

DIGITAL INSIGHT

Digital Insight 2024

반도체 기술발전과 미래컴퓨팅 기술의 진화(상)

Contents

I 반도체 기술 개요 및 동향	1 반도체 기술의 현재와 미래 ·
	2 글로벌 반도체 산업 동향 · 3

II 반도체 제조 공정 혁신	1 미세공정 기술의 진화 · 6
	2 3D 적층 및 첨단 패키징 기술 · 10

III 차세대 반도체 아키텍처	1 GAA (Gate-All-Around) 트랜지스터 · 15
	2 3차원 집적 기술 · 18

IV 모바일 엣지 컴퓨팅을 위한 저전력 반도체 기술	1 저전력 설계/소자 기술 · 27
	2 공정 기술 최적화 · 29
	3 시스템 레벨 최적화 · 30

V 차세대 컴퓨팅 패러다임을 위한 반도체 기술 개발 방향	1 엣지 AI를 위한 반도체 기술 · 33
	2 뉴로모픽 컴퓨팅을 위한 반도체 기술 · 35
	3 그 외, 새로운 컴퓨팅 패러다임 탐색 · 36

VI 참고 문헌	참고 문헌 · 38
-----------------	------------

VII 용어정리집	용어정리집 · 39
------------------	------------

그림 목차

[그림 1] DUV, EUV 및 High-NA EUV 기술의 노드 한계, 트랜지스터 밀도와 공정 복잡도 [1]····	8
[그림 2] ALE 기술(위)과 ALD 기술(아래) [2]····	9
[그림 3] 트랜지스터 구조의 발전 [3]····	10
[그림 4] 3D 적층 및 패키징 기술의 발전 [4]····	12
[그림 5] 나노와이어 GAA 트랜지스터(좌)와 나노시트 GAA 트랜지스터(우) [5]····	17
[그림 6] 고대역폭 메모리 (HBM) [6]····	23
[그림 7] 전통적인 벌크 실리콘 기반 트랜지스터(좌)와 FDSOI 기술이 적용된 트랜지스터(우) [7]····	30
[그림 8] Heterogeneous Computing 기술이 적용된 Qualcomm 사의 AI Engine [8] ·····	31
[그림 9] 전통적인 폰 노이만 구조의 컴퓨팅(위)과 인메모리 컴퓨팅(아래) 구조 [9]····	31



반도체 기술 개요 및 동향

1. 반도체 기술의 현재와 미래

반도체 기술은 현대 디지털 사회의 근간을 이루는 핵심 기술로, 지난 수십 년간 놀라운 속도로 발전해 왔다. 1965년 고든 무어가 제시한 '무어의 법칙'은 반도체 산업의 발전 속도를 예측하는 지표로 작용해 왔으며, 이에 따라 시스템 및 메모리 반도체 제품 내 트랜지스터의 집적도는 꾸준히 증가해 왔다. 그러나 최근 물리적 한계에 도달함에 따라 새로운 기술 혁신이 요구되고 있다.

현재 반도체 기술 이해를 위해, 몇 가지 관점에서 반도체 기술에 대한 내용을 아래와 같이 정리해 볼 수 있다.

(초미세 공정) 현재 가장 앞선 반도체 양산 공정 기술은 3nm(나노미터) 수준에 도달했다. TSMC, Samsung, Intel 등 주요 기업들이 해당 기술을 상용화하거나 상용화를 목전에 두고 있으며, 더 작은 기술 노드로의 진화를 위해 노력하고 있다. 이러한 초미세 공정은 트랜지스터의 물리적 크기 및 금속 배선 간의 피치(pitch)를 줄여 동일한 칩 크기에 더 많은 트랜지스터를 집적할 수 있으며 이로써 소자의 집적도를 높이고 전력의 효율성과 성능을 함께 향상시킨다. 첨단 반도체 공정을 개발하고 있는 기업들은 EUV(극자외선) 리소그래피와 같은 첨단 제조 기술을 도입하고 있다. 또한, GAA(Gate-All-Around) 트랜지스터와 같은 새로운 구조의 개발도 활발히 진행 중이다. 이를 통해, PPAC(Power, Performance, Area, Cost) 관점에서 이득이 생기는 방향으로 공정 기술은 진화 발전하게 된다. 그러나 수 nm 급 반도체 공정 기술로의 초미세화는 양자 효과와 같은 물리적 한계 및 양산 가능한 수준의 미세 선폴 달성의 경제적 한계에 직면하고 있다.

(3D 적층 기술) 평면적 집적의 한계를 극복하기 위해 3D 적층 기술이 널리 사용되고 있다. 가장 먼저 평면적 집적의 한계에 봉착한 NAND 제품의 경우, 약 10여년 전부터, 3D NAND 플래시 메모리 제품을 개발/양산하고 있다. 해당 기술은 메모리 셀을 수직으로 쌓아 올려 동일한 면적에서 더 많은 데이터를 저장할 수 있게 한다. 최근에는 200층 이상의 3D NAND가 상용화되었으며, 300층 이상의 3D NAND 시제품은 완성되었으며, 400층 이상의 3D NAND 시제품은 연구 중이다. 향후 DRAM 제품 역시, 4F² 시대를 지나, 2030년 전후로 3D DRAM 제품이 시장에 출시 될

것으로 기대된다. 마지막으로, 시스템 반도체 제품의 경우, 칩렛(Chiplet) 기술을 통해 여러 개의 작은 칩을 3D로 결합하는 방식이 도입되고 있어, 성능 향상과 비용 절감을 동시에 추구하고 있다. 정리하면, 기술적/경제적인 이유로 평면적 집적의 한계에 도달하였고, 이를 극복하기 위해 3차원 적층 기술이 시스템/메모리 반도체 제품에 적용/확대되고 있다.

(AI 가속기) 인공지능(특히 생성형 인공지능)과 기계학습(특히 딥러닝)의 발전에 따라, 이를 위한 특화 반도체인 AI 가속기의 개발이 활발히 이루어지고 있다. NVIDIA의 GPU, Google의 TPU, AMD의 MI 시리즈, 인텔의 가우디, 리벨리온의 아톰 및 리벨, 퓨리오사 AI의 워보이 및 레니게이드 등이 대표적이다. 이러한 AI 가속기는 대규모 병렬 처리와 행렬 연산에 최적화되어 있어, 딥러닝 모델의 학습과 추론 속도를 크게 향상시킨다. 최근에는 엣지 디바이스용 저전력 AI 칩 개발도 활발히 진행되고 있으며, 신경망 구조를 하드웨어 레벨에서 모방한 뉴로모픽 칩 등 새로운 개념의 AI 반도체도 연구되고 있다.

(저전력 기술) 모바일 기기와 IoT(Internet of Things) 장치의 확산으로 인해 저전력 반도체 기술의 중요성이 더욱 커지고 있다. 이는 엣지 컴퓨팅의 발전과도 밀접한 관련이 있다. 저전력 기술은 다양한 레벨에서 구현되고 있는데, 트랜지스터 레벨에서는 FD-SOI(Fully Depleted Silicon On Insulator) 기술이, 회로 설계 레벨에서는 동적 전압 및 주파수 조절(DVFS) 기술이 적용되고 있다. 또한, 비휘발성 메모리를 활용한 노멀리-오프 컴퓨팅 기술도 저전력화에 기여하고 있다. 이러한 기술들은 배터리 수명 연장뿐만 아니라, 데이터 센터의 에너지 효율성 향상에도 큰 역할을 하고 있다.

향후 우리가 접하게 될 미래 반도체 기술은 다음과 같은 방향으로 발전할 것으로 예상된다.

(차세대 트랜지스터 구조) 핀 전계 효과 트랜지스터(FinFET)를 넘어 GAA(Gate-All-Around) 구조의 트랜지스터, 그리고, GAA 트랜지스터를 위아래로 적층한 CFET 등이 주목받고 있다. 이는 주어진 칩 면적에서, 더 높은 성능과 에너지 효율성을 제공할 것으로 기대된다.

(새로운 반도체 소재의 활용) 현재 대부분의 반도체는 실리콘으로 만들어지지만, 고성능과 저전력 요구 등 실리콘 소재의 한계 극복을 위해 화합물 반도체(GaN, SiC 등)와 2D 소재(그래핀, MoS2 등)에 대한 연구가 활발히 진행되고 있다. 개별 반도체 칩에서 개별 소재를 활용한 반도체 제품이 완성된다고 가정한다면, 하이브리드 본딩 기술(Cu-to-Cu and Oxide-to-Oxide 접합 기술)을 활용하여, 서로 다른 종류의 반도체 칩을 붙임으로써, 새로운 반도체 소재 기반의 칩과 기존 실리콘 반도체 소재 기반의 칩이 하나의 제품으로 통합될 수 있을 것으로 보인다.

(양자 컴퓨팅) 기존의 디지털 로직을 넘어서는 양자 컴퓨팅을 위한 반도체 기술 개발이 진행 중이다. 물론 양자 컴퓨팅이 디지털 컴퓨팅을 압도하는 기술은 아니며, 일부 문제를 푸는 과정의 경우, 양자 컴퓨팅이 디지털 컴퓨팅보다 압도적인 우위를 보이는 것이다. 실제, 양자 컴퓨팅을 디지털 컴퓨팅보다 더욱 효율적으로 활용할 수 있는 암호화, 신약 개발, 기후 모델링 등 복잡한 문제 해결에 혁명적인 변화를 가져올 것으로 기대된다.

(뉴로모픽 컴퓨팅) 인간의 뇌를 모방한 뉴로모픽 컴퓨팅 기술이 발전하고 있다. 이는 저전력으로 효율적인 인공지능 처리를 가능하게 할 것으로 예상된다. 기존 디지털 컴퓨팅과 달리 뉴로모픽 컴퓨팅은 연산과 기억을 동시에 수행함으로써, 에너지 효율적인 연산이 가능하다. 뿐만아니라, 미래 뉴로모픽 컴퓨팅은 짧은 전기신호(spiking signal)를 사용함으로써, 디지털 컴퓨팅의 클럭 신호(clock signal)보다 훨씬 낮은 수준의 에너지를 사용한 연산/기억 기능을 구현할 수 있을 것이다.

(첨단 패키징 기술) 칩렛(Chiplet) 기술 및 하이브리드 본딩 기술과 같은 혁신적인 첨단 패키징 방식을 통해 성능 향상과 비용 절감을 동시에 달성하려는 노력이 계속될 것이다.

2. 글로벌 반도체 산업 동향

반도체 산업은 기업 간 경쟁을 넘어 국가 간 경쟁으로 인식될 만큼, 글로벌 경제에서 핵심적인 위치를 차지하고 있으며, 최근에는 지정학적 요인, 경제 안보 요인 등으로 기술 패권 경쟁의 중심에서 있는 첨단 산업이다.

