 생산(파운드리)의 분업화 확대로 생산(파운드리) 기업이 국가 간 반도체산업 경쟁의 핵심 열쇠를 쥐고 있다. 이는公正이 미세화될수록 IP 도입 비용이 증가하고, 개발비가 기하급수적으로 늘어나면서 생산기업 주도의 밸류체인이 강화되기 때문이다. 하지만 한국의 경우 메모리반도체 분야의 삼성전자, SK하이닉스 등의 IDM 기업 중심으로 생태계가 구성되어 있어, AI반도체의 분업화 흐름에 맞는

150) 전문가 인터뷰 내용 중 발췌

151) 중소벤처기업부·중소기업기술정보진흥원(2024), 「중소기업전략기술로드맵 2024~2026 「시스템 반도체」」.

선도적인 팹리스, 파운드리, 디자인하우스, IP 기업들의 역량이 부족한 실정이다([그림 3-7] 참조).

[그림 3-7] AI반도체 산업의 가치사슬 및 참여기업



자료: 채명식(2022), 「시스템반도체」, 「KISTEP 브리프」, 01, 한국과학기술기획평가원, p. 4..

실제로 팹리스 기업들은 미국 또는 대만 기업들이 선도적이며, 파운드리는 TSMC라는 거대한 장벽에 가로막혀 있고, 디자인하우스는 팹리스 기업과의 긴밀한 협력 관계의 특성에 따라 주로 대만기업들이 우세하며, IP 기업은 미국, 영국 등 반도체 선진국에게 절대적 약세를 나타낸다. 게다가 반도체 양산을 위한 장비, 재료, 부품 등의 후방산업의 경우에도 일본의 소재·부품 전문기업이나 유럽의 거대 반도체 장비기업들에 밀려 경쟁력 확보에 어려움을 겪고 있다.

한편 가전, 자동차, IT 기기, 플랫폼기업 등 전방산업의 경우에도 글로벌 선도기업에 비해 국내 기업의 AI 기술 관련 변화대응 역량이 떨어지면서 국내 AI반도체 산업 공급망 강화에 한계를 노출하고 있다. 해외에서는 알파벳(구글), 메타(페이스북), 아마존과 같은 거대 플랫폼 기업이 자체 생성AI 모델 개발 및 데이터 학습을 위해 AI반도체 수요를 폭발적으로 발생시키고 있으며, 글로벌 자동차, 가전기업들은 자사의 제품에 활용될 최적의 성능을 가진 AI반도체 개발을 위해 반도체 기업과 전략적 제휴를 강화하고 있다. 하지만 국내에서는 아직 AI반도체의 구체적인 활용 수요를 파악하지 못한 채, 미래 기술개발을 위한 단순 협력에 그치고 있어, 전방산업의 성장에 따른 AI반도체 산업경쟁력 강화의 긍정적인 파급효과를 기대하기 어려운 실정이다.

다. 산업 내 이해관계자와의 신뢰 기반 협력생태계 구축 미흡

1) 다양한 수요기업과의 신뢰·협력 관계 구축 노력 부족

AI연산을 위한 반도체는 CPU 수준의 복잡한 설계기술이 필요하다기보다는 수요기업의 환경조건에 맞는 맞춤형 제품을 제공하는 것이 핵심이기 때문에 무엇보다도 제품 설계 단계에서부터 반도체 설계·생산기업과 수요기업과의 긴밀한 협력이 필요하다. 즉, AI반도체를 만드는 공급기업과 클라우드 사업자, 최종 AI서비스 사업자 간의 원활한 의사소통이 매우 중요하다([그림 3-8] 참조).

[그림 3-8] AI반도체 산업생태계 내 협력 구조



자료: 유병두(2024.3.25.), 「시스템반도체 산업동향」(발표자료), 시스템반도체설계지원센터.

물론 국내 AI반도체 선도 스타트업들도 물론 네이버, 카카오, KT, SKT 등 국내 대표 플랫폼 및 클라우드 기업들과 투자 파트너십을 구축하고 공동 기술개발 및 실증 등에 있어서의 협력 관계를 강화하고자 노력하고 있다. 하지만 국내 AI반도체 기업의 경우 칩 설계 단계부터 제품 구현에 이르기까지의 과정에서 해외기업에 비해 상대적으로 수요기업과의 협력보다는 주로 폐쇄형의 자체 기술개발을 선호하는 경향이 있다. 또한 기업 간 협력에 있어서도 수요기업보다는 같은 칩 설계기업끼리의 순수한 기술 교류를 선호한다. 반면 글로벌 선도 칩 공급기업인 NVIDIA의 경우에는 수요기업과 적극적으로 소통하여, 수요기업이 제기한 문제는 반드시 다음 제품 설계에 반영하려고 노력한다. 따라서 우수한 성능을 가진 칩이 아닌 수요기업이 필요로 하는 성능을 갖춘 칩을 만들어낼 수 있다.¹⁵²⁾

한편 AI반도체 칩 설계의 사업화를 위한 실증 및 양산에 있어서의 파운드리와의 협력 관계 구축도 중요한 문제이다. 앞서 언급한 바와 같이 AI반도체의 PoC 및 실증을 위해서는 파운드리와의 협력 관계가 필수적이나, 삼성전자의 MPW 서비스 제공 및 AI반도체 스타트업 지원생태계 구축에 있어서의 소극적인 자세로 인해 공급 차원의 신뢰 기반 벤류체인 구축에 애로가 발생하고 있다. 실제로 삼성전자 파운드리와의 협력관계를 우선시했던 국내 주요 AI반도체 팝리스들이 새로운 칩 양산을 위해 TSMC

152) 전문가 인터뷰 내용 중 발췌

와의 협력으로 전환하거나 병행하는 사례가 늘어나고 있다. 퓨리오사AI의 경우 1세대 칩 ‘워보이’는 삼성전자 14나노 공정을 사용했지만, 2세대 칩 ‘레니케이드’는 TSMC 5나노 공정을 선택했으며, 퓨리오사AI가 4분기 출시를 목표로 하고 있는 차세대 AI칩 ‘레니케이드S’ 또한 TSMC 5나노 공정을 선택할 예정이다.¹⁵³⁾ 딥엑스 역시 기존에는 삼성전자 파운드리와 협력하였지만, 2024년 신규로 개발한 ‘DX-V3’ SoC(시스템온칩)는 TSMC의 12나노 공정을 도입하여 삼성전자와 TSMC 파운드리 양쪽을 모두 활용할 계획이다.¹⁵⁴⁾ AI반도체는 개발에서 양산에 이르기까지 수천억 원 이상의 막대한 자금이 투입되는 산업이고, 칩 하나의 양산과 매출에 기업의 생존이 달려있기 때문에, 결국 기업들은 설계된 칩의 최적화가 가능한 파운드리 공정을 선택할 수밖에 없다. 따라서 삼성전자 파운드리는 TSMC와 같은 소형 팝리스 스타트업들과 상생하여 성장하는 생태계 조성 전략을 심도있게 고민해야 할 시점이다.

또한 앞서 레퍼런스 확보 이슈에서도 언급한 바와 같이, AI반도체 기업은 다양한 수요기업과의 협력 관계 속에서 제품의 상세한 피드백을 받아 이를 제품 개선에 반영하고자 노력할 필요가 있다. 기술개발에 치중했던 사업 초기에는 개발 엔지니어 중심의 기업운영이 필요했을지라도, 제품 양산과 수요기업에의 활용 단계에 접어든 현재의 기업성장 단계에서는 수요기업과의 능동적인 의사소통 및 협력관계 구축, 원활한 고객 피드백에 필요한 전문인력 및 조직 구축에 더 많은 관심을 기울여야 한다.

2) 공동 기술개발을 위한 산학연 협력 체계 구축 미흡

AI반도체를 연구하는 대표적인 공공연구기관에는 한국전자기술연구원, 한국전자통신연구원 등이 있다. 한국전자기술연구원의 반도체·디스플레이연구본부 SoC플랫폼연구센터에서는 NPU 아키텍처 및 지능형 SoC 설계 기술 연구를 수행하고 있으며, 한국전자통신연구원의 경우 지능형반도체연구본부에서 초거대AI반도체, 지능형엣지반도체, PIM인공지능반도체, 지능형센싱반도체 설계 기술을 연구하고 있다. 그밖에 한국

153) 디디넷코리아(2024.10.2.), 「국내 AI 반도체, 삼성·TSMC 파운드리 다각화」, <https://zdnet.co.kr/view/?no=20241002152042>(검색일: 2025.1.1.)

154) 상동

전기연구원은 전력반도체연구단에서 손실이 적은 차세대 전력반도체와 다양한 전력반도체 집적회로(파워 IC)를 개발하고 있으며, 대학에서는 주로 차량용 반도체, 바이오·헬스케어용 반도체, 보안용 반도체 등 AI 활용 산업에 필요한 특화형 반도체 기술 개발에 주력하고 있다(〈표 3-3〉 참조).¹⁵⁵⁾

〈표 3-3〉 AI반도체 관련 주요 분야별 연구조직(대학·출연연) 현황

분야	기관/소속	연구분야
전력반도체	한국전자기술연구원	고온동작반도체용 소자·공정 기술, 고효율 PMIC 기술
	한국전기연구원	SiC 기반 집적회로 기술, 전력반도체 소자·공정 기술
아날로그·디지털 제어 반도체	한국전자기술연구원	전동화 기반 동력전달 고도화 설계 기술
	한국전기연구원	파워시스템 온-칩 및 칩형 전원장치 기술
인공지능 및 시스템반도체 설계 IP	한국전자통신연구원	인공지능 반도체 프로세서 코어 설계 기술, 반도체-SW 융합을 위한 SW-SoC 기술
	한국전자기술연구원	인공지능 반도체 프로세서 코어 설계 기술, 전력반도체 신뢰성 기술 개발 및 시험 인증

자료: 중소벤처기업부·중소기업기술정보진흥원(2024), 「중소기업전략기술로드맵 2024~2026 「시스템 반도체」」의 내용을 바탕으로 연구진 재작성

하지만 AI반도체의 기술개발의 속도가 빠르고, PoC(MPW)에 막대한 자금이 필요하며, 고객기업의 활용 수요에 따라 맞춤형 제작이 필요하기 때문에 과거의 메모리반도체 기술과는 달리 산업생태계 내에서 공공연구기관의 역할이 다소 모호해진 측면이 존재한다. AI반도체 설계기업들도 칩 설계 등에 있어서의 직접적인 연구협력보다는 실증·사업화 과정에서의 지금 및 설비 지원에 더 큰 수요를 제기하고 있으며, 따라서 공공연구기관보다는 수요 대기업을 기술협력의 파트너로서 더 선호하는 경향이 있다.

한편 민간에서는 최근 AI반도체 설계 분야에서 두각을 나타내는 팹리스 기업들이 탄생하고 본격적인 제품 실증 및 양산을 준비하는 단계에서 기업들과의 다양한 협력 및 교류 수요가 발생하고 있으며, 이러한 수요 충족을 위해 AI반도체를 비롯한 국내 시스템반도체 산업 및 기업을 대표하는 협·단체들이 활발히 조직되고 있다. 먼저 한국 패리스산업협회(KFIA)는 2020년 시스템반도체 포럼으로 시작, 이듬해 한국 패리스 연합을 거쳐 2022년 협회로 공식 출범하였다. 정부 및 의회에 패리스 업계의 목소리를

155) 중소벤처기업부·중소기업기술정보진흥원(2024), 「중소기업전략기술로드맵 2024~2026 「시스템 반도체」」.

대변하고, 한국 시스템반도체 설계 산업이 성장할 수 있는 생태계 구축을 위해 설립되었다. 협회 내 Automotive-Industry, Mobile-IoT, Intellectual Property, System Interface Solution, Artificial Intelligence 분과에 129개사가 가입되어 있다.¹⁵⁶⁾

한편 기존의 반도체 산업 분야의 대표적인 기업 단체인 한국반도체산업협회(KISA)는 시스템반도체 연구개발 지원의 일환으로 시스템반도체 융합얼라이언스, 차세대반도체 기술개발사업, 파워반도체 상용화 기술개발, 차세대전력반도체 소자·제조 전문인력 양성사업, 글로벌 수요연계 시스템반도체 기술개발 등의 다양한 사업을 지원하고 있다. 또한 KISA에서는 시스템반도체설계지원센터(ICS)를 설립하여 국내 시스템반도체의 역동적 생태계 기반 구축과 산업 경쟁력 강화를 위해 기업성장 환경조성, IP-Bank 플랫폼, 설계전문교육 프로그램 등을 운영하고 있다.¹⁵⁷⁾

하지만 이러한 AI반도체 기업 대상 협·단체 조직 노력에도 불구하고, AI반도체 산업 생태계 내의 신학연 협력은 활발하게 이루어지지 못하고 있는 실정이다. 이는 AI반도체 분야의 선도기업들조차도 아직 시장성을 검증받지 못할 만큼 기술개발 단계와 역량이 초기 단계이며, 기술개발을 통한 제품의 활용수요도 불명확하기 때문이다. 아울러 AI반도체 개발 인력 자체가 양적으로 부족하여 산학연 각 기관에 다양하게 분포되어 있지 못하며, 상대적으로 나은 처우를 보장받는 글로벌 빅테크 기업들에 우수한 인력이 많이 포진되어 있어 양질의 산학연 연구협력이 이루어지기 어렵다. 또한 개발자들의 지식공유 및 교류를 위한 커뮤니티나 정부 주도의 AI반도체 스타트업 간 혹은 수요 대기업과의 공동 R&D 프로젝트 등 협력 플랫폼 구축 노력도 부족한 실정이다.

3) 전략적·협력적 동반자로서의 대기업의 역할 미흡

앞서 여러 차례 언급한 바와 같이 AI반도체 분야는 칩 자체의 성능보다는 누가 가져다 쓸 것인가가 더 중요하므로 수요처와의 파트너십 구축이 필수적이다. 하지만 한국의 반도체 대기업이나 선도 플랫폼 기업들은 AI 시대에 대비한 자체 칩 개발에도 늦었

156) 한국팹리스산업협회 홈페이지, <https://www.k-fabless.com/kr/index.php>(검색일: 2024.10.21.); 한국팹리스산업협회 내부자료(2024.4.23.), 「팹리스 스타트업 육성을 위한 정책방향성 제언」.

157) 한국반도체산업협회 홈페이지, <https://www.ksia.or.kr/index.php>(검색일: 2024.10.21.)

을 뿐만 아니라, AI반도체 기업들과의 협력 관계 구축에도 매우 소극적인 모습을 보였다. 특히 사업화를 위한 막대한 자금이 투입되는 AI반도체의 특성 상 설립 초기 단계에서부터 재무적 투자자보다는 국내 대기업의 전략적 투자자로서의 적극적인 참여가 필요하다. 대기업들이 자체 스타트업 투자 조직인 기업벤처캐피탈(CVC: Corporate Venture Capital) 등을 활용하여, AI반도체 스타트업들의 제품개발은 물론 실증을 통한 사업화를 적극적으로 지원함으로써, 궁극적으로 대기업들의 자체 수요에 적합한 성능을 가진 칩을 개발할 수 있도록 유도해야 한다.

AI반도체 글로벌 선도기업인 영국의 그래프코어(Graphcore)의 경우에도 Dell, Bosch, 삼성전자 등은 Series A부터, Microsoft, BMW 등은 Series D부터 전략적으로 투자에 참여하여, 실증 및 양산을 위한 대규모의 자금을 지속적으로 확보할 수 있었다. 국내의 경우에도 AI반도체 스타트업 중 가장 일찍 창업한 퓨리오사AI의 경우 네이버(D2SF) 외 별다른 대기업의 관심을 받지 못해, Series C 단계까지 가서야 800억 원의 투자를 유치하였으나, 후발주자였던 리밸리온의 경우 Series A 단계에서 KT가 전략적 투자자로 참여하면서 단독으로 300억 원(같은 Series A 단계의 클럽딜 투자액까지 포함하면 920억 원 규모), Series B 단계에서는 1,650억 원의 투자유치에 성공하였다.¹⁵⁸⁾ 또한 SKT의 사내벤처 형태로 탄생한 사피온코리아의 경우에도 분사(Spin-off) 과정에서 SKT, SK하이닉스, SK스퀘어 등 SK 계열사를 통해 800억 원의 초기 투자를 유치하였으며, 이후 Series A 단계에서의 투자 과정에서도 높은 기업가치를 인정받아 600억원의 투자금을 유치할 수 있었다.¹⁵⁹⁾ 이와 같이 선도기업으로서의 기술력은 비슷하게 여겨지더라도 수요 대기업이 전략적 투자자로 얼마나 적극적으로 참여하는지에 따라 미래 성장가능성을 대변하는 기업가치가 크게 달라질 수 있다.

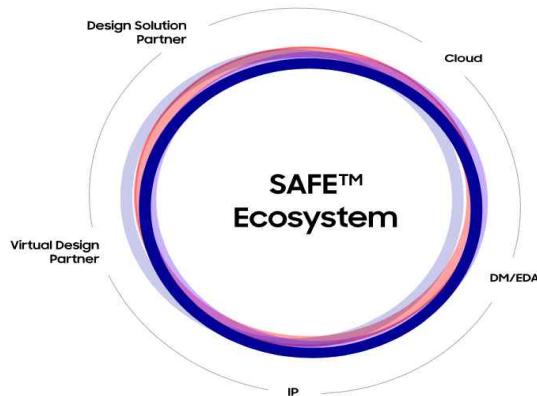
AI반도체 산업 내에서의 또 다른 문제점은 바로 국내 대-중소기업 간 하청구조에서 발생하는 고질적인 기술탈취 및 갑질문제이다. 아직도 일부 대기업에서는 납품처가 새로운 제품을 개발하기 위해 협력을 제안해 오면, 납품처의 제품에 대한 전담팀을 두고 똑같은 제품을 개발하여 나중에 비용절감을 위한 협상의 카드로 활용하려는 관행이 존재한다. AI반도체의 경우에도 파운드리를 통한 칩 설계 기술의 유출, 수요 대

158) 반면 퓨리오사AI는 Series A 단계에서 80억 원, Series B 단계에서 800억 원의 투자를 받음

159) 혁신의 숲 홈페이지, <https://www.innoforest.co.kr>(2024.5.16. 기준)

기업의 협력 R&D를 빌미로 한 기술탈취의 가능성성이 존재하며, 따라서 현재 추진 중인 AI 반도체 스타트업과 대기업과의 협력 관계에도 향후 잠재적인 위험 요소로 작용할 가능성이 있다.¹⁶⁰⁾ 다만, 최근 삼성전자의 경우 협력적 파운드리 생태계 구축을 위한 SAFE™(Samsung Advanced Foundry Ecosystem) 프로그램을 운영, 생태계 파트너, 고객과 긴밀한 협력 관계를 형성하여 공정 설계 키트(PDK), 설계 방법론(DM)을 따르는 레퍼런스 플로우, 설계자산(IP)과 설계지원을 포함한 인증된 핵심 설계 구성요소를 기반으로 경쟁력 있고 강력한 단일칩 체제(SoC) 설계를 제공하기 위해 노력하고 있다([그림 3-9] 참조).

[그림 3-9] 삼성 파운드리 SAFE 프로그램의 개념



자료: 삼성전자 파운드리 SAFE 소개, <https://semiconductor.samsung.com/kr/foundry/safe/>

2. 투자

가. 막대한 초기 투자 규모 및 시기

1) 막대한 초기 투자 필요성

AI 반도체 팩리스 기업들은 설계 과정에서 해외 IP(Intellectual Property) 및 EDA (Electronic Design Automation) 사용과 관련해 발생하는 높은 라이선스 비

160) 전문가 인터뷰 내용중 발췌

용과 동 설계의 제작을 위한 파운드리의 선단 공정 (5nm이하)에서 천문학적인 비용을 감수해야하는 문제점을 직면하고 있다. IP는 반도체 칩 설계에 활용하는 논리회로나 집적 회로 (Integrated circuit)의 설계 단위를 의미¹⁶¹⁾하며, AI 반도체 설계에서는 Processor, memory controller, interface protocol 과 같은 블록으로 나뉜 형태로 존재하며, Synopsys의 설명에 따르면 복잡한 형태의 시스템온칩¹⁶²⁾ (System-on-Chip, SoC) 형태의 반도체 설계를 가능하게 하는 주요한 원천이다, EDA는 전자 설계 자동화 도구의 줄임말로서 IP를 활용하여 설계된 칩들이 Design House나 Foundry를 통한 생산 과정 이전에 설계가 구조적으로 작동하는지, 규격에 맞는지의 검증을 하는 소프트웨어이다.

AI 반도체 설계의 핵심 요소 중 하나는 최신 제조 공정의 활용이다. 이는 SoC의 설계에서 집적도를 높여서 다양한 부문에 활용될 수 있도록 블록화된 IP를 추가하며, 이를 생산하기 위하여 향상된 선단 공정 (초미세화 공정)을 활용하기 때문이다. 이러한 선단 공정의 사용은 매우 높은 비용을 수반한다. 3나노 공정을 활용한 반도체 디자인 비용은 약 7200억 원에 이르며, 5나노 공정을 사용하는 경우에도 많은 비용이 필요하다¹⁶³⁾. TSMC와 같은 글로벌 선단공정의 61.2%의 점유율을 지닌 선도 파운드리 회사 역시 3%에서 8%까지 3나노 및 5나노 공정의 가격을 인상하였으며, 이는 선단 공정의 사용 자체가 반도체 설계에서 큰 비용을 차지하고 향후에도 증가할 추세임을 단적으로 보여준다¹⁶⁴⁾. 즉, 선단 공정의 사용은 막대한 자본 투입을 의미하며, 앞선 설계 자산(IP) 구입, 설계 자동화(EDA) 툴 구매, 시제품 생산을 위한 포토 마스크 비용 까지 포함하는 경우 막대한 초기 투자의 필요성은 더욱 중요한 문제점으로 작용할 소지가 크다¹⁶⁵⁾.

161) Tech Target, “Definition intellectual property core (IP core)”. <https://www.techtarget.com/whatis/definition/IP-core-intellectual-property-core> (검색일 24.11.25.)

162) Synopsys(22.11.14.), “Understanding System-on-Chip (SoC): Components, Construction, & Capabilities”, <https://www.synopsys.com/blogs/chip-design/system-on-chip.html> (검색일 24.11.25.)

163) 파이낸셜뉴스(24.08.07.), “TSMC, 3나노 공정 단가 인상 ‘만지작’… 삼성전자 기회 잡나.” <https://www.fnnews.com/news/202408071819127100>. (검색일 24.11.25.)

164) 머니투데이(24.08.08). “TSMC, 3나노·5나노 반도체 가격 곧 최대 8% 인상” <https://news.mt.co.kr/mtview.php?no=2024080821574436402>. (검색일 24.11.25.)

165) 한국무역협회(24.11.04.), “출발점에 선 韓 AI 반도체, 생태계 구축 시급.” https://kita.net/board/totalTradeNews/totalTradeNewsDetail.do;JSESSIONID_KITA=621FCA8F01BD5AFDD728878491DA120C.Hype

AI 반도체 설계에는 IP 라이센스와 EDA 툴이 필수적으로 사용된다. 특히, 국산 반도체 설계 기업들은 케이던스(Cadence)와 시냅시스(Synopsys)와 같은 해외 IP 및 EDA 툴에 종속적¹⁶⁶⁾일 수밖에 없으며, 대안적으로 사용할 국내 기업도 경쟁력이 높다고 볼 수 없다¹⁶⁷⁾, 하지만, EDA 툴과 IP 자산들은 반도체 설계, 검증, 제조 과정에서 필수적이며, 국내 팹리스 기업이 설계에 사용하는 IP는 라이센스나 로열티 비용을 지급해야하며, 사용을 위하여 막대한 비용(21년 기준 10나노이하의 공정에 쓰이는 SerDes IP의 가격은 80억원 안팎 수준¹⁶⁸⁾), 달러 가격 증가 및 생성형 AI보급 등 외부적 요인 고려시 상승 확률이 높음)을 지불해야한다. 시냅시스의 EDA 툴은 반도체 설계 전반에 걸쳐 활용되며, 케이던스와 같이 다양한 공정과 호환되는 솔루션을 제공하고 있다. 더불어, 한국의 팹리스 기업들에 대한 지원도 강화¹⁶⁹⁾하고 있으나, 장기적 관점에서는 한국의 팹리스 부문의 국외 기업 의존도 측면에서는 긍정적인 측면만 존재한다고 볼 수 없다.

기술 발전과 시장의 수요 기업 및 최종 소비자의 변화에 따라 지속적인 업데이트와 개선이 필수적이기 때문에, 설계 부문만을 보더라도 제품의 출시 전 뿐만 아니라 출시 이후에도 장기간으로 지속 가능한 연구개발(R&D) 투자가 필요하다. 서버용 반도체의 95% 이상의 시장점유율을 자랑하는 NVIDIA 역시 자사의 AI 반도체 H100을 지속적으로 업데이트하면서 시장에서 경쟁력을 유지하기 위해 끊임없이 투자¹⁷⁰⁾하고 있다. 다수의 선행연구와 동 분야 전문가 인터뷰를 통해서도 지적되었듯이, 국내 AI 반도체 스타트업들도 수익성 개선의 필요성에 직면하고 있으며, 많은 자금을 조달했음에도 불구하고 AI 반도체 개발에 필요한 비용과 인력을 충분히 확보하기 어렵다. 이러한 상황에서 장기간의 지속적 투자는 기업 생존에 필수적이며, 이를 통해 기술적 우위를 확보하고 시장에서 경쟁력을 유지할 수 있다¹⁷¹⁾.

r?no=86511&siteId=1. (검색일 24.11.25.)

166) 디일렉(23.05.23), “‘국내 시스템 반도체 생태계 육성 위해 EDA 툴 국산화해야’.”

<http://www.thelec.kr/news/articleView.html?idxno=21239> (검색일 24.12.16)

167) 차명식(2022), 「시스템반도체」, 『KISTEP 브리프』, 01, 한국과학기술기획평가원.

168) 전자신문(2021.12.21.), “해외 ‘반도체 IP’ 가격 급등…‘첨단공정’ 속탄다”. <https://www.etnews.com/20211221000176> (검색일 24.12.16.)

169) 디일렉(24.07.11.), “시냅시스 “반도체 설계에 생성형 AI 도입”.” <http://www.thelec.kr/news/articleView.html?idxno=29028>. (검색일 24.12.16.)

170) kotra(24.11.04.) “AI 반도체에 투자가 쏟아지는 이유 – ① 미래 먹거리 좌우하는 AI 반도체.” .

https://dream.kotra.or.kr/kotranews/cms/news/actionKotraBoardDetail.do?SITE_NO=3&MENU_ID=180&CONTENTS_NO=1&pNttSn=212905. (검색일 24.12.16.)

