

2022 August | Vol. 8

08



융합연구리뷰

Convergence Research Review

차세대 전력반도체 소자 기술

경신수(파워큐브세미(주) 연구소장)

프로세싱 인 메모리 기반 뉴로모픽 기술 동향

박성모(한국전자통신연구원 책임연구원)

박필재(한국전자통신연구원 책임연구원)

CONTENTS

- 01 편집자 주
- 03 차세대 전력반도체 소자 기술
- 29 프로세싱 인 메모리 기반 뉴로모픽 기술 동향
- 53 국가R&D 현황 분석



융합연구리뷰 | Convergence Research Review
2022 August vol.8 no.8

발행일 2022년 8월 8일

발행인 김현우

발행처 한국과학기술연구원 융합연구정책센터
02792 서울특별시 성북구 화랑로 14길 5
Tel. 02-958-4977 | <http://crpc.kist.re.kr>

펴낸곳 한빛사회적협동조합



차세대 전력반도체 소자 기술

이상 기후로 전 세계가 위협을 받고 있는 현재, 탄소 배출 저감을 위해 전력반도체의 중요성이 부각되고 있다. 전력반도체란 전자 기기 등에 필요한 전력을 각 기기에 맞게 변환, 변압, 안정, 분배, 제어하는데 사용되는 반도체를 의미한다. 태양광 패널에서 생산된 전력은 수천 kV의 고전압으로, 우리가 사용할 수 있는 전기로 바꾸기 위해서는 인버터를 통해 전력변환 과정을 거쳐야 하는데, 이때 인버터의 핵심 부품이 전력반도체이다. 또한 전기차 배터리의 직류 전기를 교류 전기로 바꾸어 전동기에 공급하는 인버터의 핵심부품 역시 전력반도체이다. 이처럼 전력반도체는 2050년 탄소중립을 실현하기 위한 신재생에너지, 전기차 시장 등에서 수요가 급증하고 있고 최근 들어 와이드 밴드갭(WBG, Wide BandGap) 물질 기반의 소자로 제작되는 차세대 전력반도체가 부상하고 있다.

1960년대부터 실리콘 기반의 반도체 소자가 전력반도체에 활용되어 왔으나 효율성 및 신뢰성의 문제로 인해 넓은 에너지 준위 특성(와이드 밴드갭)의 차세대 전력반도체가 등장하였다. 본 호 1부에서는 탄화규소, 질화갈륨, 산화갈륨 기반의 차세대 전력반도체 소자에 대해 소개한다.

차세대 전력반도체는 5G 이동통신, 사물인터넷, 모바일 디바이스 등 관련 산업 성장으로 향후 수요가 크게 성장할 것으로 전망되어 반도체의 시장의 블루오션으로 주목을 받고 있다. 지속적인 연구개발로 우리나라가 4차 산업혁명 시대를 이끌어 갈 차세대 전력반도체 산업을 선점할 수 있기를 기대해 본다.

프로세싱 인 메모리 기반 뉴로모픽 기술 동향

2016년 구글 딥마인드사의 '알파고'와 이세돌과의 대국은 컴퓨터와 인간의 대결로 주목을 받은 한편, 기존의 컴퓨팅 하드웨어의 한계를 드러냈다. 이세돌은 커피 한 잔과 바나나 두 개로 에너지를 충당했지만 알파고는 무려 3,000여대의 기업용 서버를 연결해야 했고, 176개의 그래픽 처리 장치, 1,202개의 중앙 처리 장치, 103만개의 메모리 반도체, 100여 명의 과학자 등 물적·인적 자원이 동원되었으며 시간 당 170kw의 막대한 전력을 소모했다. 알파고의 하드웨어는 폰 노이만(Von Neumann) 방식의 아키텍처로, 중앙 처리 장치와 메모리 사이에 있는 하나의 통로로 데이터가 순차적으로 이동하며 처리되는데, 저장 장치와 연산 장치가 별도로 되어있어 처리 양이 많아지면 지연현상이 발생하는 구조적 한계가 있다. 이러한 컴퓨팅의 기술적 한계를 극복하기 위한 대안으로 뉴로모픽 반도체가 등장했다.

뉴로모픽(Neuromorphic, Neuro: 뇌 신경 + Morphic: 모방하다) 반도체는 인간의 뇌신경 구조와 기능을 모방한 인공지능 반도체로 내부에 뉴런과 시냅스를 모방한 소자가 들어 있어 정보의 저장·연산·전송이 동시에 가능하여 기존 반도체에 비해 성능은 뛰어나고, 초저 전력으로 구동이 가능해 전력 소모량은 1억 분의 1에 불과하다.

전 세계 각국과 함께 우리 정부도 3세대 인공지능 반도체인 뉴로모픽 반도체를 개발하고 있으며, 특히 인공지능 과 빅데이터 처리 분야에서 데이터 이동 정체 문제를 풀 수 있는 기술인 프로세싱 인 메모리(PIM, Processing in Memory) 개발에 주력할 예정이다. 본 호 2부에서는 인공지능 연산에 최적화된 프로세싱 인 메모리(PIM)에 대해 소개한다.



융합연구리뷰

Convergence Research Review 2022 August vol.8 no.8



01

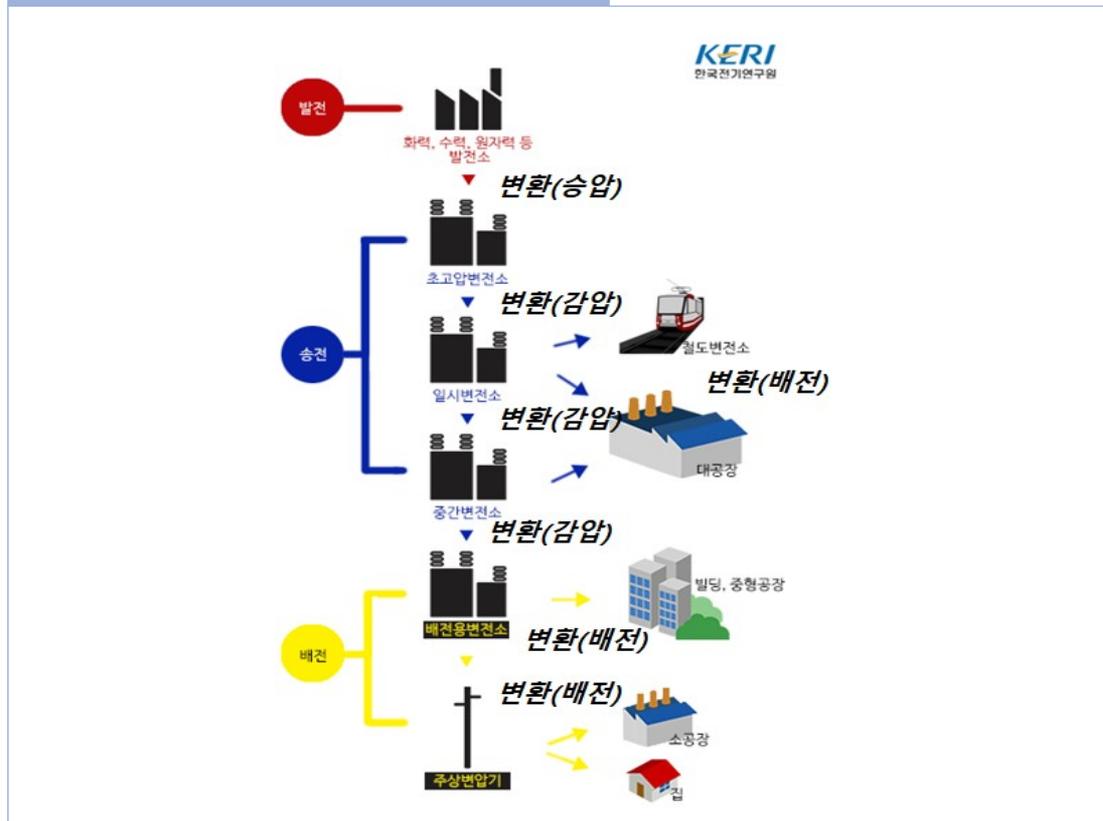
차세대 전력반도체 소자 기술

경신수(파워큐브세미(주) 연구소장)

I 서론

현대는 전기 기반 문명이라고 해도 과언이 아닐 정도로 거의 모든 생활기 기구가 전기에너지를 기반으로 한 장치에 근거하고 있다. 전기 에너지는 발전소를 통해 생산된 전력을 <그림 1>에서 보는 바와 같이 필요에 따라 전압, 전류, 주파수, 위상 등을 변환하여 사용한다. 변환 과정마다 전력변환장치(power conversion system)가 필수적으로 사용되는데 이 전력변환장치의 핵심 부품이 전력반도체 소자이다.

그림 1. 전력변환을 통한 전기에너지 사용 흐름



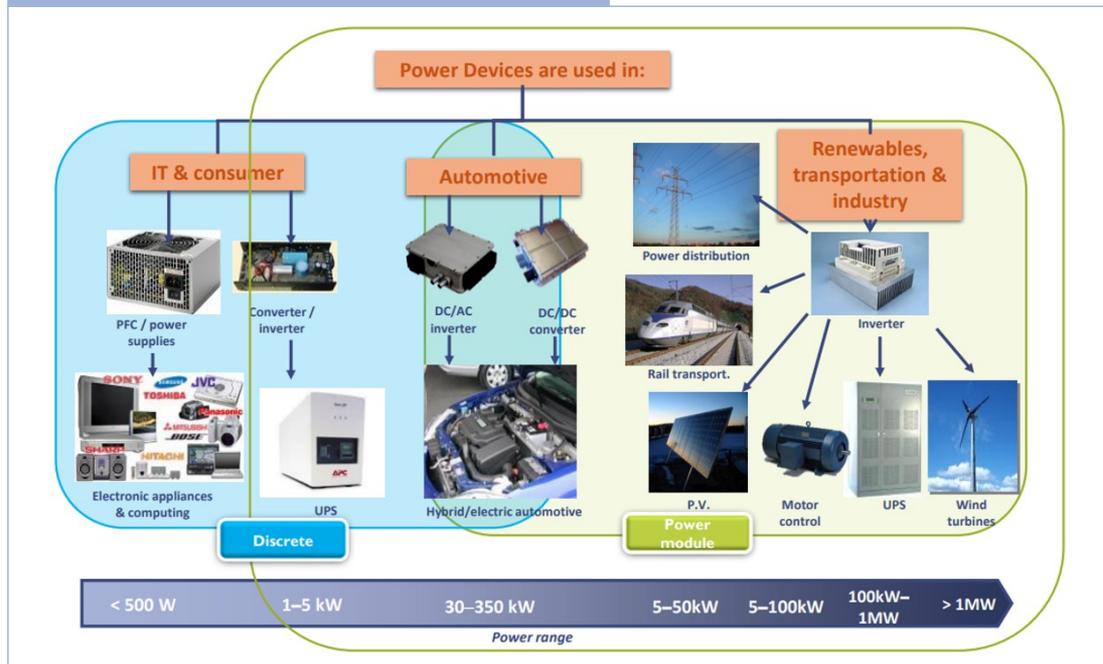
* 출처: 한국전기연구원 블로그

최근 기존의 산업용 모터 인버터, 가전기기 파워 등에서 IT 기기의 발달과 데이터 사용량 증가에 따른 데이터 센터 등의 신규 에너지 수요, 전기자동차 산업의 급속한 성장, 신재생 에너지 등으로 인해 전력 사용량이 폭발적으로 증가하고 있다. 이로 인해 전력 사용량에 대한 우려와 함께 에너지 위기와 환경 규제 강화에 대비하여 전력변환장치 및 전력 시스템의 성능 향상과 에너지 절감의 요구가 높아지고 있다. 특히 탄소 중립 및 환경위기 대응을 위해서 전력에너지 사용의 효율화가 필수적이며, 이산화탄소를 배출하지 않는 전력시스템(전기자동차, 신재생에너지)이 강하게 요구되고 있다. <그림 1>에서 보는 바와 같이 전력의 생산에서부터 소비까지의 전기에너지 사용 흐름에서 전력변환 기술이 필수적으로 사용되며, 전력에너지 사용의 효율화는 결국 전력변환 기술의 혁신을 의미한다. 전력변환 기술의 혁신은 전력변환 장치 자체의 구조 및 회로기술에서도 이루어지지만 많은 부분이 핵심부품인 전력반도체를 통해 이루어지며, 이를 위해서 전력반도체 소자의 혁신 요구가 커지고 있다. 그러나 반세기 넘게 사용되어온 Si(Silicon, 실리콘) 기반의 전력반도체 소자의 성능 향상이 한계에 다다르고 있어, 새로운 대안으로 차세대 전력반도체 소자들이 개발되어, 최근 들어 시장에 점차적으로 도입되고 있다. 융합연구리뷰에서는 차세대 전력반도체 소자로 각광받고 있는 와이드밴드갭(WBG, Wide BandGap) 전력반도체 물질인 SiC(Silicon carbide, 실리콘 카바이드), GaN(Gallium Nitride, 갈륨 나이트라이드)을 기반으로 한 전력반도체 소자 기술에 대해 소개하고, 그 다음의 전력반도체 소자로 각광받고 있는 Ga₂O₃(Gallium Oxide, 산화갈륨) 전력반도체 소자에 대해서 소개한다.

II 차세대 전력반도체 소자 기술

1. 전력반도체 소자와 발전 방향

그림 2. 전력반도체 소자의 용량대별 사용 분야



* 출처: Yole Development(2018)

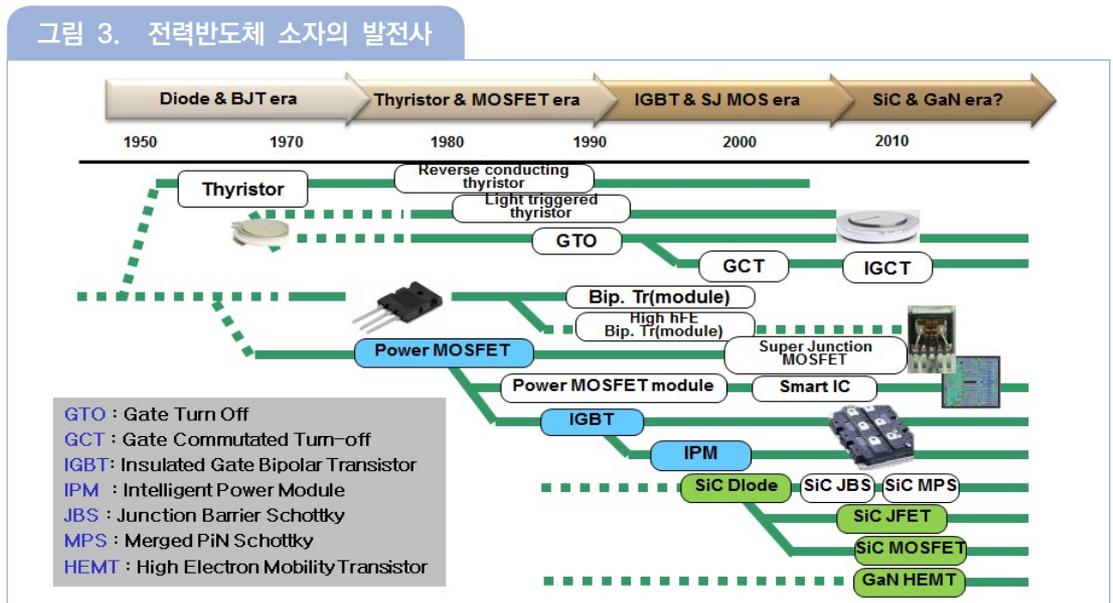
전력반도체(power semiconductor)란 전기 에너지를 활용하기 위해 직류교류 변환, 전압, 주파수 변화 등을 제어하고, 처리하는 반도체이다. 전력의 변화 및 제어에 사용되는 1 [W, 와트] 이상의 전력을 담당하는 스위칭 소자 및 정류소자들을 통칭하는 말로써 직류와 교류의 변환, 변압, 충전, 모터 구동, 전력 안정, 전력 분배 및 제어 등에 사용된다. 사용단계에서는 가전, 스마트폰, 자동차 등 전기로 작동하는 제품의 작동 여부 및 성능을 결정짓는 핵심 부품으로 사용된다.

전력반도체 소자는 전력의 전달 및 제어 과정에서 에너지 효율을 향상시키고 전압의 변화를 제어하여, 시스템 안정성 및 신뢰성을 제공해야 하기 때문에, 일반적인 반도체 소자에 비해 고내압화, 대전류화, 고주파수화 되도록 발전해 왔으며, 이를 식 (1)과 같이 정량화할 수 있다(경신수, 성만영, 2016).

$$\Delta T = R_{TH} \left[(V_{on} \times I_{on}) + \left(\int V(t) \times I(t) dt \right) f \right] \dots\dots\dots (1)$$

식 (1)에서 ΔT 는 전력소모로 발생한 열로 인한 온도 변화를, R_{TH} 는 전력반도체 소자의 열저항을, V_{on} 과 I_{on} 은 온(On) 상태 동작시의 전압과 전류를 나타내며, $V(t)$ 와 $I(t)$ 는 시간에 따른 전압과 전류를, f 는 주파수를 나타낸다.

전력반도체 소자는 전력소모를 줄이기 위하여 저항을 줄이고, 스위칭을 빠르게 하는 방향으로 발전하여 왔고, 안정성을 위해 동작 시 열 발생을 최소화 하는 방향으로 발전해 왔다. 이런 발전 방향에 따라 <그림 3>에서 보듯이 사용처가 요구하는 정격과 동작 주파수에 따라 GTO(Gate Turn-Off thyristor, 게이트 턴 오프 사이리스터), IGBT(Insulated Gate Bipolar Transistor, 절연 게이트 양극성 트랜지스터), BJT(Bipolar Junction Transistor, 쌍극 접합형 트랜지스터), MOSFET(Metal-Oxide-Semiconductor Field-Effect Transistor, 금속 산화막 반도체 전계효과 트랜지스터), 다이오드(Diode) 등 여러 종류의 전력반도체 소자가 개발되어 사용되고 있다.



* 출처: 경신수, 성만영(2016)

2. 차세대 전력반도체를 위한 와이드밴드갭(Wide Bandgap) 물질

전력변환장치의 효율화에 따라 전력반도체 소자의 성능 향상에 대한 요구가 점차 높아지면서 Si 기반의 전력 반도체 소자로는 요구 사항을 만족할 수 없는 상황이 발생하고 있다.

식 (2)는 전력반도체의 온(On) 상태 특성을 나타내는 온 저항($R_{on,sp}$)과 오프 상태 특성을 나타내는 항복전압(V_B , 다이오드 또는 트랜지스터가 파괴되기 전 견딜 수 있는 역방향 전압의 최대 한계) 사이의 관계를 나타낸다(경신수, 성만영, 2016). 온 상태 특성인 저항 $R_{on,sp}$ 는 줄어드는 방향으로 발전하려 하고, 오프 상태 특성인 항복전압 V_B 는 커지는 방향으로 발전하고자 하여 서로 트레이드 오프(trade off, 하나를 움직이면 다른 하나가 반대되는 방향으로 움직이게 되는 관계) 관계이다. 식 (2)의 분모 식이 온 저항과 항복전압 트레이드 오프 관계를 결정짓는데, 이 매개식을 Baliga의 성능지수(BFOM, Baliga's Figure of Merit)라 한다.

$$R_{on,sp} = \frac{4 V_B^2}{\epsilon_s \mu E_C^3} \dots\dots\dots (2)$$

그러나 Si 기반의 전력반도체 소자는 BFOM을 개선시키는 기술 개발에 한계가 있어, 재료 자체의 물성적 특성으로 전력반도체 소자의 성능을 향상시킬 수 있는 WBG 물질을 기반으로 하는 전력반도체 소자들이 대안으로 부상하면서, 2010년대부터 상용화가 시작된 후 최근에는 시장에 급속도로 확장되고 있다.

〈표 1〉에 Si와 WBG 물질인 SiC와 GaN 등의 물성 특성을 도시하였다(Hikita et al, 2009). 〈표 1〉에서 보듯이 Si 또는 GaAs(Gallium Arsenide, 비소화 갈륨)에 비해 SiC와 GaN이 에너지 밴드갭(Eg, band gap energy)이 높아 임계전계(critical electric field) 값이 높다. 이로 인해 BFOM이 SiC와 GaN이 Si 대비 월등히 높은 것을 확인할 수 있다.

또한 SiC의 경우 열전도도(thermal conductivity)가 타 물질 대비 높아 소자의 동작속도와 열방출 성능에 초점을 둔 Keye의 성능지수(KFOM, Keye's Figure of Merit)가 월등히 높은 것을 확인할 수 있다. 이를 통해 SiC와 GaN이 차세대 전력반도체 물질로써 탁월하다는 것을 알 수 있다.