2.1. 글로벌 반도체 시장 동향 개요

(시장 규모) 세계반도체무역통계기구(WSTS)에 따르면, 2022년 글로벌 반도체 시장 규모는 약 5,800억 달러에 달했으며, 2030년까지 1조 달러 규모로 성장할 것으로 전망된다. 1990년 PC 시대(전 세계 2천 5백만대 이상 판매), 2003년 모바일 시대 개화(블랙베리 스마트폰 출시), 2008년 모바일 시대 본격 시작(애플 iPhone 및 삼성 갤럭시 출시), 2018년 AI 시대 개화(사물인터넷 만들어 낸 데이터 >> 인류가 만든 데이터), 2022년 AI 시대 본격 시작(OpenAI사의 2022년 11월 ChatGPT 출시 이후) 등 모든 정보통신 기술의 발전 이면에는 반도체 시장의 동반 성장이 있었다.

(수요 증가) 차세대 통신(6G), 인공지능, 자율주행차, IoT, 휴머노이드 로봇 등 신기술의 발전으로 반도체 수요는 앞으로도 꾸준히 증가할 것으로 기대된다. 특히 데이터 센터와 엣지 컴퓨팅 분야에서 수요가 급증하고 있다. AI시대 도래와 함께, 다양한 산업 분야에 인공지능 서비스가 보편화

되면서, 반도체 수요는 더욱 더 폭발적으로 성장할 것으로 예상된다.

(공급망 재편) COVID-19 팬데믹과 미중 무역 갈등 등으로 인해 반도체 공급망의 안정성이 중요한 이슈로 부각되었다. 이에 따라 각국 정부와 기업들이 자국 내 반도체 생산 능력 확보에 나서고 있다. 특히, 중국의 one-china라는 정책 하에, 첨단 반도체 생산 시설을 갖춘 대만은 지정학적 갈등의 중심지로 주목받고 있다. 이를 극복하고자, 대만은 일본과 mini-GVC(global value chain)을 형성 하였고, 향후 대만의 반도체 영토를 일본으로 확대하고 있다. 또한, 미국은 첨단 반도체 제조 시설을 북미 본토 내에 확보하고자, 대규모 직접 보조금을 해외 반도체 기업에게 지원하고 있다. 반도체 산업의 특성상, 모든 공급망을 하나의 국가가 다 가지는 것은 현실적/경제적으로 불가능하나, 핵심적인 기술적 요소나 공급망의 주요 요인이 될 수 있는 기술을 각 국가가 확보하는 일은 앞으로 더욱 치열해질 것이다.

2.2. 주요 기업 및 국가 동향

(기업 동향) 파운드리 분야에서는 TSMC가 선두를 유지하고 있으며, Samsung과 Intel이 추격하고 있다. 하지만, 그 격차가 나날이 확대되고 있고, 2위/3위 업체의 기술 개발에 어려움이 있기 때문에, 당분간 TSMC의 독과점 체제는 지속될 것이다. 메모리 분야에서는 Samsung, SK hynix, Micron이 시장을 주도하고 있다. 특히 DRAM은 3개 회사의 과점 체제가 안정적으로 유지되면서, 특정 메모리 제품에 있어서는 SK하이닉스가 가장 높은 수준의 압도적인 기술력을 갖추고 있다. NAND는 여전히 미국/일본/한국/중국 기업간의 경쟁이 상존하며, 향후 NAND 시장의 지형 변화가 있을 것으로 예상된다. 팹리스 기업으로는 Nvidia, Qualcomm, AMD 등이 높은 성장세를 보이고 있다. 특히 대만의 미디어텍은 자국 내 파운드리 생태계를 심분 활용하여, 더욱 더 빠른 속도로 성장할 것으로 예상된다.

(국가별 전략) 미국의 경우, 반도체 및 과학법(CHIPS and Science Act)를 통해 국내 반도체 생산 확대 및 기술 혁신을 지원하고 있다. NSTC 설립을 가시화하였고, 비경쟁 초기 단계의 기술 연구에 박차를 가하고 있다. 중국의 경우, '중국제조 2025' 전략의 일환으로 반도체 자급률 향상을 위해 노력하고 있으나, 미국의 기술 제재로 어려움을 겪고 있다. 하지만, 중국은 자국의 충분한 레거시(legacy) 반도체 칩에 대한 수요가 상당하기 때문에, 미국과의 경쟁이라기보다는 중국 내 반도체 자급화를 위한 다양한 국가 전략을 수립하고 있다. 미국과의 경쟁이라고 바라보는 관점도 일견 맞으나, 중국 반도체 전략은 레거시 반도체 칩위주의 국산화에 있다. 유럽의 경우, '(유럽 반도체 법)European Chips Act'를 통해 2030년까지 글로벌 반도체 생산의 20% 점유를 목표로 하고 있다. 다만, 생산 역량을 갖춘 기업이

없기 때문에, 소재/장비 산업, 그리고 자동차 반도체 위주의 성장이 기대된다. 한국의 경우, 'K-반도체 전략'을 수립하여 세계 최고의 반도체 공급망 구축을 목표로 하고 있다. 메모리 강국의 위상을 지속적으로 지켜나가는 일이 그 무엇보다 중요할 것이며, 이를 위해 DRAM/NAND의 미래 기술 리더십의 지속적인 확보와 메모리 반도체 생태계 확충에 노력해야 할 것이다.

2.3. 기술 및 산업 트렌드

(특화 반도체 수요 증가) 범용 반도체보다는 AI, 자율주행, 엣지 컴퓨팅 등 특정 용도에 최적화된 반도체의 수요가 증가하고 있다. 즉 범용 반도체 제품의 시대는 지난 수십년 간의 성장만큼의 성장을 기대하기 어려울 것이다. 고객이 희망하는 고객맞춤형(customized) 제품을 설계/생산할 수 있는 기업이 다가오는 시대의 승자가 될 것이다.

(친환경 반도체) 기후 변화를 최소화 하기 위한 탄소중립 목표 달성을 위해 저전력, 고효율 반도체 개발이 강조되고 있다. 또한 제조 과정에서의 환경 영향을 줄이기 위한 노력도 계속되고 있다. ESG 경영 지표가 투자자들의 큰 관심이 되었으며, 이에 발맞추어서, 반도체 제조 공정상 기후변화 유발 가스를 최소화하는 차세대 공정 기술 개발이 매우 중요해지고 있다.

(오픈소스 하드웨어) RISC-V와 같은 오픈소스 반도체 아키텍처가 주목받고 있으며, 이는 반도체 설계의 민주화와 혁신 가속화로 이어질 것으로 예상된다. 과거에는 특정 기업이 컴퓨팅 패러다임을 독점하였으나, 향후 모든 기업들이 자사 전용 칩을 가지는 시대가 개화될 것이다. 실제 애플은 자사 제품에 적합한 M1, M2 등의 칩을 설계하였다. AI시대의 개화에 발맞추어서, AI 서비스에 특화된 칩을 설계해주는 business model이 주목 받을 것이며, 이를 위해 오픈소스 기반의 하드웨어 칩 설계는 상당한 수준으로 성장할 것이다.

(인공지능의 활용) 반도체 설계 및 제조 과정에 인공지능 기술이 적극적으로 도입되면서 개발 기간 단축 및 성능 최적화가 가능해지고 있다. 특히 디지털 집적회로 설계는 숙련된 엔지니어가 설계하는 수준에 버금가는 수준으로 시가 칩 설계를 할 수 있을 정도이다. 칩 설계 자동화에 AI 기술의 활용이 어디까지 적용될지 귀추가 주목된다.

결론적으로, 반도체 기술은 끊임없는 혁신을 통해 발전하고 있으며, 이는 글로벌 산업 구조와 국가 간 경쟁 구도에도 큰 영향을 미치고 있다. 앞으로도 반도체 기술은 디지털 전환과 4차 산업혁명을 이끄는 핵심 동력으로 작용할 것이며, 이에 따른 기술적, 산업적, 정책적 대응이 중요할 것으로 보인다.



반도체 제조 공정 혁신

반도체 제조 공정의 혁신은 반도체 산업 발전의 핵심 동력이다. 지속적인 공정 혁신을 통해 더 작고, 더 빠르며, 더 효율적인 반도체 생산이 가능해지고 있다. 이 장에서는 미세공정 기술의 진화와 3D 적층 및 첨단 패키징 기술의 발전에 대해 자세히 살펴보겠다.

1. 미세공정 기술의 진화

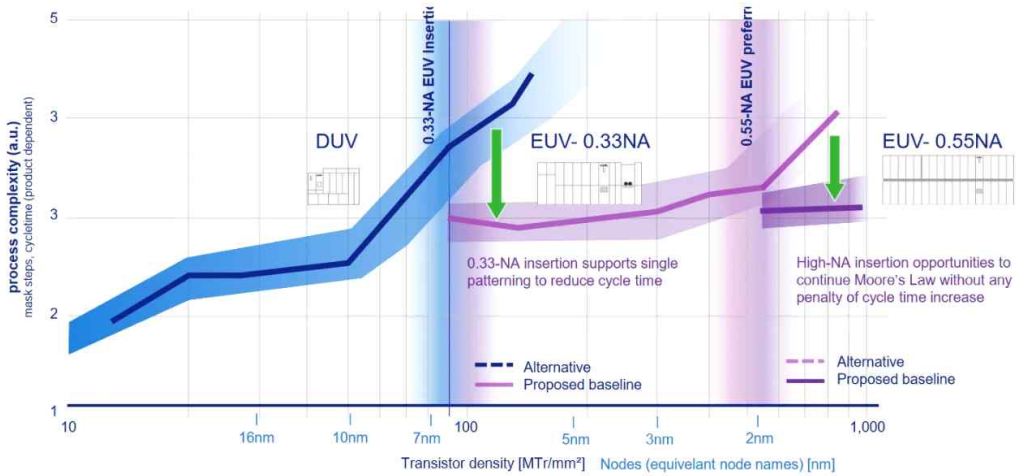
반도체 제조 공정 기술은 지속적인 혁신을 통해 발전해왔으며, 이는 현대 전자기기의 성능 향상과 소형화를 가능케 하는 핵심 요소이다. 특히 미세공정 기술의 진화는 이러한 발전의 중심에 있다. 미세공정 기술은 반도체 칩 상의 트랜지스터와 기타 구성 요소의 크기를 줄이는 기술을 의미한다. 이는 단순히 크기를 줄이는 것에 그치지 않고, 더 많은 기능을 더 작은 공간에 집적할 수 있게 함으로써 전자기기의 성능을 비약적으로 향상시키는 역할을 한다. 이러한 미세공정 기술의 발전은 무어의 법칙을 실현하는 핵심 동력이 되어왔으며, 현재도 계속해서 진화하고 있다.

1.1 포토리소그래피 기술의 발전

미세공정 기술의 발전에 있어 가장 중요한 요소 중 하나는 포토리소그래피 기술이다. 포토리소그래피는 빛을 이용해 반도체 웨이퍼 위에 회로 패턴을 그리는 기술로, 반도체 제조 공정에서 가장 핵심적인 기술 중 하나이다. 이 기술의 발전은 미세공정 실현의 핵심이라고 할 수 있다.