171) 시사저널e, (23.09.25.) “‘AI반도체 초기시장 선점 중요…정부, 실증사업·R&D 확대해야’”, <https://www.sisajo>

2) 어려운 중후기 투자 연계

국내 주요 AI 반도체 팝리스 스타트업들은 제품 출시 직전 단계에서 다양한 기술적, 시장적, 재정적 리스크를 해결해야하는 상황이다. 동 리스크들은 기업들의 성장과 생존에 직간접적으로 영향을 미치고 있으며, 가장 크게 영향을 미치는 점은 매출이 증빙되지 않은 상황에서의 R&D를 위한 대규모의 투자유치에서의 문제를 야기한다는 점이다. 우선 기술적 리스크는 신제품 개발이 초미세 공정(예: 4~5 나노미터)으로 이동함에 따라 급격히 증가할 가능성이 높다. 앞서 침단 반도체 설계에 필요한 IP와 EDA공정에 대한 높은 비용을 언급하였다. 침단 반도체의 구현에는 설계자신(IP) 및 전자설계자동화(EDA) 툴의 구매가 필요하며, 선도적 R&D를 위하여 계속된 기업의 자체적 자금 조달이 절대적임을 명시한다. 예를 들어, 퓨리오사AI, 리밸리온, 딥엑스는 각각 초미세 공정을 사용하여 AI 반도체를 제조하면서 수백억 원의 연구개발(R&D) 비용을 지출¹⁷²⁾하고 있다. 동 기업의 기존 누적 투자액이 높은 상황이더라도, 급변하는 AI반도체 생태계 내에서 기술 부문에 대한 투자는 장기적으로 증가하기 때문에, 연계되는 투자의 중요성은 극대화된다. 시장 리스크는 NVIDIA와 같은 글로벌 경쟁자들이 강력한 생태계를 구축하고 있어 신생 기업들이 진입하기 어려운 시장의 구조에 기인¹⁷³⁾한다. NVIDIA는 자체 소프트웨어 플랫폼인 쿠다(CUDA) 및 cuDDN 등을 통해 하드웨어¹⁷⁴⁾ 뿐만 아니라 소프트웨어 생태계까지 독점하고 있어서, 동 영역 부문으로 사업을 확장하려고 하는 국내 AI반도체 팝리스 스타트업이 경쟁력을 갖추었더라도 성장하는데 제약¹⁷⁵⁾이 존재한다. 재정적 리스크는 초기 투자비용의 회수 문제와 관련이 있다. 국내 AI 반도체 스타트업은 기술적 리스크에 대한 해소 및 급변하는 시장 상황과 수요에 대응하기 위하여 대규모 자본 투자가 필요한 반면, 수익성을 빠르게 달성하기 어렵다. 리밸리온의 경우, 2020년 설립 이후 누적

¹⁷²⁾ [urnal-e.com/news/articleView.html?idxno=303677](http://sisajournal-e.com/news/articleView.html?idxno=303677). (검색일 24.12.16.)

¹⁷³⁾ 시사저널(24.08.18.), “韓 기술력 뛰어난데···NVIDIA 같은 기업 나오지 못하는 까닭”, <https://www.sisajournal-al.com/news/articleView.html?idxno=305228>. (검색일 24.12.16.)

¹⁷⁴⁾ ITDAILY(24.10.31.), “[시장동향] 격동의 ‘AI 반도체’ 시장···엔비디아 아성에 도전장”, <http://www.itdaily.kr/news/articleView.html?idxno=228271> (검색일 24.12.19.)

¹⁷⁵⁾ BusinessKorea(24.09.10), “Korean AI Startups Challenge NVIDIA’s GPU, Indicating Its Inherent Limitations”, <https://www.businesskorea.co.kr/news/articleView.html?idxno=224906> (검색일 24.12.19.)

¹⁷⁵⁾ Business Korea(2024.06.04.), “Korean AI Semiconductor Industry Lagging Behind in Design, Foundry, Packaging,” <https://www.businesskorea.co.kr/news/articleView.html?idxno=218390> (검색일 24.12.19.)

투자금 2800억 원을 유치했으며, 기업가치는 8800억 원에 달하지만, 2023년 기준 영업손실 158억 원, 당기순손실 136억 원을 기록하며 수익성 확보에 어려움을 겪고 있다¹⁷⁶⁾. 앞선 리밸리온의 사례와 같이, 국내 AI 반도체 팹리스 스타트업들은 가시적 성과 부족으로 인해 연계 투자 유치에 큰 어려움을 겪고 있다. 이는 투자자들의 불안감을 초래하고, 장기적인 성장에 치명적인 장애물로 작용한다.

국내 AI 반도체 스타트업들은 대부분 초기 연구개발 비용을 정부 과제를 통해 충당하고 있지만, 정부 과제 예산은 첨단 반도체 제작비용에 비하면 작은 수준¹⁷⁷⁾이다. 중기부의 딥테크 TIPS 사업의 경우도 최대 지원 금액은 17억 원¹⁷⁸⁾ 수준으로, 몇 백억의 소요가 필요한 반도체 설계 및 칩 생산의 금액에는 비교하여 작은 금액임을 알 수 있다. 정부는 R&D 연구와 같은 기술 개발 사업에도 '차세대 지능형 반도체 사업단' 사업과 같은 국책 사업을 진행하고 있으며, 'K클라우드' 사업과 같은 실증 사업과 온 디바이스 AI, edge 부문 사업도 진행하고 있으나, 연계 생태계 구축 측면에서 더 큰 금액의 정책 연구 사업이 필요하다. 국내 주요 AI 반도체 기업들은 높은 R&D 비용으로 인한 수익성 부족 문제에 직면해 있다. 예를 들어, 23년 기준 퍼리오사AI의 영업손실은 601억 원, 리밸리온은 159억 원, 딥에스는 91억 원에 달한다¹⁷⁹⁾.

시장 점유율 확보 측면에서도, 서버향 학습 목적의 반도체 같은 경우 NVIDIA가 AI 반도체 시장에서 70%~95%의 점유율을 차지¹⁸⁰⁾하고 있으며, 이는 국내 스타트업들이 기술적 격차와 자본력 부족으로 인해 경쟁력을 갖추기 어렵게 만드는 부분이다. 더불어, 국내 AI 반도체 스타트업들은 특정 용도의 AI 반도체를 기반으로 매출을 확대하고, 이후 범용 AI 반도체 시장을 공략하는 전략을 취하고 있다¹⁸¹⁾. 하지만, 이는 독과점 기업이 존재하는 시장 하에서

176) 뉴스핌(24.04.18.), '한국판 NVIDIA' 리밸리온, 수익성 개선 방안은?, <https://www.newspim.com/news/view/20240417001079> (검색일 24.12.19.)

177) 조선비즈(24.11.04.), "돈 갈퀴로 굽어 모으던 'K-반도체 스타트업'에 서리 내렸다... 파두 때문에." https://bizchosun.com/stock/market_trend/2023/11/17/KJ6TYX553ZF7HHSSN2ELO36SIQ/. (검색일 24.12.19.)

178) hitnews(24.07.19.), "단독 | 팀스(TIPS) 중단 사태... '시험비용+인건비+지적재산권'까지 위태로운 벤처기업들", <http://www.hitnews.co.kr/news/articleView.html?idxno=56197> (검색일 24.12.19.)

179) 혁신의 숲 기업 검색 자료 이용, <https://www.innoforest.co.kr/company/CP00003249/%EB%94%A5%EC%97%91%EC%8A%A4> (검색일 24.12.19.)

180) CNBC(24.06.02.), "Nvidia dominates the AI chip market, but there's more competition than ever", <https://www.cnbc.com/2024/06/02/nvidia-dominates-the-ai-chip-market-but-theres-rising-competition-.html> (검색일 24.12.19.)

181) 전자신문(24.09.23.), "[창간기획] 출발점에 선 韩 AI 반도체, 생태계 구축 시급.",

<https://www.etnews.com/20240906000243?SNS=00002> (검색일 24.12.19.)

'매출을 확대하고 이를 기반으로 상품 다양한 전략을 통한 매출 증대'를 할 수가 없을뿐더러, 추격자 입장에서는 범용으로 전환이후에 매출확대가 더욱 어려워지는 문제점이 발생할 소지가 있다.

나. 민간 투자 전문성의 부족

AI 반도체 기술은 고도의 전문지식과 복잡성을 필요로 하며, 이는 정확한 기술 평가를 어렵게 만든다. 빠른 기술 혁신 속도, 직접도의 향상과 선단공정 활용 증가¹⁸²⁾ 등 기술과 산업의 생태계 진화 속도(GPT3-GPT3.5-GPT4 등 급격한 모델의 발전에 요구되는 칩 성능 확대 요구, 이로 인한 짧은 제품 수명 주기)가 다른 산업에 비하여 빠른 축에 속한다. 또한 선단 공정, 해외IP, EDA 등 매우 높은 초기 비용¹⁸³⁾이 소요되며, 이와 같은 기술의 도입과 비용의 증가는 기술에 대한 이해가 부족한 투자자에게 큰 도전 과제가 될 수 있다. 예를 들어, 전체적인 프레임워크와 아키텍처를 이해하지 못하면 AI 모델의 성능을 최적화하는 데 필요한 하드웨어 설계 및 제조 과정을 평가하는 것이 어렵다¹⁸⁴⁾. 따라서 AI반도체 기술에 대한 이해 부족은 동 산업의 동학(Dynamics)에 대한 이해를 어렵게 만들고, 추가적인 투자 자체가 어려워지게 만드는 부작용을 낳고 있다.

AI 반도체 분야는 시스템 반도체 부문과 비슷하지만 AI산업의 특별적 성장과 맞물려 산업별 특화 ASIC 증가, FPGA 의 도입, 선단 공정 사용 유무 등, 이 분야에 대한 전문성을 보유한 VC(벤처 캐피탈)와 PE(프라이빗 에쿼티) 투자자가 부족한 문제를 발생 시킨다. 이는 기술에 대한 이해도 부족으로 이어져 평가 과정에서 정보 비대칭이 발생할 가능성을 높인다. 예를 들어, AI 반도체는 기존 반도체 기업들의 사업 포트폴리오 확장 또는 자체 칩 개발을 통해 생존, 성장이 이루어지기도 하며, 신규 팝리스 스타트업 진출의 확대 또한 AI반도체 산업 생태계 발전의 원천¹⁸⁵⁾이다. 그러나 이와 같은 새로운 기술의 상용화는 다소 시간이 필요하며,

182) W브릿지(24.05.29.), “[전문가칼럼] AI 기술과 최신 반도체 기술의 융합”, https://www.wbridge.or.kr/platform/careersport/info/selectTrendDetail.do?ntt_sn=640. (검색일 24.12.19.)

183) 전자신문(22.03.20.), “중소 팝리스, AI반도체 7200억 ‘그림의 떡’” <https://n.news.naver.com/mnews/article/030/0003006897?sid=001> (검색일 24.12.19.)

184) 176) 상동

185) 장은현 (2024), “AI 반도체 기술 및 산업 동향”, KDB산업은행 미래전략연구소(2024.7, 제824호 3), <https://rd.kdbi.go.kr/>

장기적으로 매출을 통해 시장 내 수요자들의 반응도 분석해야하는 부분이 존재한다. 이러한 환경에서 전문성을 갖춘 투자가 부족한 것은 투자를 통한 생태계 주요 혁신 주체인 기업의 성장에 장애물이 될 수 있다¹⁸⁶⁾.

기술특례상장은 기업의 기술력만을 평가하여 상장을 허용하는 제도로, 실질적인 수익성과는 무관하게 기업 가치를 인위적으로 상승시키는 경우도 존재한다¹⁸⁷⁾. 예를 들어, 2023년 기술특례상장으로 상장된 기업들의 약 97%가 매출 추정치가 미달인 상황¹⁸⁸⁾이다. 이는 기업이 상장 시 예상했던 미래 실적과 실제 실적 사이의 괴리에 기인한다. 이에 따라 투자자들은 실제 기업 가치보다 부풀려진 가치를 기준으로 투자하게 되며, 결과적으로 큰 손실을 입을 위험이 있다. 이 역시, 특례상장의 좋은 의도와 다르게 기업의 IPO를 위한 무리한 밸류에이션에 기인한 경향도 존재할 수 있음을 고려할 필요가 있다. 더불어, 높은 밸류에이션은 투자 리스크를 증가시키는 주요 요인 중 하나이다. 밸류에이션이 높게 책정될수록 투자자는 예상 수익이 그만큼 높아야 합리화 될 수 있으며, 그렇지 않을 경우 큰 손실을 입을 수 있다¹⁸⁹⁾. 예를 들어, 리밸리온의 경우 사피온과의 합병을 통해 기업 가치를 2조-3조 원 수준에서 최대 4조원까지 바라보고 있으나, 23년 사피온의 매출은 56억 원에 영업손실 259억 원, 동기간 리밸리온은 매출 27억 원에서 영업손실 159억 원을 기록¹⁹⁰⁾하였다. 합병¹⁹¹⁾ 이후의 국내 AI 반도체 1위 기업의 탄생 및 기술력과 인재 부문의 시너지는 긍정적인 부분으로 평가

kdb.co.kr/fileView?groupId=E49EC0E6-F243-5B37-9586-D097FD724BD2&fileId=59D62619-83C0-01EC-F6FC-83BCAAC00856. (검색일 24.12.19.)

186) ZUBRCapital(24.03.07.), "The AI Boom: Investment Opportunities and Predictions for 2024", <https://zubrcapital.com/news/the-ai-boom-investment-opportunities-and-predictions-for-2024> (검색일 24.12.19.)

187) KolonPharma(23.05.25.), "[김선진의 바이오 인사이트 15] 코스닥 시장 특례 상장 가이드라인의 한계", https://kolonpharm.co.kr/bbs/board.php?bo_table=sub042_1&wr_id=35

188) 서울경제(24.05.17.), "작년 상장 기술특례기업 97%가 매출 추정치 미달…'IR 의무화로 정보 접근성 높여야'[시그널]", <https://www.sedaily.com/NewsView/2D982R7E2O> (검색일 24.12.19.)

189) sfccapital(23.03.02.), "The high valuation trap", <https://sfccapital.com/blog/high-valuation-trap> (검색일 24.12.19.)

190) Topdaily(24.08.19.), "'사피온+리밸리온' 합병, 몸값 4조 노리나", <https://www.topdaily.kr/articles/98735> (검색일 24.12.19.)

191) 비즈조선(24.07.25.), "[단독] 리밸리온 몸값, 10조 써낸 주관사 후보도 있다... 사피온 합병은 2.5대 1 이하 예상", https://biz.chosun.com/stock/market_trend/2024/07/25/R66CJRXYISBCL7BOCS45AH724IU/ (검색일 24.12.19.)

될 수 있으나, 기존의 영업 손실과 매출 부문의 흑자 전환과 같은 이슈와 우려는 여전히 존재하고 있다.

AI반도체 팹리스 스타트업 투자 이후, 장기적 관점에서의 수익 리턴 구조가 미흡한 점은 투자자들의 투자를 주저하게 만들 수 있는 원인이 된다. IPO, M&A 이후의 거래가 아닌 투자 단계에서의 지분 이동 등의 이슈는 투자 대상의 기업의 안정적 성장에 대한 의문을 제기할 수 있는 통로가 될 수 있다. 클럽딜은 다수의 투자자가 협력하여 특정 프로젝트에 투자하는 방식으로, 초기 투자 단계에서는 리스크를 나눈다는 측면에서 유용할 수 있다. 하지만, IPO나 M&A 이후의 거래가 아닌 투자 단계에서의 지분 이동이 빈번하게 이루어질 경우, 장기적인 수익 리턴 구조는 미흡할 수 있다. 예를 들어, SKT가 AI 반도체 스타트업 리밸리온과 사피온코리아의 합병을 발표했으나, 이 과정에서의 지분 이동이 투자자들에게 안정적인 수익을 제공하지 못할 수 있다¹⁹²⁾. 이는 기존 투자자(VC, PE)들이 이 후속 투자를 계속할 수 없는 환경을 만들며, 장기적 관점에서 투자 리스크가 증가할 수 있다. 단계별 투자를 받으며 성장하는 기업 가치에 따라 밸류에이션이 증가하며, 투자자 증가 및 자금 조달에 따른 창업자의 지분율은 하락이 될 수 있다. 이에 따른 기업의 적대적 M&A 방어수단으로 ‘신주인수선택권’을 사용하는 포이즌 필(Poison Pill)¹⁹³⁾ 제도의 활용도 IPO에서 주요한 이슈로 떠오르고 있다. AI반도체 팹리스 스타트업의 경우에는 앞서 언급한 바와 같이 막대한 투자가 필요하며, 성장의 단계별로 투입이 필요한 자금 조달에 따라 창업자의 지분이 계속하여 하락되는 경우, 기 계획되어 진행되는 Business Model이 아닌 형태로 기업의 사업방향과 IPO에 영향을 미칠 수 있음을 고려해야 한다.

다. 전략적 투자의 부재

국내 M&A 시장은 2010년부터 2021년까지 꾸준한 성장세를 보이며 2021년에는 거래 규모 123.4조 원으로 정점을 찍었지만, 2023년에는 거래 건수 및 거래 금액의 급감으로 인해 시장이 위축¹⁹⁴⁾되었다. 이는 금리인상, 인플레이션 상승, 정부정책의

192) Numbers(24.08.19.), “리밸리온·사피온, 합병 후 ‘기업가치’ 제고...실적 개선 과제”, <https://www.numbers.co.kr/news/articleView.html?idno=5932> (검색일 24.12.19.)

193) KDI경제정보센터(09.11.27.), “적대적 M&A 방어장치 ‘신주인수선택권(포이즌필)’”, https://eiec.kdi.re.kr/material/clickView.do?click_yymm=201512&cidx=1119 (검색일 24.12.19.)

불확실성 및 러시아의 우크라이나 침공 등이 주요 원인으로 볼 수 있다¹⁹⁴⁾. 기업의 기술 부문으로 초점을 맞추어 보면 아직 칩의 검증이 끝나지 않은 점(양산 이전), 연계된 매출에서의 증명(양산 이후) 등 이유로 국내 대기업들은 검증되지 않은 AI 반도체 스타트업 인수에 신중한 입장을 보이고 있다. 국외에서는 소프트뱅크가 영국의 Graphcore의 인수 등 다방면으로 AI반도체 기업을 인수하고 있으나, 김정호 (2023)의 연구에 따르면 그래프코어가 가진 가속기 부문의 내외부적 역량은 충분히 앞선 조건의 수준을 넘는 것으로 알 수 있다. 인텔의 하바나 랙스 인수와 같은 경우에도 기업 가치가 2조 원이 상회하더라도, 기업의 성장과 수익창출에 긍정적인 요소로 작용하면 충분히 인수가 가능한 측면이 있으나, 현재 국내 AI반도체 기업의 높은 밸류에이션과 이에 따른 성장의 가능성에 대한 부문은 정확히 예단할 수 없는 환경에 있다.

[그림 3-10] 주요 국내 AI반도체 팝리스 유치 실적



업체명	리밸리온	퓨리오사AI	사피온
	rebellions_	FURIOSA	SAPEON
대표	박성현	백준호	류수정
누적투자	2800억원 (시리즈 B)	1700억원 (시리즈 C)	600억원 (시리즈 A)
기업가치	8700억원	6800억원	5000억원
주요 레퍼런스	KT클라우드 (K-클라우드, 초거대AI '믿음')	<ul style="list-style-type: none"> • 네이버 클라우드 (K-클라우드) • 카카오엔터프라이즈 (카카오클라우드) 	NHN 클라우드 (K-클라우드)

그래픽: 윤선정 디자인기자

출처: 유니콘팩토리(24.01.31.), “‘실력’ 입증한 K-팝리스에 수천억 뭉칫돈…다음 과제는 ‘실적’”, <https://www.unicorfactory.co.kr/article/2024013013150735004> (검색일 24.12.28.)

194) PwC(24.02.01), “PricewaterhouseCoopers. “2024 Outlook – Silver Lining.”

https://www.pwc.com/kr/ko/insights/samil-insight/deals_trends2024.html. (검색일 24.12.19)

195) 김앤장(23.03.09), “Korea M&A – 2023 Outlook”, https://www.kimchang.com/en/insights/detail.kc?idx=27288&sch_section=4 (검색일 24.12.19)

높은 기업 가치 대비 실적 부족으로 인한 리스크 (검증된 실적 부족, 투자 근거 확보 필요) 측면에서, AI 반도체 스타트업들은 종종 높은 기업 가치를 인정받으나, 이들 기업 대부분이 아직 실적이 충분히 검증되지 않았다는 점은 전략적 투자를 방해하는 요소로 작용한다. 리밸리온과 사피온은 각각 137억 원과 180억 원의 순손실을 기록했으며, 아직 수익을 실현하지 못하고 있다¹⁹⁶⁾. 이는 투자자들로 하여금 투자 근거를 확보하는 데 어려움을 겪게 만든다.

AI 반도체 분야에서의 IPO 성공사례는 상대적으로 적다. 퓨리오사AI와 같은 일부 기업이 성공적으로 펀딩을 유치하고 IPO를 준비 중이지만, 장기적인 ROI에 대한 불확실성은 여전히 존재¹⁹⁷⁾한다. 이들 기업들이 실제로 지속 가능하고 수익성 있는 비즈니스 모델을 통해 성공적인 롤아웃을 할 수 있을지는 아직 미지수이며, 이는 양산 예정 칩들이 시장에서 충분한 매출을 일으켜야하는 점과 연결되어 있다.

라. 공공투자 부문의 한계 (정부 정책으로 생성된 펀드)

정부 지원 펀드의 구조적 한계는 다양한 문제점으로 나타난다. 우선 정부는 AI 반도체 산업에 8.8조 원 규모의 지원을 약속했으나, 이 중 절반 이상이 대출 형태로, 실질적 직접 지원은 제한적¹⁹⁸⁾이며, 대규모 자금이 필요한 AI 반도체 산업의 특성상, 광범위한 지원보다는 선별적 집중을 통한 PoC, MPW, Single Run등의 단계적 지원이 필요하나 현재 정책은 이를 충분히 반영하지 못하고 있다¹⁹⁹⁾.

더불어 지원하는 예산 규모가 AI반도체 설계와 생산을 지원하기에 상대적으로 적은 규모인 점도 문제점이다. 정부는 AI 반도체 산업 육성을 위해 2029년까지 총 1.02조 원 규모의 R&D 투자를 계획했으나, 이는 글로벌 경쟁사들과 비교했을 때 매우 부족

196) Invest Chosun(24.06.18.) “사피온 합병에 리밸리온 IPO ‘올 스톰’...연내 재개도 안갯 속.” https://www.investchosun.com/site/data/html_dir/2024/06/17/2024061780162.html. (검색일 24.12.28.)

197) 이뉴스투데이(24.01.09.), “올해 핵심 테마는 ‘AI반도체’⋯⋯주요 스타트업 IPO ‘시동.’” <http://www.enewstoday.co.kr/news/articleView.html?idxno=2077161>. (검색일 24.12.28.)

198) 매경(24.10.16.), “정부, 반도체에 8.8조 지원한다지만 절반은 저리대출⋯⋯ 직접 보조금 ‘0’”, <https://www.mk.co.kr/en/business/11142089> (검색일 24.12.28.)

199) Communications of the ACM(23.07.01.), “South Korea’s Nationwide Effort for AI Semiconductor Industry, East Asia and Oceania Region Special Section: Big Trends” <https://cacm.acm.org/research/south-koreas-nationwide-effort-for-ai-semiconductor-industry/> (검색일 24.12.28.)

한 수준²⁰⁰⁾이다. 더불어, NVIDIA의 경우 2024년 상반기에만 R&D 비용으로 7조 8000억 원을 지출한 반면, 국내 주요 AI 반도체 기업들의 R&D 투자액은 이의 1% 미만 수준에 그치고 있는 점도 기업의 자체 R&D비용에 정부 지원 R&D금액을 합한 것을 감안하더라도, 작은 규모임을 알 수 있다²⁰¹⁾.

정책 자금 지원의 비효율성을 고려하면, 현재 정부의 연구개발 지원이 과기부, 산업부, 중기부 등으로 분산되어 있어 체계적인 단계별 지원이 어려운 상황이며, 기술력은 있으나 업력이 짧은 신규 스타트업들이 지원받기 어려운 구조적 문제²⁰²⁾ 존재한다. 해외 IP와 EDA 사용으로 인해 반도체 하나 개발에 수백억 원의 비용이 소요되나, 정부 지원금은 이를 감당하기에 턱없이 부족하며, 장기적으로 보더라도 한계점이 명확한 지원 구조이다.

AI반도체와 연계된 다양한 기술군의 개발에 지원하는 것은 필수적 요소이나, 서버용 학습목적의 AI반도체는 NVIDIA가 글로벌 점유율의 95% 이상을 상회하는 환경에서, 추론 최적화 반도체 및 옛지 AI 반도체 지원 강화, 실증(POC) 중심의 산학연(기업-학계-연구소) 협력을 강화하여 실증을 최우선적으로 진행할 수 있는 정부 사업을 발전시켜야 한다. 민간 투자 연계 미흡의 측면에서 정부 지원 이후 민간 투자로의 연결이 원활하지 않아 지속가능한 생태계 구축에 한계가 있다. 이는 아직 각 기업들의 양산 제품이 시장에서 검증을 받는 과정이 없는 부분에도 기인할 수 있으나, 메가 연계투자가 부족한 AI반도체 생태계의 특성에도 기인한다. 더불어, 장기적인 관점에서 칩 설계의 주요 요소로서 반도체 설계 SW를 고려하여 EDA, IP등 팝리스의 경쟁력을 강화하는 방향으로 국가적 지원의 확대²⁰³⁾도 고려할 필요가 있다.

200) 과기정통부(22.06.28.), “인공지능 반도체 산업 성장 지원대책 발표” [https://www.msit.go.kr/eng/bbs/view.
do?bbsSeqNo=42&mId=4&mPId=2&nttSeqNo=702&sCode=eng&searchOpt=ALL](https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPId=2&nttSeqNo=702&sCode=eng&searchOpt=ALL) (검색일 24.12.28.)

201) 전자신문(24.09.24.), “[창간기획] 출발점에 선 韓 AI 반도체, 생태계 구축 시급”, <https://www.etnews.com/20240906000243> (검색일 24.12.28.)

202) 시사저널e(24.08.05.), “‘韓팹리스 반도체, 지원 부족으로 성장 한계’” <https://www.sisajournal-e.com/news/articleView.html?idxno=404763> (검색일 24.12.28.)

203) 한경(24.07.01.), “[단독] ‘이 장관 아니면 몰랐을 것’…반도체 EDA’ 각성한 정부 [강경주의 IT카페]”, <https://www.hankyung.com/article/202407012364i> (검색일 24.12.28.)

3. 인재

AI반도체를 포함한 반도체 산업의 성장과 함께 관련 산업에서 필요로 하는 고급 인력의 공급 부족이 심화되고 있다. 이에 정부는 2022년 7월 “반도체 초강대국 달성전략”에서 2031년까지 반도체 관련 인력 15만명을 양성하겠다고 약속하고, 이에 빌맞추어 산업계도 반도체 인력 양성을 위한 ‘반도체 아카데미’ 설립을 발표²⁰⁴⁾하였다. 또한, 2024년 4월에는 AI-반도체 이니셔티브(안)을 발표하며, AI반도체 분야 인재 양성을 위해 도전적 연구를 지원하고, 관련 대학원 운영 및 R&D 지원 확대, 실무·글로벌 인재 양성을 추진하고 있다. 그럼에도 불구하고 전문가 인터뷰를 통해 확인한 결과 AI반도체 업계는 고급 설계인력의 공급 부족 해소와 기존 반도체 설계 교육 훈련 양성시스템의 고도화의 필요성이 제기되고 있음을 확인하였다.

이러한 인력 부족 이슈는 글로벌 차원의 고급인력 확보를 위한 경쟁, 국내 차원의 고급인력 확보를 위한 경쟁이 동시에 일어나고 있는 등 세부적으로는 단순한 수요-공급 불균형보다 훨씬 복잡한 상황이다. 여기서는 고급 설계인력의 공급 부족 문제와 반도체 설계 교육 훈련 양성시스템 문제, 크게 두 가지 측면을 중심으로 살펴보기로 한다.