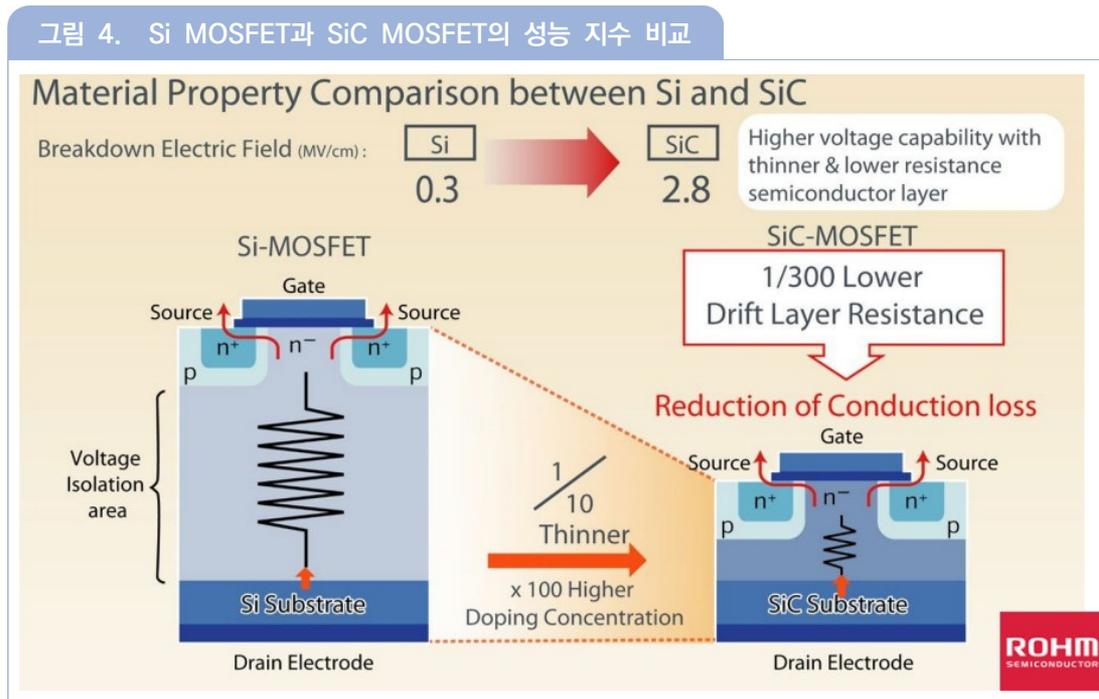
표 1. 반도체 소재의 물성 비교표

구분	Si	GaAs	SiC(4H)	GaN (Hexagonal)	Diamond (C)
에너지 밴드갭(Energy Band Gap) [eV]	1.12	1.42	3.25	3.4	5.45
절연파괴전계 (Critical Electric Field for Breakdown) [MV/cm]	0.3	0.4	3	4	5.6
전자 이동도(Electron Mobility) [cm ² /Vs]	1,500	8,500	1,000	1,250	1,900
유전상수(Dielectric constant)	11.8	12.8	9.7	9	5.5
포화전자속도 (Saturated Electron Velocity) [10 ⁷ cm/s]	1	2	2	2.5	2.7
열전도율(Thermal Conductivity) [W/cm·K]	1.5	0.5	4.9	1.3	20
Baliga의 성능지수 (BFOM, Baliga's Figure of Merit)	1	14.6	548	1,507	3,840
Keye의 성능지수 (KFOM, Keye's Figure of Merit)	1	0.5	5.1	1.6	32.1

* 출처: Hikita et al(2009)

3. SiC 전력반도체 소자

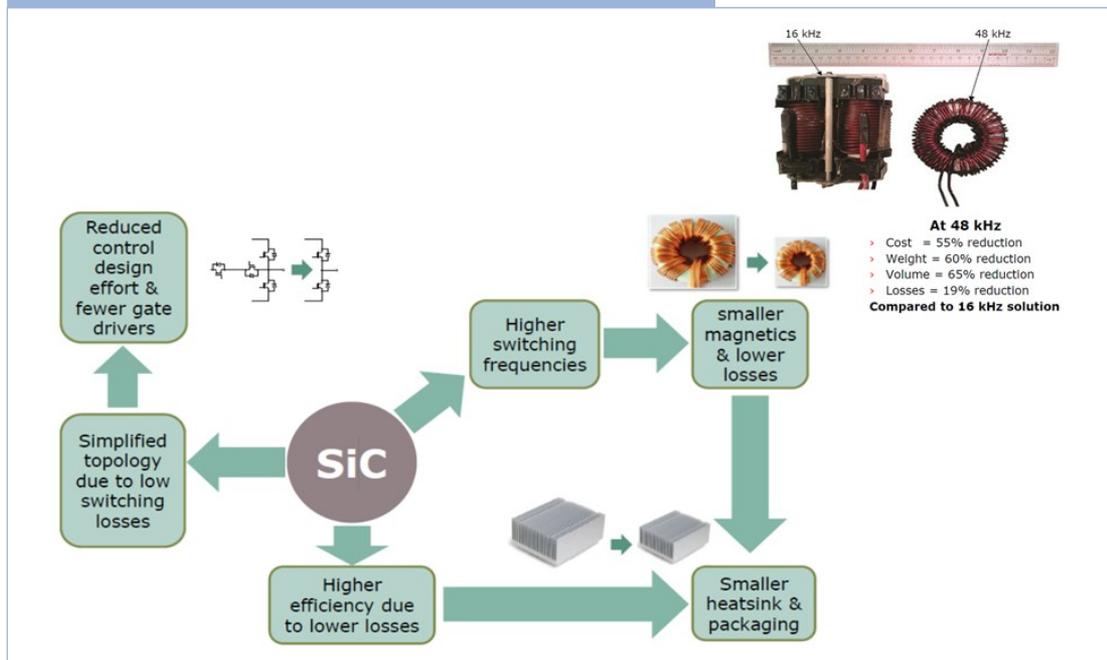
전력반도체 소재로써 SiC의 탁월성이 <그림 4>에 잘 나타나 있다(Mashaly, 2017). <그림 4>는 Si 전력 MOSFET(저비용, 고집적, 저전력, 단순 공정의 반도체 트랜지스터 소자)의 일반적인 수직 구조와 이를 SiC로 만들었을 때의 수직 구조를 비교하고 있다. <그림 4>에서 보듯이 전력 MOSFET이 높은 항복전압을 만들기 위해서는 드리프트 레이어(drift layer)라고 하는 공간의 확보가 필요하다. 오프 상태의 항복전압을 위한 이 드리프트 레이어는 온 상태에서는 저항을 높이는 역할을 하기 때문에 전력반도체 소자의 온 상태 성능을 떨어뜨린다. 이에 반해, SiC는 항복전압을 생성하기 위한 드리프트 레이어 길이를 1/10까지 줄이는 것이 가능해 Si와 같은 항복전압 성능에서 훨씬 작은 온 저항을 구현할 수 있다. 또한 SiC MOSFET은 항복전압을 높이는 것이 용이하여, 일반적으로 600~900V급의 항복전압 레벨을 구현하는 Si에 비해 1200V 이상 10,000~15,000V까지의 항복전압 구현이 가능하다. 이를 활용하여 일반적으로 Si-IGBT(게이트 전압을 통해 전류를 제어하는 전력 장치용 반도체 소자) 소자가 대부분을 차지하고 있는 초고전압 전력반도체 소자 영역에서 탁월한 성능을 보인다.



* 출처 : Mashaly(2017)

일반적으로 Si-IGBT의 구동 주파수는 고전압과 전류 용량대에 따라 10~40kHz 수준이지만, SiC MOSFET의 구동 주파수는 소자적 특성에 기인하여 고전압에서도 50~200kHz까지 가능하다. 이런 특성으로 인해 <그림 5>에서 보는 바와 같이 고주파수에 기인하여 Cap(Capacitor의 줄임말, 전자회로에서 전기를 일시적으로 저장하는 장치로 콘덴서(condenser) 또는 축전기라고도 함)이나 인덕터(Inductor, 자성체에 전선을 감은 단순한 구조의 부품)의 소형화가 가능하며, 전력변환장치의 부피를 줄이는 것이 가능하다. 또한 SiC MOSFET의 낮은 온 저항 특성과 고주파 특성을 통해 에너지 효율을 향상시킬 수도 있다. 더불어서 앞서 언급하였듯이 SiC 소재 자체의 열적 특성이 탁월하여, 열저항 R_{TH} 가 낮아져 전력변환 시스템 상에서 방열 시스템 구축에 더 낮은 비용과 공간이 필요하여 시스템적 효율성이 크다.

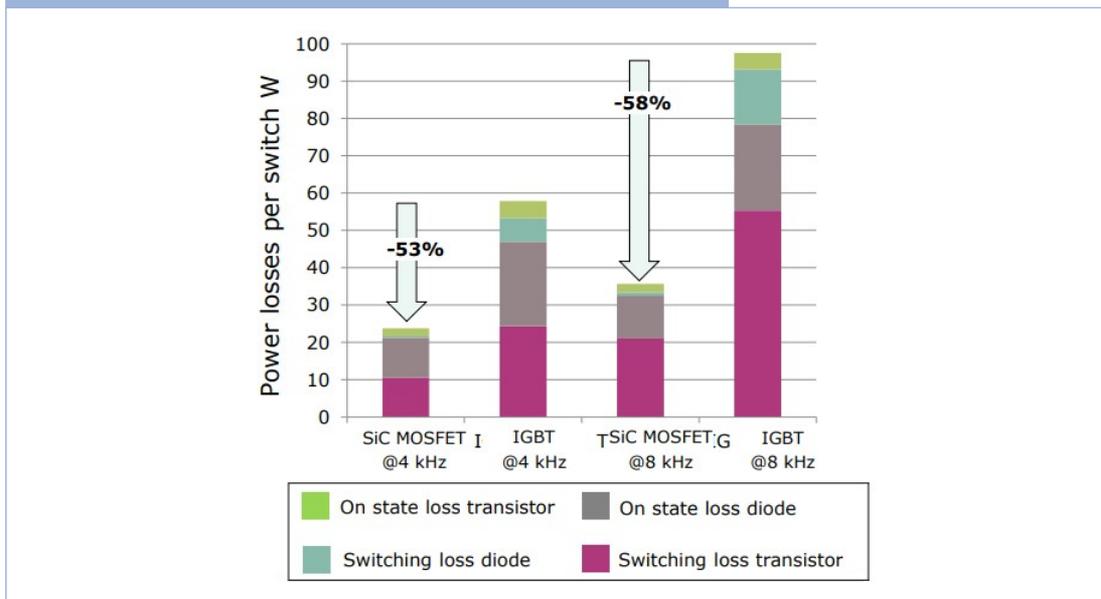
그림 5. SiC 전력반도체 소자의 전력변환 시스템 효율성



* 출처: Levett & Ma(2019)

〈그림 6〉은 Infineon社에서 발표한 SiC MOSFET과 Si-IGBT를 적용했을 때의 전력 소모를 비교한 실험 결과이다. 그림에서 보는 바와 같이 SiC MOSFET이 Si-IGBT에 비해 온 저항과 스위칭 전력 소모가 낮다. 이로 인해, 구동 주파수가 빠른 8kHz 대에서는 스위칭 전력 소모의 큰 차이에 기인하여 더 큰 전력소모 차이를 보이는 것을 확인할 수 있다.

그림 6. SiC MOSFET과 Si-IGBT의 전력소모 비교 실험



* 출처: Levett & Ma(2019)

이런 SiC MOSFET의 효율성에도 불구하고 SiC MOSFET은 상업화를 위해 고비용 문제를 극복해야 하는 과제가 있다. 현재의 Si 전력반도체 대비 최소 2배에서 최대 5배의 가격 차이가 있어 전력변환장치에 적용이 어렵다. SiC MOSFET이 Si 대비 비싼 이유는 원재료인 웨이퍼(wafer, 원통 모양의 실리콘을 잘라 만든 판)의 비싼 가격과 어려운 공정에 기인한다. 그리고 SiC는 녹지 않는 재질의 물질이므로 기존 Si의 초클라스키 잉곳 공법(Czochralski method, 다결정의 실리콘을 녹여 액체상태로 만든 후 단결정 실리콘(seed)를 접촉시켜 잉곳(ingot, 실리콘 기둥)을 성장시키는 방법)이 아닌 물리적 증기 수송법(PVT, Physical Vapor Transport)을 채택하고 있기 때문이다. 또한 SiC 공정의 경우, 〈표 2〉에서 보는 바와 같이 Si에 비해 1,700℃ 이상의 고온 공정을 쓰거나, 게이트 산화막 공정, 에칭 공정 등 고난이도의 공정이 필요해 공정 수율도 어려운 점이 고비용의 결과로 돌아온다.

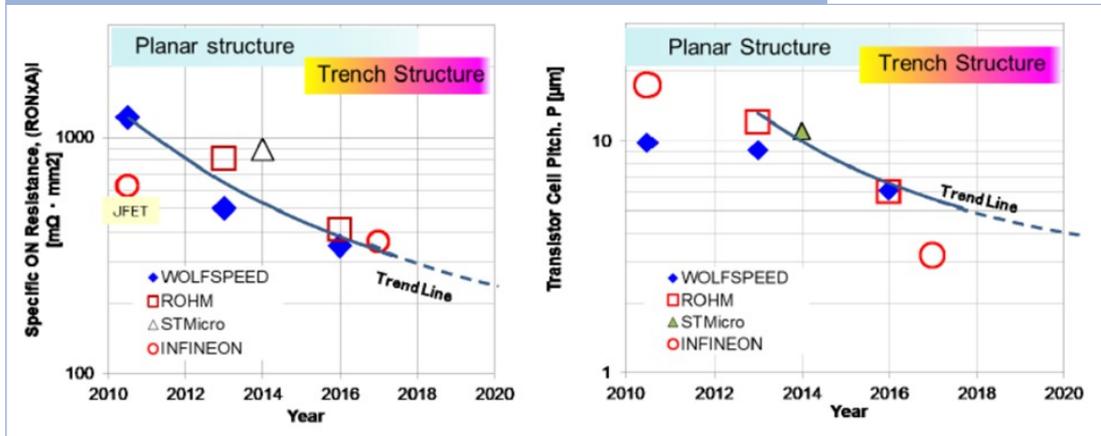
표 2. Si과 SiC 공정 비교

공정		Si	SiC	SiC 공정 이슈
기판 (Substrate)	기판 투명도	불투명	반투명	포토공정 : 포커싱 이슈 Backside reflection Defect Inspection
	공정 방식	초클라스키 (잉곳)	PVT (보울)	생산성 차이 SiC = 느린 성장속도(<2mm/h)
에피(Epitaxy)		< 1,200°C	< 1,500°C	-
도핑 (Doping)	N-type	P, As, ...	N(P)	공정 조절 어려운 소스
	P-type	B	Al(B)	공정 조절 어려운 소스
확산공정		Yes	No	도핑영역이 Flat → 도핑 Implant 공정을 여러 step으로 구성해야 함
어닐링(Annealing)		800~1,200°C	> 1,700°C	초고온 공정
에칭공정		Wet/Dry 가능	Dry만 가능	고경도로 인한 Etch 공정 어려움 Etch Mask 선택비 조절 어려움.
게이트 산화막		Thermal (800~1,200°C)	Thermal (1,500~1,700°C)	게이트 산화막 품질관리를 위한 특별한 공정 필요 (NO, N ₂ O 혼입)
오믹 공정(Ohmic)		< 400°C	< 1,000°C	Metal 녹는 점을 고려한 Metal 선택 필요

* 출처: Levett & Ma(2019)

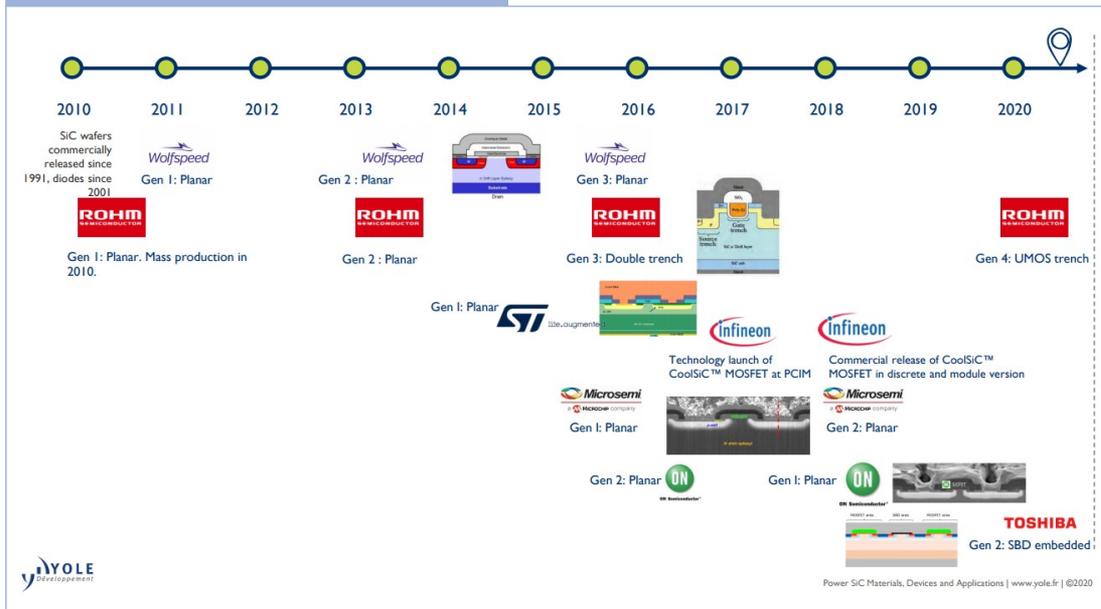
따라서 SiC MOSFET의 상업화를 위해서는 상기에서 언급한 고가의 웨이퍼 비용 및 공정 비용을 극복할 수 있도록 웨이퍼의 생산성을 높이는 기술 개발이 필요하다. 이에 대해 유수의 기업들은 SiC MOSFET의 단위 셀(Cell)의 크기를 축소하는 방식으로 기술 개발을 진행했다. <그림 7>에서 보듯이 초기에 Wolfspeed(구 Cree社에서 SiC 사업부분이 분사된 기업)社에서 처음으로 플래너 게이트(planar gate) 구조로 상업화를 성공한 이래 일본의 Rohm社와 독일의 Infineon社 등이 차례로 상업화에 성공하였다. 특히, 2016년 이후에는 기존의 planar gate보다 단위 Cell의 크기를 줄일 수 있는 트렌치 게이트(trench gate) 공정을 적용하여 보다 더 발전된 형태의 SiC MOSFET을 구현하였다. 상용화된 SiC MOSFET의 공정 구조들을 도시한 <그림 8>에서 보는 바와 같이, Planar Gate MOSFET의 Cell Pitch를 6 μ m 수준까지 줄여 4m Ω .cm²까지 줄이거나, trench gate 공정을 적용하여 Cell Pitch를 3 μ m 수준까지 줄여 4m Ω .cm² 이하까지 확보하기도 하였다.

그림 7. SiC MOSFET의 기술 개발 현황($R_{on,sp}$ 와 단위 Cell 크기)



* 출처: LTEC Co.(2019)

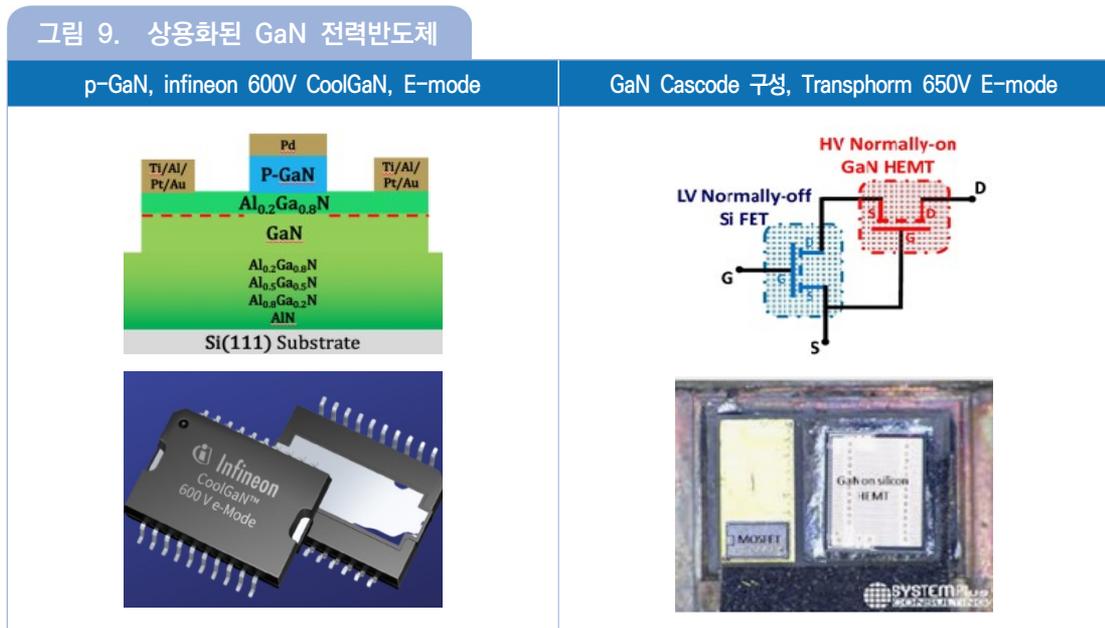
그림 8. 상용화된 SiC MOSFET 현황



* 출처: Yole Development(2020)

4. GaN 전력반도체 소자

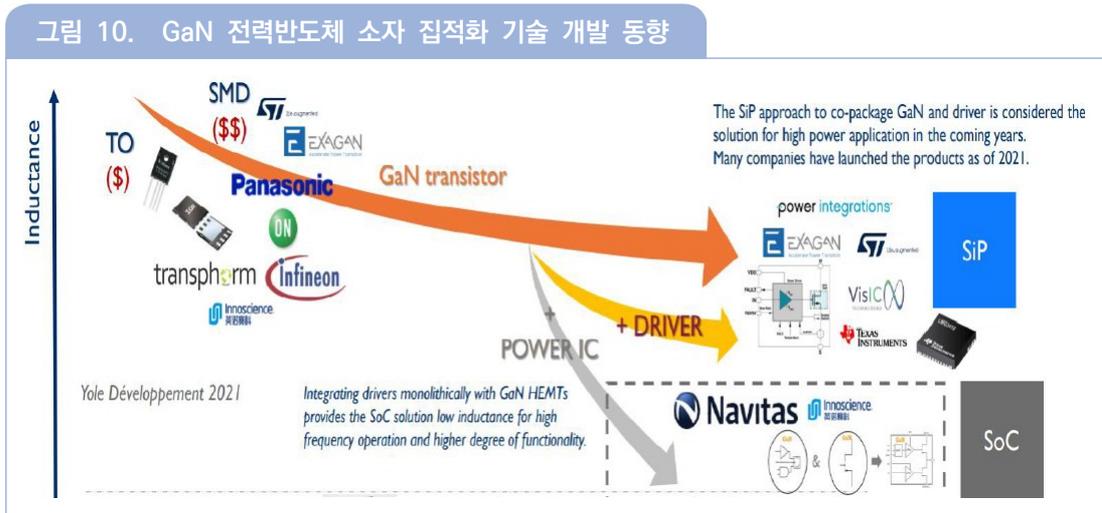
GaN 소재는 앞서 <표 1>에서 확인할 수 있듯이, KFOM은 SiC 소자보다 떨어지지만, 높은 전자 이동도 특성으로 인해 BFOM이 높아 고효율 전원 공급 장치 및 소형/고주파 응용 분야에서 이점을 가지는 전력반도체 소자를 제작할 수 있다. 현재 상용화된 GaN 소자는 모두 수평형 소자이며, <그림 9>에서 보는 바와 같이 Normally-off 특성을 가진 E-mode(Enhancement-mode, 증가형) 소자는 650V급까지 상용화되었으며, D-mode(Depletion-mode, 공핍형) 소자는 Si 소자를 캐스코드(cascode) 방식으로 조합한 형태로 상용화되고 있다.



* 출처: (좌) 전자과학(2018), (우) Power Electronic Tips(2022)

GaN 전력반도체 소자가 BFOM이 높음에도 불구하고, 650V급의 항복전압으로만 상용화가 되는 이유는 GaN 소재에서는 아직 도핑 공정이 확보되지 않았기 때문이다. 즉, N-type의 GaN 레이어(layer)와 P-type의 GaN layer로 적층 구조는 제작을 할 수 있지만, 항복전압을 향상시키는데 있어 필수적인 엣지 종단(edge termination) 구조 또는 전계 분산 구조를 도핑구조를 통해 만들기 어려워 650V급보다 상향된 소자를 효율성 있게 제작하기가 어렵다. 또한 매우 높은 전자 이동도를 가진 HEMT(High Electron Mobility Transistor) 소자 구조를 구현할 수 있지만, 현재는 Normally-on 상태로 게이트 구동 전압이 음의 전압이 되므로, 이를

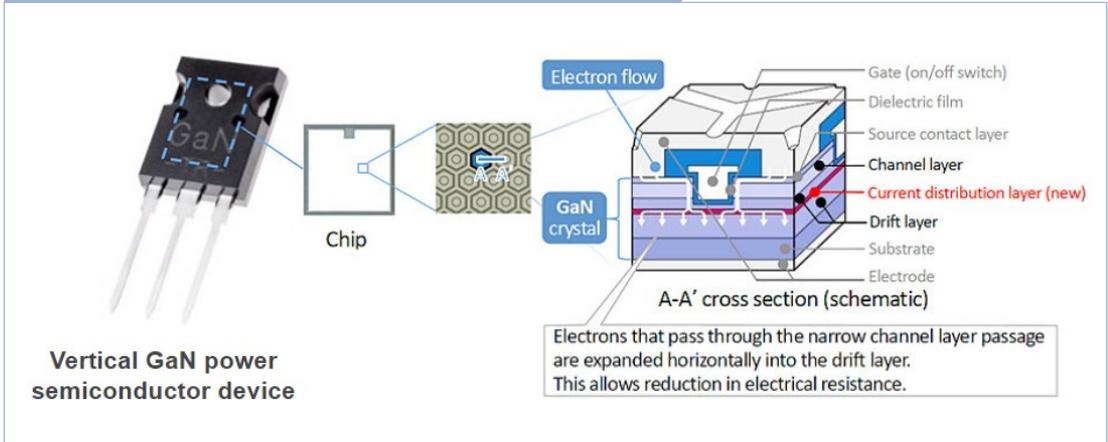
양의 전압으로 조절하기 위해 Si 소자를 연결하여 전력 모듈에 적합한 증가형(E-mode) 소자를 구현하기도 한다. 그러나 GaN 소자 특유의 높은 전자 이동도와 높은 스위칭 구동 능력에도 불구하고, Si MOSFET과 연결하는 부위에서 발생하는 기생 인덕턴스(parasitic inductance) 성분으로 인해 스위칭 구동 능력과 효율이 제한되는 문제가 있어, 이를 해결하기 위해 GaN 전력반도체 소자는 고효율 패키징 기술 또는 GaN IC(Integrated Circuit)가 집적된 SoC(System on Chip, 여러 가지 기능을 가진 시스템을 하나의 칩으로 구현한 기술이 집중된 반도체) 형태의 기술로 발전하고 있다.



항복전압 1,000V 이상의 고전압 응용을 위한 GaN 전력반도체 소자는 수직형 GaN 전력소자로 구현이 가능하지만, 다이오드 이외에 FET(Field Effect Transistor, 전계효과 트랜지스터) 기술에서 기술적 난관이 크며, GaN 기판 가격과 해결해야 할 기술적 문제점으로 단기간 시장진입은 어려울 것으로 예측되고 있다. 현재 수직 GaN on GaN 구조의 기판은 4인치까지만 상용화된 상태이며, 고가의 기판 가격을 형성하고 있다. 또한 앞서 언급했던 대로 P-type 도핑(반도체의 전기전도도를 높이기 위해 순수한 일반 반도체에 불순물을 투입하는 과정)이 어렵기 때문에 고전압 형성을 위해 드리프트 레이어 층의 농도를 저농도화 하는 형태로만 고전압 구현이 가능해 실효성이 낮은 상태이다.