그림 1은 반도체 기술 발전과 포토리소그래피 기술 발전을 함께 보여주고 있다. 2000년대 초반부터 주류 기술로 사용된 DUV(Deep Ultraviolet) 리소그래피는 193nm 파장의 빛을 사용한다. 이 기술은 오랫동안 반도체 산업을 이끌어왔지만, 10nm 이하의 초미세 공정에서는 물리적 한계에 도달하게 되었다. 이러한 한계를 극복하기 위해 개발된 것이 EUV(Extreme Ultraviolet) 리소그래피다. EUV 리소그래피는 13.5nm의 매우 짧은 파장의 빛을 사용하여 더욱 정밀한 패턴 형성을 가능하게 한다. 이 기술은 7nm 이하의 초미세 공정에서 사용되고 있으며, 멀티 패터닝 공정을 줄여 생산성을 크게 향상시킬 수 있다는 장점이 있다. 그러나 EUV 장비의 고가격과 광원 출력 및 안정성 개선 등의 과제도 여전히 남아있다. 포토리소그래피 기술은 계속해서 발전하고 있으며, 현재 EUV의 개선된 버전인 High-NA EUV 기술이 개발 중에 있다. 초기 High-NA EUV 제품은 인텔에 최초

납품이 되었다. 이 기술은 더 높은 해상도를 제공하여 더욱 미세한 패턴 형성을 가능하게 할 것으로 기대된다. 또한, 분자 수준의 자기 조립을 이용한 나노 패터닝 기술인 Directed Self-Assembly (DSA) 등의 차세대 리소그래피 기술도 연구되고 있다. 간단히 각 기술별로 장단점 등을 정리하면 아래와 같다.



[그림 1] DUV, EUV 및 High-NA EUV 기술의 노드 한계, 트랜지스터 밀도와 공정 복잡도 [1]

[DUV(Deep Ultraviolet) 리소그래피]

- 파장: 193nm
- 특징: 2000년대 초반부터 주류 기술로 사용됨
- 한계: 10nm 이하의 공정에서 한계에 도달

[EUV(Extreme Ultraviolet) 리소그래피]

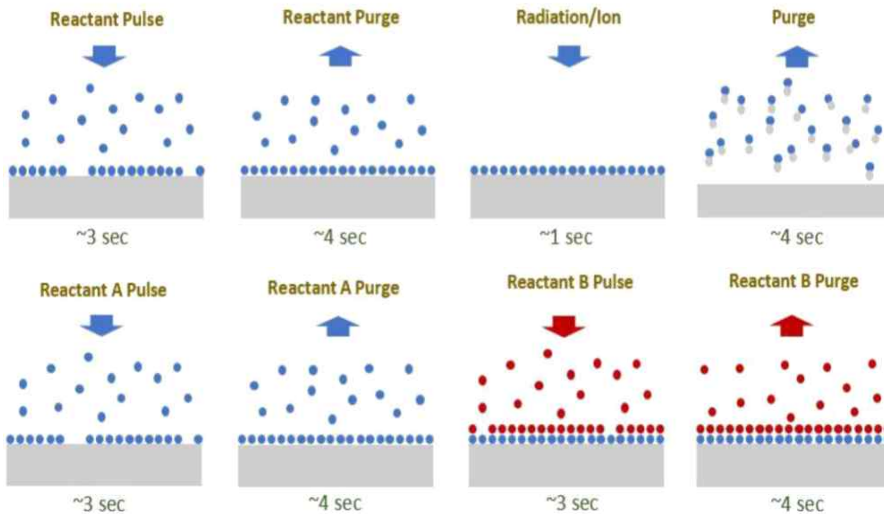
- 파장: 13.5nm
- 특징: 7nm 이하 공정에서 사용, 더 정밀한 패턴 형성 가능
- 장점: 멀티 패터닝 공정 감소, 생산성 향상
- 과제: 고가의 장비, 광원 출력 및 안정성 개선 필요

[차세대 리소그래피 기술]

- High-NA EUV: 현재 EUV의 개선된 버전, 더 높은 해상도 제공
- Directed Self-Assembly (DSA): 분자 수준의 자기 조립을 이용한 나노 패터닝 기술

1.2 식각 및 증착 기술의 혁신

미세공정을 위해서는 정밀한 식각과 증착 기술도 필수적이다. 그림 2는 원자 수준의 정밀한 식각 기술과 증착 기술을 함께 보여주고 있다. 원자층 식각(Atomic Layer Etching, ALE) 기술은 원자 단위로 정밀한 식각을 가능하게 하여, 반도체 소자 표면의 손상을 최소화하고 높은 선택비를 제공한다. 이와 쌍을 이루는 원자층 증착(Atomic Layer Deposition, ALD) 기술은 원자 단위로 박막을 형성하여 균일한 두께와 우수한 단차 피복성을 제공한다. 최근에는 선택적 증착 기술도 주목받고 있다. 이 기술은 특정 영역에만 선택적으로 물질을 증착할 수 있어 패터닝 공정을 간소화하고 정밀도를 향상시킬 수 있다. 이러한 기술들의 발전은 나노미터 수준의 초미세 구조를 형성하는 데 크게 기여하고 있다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.



[그림 2] ALE 기술(위)과 ALD 기술(아래) [2]

[원자층 식각 (Atomic Layer Etching, ALE)]

- 특징: 원자 단위의 정밀한 식각 가능
- 장점: 데미지 최소화, 높은 선택비 실현

[원자층 증착 (Atomic Layer Deposition, ALD)]

- 특징: 원자 단위로 박막을 형성
- 장점: 균일한 두께, 높은 종횡비 구조에서도 우수한 단차 피복성

[선택적 증착 기술]

- 특징: 특정 영역에만 선택적으로 물질을 증착
- 장점: 패터닝 공정 간소화, 정밀도 향상

1.3 계측 및 검사 기술의 발전

나노미터 수준의 공정에서는 정밀한 계측과 검사가 더욱 중요해졌다. 전자빔 계측 기술은 나노미터 수준의 정밀도로 선폭을 측정하고 오버레이 정렬을 검사할 수 있다. X선 계측 기술은 비파괴 검사가 가능하여 3D 구조 분석과 결함 검출에 활용된다. 최근에는 AI 기반 검사 기술도 도입되어 머신러닝을 이용한 패턴 인식 및 불량 검출로 검사 속도와 정확도를 크게 향상시키고 있다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[전자빔 계측 기술]

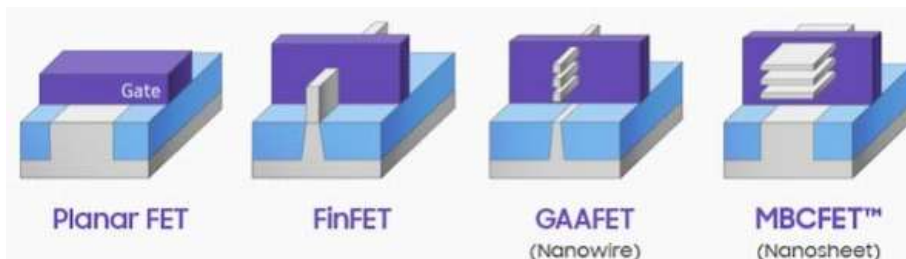
- 특징: 나노미터 수준의 정밀도
- 용도: 선폭 측정, 오버레이 정렬 검사

[X선 계측 기술]

- 특징: 비파괴 검사 가능
- 용도: 3D 구조 분석, 결함 검출

[AI 기반 검사 기술]

- 특징: 머신러닝을 이용한 패턴 인식 및 불량 검출
- 장점: 검사 속도 및 정확도 향상

1.4 새로운 트랜지스터 구조

[그림 3] 트랜지스터 구조의 발전 [3]

미세화가 진행됨에 따라 기존 평면 트랜지스터의 한계를 극복하기 위한 새로운 구조들도 등장했다. FinFET(Fin-shaped Field-Effect Transistor)은 채널이 수직으로 돌출된 구조로, 누설 전류를 감소시키고 전류 구동력을 향상시켜 22nm 이하 공정에서 널리 사용되고 있다. 더 나아가 GAA(Gate-All-Around) 구조는 채널을 게이트가 완전히 감싸는 형태로, FinFET보다 더 우수한 전류 제어 능력을 제공한다. 나노와이어(nano-wire)나 나노시트(nano-sheet) 형태의 GAA 트랜지스터는 3nm 이하 공정에서 도입될 예정이다. 또한, NMOS와 PMOS를 수직으로 쌓은 구조인 CFET(Complementary FET)도 차세대 로직 소자로 주목받고 있다. 이 구조는 집적도를 향상시키고 배선 복잡도를 감소시킬 수 있어(물론 일부 standard cell은 배선 복잡도 감소 효과 미미할 수 있음), 미래의 초고집적 반도체 제조에 중요한 역할을 할 것으로 기대된다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[FinFET (Fin Field-Effect Transistor)]

- 특징: planar MOSFET 대비, 채널이 수직으로 돌출된 구조
- 장점: 누설 전류 감소, 전류 구동력 향상
- 적용: 22nm 이하 공정에서 널리 사용

[GAAFET (Gate-All-Around Field-Effect Transistor)]

- 특징: 채널을 게이트가 완전히 감싸는 구조
- 장점: FinFET 대비 더 우수한 전류 제어 능력
- 종류: 나노와이어, 나노시트
- 적용: 3nm 이하 공정에서 도입 예정

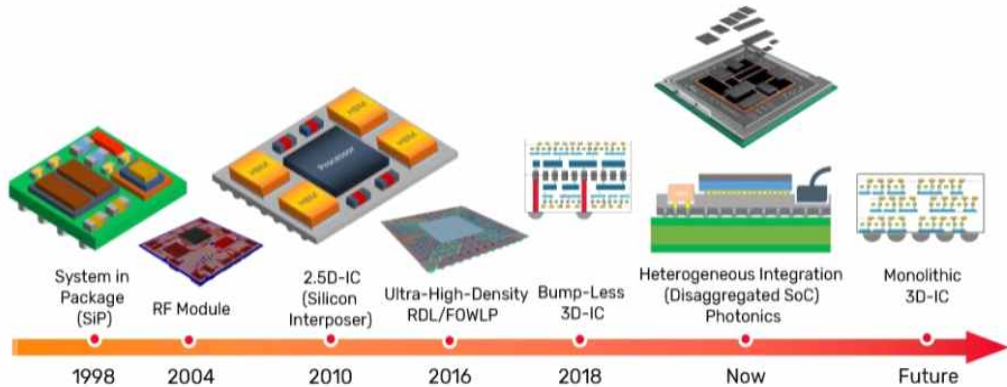
[CFET (Complementary FET)]

- 특징: NMOS와 PMOS를 수직으로 쌓은 구조
- 장점: 집적도 향상, 배선 복잡도 감소
- 전망: 차세대 로직 소자로 주목받고 있음

2. 3D 적층 및 첨단 패키징 기술

2D 평면상의 집적 한계를 극복하기 위해 3D 적층 기술과 첨단 패키징 기술도 빠르게 발전하고 있다. 3D 적층 기술은 반도체 칩을 수직으로 쌓아 올려 집적도를 높이는 기술이다.