가. 고급 설계인력의 공급 부족

AI반도체 설계 및 개발(팹리스)에는 석사급 이상의 고급설계 인력을 중심으로 필요로 하고 있으나, 국내 반도체 산업이 메모리 중심으로 구성되어 기존 팹리스/파운드리 센터가 취약한 특성상 AI반도체 설계 및 개발을 위한 경험 있는 인재가 전반적으로 부족한 것으로 알려져 있다. 또한, 이공계 기피현상과 의대 쏠림현상으로 인해 대학에서부터 반도체 관련 학과 지원이 감소하고 있는 추세이다. 또한, 반도체 산업 내에서는 대기업(삼성전자, SK하이닉스) 취업에 대한 선호도가 높으며, 그 안에서도 기반이 탄탄한 메모리

204) 반도체 업계와 산업부는 반도체 산학협력 4대 인프라로서 (1) 기업주도 전문교육기관인 반도체 아카데미 설립, (2) 정부-기업 공동 투자를 통해 R&D 및 인력양성을 지원하는 한국형 Semiconductor Research Corporation (SRC, 미국의 민관 반도체 연구 컨소시엄) 운영, (3) 현장형 교육연구를 위한 한국형 Interuniversity Microelectronics Centre (IMEC, 벨기에 소재 세계 최대의 반도체 연구기관) 운영, (4) 대학-기업 취업 연계를 위한 소부장 계약학과 신설 등에 대해 협력키로 하고 MOU를 체결(관계부처 협동(2022.7.21.), 「반도체 초강대국 달성전략」)

반도체 분야에 대한 선호가 높는 실정이다. 특히, 글로벌 선진기업으로의 AI반도체 설계 관련 인재의 해외 유출이 심화되고 있는 반면, AI반도체 설계 관련 해외 인재에 대한 국내로의 유치방안은 미흡한 실정이다.

1) 설계 및 개발(팹리스)에 필요한 석사급 이상의 고급설계 인력 요구

AI반도체의 경우 반도체 설계 경험을 보유한 석사급 이상의 고급설계 인력이 많이 필요로 하는 영역으로서 AI반도체 스타트업에서 개발을 위한 충분한 인력 확보에 어려움을 겪고 있다.

국내 반도체 설계인력의 부족은 기본적으로 기존 팹리스/파운드리 센터 취약으로 인한 반도체 설계 경험이 있는 인재가 부족한 것에 기인한다고 볼 수 있다. 국내 시스템반도체 세계시장 점유율은 고작 3% 수준에 머물러 있다. 또한, 파운드리 분야의 경우에도 삼성이 힘을 쓰지 못하고 있는 상황에서, 대만의 TSMC가 글로벌 선도기업 대부분의 반도체 생산 물량을 확보하며 전세계 시장을 주도하고 있는 상황이다. 국내 팹리스 반도체설계 회사의 경우 글로벌 상위 기업 대비 매출액도 2천억 내외 정도로 양적, 질적 측면에서 큰 차이를 보이고 있다²⁰⁵⁾. 따라서, AI반도체 관련 업계에서 요구되는 경험이 풍부한 인적 자원이 기본적으로 부족한 상황이다.

이러한 인력 부족 상황을 타개하기 위해 정부 차원에서 관련 인력 양성을 위한 산학 협력 정책을 지원하고 있으며, 그 핵심은 반도체 계약학과이다.

국내 반도체 기업의 지원에 기반한 반도체 계약학과²⁰⁶⁾는 전국적으로 10개 대학에서 500명 이상 규모로 운영되고 있다. 삼성전자의 경우 2006년 처음으로 성균관대에서 운영을 시작하였고, 이후 2021년 연세대, 2022년 KAIST, 2023년 포항공대에서 계약학과를 설치하였다. 또한, 2024년부터는 UNIST, DGIST, GIST 등 지역의 과학기술원과 반도체 계약학과를 신설하였다. SK하이닉스의 경우 삼성전자보다는 다소 늦은 2020년 고려대학교에 처음 반도체공학과를 개설하였고, 한양대, 서강대, 중앙대

205) 전문가 인터뷰 내용 중 발췌

206) 기존 정원내에서 모집하는 일반학과와 달리 계약학과는 산학협력법(산업교육진흥 및 산학연협력촉진에 관한 법률)에 근거해 별도로 운영되며, 기업과 대학간 사업 계약을 바탕으로 기존의 정원내 인원과 별도로 선발됨

에 계약학과를 신설하여 운영하고 있다²⁰⁷⁾. 2024년 기준 국내 반도체 계약학과의 현황을 정리하면 <표 3-4>와 같다.

<표 3-4> 국내 반도체 계약학과 현황

구분	운영 대학(인원)	총 인원('24년 기준)
삼성전자	KAIST(100명), 포스텍(40명), GIST(30명), DGIST(30명), UNIST(40명), 연세대(100명), 성균관대(70명)	410명
SK하이닉스	고려대(30명), 서강대(30명), 한양대(40명)	100명

자료: 베리타스 알파(2023.3.27.), “2024 반도체 계약학과 10개교 510명 체제”, <https://www.veritas-a.com/news/articleView.html?idxno=451192>, (검색일: 2024.10.23.) 내용 기반으로 연구진 정리

그러나, AI반도체 인재 양성 방안이 제대로 작동하지 않는 이유는 국내적으로는 크게 이공계 기피 및 의대로의 쏠림 현상으로 인해 반도체 관련 학과에 대한 등록율이 저조한 이슈다(ZDNET, 2024.8.5.). 또한, 반도체 산업 내에서도 스타트업 대비 대기업 쏠림 현상, 메모리반도체 쏠림 현상 등으로 설명된다.

2) 산업내 인력확보 경쟁 심화

반도체 산업에 대한 전략적 중요성이 높아지고, 특히 AI의 기술적, 시장적 성장세가 높아지면서 기존 반도체 업계와 AI반도체 설계 관련 인력확보에 대한 경쟁도 심화되고 있다. 국내 반도체 관련 업계 내에서는 학계와 산업체의 경쟁, 산업체 내에서는 대기업과 스타트업 간의 경쟁으로 구분할 수 있다. 연봉 수준 등을 고려할 때 먼저 학계 대비 대기업 선호로 인해 우수 인력들이 삼성전자, SK하이닉스 대기업 쪽을 선호하고 있다. 이로 인해 학계에서 반도체 설계, 특히 디지털 쪽 경험이 풍부한 인력이 부족하며, 인력 양성의 선순환 구조가 이루어지기 힘들어 인력양성에 어려움이 존재한다²⁰⁸⁾.

국외적으로는 우리나라 설계 인재에 대한 글로벌 기업과의 경쟁, 베트남, 인도 등 개발도상국 인재에 대한 글로벌 기업과의 경쟁으로 구분할 수 있다. 먼저 AI시대 GPU 시장을 거의 독점하고 있는 NVIDIA로 삼성전자 및 SK 하이닉스에서 수백 명의

207) TECHWORLD (2023.8.10.) 글로벌 반도체 인재 확보전 본격화… 뺏거나 혹은 뺏기거나, <https://www.epnc.co.kr/news/articleView.html?idxno=235314>, (검색일: 2024.10.22.)

208) 전문가 인터뷰 내용 중 발췌

인재가 빠져나간 것으로 알려져 있다²⁰⁹⁾. 이러한 경향은 앞으로도 지속될 것으로 국내 AI반도체 인력 부족 상황의 악화로 이어질 것이다.

반대로, AI반도체 설계 관련 해외 인재 유치 상황은 그리 밝지 못하다. <표 3-5>에서 보는 바와 같이 주요 선진국을 중심으로 첨단 산업 관련 외국인 대상 비자 프로그램이 다양하게 운영되고 있다. 국내의 경우 산업 및 과학 관련 기술을 보유한 외국인에게 발급되는 E-3 비자, E-7-S 비자 등의 비자 프로그램이 있다. E-3 비자의 경우 석사학위 혹은 박사학위 취득 후 3년 이상의 경력을 지닌 외국인에게 발급되며, 이들은 최대 5년 동안 한국에 머무를 수 있다. 또한, E-7-S 비자는 반도체 포함한 첨단산업 전문 인력에 글로벌 톱 500 대학 출신에 비자 심사 시 가산점을 주는 비자이다. 결론적으로 주요국 대비 취업비자의 실효성이 떨어진다고 볼 수 있다.

<표 3-5> 주요국별 첨단 산업 관련 외국인 대상 비자 프로그램 현황

국가	비자 특성
영국	• 글로벌 톱 50 대학 출신에 2~3년간 거주하며 반도체 포함 첨단 산업 구직활동이 가능한 High Potential Individual (HPI) 비자 신설
독일	• 반도체 등 고급 기술 보유 인재에 대해 최대 1년간 독일에서 구직 기회 제공
대만	• 글로벌 톱 500 대학 학생에 대만 반도체 기업 면접 통과 시 비자 발급
일본	• 글로벌 톱 100 대학 출신에 반도체 포함 첨단 산업 분야에서 2년간 구직활동 가능한 비자 발급 • 석박사학위 소지자에 허용된 연구원(E-3) 비자의 경우 국외 석사학위 소지자에 대해 2024년 7월부터 세계 우수대학 졸업자 또는 우수 학술논문 저자의 경우 경력 요건을 삭제
한국	• 반도체 포함한 첨단산업 전문 인력에 글로벌 톱 500 대학 출신에 비자 심사 시 가산점을 주는 비자 E-7-S 신설

자료: 법무부(2024.7.1.), 과학기술분야 우수 글로벌 인재 유치 위해연구유학생(D-2-5), 연구원(E-3) 비자 확대, 법무부 보도자료, 조선일보(2023.4.7.), 대만·일본, 심각한 인력난에 파격적 '외국인 모시기', https://www.chosun.com/economy/economy_general/2023/04/07/M3MXDDIWMFFT5MUPVLUFUDJTTE/, (검색일: 2024.10.22.), TECHWORLD(2023.8.10.) 글로벌 반도체 인재 확보전 본격화… 뺏거나 혹은 뺏기거나, <https://www.epnc.co.kr/news/articleView.html?idxno=235314>, (검색일: 2024.10.22.).

이에 따라 반도체 설계 및 패키징 관련 일부 국내 시스템반도체 업체²¹⁰⁾의 경우 베트남 현지에 R&D센터를 설립 후 현지 설계 인력을 확보·활용하고 있다²¹¹⁾. 이미

209) 조선일보 (2024.6.19.), AI 인재 1명 잡으려... MS, 회사 통째로 샀다, https://www.chosun.com/economy/tech_it/2024/06/19/DTG6YXZBMFBVJCLBK4QS6LUGLQ/, (검색일: 2024.10.25.)

210) 에이디테크놀로지(설계, 2018년), 코아시아세미(설계, 2020년), 세미파이브(설계, 2022년), 보스반도체(디자인센터, 2023년), 하나마이크론(패키징, 2022년)

미국, 일본 대만, 중국 등 글로벌 반도체기업 생태계가 베트남에 구축되어 있으며, 현지 인력을 활발히 활용하고 있다.

3) 고급 설계인력에 대한 수요 확대

AI반도체 산업의 경우 시장의 성장에 따른 인력수요가 학계, 출연연, 기업 등을 포함하여 지속적으로 확대되고 있다. AI반도체에 퓨리오사AI, 딥엑스 등과 같은 스타트업 외에 KT가 주요 주주사로 포함된 리밸리온, SK그룹의 지원을 받았던 사피온과 같은 스타트업, 그리고 현대자동차 차량용 반도체 개발 등 기존 대기업에서 신규시장의 진입까지 이루어지면서 반도체 설계인력 확보 경쟁이 심화되고 있는 것이다.

나. 반도체 설계 교육 훈련 양성시스템 고도화 필요

반도체 설계 관련 고급 인력의 부족 상황은 비단 국내의 이슈만은 아니다. 미국 또한 2030년까지 6.7만 명이 부족할 것으로 예상²¹²⁾되고 있으며, 일본도 2034년에 4만 명이 부족할 것으로 예상되고 있다. 글로벌 반도체 산업의 성장 및 각 국가별 첨단 전략산업으로 반도체 산업을 주목함에 따라 관련 인력의 부족이 전세계적으로 예상되고 있으며, 국가간 고급 인력 확보를 위한 경쟁 또한 치열해지고 있는 상황이다.

비록 관련 부처별로 다양한 반도체 설계 교육 훈련 프로그램을 운영중에 있음에도 불구하고, 체계적이지 못한 단기적 실무교육 중심으로 프로그램이 운영되고 실무에서 요구하는 고급인재 육성 프로그램이 부족한 상황이다. 또한, 실무교육이 가능한 산업체 출신의 교원 확충이 필요하며, 해외 유학생에 대한 국내로의 유치 및 국내에서 학위 후 취업으로 연계는 실효성있게 진행되지 못하고 있다. 마지막으로 인재육성 정책을 통합 관리하는 컨트롤 타워 또한 부재한 상황이다.

211) ZDNET Korea(2024.7.26.), “베트남, 韩반도체 설계 인력 공급망으로 급부상”, <https://zdnet.co.kr/view/?no=20240724170502> (검색일 24.11.28.)

212) 조선일보 (2024.6.19.), AI 인재 1명 잡으려... MS, 회사 통째로 샀다, https://www.chosun.com/economy/tech_it/2024/06/19/DTG6YXZBMFBVJCLBK4QS6LUGLQ/, (검색일: 2024.10.25.)

1) 교육 훈련 양성 프로그램의 한계

반도체 설계 교육과 관련하여 부처별로 다양한 교육 프로그램이 시행되고 있다. 그러나, 국내 시스템반도체 업계가 요구하는 핵심 인력은 석사급 이상의 고급 설계 인력이나, 체계적인 장기성 프로그램 부재로 인해 실제 반도체 업계에 투입 가능한 고급 인재의 육성이 제대로 이루어지지 못하고 있는 것으로 알려져 있다²¹³⁾. 따라서, 반도체 관련 학과의 확대에만 그치는 것이 아니라, 학과 내 교육 프로그램을 수요 기업과 연계를 통한 실무 교육 중심으로 바꾸고 교수 인력 또한 산업계 경험이 풍부한 실무 중심의 교수진을 함께 구성하여야 할 것으로 전문가들은 제안하고 있다. 즉, 연구 중심 교수진 외에 교육 중심 교수진의 확대가 필요하다. 이를 위해서는 대학교수 채용에 대한 정부의 유연한 대응이 요구된다. 또한, 학위 과정에서 칩 양산 경험 확보를 위해 MPW 제공을 위해 파운드리 업계와의 협업도 필요할 것이다. 이와 함께 졸업 후 취업으로 연계될 수 있는 체계를 함께 마련할 필요도 있다.

2) 인재육성 정책을 통합 관리하는 컨트롤타워 부재

반도체 인재 육성과 관련해서 과학기술정보통신부, 산업통상자원부, 교육부, 중소벤처기업부, 고용노동부 등 다양한 부처가 부처의 특성 및 역할에 따라 다양한 육성 사업을 <표 3-6>과 같이 추진하고 있다. 그러나, 부처별 주요 사업을 통합적으로 관리하고 효율적으로 수행 가능한 ‘컨트롤타워’가 부재한 것으로 알려져 있다²¹⁴⁾.

213) ZDNET Korea(2024.8.5.), '의대 쓸림' 반도체 설계 인재·교수가 없다...통합 컨트롤타워 필요, [이슈진단+] 시스템반도체 설계 인력·인재 부족 심각 (하). <https://zdnet.co.kr/view/?no=20240802165218>, (검색일: 2024.10.22.)

214) 상동

<표 3-6> AI반도체 관련 부처별 주요 인재양성 사업

구 분	AI반도체 고급인재양성 (과기정통부)	시스템반도체 융합전문 인력육성(과기정통부)	반도체 특성화대학원 (산업부)
지원내용	AI반도체 대학원 설치·운영 지원	시스템반도체 융합전문 인력양성센터 설치·운영 지원	반도체특성화대학원 설치·운영 지원
예산규모 ('23)	42.5억	95억	90억
지원대상 (단위)	석·박사 (대학원 단위)	석·박사 (대학, 연구기관 등 컨소시엄 단위)	석·박사 (대학원 단위)
지원분야	AI반도체 설계, AI SW 등 AI반도체 전문과정	AI·IoT·바이오·자동차·지능형 사물에너지 등 5개 융합과정	메모리, 비메모리, 패키징 등 8개 기술분야*
구 분	반도체특성화대학 (교육부)	반도체설계구현인재양성 (과기정통부)	스타트업 AI기술인력 양성 (중소벤처기업부)
지원내용	반도체 우수인재의 안정적 양성을 위한 반도체 교육 특화 대학 집중 육성	국내 중소 펩리스 인력부족 해소 위한 반도체 설계분야 특화인력 양성	혁신 벤처·스타트업이 필요로 하는 인공지능 실무인력 양성
예산규모 ('23)	480억	58억	27억
지원대상 (단위)	1~4학년 학부생 (대학 단위)	3~4학년 학부생, 재직자 (대학, 연구기관 등 컨소시엄 단위)	39세 이하 청년 구직자
지원분야	반도체 관련 전 분야(비메모리 반도체, 메모리반도체 포함)	시스템 반도체 설계 특화	인공지능 기술이 많이 접목되는 게임, 금융, 유통, 바이오 분야

출처: 과기정통부 (2022.11.5.), 부처별 반도체 인재양성 사업은 유사 중복이 아닙니다. 과기정통부 설명자료, 중소벤처기업부 (2023.1.26.), 이어드림 학교(스쿨) 보도자료 중심으로 연구진 정리

〈대만의 반도체 인력 부족 이슈 대응 현황〉²¹⁵⁾

대만은 인력 부족 측면이 여러 요소에서 한국과 유사한 면이 많은 국가이다. 먼저, 한국과 유사하게 저출산 고령화 등으로 산업 전반적인 인력난이 진행 중이다. 과거 중국의 고도 성장에 따른 중국 기업으로의 인력유출이 크게 있었으나, 미중 경쟁으로 대변되는 글로벌 패권경쟁 이후 중국과의 단절이 일부 이루어지고 있다.

대만의 인력 부족 문제는 국내적으로는 대기업, 특히 TSMC 쓸림 현상으로 요약된다. 반도체 파운드리로는 TSMC, 팹리스로는 Mediatek 등의 글로벌 기업이 대만 내 고급 엔지니어 인력을 독점하면서 인력 쓸림현상이 갈수록 심화되고 있다. 이로 인해 대만 내 스타트업 및 중소기업도 극심한 인력 부족에 시달리고 있다. TSMC 또한 관련 전공자가 부족하니 물리, 화학, 생물 등 전공을 가리지 않고 뽑아서 가르친다는 식으로 접근하고 있는데, 일단 실무 경험을 쌓은 후에는 높은 연봉 수준으로 쉽게 이탈하지는 않기 때문이다. 이로 인해, 기존 기계, 화학 등 대만 내 타 산업 분야 경쟁력 유지 이슈가 발생 중이며, 학문적으로도 반도체 산업을 제외한 분야에 박사과정 및 교수직 진출 꺼리는 경향이 심화되고 있는 실정이다. 추가적으로 학계 대비 TSMC, Mediatek 등의 기업 연봉이 상대적으로 높다 보니 박사학위 진학률이 상대적으로 낮아지는 상황도 발생하고 있다. 이러한 상황은 반도체 외 분야에서 더욱 악화되어 대학원 연구실의 유지도 갈수록 힘들어지고 있다고 알려진다.

대만 또한 우리나라와 마찬가지로 대만 정부 또한 해외 인력 활용에 관심이 많으며, 특히, 화교 네트워크가 긍정적으로 작용하고 있다. 말레이시아, 홍콩, 인도네시아 등은 화교(중국어 가능, 문화적으로 친숙)들이 많기 때문에 그들이 대만에서 취업하거나 창업 등에 기여하고 있다. 실제 학계, 산업체에서 동남아 국가에서 온 화교출신을 만나기는 어렵지 않다. 이를 뒷받침하기 위해 해외 우수인력을 위한 별도 비자제도를 신설(Gold Card)하였다. 유효기간이 3년으로 길며, 이직하더라도 귀국하지 않고 그대로 다시 구직이 가능한 장점이 있다. 가족 등 별도 비자가 불필요하고, 주택 지원과 함께 구직 전에도 입국이 가능하다.

4. 기술

현재 퓨리오사AI²¹⁶⁾, 리밸리온²¹⁷⁾ 등 국내 AI 반도체 기업들은 저마다 NVIDIA 제품 대비 자사 제품이 속도가 빠르거나 전력소모량이 적다는 식의 홍보를 계속하며 투자 유치 및 판로 확보를 위해 활동하고 있다. 그러나 현재로서는 투자금액에 비해 실제로 유의미한 매출을 올리고 있는 업체는 거의 없다시피 하다²¹⁸⁾. 기술력이 매출로 연결되지 않는 문제는 물론 앞서 기업 경쟁력 및 생태계 부문에서 지적했듯이 BM(Business Model)이나 사업화 역량 부족 등이 원인이지 기술력 부족 자체가 원인이 아닐 수도 있으며, 현재 대부분의 기업들이 창업 후 오랜 시간이 지나지 않은 스타트업인 만큼 시간이 더 필요할 뿐일 수도 있다. 그러나 제품이 실제로 타사 대비 기술적으로 우위에 있고 고객에게 더 큰 이익을 가져다줄 수 있다면 판로 개척 가능성이 높아지며, 현재 시장지배자인 NVIDIA의 제품 가격이 치솟고 그마저도 수요가 공급을 초과해 품귀현상이 빚어질 만큼²¹⁹⁾ 각종 AI 서비스 개발을 위한 반도체 수요가 높아지고 있는 것을 감안하면 이러한 매출 부진은 쉽게 이해되지 않는 면이 있다.

아래에서는 국내 AI 반도체 산업 생태계 전반에 가지고 있는 문제점으로서 기업들의 기술 측면을 간략히 분석해보고자 한다. 전문가들의 의견을 최대한 반영하고 참고 자료를 보강하고자 노력하였으나, 연구진이 AI 반도체 설계의 전문가가 아닌 만큼 일부 부정확하거나 전문가들 사이에 이견이 있을 수 있는 부분이 있는 점 먼저 양해 부탁드린다.

215) 대만 전문가 인터뷰 내용을 정리

216) 한국경제(2024.10.26.), “퓨리오사AI가 엔비디아와 경쟁할 수 있다는 이유 [각스]”

<https://www.hankyung.com/article/202410218895i> (검색일: 2024.11.6.)

217) Zdnet(2024.2.26.), “리밸리온 AI칩 ‘아톰’ 첫 시연 빈승 뜨거워..세계 무대 진출 신호탄”,

<https://zdnct.co.kr/view/?no=20240226081651> (검색일: 2024.11.6.)

218) 2023년 기준 퓨리오사AI 36.2억 원, 리밸리온 27.3억 원, 사피온코리아 24.8억 원, 딥엑스 0.5억 원, 모빌린트 4.4억 원, 하이파엑셀 0.2억 원 등. 출처: 혁신의숲(<https://www.innoforest.co.kr/>), 검색일: 2024.11.6.)

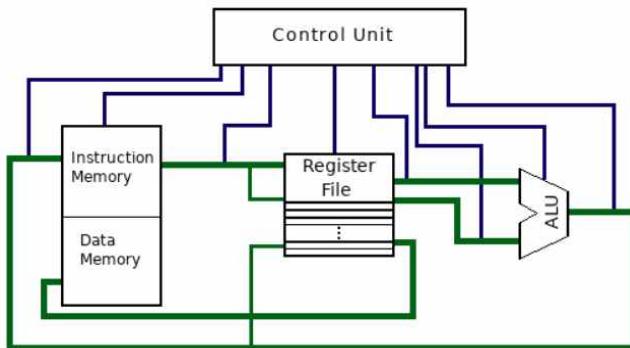
219) 전자신문(2024.4.25.), “덩치 커진 AI, GPU 부족 몸살”, <https://www.etnews.com/20240425000234> (검색일: 2024.11.6.)

가. 칩 설계 기술 격차

1) 아키텍처 설계 기술 격차

CPU, GPU 등 칩의 아키텍처(일반적으로 ‘마이크로아키텍처’라고 부름)가 무엇인가에 대해 아주 간략히 요약하면, 명령어(어떤 연산 처리를 수행하도록 정의됨) 뮤음(set)과 그것이 잘 처리되도록 하는 디지털 논리구조이다. 아키텍처는 아래 그림과 같이 명령어 메모리에서 레지스터(명령어 일시 저장 및 연산 지시)에 명령을 입력하고, 그 명령에 따라 메모리(데이터 저장)에서 데이터를 받아 ALU(실제 연산 수행 장치)로 이동시켜 실제 연산을 수행하고, 연산 결과를 다시 메모리에 저장하고, 다시 다음 연산을 수행하고… 등의 일련의 과정 처리를 구성하는 방식이라고도 볼 수 있다²²⁰⁾.

[그림 3-11] 아키텍처 논리흐름도



출처: St. Micheal(2019.2.7.)

AI 반도체 칩 아키텍처의 성능은 특정 연산에 대한 처리 속도, 처리 가능한 연산의 규모, 전력소모 등에 따라 결정된다. 그러나 AI 반도체 칩의 성능을 일률적으로 측정하는 것은 어려운 일이며, 이는 앞서 언급된 정보 비대칭성을 심화시키는 원인 중 하나다. 실제로 MLPerf 등 벤치마크가 존재하기는 하지만, 인터뷰에서 전문가들은 그 결과가 실제 업계에서 AI 반도체 칩이 활용되는 다양한 상황을 제대로 반영하고 있지

220) 아키텍처 부분의 설명은 St. Micheal(2019.2.7.)을 참조하였다.

는 못하다는 견해를 보였다.

AI 반도체에서 주로 다루는 연산 처리는 대량의 행렬 및 벡터에 대한 선형대수 연산으로 구성되어 있으며, 단순히 병렬화했다가 결과를 합치는 것으로 끝나는 것이 아니라 각각의 연산 결과가 다음 단계의 연산에 그대로 사용되는 등 연산 단계 간에 복잡한 데이터의 이동이 발생할 수 있다. 그리고 최근 AI 반도체 칩의 수요가 LLM 학습에 집중된 경향이 있기는 하지만 LLM 간에도 연산의 특성에 조금씩 차이가 있고, 그 외에도 이미지나 음성 학습 등 다양한 연산 종류가 존재한다. 또한 학습 외에 국내 AI 반도체 업체들의 타겟으로 여겨지는 추론 연산 역시 학습과는 다른 특성이 있다. 그리고 유사한 종류의 연산이라 하더라도 데이터 크기, 그 데이터의 단계별 이동, 필요한 처리 정밀도(이에 따라 32-bit, 16-bit, 8-bit, 4-bit 등 연산 단위의 크기, 정수, 부동소수점과 같은 데이터 종류 등이 달라질 수 있음, 제2장 참조)등에 따라 발생할 수 있는 경우의 수는 매우 다양하다.

따라서 특정 조건에서만이 아니라 제조사에서 예상할 수 없는 상황을 포함한 다양한 경우의 수가 발생하더라도 일관적으로 빠른 속도와 낮은 소비전력을 발휘할 수 있느냐가 실제 아키텍처 성능을 평가하는 잣대가 된다. 이러한 관점에서 봤을 때, 최근 국내 AI 반도체 업체들이 주장하는 성능 혹은 효율 우위는 특정 조건에서만 달성 가능하며, 일반적인 경우에 모두 해당되는 것은 아닌 것으로 보인다. 물론 국내 AI 반도체 기업들이 NVIDIA와 같이 거의 모든 경우에 원활히 사용할 수 있는 제품을 추구할 필요는 없으나, 실제로 고객이 제품을 사용해 봤더니 자사에서 필요한 연산에서는 에러가 발생해 결과를 얻을 수 없거나 속도가 큰 폭으로 떨어지는 경우가 다발한다면 고객이 믿고 제품을 구입하기 어려울 것이다.