일본의 Toyoda Gosei社에서 2019년 1,200V/100A급의 수직형 GaN 전력반도체를 개발하였으며, 이는 기존 WBG 소자보다 크기가 작고 스위칭 속도가 빠르다(>10MHz)고 발표하였다.

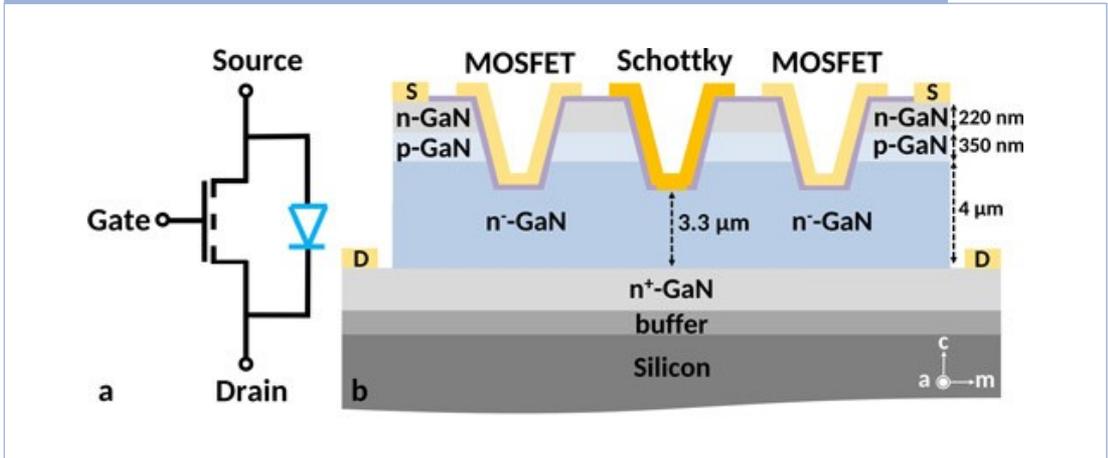
그림 11. Toyoda Gosei社의 GaN 수직형 전력 소자



* 출처: Toyoda Gosei社 사이트

미쓰비시社에서는 GaN 단결정 기판 위에 PiN 구조와 쇼트키(Schottky) 구조를 동시에 갖는 수직형 MPS(Merged PiN) 다이오드와 누설전류 감소 및 2 kV 이상의 고 항복전압 특성에 대한 연구 결과를 발표하였다.

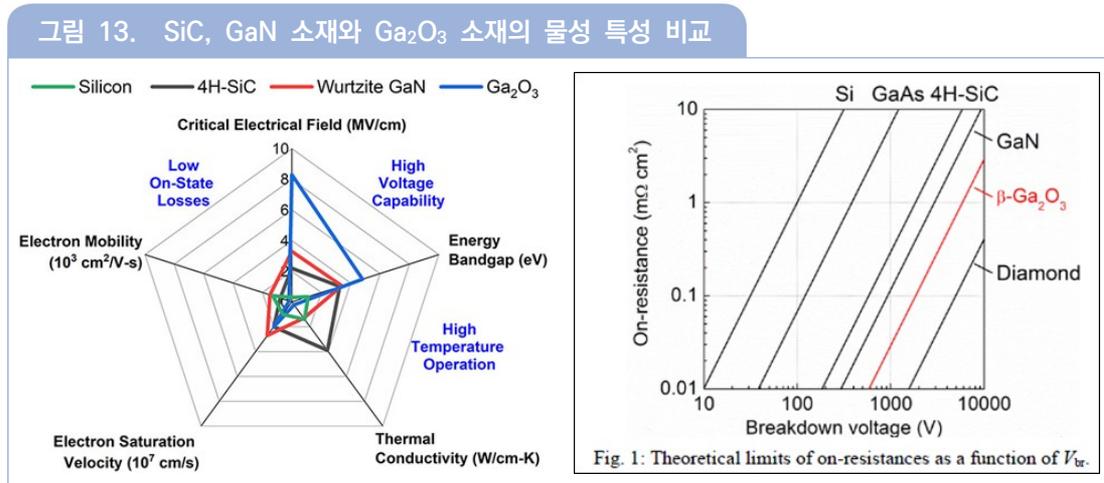
그림 12. GaN-on-Si 웨이퍼를 이용한 Quasi vertical GaN MOSFET 소자



* 출처: Liu et al(2018)

5. Ga₂O₃ 전력반도체 소자

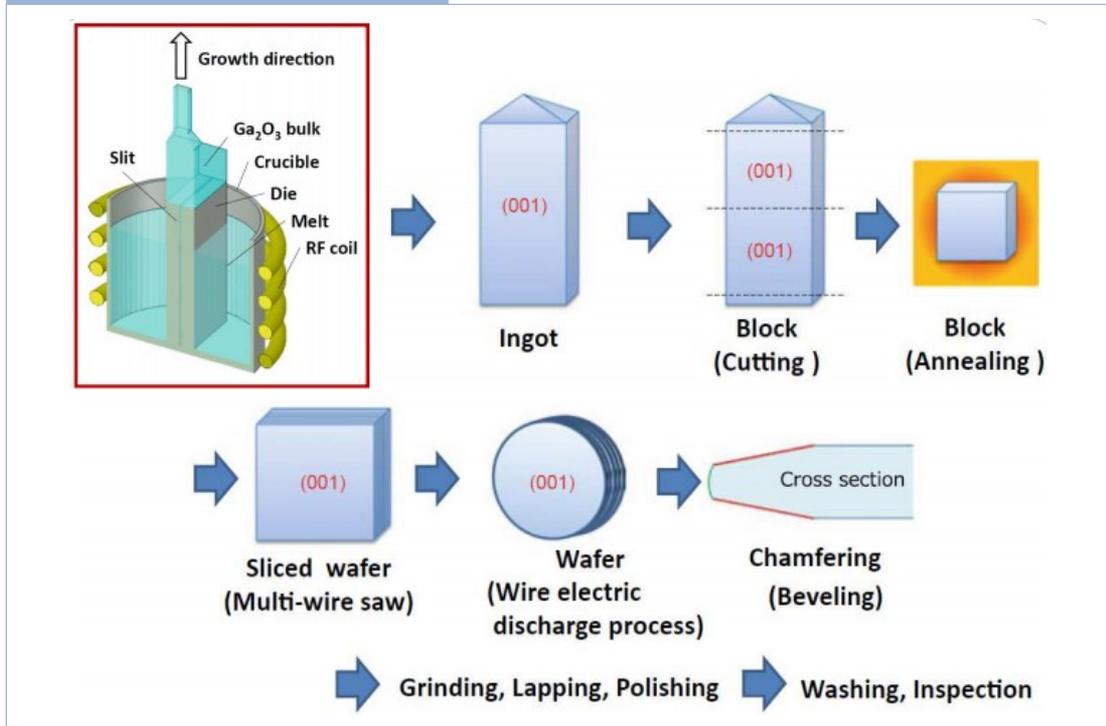
앞서 살펴보았던 SiC 전력반도체 소자와 GaN 전력반도체 소자는 WBG 소재의 물성을 기반으로 한 전력반도체 로써 상용화에 성공하였다. 그러나 소재 자체의 가공 특성과 공정적 어려움으로 인한 높은 가격으로 시장 확장에는 어려움을 겪고 있다. SiC, GaN 전력반도체 소자의 높은 가격 문제를 소재적 장점으로 극복하기 위한 차세대 전력반도체 소재로 Ga₂O₃(산화갈륨) 소재가 있다. Ga₂O₃은 <그림 13>에서 보는 바와 같이 기존의 SiC와 GaN 보다도 더 큰 5eV에 가까운 에너지 밴드갭 특성과 그에 따른 임계전계 값을 가지고 있다(Pearnton et al, 2018). 이로 인해 SiC, GaN에 비해 고전압에 유리한 WBG 소재이다.



* 출처: Pearnton et al(2018)

전력반도체로써의 유리한 특성과 함께 Ga₂O₃은 웨이퍼 제작에 있어서 SiC, GaN과 다르게 대량 생산할 수 있는 기판 제조 공정 방법을 확보하고 있다. <그림 14>에 Ga₂O₃의 양산화된 기판 제조 공정을 도시하였다. EFG(Edge Film Growth) 공정으로 일본의 Noble Crystal社에서 상용화에 성공하여 최근 4인치 웨이퍼까지 판매하고 있다. EFG 방법은 Si의 초클라스키 공법과 유사하게 대형 단결정 기판 판상 잉곳을 제조할 수 있다. 그림에서 보는 바와 같이 용융된 Ga₂O₃을 판상으로 뽑아낸 후 적절한 크기로 가공하여 웨이퍼 기판을 제조할 수 있다. 이로 인해 Ga₂O₃ 웨이퍼의 가격은 향후 SiC 웨이퍼의 20% 수준이 될 것으로 예측되고 있다.

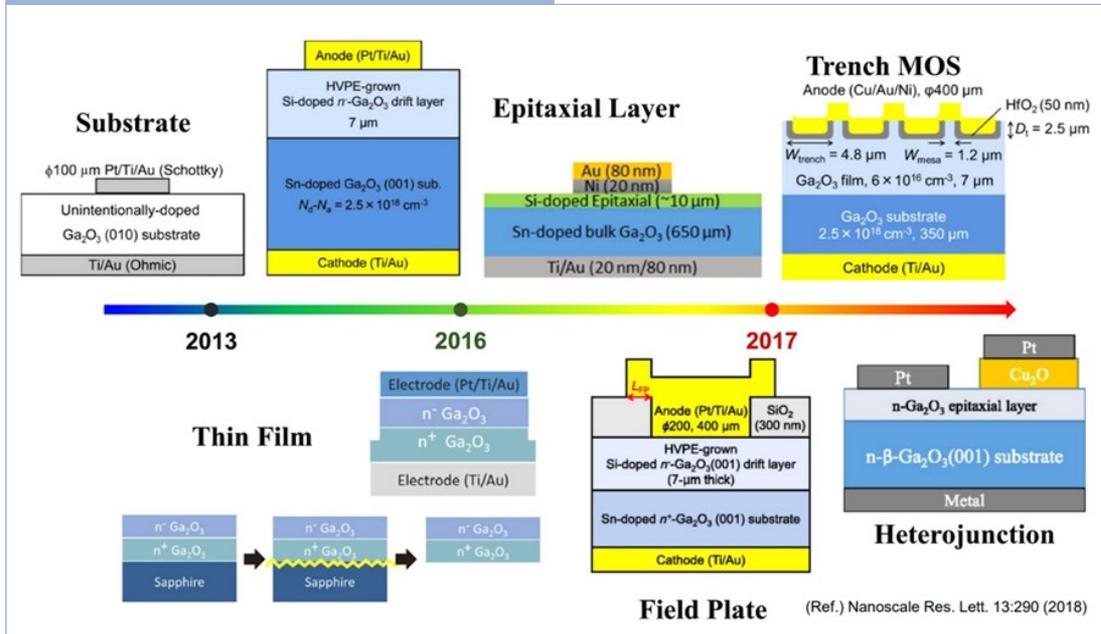
그림 14. Ga₂O₃ 기판 제조 공정



* 출처: 배시영(2021)

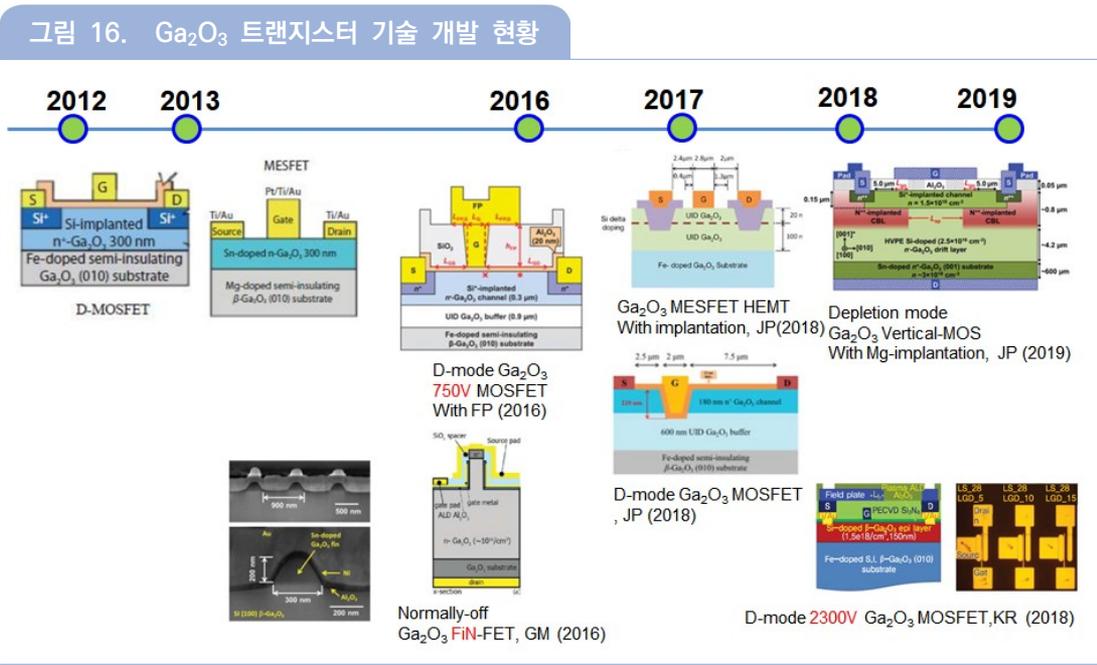
또한 Ga₂O₃는 SiC, GaN 대비 재료의 가공이 상대적으로 용이하여 전력용 쇼트키 배리어 다이오드(SBD Shottky Barrier Diode, 금속 전극과 탄화규소의 접촉면에 형성되는 쇼트키 배리어를 이용해, 정류 작용을 실현시키는 소자)와 전력용 트랜지스터까지 많은 연구개발이 이루어지고 있다. 초기에 SiC나 GaN 다이오드가 개발될 때와 마찬가지로 SBD에 대한 연구부터 최근에는 산화구리(CuO, Copper Oxide) 또는 니오(NiO) 등의 P-type 이종접합 물질을 적용한 PN Diode 연구 결과들까지 발표되고 있다(Xue et al, 2018).

그림 15. Ga₂O₃ 다이오드 기술 개발 현황

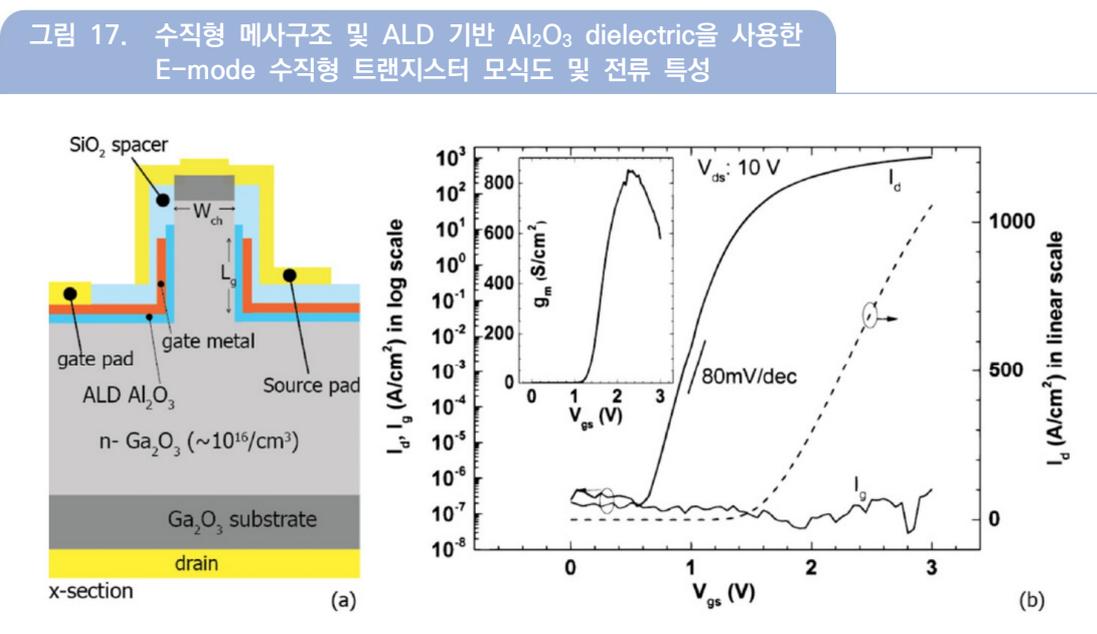


* 출처: Xue et al(2018)

Ga₂O₃ 트랜지스터 또한 많은 연구개발이 이루어지고 있다. 초기에는 Schottky 접합 게이트 채널 제어를 활용한 MESFET(Metal-Semiconductor Field Effect Transistor, 금속-반도체 전계효과 트랜지스터) 연구가 이루어졌으나, Normally-on 소자가 주로 만들어졌다. 이후에 FinFET 공정을 활용한 최초의 Normally-off 제어 Transistor 연구결과가 발표되기도 하였다(Dong et al, 2019; Hu et al, 2018).



* 출처: Dong et al(2019)



* 출처: Hu et al(2018)

III 차세대 전력반도체 국내·외 정책 동향

1. 해외 동향

앞서 살펴본 SiC, GaN, Ga₂O₃ 등 WBG 물질을 기반으로 하는 차세대 전력반도체 기술의 발전은 개별 기업 또는 대학 수준에서 개발되기에는 규모와 기술 난이도에서 어려움이 많다. 이런 어려움을 극복하고 차세대 전력반도체 산업에서 기술을 선도하기 위해 미국, 유럽, 중국, 일본 등의 국가들은 WBG 반도체 기반의 전력반도체 육성 정책을 펴고 있다.

미국은 에너지부(DOE, Department of Energy) 첨단제조 사무국(Advanced Manufacturing Office)에서는 WBG 반도체를 개발하는 ‘파워 아메리카(PowerAmerica) 사업’을 2014년에 출범하여 지속적으로 차세대 전력반도체 산업을 지원하고 있다. PowerAmerica의 중점 사업 목표는 SiC 및 GaN 전력반도체 개발이며, 현재 45개 기업 및 18개 대학이 참여하고 있다(KISTEP, 2017).

유럽연합에서는 9개국 26개의 파트너로 이루어진 ‘Ultimate GaN’이라는 연구 컨소시엄을 구성하여 유럽연합 주관의 민·관 지원기관으로부터 약 4,800만 유로의 펀딩을 조성하여 GaN 전력반도체 연구 협력을 대규모로 추진하고 있다(KISTEP, 2017).

일본은 2014년도부터 전략적 이노베이션 창조프로그램(SIP, Cross-ministerial Strategic Innovation Promotion Program) 중 ‘파워일렉트로닉스 사업’을 통해 일본 기업의 세계 시장 점유율 40%(매출액 1조 7,000억엔)을 목표로 WBG 기반 소재, 전력반도체, 소자/모듈 및 응용기술 개발을 추진하고 있다. 특히 일본은 SiC, GaN 이외에 차세대 전력반도체 물질인 산화갈륨을 지원 내역에 포함시킨 바 있다(KISTEP, 2017).

중국은 ‘14차 5개년 계획 및 2035 중장기 목표’에서 반도체를 중점 과학 기술분야로 선정하고, 2021년 양회에서는 3세대 반도체(화합물 반도체)를 ‘7대 첨단 과학기술’에 포함시켜 자립화한다는 계획을 발표하였다. 이 계획에는 SiC와 GaN 등 3세대 반도체의 개발과 반도체 가공라인 건설 등 1,000억 달러 규모의 펀드를 조성하는 자강책이 포함되어 있다.

이처럼 글로벌 주요 국가들은 차세대 전력반도체의 성장 가능성과 시장 폭발력을 인지하고, 선제적으로 지원 정책을 추진하고 있다(관계부처 합동, 2021).

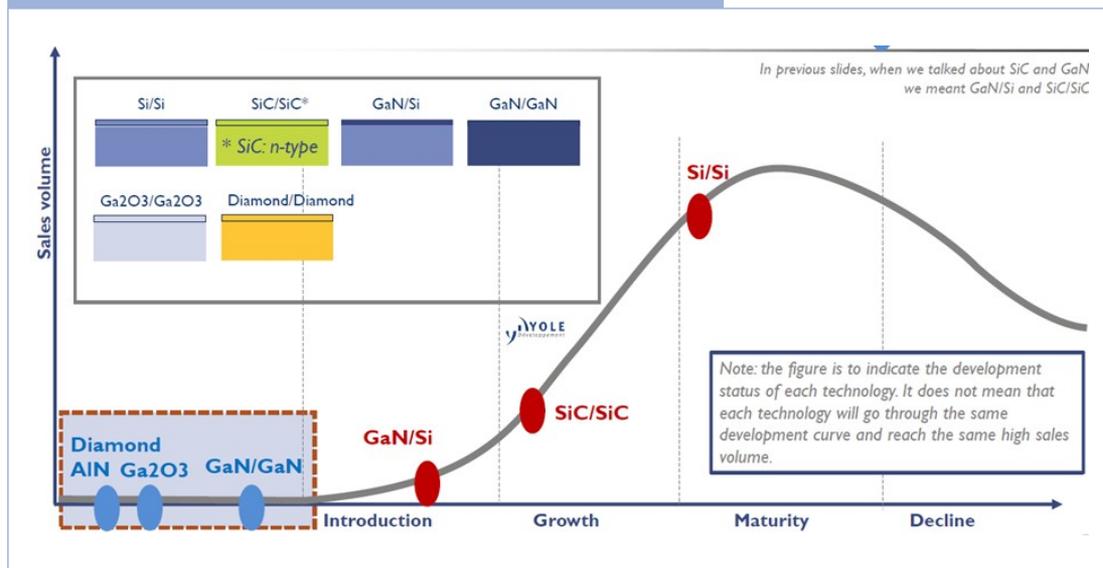
2. 국내 동향

우리나라 정부도 전력반도체의 시장 성장성과 중요성을 인지하고, 전력반도체 기술력 제고를 위해 2017년부터 총 830억 원 규모의 '신산업 창출 파워반도체 상용화 사업'을 추진 중이다. 파워반도체 상용화 사업을 통해 Si 외에도 SiC, GaN 전력반도체의 성장 기초를 마련하고, 제품 기술을 개발하는 것을 지원하는 한편, 6인치 SiC 전력반도체 전용 제작 시설 인프라(부산 파워반도체 상용화 센터)를 구축하였다. 2021년에는 '2050 탄소중립' 선언에 맞춰 2030년까지 반도체 선두국가의 자리를 굳히는 13억 달러 규모의 K-벨트 반도체 정책을 발표하였다(연합뉴스, 2022). K-벨트 반도체 정책에는 차세대 전력반도체 산업 생태계 구축을 위한 '차세대 전력반도체 기술개발 및 생산역량 확충 방안'의 세부 정책이 포함되었다. 더불어 산업통상자원부에서는 '신산업 창출 파워반도체 상용화 사업'의 후속 사업으로 '파워반도체 고도화 사업(가칭)'을 계획 중에 있다. 산화갈륨 반도체와 관련해서는 한국반도체협회에서 2021년 K-GOAL(한국 산화갈륨 얼라이언스)를 출범하여 소재-부품-시스템으로 연계되는 연구개발을 지원하는 역할을 수행하고 있다.

IV 결론

지금까지 차세대 전력반도체 소자로 각광받고 있는 SiC, GaN 전력반도체 소자 기술의 현황에 대해 설명하고, 그 이후의 전력반도체 소자로 기대되고 있는 Ga₂O₃ 전력반도체 소자에 대해 소개하였다. <그림 18>에 차세대 전력반도체인 SiC, GaN과 Ga₂O₃ 전력반도체 소자 기술의 기술 주기를 나타내었다. SiC와 GaN 전력반도체 소자는 기술 이슈가 해결되어 시장 상용화가 진행되고 성장단계에 있다. 이후의 차세대 전력반도체 소재인 Ga₂O₃의 경우 초기 원천기술 단계에 머물러 있지만, 기술 성장속도가 매우 빠른 상황이다.