TSV(Through-Silicon Via) 기술은 실리콘 웨이퍼를 관통하는 전극을 형성하여 칩 간 연결 거리를 단축하고 신호 지연을 감소시킨다. 이 기술은 고대역폭 메모리(HBM)나 이미지 센서 등에 널리 활용되고 있다.



[그림 4] 3D 적층 및 패키징 기술의 발전 [4]

2.1. 3D 적층 기술

몰딩 컴파운드를 활용한 기술은 칩들을 몰딩 컴파운드로 감싸 적층함으로써 기계적 강도를 향상시키고 열 방출을 개선한다. 이는 모바일 AP나 저밀도 패키지 등에 응용되고 있다. 다이 대 다이 (Die-to-Die) 본딩 기술은 개별 다이를 직접 연결하는 기술로, 마이크로범프나 하이브리드 본딩 등의 방식이 있다. 이 기술은 초미세 피치 연결을 가능하게 하여 전기적 특성을 크게 향상시킬 수 있다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[TSV (Through-Silicon Via) 기술]

- 특징: 실리콘 웨이퍼를 관통하는 전극을 형성
- 장점: 칩 간 연결 거리 단축, 신호 지연 감소
- 응용: 고대역폭 메모리(HBM), 이미지 센서 등

[몰딩 컴파운드 활용 기술]

- 특징: 칩들을 몰딩 컴파운드로 감싸 적층
- 장점: 기계적 강도 향상, 열 방출 개선
- 응용: 모바일 AP, 저밀도 패키지 등

[다이 대 다이 (Die-to-Die) 본딩]

- 특징: 개별 다이를 직접 연결
- 종류: 마이크로범프, 하이브리드 본딩
- 장점: 초미세 피치 연결 가능, 전기적 특성 향상

2.2. 첨단 패키징 기술

첨단 패키징 기술의 발전도 시스템의 성능과 집적도를 높이는 데 중요한 역할을 하고 있다. 팬아웃 웨이퍼 레벨 패키징(Fan-Out Wafer Level Packaging, FOWLP)은 칩 크기보다 큰 영역에 재배선층을 형성하여 고밀도 I/O를 구현하고 소형화와 열 특성 개선을 가능하게 한다. 이 기술은 모바일 AP나 RF 모듈 등에 널리 사용되고 있다. 2.5D 패키징 기술은 실리콘 인터포저라는 얇은 실리콘 층을 통해 여러 칩을 연결하는 방식으로, 이종 집적(heterogeneous integration)을 용이하게 하고 칩 간 데이터 대역폭을 크게 향상 시킬 수 있다. 이 기술은 고성능 GPU나 FPGA 등에 적용되고 있다. 시스템 인 패키지(System in Package, SiP) 기술은 여러 기능의 칩들을 하나의 패키지에 통합하여 시스템을 소형화하고 성능을 최적화할 수 있어 IoT 디바이스나 웨어러블 기기 등에 널리 사용되고 있다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[팬아웃 웨이퍼 레벨 패키징 (Fan-Out Wafer Level Packaging, FOWLP)]

- 특징: 칩 크기보다 큰 영역에 재배선층 형성
- 장점: 고밀도 I/O, 소형화, 열 특성 개선
- 응용: 모바일 AP, RF 모듈 등

[2.5D 패키징]

- 특징: 실리콘 인터포저를 이용해 여러 칩을 연결
- 장점: 이종 집적 용이, 대역폭 향상
- 응용: 고성능 GPU, FPGA 등

[시스템 인 패키지 (System in Package, SiP)]

- 특징: 여러 기능의 칩들을 하나의 패키지에 통합
- 장점: 시스템 소형화, 성능 최적화
- 응용: IoT 디바이스, 웨어러블 기기 등

2.3. 첨단 패키징을 위한 소재 기술

앞서 언급한 첨단 패키징 기술의 발전에는 새로운 소재 기술의 개발이 필수적이다. 저유전율 물질은 신호 간섭을 감소시키고 전력 소비를 줄여 재배선층(RDL)의 절연체로 사용된다. 고열전도성 소재인 그래핀이나 붕화알루미늄 등은 효율적인 열 방출을 가능하게 하여 방열 기판이나 열전달 인터페이스에 활용된다. 저온 접합 소재는 낮은 온도에서 접합이 가능해 열응력을 감소시키고 미세 피치 접합을 가능하게 하여 3D 적층이나 다이 접합 등에 사용된다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[저유전율 물질]

- 특징: 신호 간섭 감소, 전력 소비 감소
- 응용: 재배선층 (RDL) 절연체

[고열전도성 소재]

- 특징: 효율적인 열 방출
- 종류: 그래핀, 붕화알루미늄 등
- 응용: 방열 기판, 열전달 인터페이스

[저온 접합 소재]

- 특징: 낮은 온도에서 접합 가능
- 장점: 열응력 감소, 미세 피치 접합 가능
- 응용: 3D 적층, 다이 접합 등

2.4. 첨단 패키징의 미래 전망

첨단 패키징 기술은 앞으로도 계속 발전할 것으로 예상된다. 칩렛(Chiplet) 기술은 대형 칩을 여러 개의 작은 칩으로 분할하여 수율을 높이고 비용을 절감할 수 있으며, 다양한 용도에 맞춰 유연한 제품 구성을 가능하게 한다. 단, 각 칩렛들이 하나의 시스템처럼 동작하려면 서로 원활하게 통신할 수 있는 표준화된 인터페이스의 개발이 필요하다. 광 인터커넥트 기술은 전기 신호 대신 광 신호를 사용하여 초고속 데이터 전송과 저전력 소비를 실현할 수 있지만, 집적화와 비용 절감이 과제로 남아있다.

더 나아가 자가 조립(Self-Assembly) 기술은 나노 스케일 구조체의 자발적 조립을 통해 초미세 구조를 형성하고 대량 생산을 가능하게 할 것으로 기대된다. 이는 3D 나노 구조 패키지 등에 응용될 수 있다. Bio-inspired 패키징은 생물학적 구조를 모방한 패키징 설계로, 효율적인 열 관리와 응력 분산을 가능하게 한다. 생체모방 방열 구조나 자가 치유 소재 등이 이 분야의 주요 연구 대상이다. 각 기술별로 주요 사항을 간단히 정리하면 아래와 같다.

[칩렛 (Chiplet) 기술의 확산]

- 특징: 대형 칩을 여러 개의 작은 칩으로 분할
- 장점: 수율 향상, 비용 절감, 유연한 제품 구성
- 과제: 표준화된 인터페이스 개발 필요

[광 인터커넥트 기술]

- 특징: 전기 신호 대신 광 신호를 사용
- 장점: 초고속 데이터 전송, 저전력 소비
- 과제: 집적화, 비용 절감

[자가 조립 (Self-Assembly) 기술]

- 특징: 나노 스케일 구조체의 자발적 조립
- 장점: 초미세 구조 형성, 대량 생산 가능성
- 응용: 3D 나노 구조 패키지

[바이오 인스파이어드 패키징]

- 특징: 생물학적 구조를 모방한 패키징 설계
- 장점: 효율적인 열 관리, 응력 분산
- 연구 분야: 생체모방 방열 구조, 자가 치유 소재

결론적으로, 반도체 제조 공정의 혁신은 미세공정 기술의 지속적인 발전과 3D 적층 및 첨단 패키징 기술의 혁신을 통해 이루어지고 있다. 이러한 기술들은 더 작고, 더 빠르며, 더 효율적인 반도체 제품을 가능하게 하며, 다양한 응용 분야에서 혁신을 이끌고 있다. 앞으로도 새로운 소재, 공정 기술, 장비 개발을 통해 반도체 제조 기술은 계속해서 발전할 것으로 예상된다.



차세대 반도체 아키텍처

반도체 기술의 지속적인 발전을 위해서는, 반도체 칩을 구성하는 반도체 소자 아키텍처의 혁신 기술 개발이 필수적이다. 본 장에서는 차세대 반도체 아키텍처 기술의 두 가지 주요 혁신 요소 기술인 GAA(Gate-All-Around) 트랜지스터와 3차원 집적 소자 기술에 대해 자세히 살펴보자.

1. GAA (Gate-All-Around) 트랜지스터

GAA 트랜지스터는 현재 널리 사용되고 있는 FinFET의 후속 기술로, 3nm 급 공정 노드에서부터 도입되고 있으며, 기존 소자 대비 우수한 성능 및 향상된 전력 효율성을 제공할 수 있는 혁신적인 반도체 소자 구조이다. 이 기술은 반도체 산업의 미래를 좌우할 핵심 소자 기술로 주목받고 있다.

1.1. GAA 트랜지스터의 기본 구조

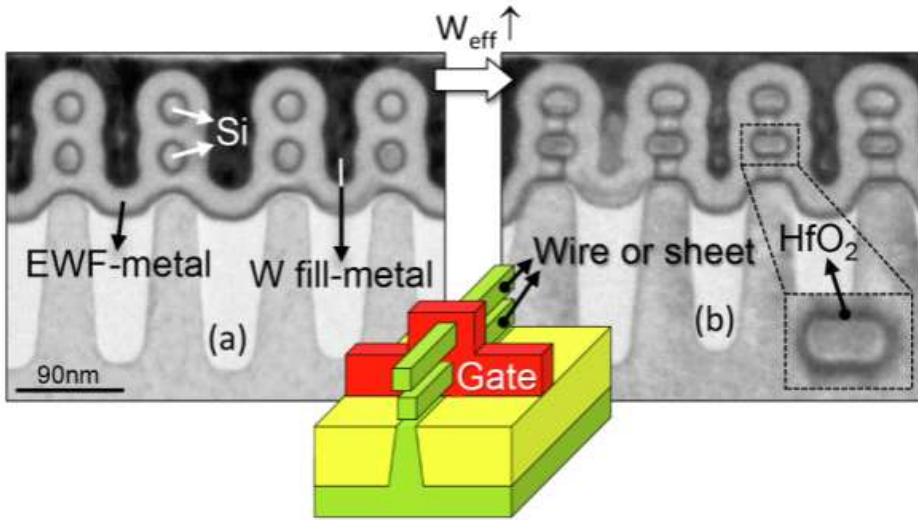
GAA 트랜지스터의 기본 개념은 채널을 게이트 물질이 완전히 감싸는 구조를 가지고 있다는 것이다. 이는 채널의 형태에 따라 나노와이어(nano-wire)와 나노시트(nano-sheet)로 구분된다. 현재 널리 사용되고 있는 FinFET과 비교했을 때, FinFET은 채널의 3면이 게이트로 둘러싸인 구조인 반면, GAA는 채널의 4면 모두가 게이트로 둘러싸인 구조를 가진다. 이러한 구조적 특징으로 인해 GAA 트랜지스터는 더 우수한 전류 제어 능력을 가지며, 단채널 효과를 크게 감소시킬 수 있다.