이러한 아키텍처의 강건성은 설계의 혁신성도 중요하지만 경험, 즉 업력이 중요하게 작용하는 경우가 많다. 다시 말해 고객의 사용 경험에 대한 피드백을 받고 여러 가지 경우에 대응할 수 있도록 자사 아키텍처를 다듬는 과정이 필요하다. 또한 이는 소프트웨어 개발 역량과도 관련이 있는데, 최종 사용자(AI 서비스 개발자)가 작성하는 SW가 실제로 HW(AI 반도체 칩)를 원활하게 구동할 수 있도록 하기 위해서는 CUDA와 같은 API 뿐만이 아니라 SW in HW 등 SW와 HW 사이의 전 과정이 매끄럽게 처리되어야 하기 때문이다.

또한 AI 반도체의 경우, 특히 LLM과 같이 단일 칩 혹은 카드로 처리할 수 없는 양의

데이터를 처리해야 할 경우에는 수백, 수천 개의 칩 혹은 카드를 연결해서 방대한 양의 데이터를 서로 자유롭게 교환하며 처리할 수 있도록 하는 아키텍처의 확장 가능성이 성능 지표가 된다. 이러한 부분은 과거부터 다수의 장치를 연결해서 사용하는 슈퍼컴퓨터, 데이터 센터용 제품들을 다수 만들어 온 NVIDIA, AMD 등이 유리한 면이 있다.

NVIDIA는 이러한 아키텍처의 속도 및 효율성, 강건성, 확장성을 모두 갖추고 있는 것으로 평가되며, 이는 곧 'NVIDIA 제품을 구입하면 어떤 경우에도 문제를 겪을 가능성이 낮다'는 고객의 인식과 절대적인 시장지배력으로 이어지고 있다. 현재로서는 국내 AI 반도체 기업들의 아키텍처 경쟁력이 그 정도 수준에 올라와 있는 것으로 보기는 어렵다. 따라서 국내 기업들은 제한된 자원을 활용해서 우위를 확보할 수 있는 시장을 개척하는 전략이 필요해 보인다. 현재 NVIDIA는 대규모 LLM 학습을 위한 제품부터 개인용 GPU까지 제품 라인업을 모두 갖추고 있지만, 최근 제품으로 올수록 B100의 TDP가 700W에 이르는 등(2장 참조) 빅테크 고객을 겨냥해 대규모 LLM 학습 중심으로 제품을 설계하는 추세이다. 따라서 추론, 엣지 디바이스 등의 기타 틈새 시장을 노리는 전략이라면 NVIDIA와의 직접 경쟁 부담이 다소 적어질 수 있을 것이다.

2) 최신 공정 활용 설계 격차

국내 반도체 설계 분야 인재양성이 충분하지 못하다는 점은 앞서 인재 부문의 문제점에서 제시한 바 있으나, 조금 더 부연해서 설명하면 현재 국내 반도체 설계 분야에서 배출되는 인력, 혹은 현재 활동 중인 인력들 중에 초미세공정(FinFET을 사용하는 14nm 이후 공정) 및 chiplet(HBM)을 활용한 설계 경험을 보유한 인력이 많지 않다는 점을 지적할 수 있다. 국내 대학의 경우 디지털이 아닌 아날로그 회로 설계를 연구하는 연구실의 비율이 높을 뿐만 아니라 대학에서 고가의 FinFET 이후 공정을 활용해 칩을 설계하는 경험을 얻기 쉽지 않다는 점이 업계 관계자 인터뷰에서 지적되었다. 또한 국내에서 초미세공정을 활용한 칩을 설계하는 업체 자체도 많지 않으며, HBM 및 chiplet의 경우는 전 세계적으로도 적용된 제품이 아직 극히 소수이므로 더욱 그럴 가능성이 높다.

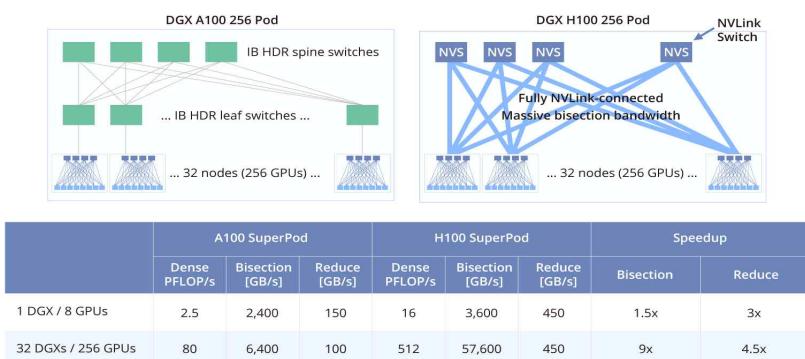
현재 국내 AI 반도체 기업을 이끄는 대표, CTO 등 핵심 인력들의 경우 IBM,

AMD, Apple, 삼성전자 등 유력 기업에서 근무한 이력이 있는 경우가 많아 이러한 우려와는 무관할 수 있으나, 그들 핵심 인력 수 명을 제외하면 전반적으로는 최신 공정을 활용한 설계 경험이 부족할 가능성이 높다. 그리고 핵심 인력들 역시 전반적으로 업계 경력이 충분히 긴 편이 아닌 점도 우려스러운 부분이다.

3) 인터커넥트 기술 격차

현재 NVIDIA의 절대적인 시장지배력의 근간을 형성하는 요소 중 하나가 칩과 칩을 연결하는 인터커넥트 기술이다. GPU 혹은 AI 칩 간 데이터 통신에 NVIDIA는 독점 규격인 NVLink를 활용하는데, 타 업체들이 활용하는 범용 규격인 PCI Express에 비해 1레인당 대역폭이 3배 넓다. 그리고 NVIDIA의 NVSwitch 칩을 활용하면 칩 간 데이터 교환의 유연성을 확연히 높일 수 있다. 또한 AI 칩을 여러 대 장착한 서버 간 통신에는 NVLink Network를 사용하는데, 타 업체들이 사용하는 InfiniBand Network 대비 대역폭이 수 배 넓고 지연시간(latency)이 낮다²²¹⁾.

[그림 3-12] NVIDIA 인터커넥트 기술 개괄



주: 원쪽이 InfiniBand Network를 사용한 경우(A100), 우측이 NVLink Network를 사용한 경우(H100).

출처: <https://community.fs.com/article/an-overview-of-nvidia-nvlink.html>

221) Howard(2024.1.29.), An Overview of NVIDIA NVLink, <https://community.fs.com/article/an-overview-of-nvidia-nvlink.html> (검색일: 2024.11.6.)

따라서 단일 칩이 아니라 수백, 수천 개의 칩을 묶어서 사용할 경우 NVIDIA의 강점이 크게 드러나게 된다. AMD 또한 칩 간 통신에 자사 CPU에서 검증된 Infinity Fabric을 사용하며, AMD와 Intel은 NVIDIA의 독자 규격에 비해서는 대역폭, 속도, 구성의 유연성이 부족하나 기존의 슈퍼컴퓨터, 데이터센터 납품 경험을 통한 노하우가 있으며, 다수의 칩을 확장해서 활용하는 데 큰 무리가 없는 것으로 알려져 있다. 이에 비해 국내 업체들의 경우 최근 리밸리온이 하나의 서버에 다수의 아톰 카드를 장착하는 멀티카드 환경에서의 검증을 받았고²²²⁾ 퓨리오사AI의 레니케이드는 최대 20대까지의 구성이 가능한 것으로 알려져 있으나²²³⁾ 멀티카드 환경에서의 속도 등 성능이나 다수 서버를 사용한 추가 확장 가능성에 대해서는 자료를 찾을 수 없었다.

나. SW 기술 격차

1) SW 기술 격차

현재 NVIDIA의 시장지배력을 형성하는 데 가장 큰 역할을 하고 있다고 평가되는 것이 CUDA이다. CUDA는 API(Application Programming Interface)로서, 사용자가 머신러닝이나 과학 연산 등의 각종 연산을 GPU에 맞춰서 프로그래밍하지 않아도 일정 규격에 따라 간략한 프로그램을 작성하면 GPU의 병렬 연산 능력을 충분히 활용할 수 있도록 HW인 GPU를 원활히 구동해 주는 SW 계층이다. 기본적으로 병렬 프로그래밍 자체가 쉽지 않고 GPU는 CPU와 달리 프로그래밍을 간편하게 해 주는 요소들이 적기 때문에 병렬 프로그래밍이 더욱 어려운데, CUDA가 이러한 작업을 간편하게 해 주기 때문에 핵심적인 연산 부분을 잘 프로그래밍해 주면 GPU를 통해 병렬 연산을 편리하게 처리할 수 있게 된다.

NVIDIA는 일찍부터(2000년대 후반) 자사 GPU를 과학기술 병렬 연산에 활용하는 데 주목하여 CUDA를 개발함으로써 시장을 선점하였다. 그리고 시간이 지나면서 머

222) AI타임스(2024.10.10.), “리밸리온, 서버용 반도체 ‘아톰’ 인정성 인증 통과”, <https://www.aitimes.com/news/articleView.html?idxno=164084> (검색일: 2024.11.6.)

223) CHESTER LAM(2024.9.12.), “FuriosaAI’s RNGD at Hot Chips 2024: Accelerating AI with a More Flexible Primitive” <https://chipsandcheese.com/p/furiosaaais-rngd-at-hot-chips-2024-accelerating-ai-with-a-more-flexible-primitive> (검색일: 2024.11.6.)

신러닝에도 GPU의 활용 가능성이 열리고 많은 머신러닝 연산 프로그램 역시 CUDA를 기반으로 작성되면서 강력한 네트워크 효과가 발생하게 되었다. 현재 개발자들이 많이 사용하는 PyTorch, Tensorflow와 같은 머신러닝 라이브러리는 CUDA와 NVIDIA GPU의 사용을 사실상 강제하고 있다²²⁴⁾. 최근 각광받는 attention 연산을 획기적으로 빠르게 해 주어 주목받는 FlashAttention 기법 라이브러리 역시 CUDA와 AMD ROCm만을 지원하고 있다²²⁵⁾. 그 결과 CUDA가 없이 범용 프로그래밍 언어를 사용해 GPU 혹은 AI 칩을 효과적으로 활용하는 전용 SW프로그램을 처음부터 작성하는 데에는 막대한 시간과 돈(개발자 자원)이 필요하므로²²⁶⁾, 업계 전체가 CUDA의 lock-in 효과에 걸려 있는 상황이다²²⁷⁾.

아직 완성도가 낮지만 Intel의 OneAPI, 오픈소스 규격인 OpenCL 등 대안이 늘어나고 있어 NVIDIA의 독점력이 다소 약화되고 있는 정황도 엿보이나, 여전히 많은 개발자들이 CUDA에 의존하고 있다. 더불어, 24년 초 AMD에서 발표한 MI300X과 NVIDIA의 H100, H200의 비교를 보면²²⁸⁾, 발표된 spec상의 수치는 압도할 수 있는 성능으로 보이나 호환성, 사용편의성 등을 고려한 소프트웨어 부문에서 낮은 점수를 받았다²²⁹⁾. 퓨리오사AI²³⁰⁾ 등 국내 업체들은 자체 SDK를 제공하고 ONNX, 파이토치 연결, Meta의 오픈소스 LLM ‘LLaMa 3.1’을 추론하는 등 SW 개발 환경에 주력하고 있다고 홍보하고 있으나, 고객 운용 사례가 없어 다양한 환경에서 실제 작동이 원

224) 최근 PyTorch가 CUDA에 대응되는 AMD의 ROCm을 지원하기 시작하였다(Linux 환경에서만 가능, 2024년 11월 6일 기준).

225) 출처: Dao(2024), “FlashAttention”, <https://github.com/Dao-AI-Lab/flash-attention> (검색일: 2024.11. 6.).

226) 경향신문(2024.3.20.), “엔비디아가 ‘넘사벽’인 이유... “다들 ‘구다’만 찾아”, <https://www.khan.co.kr/economy/economy-general/article/202403200700001> (검색일: 2024.11.6.)

227) 자체 칩과 충분한 개발자 자원을 가지고 있는 Google 등 대기업의 경우 이러한 제약에서 다소 자유로운 편이며, CUDA의 지배력을 그리 높게 보지 않는 인터뷰 대상자도 있었다.

228) 김우용(2025.1.2.), AMD GPU가 엔비디아를 못 따라잡는 이유 '그놈의 SW', <https://byline.network/2025/01/2-239/> (검색일: 2025.1.6.)

229) Patel et al.(2024.12.22.), MI300X vs H100 vs H200 Benchmark Part 1: Training – CUDA Moat Still Alive Training Performance, User Experience, Usability, Nvidia, AMD, GEMM, Attention, Networking, InfiniBand, Spectrum-X Ethernet, RoCEv2 Ethernet, SHARP, Total Cost of Ownership, <https://semanalysis.com/2024/12/22/mi300x-vs-h100-vs-h200-benchmark-part-1-training/> (검색일: 2025.1.6.)

230) 아주경제(2024.8.28.), “‘레나케이드’, 엔비디아 AI칩보다 전성비 60% 우수”... 백준호 퓨리오사AI 대표 자신감”, <https://www.ajunews.com/view/20240828101412571> (검색일: 2025.1.6.)

활한지는 아직 검증되지 않았다.

한편 CUDA와 같은 API 뿐만이 아니라 AI 반도체의 성능을 최대한 이끌어내기 위해서는 compiler, driver, firmware 등 SW in HW의 최적화가 매우 중요하다. 이러한 SW in HW의 개발을 위해서는 SW와 HW에 모두 정통한 인력이 다수 필요하며, 단기간에 완성도를 높이는 것이 쉽지 않다. 그리고 이러한 인력은 세계적으로도 드문 편이며, 우리나라의 경우 HW(회로) 설계 인력 대비 더욱 구하기 어려운 것으로 알려져 있다. 인터뷰 대상자 중에도 이러한 SW 인력을 구하기 힘들다고 응답한 경우가 있었다. 따라서 향후 국내 기업들이 본격적으로 판로를 개척하는 데 있어 이러한 SW 역량 부족이 발목을 잡을 가능성이 있다.

2) SW 관련 고객지원 역량 격차

어쩌면 NVIDIA의 시장지배력의 근간은 CUDA 자체보다 CUDA와 기타 SW in HW 를 개선해 나가는 SW 관련 고객지원 역량의 격차일 수도 있다. SW는 일반적으로 다양한 고객들이 다양한 활용 상황에서 예상치 못한 에러에 맞닥뜨리면 제조사가 상황을 파악하고 SW 업데이트를 통해 문제를 해결해 나감으로써 개선되는 과정을 거친다. 그리고 이러한 과정은 대체로 노동집약적인 과정으로, 다수의 SW 관련 고객지원 인력을 보유한 대기업에게 유리하다. 한 인터뷰 응답자에 따르면 NVIDIA는 현재도 고객이 문제를 보고하면 매우 적극적으로 대응하며 문제를 빠르게 수정한다고 한다. 즉 NVIDIA의 CUDA나 기타 SW in HW 등은 지난 수 년 간의 업데이트를 거쳐 현준하는 대부분의 연산을 지원하고 원활히 작동했던 실적이 있으므로, 고객들은 NVIDIA의 제품을 사용하면 대부분의 상황에 대처 가능할 것이라는 믿음을 갖고 있다.

국내 AI 반도체 업체들의 경우 아직 고객의 피드백을 제대로 받고 있지 못하기 때문에 SW 관련 고객지원 역량 역시 검증되었다고 보기 어려우며, 고객지원에 대한 경험 자체가 부족한 것도 우려스러운 점이다. 각 사 홈페이지를 살펴보면 일부 LLM 구동 사례를 홍보하는 것 외에 인상적인 서비스 개발 예시 제시를 찾아보기 어려우며, 개발문서의 양과 질 역시 NVIDIA는 물론 AMD 등에 비해서도 많이 부족한 편이다. 또한 개발자 커뮤니티 지원(Q&A) 등은 현재까지는 B2B 비즈니스인 특성상 홈페이지

등에 적극적으로 표시하지 않는 것은 이해할 수 있으나, 앞으로 개선되어야 할 부분으로 생각된다.

다. 실제 칩 양산 및 검증 경험 부족

1) PoC(Proof of Concept) 및 대규모 실증 경험 부족

앞서 기업 경쟁력 및 생태계 부분에서도 유사한 문제가 언급되었는데, 여기서는 보다 기술적인 관점에서 해당 문제를 접근하고자 한다. 실제 칩 양산 경험이 부족한 경우, PoC, MPW, 양산, test 및 대규모 성능 검증 등 각 과정에서 불량이 발생하여 재설계(revision)을 피할 수 없는 상황에 처하거나, 생산된 칩이 목표 성능에 미달하거나, 불량이 다수 발생하거나(낮은 수율), 개발 지연 등의 치명적인 추가비용이 발생하는 경우가 있다. NVIDIA 역시 최근 Blackwell 제품의 재설계를 거치는 등²³¹⁾ 대기업에게 조차 드문 사례는 아니지만, 국내 스타트업들의 경우 재정적으로 여력이 크지 않아 이러한 리스크에 취약하다고 할 수 있다.

한편 실제 제품을 양산해 고객에게 제공하면 고객은 짧게는 수개월에서 길게는 수년에 걸쳐 대규모로 성능 및 신뢰성을 검증하게 된다. 이러한 검증은 온도, 전압, 연속 구동 등 각종 극한 환경에서의 신뢰성, 장기간 사용에 대한 신뢰성, 각종 기능 검증 등 다양한 항목에 대해 이뤄진다. 기존 업체들은 시행착오를 거쳐 이러한 성능 및 신뢰성 문제를 수정한 경험이 있어 문제를 미연에 방지할 수 있는 노하우를 가지고 있으나(예를 들어 이전의 경험으로부터 체크리스트를 축적하고 추후 개발과정에서 이를 준수하는 시스템을 갖춤), 국내 스타트업들의 경우 처음으로 제작하는 칩인 만큼 이러한 경험을 갖고 있지 못한 경우가 많아 검증 과정에서 예기치 못한 문제에 직면할 가능성이 있다.

231) 전자신문(2024.10.24.), “젠슨 황 “엔비디아-TSMC 불화설은 ‘가짜뉴스’… 블랙웰 설계 결함 고쳤다””, <https://www.etnews.com/20241024000342> (검색일: 2025.1.6.)

2) 호환성 검증 부족

먼저 SW 측면의 호환성은 앞서 SW 경쟁력 부분에서 상당부분 논의한 바 있다. 거기에 추가로 덧붙이자면, 윈도, 리눅스 등 다양한 OS에 대한 지원이 취약할 가능성이 있다. AMD조차 HPC라이브러리를 모든 OS에서 원활하게 지원하기 어려운 상황을 감안하면 NVIDIA 외에 SW 지원이 잘 되는 기업이 오히려 드물다고도 할 수 있다. PyTorch, Tensorflow 등에 대한 지원뿐 아니라 각종 프로그래밍 언어에 대한 지원도 각 기업 별 반도체 활용의 차원에서 기본적으로 지원해야하는 필수 요소라고 볼수 있다.

한편 HW 측면에서는 각종 CPU(x86 or Arm), GPU 등과의 호환성이 검증되지 않은 상태로 볼 수 있다. 또한 서버 구성에서의 동작은 앞서 언급하였듯 검증이 되고 있으나, 이를 더욱 확장한 데이터센터 수준에서의 동작은 검증되지 않은 것으로 보인다.

라. 기술 조달 가능한 커뮤니티 미비

국내 logic chip 설계 관련 국내 대학(학계), 연구소, 기업 등 주변 커뮤니티 규모는 해외에 비해 작은 편이다. 또한 대기업과 스타트업 간 인력 이동도 쉽지 않은 편으로, 기술 교류 및 피드백이 일어나기 쉽지 않은 환경이며 인력조달에도 어려움이 있다. 이러한 상황은 당장 쉽게 극복될 수 없기 때문에, 장기적인 관점에서 바라볼 필요가 있다. 한국팹리스산업협회가 발족된 것은 지난 2022년 8월²³²⁾으로, 향후 국내 반도체 설계 관련 커뮤니티의 활성화에 중요한 역할을 할 것으로 기대된다.

232) 전자신문(2022.8.17.), 팹리스산업협 출범… “시스템 반도체 강국 도약”,

<https://www.etnews.com/20220817000224> (검색일: 2025.1.6.)

제3절 기업의 성장 요인 도출

1. 서면 자문 및 분석

마인드맵의 분석에 대한 서면 자문을 통하여, 지적된 문제점에서 우선적으로 대응하여 해결해야하는 문제점을 분석하였다. 동 분야에 대해 투자기업을 가지고 있는 VC, PE 심사역, 협회 종사자로서 AI팹리스 스타트업의 지원, 평가, 육성을 담당하고 실무자, 동 분야 연구를 진행하고 있는 정출연 연구자 및 학계의 교수, 동 분야 R&D를 진행하여 직접 칩을 설계하는 엔지니어, 스타트업 연구자 등 다양한 배경을 가진 전문가 서면 자문을 통해, 연구진이 진단한 문제점의 feasibility와 추가되거나 수정해야 할 문제점에 대한 자문과 1단, 2단, 3단으로 분석된 마인드맵에서의 문제점 중 우선으로 고려해야 할 사항에 대하여 의견을 받고 분석을 진행하였다.

<인재>

구분	사유
인재	<ul style="list-style-type: none"> 국내 반도체 생태계는 메모리 반도체 IDM 2개 사 중심으로 형성, 상대적으로 영세한 팹리스에 석·박사 이상의 고급인재 공급이 어려움 생태계 확산을 통한 직업의 안정성, 수입 등 현실적인 차우 개선 역시 필요 글로벌 경쟁력 있는 인력이 매우 부족하며 특히 소프트웨어 인력이 크게 부족 → 1) 스타급 업체를 육성하여 그나마 있는 글로벌 인재를 모아야 하며 2) 해외에 있는 글로벌인재가 한국에 들어올 유인을 만들어야 한다 반도체 관련 인력 노후화(고령화)/감소, 일본 사례같이 반도체 전문인력 감소 AI반도체 설계를 위해서는 설계 고급인력 투입이 필수적으로, 석박사 이상의 인력공급이 필요함 → 중장기적으로 석박사 인력을 공급할 수 있는 대학원 중심 연구환경 구축이 필수 대출 수준의 초급전문가 또한, 기업수요에 맞는 교육시스템이 갖추어져야 하며, 이를 위해서 팹리스기업들의 맞춤형 수요를 반영하여 교육할 수 있는 산업협회 등에서 교육을 전담하는 것이 필요 또한, 시스템반도체를 위한 임베디드 SW, 시스템 SW 개발 전문가 양성도 꼭 필요함 → 우리나라 시스템 SW개발 전문가도 매우 부족 고급 인력이 국내 대기업 또는 해외 대기업으로 취업을 선호하는 경향이 강함. 그러나 이러한 경향은 국내 반도체 스타트업의 성공 사례가 증가하게 되면 자연스럽게 해소될 것으로 생각됨. 기술 역시 학계에서 연구되는 수준은 해외와 큰 차이가 없을 것으로 생각되나, 실제 양산 및 기업 측면에서의 연구가 부족한 것으로 사료됨. 따라서, 기술, 인재 부문에서는 현재 다소간 부족한 점은 존재하나, 관련 산업이 성장함에 따라 자연스럽게 극복 가능할 것으로 기대.

구분	사유
	<ul style="list-style-type: none"> • 반도체의 경우 초기 설계단에서부터 성공 가능성 여부가 중요한 산업이며, 설계부터 양산까지 선진기업에서 경험해 본 사람을 보유하고 있으면, 성공 가능성은 더욱 높아질 것으로 판단됨. 반도체 산업에서 설계부터 양산까지 풍부한 경험을 보유한 인재를 확보하기 위해서는 해외에서 국내로 복귀하려는 인재 빌굴과 함께 미국, 독일, 일본, 인도 등의 인재를 확보할 수 있도록 전폭적인 창업 지원제도를 구축할 필요가 있음. • AI반도체 기술분야는 고도의 전문성을 요구함. 설계, 제조, 응용 전 단계 별 깊이 있는 지식/경험 요구 • HW 및 SW, 알고리즘 등 다양한 분야가 함께 고민되어야 함. 융합적 사고를 할 수 있는 고급 인력이 필요 → 급변하는 기술변화를 추종하고 선도할 수 있는 인재 필요 • AI 반도체 산업의 성장에 필요한 핵심은 우수한 인재임 → 기술 개발과 제품의 시장 적응력 강화에도 직접적인 영향을 미침 • 기술을 증진시키기 위한 연구개발의 핵심 중 하나는 인재임. 뿐만 아니라 사업화 및 검증 및 실증에도 인재들이 필요함. 이는 AI 반도체 산업이 첨단기술과 지식을 요하는 고부가가치 산업이기 때문임. 인재를 확보하지 못하면 한국의 AI 반도체 산업의 성장에는 한계가 있을 수 밖에 없음 • AI반도체관련 분야를 연구하고 학생을 교육하는 교수가 부족하고, 졸업생은 외국기업이나 대기업을 선호 • 국내에 AI반도체 설계/개발/양산/SW 전문가 수가 적고, 역량있는 구직자의 극소수가 AI반도체 팝리스에 취업 • 경력 및 역량있는 AI반도체 팝리스 재직자들이 지속적으로 대기업으로 이직 • 창업 그룹 임직원 성과 보상 체계: 과거 INTEL, AMD 등 신화적 기업은 예외없이 전 임직원 대상 충분한 스톡옵션 부여, 고급 인력의 확보는 우선 팝리스 스타트업 창업그룹의 보상 수준 제고 외 다른 방법은 그닥 효과적이지 못할 듯. 일부 임금 수준과 대우를 올려주었다는 주장이 많으나, 직업 안정성과 커리어 관리 측면에서 본인의 자제들을 대기업이 아닌 스타트업에 취업시킬 것인지 고민해 볼 필요

• 인력 수급 현황 및 문제점

- 국내 반도체 생태계가 메모리 반도체 IDM 2개사 중심으로 형성되어 팝리스 기업의 고급인력 확보가 어려움, AI반도체 분야의 석박사급 전문인력이 극히 부족하며, 특히 시스템 SW 개발 전문가가 매우 부족, 기존 인력의 고령화 및 전문 교수진 부족으로 인재 양성에 어려움

• 핵심 인재 유출 문제

- 우수 인재들이 대기업이나 해외 기업 취업을 선호, 팝리스 스타트업 재직자들의 대기업 이직이 지속적으로 발생, 직업 안정성과 처우 문제로 인한 인재 유출이 심각

• 우선 고려 사항

- 대학원 중심의 연구환경 구축으로 중장기적 석박사 인력 양성 필요, 글로벌 인재 유치를 위한 해외 Returnee 발굴 및 외국 전문가 영입 지원, 스타트업의 성과보상체계 개선(스톡옵션 등)과 처우 개선을 통한 인재 유치, 협회(학계, 산업계 Bridge 기능) 추진 맞춤형 교육시스템 구축으로 실무형 인재 양성 등

<비즈니스 모델 & 생태계>

구분	사유
비즈 니스 모델 & 생태 계	<ul style="list-style-type: none"> PoC 및 대규모 실증 역량 부족: 데이터센터 용 NPU 등 AI 가속기(HW) 설계/개발이 이루어져도 실질적으로 적용 가능한 수요처 확보에 어려움이 존재 → 글로벌 수요처 확보를 위한 노력, 해외 빅테크 데이터센터 수요에 부합하는 HW/SW 스펙을 제시 수요 다변화 중인 Edge向 AI반도체 개발·양산에 집중 필요 제품 및 사업 기획 단계 기술사업화 가능성 검증 필요 거래관계에 대한 경로 의존성 강한 B2B 판로확보 방안 및 개발 제품의 지식재산(IP) 전략 등 검토 미흡 서버 쪽은 1) 칩 개발, 생산에 들어가는 비용은 매우 큰데, 2) 이를 개발하려는 창업팀의 역량이 제한적이라서 (글로벌 리딩 업체 대비) 경쟁력 있는 칩을 개발하기 어렵고, 3) 그래서 수요처와 협력 등이 잘 이루어지지 않음. Edge 쪽은 1) 개발하려는 업체가 매우 많은데 비하여 (경쟁이 매우 심함), 2) 개발된 칩이 경쟁사 대비 월등하기 어렵기에, 3) 수요쪽과 밀접하게 일하는 것이 중요 → 한국보다는 중국, 대만이 유리한 상황으로 판단 Value Chain: AI반도체에 집중되어 있고, Value chain상 End Customer의 Needs Speedy하게 반영 어려움 글로벌 진출능력 부재: 개발제품의 글로벌 진출을 위한 마케팅 및 네트워크 능력 부족 인공지능 반도체에 대한 적용 시장: 다양한 분야에 적용 가능할 수 있는 분야인 만큼, 이 반도체들이 활용될 수 있는 보완(Complementary) 산업이 함께 성장해야 함. PoC를 통과하면 이를 시장에 적용을 해야 하는데 이 시장 테스트를 위해서는 다양하며 활발한 산업이 있어야지만 지속적인 비즈니스가 가능. 국내시장이 협소한 만큼, 관련하여 국제협력이 필요할 수 있어 보임. 주로 글로벌 빅테크(하이퍼스케일러 포함)를 대상으로 하는 제품을 개발하고 있어서 연구소, 대학 및 중소기업을 대상으로 진행되는 PoC 또는 검증은 의미없는 상황임. 대규모 실증을 하기 위해서는 현재 운영 중인 데이터센터에 적용해야 하나, 국가 과제 등이 아닌 다음에야 일반 기업이 테스트용으로 스타트업 제품을 시험 또는 적용하는 것은 매우 큰 위험 요인일 수 있음. 단순 레퍼런스 체크 이외의 생태계가 구축되지 않아 실증이 어렵다는 점이 개발에 난도를 높이고 있는 것으로 판단. 대기업 수요자 인센티브 필요: 국내 AI반도체는 미국과 달리 태생적으로 정부의 지원없이는 성장이 힘든 구조로 대기업 수요자가 국내 스타트업과 초기 설계단에서 구매조건부로 협력, 투자 참여한다면 해당 기업에게 세제혜택 등의 인센티브 제도 활성화 필요할 것으로 판단됨 국내 AI반도체 기업은 대부분 NPU 기반의 Inference(추론) AI반도체를 개발 및 생산하고 있어, SoC 수요기업에 NVIDIA와 같이 기존 Training AI반도체와의 범용성, 확장성, 호환성에 대한 다양한 레퍼런스를 요구하고 있습니다. 수요기반의 실증 및 사업화 역량 부족: 국내 기업이 주력으로 개발 및 생산하고 있는 추론용 AI반도체(NPU)는 연산속도와 전력효율에 강점을 가지고는 있으나, PoC 및 대규모 실증 역량 부족은 결과적으로 사업화에 도달할 수 없는 어려움으로 이어지고 있습니다. 기술개발에 성공하였어도 수요처 부족과 레퍼런스 확보 실패는 종장기적으로 가장 큰 걸림돌이라고 생각합니다. 실제 칩 양산 및 검증 경험 부족: MPW, 양산 등 개발 및 생산과정에서는 이슈가 작으나, 이를 시스템 Level에서 적용 및 검증하는 과정에서 다양한 주변기기 또는 기존 Training AI 칩과의 호환성 부족은 개발된 칩의 성장 저해 요인이라고 생각됩니다.