그림 18. 차세대 전력반도체 소자의 기술 개발 주기 현황



* 출처: Yole Development(2018)

현재 차세대 전력반도체가 주목받고 있는 이유는 전 세계적으로 신규 전력 수요 시장의 성장 및 인구 증가로 전력수요가 급상승하고 있고, 에너지 위기와 기후위기 등으로 전력 사용에 대한 효율화와 탄소중립 이슈가 증대되고 있는 상황에서 전력반도체 소자의 혁신 성장이 새로운 돌파구가 될 수 있을 것으로 기대되고 있기

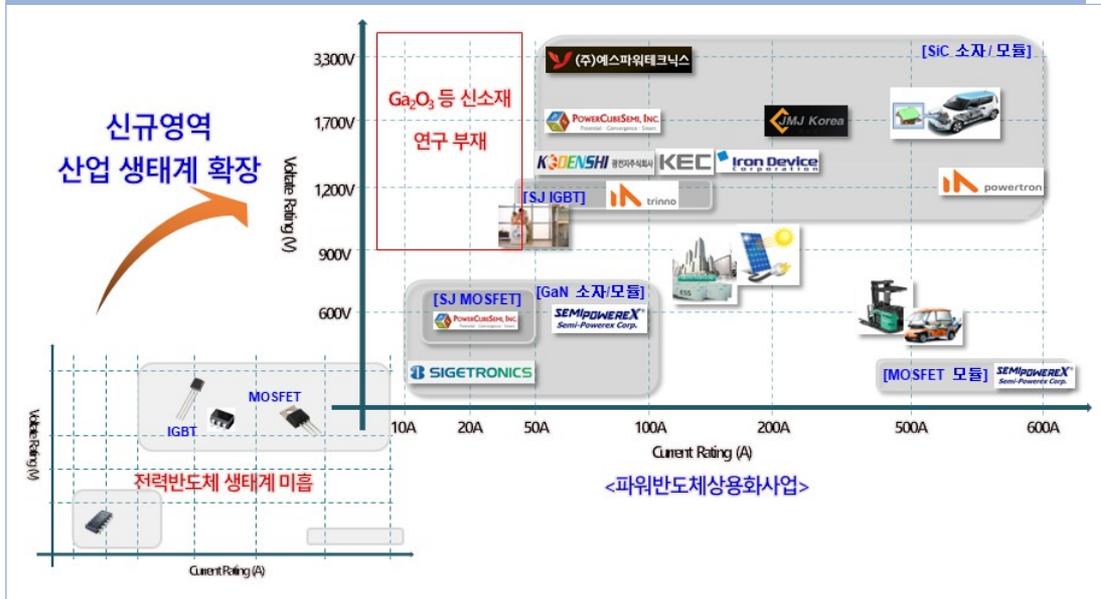
때문이다. 그러나 앞서 살펴보았듯이 새로운 혁신적인 제품이 시장에 안착되기 위해서는 시장 적용 시간과 제품의 최적화 등이 필요하다. 현재 SiC와 GaN 전력반도체 소자의 기술적 이슈들은 대부분 해결된 상황에서 시장에 안착하기 위한 신뢰성 개선, 가격 효율성 개선 등의 방향으로 기술개발과 제품 개선이 이루어지고 있다. 이를 통해 제품의 신뢰성과 가격 효율성이 통하는 전기자동차나 신재생 에너지 시장에서는 SiC, GaN 전력반도체 소자의 사용량이 급격하게 증가되고 있다.

또한 새로운 전력반도체 소재로 급부상하고 있는 Ga_2O_3 의 경우는 현재 초기 원천기술에 가깝지만, 초기 상용화 제품이 발표되고 있고, 상용화에 근접한 연구결과들이 발표되고 있어 다른 어떤 소재들보다 기술개발 속도가 빨라 빠른 시장 상용화를 기대하고 있는 상황이다.

전 세계적으로 주목받고 있는 전력반도체 소자 산업에서 국내 산업계는 메모리 반도체 산업에 치여 불모지와 같은 수준으로 외면 받고 있었다. 그러나 2017년 발의된 ‘신산업 창출 파워반도체 상용화 사업’으로 국가적으로 대규모 연구개발 및 인프라 구축 지원이 이루어져 국내 전력반도체 산업 생태계가 지난 5년 동안 부흥할 수 있었다.

〈그림19〉에 전력반도체 상용화 사업을 통해 성장한 국내 전력반도체 기업들을 도시하였다. 기존에 2~3개의 중소기업만 있었던 전력반도체 산업 생태계는 현재 20여 개의 전력반도체 기업이 성장하는 발판이 되었다. 더불어 부산에 SiC 전력반도체 전용 Fab(제조라인)과 신뢰성 평가 센터 등 인프라 및 기반 구축도 진행되어 연구개발 저변이 확대되었다. 향후에도 이를 바탕으로 국내 전력반도체 산업이 더욱 성장하기를 기대한다.

그림 19. 파워반도체 상용화 사업을 통한 SiC, GaN 전력반도체 분야 국내 기업들의 성장



* 출처: 한국반도체산업협회 사이트

저자_ 경신수(Sin Su Kyoung)

- 학력
 - 고려대학교 전기전자공학 박사
 - 고려대학교 전기전자공학 석사
 - 고려대학교 전기전자전파공학 학사

- 경력
 - 現) 파워큐브세미(주) 연구소장
 - 前) 삼성전자 System LSI 선임연구원

참고문헌

〈국내문헌〉

- 1) 경신수, 성만영. (2016). 파워 디바이스 트렌드와 전망, 반도체설계교육센터 IDEC Newsletter, Vol. 230.
- 2) 한국과학기술기획평가원. (2017). 2015년도 예비타당성조사 보고서-신산업 창출 파워반도체 상용화 사업.
- 3) 관계부처 합동(2021). K-반도체 대규모 예타사업 본격 추진방안. 산업통상자원부.

〈국외문헌〉

- 4) Baliga, B. J.. Fundamentals of Power Semiconductor Devices. Springer Link. 2009.
- 5) Dong, H., Xue, H. W., He, Q. M., Qin, Y., Jian, G. Z., Long, S. B., Liu, M., (2019). Progress of power field effect transistor based on ultrawide bandgap Ga₂O₃ semiconductor material. Journal of Semiconductors, 2019, 40(1), 011802, <http://doi.org/10.1088/1674-4926/40/1/011802>.
- 6) Hikita, M., Manabu, Y., Yasuhiro, U., Tetsuzo, U., Tsuyoshi, T., Daisuke, U. (2009). GaN-Based Power Devices. Panasonic Technical Journal, 55(2).
- 7) Hu, Z., Nomoto, K., Li, W., Zhang, Z., Tanen, N., Thieu, Q. T., Sasaki, K., Kuramata, A., Nakamura, T., Jena, D., Xing, H. G., (2018). Breakdown mechanism in 1 kA/cm² and 960 V E-mode β -Ga₂O₃ vertical transistors. Applied Physics Letters, 113, 122103, <https://doi.org/10.1063/1.5038105>.
- 8) Liu, C., Khadar, R. A., Matioli, E. (2018). GaN-on-Si Quasi-Vertical Power MOSFETs. IEEE Electron Device Letters, Vol. 39, p. 71.
- 9) Mashaly, A. (2017.02.22). The Creation and Potential Cell Structures of SiC Devices. EE POWER, <https://eepower.com/technical-articles/the-creation-and-potential-cell-structures-of-sic-devices/>
- 10) Pearton, S. J., Ren, F., Tadjer, M., Kim, J. (2018). Perspective: Ga₂O₃ for Ultra-High Power Rectifiers and MOSFETS. Journal of Applied Physics, 124, 220901.
- 11) Xue, H., He, Q., Jian, G., Long S., Pang, T., Liu, M. (2018). An Overview of the Ultrawide Bandgap Ga₂O₃ Semiconductor-Based Schottky Barrier Diode for Power Electronics Application. Nanoscale Research Letters, 13: 290, <https://nanoscalereslett.springeropen.com/articles/10.1186/s11671-018-2712-1>

〈기타문헌〉

- 12) 배시영, 2021 K-GOAL 세미나 자료 2021, 한국세라믹기술원
- 13) 박상돈. (2022.07.01). 340조 투자해 반도체 초강대국 달성...세제 혜택-반도체 아카데미 신설, 연합뉴스, <https://n.news.naver.com/mnews/article/001/0013325396>.
- 14) 박종배. (2018.22.20). 인피니언, 갈륨 나이트라이드(GaN) 솔루션 양산 시작. 전자과학, <https://elec4.co.kr/article/articleview.asp?idx=21926>.
- 15) 한국전기연구원 블로그, https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=keri_on&logNo=220903597687.
- 16) 한국반도체산업협회 사이트, <https://www.ksia.or.kr/index.php>.
- 17) Ben-slimane, A., Chiu, P. (2021). GaN Power 2021: Epitaxy, Dvices, Applicatios Technology Trends. Yole Development.
- 18) Dogmus, E., Ben-slimane, A., Chiu, P. (2020). Power SiC Materials, Devices, and Application. Yole Development.
- 19) Levett, D. & Ma, K. (2019). APEC 2019 Professional Education Seminar 자료 : Ten Most-Commonly Asked Questions on Migrating from Si to SiC MOSFET-Based Converter Designs.
- 20) LTEC Corporation. (2019). 1200V SiC MOSFETs Technology Assessment by Means of Deep Structural and Physical Analysis. APEC 2019 Industry Session.
- 21) Shepard, J. (2022.02.15). What is d-GaN, e-GaN and v-GaN power?, Power Electronic Tips, <https://www.powerelectrontips.com/what-is-d-gan-e-gan-and-v-gan-power-faq/>.
- 22) Villamore, A., Rosina, M., Suarez, A. F. (2018). Status of the Power Electronics industry 2018. Yole development.
- 23) Toyoda Gosei社 사이트. (2019). Toyoda Gosei Develops Vertical GaN Power Device With Current Operation of 100 Amperes. <https://www.toyoda-gosei.com/news/detail/?id=228>.



02

프로세싱 인 메모리 기반 뉴로모픽 기술 동향

박성모(한국전자통신연구원 책임연구원)
박필재(한국전자통신연구원 책임연구원)

I 서론

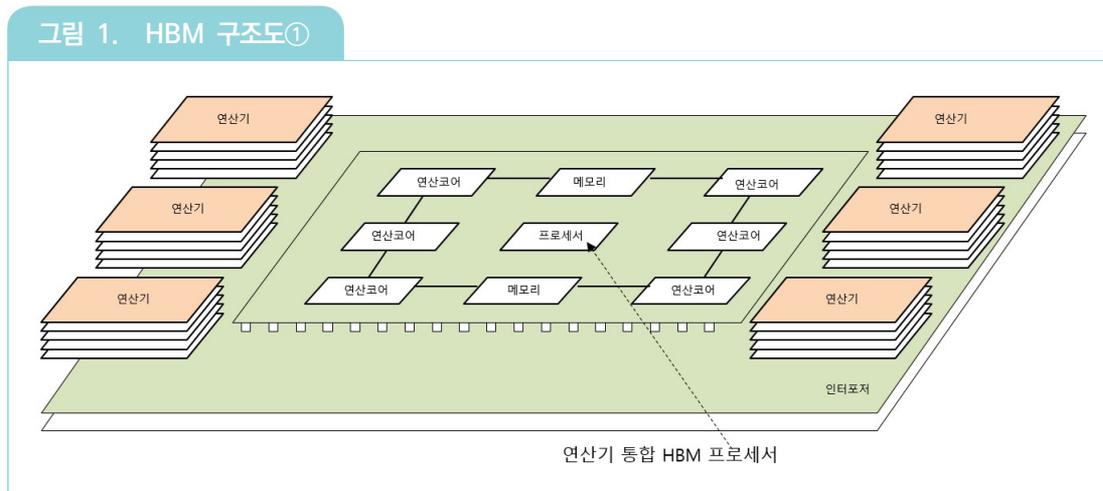
다양한 방법으로 뇌를 닮은 인공지능 반도체(뉴로모픽, Neuromorphic)에 대한 연구개발이 이루어지고 있다. 뉴로모픽의 아키텍처는 초기에는 연산과 저장이 분리되어 하나의 반도체가 모여서 연산을 수행하는 형태로 진행이 되었고, 그 후에는 공정기술이 발전하고 메모리의 집적도가 향상되면서 새로운 서비스에 대한 솔루션이 필요하게 되었다. 기존의 폰 노이만 아키텍처(Von Neumann architecture)는 메모리가 명령어 및 연산자의 저장을 수행하고 프로세서는 메모리에서 명령어와 연산자를 가져와서 연산을 수행하였다. 그러나 프로세싱 인 메모리(PIM, Processing In Memory)는 메모리에서 연산이 이루어지는 구조로써 뉴럴네트워크(Neural Network, 신경망)와 딥러닝(Deep Learning)은 영상/음성 인식 등 최적의 솔루션을 제공할 수 있으며 많은 연구가 수행되고 있다. 다층 퍼셉트론(MLP, Multilayer Perceptron) 및 합성곱 신경망(CNN, Convolutional Neural Network) 같은 뉴럴네트워크 및 딥러닝 알고리즘은 뉴럴 네트워크의 크기가 급격히 증가함에 따라 큰 메모리 용량이 필요하며 프로세싱 유닛(processing unit)에서 가중치 값을 메모리로부터 가져오기 위해 높은 메모리 대역폭을 필요로 한다(Chen et al, 2014). 이러한 솔루션이 프로세싱 유닛과 오프칩(Off-Chip) 메모리 사이의 시냅스 가중치(synaptic weight) 전달을 효과적으로 줄이지만, 시냅스 가중치 외에 입출력 데이터를 포함한 데이터 이동은 여전히 성능 향상 및 많은 전력소모의 문제점이 있다. 이런 문제점을 해결하기 위해서 프로세싱 인 메모리 기반의 뉴로모픽은 더 많은 온 칩(On-Chip) 메모리를 통합하는 대신 연산 로직을 메모리 칩에 넣어 이 문제를 해결할 수 있는 솔루션이다(Chi et al, 2016). 폰노이만 구조의 한계를 넘어 프로세서와 메모리를 통합하는 신개념 반도체 기술로써 딥러닝에 최적화된 인공지능 반도체는 프로세싱 인 메모리 구조를 통해 반도체 성능과 전력효율을 대폭 향상시켜 초고속 연산을 저전력으로 실현 가능하며 메모리 반도체 기반 프로세싱 인 메모리, 로직(프로세서) 기반 프로세싱 인 메모리, 공정기술과 양산에 혁신이 요구되는 신물질 소재(RRAM, PRAM 등) 기반 프로세싱 인 메모리 등으로 구분할 수 있다. 뉴로모픽 개발은 상당히 많은 발전을 거듭하여 왔으나 아직은 인간의 뇌를 그대로 모방하기에는 많은 난제들이 있다. 인간의 뇌는 시냅스와 뉴런의 조합이며 이들의 동작은 스파이크의 자극으로 상호 추론에 의한 복잡한 연결로 확장을 하는데, 자기 학습에 대한 연구 등은 아직 초기 단계이다(사이언스타임즈, 2019; Bengio, et al, 2013).

융합연구리뷰에서는 뉴로모픽 대역폭의 문제점을 해결하기 위한 프로세싱 인 메모리의 다양한 솔루션을 살펴보고자 한다. 이를 위해, 먼저 2장에서는 적층 DRAM(Dynamic Random-Access Memory, 동적 램) 기반의 프로세싱 인 메모리 반도체에 대한 사항을 먼저 살펴보고 3장에서는 SRAM(Static Random Access Memory, 정적 램), eDRAM(embedded DRAM, 임베디드 DRAM) 기반 프로세싱 인 메모리 반도체 동향을 분석한다. 이어서 4장에서는 신경망 가속을 위한 프로세싱 인 메모리 아키텍처에 대한 사항을 언급하고 5장에서는 프로세싱 인 메모리를 위한 인터페이스를 설명한다. 6장에서는 프로세싱 인 메모리 시장 및 정책 동향에 대해 살펴보고 마지막 장에서 결론을 제시한다.

II 적층 DRAM 기반의 프로세싱 인 메모리 반도체

1. HBM 기반 프로세싱 인 메모리

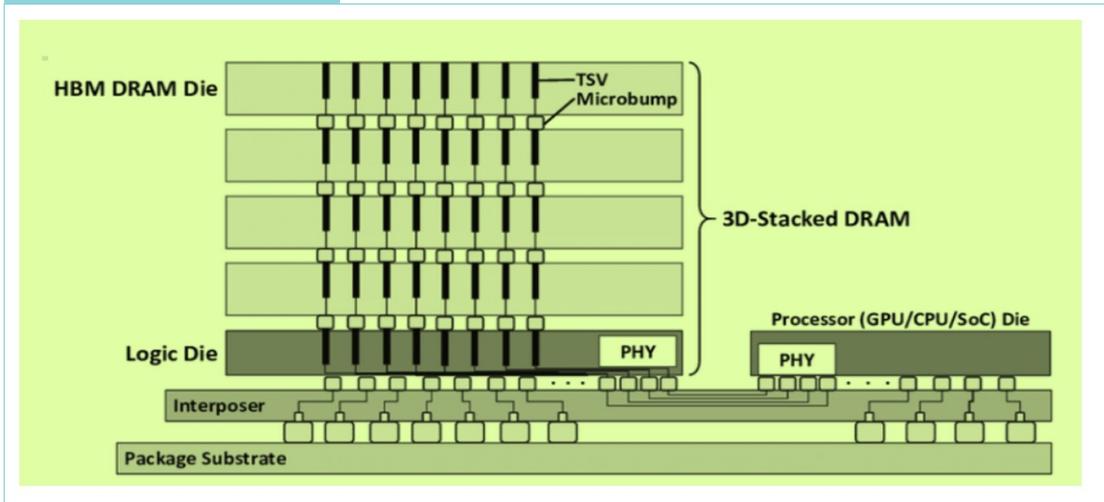
고대역(HBM, High Bandwidth Memory) 기반의 프로세싱 인 메모리 반도체 기술은 HBM 메모리와 반도체 내부에 내장된 연산 기능을 통합, 메모리 데이터 접근과 데이터 처리 성능을 극대화하여 대규모의 실시간 인공지능 연산을 가능하게 한다. 이 초저전력 인공지능경망 연산을 위한 반도체 칩 기술은 인공지능이 요구하는 기존 기술 대비 1,000배 이상인 100 테라플롭스(TeraFLOPS) 이상의 성능을 갖고 있다(권영수, 2021).



* 출처 : 권영수(2021)

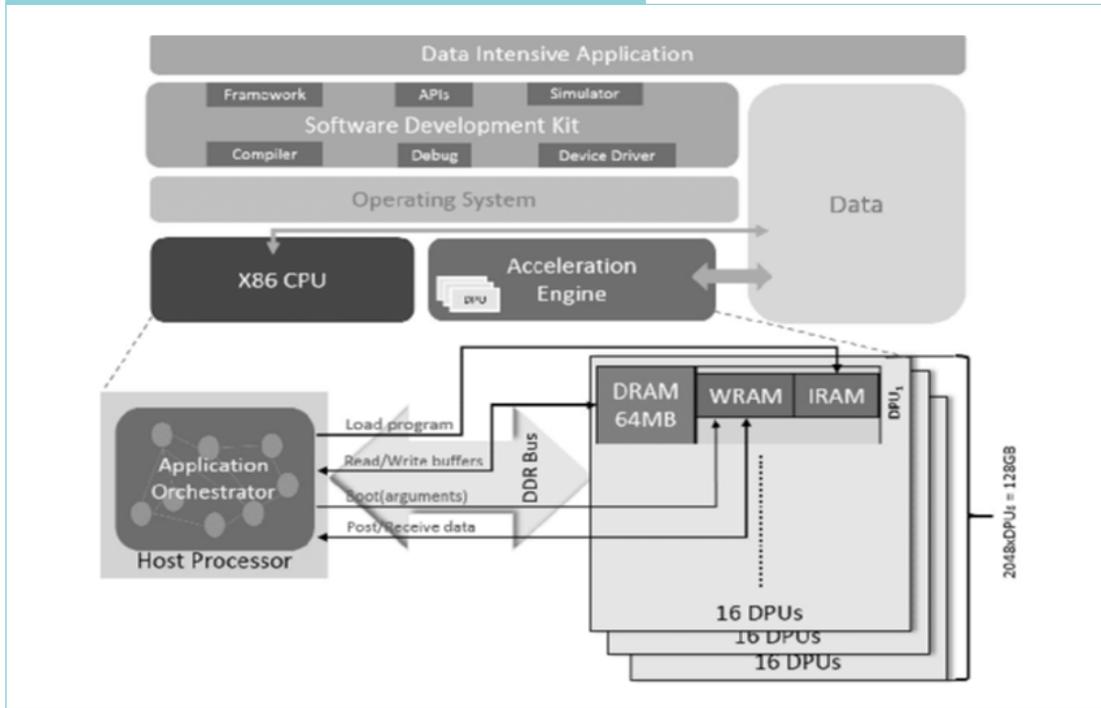
〈그림 2, 3, 4〉의 HBM은 DRAM 칩과 패키징 간의 대역폭을 높이는 방법으로, 특히 메모리 인터페이스 TSV(Through Silicon Via, 실리콘 관통 전극) 기술을 기반으로 수직적인 구조를 통해서 데이터를 전송하며, 최근에는 뉴로모픽의 대역폭 문제를 해결하는데 사용된다. 인터포저(Interposer, 집적회로)에 구현된 다수 DRAM에 최소의 논리회로를 추가한 DPU(DRAM Processing Unit) 64MB DRAM과 32-bit RISC(Reduced Instruction Set) 코어 기반 프로세싱 인 메모리 기반의 아키텍처가 제안되었다. 프로세싱 인 메모리는 메모리의 대역폭을 우선적으로 해결하기 위해 DRAM 내의 프로세서를 통합하여 낮은 데이터 액세스, 대기 시간, 고대역폭, 대규모 병렬 처리 및 낮은 전력 소비를 구현한다. 연산자는 광범위하게 컴퓨팅 및 데이터 집약적인 구조로써, 프로세싱 인 메모리 아키텍처용 알고리즘을 이용하여 대규모 병렬 및 처리량 효율성을 높인다. 기존 솔루션에 비해 14배 더 높은 처리량과 에너지 소비를 보여주고 있다(Zois et al, 2018).