GAA 트랜지스터의 주요 구성 요소로는 전류가 흐르는 경로인 채널, 채널을 둘러싸며 전류를 제어하는 게이트, 그리고 전류의 유입과 유출 지점인 소스와 드레인이 있다. 채널은 주로 실리콘이나 게르마늄 기반의 재료로 만들어지며, 이러한 구조는 더 작은 크기의 트랜지스터를 만들 수 있게 해주어 집적도를 높이는 데 큰 도움이 된다.

1.2. GAA 트랜지스터의 종류

GAA 트랜지스터는 크게 나노와이어 GAA와 나노시트 GAA로 구분할 수 있다. 나노와이어 GAA는 원통형 채널 구조를 가지고 있어 우수한 전류 제어와 낮은 누설 전류 특성을 보인다. 하지만 복잡한 제조 공정과 전류 밀도의 한계라는 과제를 안고 있다.

반면, 나노시트 GAA는 넓고 얇은 채널 구조를 가지고 있어 높은 전류 밀도를 제공하며, 나노와이어 대비 제조가 상대적으로 용이하다. 이러한 특성으로 인해 3nm 이하의 공정에서 주로 채택어서 활용 중이다.



[그림 5] 나노와이어 GAA 트랜지스터(좌)와 나노시트 GAA 트랜지스터(우) [5]

또한, 수직 나노와이어/나노시트 구조도 개발되고 있는데, 이는 수직으로 세워진 채널 구조를 가지고 있어 높은 집적도를 달성할 수 있으며, 3D 집적 기술과의 시너지 효과를 기대할 수 있다. 그러나 이 구조 역시 복잡한 제조 공정과 열 관리라는 과제를 해결해야 한다.

1.3. GAA 트랜지스터의 제조 기술

GAA 트랜지스터의 제조는 여러 단계의 복잡한 공정을 거칩니다. 먼저, 채널 형성 기술에는 선택적 에피택시 성장(Selective Epitaxial Growth, SEG)과 다층 박막 증착 및 패터닝 기술이 사용된다. 이 과정에서 나노미터 수준의 정밀한 제어가 필요하며, 균일한 채널 형성이 GAA 트랜지스터의 성능을 좌우하는 핵심 요소가 된다.

게이트 스택 형성 과정에서는 고유전율 게이트 절연막(High-k Dielectric)을 증착하고, 금속 게이트(Metal Gate)를 형성한다. 이때, 게이트 금속의 일함수 조절을 위한 다층 금속 구조를 사용하거나 dipole 소재를 활용하여 트랜지스터의 전기적 특성을 최적화한다. 이 과정은 나노미터 수준의 두께 제어와 균일성 확보가 중요하다.

Channel Release 공정은 GAA 트랜지스터 제조의 핵심 단계 중 하나이다. 이 과정에서는 희생층(Sacrificial Layer)을 제거하는 기술과 등방성 식각 기술의 정밀 제어가 필요하다. 이 단계에서 채널 주변의 게이트 물질이 채널을 완전히 감싸도록 하는 것이 중요하며, 이는 GAA 트랜지스터의 성능을 결정짓는 중요한 요소이다.

마지막으로, 소스/드레인 형성 과정에서는 선택적 에피택시 성장 기술과 저저항 콘택 형성 기술이 사용된다. 이 과정에서 소스/드레인 영역의 도핑 프로파일과 금속/반도체 접합 특성을 최적화하여 트랜지스터의 전체적인 성능을 향상시킨다.

1.4. GAA 트랜지스터의 성능 향상 기술

GAA 트랜지스터의 성능을 더욱 향상시키기 위해 다양한 기술이 연구되고 있다. 채널 엔지니어링 기술은 채널의 두께와 폭을 최적화하고, stress engineering을 통해 캐리어의 이동도를 향상시킨다. 이를 통해 트랜지스터의 전류 구동 능력을 높이고 스위칭 속도를 개선할 수 있다.

게이트 스택 최적화 기술은 게이트 물질의 일함수 조절을 통한 문턱 전압 제어와 계면 특성 개선을 통한 신뢰성 향상에 초점을 맞춘다. 이는 트랜지스터의 전기적 특성을 미세하게 조절하고 장기적인 안정성을 확보하는 데 중요한 역할을 한다.

소스/드레인 최적화 기술은 기생 저항 감소를 위한 도핑 프로파일 제어와 금속/반도체 접합 특성 개선에 중점을 둔다. 이를 통해 트랜지스터의 전체적인 성능을 향상시키고 전력 소비를 줄일 수 있다.

특히 주목할 만한 것은 다중 채널 구조 기술이다. 이 기술은 단일 트랜지스터 내에 여러 개의 채널을 형성하여 전류 구동력을 향상시키고 면적 효율성을 증대시킨다. 이는 고성능 로직 회로에서 특히 유용하게 활용될 수 있다.

1.5. GAA 트랜지스터의 응용 및 전망

GAA 트랜지스터는 다양한 분야에서 응용될 것으로 예상된다. 로직 소자 분야에서는 고성능 CPU, GPU 등에 적용되어 연산 속도를 높이고 전력 효율을 개선할 수 있다. 또한 저전력 모바일 AP(Application Processor)에도 활용되어 모바일 기기의 배터리 수명을 연장시킬 수 있다.

메모리 소자 분야에서는 3D DRAM의 셀 트랜지스터로 활용될 가능성이 있으며, 3D NAND의 수직 채널 구조와의 융합을 통해 더 높은 집적도와 성능을 달성할 수 있다. 이는 대용량 저장 장치의 발전에 큰 기여를 할 것으로 예상된다.

아날로그/RF 소자 분야에서는 우수한 전류 제어 능력을 활용한 고주파 특성 개선이 가능하며, 저 잡음 증폭기(LNA) 등에 적용될 수 있다. 이는 5G 및 향후 6G 통신 시스템의 성능 향상에 기여할 것이다.

전력 반도체 분야에서도 GAA 트랜지스터의 적용이 검토되고 있다. 고전압, 대전류 응용 분야로의 확장 가능성이 있으며, 전력 변환 효율 향상에 기여할 수 있다. 이는 전기 자동차, 신재생 에너지 시스템 등의 발전에 중요한 역할을 할 것으로 기대된다.

향후 GAA 트랜지스터는 2nm 이하 공정에서 주류 기술로 자리매김할 것으로 예상된다. 또한 새로운 채널 소재(예: 2D 소재)와의 결합을 통한 성능 향상이 기대되며, 양자 효과를 활용한 새로운 동작 원리의 소자로 발전할 가능성도 있다. 이는 포스트 무어 시대의 반도체 기술 발전을 이끌어갈 핵심 기술이 될 것이다.

2. 3차원 집적 기술

3차원 집적 기술은 반도체 산업의 혁신을 이끄는 핵심 기술로, 2차원 평면상의 집적 한계를 뛰어넘어 시스템의 성능과 효율을 극대화하기 위한 혁신적인 접근 방식이다. 이 기술은 반도체 소자를 수직으로 쌓아 올림으로써 기존의 2차원 집적 방식에서는 달성하기 어려웠던 고집적도와 고성능을 실현할 수 있게 한다.

2.1. 3차원 집적 기술의 개요

3차원 집적 기술의 기본 개념은 여러 층의 반도체 소자를 수직으로 쌓아 올리는 것이다. 이는 단순히 소자를 쌓는 것에 그치지 않고, 칩 간, 웨이퍼 간 연결을 통해 시스템을 통합하는 것까지 포함한다. 이러한 접근 방식은 기존의 2차원 평면 구조에서는 불가능했던 새로운 차원의 집적을 가능하게 한다.

3차원 집적 기술의 주요 이점은 크게 네 가지로 요약될 수 있다. 첫째, 반도체 소자의 집적도

향상이다. 3차원 구조를 통해 단위 면적당 더 많은 소자를 집적할 수 있게 되어, 동일한 평면 크기에서 더 높은 성능과 기능성을 구현할 수 있다. 이는 특히 모바일 기기나 웨어러블 디바이스와 같은 소형 전자기기에서 중요한 이점이 된다. 물론 NMOS와 PMOS를 위 아래에 적층하니, 절반의 면적에 1개의 트랜지스터를 집적한다고 생각할 수 있으나, 사실은 NMOS와 PMOS 간의 금속 배선으로 연결해야 하는 공간이 필요하기 때문에, 2배 이상의 면적 효율이라는 표현은 틀렸고, 1.5배~2배 수준의 집적도 향상을 기대해볼 수 있다.

둘째, 성능 향상이다. 3차원 구조는 신호 경로를 단축시켜 신호 지연을 감소시키고, 동시에 대역폭을 증가시킬 수 있다. 이는 고성능 컴퓨팅 시스템이나 데이터 센터에서 특히 중요한 요소로 작용한다. 물론 parasitic capacitance를 정확히 모델링 하고, 그로 인해 부작용을 최소화하는 기술이 필요할 것이다.

셋째, 전력 효율성 증대이다. 인터커넥트 길이가 감소함에 따라 신호 전송에 필요한 전력 소비가 줄어들게 된다. 이는 배터리 수명이 중요한 모바일 기기나 에너지 효율이 중요한 데이터 센터 등에서 큰 이점이 된다. 또한, backside power distribution network를 활용하여, 전력 배선과 신호 배선을 분리하여, 신호 무결성(signal integrity) 향상도 있을 것이다.

마지막으로, 이중 집적 기술의 가능성이다. 3차원 집적 기술을 통해 서로 다른 공정이나 기술로 만들어진 소자들을 하나의 패키지에 통합할 수 있게 된다. 이는 시스템의 다기능화와 소형화를 동시에 달성할 수 있게 해주며, 특히 시스템 온 칩(SoC) 설계에서 큰 유연성을 제공한다.

그러나 3차원 집적 기술은 여러 가지 기술적 과제도 안고 있다. 가장 큰 문제 중 하나는 열 관리이다. 수직으로 적층된 구조에서는 열이 집중되기 쉬워 효과적인 냉각 시스템의 설계가 필수적이다. 또한, 3D 구조에서의 효과적인 테스트 방법론 개발도 중요한 과제이다. 기존의 2D 구조에서 사용되던 테스트 방법을 그대로 적용하기 어려우며, 각 층별로 또는 전체 시스템에 대한 새로운 테스트 방법이 필요하다. Metrology and Inspection 기술을 3차원 집적 기술 고도화와 함께 개발해야 할 것이다.

마지막으로, 복잡한 제조 공정으로 인한 비용 증가도 중요한 과제이다. 3D 집적 기술은 여러 단계의 정밀한 공정을 필요로 하며, 이는 제조 비용의 상승으로 이어질 수 있다. 따라서 비용 효율적인 제조 기술의 개발이 3D 집적 기술의 광범위한 적용을 위해 매우 중요하다.

2.2. 3차원 집적 기술의 주요 방식

3차원 집적 기술은 다양한 방식으로 구현될 수 있으며, 각 방식은 고유한 특징과 장단점을 가지고 있다. 주요 방식으로는 TSV(Through-Silicon Via) 기술, 모놀리식 3D 집적, 웨이퍼 본딩 기술, 그리고 칩렛(Chiplet) 기술 등이 있다.