구분	사유
	<ul style="list-style-type: none"> 또한, 시스템 레벨에서의 SW(파이토치 등)와 HW(윈도우, 리눅스 등)와의 호환성 부족 역시 국산 AI반도체의 검증 및 사업화에 큰 어려움이라고 생각됩니다. 협력 생태계 구축 및 상호 협력을 통한 글로벌 시장경쟁력 확보: AI 반도체 시장에서 생존하고 성장하려면 글로벌 시장에서도 경쟁력을 갖춰야 함. 협력 생태계를 통해 소프트웨어 스택, 호환성, 고객 지원 등 다양한 생태계 요소를 강화하여 글로벌 시장에 적합한 제품을 개발하는 것이 중요함 기술이 선도적일 경우, 투자와 기업경쟁력을 향상 시키는 방안 등은 시장에의해서 자연스럽게 따라올 수가 있다고 봄. 특히 AI 반도체에 대한 수요가 급증하고 있는 현 상황에서 기술력 있는 기업에 대해서 시장이 반응하지 않을 수 없음 수요연계: 네트워크, 기존 성과 등이 미흡하여 빅테크 기업 등과 교류 및 협력이 부족하여 향후 기술 발전 방향, 수요처의 요구사항 등에 대한 파악이 지연되어 선행기술/제품 개발, Time-to-Market 등에 한계 봉착, 대부분 수요처는 퍼스트 무버 역할에 미온적임 다양한 극초기 투자주체(CVC, Angels 대학/출연연 등): 인공지능 반도체의 경우 일반 반도체의 경우와 달리 아주 특정한 계산 능력에 초점을 두기 때문에 다양한 칩 설계가 나올 수 있음. 이는 주로 연구개발 주체(중소기업을 포함한 기업, 대학, 출연연 등)에서 나올 가능성이 높은데, 실증을 위한 높은 진입장벽으로 인해서 논문 수준에서 그칠 가능성이 높다. 따라서 이러한 기술을 '극' 초기에 발견해서 지원할 수 있는 체계가 필요. 다양성이 많아야 전체 생태계가 살아날 수 있을 것으로 보임. 정부 등 지원주체의 역할: 아이디어의 실증을 위해서 이를 지원한 체계가 필요할 것으로 생각됨. 타 산업과는 달리 PoC가 아주 중요하고 이를 위해서는 아주 많은 자금이 필요. 따라서 이를 지원 할 수 있는 조직(누구나 쉽게 관련 장비를 사용할 수 있는) – 특히, 비영리 Incubator/Accelerator 의 역할이 중요.

• PoC(Proof of Concept) 및 실증 문제

- 개발된 AI 가속기의 실질적 수요처 확보 어려움, 대규모 실증을 위한 데이터센터 테스트 환경 부족, 스타트업 제품 적용에 대한 기업들의 위험 회피 성향, 레퍼런스 확보 실패가 사업화의 주요 걸림돌

• 수요 기반 한계

- 국내 시장 규모의 혐소성, 대부분의 수요처가 퍼스트 무버 역할에 미온적, 빅테크 기업과의 교류/협력 부족으로 인한 수요 파악 지연, 보완(Complementary) 산업 생태계 미성숙, Value Chain 상 최종 고객 니즈 반영 지연

• 시장 진입 장벽

- 서버 시장: 높은 개발비용 대비 제한적 창업팀 역량, Edge 시장: 과다 경쟁 및 중국/대만 대비 열위, B2B 판로확보의 어려움과 IP 전략 미흡, 글로벌 마케팅 및 네트워크 역량 부족

• 기술적 호환성 문제

- 기존 Training AI 칩과의 호환성/범용성 부족, 시스템 레벨에서 SW(PyTorch 등)와 HW 호환성 부족, 주변기기와의 통합 및 검증 경험 부족,

• 우선 고려 사항

- 대기업 수요자 인센티브 제도 도입(세제혜택 등), 극초기 투자 지원체계 구축(CVC, Angels, 대학/출연연 등), 비영리 Incubator/Accelerator 통한 실증 지원, 글로벌 협력 생태계 구축을 통한 경쟁력 강화

〈투자〉

구분	사유
투자	<ul style="list-style-type: none"> • 투자유치: 기존 국가 사업 또는 프로그램 등을 활용하여 MPW를 제작하여 기술력을 검증하고 제한적으로 PoC를 진행중이나, 양산매출을 달성하여 중소기업에서 중견기업으로 성장하기 위해서는 장시간이 소요되므로 대규모 투자유치가 필요 • 막대한 자금 : AI 반도체의 연구개발, 기술사업화에는 막대한 자금이 필요하면 이를 공급해주는 것은 민간 및 공공 투자임. 모험자본의 투자없이 새로운 제품을 개발하는 팝리스는 연구개발을 할수도, 인재를 유치할 수도 없어 기업이 성장 할 수 없음. • 지속적 자금 투입: 기술개발 및 제품의 시장가치를 제공할 수 있을 때 까지 많은 시간, 자본을 요구함. 장기적 지속적인 투자가 필수적. 공공투자 규모의 한계로 민간투자 시장의 확대가 필수 • 막대한 초기 투자 필요성: 10nm이하 선단공정을 사용하여 칩 개발시 높은 개발비용(IP, DSP, PKG&TEST 등)을 감당하기 위해서는 막대한 자금이 필요하나, 현재 정책편드는 VC 레벨의 투자규모로(10~20억내외), 칩 한번 개발시 필요한 4~500억원의 투자금을 모으기에는 쉽지 않은 구조. • 후속투자(Follow on 투자편드), 대규모 AI 프로젝트 편드(4~500억원 규모) 등이 조성되어 적기에 자금을 확보해야, PoC 및 개발, 인력확보, 칩양산 및 실증 과정을 통해 사업을 진행할 수 있음 • 종후기 원활한 펀딩 필요: 해외 VC들의 경우 국내 PE 규모의 세자리 수 투자가 빈번하게 일어나고 있으나, 국내 VC들의 경우 초기 투자 이후에는 운용 펀드 규모 및 펀드 포트폴리오 리스크 관리 측면에서 대규모 추가투자가 일어나기 어려움. 따라서, 매출이 발생하지 않는 상황에서도 PE들이 과감하게 투자를 이어받을 수 있는 환경 (특수목적 펀드 조성 등) 조성이 요구됨 • 국내에서는 M&A 사례가 많지 않기 때문에, 시장에서 궁극적인 Exit 전략은 IPO 밖에 없는 상황임. 이에 국내 시장 상장을 위해 한국거래소의 기아드라인을 따르고, 특히 창업자 및 우호자분의 지분율을 일정 수준 이상 확보하여야 함. 또한, 선단 공정 사용을 위해 최소 500억 내외의 개발 비가 필요함. 이 두 가지 요인으로 인해 투자 규모와 회사 가치가 급격하게 상승하게 되면서, 타 분야 스타트업 대비 투자 결정이 어려워짐. 반도체 개발에 일정 기간이 소요되며, 이 기간 동안에는 매출이 부진하므로 투자금으로 운영될 수 밖에 없기 때문에, AI반도체 스타트업에 있어서 투자 유치 중요성이 갈수록 증대되고 있음. • 개발자금 부족: AI반도체는 대부분 선단공정으로 제작되므로, 많은 개발비용 이외에서 PoC 등 양산을 위하여 막대한 자금이 소요됨. 또한, 하이리스크-하이리턴 형 사업으로 전폭적인 투자가 이루어져야 하므로, 팝리스산업에 특화된 투자환경 혁신이 필요 • exit 시장: 국내 Exit 시장 감안시(M&A가 부족한), 조단위 상장을 위해서는 자본 시장 크기와 펀드가 장기간 support 할 필요 • 투자금이 많이 필요하기에 벤처에이션이 높을 수밖에 없으며, 따라서 투자 난이도가 매우 높은 상황, 정부 지원 등으로 개발자금 일부를 지원해준다면 (글로벌 경쟁력 있는 곳만 선별하여) 의미 있을 수 있음 • 장기 퇴임기간 인내 자본의 필요: 현재 투자가 물린 주요 3사 제품의 사업화 결과가 부진 시 AI 반도체 뿐 아니라 시스템반도체, 팝리스 부문 벤처투자 등 기술금융 규모 급감 우려, 현재 AI 반도체 스타트업의 단계별 기술사업화에 소요되는 자금이 단위 기업 당 충분히, 장기간 제공되지 못함

- 대규모 초기 투자 필요성

- AI 반도체 개발은 10nm 이하 선단공정 사용시 IP, DSP, PKG&TEST 등에 400~500억원 규모의 막대한 초기 자금이 필요, 현재 정책편드는 VC 수준인 10~20억원 규모로, 실제 필요 자금과 큰 격차가 있음, 기술개발과 시장가치 창출까지 장기간의 지속적 투자가 필수적이나 공공투자만으로는 한계가 있음

- 중후기 투자 환경의 한계

- 해외 VC들은 대규모 후속 투자가 활발하나, 국내는 초기 투자 이후 대규모 추가투자가 어려운 구조, 매출이 없는 상황에서도 PE들이 과감한 투자를 할 수 있는 특수목적 펀드 등의 환경 구축이 필요, 국내 M&A 시장이 미흡하여 IPO가 유일한 Exit 전략이 되는 상황이며, 이는 투자 유치를 더욱 어렵게 만듦

- 투자 활성화를 위한 과제

- 팹리스 산업에 특화된 투자환경 혁신이 필요하며, 하이리스크-하이리턴 특성을 고려한 전폭적 지원 필요, 글로벌 경쟁력 있는 기업을 선별하여 정부 지원으로 개발자금 일부를 지원하는 방안 검토, 현재 주요 AI 반도체 스타트업의 성공이 향후 전체 시스템반도체, 패리스 부문의 벤처투자 생태계에 중요한 영향을 미칠 것으로 예상

<기술>

구분	사유
기술	<ul style="list-style-type: none"> • 호환성 검증 부족: 정형화된 메모리 반도체 대비 SoC, NPU를 비롯한 AI 반도체는 상대적으로 프로그래밍 언어에 대한 높은 이해도가 필요, HW 설계뿐 아니라 HW-SW 간 interface, API, application 등 종합적으로 고려할 수 있는 SW 분야 기술력 강화 필요 • 첨단 기술 확보: AI 분야에서는 현존 최고의 고성능, 고효율 반도체 사양을 요구하고 있음, 글로벌 기업과 차별화 된 가치를 제공하는 기술경쟁력 필요, 분야별 빠르게 발전하는 기술 트렌드에 맞춰 지속적인 기술개발을 통한 선도기술의 개척 또한 개발 중요 • SW 기술 격차: 국내 AI반도체 기업은 업계선두 대비 SW경쟁력이 매우 부족한 것으로 평가 받고 있습니다. 칩 SW(SDK 등) 경쟁력이 떨어져 테스트 및 성능 검증 등의 과정에서 Trial & Error가 증가 <p>• 다수의 내용이 인재, Business model, 투자 부분과 연계되어 기술 되어 있음. 그럼에도 주요하게 기술 부분으로 특정 지을수 있는 문제점은 호환성과 SW부분의 기술을 발전 시키고 확보하는 것에 대한 중요성을 강조하는 점</p>

2. 주요 기업 인터뷰

앞선 도출된 AI 반도체 생태계의 문제점에 대한 주요 기업의 의견을 청취하고, 보완하고자 연구 대상에 부합하는 IP, 디자인하우스, 팹리스 스타트업 (엣지 AI, 온디바이스 AI 등) 분야의 국내 주요기업의 인터뷰를 진행하였다. 특히, 기업들의 창업 동기, Business model, 투자, 인재, 기술 부문의 주안점, 애로사항, AI 반도체 생태계 발전 방향에 집중하여 인터뷰를 진행하였다.

□□□□ 대표 이사 ○○○	
	<input type="checkbox"/> AI 반도체 시장과 NVIDIA의 우위성 <ul style="list-style-type: none"> - NVIDIA가 시장을 선도하는 이유는 단순히 쿠다(CUDA) 생태계 때문이 아님, CUDA 생태계의 Lock-in 효과도 분명 존재하나, 빅테크는 각자의 칩을 생산하고 있으며, 자사 칩을 활용하는 SW를 구축 (구글 TPU) - 본사의 경험을 비춰보면, NVIDIA는 고객 요구사항을 1~2주 내에 신속하게 반영하고 제품에 구현하는 강점 보유 - 특히 LLM(대규모 언어모델) 시대에 맞춰 H100, A100 등을 트랜스포머 아키텍처에 최적화하여 시장에 출시한 전례가 있음 - 실제 서비스 환경에서 요구되는 성능/가격비가 가장 뛰어남 - 테라바이트급 메모리 처리, HBM 컨트롤러 등 핵심 기술을 보유하고 있음
주요 내용	<input type="checkbox"/> 국내 AI 반도체 기업들의 문제점 <ul style="list-style-type: none"> - 실제 수요 기업과의 소통/피드백이 부족함 - 성능 수치 과대 홍보에만 치중(예: "NVIDIA 대비 600배 성능" 등) - MLPerf 등 벤치마크 수치에만 집중하고 실제 서비스 요구사항 반영 미흡 - 학술적/이론적 접근에 치중하고 실용성/사업성 고려 부족
	<input type="checkbox"/> 투자 생태계의 문제 <ul style="list-style-type: none"> - VC들이 기술 평가 역량 부족으로 제대로 된 심사 어려움 - 다른 투자자들이 투자했다는 이유로 따라 투자하는 관행 (클럽딜) - 과기부 중심의 R&D 지원으로 산업부/중기부 등 기술개발 외에 수요 연계 등 비즈니스 적인 측면에서의 지원 부족 - 실제 서비스/시장 검증 없이 칩 제작부터 시작하는 잘못된 접근
	<input type="checkbox"/> 개선 방향 <ul style="list-style-type: none"> - 수요 기업들과 적극적인 소통을 통한 실제 요구사항 파악 필요 - 성능 수치보다 실제 개발된 칩을 활용한 서비스 및 비즈니스에서의 활용성 중심 접근 - 칩 제작 전 충분한 시장/수요 확인하고 제품의 검증이 필요 - NVIDIA의 한계점/문제점을 파악하여 차별화 포인트를 발굴하는 것도 좋은 전략으로 볼 수 있음

□□□□ 대표 이사 ○○○	
	<ul style="list-style-type: none"> - 학교/연구소는 연구용 칩 개발에 집중하고, 기업은 실용화에 주력하는 방향으로 정책 지원을 다양하게 진행하는 방향도 고려할 수 있음
	<p>□ 기술 개발 방향</p> <ul style="list-style-type: none"> - 단순 성능 향상보다 실제 서비스 최적화에 초점을 둔 기술 개발 필요 - 현재는 5나노 공정으로도 충분한 성능 구현 가능한 점을 고려할 필요 (2나노, 3나노 공정 활용만이 해답은 아님) - HBM 컨트롤러 등 AI반도체 생태계 내에서의 핵심 기술 확보 중요 - 칩 설계 능력보다 비즈니스 모델과 컨셉 기획력이 더 중요 (기술적으로 개발이 되는 것은 AI 반도체 생태계 내에서의 Default 조건) - 학술적 접근보다 실용적 문제 해결 중심으로 전환 필요
	<p>□ 기업 측면</p> <ul style="list-style-type: none"> - 실제 서비스 제공자들의 요구사항 파악, 부정적 피드백도 수용하는 자세, 실질적인 성능 검증 필요
	<p>□ 정책 측면</p> <ul style="list-style-type: none"> - R&D 지원 방향성이 명확한 제시, 민간 전문가의 평가 참여 확대 (수요기업의 참여는 적극 환영), 실용적 성과 중심의 지원체계 구축

□□□□ 대표 이사 ○○○	
	<p>□ 창업 배경과 동기</p> <ul style="list-style-type: none"> - 반도체 연구로 박사학위 취득, 반도체 업계 경험은 전무 - 당시 AI 반도체 인력 공급이 부족했고, 대부분의 연구실 동료들은 실리콘밸리 기업이나 해외 포털을 선택했으나, 대기업보다는 AI 반도체를 직접 처음부터 끝까지 연구개발하고 상용화하고 싶은 욕구가 있었음 - 2019년에 창업했으며, 당시에는 해외 AI 반도체 스타트업들이 이미 많이 설립된 상황이었음 (Groq, Graphcore 등) - 기술적 경쟁력을 갖출 수 있다는 판단과 한국에서의 창업이 의미 있다고 생각해 시작한 것이 계기
주요 내용	<p>□ 인재 영입과 조직 구성의 특징</p> <ul style="list-style-type: none"> - 전체 인력의 약 25%만이 반도체 관련 인력이며, 75%는 소프트웨어/컴파일러/알고리즘 인력으로 구성 - 코파운더들도 수학과 출신으로 차별화된 접근법 추구(AI 반도체 컴파일러 개발 경험자가 극히 드물어, 수학적 배경을 가진 인재를 선별해 육성하는 전략 채택) - 기술적 차별화와 글로벌 경쟁력을 위해 수학적 백그라운드가 필수적이라고 판단 <p>□ 투자 유치와 자금 조달</p> <ul style="list-style-type: none"> - 2019년 당시보다 현재는 딥테크 투자 환경이 개선됨 - 초기 투자 유치는 쉬워졌으나, 글로벌 경쟁을 위한 대규모 자금(수백억~천억 원) 조달은 여전히 어려움 - 예비창업파기지부터 텁스까지 단계적으로 정부 지원을 받았으나, 빠른 기술변화에 대응하기에는 부족

□□□□ 대표 이사 ○○○	
	<ul style="list-style-type: none"> - 초기에 대규모 자본을 투자받아 빠르게 기술 개발과 상용화를 진행하는 것이 성공 가능성을 높일 수 있음 <p><input type="checkbox"/> 제품 개발과 사업 전략</p> <ul style="list-style-type: none"> - 특정 고객이나 니치 마켓보다는 AI 가속기 표준 제품 형태로 개발 추구 (ASIC와 같은 ASSP로 다양한 산업군의 표준 제품 형태로 시장을 돌파하는 전략) - 기존 솔루션 대비 고객에게 가성비/전성비 측면의 베네핏 제공 목표 - 첫 번째 제품은 삼성 파운드리에서 양산 예정 → 두 번째는 TSMC 이용 예정 - 수요 연계보다는 기술 개발에 우선 집중, 다양한 수요자 의견을 종합해 시장 필요 제품 개발 <p><input type="checkbox"/> AI 반도체 생태계 발전을 위한 제언</p> <ul style="list-style-type: none"> - 기술력 평가를 위한 객관적 지표와 평가체계 필요 (선별적 지원의 도입을 위해서는 누구나가 납득할 수 있는 평가 지표 필요, 주요 5개사 10개사 모두가 참여하여 평가지표를 같이 제안할 수 있는 제안의 장이 필요) - 기술 평가에는 Business Model에 대한 부분도 필요 ('기술력의 높음'이 실제 기업의 반도체 이용에서 혜택을 제공받는다고 맹신할 수 없음) 수요 기업들이 평가에 참여하는 방안 필요 - 단계별로 차등화된 지원 필요 (아이디어 단계 20-30억, PoC 준비 단계 100억, MPW 횟수 등), 이를 위해서는 단계별 성과를 증빙하는 방안 또는 2-3년 주기의 성과 평가의 체계가 마련될 필요 - 상장 이후 우수 인력 유지와 추가 확보를 위한 중장기적 전략 필요 → 상장 이후에 투자를 받는 것이 어려운 부분이 존재하며, 기업의 규모가 커지는 경우 Skilled and Qualified talent 유치는 더욱 어려운 부분이 될 가능성성이 높후 - 팝리스와 디자인하우스가 생태계의 중심이며, 이들의 경쟁력 있는 인력 유지가 핵심 과제 <p>* 그래프코어가 소프트뱅크에 상대적으로 낮은 가격에 인수된 사례를 통해서 알 수 있듯이, 기술 중심의 접근에는 한계가 존재한다고 생각함. 기술만큼이나 시장에서의 제품 경쟁력과 산업과 전반적 기업의 수요를 충족시킬 수 있는 제품의 생산이 필요한 시점</p>
주요 내용	<p>□ 창업 배경 및 과정</p> <ul style="list-style-type: none"> - 글로벌 빅테크 기업에서 근무 경험 - 2021년 트랜스포머 기반 모델의 중요성을 인지하고 관련 프로세서 연구 시작하며, FPGA 기반으로 8개 멀티 가속기 서버 시스템 프로토타입 개발 - 2022년 말 ChatGPT 등장으로 시장성을 확인하고 2023년 1월 창업을 시작 <p>□ 교원창업의 장단점</p> <ul style="list-style-type: none"> - 장점: <ul style="list-style-type: none"> · 학교 연구실에서 프로토타입 개발 가능, 우수한 학생인력 활용 가능 · 학교의 창업 지원제도 활용(자문 제공 방식), 초기 자금난 해소(Angel 투자 유치)

□□□□ 대표 이사 ○○○	
	<ul style="list-style-type: none"> - 단점/어려움: <ul style="list-style-type: none"> · 비즈니스/회사 운영 관련 경험 부족 및 관련 인재 수급이 쉽지 않음 · 산업계 출신 인력과의 문화적 충돌이 존재함 (학교(Lab)와 기업 문화) · <u>소프트웨어</u> 인력 확보의 어려움 → 컴파일러 개발자
	<p><input type="checkbox"/> 비즈니스 모델과 차별점</p> <ul style="list-style-type: none"> - LLM에 특화된 AI 반도체 개발에 집중, 타기업이 HBM과 선단 공정 사용으로 인한 고비용이 드는데 비해, 본사는 LPDDR 기반으로 가격 경쟁력 확보 - 서버제품, 칩, IP 등 전 레이어 사업 전개, AMD 등 글로벌 기업과 협업 - 주요 타겟: 데이터센터, 금융, 공공기관 등
	<p><input type="checkbox"/> 정부 지원정책 관련 의견</p> <ul style="list-style-type: none"> - 긍정적 측면: 반도체 분야 정부 R&D 지원 풍부 (다양한 반도체 분야를 지원), 실증 사업 제공은 효과적 - 개선 필요 사항: <ul style="list-style-type: none"> · NRE(초기 개발비용) 직접 보조 필요, R&D 과제기간(4~5년)과 시장변화 속도 불일치, 팝리스 기업 대상 실질적 지원책 부족 * NRE²³³⁾는 제품 설계에 필요한 IP, EDA 라이선스 비용과 마스크(MASK) 비용이 포함된 것으로, 개발의 성공 유무와 상관없이 무조건 내야하는 돈을 의미), PoC(개념검증)을 통한 IP간의 신호 전달 여부 확인, 프로토 타입을 통한 시제품 생산 등
	<p><input type="checkbox"/> 향후 전망과 과제</p> <ul style="list-style-type: none"> - AI 반도체 업계 전반적으로 실적 검증 시기 도래하고 있으며, 대기업 주도 사업은 이미 구조조정 진행(사피온, 뉴블라 등) - 팝리스 스타트업 중심으로 재편 진행 중이나, 투자자들의 실질적 성과 요구 증가하고 있음 - 인력 확보와 양산 성공이 핵심 과제
주요 내용	□□□□ 대표 이사 ○○○
	<p><input type="checkbox"/> 반도체 용어 및 개념 정립</p> <ul style="list-style-type: none"> - 현재 업계를 바라보는 관점에서 메모리/비메모리 반도체의 구분이 부적절함 - 비메모리를 시스템 반도체로 통칭하는 것은 부정확한 정의이며, 실제로는 프로세서 포함 여부가 중요한 구분 기준 - 즉, AI 반도체의 정의도 재정립 필요: 단순히 AI 용도로 쓰이는 모든 반도체가 아님 - 딥테크 SOC(System on Chip)라는 새로운 개념 제시: 복잡한 프로세서가 포함된 고난도 기술의 반도체
	<p><input type="checkbox"/> IP 비즈니스 모델의 특성</p> <ul style="list-style-type: none"> - IP 벤더는 반도체 설계 자식재산권을 라이센싱하는 비즈니스

233) 키포스트(20.07.20), “팝리스의 가장 큰 고민, 개발비(NRE)를 줄이는 방법”
<https://www.kipost.net/news/articleView.html?idxno=204801> (검색일 24.11.26)

□□□□ 대표 이사 ○○○	
	<ul style="list-style-type: none"> - 신생 IP 업체의 시장 진입이 매우 어려움 (신뢰성 입증 필요), 즉시성, 신뢰성, 가격 메리트가 IP 비즈니스의 핵심 요소이기 때문 - 한국에서는 삼성 파운드리의 태동기에 일부 IP 벤더들이 성장 가능했음 - 현재는 새로운 IP 벤더의 등장이 거의 불가능한 상황 → SAFE와 같은 삼성 IP, 디자인하우스 등이 이미 구축이 되어있고, 새로운 파운드리의 태동은 자국 중심으로 개편될 가능성성이 있기 때문
<p><input type="checkbox"/> AI 반도체 시장 현황</p> <ul style="list-style-type: none"> - AI 반도체 시장이 응용 분야별로 파편화되어 있고, 대형 시장은 NVIDIA 등 거대기업이 장악 - 중소형 시장은 범용 프로세서로 대체되는 경향이 있어서, 실제 AI 반도체가 필요한 시장은 제한적 (노트북, 자동차 등) - 한국의 AI 반도체 정책 지원이 실효성이 낮음 (규모가 작고 실질적 도움 안 됨) 	
	<p><input type="checkbox"/> 인재 수급 문제</p> <ul style="list-style-type: none"> - 대기업의 폐쇄적 인사정책이 인재 순환을 저해하고 있음. 단적으로 퇴사 후 스타트업/중소기업 이직 후 다시 돌아오는 경우를 확인하기 어려움 - 인재들이 스타트업/중소기업에 대한 부정적 인식을 가지고 있음 - 실패에 관대하지 않은 사회 문화 및 대출자들의 대기업 선호 현상이 극대화 되어 있음
<p><input type="checkbox"/> 한국 반도체 산업의 문제점</p> <ul style="list-style-type: none"> - 메모리 반도체 중심의 왜곡된 산업구조 - 중국과의 기술격차 심화 (특히 시스템 반도체 분야)에 대한 인식 부재 - 정부 지원정책의 비효율성 (사업 당 낮은 지원 금액) - IP, EDA 등 기반 기술 부족 (해외기업에 대한 의존성이 높음)_ - 직접 보조금 등 적극적 지원 필요성 	
□□□□ 대표 이사 ○○○	
주요 내용	<p><input type="checkbox"/> 성장 경로 분석 및 AI 반도체 사업 성공 요인</p> <ul style="list-style-type: none"> - 전용 반도체 (ASIC) 시장 확대가 창업의 동기 - 대학원 설계 공부 후 컨설팅 업체에서 글로벌 반도체 기업 기술 및 사업 전략 관련 프로젝트 진행 - AI 부상 이전부터 전용 반도체 수요 증가를 예측, 애플, 테슬라, 아마존 등의 사례 존재 - 전용 반도체 시장 확대의 주요 원인으로 반도체 공정 기술 개발의 한계 지적 <p><input type="checkbox"/> 무어의 법칙 한계와 전용 반도체 부상</p> <ul style="list-style-type: none"> - 무어의 법칙, 즉 2년마다 반도체 성능이 두 배로 향상되는 현상이 30년 이상 지속되었으나, 최근 한계에 직면. - 소형화에도 성능 향상과 전력 효율 개선이 더뎌지고, 제조 비용이 급증하면서 기존 방식의 한계 도래. - 이에 따라 특정 운영체제 및 사용자 성향에 최적화된 전용 반도체 설계 필요성 증대 <p><input type="checkbox"/> 전용 반도체 활용한 제조 단가 절감 효과</p>

□□□□ 대표 이사 ○○○

- 전용 반도체는 대량 생산이 어려워 범용 반도체보다 비용이 높지만, 예를 들어 전자담배와 같은 특정 제품에 특화하면 제조 단기를 낮출 수 있음

- 전자담배의 경우, 기존 3개의 칩을 하나로 통합하여 제조 단기를 1달러에서 50센트까지 절감 가능
- 단가 절감으로 인해 제품 경쟁력 향상 및 기능 확장(기존의 칩에 AI, 음악 기능 추가 등)이 가능

□ 전용 반도체 개발의 어려움과 디자인 하우스의 역할 변화

- 반도체 개발에 전문학적인 비용이 소요되어 개발 비효율성이 심각해짐.

