

PIM 인공지능반도체 발전 현황 및 시장 전망

송근혜·설성호

본 보고서는 ETRI ICT전략연구소 기본사업인
“국가 지능화 기술정책 및 표준화 연구”를 통해 작성된 결과물입니다.



Executive summary

연구 목적

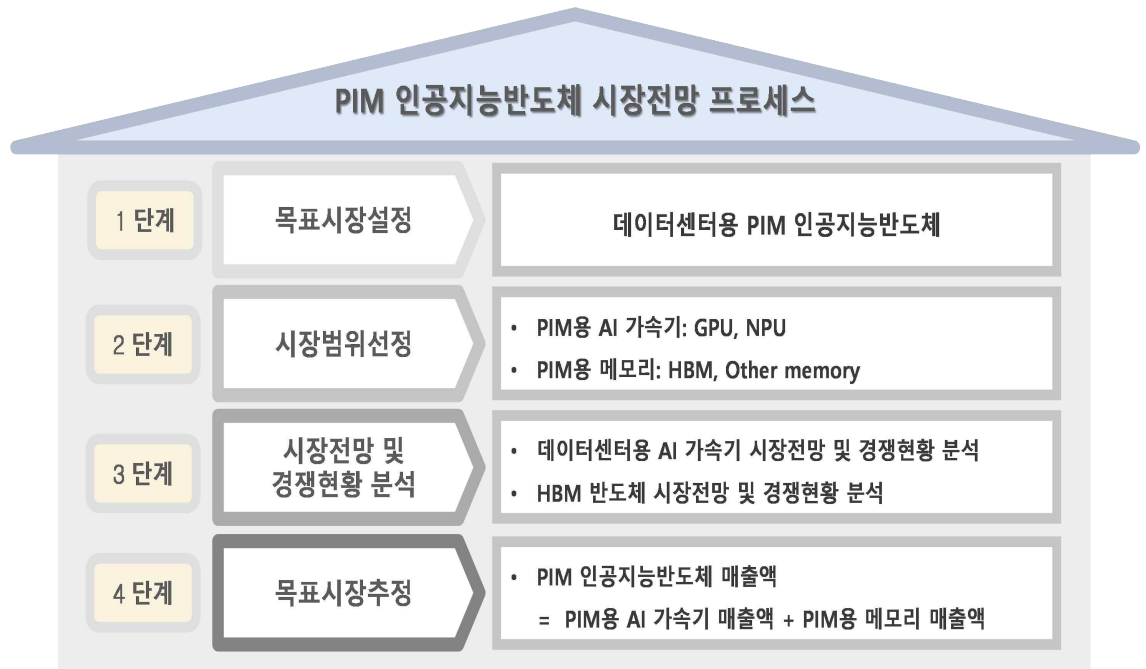
- 차세대 AI반도체인 PIM 인공지능반도체의 발전 방향, 개발 현황, 성장가능성 분석을 통한 전략 제언
 - 어떻게 발전해 나갈 것인가? → AI반도체 측면에서의 발전 방향
 - 누가 관심을 두고 있는가? → 주요 제품 및 R&D 현황
 - 성장가능성이 있는가? → 시장 전망
 - 무엇을 해야 하는가? → 전략 제언