TSV(Through-Silicon Via) 기술은 실리콘 웨이퍼를 관통하는 전극을 형성하는 기술로, 3차원 집적 기술의 핵심 요소 중 하나이다. 이 기술은 높은 집적도와 우수한 전기적 특성을 제공하며, 특히 고대역폭 메모리(High Bandwidth Memory, HBM)와 이미지 센서 분야에서 널리 사용되고 있다. TSV 기술의 제조 과정은 복잡하고 정밀한 공정을 필요로 한다. Via 홀 형성, 절연막 증착, 배리어/시드 레이어 형성, Via 충전, 그리고 CMP(Chemical Mechanical Polishing) 등의 단계를 거치게 된다. 이러한 복잡한 공정은 TSV 기술의 성능을 결정짓는 중요한 요소이며, 지속적인 공정 개선을 통해 더 높은 성능과 신뢰성을 확보하고 있다. SK하이닉스 HBM3/3E 제품을 보면, 신호전달을 위한 TSV가 1024, 2048개이나, Power를 전달하는 TSV까지 포함하면 무려 5,000 여개가 들어가 있다.

모놀리식 3D 집적 기술은 단일 웨이퍼 상에서 여러 층의 트랜지스터를 형성하는 방식이다. 이 기술은 초고밀도 집적이 가능하고 층간 정렬이 매우 정밀하다는 장점이 있다. 그러나 상부 층을 형성할 때 하부 층이 손상되지 않도록 하는 것은 매우 큰 도전적 과제이다. 이를 해결하기 위해 저온 공정 기술, 레이저 어닐링, 이온 주입을 통한 도핑 활성화 등의 기술이 개발되고 있다. 특히 저온 공정 기술은 하부 층의 열 손상을 방지하면서도 상부 층의 성능을 보장할 수 있는 핵심 기술로 주목받고 있다.

웨이퍼 본딩 기술은 개별적으로 제작된 웨이퍼들을 상호 간에 접합하는 방식이다. 이 기술은 직접 본딩, 중간층 본딩, 하이브리드 본딩 등 다양한 방식으로 구현될 수 있다. 웨이퍼 본딩 기술의 큰 장점은 이종 소자 통합에 유리하고 상대적으로 높은 수율을 얻을 수 있다는 점이다. 서로 다른 소재를 사용해서 만들어야 하는 소자들을 실리콘 웨이퍼 위에 집적하는 일은 매우 힘든 과제였으나, 웨이퍼 본딩 기술을 이용하면, 상대적으로 매우 쉬운 과제로 변하게 된다. 그러나 정밀한 정렬과 본딩 계면의 품질 관리가 중요한 과제로 남아있다. 최근에는 나노 수준의 정밀도를 가진 정렬 기술과 고품질 본딩 계면을 형성할 수 있는 새로운 재료 및 공정 기술이 개발되고 있어, 웨이퍼 본딩 기술의 적용 범위가 더욱 확대될 것으로 예상된다.

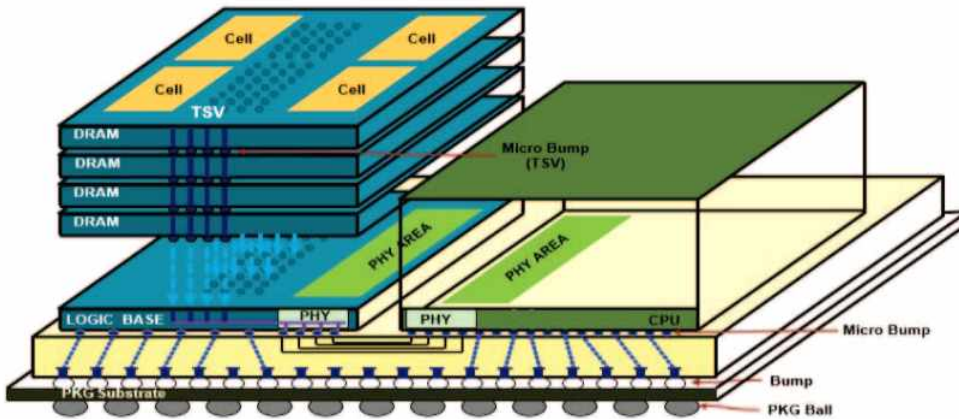
칩렛(Chiplet) 기술은 여러 개의 작은 칩들을 하나의 패키지에 통합하는 방식이다. 이 기술은 유연한 제품 구성이 가능하고, 수율 향상과 비용 절감 효과를 얻을 수 있다는 장점이 있다. 특히 대형 칩을 제작할 때 발생하는 수율 문제를 해결할 수 있어, 고성능 프로세서나 AI 가속기 등의 분야에서 주목받고 있다. 그러나 칩렛 간의 효율적인 연결을 위한 표준화된 인터페이스 개발과 칩렛 간 연결 최적화가 중요한 과제로 남아있다. 이를 해결하기 위해 업계에서는, 인텔이 주도하고 있는 UCIe(Universal Chiplet Interconnect Express)와 같은 표준을 개발하고 있으며, 이를 통해 칩렛 기술의 보편화와 확산이 가속화될 것으로 예상된다.

이러한 다양한 3차원 집적 기술들은 각각의 장단점을 가지고 있으며, 응용 분야와 요구사항에 따라 적절한 기술이 선택되어 사용된다. 앞으로는 이들 기술의 장점을 결합하고 단점을 보완하는 하이브리드 접근 방식이 더욱 발전할 것으로 예상되며, 이를 통해 3차원 집적 기술의 적용 범위가 더욱 확대될 것으로 기대된다.

2.3. 3차원 집적 기술의 응용 분야

3차원 집적 기술은 다양한 분야에서 혁신적인 변화를 이끌고 있다. 이 기술의 주요 응용 분야로는 고대역폭 메모리, 이미지 센서, RF/아날로그 소자, 뉴로모픽 칩, 그리고 파워 일렉트로닉스 등이 있다. 각 분야에서 3차원 집적 기술은 고유한 장점을 제공하며, 기존 기술의 한계를 뛰어넘는 성능 향상을 가능하게 한다.

고대역폭 메모리(HBM, High Bandwidth Memory)는 3차원 집적 기술의 대표적인 응용 사례이다. HBM의 구조는 DRAM 다이를 수직으로 적층하고, 이를 로직 칩과 인터포저를 통해 연결하는 방식을 취한다. 이러한 구조는 초고속, 저전력 메모리 솔루션을 제공한다. HBM은 특히 고성능 그래픽 카드와 AI 가속기에서 널리 사용되고 있다. 이 기술은 기존의 2D 구조 메모리에 비해 훨씬 높은 대역폭과 낮은 전력 소비를 실현할 수 있어, 데이터 집약적인 애플리케이션의 성능을 크게 향상시킬 수 있다. 또한, 좁은 면적에 고용량 메모리를 구현할 수 있어 시스템의 소형화에도 기여한다.



[그림 6] 고대역폭 메모리 (HBM) [6]

이미지 센서 분야에서도 3차원 집적 기술이 중요한 역할을 하고 있다. 3D 구조의 이미지 센서는 픽셀 어레이와 신호 처리 회로를 수직으로 분리하여 구성한다. 이러한 구조의 주요 장점으로서는 감도 향상, 노이즈 감소, 그리고 소형화를 들 수 있다. 픽셀 어레이와 신호 처리 회로를 분리함으로써 각 부분을 최적화할 수 있으며, 이는 결과적으로 이미지 품질의 향상으로 이어진다. 이 기술은 스마트폰 카메라, 보안 카메라, 의료 영상 기기 등에서 널리 사용되고 있으며, 더 높은 화질과 더 나은 저조도 성능을 제공한다. 또한, 3D 구조를 통해 픽셀 크기를 줄이면서도 충분한 광 흡수 영역을 확보할 수 있어, 고해상도 이미지 센서의 개발에도 크게 기여하고 있다.

RF/아날로그 소자 분야에서 3차원 집적 기술은 디지털 로직과 RF/아날로그 회로의 수직 통합을 가능하게 한다. 이러한 통합의 주요 장점으로서는 신호 무결성 향상, 노이즈 감소, 그리고 소형화를 들 수 있다. 디지털 회로와 아날로그 회로를 분리하여 각각 최적화할 수 있으며, 이는 전체 시스템의 성능 향상으로 이어진다. 이 기술은 5G 통신 모듈, IoT 센서 노드 등에서 널리 응용되고 있다. 특히 5G 통신에서 요구되는 고주파 신호 처리와 저전력 동작을 동시에 만족시킬 수 있어, 차세대 통신 시스템 구현에 핵심적인 역할을 하고 있다.

뉴로모픽 칩은 인간의 뇌 구조를 모방한 3D 구조를 가진다. 이 구조는 뉴런과 시냅스의 고밀도 연결을 3차원적으로 구현함으로써, 기존의 2D 구조에서는 달성하기 어려웠던 고밀도 연결, 낮은 지연 시간, 저전력 소비 등의 장점을 제공한다. 이러한 특성은 엣지 AI, 패턴 인식, 자율 주행 시스템 등의 분야에서 매우 유용하게 활용된다. 특히 엣지 컴퓨팅 환경에서 실시간 데이터 처리와 저전력 동작이 요구되는 상황에서 뉴로모픽 칩은 큰 강점을 발휘할 수 있다.

마지막으로, 파워 일렉트로닉스 분야에서도 3차원 집적 기술이 적용되고 있다. 이 기술은 전력 소자와 제어 로직을 수직으로 통합하는 것이 특징이다. 이를 통해 전력 밀도 향상, 효율성 증가, 그리고 소형화라는 장점을 얻을 수 있다. 전력 소자와 제어 로직을 긴밀하게 통합함으로써 전력 변환 효율을 높이고, 시스템의 크기를 줄일 수 있다. 이 기술은 전기 자동차 인버터, 산업용 전력 변환 장치 등에서 널리 응용되고 있다. 특히 전기 자동차 분야에서는 고효율, 고전력 밀도의 인버터 개발이 차량의 주행 거리와 성능 향상에 직접적인 영향을 미치기 때문에, 3차원 집적 기술의 적용이 활발히 이루어지고 있다.

다양한 응용 분야에서 3차원 집적 기술은 기존 기술의 한계를 극복하고 새로운 가능성을 열어가고 있다. 앞으로도 이 기술은 계속해서 발전하여 더 많은 분야에서 혁신을 이끌어 낼 것으로 기대된다. 특히 인공지능, 자율주행, 사물인터넷 등 미래 핵심 기술 분야에서 3차원 집적 기술의 중요성은 더욱 커질 것으로 예상된다.

2.4. 3차원 집적 기술의 미래 전망

3차원 집적 기술은 반도체 산업의 미래를 선도할 핵심 기술로 주목받고 있다. 이 기술의 발전 방향은 다양한 측면에서 예측되고 있으며, 각 영역에서 혁신적인 발전이 기대된다.