- 다양한 전용 반도체 개발 필요성 증대, 기존 팹리스의 한계 드러남.

- 디자인 하우스는 팹리스의 설계를 바탕으로 제작하는 역할에서 벗어나, 다양한 전용 반도체 설계 까지 담당해야 하는 상황 발생.

□ 디자인 하우스의 플랫폼화: 설계 재사용과 효율 증대

- 다수의 칩 설계에도 차산성 확보를 위해 설계 요소 재사용이 중요해짐.

- 팹리스에서 디자인 하우스로 설계 허브 포인트 이동으로 디자인 하우스의 플랫폼화 진행 예상.

- 반도체 제조의 플랫폼화(TSMC 사례)처럼 설계의 플랫폼화 필요성 인식으로 창업

□ AI 반도체 설계 플랫폼의 성장

- AI 분야의 급격한 성장으로 전용 반도체 수요 증가.

- 설계 플랫폼 전환을 통해 AI 반도체 설계의 자동화 및 재사용 용이성 증대.

- 매년 세 배씩 증가하는 수주를 통해 디자인 플랫폼의 중요성과 성장세 확인.

□ 디자인 하우스의 플랫폼 전환과 성장

- 초기에는 레이아웃 설계만 담당하던 디자인 하우스들이 이제는 설계부터 제조까지 고객을 위한 토탈 솔루션을 제공하는 방향으로 변화했다.

- TSMC와 같은 제조 플랫폼의 등장으로 다양한 칩 제작이 가능해졌고, 디자인 하우스들은 플랫폼 역할을 확대하며 성장, AI 칩 개발 등의 수요 증가로 반도체 회사뿐 아니라 세트 회사들도 칩 제작에 참여하게 되면서 디자인 하우스의 중요성이 커짐

□ 설계 분리와 한국 반도체 산업의 미래

- 미국과의 경쟁에서 불리한 상황에도 불구하고, 팹리스의 설계 가치가 독립적으로 분리되고 있음

- 이러한 설계 분리는 기존 팹리스 모델에서 벗어나 새로운 비즈니스 모델(ODM)을 가능하게 함

- 한국은 우수한 제조 인프라를 바탕으로 설계 분리 현상에서 유리한 위치를 점할 수 있으며, 삼성의 경험은 이를 뒷받침

□ 한국 반도체 산업의 새로운 기회

- 국내 제조 인프라와 풍부한 인력(설계, 소프트웨어) 기반 강조

- 팹리스 중심의 벤류 체인 변화가 한국 반도체 산업에 긍정적 영향 미칠 것이라는 전망

- 국내 전용 반도체 개발 수요 증가로 인한 국내 성장세 및 해외 진출 어려움에도 불구하고 국내 시장 성장 가능성 확인, 더불어 국내 디자인 플랫폼의 글로벌 확장 가능성 증대의 기회

- 미국과 중국 등 해외 고객과의 과제 수주를 통해 국내 팹리스 산업의 성장 가능성이 확인됨.

- 대만 TSMC처럼 국내 디자인 플랫폼이 전 세계 고객에게 설계 및 칩 공급을 제공하는 비즈니스 모델 제시.

□□□□ 대표 이사 ○○○	
	<ul style="list-style-type: none"> - 케이던스, 시놉시스 등 해외 IP 의존도 감소를 위한 국내 디자인 하우스의 플랫폼 전환 필요성 강조.
<input type="checkbox"/>	<p>IT 자원 관리 및 최신 고속 IP 활용 전략</p> <ul style="list-style-type: none"> - 최신 고속 IP(예: GDDR6 16Gb, 16Gb PCIE Gen5) 활용 시 설계, 검증, 소프트웨어 개발 역량 필요 - 단, IP 활용 시스템 구축의 어려움과 생태계 고려가 필요 - 목적에 맞는 IP 선정, 설정, 레이아웃 구성 및 시스템 소프트웨어 개발, 보드 성능 튜닝까지 필요. - 전 세계적으로 이러한 서비스를 제공할 수 있는 회사가 매우 적음 - IP 라이센스만으로 모든 문제가 해결되지 않으며, IP 활용 역량을 갖춘 고객 또는 협력업체가 필수적이기야 생태계 고려가 중요
<input type="checkbox"/>	<p>chipless 가치사슬도 주목할 필요</p> <ul style="list-style-type: none"> - IP 자원 관리의 측면과 파운드리까지 가는 과정에서의 POC 이후 다양한 공정에 대한 플랫폼화된 접근을 통하여 MPW에 이르는 과정의 지원이 필요 - 옥석 가리기에 대한 정확한 접근이 필요

국내 AI 반도체 스타트업 생태계는 높은 성장 잠재력과 함께 다양한 도전과제에 직면해 있다. 주요 기업들의 심층 인터뷰 분석 결과, 창업 초기와 배경에서 뚜렷한 특징이 나타났다. 인터뷰를 진행한 창업자뿐만 아니라 앞선 기업 분석에서 본 다수의 기업의 창업자는 반도체 설계나 동 분야의 전문가로서, 해외 선진기업 근무 경험 또는 학계 연구 경험을 보유하고 있었다. 이들은 시장의 니치를 발굴하여 차별화된 전략을 수립했으며, AI 반도체 시장 초기 진입의 중요성을 인식하고 있었다. 더불어, 인터뷰를 진행하지 못했던 리밸리온의 창업자도 GPU와 함께 사용될 수 있는 AI반도체(NPU)의 가능성을 발견하고 창업을 시작했으며, 딥엑스의 대표역시 Apple의 AP 수석엔지니어에서 AI 침의 다양한 활용 가능성을 기준으로 창업을 시작했음을 알 수 있다. 더불어, Intel, 삼성, IBM, MS 등 유수의 기업에서 시장의 변화를 선도적으로 접하거나 학계에서 동 분야의 연구를 진행하며 그 가능성을 확인하여 진입한 경우도 눈에 띈다.

비즈니스 모델 측면에서는 특정 응용분야 특화, 저전력·고효율 등 차별화된 가치 제안, IP 라이센싱과 팜리스 연계 등 다양한 접근 방식이 관찰되었다. 특히 수요처 확보 전략에서는 초기부터 특정 수요처에 종속되는 것을 경계하면서도, 프로토타입 개발을 통한 시장 검증을 중시하는 특징을 보였다. 대부분의 기업이 글로벌 시장 진출을

고려한 확장성 확보에 주력하고 있었다. ○○향의 기업들이 가질 수 있는 한계도 지적하는 동시에, 다양한 채널을 열어두고 수요처와의 긴밀한 협력의 중요성과 Business Model로 확장될 수 있는 기본요건인 제품의 초기 모델에 대한 중요성 역시 여러 번 강조되었다.

투자 측면에서 AI반도체 설계라는 하이테크 기술군의 특성으로 인한 투자자들의 이해도 부족 및 투자 검증 및 장기적 ROI 확보에 대한 불확실성 때문에 글로벌 투자를 유치하는데 어려움을 겪고 있으며, 이는 막대한 초기 비용, 우수인재 확보후 지급해야 하는 인건비, 글로벌 경쟁력 확보를 위한 지속가능한 연구개발을 위한 투자 유치에 장애 요인으로 작용한다. 다양한 정책을 통해 국내 연구개발 지원 사업은 다양해졌으나, 실증까지 진행하기에는 부족한 과제당 연구비, 기술 개발의 주기가 빠르고 제품 수명주기가 짧은 산업 특성으로 인해 과제 기획 단계에서의 예상 결과물과 연구개발 이후 시장에서 필요로 하는 기술의 갭이 생기는 단점도 존재한다. 해결책으로 민간과 공공 부문의 최종 수요자 (기업, 정부 부처 또는 공공 데이터 센터), 주요 업체들이 모두 참여하여 기술개발 가이드라인을 공동개발하고 주기적으로 업데이트 하는 방안이 제시 되었다. 이는 지속적인 투자에 대하여 정보 비대칭성을 해소할 수 있는 역할을 할 것으로 기대되며, 동 산업 분야 기업의 참여를 통해 시장에서 주목하는 요소도 반영할 수 있는 장점이 있다.

인재 확보 및 유지와 관련해서는 전문 인력의 부족, 스타트업/중소기업과 대기업 간의 인재 확보 경쟁, 소프트웨어와 하드웨어 통합 인재 부족 등이 주요 당면과제로 지적되었다. 스펙옵션과 같은 차별화된 보상, 기술 중심의 기업 문화 조성, 성장 가능성의 제시등을 통한 인재 유입의 한계점 또한 존재한다. 동 분야의 혁신 인재의 유입과 자유로운 기업간 이동에 대한 보장역시 주요한 요소로 장기적으로 혁신 생태계의 기반 강화라는 측면에서 필수적 요소이기 때문이다. AI 반도체 수요가 자체적으로 AI 반도체를 생산하는 Magnificent 7과 같은 빅테크 기업 뿐 아니라 레거시 산업으로 불리우는 철강 산업의 포스코-딥엑스 협력 등 다양한 분야로 이어지는 산업 동향을 보면, 인력의 이동, 유치, 확보 등 인재와 관련된 다양한 이슈들은 하드웨어 개발자의 이동이 아닌, 소프트웨어 전문가, 동 분야 사업의 기획과 운영인 Business자체에 강

점이 있는 인적 네트워크 전문가 등 다양한 인력의 이동이 수반 돼야 함을 시사한다.

기술 부문과 관련하여 투자 측면에서 논의된 높은 개발 비용과 실증의 한계와 더불어 시장의 빠른 수요 변화에 대응해야하는 도전과제가 제시되었다. 단, 기업별로 가지는 설계 기술과 같은 하드웨어기술은 시장에 참여하기 위한 기초 요건이며, 지속적으로 잠재 수요자의 니즈를 반영하여 지속가능한 개발과 소통은 사업에서 주요한 요소로 작용한다는 지적도 존재한다.

AI 반도체 생태계의 발전을 위해서는 정부의 실효성 있는 지원은 필수적 요소이며, 연구개발 외에도 직접 보조금의 확대, 실증 사업 확대, 및 투자 검증 체계 구축, 해외 진출 지원 등 기술개발 이외의 고려요소 또한 존재한다. 극초기 투자 측면에서 정부 지원사업이 스타트업에 주는 장점이 존재하지만, 장기적으로 PoC, MPW, Single Run, 양산의 단계와 비교하면 그 금액이 상대적으로 작다고 볼 수 있다. 이에, 극초기 이후 선별적 지원을 강화할 수 있는 가이드라인 구축이 필요하다.

기업들은 특화 분야 집중과 기술 경쟁력 강화, 시장 니즈 대응을 통한 차별화 전략을 추구하면서, 안정적 수요처 확보, 투자 유치 다변화, 인재 확보/유지 체계 구축 등을 통해 지속가능한 성장 기반을 구축해야 한다는 점에 모두 공감하였다. 더불어, 선순환적인 생태계 발전을 위해서는 단순한 R&D 지원을 넘어 실질적인 시장 창출과 성장 지원이 필요하며, 글로벌 경쟁력 확보를 위한 체계적인 접근이 요구되는 점에 공감했다.

기업의 인터뷰를 기반으로 본 AI 반도체 기업의 성장과 생존을 위한 핵심 요소는 다양하다. 우선적으로 각 기업들이 명확한 시장 포지셔닝과 차별화 전략을 가져야 한다. NVIDIA가 시장을 지배하고 있는 범용 AI 가속기가 아닌 특정 응용 분야에 특화된 접근이 필요하며, 시장의 실질적 니즈를 충족시키면서도 기존 솔루션과의 차별화 포인트를 명확히 할 필요가 있다. 특히 전력효율, 성능, 가격, 전성비 (성능/소비 전력) 등 핵심 지표에서의 우위를 확보하고 독자적 기술 경쟁력을 구축하는 것이 중요하다. 특히, 효과적인 기술 개발 및 상용화 전략을 통해, 고객 니즈에 기반한 개발 방향 설정, 빠른 시장 대응 능력, 기술과 시장 간의 균형 유지를 해야 한다. 예를 들어, 단계적 제품 출시 전략을 통해 시장의 피드백을 지속적으로 반영하고, 파트너십을 구축하여 상용화를 효과적으로 추진할 필요가 있다. 이는 안정적인 수요처의 확보에 영향을 미

칠 수 있는 요소이다.

두번째로 칩 설계 및 제작에 드는 막대한 비용, 우수 인재 확보를 위한 자금, 지속적인 R&D를 위한 운영 자금 등 대규모 투자를 위한 자금 확보가 필요하다. 정부 지원 사업의 활용, 전략적 투자자 확보, 글로벌 투자 유치 등 자금 조달 전략을 다각화해야 한다. 현재 국내 AI 반도체 기업들의 경우, 수백억 원 단위의 투자를 받더라도 글로벌 경쟁을 위해서는 부족하다는 의견이 존재하기 때문에(25년 양산 이후 시장에서의 경쟁력을 확인할 필요), 본격적 양산이후 기업의 성공/실패 사례에 대한 추적 조사와 분석을 통해 장기적으로 시장에서의 경쟁력을 강화할 수 있는 방향으로 정책 지원 또한 변화해야 한다.

언급된 핵심 요소들은 독립적으로 작용하는 것이 아니라 상호 연계되어 시너지를 창출해야 한다. 예를 들어, 우수한 기술력은 투자유치를 쉽게 하고, 이는 다시 우수 인재 확보로 이어져 기술력을 더욱 강화하는 선순환 구조를 만들 수 있다. 또한, 초기 레퍼런스 구축은 추가 고객 확보로 이어지고, 이는 매출 증가로 이어지는 선순환을 만들 수 있다. 특히 AI 반도체 산업의 특성상 자금조달 리스크, 기술개발 리스크, 시장 진입 리스크, 인재이탈 리스크 등 다양한 리스크가 존재하므로, 이러한 리스크들을 효과적으로 관리하는 것도 중요하다. 급변하는 시장 환경에서 지속 가능한 성장을 위해서는 각 요소의 지속적인 강화와 함께 환경 변화에 대한 유연한 대응이 필수적이다. 결론적으로 AI 반도체 기업이 성공적으로 성장하고 생존하기 위해서는 이러한 다섯 가지 핵심 요소들이 균형이 있게 갖춰져야 하며, 특히 초기 단계에서는 명확한 시장 포지셔닝과 차별화된 기술력을 바탕으로 안정적인 수요처를 확보하는 것이 중요하다. 이를 위한 충분한 자금과 우수 인재의 확보가 뒷받침되어야 하며, 각 요소 간의 선순환 구조를 구축하고 잠재적 리스크를 효과적으로 관리해야 한다.

3. 시사점

분석결과 AI 반도체 산업은 크게 비즈니스 모델, 투자 환경, 기술적 역량, 인재 확보라는 생태계 내 네 가지 핵심 부문에서 다양한 도전 과제에 직면해 있다. 먼저 비즈니스 모델 측면에서는 PoC(Proof of Concept) 및 대규모 실증 역량의 부족이 가장 큰 문제로 여겨진다. 특히 데이터센터용 서버 AI 반도체는 막대한 개발 및 생산 비용의 문제로 스타트업은 상대적으로 글로벌 선도기업에 비해 경쟁력 있는 제품 개발에 한계가 있다. 더욱이 B2B 시장 특성상 기존 거래처와의 관계 때문에 신생기업의 시장 진입이 매우 어려운데다, 다수의 국내 스타트업이 포진하고 있는 엣지(Edge)형 AI 반도체 시장의 경우 과당 경쟁 상황에서 차별화된 경쟁우위 확보가 쉽지 않다. 더구나 제조 인프라와 수요 기업과의 협력 측면에서는 중국, 대만 기업들이 한국기업들에 비해 상대적으로 비교우위를 점하고 있다.

투자 환경 측면에서는 AI 반도체 개발에 필수적인 대규모 자금 조달에서 구조적 한계가 존재한다. 특히 10nm 이하 선단 공정을 활용하여 칩을 개발하는 경우, IP 확보, DSP(디자인하우스), 패키징 및 테스트 등에 드는 비용이 400~500억 원 규모에 달하지만 현재 국내 정책펀드는 대부분 10~20억 원 수준의 초기 VC 투자에 머물러 있다. 한편 국내 M&A 시장의 미성숙으로 인해 투자회수(EXIT) 전략이 IPO에 편중되어 있어, 창업자 지분율의 확보와 대규모 개발비 조달에 한계가 존재한다. 특히 국내 투자자들의 투자위험 회피 등의 보수적인 투자 성향으로 인해 매출 발생 이전 단계에 필요한 대규모 투자유치에 어려움이 크다. 결국, 국내에서는 아직 본격적인 매출액을 발생 시킨 기업이 없다 보니 후속 투자가 더욱 어려워지는 악순환이 지속되고 있다.

기술적 측면에서는 SW 경쟁력 부족과 시스템 레벨의 호환성 문제가 주로 제기된다. SoC, NPU 등 AI 반도체는 메모리 반도체와 달리 프로그래밍 언어에 대한 깊은 이해가 필요하다. 또한 HW-SW 인터페이스, API, 애플리케이션 등을 종합적으로 고려한 설계가 요구되나, SDK 등 칩 SW 경쟁력이 글로벌 선도 기업 대비 현저히 부족하여 테스트 및 성능 검증 과정에서의 비효율이 발생하고 있다.

인재 측면에서는 AI 반도체 산업의 특성상 하드웨어와 소프트웨어를 아우르는 융합형 인재가 필수적이나, 국내 인재 풀이 매우 제한적이다. 특히 AI 알고리즘에 대한

이해와 반도체 설계 능력을 동시에 갖춘 고급인력이 부족한 상황이다. 또한, 글로벌 기업들의 공격적인 인재 영입으로 인해 국내 기업들의 인재 확보가 더욱 어려워지고 있으며, 스타트업의 경우 경쟁력 있는 보상 패키지 제공에 한계가 있어 우수 인재 영입과 유지에 어려움을 겪고 있다.

위와 같은 도전과제에 직면해 있는 AI 반도체 생태계 내에서의 주요 기업들의 성장경로 역시 기존의 아날로그 IP기업, 팹리스 기업과는 사뭇 다른 양상을 보이고 있다. 기존의 초기, 성장기, 성숙기를 거쳐서 MPW(Multi-Project Wafer),²³⁴⁾ Single-run,²³⁵⁾ 및 전략 파운드리²³⁶⁾(Foundry)를 통해 주력 제품인 전력반도체, 디스플레이 구동칩, 등 다양한 산업에 필요한 반도체를 생산할 수 있다 (성숙 공정에는 TSMC와 삼성 이외에도 DB하이텍, SK키파운드리 등 기업 존재).

하지만, AI 산업의 부상과 생성형 AI의 등장으로 인해 AI 컴퓨팅 파워에 대한 산업계 요구가 증가하고 있는 상황에서 기존의 성장경로를 따르기 어려운 부분이 발생한다. GPU, NPU등의 AI 산업에 필요한 저전력의 고도화된 설계의 칩들이 필요해지면서, 설계가 복잡해지는 동시에 설계된 칩의 생산에서 파운드리가 선단공정 14nm 이하의 10, 7, 3, 2nm 공정을 필수적으로 사용해야 한다. 더불어, 앞서 AI반도체의 특징에서 언급한 바와 같이, 칩 자체의 아키텍처(Architecture)가 고도화되기 때문에, 칩 설계 과정에서 고가의 케이던스, 시놉시스, 지멘스의 IP, EDA를 사용해야 한다. 즉, AI반도체 초기 투자 비용은 500~1000억 원内外로 소요된다. 이는 앞서 언급한 IP, EDA tool의 라이센스 비용이 대다수를 차지하며, 설계를 위한 다수의 전문 인력이 필요하다.

이는 AI반도체의 특징으로 언급한 software stack의 준비라는 측면과 부합한다. 우선 하드웨어적으로 설계가 고난이도가 되며, 만들어진 칩을 활용하는 측면의 소프트웨어도 고려해야 하기 때문이다. 데이터센터용 학습 목적의 반도체 시장은 NVIDIA

234) 처음 설계하고 생산하는 칩을 시스템레벨에서 sample과 같이 그냥 한번 테스트용으로 뽑아 보는게 바로 MPW, 파운드리 업체가 한 웨이퍼에 여러 고객사의 시제품 반도체를 만들어주는 것

235) 팹리스 업체는 MPW를 활용해 칩의 기술적 성능, 특성, 오류를 수정하는 과정을 거치고, 싱글런 과정을 거침. 이 과정은 MPW와 달리 한 웨이퍼에 하나의 반도체만을 생산하는 것. 공정마다 차이가 있으나 MPW 비용의 6배 이상 필요

236) legacy(성숙) 반도체 공정 활용으로 28nm 이상, 40 나노 등

가 95% 이상 글로벌 마켓을 점유하고 있는 상황에서, 대규모 추론의 영역에 참여하는 서버용 AI 시스템 온 칩 시장에서 반도체의 성능을 고도화하기 위한 설계 부문의 고도화와 이를 활용하는 수요처의 칩 활용에 대한 소프트웨어의 제공이라는 두 가지 측면의 도전과제가 있다. 성장경로라는 측면에서 AI반도체는 앞선 도전과제의 문제점을 포함하여, 창업단계에서의 충분한 하드웨어 엔지니어와 소프트웨어엔지니어 즉, 동 분야의 설계가 가능한 전문인력이 필요하다. 이는 국내 주요 스타트업의 창업자의 배경을 보면 더욱 쉽게 알 수 있는데, 삼성, 애플, AMD, NVIDIA, MS 등 주요 기업에서 경력을 쌓고 창업에 도전하고 있다. 창업 시작 전부터 최초 인력 구성에 대한 준비, 명확한 틈새시장에 대한 파악을 통해 창업을 통한 AI반도체 설계 산업에 참여하지만, 시제품의 생산에 필요한 IP, EDA 부분의 투자는 필수적 요소로 작용한다.

문제점의 해결을 위하여 각 섹션별로 고려해야 할 우선 사항은 다음과 같다. 산업 생태계 구축 측면에서는 극 초기 기술 발굴부터 대규모 실증까지 이어지는 원스톱 지원 체계 구축이 필요하며, 대기업-스타트업 간 협력을 촉진하기 위한 제도적 인센티브 방안을 도입해야 한다. 특히, 글로벌 비테크 기업들과의 전략적 협력 관계 구축을 위한 정부 차원의 지원이 필요하다. 2020년 중소벤처기업부와 ARM간의 협력사업인 ‘플렉시블 액세스 스타트업(Flexible Access Startups)’을 통해 ARM의 우수한 인프라를 공급받을 수 있었다²³⁷⁾. 초기 레퍼런스 구축을 위한 시연회 혹은 과기공동위 등 주요 국가 간 플랫폼에서의 신흥기술 부문 협력의 장 마련 등의 기회 제공을 고려 할 수 있다.

기술 경쟁력 강화 측면에서는 SW 개발 인력 확보 및 육성을 위한 체계적인 R&D 프로그램 수립이 필요하며, 시스템 레벨 호환성 검증을 위한 공용 인프라 구축이 요구 된다. 더불어 정부 차원에서의 공개 가능한 형태의 글로벌 표준 및 최신 기술 트렌드에 대한 지속적인 모니터링과 대응이 필수적 요소이다. 시장 전략 측면에서 기업들은 응용 분야별 특화된 솔루션 개발을 통한 수요-연계 측면에서 틈새시장 공략을 진행해야 한다. 특히, 수요기업과의 공동 개발 및 검증을 통한 시장 진입을 전략적으로 진행

237) 중소벤처기업부(2020.09.17.), “자상한기업 ‘암(Arm)’, 시스템반도체 혁신기업 본격 지원”<https://www.mss.go.kr/site/smbs/ex/bbs/View.do?cblidx=86&bclidx=1021592&parentSeq=1021592> (검색일 24년 12월 19일)

할 필요가 있다. 기술적으로 우위가 있더라도, 시장에서의 경쟁력은 수요 기업과의 연계를 통한 매출의 극대화이며, 시장점유율 확대를 통한 상품 다각화 등의 전략을 도모 할 수 있기 때문이다.