주요 내용

- PIM 인공지능반도체 발전 방향
 - AI반도체는 연산기 자체 발전과 메모리와 연산기의 연계 발전을 통해 진화
 - 메모리와 연산기를 융합하는 방식에 집중하는 PIM 인공지능반도체는 NM-PIM(PIM 1세대), IM-PIM(PIM 2세대), 뉴로모픽형 PIM(PIM 3세대)으로 발전 전망
 - ※ NM-PIM은 단일 칩에 메모리와 로직 다이가 구분되는 형태(NM-PIM i)와, 단일 다이에 메모리와 로직이 집적되는 형태(NM-PIM ii)로 구분
- 주요 제품 및 R&D 현황
 - PIM 인공지능반도체는 NM-PIM 유형을 중심으로 출시되고 있으며, 시스템반도체 기업, 메모리반도체 기업, 빅테크, 스타트업 등 다양한 기업들이 PIM 인공지능반도체 제품 출시
 - ※ AI 가속기 업체는 HBM을 바로 삽입할 수 있는 'NM-PIM i'를 주로 개발하는 한편, 메모리 업체는 'NM-PIM i' 뿐만 아니라 'NM-PIM ii'도 개발
 - NM-PIM i 단계에서는 HBM의 주요 수요처가 AI 가속기 업체이지만, NM-PIM ii 단계에서는 메모리 업체가 개발한 제품의 단독 AI 프로세싱 성능이 높아질수록 향후, AI 가속기 업체에 대한 의존도가 줄어들기 시작할 것으로 보임
- PIM 인공지능반도체 시장전망
 - PIM 인공지능반도체 시장전망은 ① 목표시장설정 ② 시장범위선정 ③ 시장전망 및 경쟁현황 분석 ④ 시장추정의 4단계로 구분하여 진행

※ 기존 PIM 인공지능반도체 시장전망과 달리, 시장범위를 한층 더 세분화하고 PIM 인공지능반도체 시장 규모 증가에 대한 임의성을 최소화함으로써 시장 전망의 정확도를 높임

그림 데이터센터용 PIM 인공지능반도체 시장 전망 프로세스



* 출처: 저자 작성

- 전 세계 데이터센터용 PIM 인공지능반도체 시장은 2022년~2027년 사이 연평균 28.1% 성장하여 2027년에는 약 340억 달러 규모를 형성할 전망
- PIM 인공지능반도체가 전체 AI 반도체 시장에서 차지하는 비중은 2022년 약 22.3%에서 2027년 약 28.4%로 증가하여, 약 1/3 규모를 차지할 전망

※ Gartner와 Omdia의 AI반도체 시장 데이터를 참고하여 저자 산출

📄 결론 및 시사점

- 국내 기업의 NM-PIM ii 단계 PIM 인공지능반도체의 조기 상용화 방안 마련
 - NM-PIM ii 단계 개발 프로세스 단축 지원
 - ※ AI모델 호환성을 지원하는 SW 플랫폼 구축, IP 블록 표준화 작업 및 DB 구축 등
 - NM-PIM ii 단계에 대한 실증 기회 제공
 - ※ MPW 지원 대상 범위 확대, 대기업 공조 강화, 공신력 있는 인증 부여 등



- 학연 중심의 R&D 진행 중인 IM-PIM(PIM 2세대)은 벤처기업 지원 및 육성을 통해, 차세대 PIM 제품 개발 및 조기 상용화
 - IM-PIM 인공지능반도체 R&D 결과물이 차세대제품 수요 대응으로 연결될 수 있는 발판 마련
 - 정부 지원 대상 벤처기업의 다양성 확대

목 차

C O N T E N T S

Executive summary

I. PIM 인공지능반도체 개요 및 발전 방향	1
1. PIM 인공지능반도체 개요	1
2. PIM 인공지능반도체 발전 방향	4
II. PIM 인공지능반도체 현황 분석	6
1. PIM 인공지능반도체 R&D 현황	6
2. PIM 인공지능반도체 제품 현황	11
[참고2] PIM 인공지능반도체 기술·사업화·특허경쟁력 수준	17
III. PIM 인공지능반도체 시장 전망	20
1. 시장 전망 프로세스	20
2. PIM 인공지능반도체 시장 전망	24
가. 데이터센터용 AI가속기 시장 전망 및 경쟁 현황	24
나. HBM 시장 전망 및 경쟁 현황	28
다. PIM 인공지능반도체 시장 전망	32
IV. 결론 및 제언	35
참고문헌	39