초고밀도 집적 분야에서는 수백 층 이상의 3D NAND 플래시 메모리 개발이 향후 5년간 이어질 것으로 전망된다. 이는 현재의 저장 용량을 획기적으로 증가시킬 수 있는 기술로, 빅데이터 시대의 데이터 저장 요구를 충족시킬 수 있을 것으로 기대된다. 또한, 로직-온-로직(Logic-on-Logic) 구조의 고성능 프로세서 개발도 예상된다. 이 구조는 여러 층의 로직 회로를 수직으로 쌓아 올림으로써 처리 속도와 전력 효율을 크게 향상시킬 수 있다.

이종 집적 기술의 극대화도 중요한 발전 방향이다. 서로 다른 공정 노드의 로직 회로를 통합하는 기술이 발전하여, 최적의 성능과 비용 효율성을 달성할 수 있을 것으로 예상된다. 또한, 광전자 소자, MEMS, 센서 등 다양한 기능의 수직 통합이 가능해져, 더욱 복잡하고 고성능인 시스템을 단일 패키지로 구현할 수 있게 될 것이다.

인공지능 가속 분야에서는 메모리와 연산 유닛의 긴밀한 3D 통합이 이루어질 것으로 예상된다. 이는 데이터 이동을 최소화하고 처리 속도를 극대화하여 AI 연산의 효율을 크게 높일 수 있다. 또한, 뇌의 구조를 모방한 3D 뉴로모픽 아키텍처 개발도 전망되어, 더욱 효율적이고 인간의 뇌와 유사한 방식으로 동작하는 AI 시스템을 구현할 수 있을 것이다.

양자 컴퓨팅과의 융합도 주목할 만한 발전 방향이다. 양자 비트와 제어 로직의 3D 집적이 이루어질 것으로 예상되며, 이는 양자 컴퓨터의 성능과 안정성을 크게 향상시킬 수 있다. 또한, 극저온 냉각 시스템과의 통합도 이루어져, 양자 컴퓨터의 실용화를 앞당길 수 있을 것이다.

첨단 패키징 기술과의 시너지도 기대된다. 팬아웃 웨이퍼 레벨 패키징(FOWLP)과 3D 집적의 결합을 통해 더욱 복잡하고 고성능인 시스템을 구현할 수 있을 것이다. 또한, 액티브 인터포저를 활용한 시스템 통합 기술이 발전하여, 다양한 기능을 가진 칩들을 효율적으로 연결하고 통합할 수 있게 될 것이다.

새로운 소재 및 구조의 도입도 3차원 집적 기술의 미래를 좌우할 중요한 요소이다. TMDC, MoS₂ 등의 2D 소재를 활용한 초박막 층간 연결 기술이 개발되어, 더욱 얇고 효율적인 3D 구조를 만들 수 있을 것이다. 표면은 도체이나 내부는 부도체인 위상학적 절연체(topological insulator)를 이용한 저전력, 고효율 인터커넥트 기술도 주목받고 있다. 또한, 자기조립(self-assembly) 기술을 활용한 나노스케일 3D 구조 형성 기술이 발전하여, 더욱 정밀하고 복잡한 3D 구조를 효율적으로 만들 수 있게 될 것이다.

열 관리 기술의 혁신도 3차원 집적 기술의 발전에 중요한 역할을 할 것이다. 다이아몬드 기반 열 전도층 삽입 기술이 개발되어, 3D 구조에서 발생하는 열을 효과적으로 제거할 수 있을 것이다. 마이크로채널 냉각 시스템의 3D 구조 내 통합도 이루어져, 더욱 효율적인 냉각이 가능해질 것이다. 또한, 상변화 물질을 이용한 능동적 열 제어 시스템이 개발되어, 동적으로 변화하는 열 발생에 대응할 수 있게 될 것이다.

초고속 인터커넥트 기술도 3차원 집적 기술의 미래를 이끌 중요한 요소이다. 광 인터커넥트의 3D 구조 내 통합이 이루어져, 초고속 데이터 전송이 가능해질 것이다. 테라헤르츠 통신을 위한 3D 안테나 구조가 개발되어, 더욱 빠르고 효율적인 무선 통신이 가능해질 것이다. 또한, 무선 층간 통신 기술이 도입되어, 복잡한 배선 없이도 3D 구조 내에서 효율적인 통신이 가능해질 것이다.

유연성과 신축성을 갖춘 3D 집적 기술의 발전도 예상된다. 웨어러블 디바이스를 위한 유연한 3D 구조가 개발되어, 더욱 편리하고 기능적인 웨어러블 기기를 만들 수 있게 될 것이다. 뇌-기계 인터페이스를 위한 생체 적합성 3D 전자 소자도 개발되어, 인간의 신경계와 직접 연결될 수 있는 전자 기기가 가능해질 것이다. 또한, 자가 치유 능력을 갖춘 3D 집적 회로가 개발되어, 내구성과

신뢰성이 크게 향상된 전자 기기를 만들 수 있게 될 것이다.

극한 환경을 위한 3D 집적 기술도 발전할 것으로 예상된다. 우주 환경에 적합한 방사선 내성 3D 구조가 개발되어, 우주 탐사와 위성 기술에 혁신을 가져올 수 있을 것이다. 초고온, 초저온 환경에서 동작 가능한 3D 시스템이 개발되어, 극한 환경에서의 전자 기기 사용 범위가 크게 확대될 것이다. 또한, 극한 압력 하에서 작동하는 3D 센서 시스템이 개발되어, 심해 탐사나 지하 자원 탐사 등에 활용될 수 있을 것이다.

Bio-inspired 3D 집적 기술도 주목받고 있다. 뇌의 구조를 모방한 3D 뉴럴 네트워크가 개발되어, 더욱 효율적이고 인간의 뇌와 유사한 방식으로 동작하는 AI 시스템을 구현할 수 있을 것이다. 생체 시스템의 자기 조직화 능력을 모방한 적응형 3D 구조도 개발되어, 환경 변화에 유연하게 대응할 수 있는 전자 시스템을 만들 수 있게 될 것이다. 또한, DNA 오리가미 기술을 활용한 나노스케일 3D 패터닝 기술이 발전하여, 분자 수준의 정밀한 3D 구조를 만들 수 있게 될 것이다.

재구성 가능한 3D 아키텍처의 발전도 예상된다. 동적으로 기능을 변경할 수 있는 3D FPGA 구조가 개발되어, 더욱 유연하고 효율적인 컴퓨팅 시스템을 구현할 수 있을 것이다. 인공지능 기반의 자가 최적화 3D 시스템도 개발되어, 작업 부하에 따라 자동으로 최적의 구조로 변경되는 시스템이 가능해질 것이다. 또한, 포스트 무어 시대를 위한 적응형 컴퓨팅 아키텍처가 개발되어, 기존의 반도체 기술의 한계를 뛰어넘는 새로운 컴퓨팅 패러다임을 제시할 수 있을 것이다.

마지막으로, 지속가능성을 고려한 3D 집적 기술의 발전도 중요한 미래 전망이다. 재활용 및 분해가 용이한 모듈형 3D 구조가 개발되어, 전자 폐기물 문제를 해결하는 데 기여할 수 있을 것이다. 생분해성 재료를 활용한 일회용 3D 의료기기가 개발되어, 의료 분야에서의 환경 부담을 줄일 수 있을 것이다. 또한, 에너지 하베스팅 기능이 통합된 자급자족형 3D 시스템이 개발되어, 외부 전원 공급 없이도 지속적으로 동작할 수 있는 전자 기기가 가능해질 것이다.

이러한 미래 기술들은 3차원 집적 기술이 단순히 집적도 향상을 넘어, 새로운 기능과 응용 분야를 창출할 수 있는 잠재력을 보여준다. 이 기술은 전자 산업뿐만 아니라 의료, 우주 탐사, 환경 모니터링 등 다양한 분야에 혁명적인 변화를 가져올 것으로 예상된다. 그러나 이러한 혁신적인 기술들을 실현하기 위해서는 여전히 많은 기술적 과제들을 해결해야 한다. 열 관리, 신뢰성, 제조 비용, 테스트 및 수율 관리 등의 문제들이 지속적으로 연구되고 개선되어야 할 것이다. 또한, 이러한 첨단 기술들의 상용화를 위해서는 산업 표준의 확립과 생태계 조성도 중요한 과제가 될 것이다. 결론적

으로, 3차원 집적 기술은 반도체 산업의 미래를 형성하는 핵심 기술 중 하나로 자리잡을 것이며, 이를 통해 우리가 상상하지 못했던 새로운 응용 분야와 서비스들이 등장할 것으로 기대된다.

IV

모바일 엣지 컴퓨팅을 위한 저전력 반도체 기술

모바일 엣지 컴퓨팅 기술의 급속한 발전과 함께 저전력 반도체 기술의 중요성이 그 어느 때보다 부각되고 있다. 모바일 기기와 IoT 디바이스의 확산으로 인해 엣지에서의 데이터 처리 수요가 급증하고 있으며, 이에 따라 효율적이고 지속 가능한 컴퓨팅 솔루션의 필요성이 대두되고 있다. 본 장에서는 모바일 엣지 컴퓨팅 환경에 최적화된 초저전력 반도체 기술의 현재와 미래에 대해 심도 있게 살펴보고자 한다. 특히 전력 소비를 최소화하면서도 높은 성능을 유지할 수 있는 혁신적인 반도체 소자 기술에 주목할 것이다.

초저전력 반도체 기술은 모바일 엣지 컴퓨팅의 핵심 요소로, 제한된 배터리 용량 내에서 최대한의 성능을 이끌어내는 것을 목표로 한다. 이를 위해 다양한 설계 기술과 혁신적인 접근 방식이 개발되고 있다.

1. 저전력 설계/소자 기술

[동적 전력 관리 (Dynamic Power Management, DPM)]

동적 전력 관리는 실시간으로 시스템의 전력 소비를 최적화하는 기술이다. 클록 게이팅(Clock Gating) 기술은 현재 사용되지 않는 회로 부분의 클록 신호를 선택적으로 차단한다. 클록 신호가 차단된 회로는 동작을 멈추고 전력 소비를 크게 줄일 수 있다. 예를 들어, 모바일 프로세서에서 그래픽 처리 유닛이 사용되지 않을 때 해당 부분의 클록을 끄는 방식으로 적용된다. 전원 게이팅(Power Gating) 기술은 클록 게이팅보다 한 단계 더 나아가, 완전히 사용하지 않는 블록의 전원을 물리적으로 차단하는 기술이다. 이는 대기 전력 소비를 획기적으로 줄일 수 있지만, 전원을 다시 켜는 데 시간이 걸리는 단점이 있다.

DVFS(Dynamic Voltage and Frequency Scaling)기술은 시스템의 작업 부하에 따라 프로세서의 전압과 주파수를 동적으로 조절하는 기술이다. 낮은 부하 상태에서는 전압과 주파수를 낮춰 전력 소비를 줄이고, 높은 성능이 필요할 때는 이를 높여 처리 속도를 개선한다. 최신 모바일 프로세서들은 이 기술을 적극 활용하여 전력 효율성을 극대화하고 있다.