인재 육성 및 확보 측면에서는 산학연 협력을 통한 융합형 인재 양성 프로그램 확대 및 산업에서 즉시 활용이 가능한 인재를 교육할 수 있는 프로그램이 필요하다. 특히, 실무 중심(칩 제작 경험, 설계 경험, 설계-제작까지 경험 등)의 전문인력 양성을 위한 교육 프로그램 개발이 시급한데, 이를 위하여 기업-대학 간 산학협력 프로그램 강화를 통한 실무형 인재 육성 방안도 고려할 필요가 있다.

투자 생태계 혁신 측면에서는 AI 반도체 특화 대규모 프로젝트 펀드(400~500억 원 규모) 조성이 시급하며, PE의 적극적 참여를 유도하기 위한 리스크 분담 메커니즘 도입이 필요하다. 특히, 장기 투자가 가능한 인내 자본 육성을 위한 제도적 기반 마련이 요구된다. 정책적 지원 측면에서는 칩의 설계 이후 실증 인프라 구축을 위한 공공 투자 확대가 필요하며, 수요 창출을 위한 공공조달 제도 개선을 고려해야 한다. 글로벌 시장 진출을 위한 마케팅 및 네트워크 구축 지원이 필요하지만, 이는 기업이 구축 한 네트워크를 확장하는 측면으로 접근할 필요가 있다.

이러한 분석을 통해 볼 때, AI 반도체 산업의 성공적 발전을 위해서는 기술, 자본, 시장, 인재라는 네 가지 요소의 유기적 결합이 필수적이며, 이를 위한 정부, 기업, 투자자, 교육기관의 협력적 생태계 구축이 시급한 과제라 할 수 있다. 특히 인재 육성 및 확보 측면에서는 장기적인 관점의 전략적 접근이 필요하며, 글로벌 경쟁력을 갖춘 인재 풀 형성을 위한 체계적인 노력이 요구된다.

| 제4장 | 결론

제1절 결론

1. 시사점

가. 인공지능 반도체 기술의 발전 전망

NVIDIA는 데이터센터 규모의 인공지능 학습에 필요한 하드웨어 시장을 90% 이상 독점하고 있다. 하지만, 국내외 다양한 그룹에서 거대 언어 모델을 목표로 개발되고 있는 인공지능 반도체들은 그 하드웨어 성능에서 NVIDIA GPU를 맹추격하고 있다. 그뿐만 아니라, 후발 주자들은 다각도의 소프트웨어 개발 환경을 지원함으로써 NVIDIA GPU 및 CUDA에 락인된 인공지능 개발자들을 끌어들이고자 노력하고 있다. 하지만, 이미 자리 잡은 현시점의 거대 인공지능 모델을 학습하는 과정에서 인공지능 개발자가 NVIDIA GPU가 아닌 다른 인공지능 반도체를 선택하는 것은 흔치 않을 것으로 보인다. 그럼에도 불구하고, 다양하게 개발되고 있는 인공지능 반도체에는 아래와 같은 기회가 있을 것으로 전망된다.

- ① (자체 데이터센터 또는 클라우드 구축) NVIDIA의 시장 파이를 뺏어, 판매를 목적으로 하는 것이 아닌, 자체 거대 인공지능 학습을 위한 인공지능 데이터센터에 한정적으로 사용하기 위해 칩을 개발할 수 있다. 일반적으로는 판매 목적이 아닌 초고성능의 반도체를 자체 설계하는 것은 경제성이 맞지 않는 경우가 많다. 하지만, 거대 인공지능 학습을 위한 인공지능 반도체는 매우 높은 가격에 공급되고 있어, 자체 개발로 해당 비용을 절감하는 것이 유리할 수 있다. 또한, 미국-중국 간 NVIDIA GPU 수출 제한의 예시에서 볼 수 있듯이, 인공지능 반도체는 일종의 전략 물자로 판단될 수 있기 때문에, 자체 설계 능력을 갖추는 것이 필요할 수 있다. Google은 TPU를 지속해서 개발해 온 덕분에 NVIDIA의 GPU가 아닌

자체 설계 TPU를 활용해 Gemini의 학습을 수행할 수 있었다. 또한, Meta, Microsoft 등, 자체 데이터센터를 지니고 있는 기업에서 MTIA, MAIA100 등 의 인공지능 반도체를 개발하는 목적으로 비슷할 것으로 예상된다. 대한민국 역시 NVIDIA 공급망에 대한 의존을 벗어나 자체 생산할 수 있는 반도체로 거대 인공 지능을 학습할 수 있는 환경이 필요해질 수 있다.

- ② (인공지능 프레임워크의 발전) 초기 인공지능 개발자는 CUDA를 직접 활용하는 프로그래밍 능력이 필요했다. 하지만 최근에는 PyTorch, Tensorflow 등의 인공지능 프레임워크가 제공하는 기능과 편의성이 고도화됨에 따라, 인공지능 모델을 설계하는 연구자가 CUDA를 직접 사용하지 않아도 프레임워크에서 제공하는 기능만으로도 NVIDIA GPU의 사용이 용이하다. 후발 주자의 SDK 및 소프트웨어 스택이 해당 프레임워크와 높은 호환성만 보장해 준다면, 프레임워크 안에서 기존 인공지능 개발자들의 CUDA 라인을 깨고 새로운 인공지능 반도체의 사용으로 끌어들일 수 있다.
- ③ (새로운 인공지능 모델의 등장) AlexNet을 시작으로 CNN 기반 인공지능이 널리 쓰이고, 트랜스포머 모델을 기점으로 거대 언어 모델에 대한 학습을 지원하는 반도체가 필요해졌듯이, 완전히 새로운 인공지능 모델이 등장할 수 있다. 특히, 얀 르쿤 등의 일부 인공지능 연구자들은 현시점의 트랜스포머 구조로는 인공 일반 지능에 도달할 수 없다고 주장한다.²³⁸⁾ 새롭게 등장할 인공지능 모델에 최적화된 하드웨어 구조 및 소프트웨어 스택을 빠르게 선점해 NVIDIA의 독점 시장을 나눠 가질 기회가 있을 수 있다.

국내외에서 인공지능 반도체가 다양하게 개발되고 있지만, 아직은 NVIDIA의 시장 점유율에 큰 타격을 주고 있지 못하다. 그럼에도 불구하고, 앞서 소개된 기회들을 살리기 위해서 관련된 연구는 계속되어야 한다. 대한민국 역시 ChatGPT와 같이 해외에서 학습된 인공지능 모델에 계속해서 의존할 수는 없으며, 자체 거대 인공지능을 학습하기 위해서는 대규모의 인공지능 데이터센터를 구축할 필요가 있다. 그런 상황에서

238) AI타임즈(2024.4.10.). 메타 “AGI 도약 방법 발견...‘트랜스포머’와 다른 아키텍처 개발 중”. AI타임즈, <https://www.aitimes.com/news/articleView.html?idxno=158694> (검색일: 2024.10.25.)

고비용의 NVIDIA GPU에 의존하지 않고 국내에서 개발된 인공지능 반도체를 활용할 수 있는 가능성이 있다.

나. AI 반도체 생태계 전반의 경쟁력 진단

AI 반도체 산업의 특징은 동 산업이 글로벌 가치사슬 (Global Value Chain)에 따라 전문 분업화되어있다는 점이다. 본 연구에서 집중하고 있는 설계 영역을 제외한 생산과 패키징/테스트의 경우는 엄청난 설비투자와 경험이 필요하므로 스타트업으로서 이 분야에 들어가는 것은 현실적으로 어렵다. 따라서 많은 AI 반도체 스타트업이 상대적으로 적은 비용과 전문성으로 시작할 수 있는 펩리스²³⁹⁾ 부분에 몰려 있다. NVIDIA와 같은 기업이 GPU 부분에서 글로벌 독점의 위치에 있으나, 목적(edge 向에 추론 및 학습 목적)에 따른 다품종 소량생산 (분야별로 특화된 AI 반도체²⁴⁰⁾, ASIC, ASSP 등) 영역에서는 여전히 스타트업 기업에게 틈새(Niche) 시장이 존재한다. 이를 산업/생태계의 발전단계 및 혁신의 패턴을 나타내는 S 곡선으로 비유해보면, 틈새(혹은 하부)시장마다 전반적인 GPU가 가지는 곡선과 별개로 산업 특화 AI반도체의 S 곡선이 존재할 수 있다는 가정을 할 수 있다. 이는 스타트업으로서 연계 수요 산업에 특화된 기회요인이 될 수 있다.

생태계 전체 기준으로, 펩리스 산업으로 통칭하더라도 AI 반도체 설계 공정 안에는 IP, EDA, 디자인하우스 등 전문적인 공정들로 나누어져 있다. 스타트업의 특성을 살펴보았을 때 이 모든 설계 프로세스에서 경쟁력을 가지고 시작하기는 어렵다. 따라서, 펩리스 스타트업에 대한 지원 방안 및 육성 방안을 고려할 때, 세부 공정을 구성하는 전문 기업을 육성할 필요가 있다. EDA tool이나 IP의 경우, 설계 공정에서 해외 주요 기업에 대한 종속도가 높다. 단순히 AI 반도체 설계 기업뿐이 아닌, AI 반도체 설계 부문의 하위 생태계 기업의 경쟁력을 제고할 수 있는 정책적 방안도 고려해야 한다.

산업을 막론하고 스타트업의 성장에 있어서 투자는 아주 중요한 역할을 한다. VC 들은 단순히 재무적인 측면에서의 투자뿐만 아니라, 기업의 성장을 위한 교류 혹은

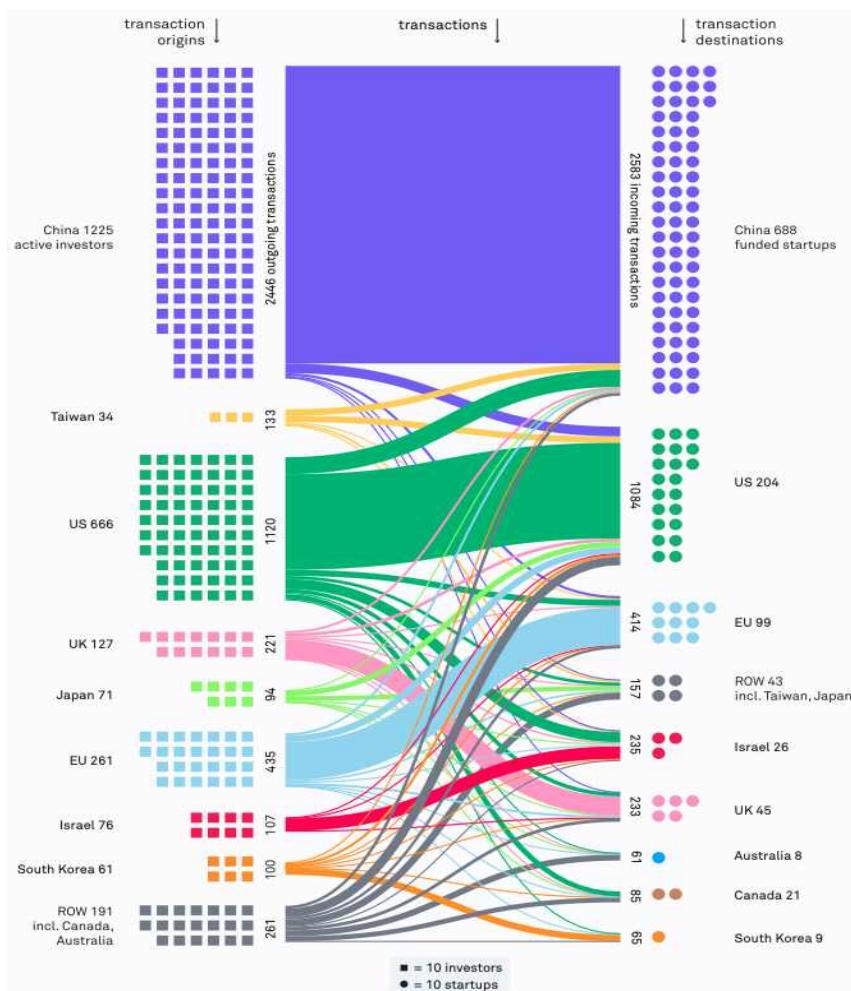
239) 여기서 펩리스는 IP에서부터 디자인 하우스 기능까지의 전체 세계 부분에 해당되는 비즈니스 모델을 의미

240) Application Specific Integrated Circuit, Application Specific Standard Product

지식 연계 등을 지원하기도 한다. 따라서 투자의 형태를 살펴보면, 각국 AI 반도체의 생태계의 특징을 살펴볼 수 있다. 2023년 4월 Hess, Denkena, Kleinhans, Maham 이 발간한 “Who is funding the chips of the future? Analysis of global semiconductor startup funding activities”의 2020~2022까지의 투자 데이터를 살펴보면 다음과 같은 특징을 알 수 있다.

1. 중국은 반도체 산업에 투자하는 투자자 수 (1225 개사), 투자받은 스타트업 기업 수 (688 개사), 그리고 투자 횟수 (822)건으로 2위인 미국에 비해서 3배에 약간 못 미치는 정도의 정량적 활동을 보여주고 있다.
2. 국가 랭킹으로는 중국, 미국, EU, UK, Israel 순의 활동을 보여주고 있으며, 한국은 스타트업 수로는 8위, 투자자 수로는 7위를 보여주고 있다. 투자자 대비 투자받은 스타트업 수 (Funded startup / active investors)의 자료를 살펴보면, 중국 56%, EU 37%, 영국 35%, 이스라엘 34%, 미국 30% 인데 반해 한국은 약 15%에 그치고 있어 관련 분야 투자자 수 대비 스타트업 수가 다른 선진국에 비해서 훨씬 떨어지는 특징을 보이고 있다.
3. 흥미로운 부분은, 국가 간의 투자가 어떻게 일어나고 있느냐이다. 과거 외국인비용(Liability of foreignness) 등의 문제로 인해 국가 간 투자가 미미했다면, 점차 세계화와 세계 시장의 통합 등으로 인해서 국가 간 VC 투자(Cross-border VC investment)가 늘어나고 있다. 특히, AI 반도체와 같이 일부 기업이 독점적 기술을 가지고 있고, 규모의 경제를 실현할 수 있는 효율적 생산 공정을 가진 기업과 연대하여 글로벌 시장의 점유율을 높이는 기업이 드물어서, 국가 간 VC 투자는 타 산업과 비교하면 더 활발하다고 할 수 있다. 같은 보고서의 데이터에 따르면 [그림 4-1]과 같이 투자의 흐름을 살펴볼 수 있는데, incoming / outgoing 비율을 따져보면, 중국(1.06), 미국(0.97), 영국(1.06), 이스라엘(2.20)으로 투자의 유입이 유출과 비슷하거나 크다는 것을 알 수 있다. (국내 투자 및 해외 투자 포함). 반면에 EU(0.54), 한국(0.65)로 타 선진국가에 비해 투자의 유출이 훨씬 더 큼을 알 수 있다.

[그림 4-1] 반도체 분야 글로벌 투자 현황



자료: Hess et al., (2023)

기업의 글로벌 진출은 수요처 발굴이라는 측면에서 중요한 부분이다. 하지만, 투자의 유치라는 관점에서, 기업의 혁신적인 기술로 만들어낸 시제품(완제품)을 가지고 시장의 수요를 창출해야 하며, 이에 대한 시장의 반응에서 초기/후속 투자의 연계를 노릴 수 있다. 이를 위하여, 칩의 POC 및 MPW와 같은 칩 설계의 검증과 시제품 제작의 지원이 향후 이를 기반으로 한 글로벌 판로 개척의 지원이 필요한 시점이다.

2. 정책 제언

가. 실증 및 수요처 발굴의 문제

AI 반도체의 특성상 특정 목적에 최적화되도록 설계된 반도체로써 그 자체로는 소비자의 효용을 얻지 못하는 특징이 있다. 그래서 이를 활용할 수 있는 기업과의 거래가 주요한 business model인 B2B가 주요 사업 모델로 작용한다. 더불어, AI 반도체를 활용하는 기업은 두 가지의 필수적인 부분이 필요한데, 첫 번째는 AI 반도체가 학습하고 처리할 데이터(Big data)의 확보 문제이다. AI 반도체는 재료가 되는 데이터를 바탕으로 각 파라미터를 설정하여 연산의 최적화가 목적이기 때문에, 파라미터 설정을 위해 기반이 되는 데이터의 확보가 필수적이다. 또한, 원 데이터를 바탕으로 파라미터를 설정하였다 하더라도 이를 활용하여 제품화 혹은 서비스화 할 수 있는 수요처가 없다면 AI 반도체 자체는 큰 의미가 없게 된다. 즉, 기술적 차원에서의 혁신성이 존재하더라도, 동 기술을 활용할 수 있는 기업과 시장이 없는 경우에는 기술의 혁신성을 통한 시장의 리더가 되기는 어렵다는 말이다. 따라서 설계 기술만으로 특정 분야의 AI 반도체 하부 시장이 성장할 수 없다. 이는 Moore가 주장한 The chasm 측면으로도 살펴볼 수 있는데, 극초기 수용자 집단(Innovators)을 대상으로 하기에는 해당 시장의 성장을 꾀할 수 없다. 그만큼 수요처 확보가 중요하다는 것이다. 하지만, 네트워크도 경험도 부족한 스타트업이 이러한 대규모 수요처를 찾기가 쉽지 않다.

미국의 정부 연구기관(예, NASA)의 경우 연구개발 뿐만 아니라 수요처로써의 역할도 크다고 할 수 있다. AI 반도체와 같이 딥테크 산업의 경우 선도적인 기술 개발을 하는 정부 출연연구기관 등에서 초기 수요처로써의 역할 및 투자를 진행하는 것도 하나의 방안이 될 것으로 보인다. 실제로 미국의 AI 반도체 분야 스타트업인 Mentium Technology²⁴¹⁾의 경우 현재 Seed 단계인데, 2017년, 2022년 두 번에 걸쳐 Pre-seed 투자로 NASA로부터 약 \$6,100,000의 투자를 받았다. 더불어, GPU시장에서 선도적인 역할을 하는 NVIDIA 역시 2010년 DARPA로부터 \$25,000,000의

241) Crunchbase, Mentium Technologies 기업 소개 https://www.crunchbase.com/discover/principal.investors/field/organizations/num_investors/mentium

Grant, 2012년 \$20,000,000 Grant, 2023년 ARPA-E로 부터 \$5,000,000의 GRANT를 받은 전례가 있다.

DARPA의 GRANT를 받은 사업을 살펴보면, 2023년에는 COOLERCHIPS²⁴²⁾ program에 참여하여 데이터 센터의 발열에 대한 연구 과제를 Boyd Corporation (cold plate technology), Durbin Group (pumping system), Honeywell (fluid selection), Vertiv Corporation (heat rejection technology), Binghamton University, Villanova University, Sandia National Laboratory과 같이 진행하였다. 이는 콘소시움 형태의 정부주도 프로젝트에 산, 학, 연이 모두 참여하여 향후 AI산업 성장에 따른 데이터 센터의 증가에 따른 에너지 및 발열에 대한 실증 연구를 하는 것으로서, 당면한 주요한 문제에 대한 연구를 진행하였다는 점에서 의의가 있다. Kcloud와 같은 데이터센터의 실증과 같은 과제도 유의미한 시도이지만, 장기적으로 부상하는 AI산업에 대한 문제점을 해결하는 R&D에 국내 AI칩 설계 기업의 참여를 유도할 필요가 있다. 2010년의 연구²⁴³⁾ 역시, Cray Inc., Oak Ridge National Laboratory, 미국내 주요 6개 대학과 함께 Ubiquitous High Performance Computing (UHPC) 프로젝트를 진행하였다. 이는 exascale의 슈퍼컴퓨터를 위한 GPU기술 개발이 목적으로 기존의 슈퍼컴퓨터보다 1000배 더 강력한 컴퓨터를 만드는 칩을 실증하는 연구로, 기업의 성장 및 사업 모델과 국가 정책의 니즈가 부합하는 사례로, 향후 국내에서 AI반도체 부문의 실증 외에 수요를 창출하는 측면에서 선별된 팜리스 스타트업과의 연구 부문의 콘소시움 구성도 좋은 정책 대안이 될 수 있다.

수요처와 더불어 중요한 것은 제품 설계를 했다고 하더라도 실증(Proof of Concept)을 제대로 마치지 못한다면 수요처를 발굴했다고 하더라도 거래로 이루어지기 어렵다. 이를 위해서는 수요처와 AI 제품기획 및 개발이 한꺼번에 이루어져야 한다는 것이다. 앞선 이론 부분에서 설명한 스타트업의 두 가지 종류 - de alio, de novo

242) NVIDIA(23.05.25.), “ Cooling Operations Optimized for Leaps in Energy, Reliability, and Carbon Hyperefficiency for Information Processing Systems”, <https://blogs.NVIDIA.com/blog/liquid-cooling-doe-challenge/> (검색일 24.12.28.)

243) NVIDIA(10.08.09.) “NVIDIA-Led Team Receives \$25 Million Contract From DARPA to Develop High -Performance GPU Computing Systems”, <https://nvidianews.nvidia.com/news/nvidia-led-team-receives-25-million-contract-from-darpa-to-develop-high-performance-gpu-computing-systems> (검색일 24.12.28.)

- 로 나누어 볼 수 있는데, 특히 AI 반도체와 같이 수요처/실증기회 확보가 생존과 성장에 필수적이면 이러한 생태계 특성에 따라 두 가지 종류의 스타트업이 가지는 장단점을 명확히 살펴볼 수 있다.

de *alio* (기존 기업에서 spinoff 되었거나, corporate venturing 등을 통해 기존 기업과의 연계가 있는 스타트업)의 경우 이러한 수요처 확보가 완전히 새로운 시작을 하는 de *novo* 스타트업에 비해서 월등히 유리하다. de *alio* 기업의 경우 모기업이나 관련 기업을 통해 개발된 AI 반도체를 실증할 수 있고, 검증된 후 수요처로써도 역할을 할 수 있다(뉴블라, 사피온, 보스 반도체 사례). de *novo*의 경우에는 관련 기업들과 직접적인 관련은 없더라도 창업자가 관련 전공 혹은 관련 산업의 경험을 통해 실증을 하거나, 주요 투자자를 통한 실증이 가능하다. 한국의 AI 반도체 스타트업 중에 하나인 리밸리온(Rebellions)의 박성현 대표도 MIT에서 전자공학 전공으로 박사를 받았고 Networks on Chip을 주제로 연구를 진행하였으며, 이후 인텔의 프로세서 디자인, 삼성의 ASIC 디자인, SpaceX의 ASIC 디자인, 모건스탠리에서 지연 없는 트레이딩을 위한 반도체 디자인, 그리고 현재 AI 반도체를 개발하고 있다²⁴⁴⁾. 현재 리밸리온스는 삼성전자와 연계를 하고 제품을 개발 중이다. 캐나다에 있는 Tenstorrent 회사 역시 AI 컴퓨팅을 위한 칩 디자인을 전문적으로 하는 스타트업인데 2023년 LG 전자와의 협력은 물론 약 1억 달러의 대규모 투자를 한 현대차그룹과 삼성카탈리스트 펀드(Samsung Catalyst Fund)와 협력을 수행할 예정이라고 발표²⁴⁵⁾했다. 이는 텐스토렌트 스타트업 자체는 de *novo* 형태의 새로운 기업이지만, ‘반도체의 전설’로 불리는 현재의 CEO인 Jim Keller의 영향력이 큰 것으로 판단된다. 최근 보스반도체²⁴⁶⁾가 텐스토렌트와의 R&D 협력을 통해 공동개발한 자동차 AI 가속기 칩셋 SoC ‘이글-N(Eagle-N)’을 발표하였다. 이는 기존의 자동차 부문의 NPU 설계 스타트업인 보스반도체(현대 자동차로부터 기 20억원 투자를 받음²⁴⁷⁾)와 NPU 설계 부문의 글로

244) 매경(24.08.30.), “CPU·GPU 넘어 ‘NPU 시대’ 도전장 [CEO LOUNGE]”, <https://www.mk.co.kr/economy/view/2024/645923> (검색일 24.12.28.)

245) 디지털투데이(24.01.26.), “[단독] 짐 켈러의 AI칩 스타트업 ‘텐스토렌트’ 한국 시장 진출”, <https://www.digitaltoday.co.kr/news/articleView.html?idxno=503516> (검색일 24.12.28.)

246) zdnet(24.12.12.), “보스반도체, 텐스토렌트와 자동차 AI 가속기 칩셋 SoC ‘이글-N’ 출시”, <https://zdnet.co.kr/view/?no=20241211100527> (검색일 24.12.28.)

247) 현대자동차 뉴스룸(23.06.28), “현대자동차·기아, 차량용 반도체 스타트업 보스반도체에 20억 투자”,

별 리딩 스타트업인 텐스토렌트와의 협업을 통해 글로벌 자율주행차 시장에 참여하는 수요처에 대한 검증을 하는 작업으로 판단할 수 있다. 즉, 스타트업 기업의 특성(de alio 기업) 혹은 창업자의 특성(전공분야 혹은 관련 분야의 경험)을 통해 실증을 위한 기업과의 협력 그리고 나아가 그들이 수요처가 될 수 있도록 하는 연계가 성장과 생존을 위한 필수적 조건 중 하나다. 기업 인터뷰, 생태계 섹션별 전문가 인터뷰, 마인드맵 분석을 통해 분석된 바로도, 동 분야의 창업은 기술 부분에서의 차별성을 만들 수 있는 창업자에 의해서 주도되는 엘리트 창업 형태가 주를 이룬다는 것이다. 동 기업들의 chip 설계 이후 PoC 및 연계되는 MPW의 비용에 대한 지원과 동시에 기업들이 수요처와 설계의 초기 단계부터 수요기업과 연계할 수 있는 기회를 확보할 수 있는 장을 마련할 수 있는 정책적 지원도 고려해야 한다.

나. AI 반도체 스타트업의 Seed 투자 역할

AI 반도체 설계에서는 특정 블록 기능에 대한 설계 (IP 기업), 이 블록을 활용한 반도체 칩을 설계 (팹리스), 설계를 검증하는 (EDA), 설계가 된 반도체 칩을 파운드리에서 제조할 수 있도록 하는 과정 (디자인 하우스 및 디자인 솔루션 파트너(Design Solution Partner)이 진행된 이후 시제품으로 반도체를 만들어서 수요처를 고려한 테스트를 진행할 수 있다. 이 말은 다른 스타트업 분야와는 달리 AI 반도체의 설계만 하더라도, 대규모의 초기 Seed 투자가 있어야지만 AI 반도체를 시작해 볼 수 있다는 점을 내포하고 있다. 앞선 설명들과 같이 IP의 가격, EDA 로열티, 선단공정 사용 등 딥테크 SOC 기반의 AI 반도체 설계 (기존의 핵심을 넘은 칩셋 기술 등을 포함)를 통한 제품 개발 비용이 설계에서 수 백억원 (10nm 이하 공정에 쓰이는 초고속 연결 기술 '서데스(SerDes)' IP 가격은 500만~700만 달러²⁴⁸⁾(80억 원 내외) 이후 칩 양산까지 수천억 원이 들 수 있음을 고려해야 한다는 점이다.