그림 2. HBM 구조도②



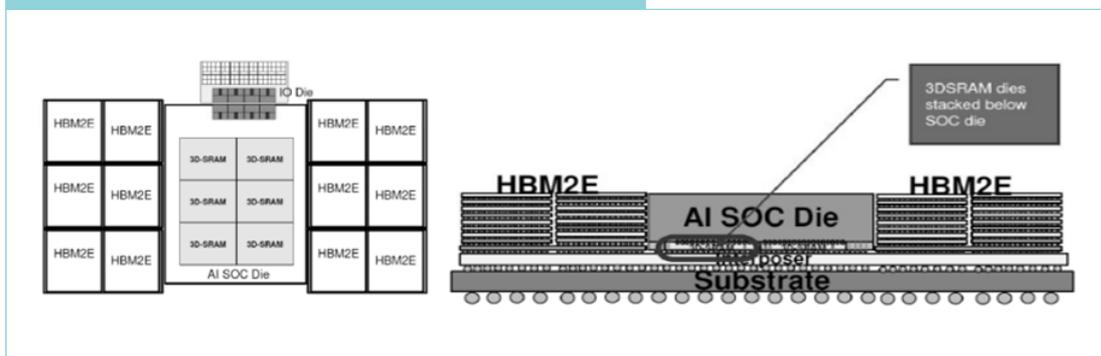
* 출처 : Imani et al(2018)

그림 3. UPMEM사의 프로세싱 인 메모리 구조도①



* 출처 : Zois et al(2018)

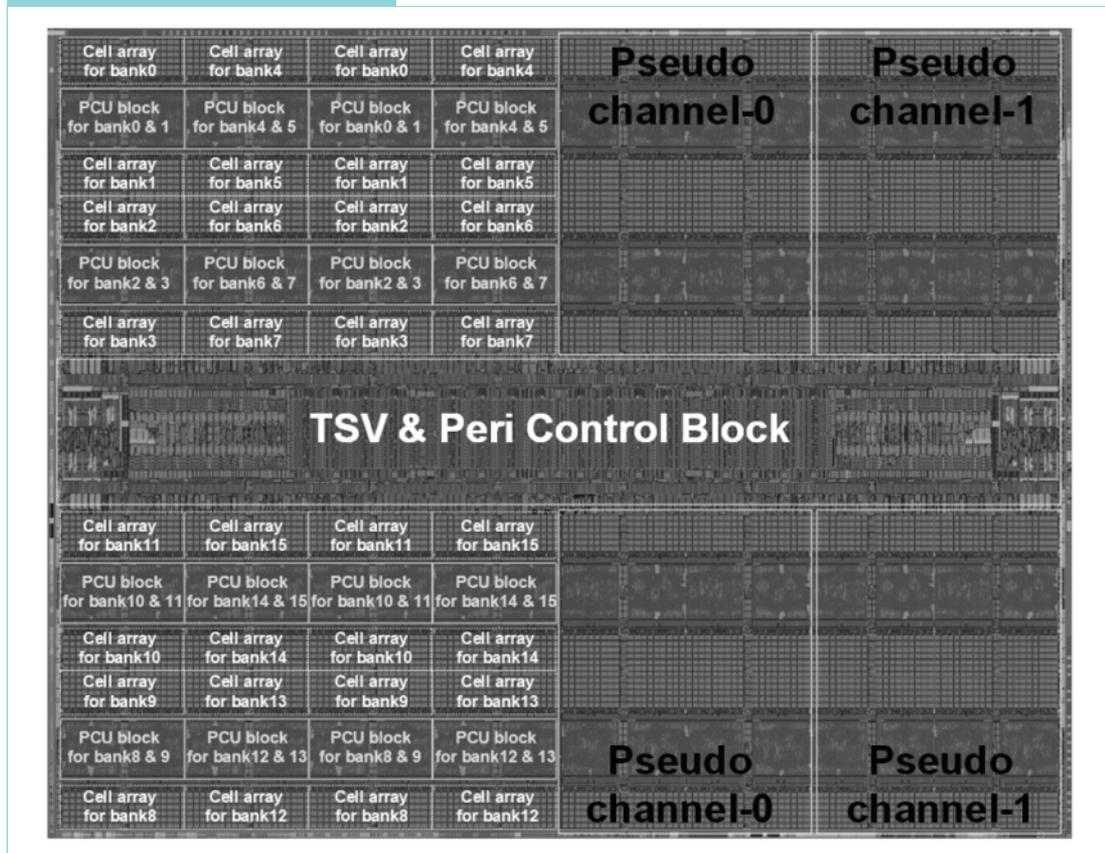
그림 4. UPMEM사의 프로세싱 인 메모리 구조도②



* 출처 : EETimes(2019)

〈그림 5〉의 FIMDRAM(Function-In-Memory DRAM) 칩은 프로세서(Die, 반도체 물질의 작은 사각형 조각)와 SRAM(Die)을 적층하고, HBM과 통합한 시스템에서 메모리와 로직을 더 가깝게 배치하여 연산 속도를 높이는 방법인 PNM(Processing Near Memory, 하나의 모듈에 메모리 반도체와 처리장치(CPU)가 탑재된 형태의 융합형 반도체) 구조를 가지며 SoC(System on Chip, 여러 가지 기능을 가진 시스템을 하나의 칩으로 구현한 기술이 집종된 반도체)의 3차원 SRAM 캐시(Cache, 주 기억장치의 접근 속도를 빠르게 하기 위해 프로세서 주변에 배치된 소 용량의 메모리)가 배열되는 특징을 가진다. 성능 측정 결과, HBM2(2세대 고대역폭 메모리)를 사용하는 일반적인 GPU 시스템에 비해 71%의 시스템 에너지 절감으로 성능이 2.1배 향상되었다.

그림 5. FIMDRAM 칩 사진

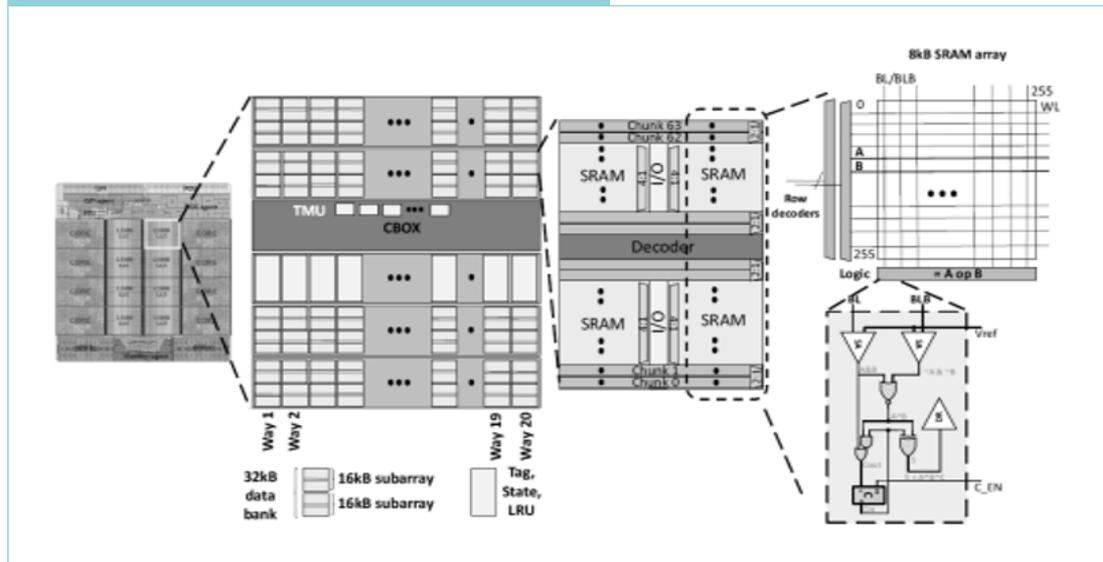


* 출처: Kwon et al(2021)

III SRAM, eDRAM 기반 프로세싱 인 메모리 반도체

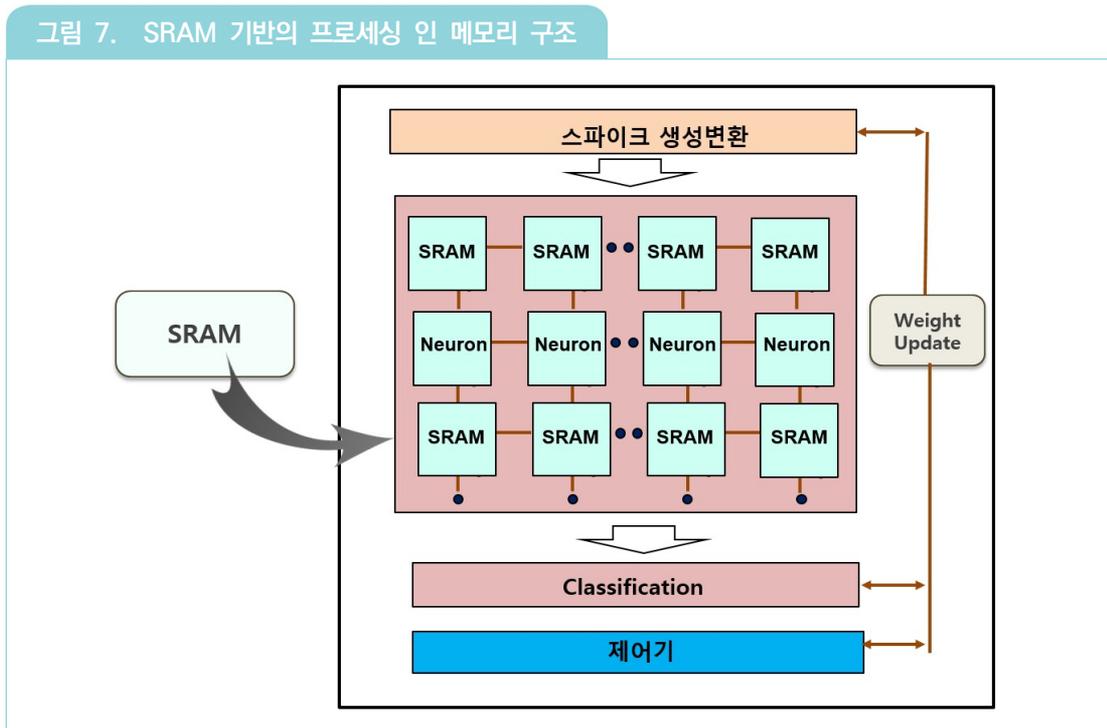
SRAM 기반 프로세싱 인 메모리 반도체(그림6)는 SRAM 셀을 기반으로 SRAM에 읽고, 저장을 할 때 동시에 연산을 할 수 있도록 하여 CPU 내에 프로세싱 인 메모리 기술을 적용한다. 8KB SRAM 어레이는 데이터 저장 기능과 동시에 열 단위로 저장된 데이터에서 비트 라인(bit line, 데이터를 읽고 쓰는 역할 수행) 연산이 가능한 연산기를 SRAM 어레이를 내부에 포함하고 있다.

그림 6. SRAM 기반 프로세싱 인 메모리 반도체



* 출처: 인텔 사이트

〈그림 7〉은 SRAM 기반의 프로세싱 인 메모리 구조이다. 뉴런, SRAM, 스파이크 생성 변환, 분류기, 제어기 등의 모듈로 구성되며, 스파이크 입력과 뉴런 어레이를 변환하는 세대 변환 블록에서 입력 및 출력 뉴런의 전압이 시간에 따라 변화한다. 뉴런의 가소성으로 인한 뉴런의 입력 및 출력 뉴런 전압이 증가하면서 누적 또는 감소하는 동작을 수행하며, 전압이 일정 수준을 초과하면 스파이크가 발생하는 연산으로 학습을 수행한다. 웨이트 값은 정기적으로 업데이트되며 SRAM을 저장하는 구조이다.

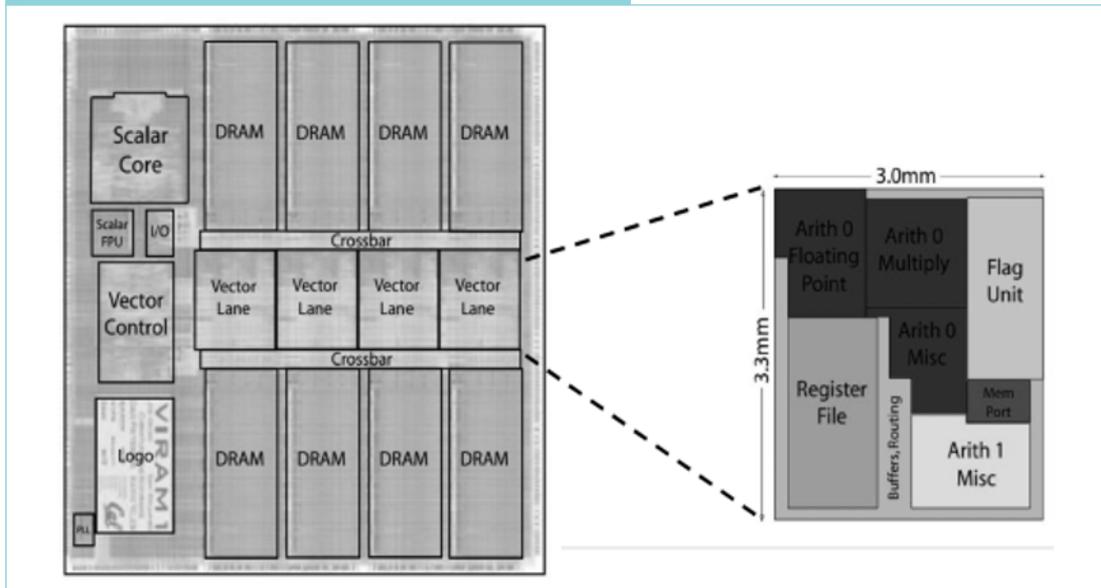


* 출처: 박성모 외(2021)

〈그림 8〉의 임베디드 기반 프로세싱 인 메모리 반도체(eDRAM, embedded DRAM)는 메모리를 병렬 메모리로 변경한 것으로 64개의 연산기 코어와 병렬처리 메모리로 구성하여 OpenMP(Open Multi-Processing) 등의 병렬 프로그래밍 모델을 지원하며, 고성능 응용 프로그램 메모리 대역폭 부족으로 인한 비효율성을 개선하였다. 벡터 방식의 지능형 램(VIRAM, Vector Intelligent RAM) 아키텍처는 새로운 프로세싱 인 메모리 기술을 사용하여 임베디드 시스템의 큰 대역폭 잠재력을 활용하기 위해 보조 벡터 프로세서가 있는 DRAM을 기반으로 다양한 메모리 계층 구조를 제공하며, SIMD(Single Instruction

Multiple Data, 하나의 명령어로 여러 개의 데이터를 처리할 수 있는 기술) 제어 VLIW(Very Long Instruction Word) 클러스터의 성능을 개선하였다. 또한 계산 특성 및 메모리 액세스 패턴을 분석하여 각각의 아키텍처에 가장 적합한 애플리케이션 특성을 가진다. eDRAM과 함께 집적된 프로세싱 어레이에서는 eDRAM의 병렬 입출력을 통해 읽은 많은 데이터를 동시에 처리가 가능한 64개의 SIMD 기능을 가지고 있는 프로세서 코어와 함께 집적도를 높였다.

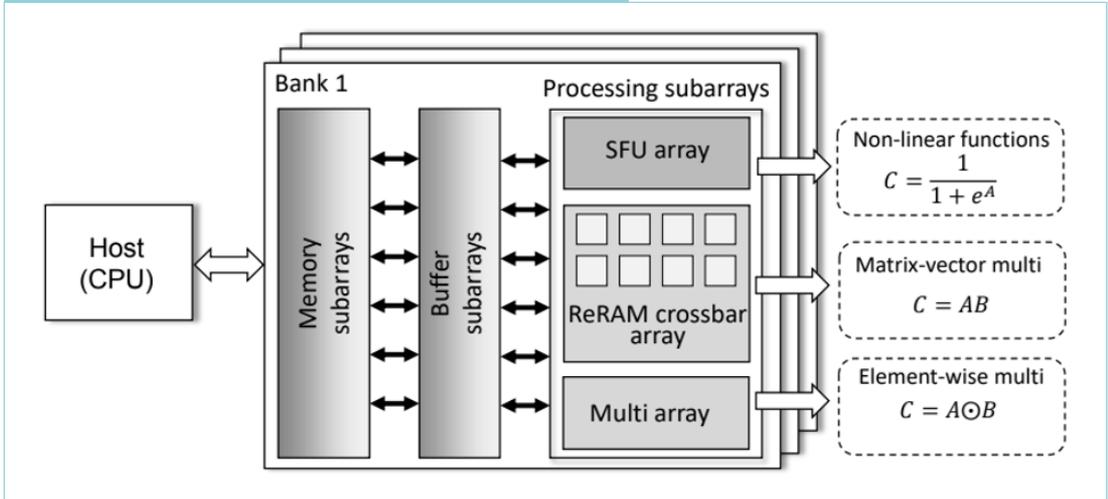
그림 8. 임베디드 기반 프로세싱 인 메모리 반도체



* 출처: The Berkeley Intelligent RAM(IRAM) Project

프로세싱 인 메모리 기술을 실현하기 위한 새로운 메모리 기술의 연구개발이 활발히 이루어지고 있으며 다양한 분야에서 연산 성능의 획기적 개선이 가능한 프로세싱 기술을 통해서 발전하고 있다. 또한 새로운 소자를 사용하여 ReRAM/MRAM 기반의 프로세싱 인 메모리 반도체를 개발 중인데, ReRAM(Resistive Random Access Memory, 저항 변화 메모리)은 금속-부도체-금속 구조의 비휘발성 메모리 소자로, 양 금속판 전압에 따라 조절되는 부도체 저항으로 데이터를 저장한다. 고속 동작, 저전력, 고집적 가능 등의 장점도 있을 뿐만 아니라, bit-lane을 공유하는 소자별 저항 차별화를 통해 ReRAM 구조 자체를 내적(dot-product) 연산에 활용할 수 있어 프로세싱 인 메모리 구성에 이점이 있다.

그림 9. ReRAM 기반 프로세싱 인 메모리 반도체

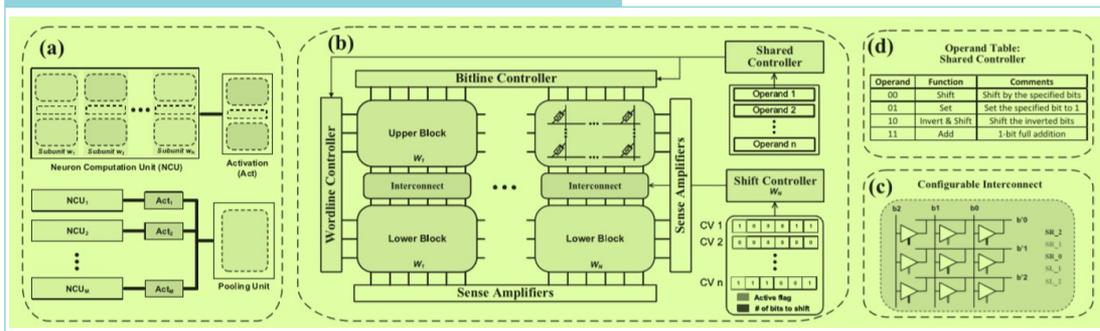


* 출처: Long et al(2018)

IV 신경망 가속을 위한 인 메모리 아키텍처

Gupta et al(2019) 연구에서는 NN프로세싱 인 메모리(Processing In-Memory Architecture for Neural Network Acceleration, 신경망 가속을 위한 인 메모리 아키텍처)라고 하는 새로운 처리 인 메모리(In-Memory) 아키텍처를 제안하였다(그림10) 참고). 신경망의 추론 단계를 가속화하며 메모리 내부에서 빠른 덧셈, 곱셈 및 검색 연산을 지원하는 크로스바 메모리 아키텍처(crossbar array architecture)를 동작하여 신경망의 성능을 크게 향상시키고 전체 에너지 소비를 줄이는 간단한 최적화 기술을 구현하였다. 또한 병렬 인 메모리 구성 요소를 사용하여 모든 뉴럴네트워크 기능의 효율성을 매핑을 통해 더욱 향상시키기 위해 메모리의 계산 수를 줄이고, 연속적으로 NN프로세싱 인 메모리 계산의 속도를 높이기 위해 가중치 공유를 통한 리소스를 최소화 하였다. NN프로세싱 인 메모리 아키텍처와 그래픽 처리 장치(GPU, Graphics Processing Unit) 및 최신 프로세싱 인 메모리 아키텍처와의 효율성을 비교평가 하였다. 비교 결과, NN프로세싱 인 메모리 아키텍처는 NVIDIA社의 GeForce GTX 1,080 GPU 아키텍처와 비교하여 131.5배 높은 에너지 효율성을 달성하고 48.2배 빠른 것으로 나타났다. 또한 신경망 가속기와 비교했을 때, NN프로세싱 인 메모리는 동일한 분류 정확도를 제공하면서 평균 3.6배 높은 에너지 효율과 4.6배 빠른 속도를 나타내었다.

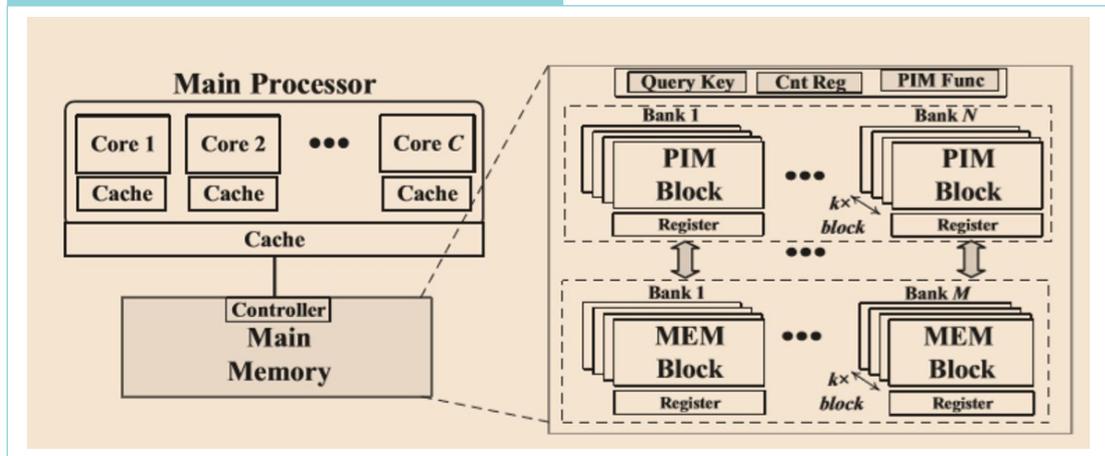
그림 10. 신경망 가속을 위한 인 메모리 아키텍처



* 출처: Gupta et al(2019)

〈그림 11〉은 프로세싱 인 메모리의 사용을 일반화하기 위해 기존 프로세서와 프로세싱 인 메모리 가속기로 구성된 일반 처리 인 메모리 아키텍처인 Gen프로세싱 인 메모리이다(Imani et al, 2018). Gen프로세싱 인 메모리는 비트 연산, 검색 연산, 덧셈 및 곱셈을 비롯한 특수 비휘발성 메모리의 기본 프로세싱 인 메모리 기능을 지원하며, 각 응용 프로그램에 대해 프로세싱 인 메모리 작업을 사용하는 부분을 식별하고 범용 코어에서 응용 프로그램의 나머지 비 프로세싱 인 메모리 작업 또는 데이터 집약적인 부분을 처리하는 특징을 가지고 있다. 또한 메모리 내 계산을 완화하여 구성 가능한 프로세싱 인 메모리 근사를 가능하게 하며, 실험 결과, 기존 코어에서 데이터를 처리하는 것과 비교하여 10.9배의 에너지 효율 개선과 6.4배의 속도 향상이 된 것으로 나타났다.

그림 11. Gen프로세싱 인 메모리 아키텍처



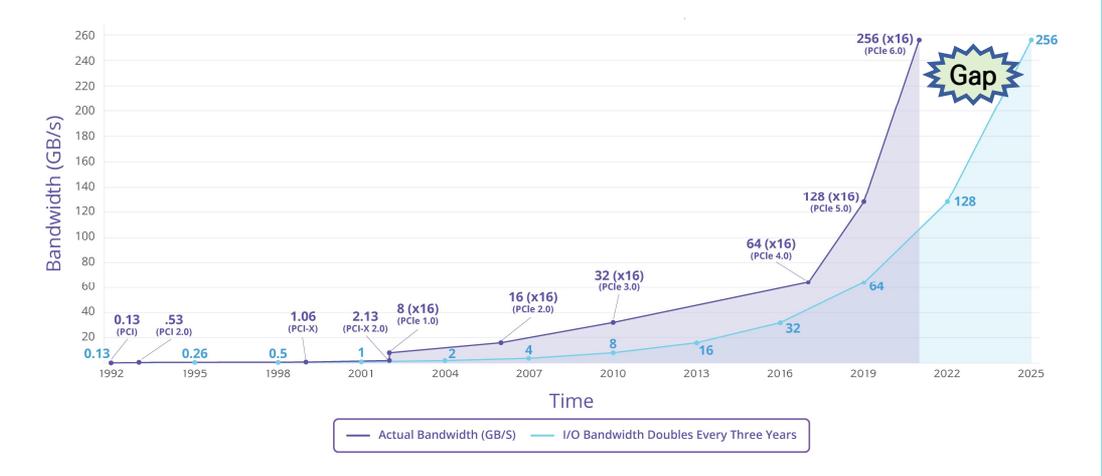
* 출처: Imani et al(2018)

V 프로세싱 인 메모리를 위한 인터페이스

PCIe(Peripheral Component Interconnect Express)는 Intel社 주도하에 만들어진 PCI(Peripheral Component Interconnect), PCI-X(Peripheral Component Interconnect eXtended)와 AGP(Advanced Graphics Port) 버스를 대체하기 위하여 개발된 고속 직렬 인터페이스이다. 높은 대역폭(PCIe 5.0의 경우 lane 당 32GT/s), 적은 I/O 핀 수, 성능 확장성, 오류 검출 및 보고 구조, 핫-플러그, 하드웨어 I/O 가상화 등을 지원한다. PCIe 5.0은 2019년 규격이 확정되었으며 2022년부터 PCIe 5.0을 적용한 상용 프로세서, NVMe M.2 SSD 혹은 그래픽 카드가 출시될 예정이다. NVIDIA社의 NVLink 상호연결 인터페이스는 자사의 GPGPU(General-Purpose computing on Graphics Processing Units, GPU 상의 범용 계산) 간의 연결뿐만 아니라 CPU와의 연결에도 사용될 수 있도록 개발되었으며, 최근에는 P2P 연결뿐만 아니라 NV Switch를 통한 네트워크 구성도 구조의 특징을 가진다. 또한, NVIDIA社의 Volta GV100은 GP100이 4개의 VLink와 160GB/s의 총 대역폭을 지원하던 것에 비해 최대 6개의 NVLink와 300GB/s의 총 대역폭을 지원하며, NVLink는 레인 당 25Gbps 대역폭을 제공, 레인 8개를 한 방향으로 데이터를 전송하는 하위링크를 형성하고 방향마다 하나씩 2개의 하위링크를 형성하여 2개의 프로세서(GPU-GPU 또는 CPU-to-GPU)를 연결하는 링크를 가진다.