참고문헌

○ 국내자료

- 권영수 (2021), “PIM 인코지능 NPU”, 전자공학회지 Vol.48, No.6. pp. 447-454.
- 정보통신기획평가원 (2020), ICT R&D 기술로드맵 2025.
- 윤종혁 (2023), “[DGIST 시리즈 1편] 메모리 기반 연산 가속기 : 저장과 연산을 동시에 하는 진정한 두뇌로”, <https://news.skhynix.co.kr/post/dgist-series-1>
- 강신욱·추민성 (2023), “인공지능 반도체(PIM) 기술 연구 및 산업 동향”, KISDI AI Outlook, 2023년 12권.
- 박성모·최병건·박필재·박경환 (2022), “Processing-in-Memory 반도체 기술 동향”, 주간기술동향 2075호, pp. 2-14.
- 신창환·김영우 (2022), “AI반도체 생태계 분석”, Digital Insight 2022, 한국지능정보사회진흥원.
- 임의철 (2023), “Cost-effective LLM accelerator system using AiM”, ETRI Conference 2023.
- 김선영·안후영·박유미·한우중 (2023), “CXL 인터커넥트 기술 연구개발 동향”, 전자통신동향분석 제 38권 제5호, pp. 23-33.
- 한국전자통신연구원 (2022), “차세대 컴퓨팅 분야 기술 분석 및 전망”
- 과학기술정보통신부 (2022), “인공지능 반도체 산업 성장 지원대책”

○ 해외자료

- N. P. Jouppi et al., "Ten Lessons From Three Generations Shaped Google's TPUv4i : Industrial Product," 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA), Valencia, Spain, 2021, pp. 1-14, doi: 10.1109/ISCA52012.2021.00010.
- Gartner, “Forecast Analysis: High-Bandwidth Memory, Worldwide, 2022-2027”, 2023.
- OMDIA, “AI Processors for Cloud and Data Center Forecast Report”, 2023.
- OMDIA, “DRAM Market Tracker-2Q23 Analysis”, 2023.
- Kazi Asifuzzaman, Narasinga Rao Miniskar, Aaron R. Young, Frank Liu, Jeffrey S. Vetter, 2023.07, A Survey on processing-in-memory techniques : Advances and challenges, Memories-Materials, Devices, Circuits and System, volume 4, pp. 1-11.

- S. Lee, S.-h. Kang, J. Lee, H. Kim, E. Lee, S. Seo, H. Yoon, S. Lee, K. Lim, H. Shin, J. Kim, O. Seongil, A. Iyer, D. Wang, K. Sohn, N.S. Kim, Hardware architecture and software stack for PIM based on commercial DRAM technology : Industrial product, in: 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture, ISCA, 2021, pp. 43-56.
- NVIDIA, “NVIDIA GH200 Grace Hopper Superchip”, 2023.08.
- NVIDIA, “NVIDIA H100 PCIe GPU: Product brief”, 2022.11.

○ 신문기사

- THEELEC(2023.09.19.), SK하이닉스 ‘HBM3E’, HBM 첫 초당 테라바이트 돌파...비결은?
(<https://www.thelec.kr/news/articleView.html?idxno=23125>)
- BusinessKorea(2023.10.30.), Intense Race to Dominate the HBM3E Market : SK, Samsung, Micron. (<https://www.businesskorea.co.kr/news/articleView.html?idxno=204588>)
- 전자신문(2021.02.17.), 삼성전자, 세계 최초 인공지능 결합 메모리 ‘HBM-PIM’ 개발...“에코 시스템 구축 시동”. (<https://www.etnews.com/20210217000148>)
- 더벨(2021.01.08.), 에이스에쿼티, 2조 반도체업체 합병...스팩 첫 성과.
(<https://www.thebell.co.kr/free/content/ArticleView.asp?key=202101081229230800107640&lcode=00>)
- Medium(2020.06.19.), OpenAI GPT-3: Language Models are Few-Shot Learners
(<https://medium.com/analytics-vidhya/openai-gpt-3-language-models-are-few-shot-learners-82531b3d3122>)
- ANANDTECH(2023.06.13.), AMD expands AI/HPC product lineup with flagship GPU-only Instinct MI300X with 192GB memory.
(<https://www.anandtech.com/show/18915/amd-expands-mi300-family-with-mi300x-gpu-only-192gb-memory>)
- 삼성전자(2023.06.02.), 삼성전자 메모리는 어떻게 AI 혁명을 이끌어 가고 있는가?
(<https://semiconductor.samsung.com/kr/news-events/tech-blog/how-samsung-memory-is-powering-the-ai-revolution/>)
- SK하이닉스(2023.09.18.), SK하이닉스, 생성형 AI에 특화된 GDDR6-AiM 기반 가속기 카드 ‘AiMX’ 시제품 최초 공개 (<https://news.skhynix.co.kr/post/ai-hw-summit-2023>)