글로벌 스타트업의 seed 투자를 고려할 때 미국 및 중국의 경우 평균적인 AI semiconductor의 Seed funding의 규모가 한국에 비해 크다. 현재 한국의 경우 TIPS 프

<https://www.hyundai.co.kr/news/CONT0000000000098823> (검색일 24.12.28)

248) 전자신문(21.12.22), “해외 '반도체 IP' 가격 급등…'첨단공정' 속탄다”,

<https://www.etnews.com/20211221000176> (검색일 24.12.28)

로그램을 통해서 초기 기업을 지원하고 있는데, 2024년 기준 텁스 R&D 최대 5억원, 딥테크 텁스 R&D가 최대 15억원이며²⁴⁹⁾, 이를 통해 직접적인 VC 투자로 연계되기가 쉽지 않다²⁵⁰⁾. 이는 AI 반도체 산업의 특징상 설계 자체도 큰 금액이 들지만, 시제품을 만들어 실증까지 시간과 비용이 많이 들기 때문이다. 이를 위하여, 향후 기획되는 정책에서 우선적으로 고려할 사항은 다음과 같다. ASIC 반도체를 사용할 수요처에 지원을 하는 방향이다.

[그림 4-2] AI 바우처 운영 개념



출처: “‘2024년 AI바우처 지원’ 사업공고 및 설명회 개최...425억원, 200개 과제 선정·지원”²⁵¹⁾,

위 그림의 개념도를 통해서 알 수 있듯이, 과기정통부에서 진행하고 있는 AI 바우처 형식으로 주문형 반도체를 수요 기업에 제공하는 형식으로 도입을 하는 방안을 계획해볼 수 있다. 기본 설계부터 POC 이전까지 지원하는 방식으로 정책을 고민하게 되면, AI반도체 스타트업 기업의 수요처 확보는 장기적으로 수요처와의 협력을 통해 진행 될 수 있으며, 정책 예산도 선순환적으로 활용될 수 있다. 대만 산업기술연구소(ITRI) 중소기업 R&D 지원프로

249) 2024년 텁스(TIPS) 창업기업 지원계획 통합 공고 – 지원사업 공고, 기업마당, 2024.03.19. https://www.bizinfo.go.kr/web/lay1/bbs/S1T122C128/AS/74/view.do?pblanId=PBLN_000000000095963 (검색일 24.12.28.)

250) 인터뷰를 통해 VC에서도 정부 투자를 받은 이력이 기업의 좋은 레퍼런스 작용한다고 언급하였지만, 몇 백억 단위의 투자금이 몰리는 상황에서 정부 사업 수행 이력만으로는 투자에 대한 확신을 주기 어렵다는 전문가 인터뷰

251) AI타임즈(24.02.26.), “‘2024년 AI바우처 지원’ 사업공고 및 설명회 개최...425억원, 200개 과제 선정·지원”

<https://www.aitimes.kr/news/articleView.html?idxno=30461> (검색일 24.12.28.)

그램²⁵²⁾에서, 중소기업이 정부 R&D 자금을 지원받고 자사의 필요 연구 분야에 해당한 R&D를 ITRI를 통해 진행하는 정책과 큰 궤를 같이한다(Jan & Chen, 2006). 예를 들어, 엣지향 반도체의 도입을 통한 사업의 다각화를 노릴 수 있는 자동차, 조선, 철강, 신업용 IoT제품 등 다양한 산업군에 매치되어 설계를 진행하거나 시제품을 제작하는 경우에 지원하는 금액이 이를 활용하는 수요 기업으로 지원이 되는 경우, 수요 기업으로서도 일부 레퍼런스가 부족한 상황에서도 퍼스트 무버로서 스타트업의 제품을 활용할 수 있는 최소한의 보호 장치를 갖기 때문이다.

다음으로는 America's Seed Fund로서²⁵³⁾ 혁신적 기술을 가진 스타트업 및 중소기업을 지원하는 미국의 SBIR(Small Business Innovation Research)과 같이 단계별 지원을 도입하여, 특정 단계부터 기업이 시장에서 투자받은 금액을 상한을 두고 정부에서 똑같이 매칭해주는 제도로 정부 자금을 지원하는 방식이다. SBIR은 Phase 1, 2, 3로 이뤄져 있는데, 아중 Phase 2에서는 외부에서 투자받은 지원금의 50%까지 매칭해주는 제도가 존재하고 있다. NSF Phase II²⁵⁴⁾에서는 제3자 투자금에 대해서 최대 50%까지 매칭해서 지원하고 있으며, NASA Phase II-E²⁵⁵⁾에서는 1:1 매칭으로 외부투자금과 동일하게 지원하며 최대 \$ 750,000 까지 지원하고 있다. 앞서 기술 평가 부문에 대한 문제점도 제시되었다시피, 정부-기업-수요자-투자자가 합의하는 기술 평가는 도입되지 않았고, 현재 시장에서 주류로 받아들여지는 기술평가 기준도 아직 없으므로, 현재 정부 지원 R&D보다 향상된 1차 지원 이후 (적어도 POC나 MPW를 진행할 수 있는), 시장에서의 매출 및 투자 유치에 대한 비율 매칭으로 정부 지원금을 투자하는 경우 주요하게 투자할 수 있는 기업을 선별하는데 도움이 될 수 있다. 이는 국가 R&D 지원을 통한 성장이후 기업이 시장에서 자생적으로 경쟁력을 갖춘 후, 기업의 제품 경쟁력, 사업 다각화, 발전 가능성, 성장 가능성을 기반으로 한 제 3자의 투자를 유치하고 이에 대한 매칭 지원이 목적이므로, 향후 정책 지원의 선별력 강화 및 집중 투자의 가능성을 전제할 때, 유용하게 활용할 수 있는 정책대안으로 사료된다.

252) Department of Industrial Technology, "A+ Industrial Innovative R&D Program", https://www.moea.gov.tw/MNS/doit_e/content/Content.aspx?menu_id=37705 (검색일 24.12.28.)

253) SBIR 홈페이지 사업소개, <https://www.sbir.gov/about> (검색일 24.12.28.)

254) America's Seed Fund-National Science Foundation 소개, <https://seedfund.nsf.gov/resources/awards/phases-phase-2-supplement/> (검색일 24.12.28.)

255) NASA 홈페이지 "Phase II-Extended (II-E)" 소개, https://www.nasa.gov/sbir_sttr/phase-ii-e/ (검색일 24.12.28.)

참고문헌

(논문/보고서)

KISTEPINI Vol.45 (2023)

S&T GPS(2022), 「미국, 인공지능 기술에 대한 중국의 접근을 제한하는 정책 추진」,
https://www.kistep.re.kr/gpsTrendView.es?mid=a30200000000&list_no=2753

S&T GPS(2023), 「미·중 첨단기술의 블록화와 우리의 대응」. 이슈분석 245호
https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=48740&nPage=3

S&T GPS(2023), 「글로벌 공급망 재편에 관한 EU의 주요 정책방향」. 이슈분석 250호,
https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=48745&nPage=3

고동환; 김민식; 김봉진; 김현태; 임종호:(2023)「정형/비정형데이터 기반 반도체 시장 이슈 분석 및 예측 모형 개발」, 정책자료 23-07, 정보통신 정책 연구원,

과기정통부 (2022.11.5.), “부처별 반도체 인재양성 사업은 유사 중복이 아닙니다”, 과기정통부 설명 자료

관계부처 합동(2019.5.1.). 「시스템반도체 비전과 전략」

관계부처 합동(2020.10.12.). 「인공지능(AI) 반도체 산업 발전전략」

과학기술정보통신부(2022.6.27.). 「인공지능(AI) 반도체 산업 성장 지원대책」

관계부처 합동(2022.7.21.), 「반도체 초강대국 달성을전략」

과학기술정보통신부(2022.12.12.). 「국산 AI반도체를 활용한『K-클라우드』추진방안」

관계부처 합동(2024.4.25.), 「AI-반도체 이니셔티브(안)」

권지연; 심은정; 안재용; 최동규 (2022). 유럽반도체법 주요 내용 및 영향 (디지털자료). KOTRA (대한무역투자진흥공사) 브뤼셀 무역관

김동수, 박병열, 최정환, 김양팽, 전정길. (2024). 대만 반도체산업 분석 및 정책적 시사점. 산업연구원. 이슈페이퍼 2024-03

김우용(2025.1.2.), AMD GPU가 엔비디아를 못 따라잡는 이유 ‘그놈의 SW’, <https://byline.network/2025/01/2-239/> (검색일: 2025.1.6.)

김정호(2023), 「하드웨어 유니콘의 스케일업 요인에 관한 연구: 해외 유니콘과 국내 예비유니콘의

사례 비교, 『한국창업학회지』, 18(3), pp. 325~352.

김현주.(2024). 「미국의 반도체 리더십 :국가반도체기술센터(NSTC) 출범」. 국회도서관 | 2024-8호 (통권 제80호)

미-중 기술패권 전쟁 2.0 [전자자료] : 반도체 전쟁(2022), 한국산업기술진흥원 산업기술정책센터 동향조사연구팀(KIaT)

윤종혁(2024.4.22.), 「엣지 디바이스에서의 AI 이용에 관한 산업분석」, 『ICT Standard Weekly』, 제1183호, 한국정보통신기술협회.

이선재(2023), AI 반도체 기술동향과 산업생태계, 『TTA Journal』, vol.205, pp.25~31.

임현 외, “2030 AI 반도체 공급망 시나리오 분석과 정책방안 스트레스 테스트”, 『KISTEP 브리프 148』, (2024.10).

장은현 (2024)., “AI 반도체 기술 및 산업 동향”, KDB산업은행 미래전략연구소(2024. 7, 제824호 3) 전원, 여준기(2023). 초거대 인공지능 프로세서 반도체 기술 개발 동향. 『전자통신동향분석』, 38 권, 5호, 1~11.

정부부처 합동.(2024.10.16.). 내년까지 반도체 산업에 8조 8000억 원 투입…생태계 경쟁력 키운다. <https://www.korea.kr/news/policyNewsView.do?newsId=148935134>, (검색일 2024.12.28.)

중국의 반도체 국산화 추진 현황과 시사점, 세계경제포커스 Vol. 6 No. 20, 대외경제정책연구원, 오종혁, 2023.06.27.

중소벤처기업부 (2023.1.26.), 이어드림 학교(스쿨), 디지털 인공지능(AI) 인재 키워 취·창업으로 연결한다, 중소벤처기업부 보도자료.

중소벤처기업부(2022.11.3.). 「첨단 미래산업 스타트업 육성전략 : 초격차 스타트업 1,000+ 프로젝트」

중소벤처기업부·중소기업기술정보진흥원, 「중소기업전략기술로드맵 2024~2026 「시스템 반도체」」.

채명식(2022), 「시스템반도체」, 『KISTEP 브리프』, 01, 한국과학기술기획평가원.

(국외문헌)

“Evolving the Electronics Resurgence Initiative (ERI 2.0)”- Briefing prepared for NDIA, Mark Rosker, MTO, 2021.04.21. DARPA https://www.ndia.org/-/media/sites/ndia/divisions/electronics/2021/eri2_ndia_20210421_releaseapproved_34584.pdf

Bahinipati, B. and Deshmukh, S. (2012). E-markets and supply chain collaboration: a literature-based review of contributions with specific reference to the semiconductor

- industries. *Logistics Research*, 4(1-2), 19-38. <https://doi.org/10.1007/s12159-012-0067-z>
- Benner, M. J., & Tripsas, M. (2012). The influence of prior industry affiliation on framing in nascent industries: The evolution of digital cameras.?Strategic Management Journal,?33(3), 277-302.
- CHESTER LAM(2024.9.12.), “FuriosaAI’s RNGD at Hot Chips 2024: Accelerating AI with a More Flexible Primitive”
<https://chipsandcheese.com/p/furiosaaais-rngd-at-hot-chips-2024-accelerating-a-i-with-a-more-flexible-primitive> (검색일: 2024.11.6.)
- Dao(2024), “FlashAttention”, <https://github.com/Dao-AI-Lab/flash-attention> (검색일: 2024.11.6.).
- Davis, F. D. (1989), “Perceived usefulness, perceived ease of use, and user acceptance of information technology”,?MIS Quarterly,?13?(3): 319?340.
- Del Bosco, B., Mazzucchelli, A., Chierici, R., & Di Gregorio, A. (2021). Innovative startup creation: The effect of local factors and demographic characteristics of entrepreneurs.?International Entrepreneurship and Management Journal,?17(1), 145-164.
- Eliakis, S., Kotsopoulos, D., Karagiannaki, A., & Pramatari, K. (2020). Survival and Growth in Innovative Technology Entrepreneurship: A Mixed-Methods Investigation. *Administrative Sciences*, 10(3), 39. <https://doi.org/10.3390/admsci10030039>
- Eric Ries, (2011), “The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses”
- Gang, K. (2018). The impact of pre-entry experiences on entry decisions and firm survival.?Technological Forecasting and Social Change,?137, 249-258.
- Hess J., Denkena, W., Kleinhans, J.P., Maham, P. (2023). Who is funding the chips of the future? Analysis of global semiconductor startup funding activities. Stiftung Neue Verantwortung (<https://www.stiftung-nv.de/en/publication/semiconductor-startup-funding-activities>).
- Howard(2024.1.29.), An Overview of NVIDIA NVLink, <https://community.fs.com/article/an-overview-of-nvidia-nvlink.html> (검색일: 2024.11.6.)
- Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, and venture capital funding.?Research policy,?36(5), 722-741.

- Hwang, B. and Lu, T. (2013). Key success factor analysis for e‐scm project implementation and a case study in semiconductor manufacturers. *International Journal of Physical Distribution & Logistics Management*, 43(8), 657-683. <https://doi.org/10.1108/ijpdlm-03-2012-0062>
- Jan, Tain-Sue & Chen, Yi-Jen. (2006). The R&D system for industrial development in Taiwan. *Technological Forecasting and Social Change - TECHNOL FORECAST SOC CHANGE*. 73. 559-574. 10.1016/j.techfore.2005.01.005.
- Julia Hess, Wiebke Denkena, Jan-Peter Kleinhans, Pegah Maham (2023). Who is funding the chips of the future? Analysis of global semiconductor startup funding activities. Data Brief. Think Tank at the Intersection of Technology and Society. Stiftung Neue Verantwortung <https://www.interface-eu.org/publications/semiconductor-startup-funding-activities>
- Kalthaus, M. (2020). Knowledge recombination along the technology life cycle.?Journal of Evolutionary Economics,?30(3), 643-704.
- Kalyanasundaram, G., Ramachandrula, S., & Mungila Hillemane, B. S. (2021). The life expectancy of tech start-ups in India: what attributes impact tech start-ups' failures?.?International Journal of Entrepreneurial Behavior & Research,?27(8), 2050-2078.
- Kapoor, R., & Furr, N. R. (2015). Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry.?Strategic Management Journal,?36(3), 416-436.
- Khessina, O. M., & Carroll, G. R. (2008). Product demography of de novo and de alio firms in the optical disk drive industry, 1983?1999.?Organization Science,?19(1), 25-38.
- Klepper, S., & Simons, K. L. (2000). The making of an oligopoly: firm survival and technological change in the evolution of the US tire industry.?Journal of Political economy,?108(4), 728-760.
- Klepper, S., & Simons, K. L. (2005). Industry shakeouts and technological change.?International Journal of Industrial Organization,?23(1-2), 23-43.
- Kumar, S. (2024). Opportunities and challenges in data-centric ai. Ieee Access, 12, 33173-33189. <https://doi.org/10.1109/access.2024.3369417>
- Li, M. (2024). Research on NVIDIA's success and socioeconomic perspective. *Advances*

- in Economics Management and Political Sciences, 67(1), 44-49. <https://doi.org/10.54254/2754-1169/67/20241260>
- Meta(2024). Introducing Llama 3.1: Our most capable models to date. <https://ai.meta.com/blog/meta-llama-3-1/>
- Moore, G. A. (1991).?Crossing the Chasm: Marketing and Selling High-Tech Goods to Mainstream Customers.?New York: Harper Business.
- NVIDIA(2024). NVIDIA Blackwell architecture technical brief: Powering the new era of generative AI and accelerated computing. <https://resources.NVIDIA.com/en-us-blackwell-architecture>
- Oh, Y., et al.(2023). Mobilint's ARIES: Chip for Edge AI. In 2023 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-4). IEEE.
- Paik, Y. and Woo, H. (2013). Economic downturn and financing innovative startup companies. Managerial and Decision Economics, 35(2), 114-128.<https://doi.org/10.1002/mde.2646>
- Patel et al.(2024.12.22.), MI300X vs H100 vs H200 Benchmark Part 1: Training ? CUDA Moat Still Alive Training Performance, User Experience, Usability, Nvidia, AMD, GEMM, Attention, Networking, InfiniBand, Spectrum-X Ethernet, RoCEv2 Ethernet, SHARP, Total Cost of Ownership, <https://semianalysis.com/2024/12/22/mi300x-vs-h100-vs-h200-benchmark-part-1-training/> (검색일: 2025.1.6.)
- Prytkova, E. and Vannuccini, S. (2022). On the basis of brain: neural-network-inspired changes in general-purpose chips. Industrial and Corporate Change, 31(4), 1031-1055. <https://doi.org/10.1093/icc/dtab077>
- Qoriawan, T. and Apriliyanti, I. (2022). Exploring connections within the technology-based entrepreneurial ecosystem (ee) in emerging economies: understanding the entrepreneurship struggle in the indonesian ee. Journal of Entrepreneurship in Emerging Economies, 15(2), 301-332.<https://doi.org/10.1108/jeee-02-2021-0079>
- Ren, Y., Yang, Y., Wang, Y., & Yi, L. (2023). Dynamics of the global semiconductor trade and its dependencies. Journal of Geographical Sciences, 33(6), 1141-1160. <https://doi.org/10.1007/s11442-023-2123-9>
- Reuters(2024.8.30). OpenAI says ChatGPT's weekly users have grown to 200 million, <https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/>

- Saha, S. (2018). Transitioning semiconductor companies enabling smart environments and integrated ecosystems. Open Journal of Business and Management, 06(02), 428-437. <https://doi.org/10.4236/ojbm.2018.62031>
- Singh, S. and Subrahmanyam, M. (2020). The financial requirements of tech startups over its lifecycle in bangalore: an analysis of why and how do they differ?. International Journal of Finance & Economics, 27(4), 4123-4141.<https://doi.org/10.1002/ijfe.2362>
- St. Micheal(2019.2.7.), What Is a Microarchitecture? Understanding Processors and Register Files in an ARM Core, <https://www.allaboutcircuits.com/technical-articles/what-is-a-microarchitecture-processor-register-files-ARM-core/> (검색일: 2024.11.6.)
- Tal, E., Viljoen N., Coburn J., and Levenstein R.(2024.04.10.). Our next-generation Meta Training and Inference Accelerator. <https://ai.meta.com/blog/next-generation-meta-training-inference-accelerator-AI-MTIA/>
- Triono, Sunu & Agus, Rahayu & Wibowo, Lili & Alamsyah, Andry. (2022). Factors Affecting Start-up Performance: A Literature Review. 10.2991/aebmr.k.220701.097.
- Tushman, Michael L. and Lori Rosenkopf (1992) “Organizational Determinants of Technological Change: Toward a Sociology of Technological Evolution,” Research in Oragnizational Behavior Vol 14: 311-347
- Vahdat, A. & Lohmeyer, M.(2023.12.7.). Enabling next-generation AI workloads: Announcing TPU v5p and AI hypercomputer. Google Cloud, <https://cloud.google.com/blog/products/ai-machine-learning/introducing-cloud-tpu-v5p-and-ai-hypercomputer>
- Vaswani, A., et al.(2017). Attention is all you need. Advances in Neural Information Processing Systems.
- Wang, J. (2024). A comprehensive analysis of NVIDIA's technological innovations, market strategies, and future prospects. International Journal of Information Technologies and Systems Approach, 17(1), 1-16. <https://doi.org/10.4018/ijitsa.344423>
- Wang, J. (2024). Exploring NVIDIA's evolution, innovations, and future stock trends. JMEI, 17(1), 21-33. <https://doi.org/10.62704/10057/28082>
- Wheelwright, S. C., & Clark, K. B. (1992). Revolutionizing product development: quantum leaps in speed, efficiency, and quality. Simon and Schuster.
- Yin, Y. (2023). Analysis of NVIDIA's acquisition of mellanox technologies, ltd. based

on peg and eva valuation model. *Highlights in Business Economics and Management*, 19, 293–300. <https://doi.org/10.54097/hbem.v19i.11892>

半導??デジタル産業?略 (改定案) (반도체?디지털 산업전략 개정안)(2023.05.30.), 경제산업성 (METI)https://www.meti.go.jp/policy/mono_info_service/joho/conference/seminar_on_digital/0009/4hontai.pdf

(발표자료)

김정호(2024.5.17.), 「AI반도체 기업의 스케일업 요인: 국내외 사례, 최근 연구 결과와 시사점」.

유병두(2024.3.25.), 「시스템반도체 산업동향」, 시스템반도체설계지원센터.

이동수(2022), 「지능형 반도체의 새로운 패러다임」, 2023 ICT 산업전망컨퍼런스, 네이버.

전원(2024.5.28.), 「Technical Trends in Hyperscale Artificial Intelligence Processors」, STEPI 세미나 자료.

한국팹리스산업협회 (2024.4.23.), 「팹리스 스타트업 육성을 위한 정책방향성 제언」.

(인터넷 홈페이지)

가트너, <https://www.gartner.com>

경향신문, <https://www.khan.co.kr>

글로벌이코노믹, <https://www.g-enews.com/ko-kr>

기업마당, <https://www.bizinfo.go.kr/>

김앤장, <https://www.kimchang.com>

뉴스와이어, <https://www.newswire.co.kr>

뉴스핌, <https://www.newspim.com>

더벨, <https://www.thebell.co.kr/>

동아일보, <https://www.donga.com>

디일렉, <https://www.thelec.kr/>

디지털투데이, <https://www.digitaltoday.co.kr>

매경, <https://www.mk.co.kr/>

매일일보, <https://www.m-i.kr/>

- 머니투데이], <https://news.mt.co.kr>
- 보안뉴스, <https://www.edge-ai-vision.com>
- 브릿지경제, <https://www.viva100.com>
- 삼성전자 파운드리 SAFE 소개 홈페이지, <https://semiconductor.samsung.com/kr/foundry/safe/>
- 서울경제, <https://www.sedaily.com>
- 서울신문, <https://www.seoul.co.kr>
- 시사저널e, <https://www.sisajournal-e.com>
- 아시아경제, <https://view.asiae.co.kr>
- 아주경제, <https://www.ajunews.com>
- 월간인물, <https://monthlypeople.com>
- 이뉴스투데이, <http://www.enewstoday.co.kr>
- 전자신문, <https://www.etnews.com>
- 정보통신산업진흥원(NIPA) AI반도체 응용실증지원 소개 홈페이지, <https://www.nipa.kr/home/bsnsAll/3/detail?bsnsDtlsItemNo=638&tab=1>
- 정보통신산업진흥원(NIPA), <https://www.nipa.kr/>
- 조선비즈, <https://biz.chosun.com>
- 조선일보, <https://www.chosun.com>
- 중소벤처기업부, <https://www.mss.go.kr>
- 키포스트, <https://www.kipost.net>
- 파이낸셜 뉴스, <https://www.synopsys.com>
- 파이토치한국사용자모임. <https://pytorch.kr>
- 퓨리오사AI 홈페이지, <https://furiosa.ai/rngd>
- 한겨레, <https://www.hani.co.kr>
- 한경비즈니스, <https://magazine.hankyung.com>
- 한국경제, <https://www.hankyung.com>
- 한국무역협회, <https://kita.net>
- 한국반도체산업협회 홈페이지, <https://www.ksia.or.kr/index.php>

한국팹리스산업협회 홈페이지, <https://www.k-fabless.com/kr/index.php>

헤럴드경제, <https://biz.heraldcorp.com>

헬로티, <https://www.hellot.net>

혁신의숲 홈페이지, <https://www.innoforest.co.kr>

현대자동차, <https://www.hyundai.co.kr>

AI타임즈, <https://www.aitimes.com>

IT데일리, <http://www.itdaily.kr>

IT비즈뉴스, <https://www.itbiznews.com>

KDI경제정보센터, <https://eiec.kdi.re.kr/material>

W 브릿지, <https://www.wbridge.or.kr>

Alan D. Thompson, <https://lifearchitect.ai/>

America's Seed Fund-National Science Foundation, <https://seedfund.nsf.gov>

ARM, <https://www.arm.com>

Aving.net, <https://us.aving.net>

BusinessKorea, <https://www.businesskorea.co.kr>

Businesswire, <https://www.businesswire.com>

CBInsight, <https://www.cbinsights.com>

Celestial AI, <https://www.celestial.ai>

Cerebras, <https://cerebras.ai/>

CHESTER LAM, <https://chipsandcheese.com>

CNBC, <https://www.cnbc.com>

Communications of the ACM, <https://cacm.acm.org>

Crunchbase, <https://www.crunchbase.com>

Data center dynamics, <https://www.datacenterdynamics.com>

Department of Industrial Technology Taiwan, <https://www.moea.gov.tw>

Design & reuse, <https://www.design-reuse.com>

Edge AI and Vision Alliance, <https://www.edge-ai-vision.com>

European Commission, <https://digital-strategy.ec.europa.eu/en>

Forbes, <https://www.forbes.com>

Forge Global, <https://forgeglobal.com>

FuriosaAI. Furiosa SDK Documentation 0.10.1, FuriosaAI SW stack introduction.
<https://furiosa-ai.github.io/docs/latest/en/software/intro.html>

Futurum, <https://futurumgroup.com>

Github, <https://github.com>

GlobalFoundries, <https://gf.com>

Google Cloud, <https://cloud.google.com>

Hailo AI, <https://hailo.ai>

hitnews, <http://www.hitnews.co.kr>

HPCwire, <https://www.hpcwire.com>

IBM, <https://www.ibm.com>

IMDT, <https://www.imd-tec.com>

Intel, <https://intel.com>

Invest Chosun, <https://www.investchosun.com>

KAIST, <https://kaist.ac.kr>

KolonPharma, <https://kolonpharm.co.kr>

KOTRA, <https://dream.kotra.or.kr/>

Laser focus world, <https://www.laserfocusworld.com>

Myscale, <https://myscale.com>

NASA, <https://www.nasa.gov>

National Applied Research Laboratories, <https://www.narlabs.org.tw/en/>

Numbers, <https://www.numbers.co.kr>

Nestrends Asia, <https://nextrendsasia.org>

Next platform, <https://www.nextplatform.com>

NVIDIA. Deep learning software. <https://developer.nvidia.com/deep-learning-software>

- PR Newswire, <https://www.prnewswire.com>
- Rebellions. Rebellions' software stack: silent support. <https://rebellions.ai/category/papers/>
- Samsung, <https://semiconductor.samsung.com>
- SAPEON. Enhanced developer experience with our comprehensive full SW stack.
<https://www.sapeon.com/products/software>
- SBIR, <https://www.sbir.gov>
- Semiconductor Today, <https://www.semiconductor-today.com>
- Synopsys, <https://www.synopsys.com>
- Tech Target, <https://www.techtarget.com>
- Tech world, <https://www.epnc.co.kr>
- Techwire asia, <https://techwireasia.com>
- Trendforce, <https://www.trendforce.com>
- TheJerusalempost, <https://www.jpost.com>
- Things Embedded Limited, <https://things-embedded.com>
- Topdaily, <https://www.topdaily.kr>
- TSMC, <https://tsmc.com>
- UMC, <https://umc.com>
- Upstage, <https://www.upstage.ai>
- Wowtale, <https://wowtale.net/>
- Yahoo Finance, <https://finance.yahoo.com>
- Zdnet, <https://zdnet.co.kr>
- ZUBRCapital, <https://zubrcapital.com>