최근 인간의 욕구가 증대되고 서비스 또한 고대역폭 응용 분야가 증대함에 따라 현재 CPU와 GPU 및 인공지능 가속기(Accelerator) 등의 입출력 디바이스는 대부분 프로토콜을 통해 연결되어 있으며, 이로 인해 입출력 대역폭은 3년마다 배로 증가한다.

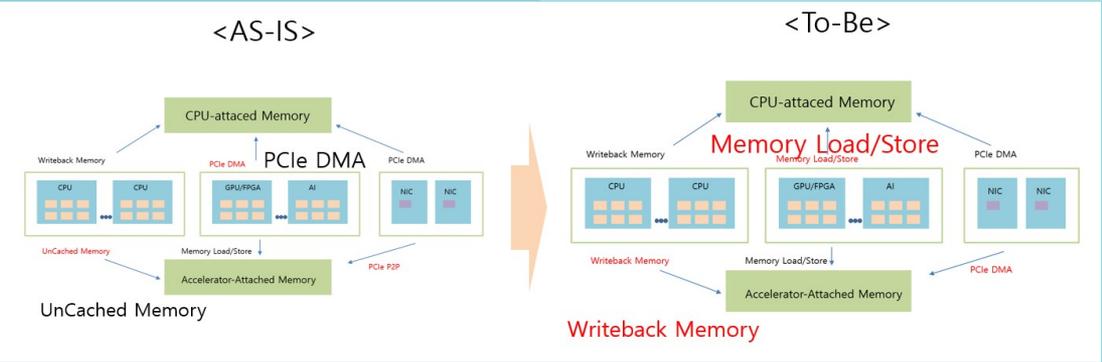
그림 12. I/O 대역폭은 매 3년마다 2배씩 증가



* 출처: PCI-SIG 사이트

CPU의 코어 개수는 지속적으로 증가하고 있고, 코어 당 PCIe 대역폭 또한 계속해서 감소하며, 대역폭의 증가로 인한 가속기 메모리 공유방법과 이에 따른 캐시 일관성 문제로 인해 GPU, FPGA(Field Programmable Gate Array, 프로그래밍이 가능한 내부 회로가 포함된 반도체 소자) 등 가속기 분야에서 새로운 해결이 필요하다. 이러한 문제점을 해결하기 위하여 새로운 프로토콜인 CXL(Compute Express Link)은 2019년 3월 Intel社가 제안한 새로운 컴퓨터 디바이스 인터커넥터 규격이며, Intel社가 주도하는 CXL 컨소시엄에는 ARM, HP, Huawei, Mellanox를 포함한 다수의 컴퓨터 기업들이 참여하고 있으며, 현재 CXL 2.0이 발표되었다.

그림 13. Compute Express Link 구조도



* 출처: Lender(2020)

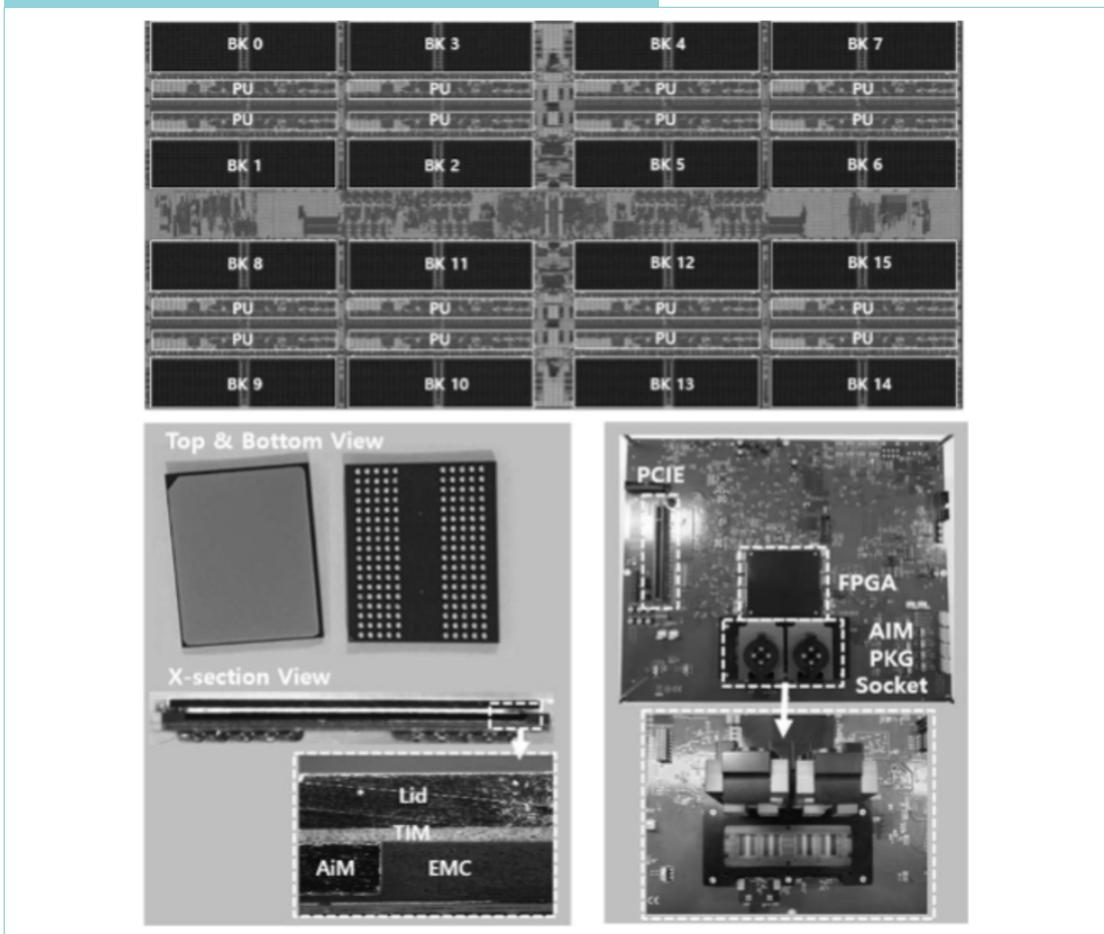
CXL 2.0에서는 프로토콜 호스트 프로세서(Protocol Host Processor)에서 간단한 명령어로 가속기의 메인 메모리에 접근이 가능하고 반대로 가속기에서 CPU의 메인 메모리에 접근 시에는 PCIe DMA를 거치는 대신 로드 스토어의 명령으로 가능해졌다. CXL은 고성능 컴퓨팅 시스템에서 CPU와 함께 사용되는 가속기, 메모리, 저장장치 등을 보다 효율적으로 활용하기 위한 인터페이스로 기존 컴퓨팅 시스템인 DRAM의 물리적 한계를 극복하고, 메모리 용량을 획기적으로 확장이 가능하다. 최근 모든 산업, 모든 서비스에 인공지능이 도입되고 있으며 이울러 처리해야 할 데이터양도 기하급수적으로 증가하기 때문에 인공지능 컴퓨팅 성능에 큰 영향을 끼치는 기존 DRAM 성능을 획기적으로 도약시킬 수 있는 혁신적 반도체 아키텍처를 제안하고 이를 구현하는 PHY(Physical Layer, 물리층)를 개발하여, 메모리 액세스 실패 대역폭 및 프로세서의 실패 성능을 향상시키고 있다. 특히, 다중 DRAM 모듈용 CXL 기반 초고속 프로세싱 인 메모리 반도체는 이기종 컴퓨팅 성능 극대화에 적합한 니어(near) 프로세싱 인 메모리 구조를 기반으로, 인공지능 연산기와 다중 DRAM 모듈 사이의 데이터 이동을 최적화하고 연산의 실패 성능을 고도화한다. 니어 프로세싱 인 메모리는 프로세서의 로직과 메모리 간의 데이터 채널을 효율적으로 연결이 가능하고 또한 상호간에 인터커넥트 표준으로 제정된 CXL 인터페이스를 이용하여 프로세서 메인 메모리와 CXL 메모리를 함께 활용할 수 있는 새로운 구조를 통해서 DRAM 메모리 대역폭을 대폭 확장시키는 초고속 프로세싱 인 메모리 기술이다. 기존 DRAM 모듈의 한계를 극복하고 용량과 대역폭을 획기적으로 확장하며 효과적으로 자원 활용을 통하여 캐시 일관성을 구현하고 저지연 고속 연결이 가능한 CXL 기반 프로세싱 인 메모리 반도체 개발과 인공지능 성능 향상 시연이 포함되며, CXL 프로토콜을 기반으로 하면 프로세서의 자기 메모리처럼 사용이 가능해졌다.

메모리 중심의 구조에서 알 수 있듯이 국내 메모리 산업에 강점이 있는 기업이 뛰어 들고 있다. SK하이닉스 뿐만 아니라 삼성전자도 투자 및 연구 개발을 통해서 미래 시장을 대비하고 있다. 인공지능 반도체를 중심으로 산업구조가 재편되면서 저전력 또는 메모리 대역폭의 문제점을 획기적으로 개선할 신기술의 필요성이 커지고 있다. 그중 하나가 프로세싱 인 메모리 기술로써 다양한 방법을 통해서 접근을 시도하고 있다. 프로세싱 인 메모리에서는 HBM, DRAM, SRAM, ReRAM, eDRAM 등 다양한 기반의 솔루션을 위한 연구개발을 통해 기존에 생각하지 못했던 새로운 방법 및 구조를 사용하여 메모리 중심의 새로운 구조가 곧 양산될 것으로 보인다. 프로세싱 인 메모리는 단일 칩이 메모리와 프로세서의 역할을 모두 수행할 수 있고, 반도체 간의 데이터를 별도로 전송할 필요가 없기 때문에 데이터 전송 과정에서 발생하는 지연 시간 및 전력 손실을 줄일 수 있다.

메모리 반도체가 연산 기능을 수행하는 프로세싱 인 메모리 시장은 확대 조짐을 보이고 있다. 프로세싱 인 메모리 사업은 정부의 중점 사업의 하나로, 인공지능 보편화와 빅데이터 처리 속도 향상에 기여할 기술로

평가받고 있다. SK Hynix가 개발한 ‘GDDR6-AiM’은 일반 DRAM 대신 CPU·GPU와 함께 탑재하여 특정 연산 속도는 최대 16배까지 빨라지며 자체 연산을 하는 프로세싱 인 메모리 구조를 통해서 CPU·GPU로의 데이터 이동을 줄여 전력 사용량이 감소하며, 에너지 소모는 80%가량 줄어든다. GDDR6(Graphics DDR6) 기반의 AiM(accelerator-in-memory)을 적용하여 상대적으로 비용이 저렴하고 GDDR6 인터페이스와 호환되는 딥러닝 응용 프로그램을 가속화하도록 설계하였다(그림 14) 참고). AiM은 16Gb/s 속도의 GDDR6의 특성을 활용하여 1GHz 속도의 처리 장치(PU, Processor Unit)로 1 플롭스(TFLOPS)의 최대 처리량을 가지며, 또한 다양한 활성화 기능을 내장하여 많은 애플리케이션을 지원한다.

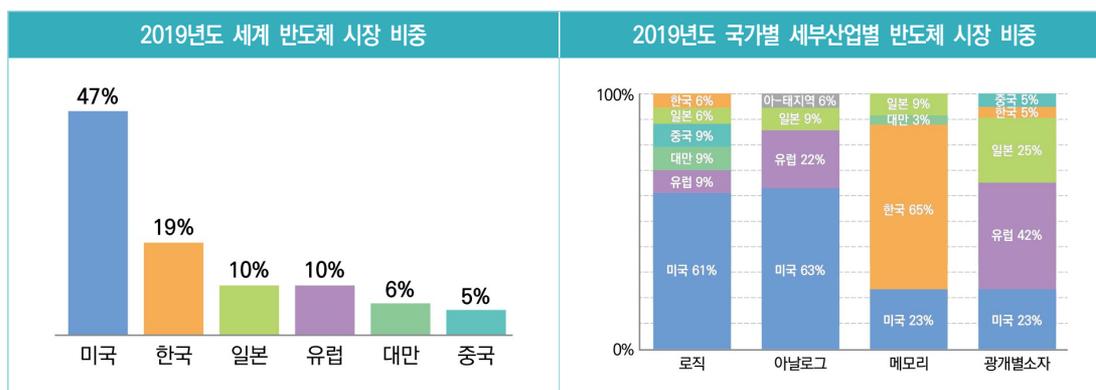
그림 14. AiM 칩 패키지 사진 및 에물레이션 시스템



* 출처: Lee et al(2022)

VI 프로세싱 인 메모리 시장 및 정책

2019년도 세계 반도체 시장은 전년 대비 12% 감소한 4,123억 달러를 기록하였으며, 단품 메모리 제품의 수요 감소에 따라 통신(-10.5%), 컴퓨터(-18.7%), 소비자(-5.2%), 자동차(-6.9%), 산업(-13.0%) 등 매출액이 급락하고 있다. 국가별 비중은 미국 47%, 한국 19%, 일본 10%, 유럽 10%, 대만 6%, 중국 5%이며, 미국이 시스템 반도체와 아날로그 반도체에서 시장을 주도하고 있고 종합 반도체 기업(IDM, Intergrated Device Manufacturer) 시장의 51%, 팹리스(반도체 설계) 시장의 65%를 차지한다. 한국은 메모리 반도체, 유럽과 일본은 광 개별 소자 시장을 주도하고 있다.



* 출처: 미국 반도체산업협회 자료(2020.06)

프로세싱 인 메모리 반도체의 주 수요처는 파운드리 시장 및 기업과 프로세서 기업으로써, 미국에 AMD, 애플, 브로드컴, 퀄컴, NVIDIA 등과 같은 대형 팹리스 기업들이 있기 때문에 가장 수요가 많으며, 글로벌 파운드리 시장에서는 상위 8개 기업이 88.4%를 점유하고 있고 그중에서 대만 기업이 약 60%를 차지한다. 최근 5년간 중국 파운드리 시장이 빠른 속도로 성장 중이며 2017년에는 전년 대비 30% 증가하며 글로벌 성장률 평균인 9%에 비해 세 배 이상 증가하였다. 국내 시장은 글로벌 메모리 반도체 강자인 삼성전자와 SK Hynix가 파운드리 사업부문을 분리하면서 파운드리 사업 강화와 함께 파운드리 시장 수요가 증가하고

있다. 삼성전자 파운드리 사업은 2018년부터 세계 2위, 2019년 1분기 점유율 19.1%이며 AMD사는 삼성전자와 모바일 그래픽 관련 전략적 제휴를 맺고 GPU와 CPU 생산 시 삼성전자 파운드리 설비를 활용할 예정이다. SK Hynix는 2017년 1월 파운드리 사업부를 분사해 'SK Hynix 시스템IC'를 설립하고, 파운드리 분야 경쟁력을 강화하고 있다.

프로세싱인 메모리 관련 국가연구 사업으로는 과학기술정보통신부에서 4개 사업, 산업통상자원부에서 3개 사업 그리고 다부처 사업으로 1개 사업이 진행되고 있다. 과학기술정보통신부는 2028년까지 프로세싱 인 메모리 기술개발에 4,000억 원을 추가로 투자하겠다고 밝혔으며 인공지능 반도체 개발에 필요한 소프트웨어, 실증사업, 데이터센터 등을 지원할 예정이다.

표 1. 부처별 인공지능 반도체 관련 사업 현황

(단위: 억 원)

부처명	사업명	2021년 예산	종료시점
과학기술정보통신부	신개념 PIM 기술개발	115.30	2024
	혁신성장연계 지능형반도체 선도기술개발	150.0	2021
	인공지능반도체응용기술개발	70.79	2024
	시스템반도체융합전문인력육성	72.00	2026
산업통상자원부	전자정보 디바이스 산업 원천기술개발	81.99	2021
	전자부품산업기술개발	164.77	2023
	시스템반도체핵심IP개발	79.92	2024
다부처 (과학기술정보통신부, 산업통상자원부)	차세대 지능형반도체 기술개발	1130.62	2029

* 출처: 한국과학기술기획평가원(2021)

VII 결론

메모리 반도체가 연산 기능을 수행하는 프로세싱 인 메모리는 인공지능 일반화와 빅데이터 처리 속도 향상에 기여할 기술로 평가받고 있다. 메모리는 명령어와 연산자를 저장하고 프로세서는 메모리에서 명령어와 연산자를 가져와 연산을 수행하였으나, 프로세싱 인 메모리는 연산이 메모리에서 수행되는 아키텍처로 다양한 서비스에 적용이 가능하게 되었다. 기존의 한계를 넘어 프로세서와 메모리를 집적한 신개념 반도체인 인공지능 반도체는 프로세싱 인 메모리 구조를 통해 반도체 성능과 전력 효율이 크게 향상되어 저전력 소모로 고속 동작이 가능하다. 인공지능 반도체는 메모리 공정의 혁신이 필요한 반도체 기반의 메모리 기반 프로세싱, 로직(프로세서) 기반의 프로세싱 인 메모리와 RRAM, PRAM 등을 기반으로 하는 프로세싱 인 메모리로 구분할 수 있다. 융합연구리뷰에서는 적층 DRAM 기반의 프로세싱 인 메모리 반도체의 특징과 구조에 대해 분석하였고, 또한, eDRAM 기반 프로세싱 인 메모리 반도체 동향을 언급하였다. 신경망 가속을 위한 프로세싱 인 메모리 아키텍처에 대한 사항 및 프로세싱 인 메모리 위한 인터페이스 표준에 대한 기술, 특히 CXL의 동향을 파악하였으며 마지막으로 관련 시장 및 정책에 대해서 언급을 하였다.

저자_ 박성모(Seong Mo Park)

• 학력

경북대학교 전자공학 박사
경북대학교 전자공학 석사
경북대학교 전자공학 학사

• 경력

現) 한국전자통신연구원 인공지능연구소 책임연구원
前) 금성일렉트론 주임연구원

저자_ 박필재(Pil Jae Park)

• 학력

캘리포니아 대학교 산타바버라
(Univ. of California, Santa Barbara)
전기컴퓨터공학 박사
광주과학기술원 기전공학 석사
영남대학교 전기공학 학사

• 경력

現) 한국전자통신연구원 인공지능연구소 책임연구원

참고문헌

<국내문헌>

- 1) 권영수. (2021). HBM 기반 프로세싱 인 메모리 NPU 반도체 기술. 정보통신기획평가원 주간기술동향, ICT R&D 동향, pp. 29-34.
- 2) 박성모, 최병건, 구본태, 박경환, 권영수. (2021). 저전력 인공지능 반도체 기술 동향. 정보통신기획평가원 주간기술동향, pp. 2-12.
- 3) 한국과학기술기획평가원. (2021). PIM 인공지능 반도체 핵심기술개발사업 예비타당성조사 보고서, pp. 166.

<국외문헌>

- 4) Chen, T., Du, Z., Sun, N., Wang, J., Wu, C., Chen, Y., Temam, O. (2014). DianNao: A small-footprint high-throughput accelerator for ubiquitous machine-learning. Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. DOI:10.1145/2541940.2541967.
- 5) Chi, P., Li, S., Xu, C., Zhang T., Zhao J., Liu, Y., Wang, Y., Xie, Y. (2016). PRIME: A Novel Processing-in-Memory Architecture for Neural Network Computation in ReRAM-Based Main Memory, 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture(ISCA), DOI: 10.1109/ISCA.2016.13.
- 6) 김준래. (2017.08.10). 사람 뇌 닮은 반도체칩 '뉴로모픽'. 사이언스탐임즈.
- 7) Bengio, Y., Courville, A., Vincent, P. (2013). Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828.
- 8) Gupta, S., Imani, M., Kaur, H., Rosing, T. (2019). NNPIIM: A Processing In-Memory Architecture for Neural Network Acceleration. IEEE Transactions on Computers, 68(9), 1325-1337.
- 9) Imani, M., Gupta, S., Rosing, T. (2018). GenPIM: Generalized Processing In-Memory to Accelerate Data Intensive Applications, Design. Automation And Test in Europe. pp. 1155-1158.
- 10) Zois, V., Gupta, D., Tsotras, V. J., Najjar, W. A., Roy, J. F. (2018). Massively Parallel Skyline Computation for Processing-In-Memory Architectures. PACT '18: Proceedings of the 27th International Conference on Parallel Architectures and Compilation Techniques. Article No. 1, pp. 1-12.
- 11) Kwon, Y. C., Lee, S. H., Lee, J. H. et al. (2021). A 20nm 6GB Function-In-Memory DRAM, Based on HBM2 with a 1.2TFLOPS Programmable Computing Unit Using Bank-Level Parallelism, for Machine Learning Applications. 2021 IEEE International Solid-State Circuits Conference(ISSCC), pp. 351-353.

- 12) Long, Y., Na, T., Mukhopadhyay, S. (2018). ReRAM based Processing-in-Memory Architecture for Recurrent Neural Network Acceleration. IEEE Transactions on Very Large Scale Integration(VLSI) Systems, 26(12), pp. 1-14.
- 13) Lee, S., Kim, K., Oh, S., et al. (2022). A 1ynm 1.25V 8Gb, 16Gb/s/pin GDDR6-based Accelerator-in-Memory supporting 1TFLOPS MAC Operation and Various Activation Functions for Deep-Learning Applications. 2022 IEEE International Solid-State Circuits Conference(ISSCC), pp. 176-177.

〈기타문헌〉

- 14) Merritt, R. (2019). Huawei: Covering all the Bases in AI, Aspencore. <https://www.eetasia.com/huawei-covering-all-the-bases-in-ai/>
- 15) 인텔 사이트, <https://www.Intel.com>
- 16) University of California, Berkeley 사이트, <http://iram.cs.berkeley.edu>. The Berkeley Intelligent RAM (IRAM) Project 자료.
- 17) PCI-SIG 사이트, www.PCI-SIG.com
- 18) Lender, L. (2020). Compute Express Link(CXL): A Coherent Interface for Ultra High Speed Transfers, CXL Marketing Work Group Intel Corporation. Flash Memory Summit 2019 Santa Clara, pp. 4.

융합연구리뷰

Convergence Research Review 2022 August vol.8 no.8



03

국가R&D 현황 분석

융합연구리뷰 8월호에서 다룬 2개의 주제(차세대 전력반도체 소자 기술 및 인공지능 반도체 기술)에 대한 각각의 국가R&D 현황을 살펴보기 위해 국가연구개발 과제 분석을 수행하였다. 연구비를 기준으로 연구비 규모별 과제수, 연구수행주체, 연구수준, 연구분야(국가과학기술표준분류, 미래유망신기술분류) 등 여러 측면에서의 분석 결과를 제시한다.