○ 웹사이트

마이크론 홈페이지, <https://www.micron.com/>

아크로닉스 홈페이지, <https://www.achronix.com>

구글 클라우드 홈페이지, <https://cloud.google.com/tpu/docs/system-architecture-tpu-vm>

NVIDIA 홈페이지, <https://www.nvidia.com/ko-kr/>

Graphcore 홈페이지, <https://www.graphcore.ai/>

Cerebras 홈페이지, <https://www.cerebras.net/>

AMD 홈페이지, <https://www.amd.com/ko.html>

AWS 홈페이지, <https://aws.amazon.com>

Cambricon 홈페이지, <https://www.cambricon.com/>

저자 소개

송근혜 ETRI ICT전략연구소 기술정책연구본부 산업분석연구실 연구원
e-mail: ghsong0227@etri.re.kr, Tel. 042-860-6702

설성호 ETRI ICT전략연구소 기술정책연구본부 통신정책연구실 책임연구원
e-mail: ssh1517@etri.re.kr, Tel. 042-860-6603

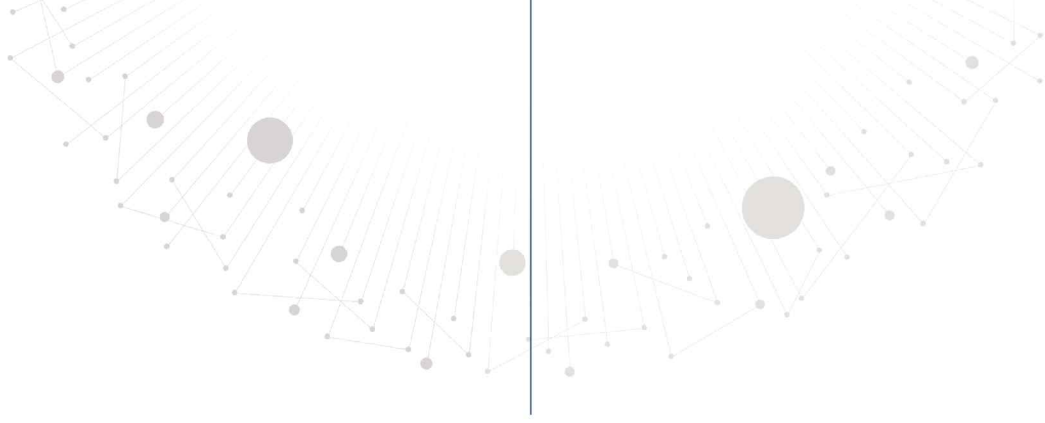
PIM 인공지능반도체 발전 현황 및 시장 전망

발행인 한 성 수

발행처 한국전자통신연구원 ICT전략연구소

발행일 2023년 12월 31일





www.etri.re.kr

본 저작물은 공공누리 제4유형:

출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.



ETRI Electronics and Telecommunications
Research Institute

34129 대전광역시 유성구 가정로 218
TEL.(042) 860-6114 FAX.(042) 860-6504