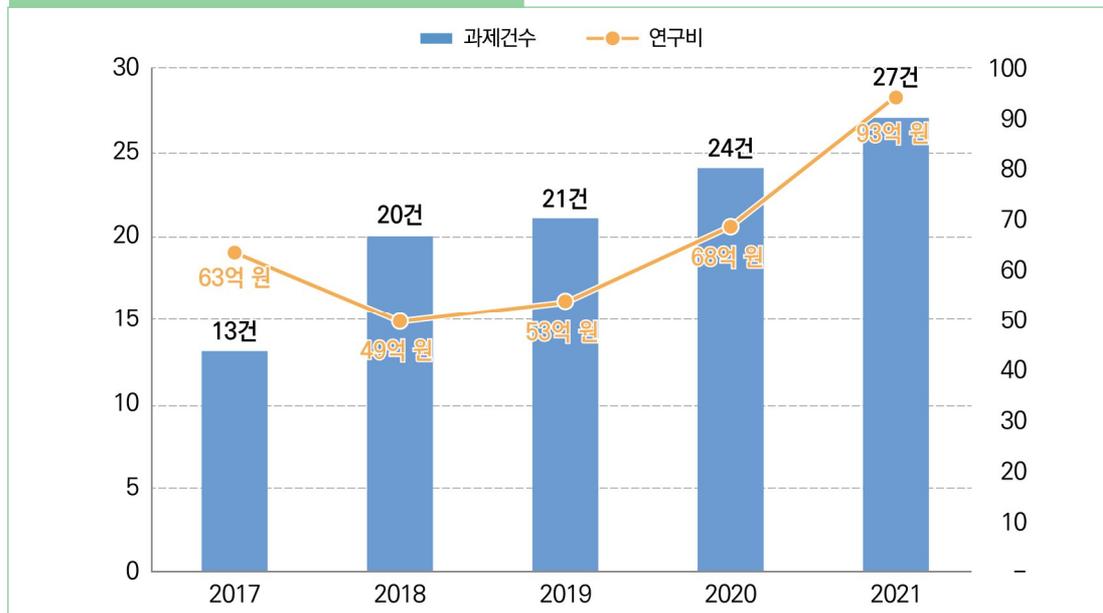
I 차세대 전력반도체 소자 기술

□ (총괄) 최근 5년간('17~'21) 총 105건의 과제에 대해 327억 원의 연구비가 투자됨

※ 국가과학기술지식정보서비스(NTIS) 플랫폼을 기반으로 관련 국가 연구개발 과제 분석 수행 : 원고의 핵심 키워드를 고려하여 검색 실시

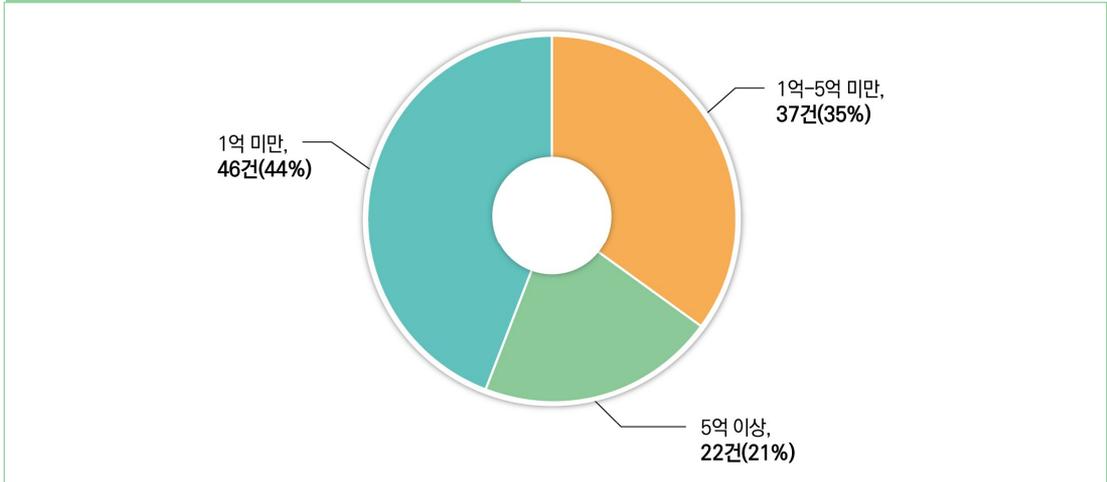
* 전력반도체 and 소자 and (SiC or GaN or 산화갈륨)

그림 1. 연도별 연구과제 건수 및 연구비



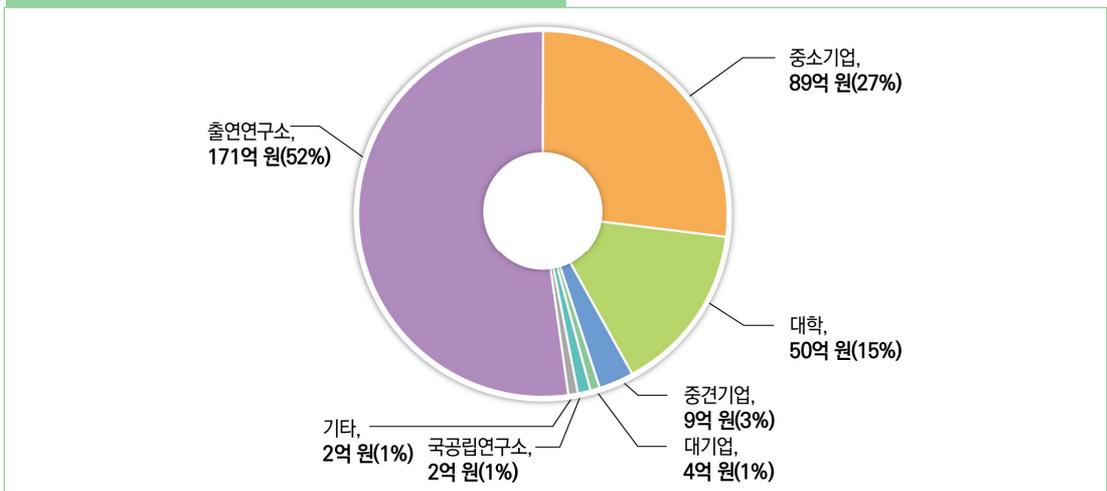
□ (연구비 규모별 과제 수) 연구비가 1억 원 미만인 과제(44%, 46건), 1억 원 이상 5억 원 미만인 과제(35%, 37건), 5억 원 이상인 과제(21%, 22건) 순으로 큰 비중을 차지하는 것으로 확인됨

그림 2. 연구비 규모별 과제 수 및 비율



□ (연구수행주체) 출연연구소가 지원받는 연구비의 비중이 차세대 전력반도체 소자 기술 전체 연구비의 절반 이상(52%, 171억 원)을 차지하는 것으로 확인되었고, 중소기업(27%, 89억 원), 대학(15%, 50억 원) 순으로 연구비 비중이 큰 것으로 확인됨

그림 3. 연구수행주체별 연구비 규모 및 비율



□ (연구수준) 연구수준을 분석한 결과, 차세대 전력 반도체 소자 기술 관련 연구는 성장기이며 개발연구 단계인 것으로 나타남

- (연구개발단계 분석 결과) 개발연구에 투자되는 연구비 비중이 64%(208억 원)로 가장 큰 것으로 확인되었고, 응용연구(22%, 71억 원)와 기초연구(11%, 35억 원) 순으로 연구비 비중이 큰 것으로 나타남
- (연구개발성격 분석 결과) 시작품 개발 관련 연구의 연구비 비중(16%, 53억 원)이 제품 또는 공정 개발(7%, 22억 원)과 아이디어 개발(1%, 3억 원)에 비해 월등히 큰 것으로 드러남
- (기술수명주기 분석 결과) 도입기(15%, 49억 원)와 성숙기(2%, 6억 원)에 투자되는 연구비 비중에 비해 성장기에 대한 연구비 비중이 31%(102억 원)로 가장 큰 비중을 차지하는 것으로 확인됨

그림 4. 연구개발단계별 연구비 규모 및 비율

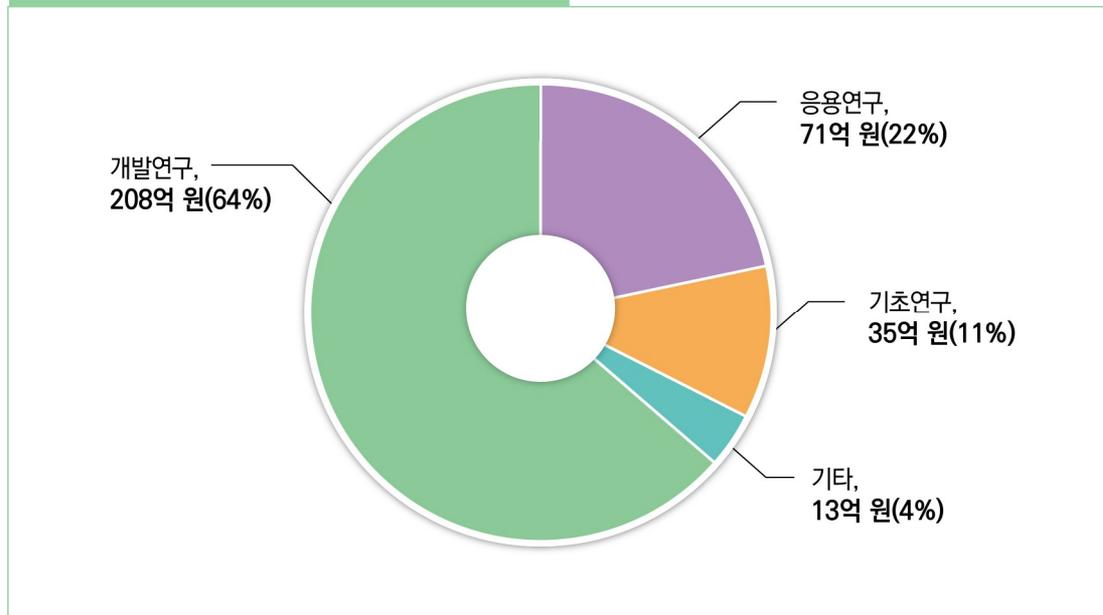


그림 5. 연구개발성격별 연구비 규모 및 비율

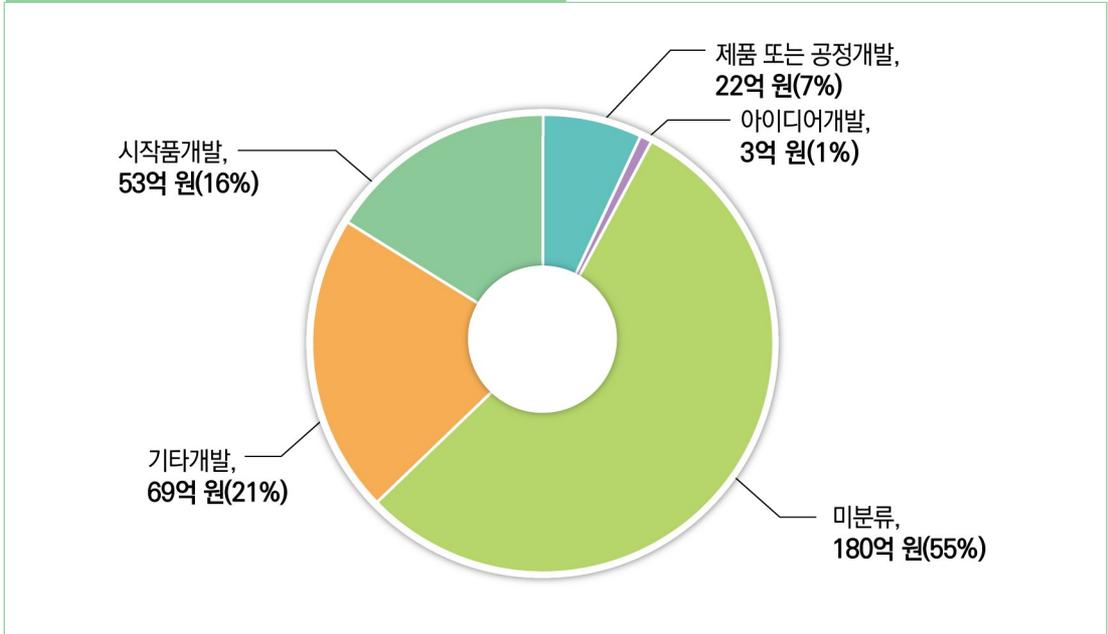
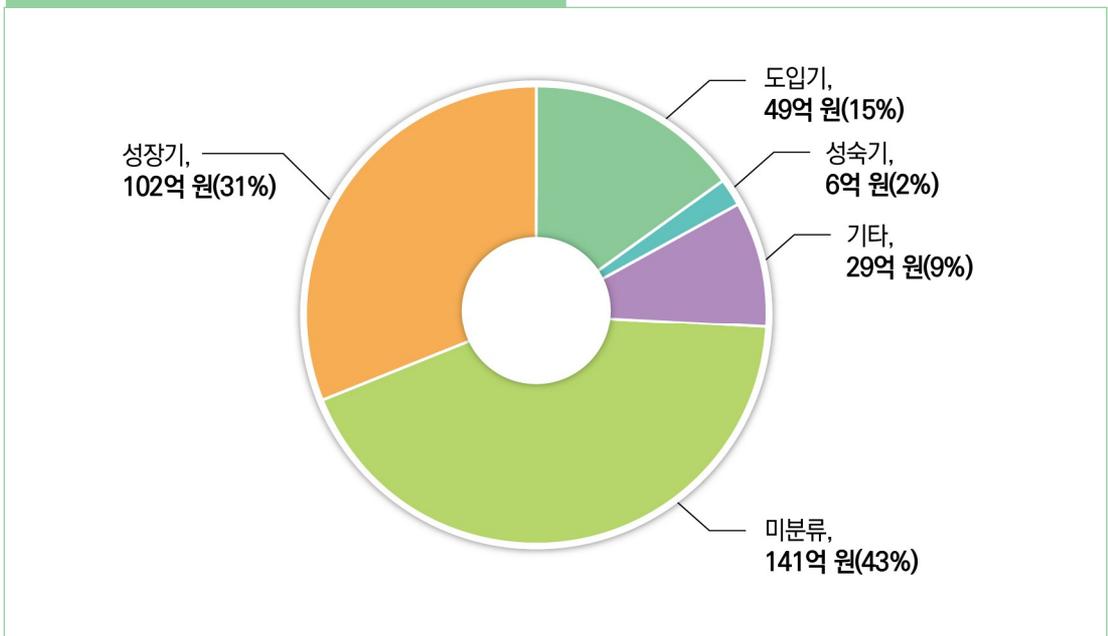


그림 6. 기술수명주기별 연구비 규모 및 비율



□ **(연구분야)** 국가과학기술표준분류와 미래유망신기술분류(6T) 분석 결과, 차세대 전력 반도체 소자 기술에 대한 연구비 투자는 정보통신 기술(IT) 및 전기/전자 분야를 위주로 연구비 투자가 이루어짐

- **(국가과학기술표준분류 분석 결과)** 국가과학기술표준분류 중 전기/전자 분야에 투자되는 연구비 비중이 57%(187억 원)로 가장 큰 것으로 확인되었으며, 에너지/자원 분야(24%, 78억 원), 정보/통신 분야(6%, 21억 원)가 그 다음으로 큰 것으로 확인됨

※ 연구책임자가 최대 3개까지 지정한 국가과학기술표준분류의 대분류에 대한 각 가중치를 고려한 결과임

- 차세대 전력 반도체 소자 기술에 대한 연구 중 융합과제에 지원된 연구비 비중은 36%(119억 원)인 것으로 나타남

※ 융합과제란 연구책임자가 지정한 국가과학기술표준분류의 대분류가 두 개 이상의 분류에 해당하는 과제를 의미함

- **(미래유망신기술분류(6T) 결과)** 정보통신 기술(IT) 관련 연구에 대한 연구비 비중이 39%(127억 원)로 가장 크고, 환경 기술(ET) 관련 연구(32%, 104억 원), 나노 기술(NT) 관련 연구(9%, 31억 원) 순으로 큰 것으로 드러남

그림 7. 국가과학기술표준분류별 연구비 규모 및 비율

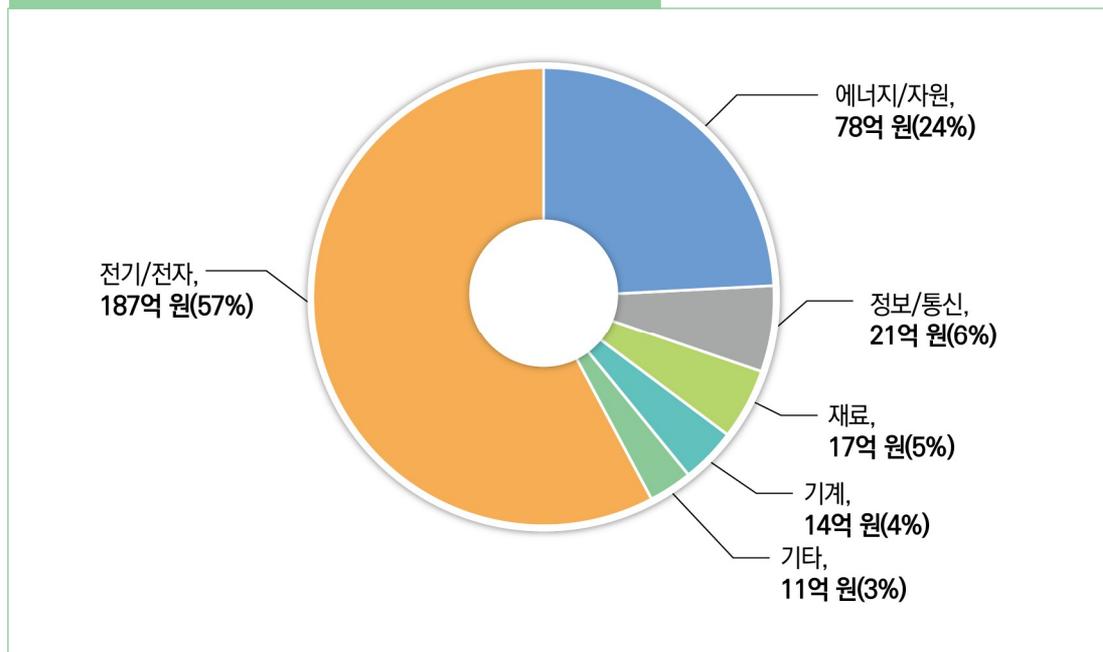


그림 8. 융합 R&D 과제 연구비 규모 및 비율

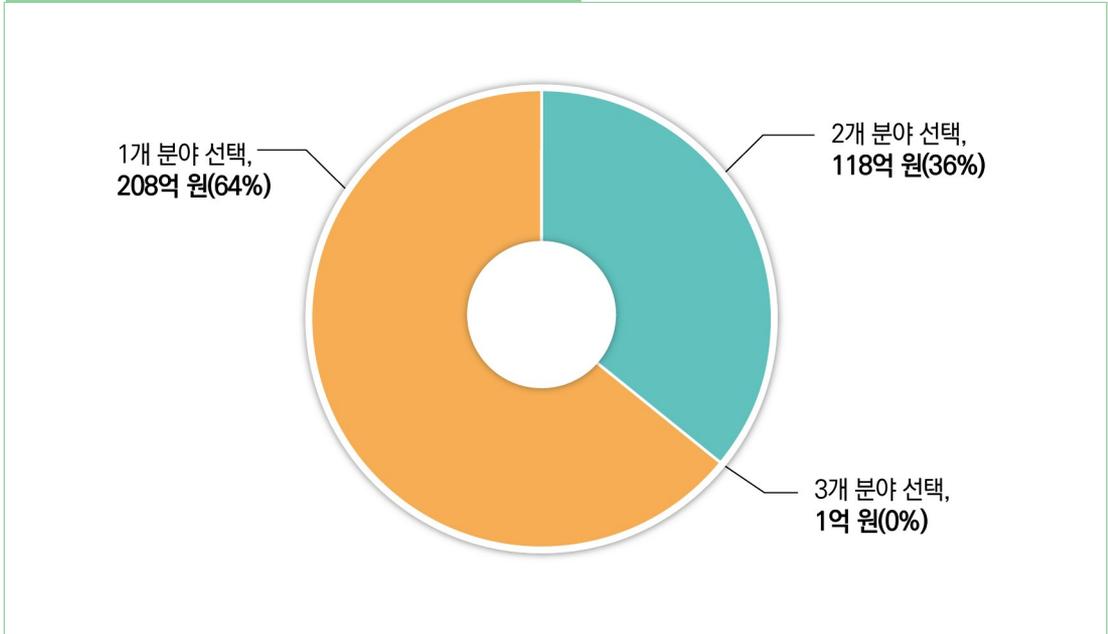
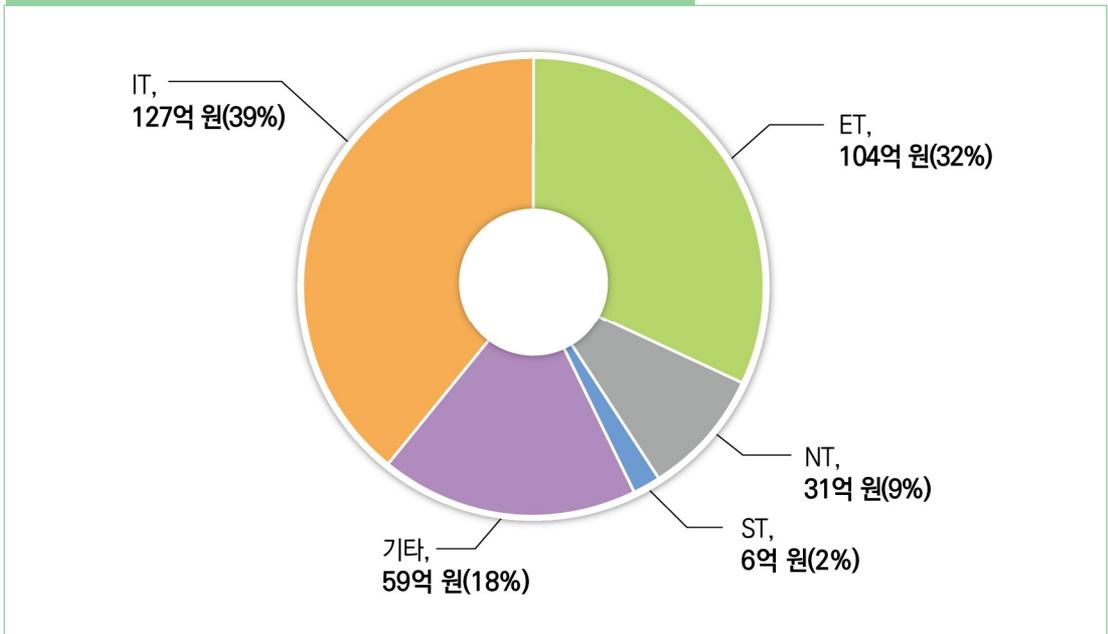


그림 9. 미래유망 신기술분류(6T)별 연구비 규모 및 비율



□ (주요 과제) 원고의 주요 내용 및 키워드 등을 기준으로 선정함

과제명 (사업명, 부처명)	수행기관, 총 연구기간, 연구비 규모	과제 주요 내용
반도체 소자 기반 대용량 스위치 모듈 국산화 개발 (한국전기연구원연구운영비지원, 과학기술정보통신부)	한국전기연구원, 2021-2023년, 7억 원('21)	IGBT 기반 15kV, 1kA급 반도체 소자 기반 대용량 스위치 개발 및 SiC 전력반도체 1.2kV/50A 단위소자 설계
초소형 고효율 반도체 변압기 기반 나노 전력그리드 게이트웨이 개발 (기후변화대응기술개발, 과학기술정보통신부)	중앙대학교, 2021-2024년, 2억 원('21)	실시간 전기요금제, 지능형 수요관리 등에 최적 대응을 위해 16 채널과 100 노드 이상을 수용하는 나노 전력그리드 게이트웨이 개발
능동전력디커플링과 와이드밴드갭 스위칭 소자를 이용한 고전력밀도 양방향 컨버터 개발 (개인기초연구, 과학기술정보통신부)	홍익대학교(세종캠퍼스), 2017-2021년, 0.6억 원('20)	지능형 전기자동차용 전력변환장치 및 신재생에너지용 전력변환 장치를 위한 고전력밀도 컨버터 기술 개발

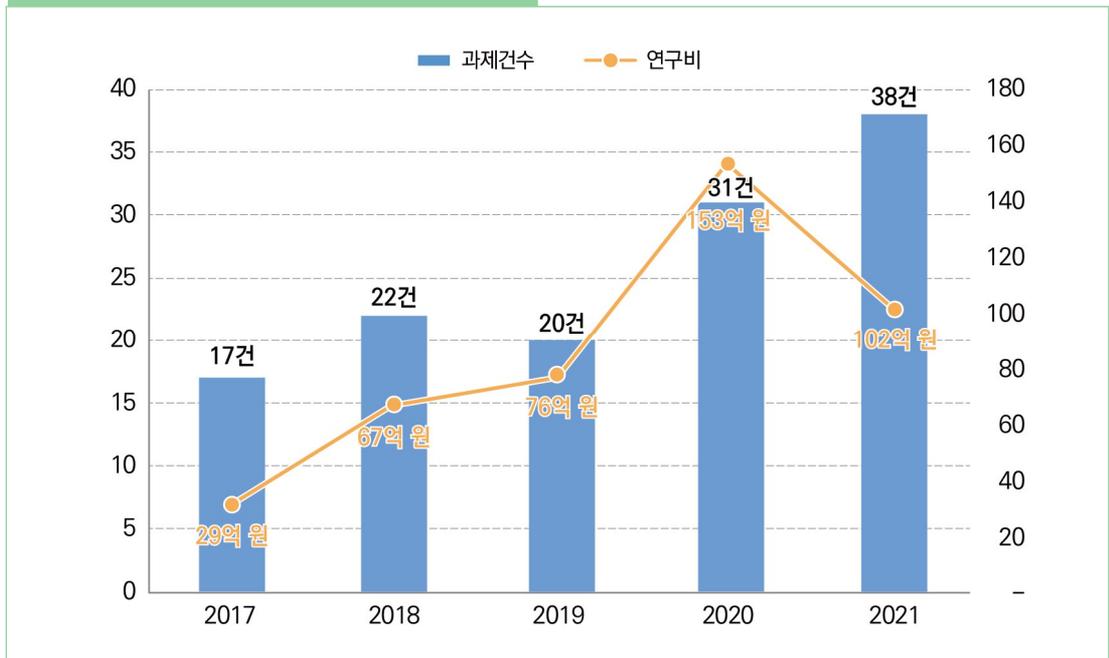
II 인공지능 반도체 기술

□ (총괄) 최근 5년간('17~'21) 총 128건의 과제에 대해 427억 원의 연구비가 투자됨

※ 국가과학기술지식정보서비스(NTIS) 플랫폼을 기반으로 관련 국가 연구개발 과제 분석 수행 : 원고의 핵심 키워드를 고려하였으며, 3세대 인공지능 반도체인 뉴로모픽 반도체에 한정하여 검색 실시

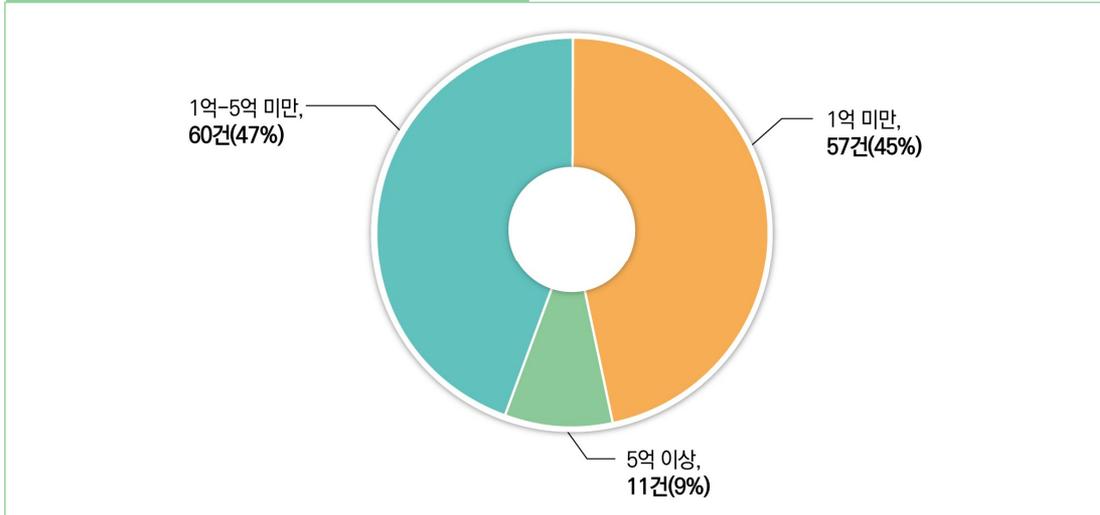
* (뉴로모픽) and (칩 or 반도체)

그림 10. 연도별 연구과제 건수 및 연구비



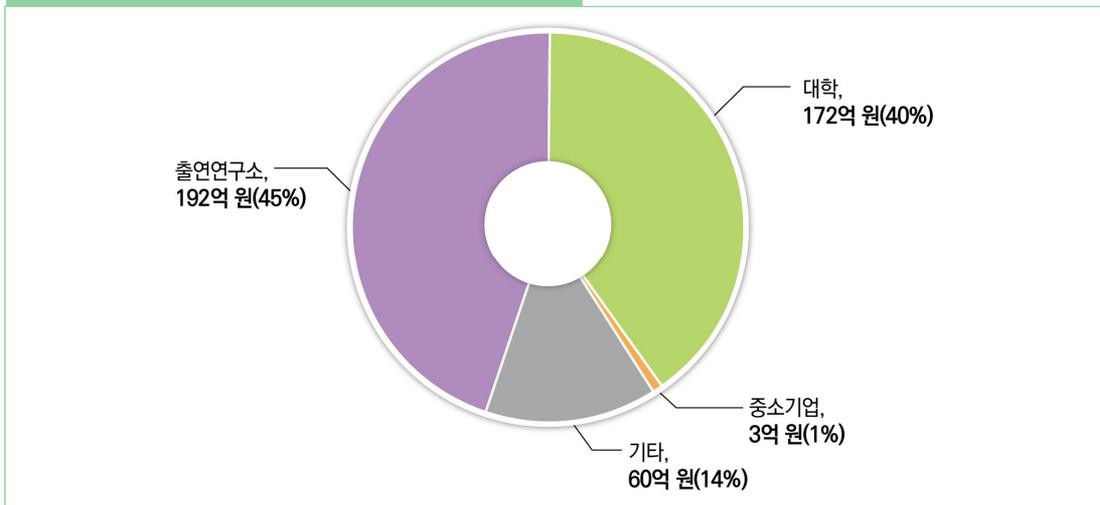
□ (연구비 규모별 과제 수) 1억 원 이상 5억 원 미만인 과제와 1억 원 미만인 과제의 연구비가 각각 47%(60건), 45%(57건)를 차지하고, 5억 원 이상인 과제의 연구비 비중은 9%(11건)에 불과한 것으로 나타남

그림 11. 연구비 규모별 과제 수 및 비율



□ (연구수행주체) 연구수행주체 별로는 출연연구소(45%, 192억 원), 대학(40%, 172억 원) 중심으로 연구비가 지원되는 것으로 드러남

그림 12. 연구수행주체별 연구비 규모 및 비율



□ (연구수준) 3세대 인공지능 반도체 기술은 도입기이며 기초연구 단계인 것으로 확인됨

- (연구개발단계 분석 결과) 기초연구에 투자되는 연구비 비중이 47%(200억 원)로 가장 크고, 응용연구와 개발연구에 투자되는 연구비 비중은 각각 38%(161억 원)과 7%(29억 원)로 나타남
- (연구개발성격 분석 결과) 시작품 개발 관련 연구에 투자되는 연구비가 29%(122억 원)로 가장 큰 비중을 차지하고, 아이디어 개발(6%, 24억 원), 제품 또는 공정개발(1%, 5억 원) 순으로 비중이 큰 것으로 나타남
- (기술수명주기 분석 결과) 도입기에 투자되는 연구비 비중(41%, 176억 원)은 성장기(4%, 16억 원) 연구 대비 약 10배 이상 큰 것으로 나타남

그림 13. 연구개발단계별 연구비 규모 및 비율

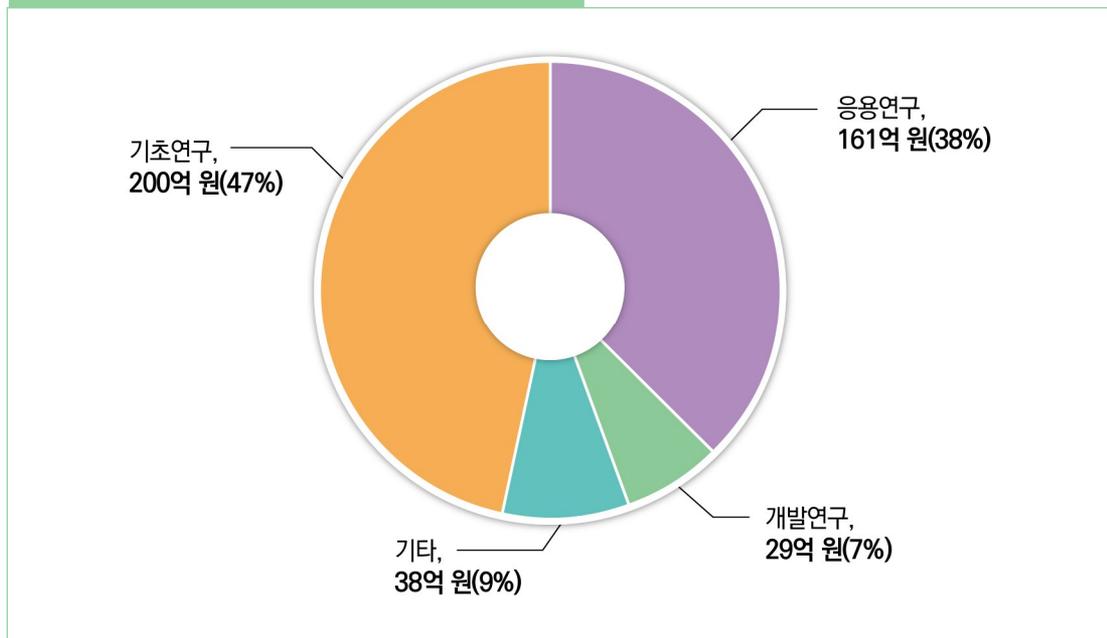


그림 14. 연구개발성격별 연구비 규모 및 비율

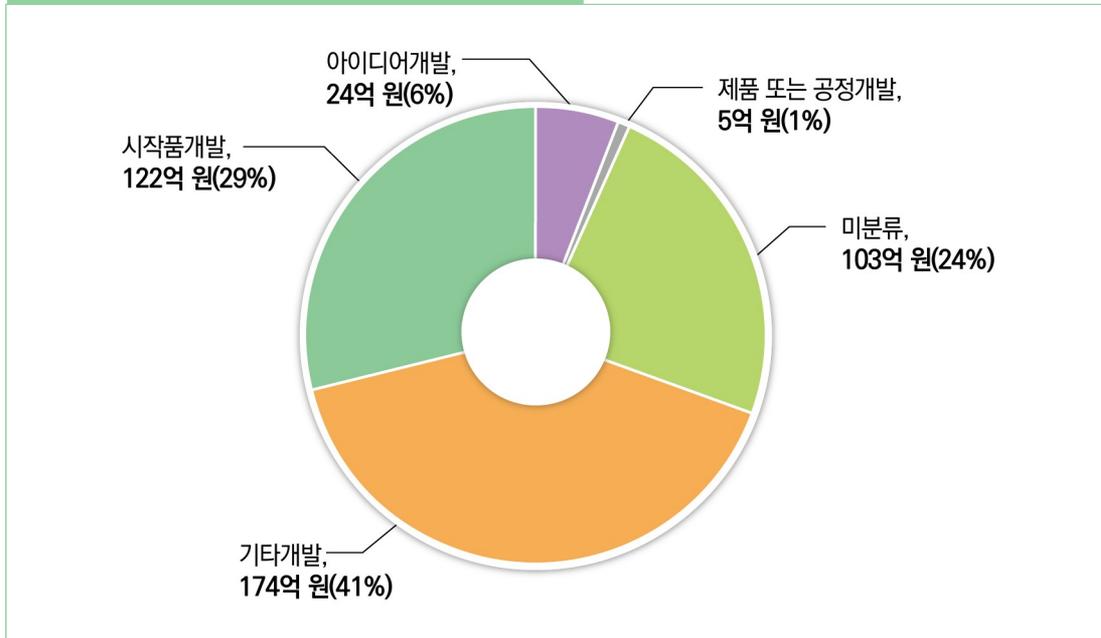
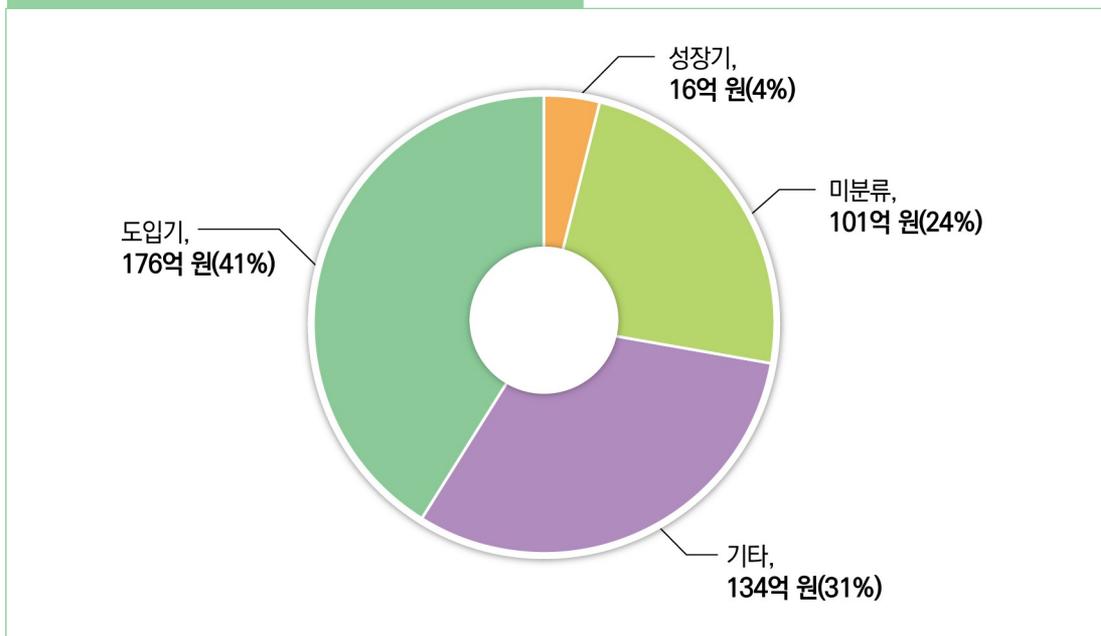


그림 15. 기술수명주기별 연구비 규모 및 비율



□ **(연구분야)** 국가과학기술표준분류와 미래유망신기술분류(6T) 분석 결과, 전기/전자 분야 및 정보통신기술(IT) 분야를 중심으로 3세대 인공지능 반도체 연구가 이루어짐

- **(국가과학기술표준분류 분석 결과)** 전기/전자 분야(69%, 295억 원), 재료 분야 10%(44억 원) 순으로 연구비 비중이 크고, 뇌과학 및 정보통신 분야의 연구비 비중은 7%로 동일한 것으로 확인됨
 - ※ 연구책임자가 최대 3개까지 지정한 국가과학기술표준분류의 대분류에 대한 각 가중치를 고려한 결과임
 - 융합과제에 지원된 연구비 비중은 3세대 인공지능 반도체 연구에 투자된 전체 연구비의 43%를 차지하며 약 184억 원이 지원됨
 - ※ 융합과제란 연구책임자가 지정한 국가과학기술표준분류의 대분류가 두 개 이상의 분류에 해당하는 과제를 의미함
- **(미래유망신기술분류(6T) 결과)** IT 관련 연구에 대한 연구비 비중이 78%(334억 원)로 가장 큰 것으로 확인됨

그림 16. 국가과학기술표준분류별 연구비 규모 및 비율

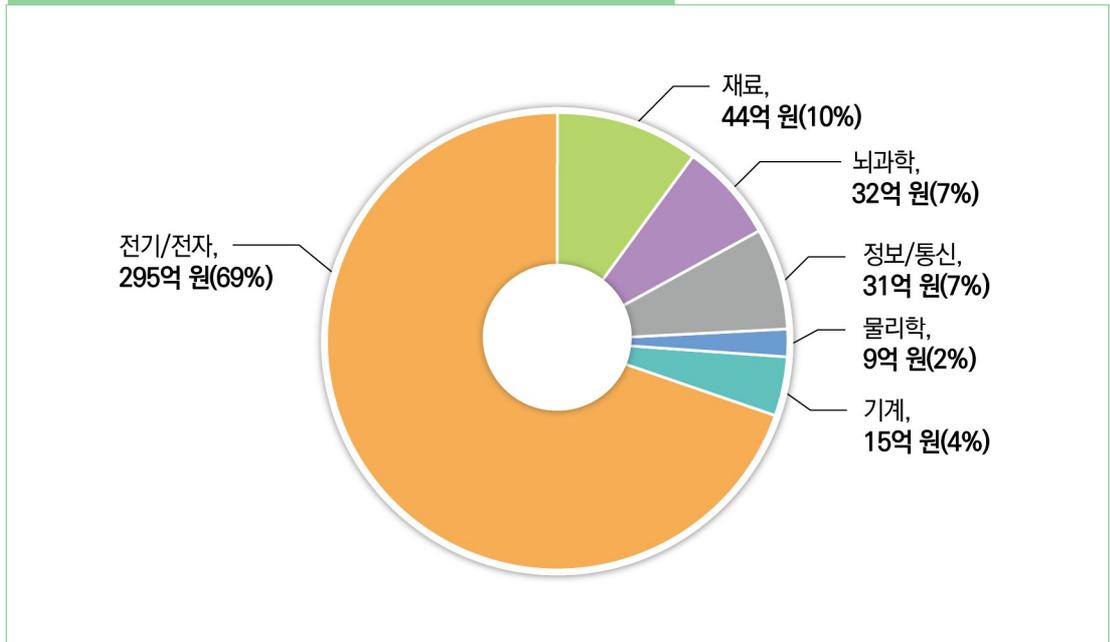


그림 17. 융합 R&D 과제 연구비 규모 및 비율

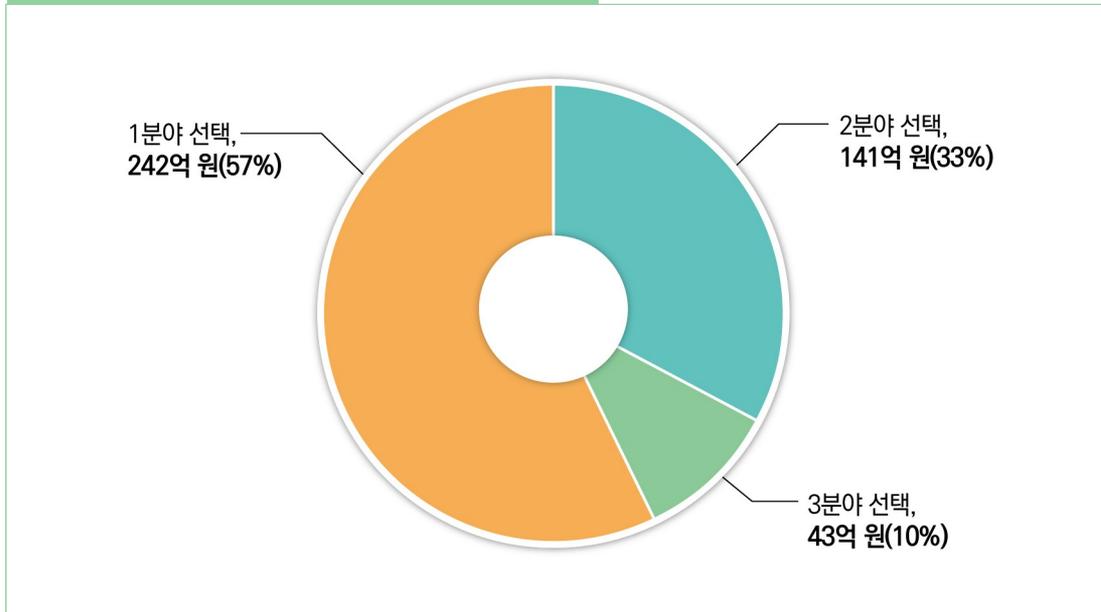
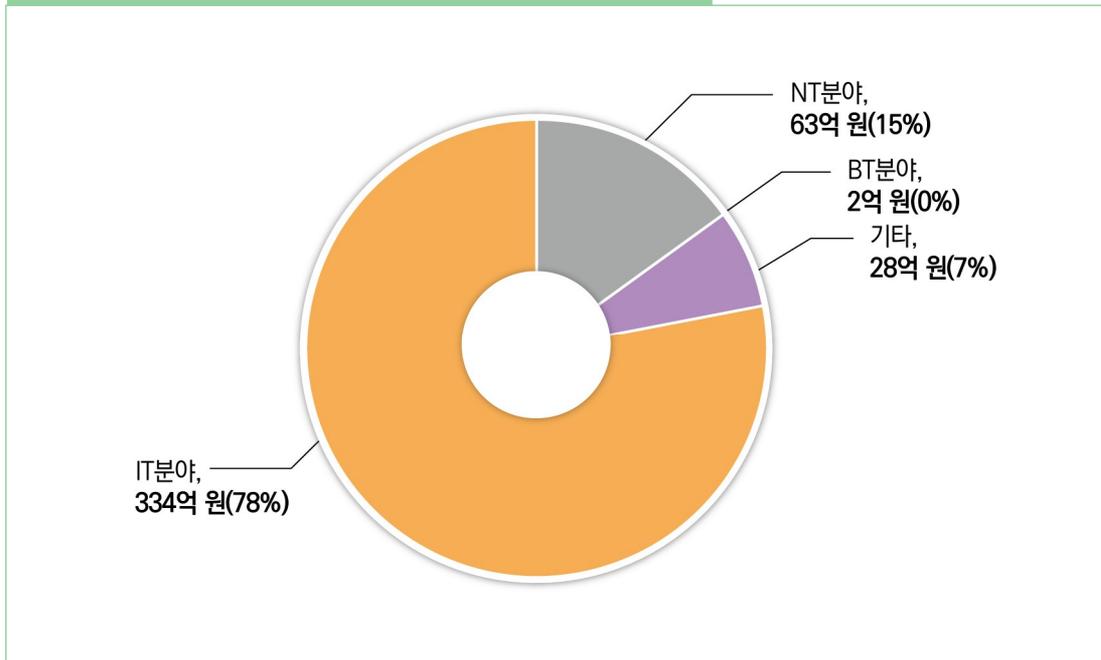


그림 18. 미래유망 신기술분류(6T)별 연구비 규모 및 비율



□ (주요 과제) 원고의 주요 내용 및 키워드 등을 기준으로 선정함

과제명 (사업명, 부처명)	수행기관, 총 연구기간, 연구비 규모	과제 주요 내용
수직 적층 인공지능 플랫폼을 위한 신소자 기반 초고집적 적층형 시냅스 어레이 및 CMOS 집적회로 (차세대지능형반도체기술개발, 과학기술정보통신부)	서울대학교, 2021-2023년, 10억 원('21)	Flash 메모리, RRAM, PRAM, 그리고 FeFET 등 다양한 메모리 플랫폼을 활용하여 뉴로모픽 시스템에 적합한 수직 적층 3D 시냅스 어레이와 시냅스 어레이 구동을 위한 회로를 설계·제작하고 VSOC(Vertical Stack Over Circuit) 구조가 통합 집적된 뉴로모픽 칩 플랫폼 제작
반데르발스 이종접합 기반 멀티터미널 플로팅게이트 멤리스터 및 이를 이용한 비지도학습 뉴로모픽 시스템 개발 (개인기초연구, 과학기술정보통신부)	성균관대학교 (자연과학캠퍼스), 2021-2026년, 2억 원('21)	멀티 터미널(multi-terminal)구조의 플로팅게이트 멤리스터 (MT-FGMEM)를 기반으로 한 인공뉴런 및 인공시냅스를 집적한 뉴로모픽 시스템 개발
듀얼 터미널 자극형 헤테로시냅스 소자 및 어레이 개발 (이공학기술연구기반구축, 교육부)	충북대학교, 2021-2024년, 0.4억 원('21)	듀얼 터미널(소스 및 게이트)에 인가된 전기 자극을 통해 시냅스의 가중치 업데이트를 최적화 할 수 있는 초박막 2차원 반도체와 산화물 기반의 저전력 멤트랜지스터 시냅스 소자·어레이 회로 개발

융합연구리뷰

Convergence Research Review 2022 August vol.8 no.8

이 보고서는 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 사업임

(No. NRF-2012M3C1A1050726)