[저전압 동작을 위한 소자 기술]

저전압 동작 기술은 반도체 소자의 동작 전압을 최소화하여 전력 소비를 획기적으로 줄이는 것을 목표로 한다. 니어-쓰레숄드(Near-Threshold) 컴퓨팅은 트랜지스터의 문턱전압 바로 위에서 반도체 소자가 동작하도록 회로를 설계하는 기술이다. 이는 전력 소비와 성능 사이의 최적의 균형점을 찾는 방법으로, 전력 효율성을 크게 높이면서도 acceptable한 성능을 유지할 수 있다. 서브-쓰레숄드(Sub-Threshold) 컴퓨팅은 니어-쓰레숄드 컴퓨팅보다 더 극단적인 접근 방식으로, 트랜지스터를 문턱전압 이하에서 동작시킨다. 이는 초저전력 동작을 가능케 하지만, 성능 저하와 신뢰성 문제가 발생할 수 있어 특수한 용도로 제한적으로 사용된다. 적응형 바디 바이어싱(Adaptive Body Biasing)은, 트랜지스터의 기판 전압을 동적으로 조절하여 누설 전류를 최소화하는 기술이다. 이는 공정 변동에 따른 전력 소비 차이를 보정하고, 칩의 수명 주기 동안 최적의 전력 효율을 유지하는 데 도움을 준다.

[저전력 회로 설계 기술]

회로 수준에서의 혁신적인 설계 기술들은 전체 시스템의 전력 효율성을 크게 향상시킬 수 있다. 비동기 로직(Asynchronous Logic) 기술은 전통적인 동기식 설계와 달리 클럭 신호 없이 동작하는 회로 설계 방식이다. 이는 불필요한 스위칭을 줄여 동적 전력 소비를 감소시키고, 전자기 간섭도 줄일 수 있다. 그러나 설계의 복잡성이 증가하는 단점이 있다. 파워 도메인 분할(Power Domain Partitioning) 기술은 칩을 여러 개의 독립적인 전원 관리가 가능한 영역으로 나누는 기술이다. 각 도메인은 필요에 따라 개별적으로 전원을 켜고 끌 수 있어, 시스템 레벨에서의 전력 관리 유연성을 크게 높일 수 있다.

저전력 메모리 기술로는 전통적인 SRAM이나 DRAM 대신 eDRAM(embedded DRAM)이나 MRAM(Magnetoresistive RAM) 같은 새로운 메모리 기술을 사용하여 대기 전력과 동작 전력을 모두 줄일 수 있다. 특히 MRAM은 비휘발성이면서도 빠른 접근 속도를 제공하여 차세대 저전력 메모리 솔루션으로 주목받고 있다.

이러한 다양한 저전력 설계 기술들은 개별적으로도 효과적이지만, 실제 시스템에서는 이들을 복합적으로 적용하여 시너지 효과를 극대화한다. 모바일 엣지 컴퓨팅 환경에서는 이러한 기술들의 지속적인 발전과 새로운 혁신이 필수적이며, 이를 통해 더욱 효율적이고 지속 가능한 컴퓨팅 솔루션을 실현할 수 있을 것이다.

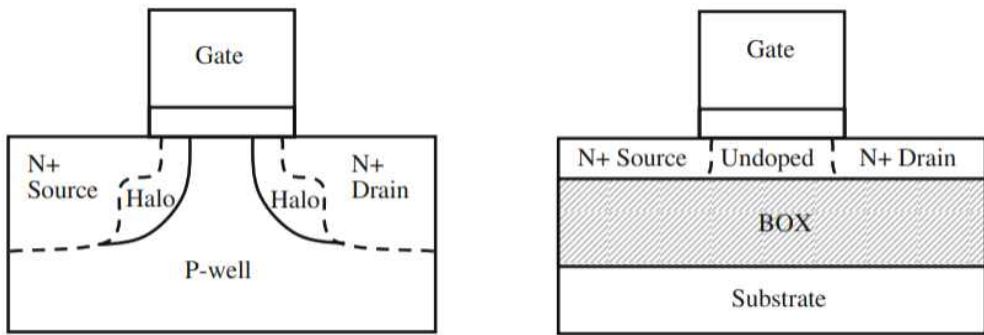
2. 공정 기술 최적화

반도체 제조 공정 기술의 최적화는 저전력 반도체 개발에 있어 핵심적인 역할을 한다. 최신 공정 기술을 활용하여 트랜지스터의 성능을 향상시키고 전력 소비를 줄이는 것이 이 분야의 주요 목표이다.

[첨단 공정 노드 활용]

반도체 산업은 지속적으로 더 작고 효율적인 트랜지스터를 만들기 위해 노력하고 있다. 이러한 노력의 결과로 최신 공정 기술이 계속해서 발전하고 있다. 최신 5nm, 4nm와 3nm 공정 기술은 트랜지스터의 크기를 극단적으로 축소시켜 칩 면적당 집적도를 크게 높였다. 이는 단순히 더 많은 트랜지스터를 집적할 수 있다는 것을 넘어서, 각 트랜지스터의 스위칭에 필요한 에너지를 줄여 전체적인 전력 효율을 대폭 개선했다. 예를 들어, 5nm에서 3nm로 전환 시 약 30-35%의 전력 효율 향상을 기대할 수 있다. 3nm이하부터 사용되고 있는, 기존의 FinFET 구조를 넘어선 GAA 구조는, 트랜지스터의 채널을 게이트가 완전히 감싸는 혁신적인 설계이다. 이 구조는 전류의 흐름을 더욱 정밀하게 제어할 수 있어, 누설 전류를 최소화하고 전력 효율을 극대화한다. 특히 3nm 이하의 공정에서 GAA 구조의 장점이 더욱 부각될 것으로 예상된다.

전통적인 벌크 실리콘 기반 트랜지스터의 한계를 극복하기 위해 새로운 구조와 동작 원리를 가진 트랜지스터 기술이 개발되고 있다. FDSOI(Fully Depleted Silicon On Insulator) 기술은 매우 얇은 실리콘 층을 절연체 위에 형성하여 트랜지스터를 만드는 방식이다. 이 구조는 기생 용량을 줄이고 바디 효과를 최소화하여 저전압에서도 효율적으로 동작할 수 있게 한다. 또한, 역 바이어스를 적용하여 누설 전류를 더욱 줄일 수 있어, 특히 모바일 및 IoT 애플리케이션에 적합하다. 소자 구조 개선과 함께, 반도체 칩의 성능 향상과 전력 소비 감소를 위해 배선 간 간섭을 최소화하는 것이 중요하다. 이를 위해 저유전율 물질의 사용이 필수적이다. 저유전율 물질을 금속 배선 사이에 삽입함으로써 배선 간 정전용량을 크게 줄일 수 있다. 이는 신호 지연을 감소시키고 동시에 충전과 방전에 소비되는 전력을 줄여준다. 최신 공정에서는 유전상수(k)가 2.5 이하인 물질을 사용하여 RC 지연을 최소화하고 있다. 더 나아가, 일부 영역에 실제 공기 간극을 형성하는 기술이 도입되고 있다. 공기의 유전상수는 1에 가까워 이론적으로 가장 낮은 유전율을 제공한다. 이 기술은 특히 중요한 신호선 주변에 선택적으로 적용되어 해당 배선의 성능을 극대화하고 전력 소비를 더욱 감소시킨다. 그러나 공정의 복잡성과 구조적 안정성 문제로 인해 전면적인 도입에는 아직 제약이 있다.



[그림 7] 전통적인 벌크 실리콘 기반 트랜지스터(좌)와 FDSOI 기술이 적용된 트랜지스터(우) [7]

이러한 공정 기술의 최적화는 지속적인 연구와 혁신을 통해 계속 발전하고 있다. 앞으로도 새로운 물질과 구조의 도입, 그리고 기존 기술의 개선을 통해 반도체의 성능과 전력 효율은 계속해서 향상될 것으로 예상된다. 이는 모바일 엣지 컴퓨팅을 비롯한 다양한 응용 분야에서 더욱 효율적이고 강력한 컴퓨팅 솔루션을 가능하게 할 것이다.

3. 시스템 레벨 최적화

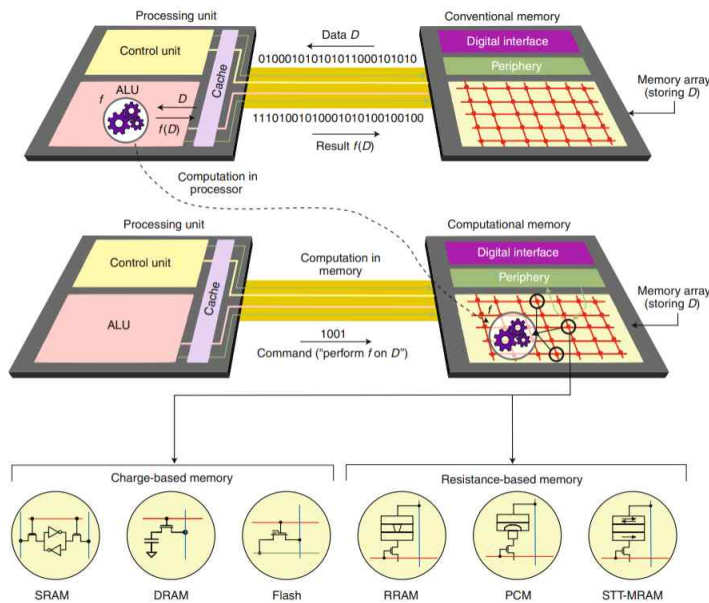
시스템 레벨 최적화는 개별 컴포넌트의 성능 향상을 넘어 전체 시스템의 효율성을 극대화하는 것을 목표로 한다. 이는 하드웨어와 소프트웨어의 긴밀한 협력을 통해 달성될 수 있으며, 특히 모바일 엣지 컴퓨팅 환경에서 중요한 역할을 한다.

이기종 컴퓨팅(Heterogeneous Computing)은 서로 다른 특성을 가진 프로세서들을 효과적으로 조합하여 시스템의 전체적인 성능과 효율성을 높이는 접근 방식이다. 현대의 모바일 시스템온칩(SoC)은 CPU, GPU, NPU 등 다양한 프로세서를 통합하고 있다. 범용 연산을 위한 CPU, 그래픽 처리에 특화된 GPU, 그리고 인공지능 연산에 최적화된 NPU(Neural Processing Unit) 등이 하나의 칩에 함께 구현된다. 예를 들어, 최신 스마트폰 프로세서들은 대부분 이러한 이기종 구조를 채택하고 있으며, 삼성의 Exynos나 퀄컴의 Snapdragon 시리즈가 대표적이다. 또한, 작업 특성에 맞는 최적의 프로세서 선택을 통해 전력 효율 극대화도 이루어지고 있다. 예를 들어, 복잡한 AI 추론 작업은 NPU에서 처리하고, 3D 렌더링은 GPU에서, 일반적인 애플리케이션 실행은 CPU에서 처리한다. 이러한 최적화된 작업 분배는 각 프로세서의 장점을 극대화하고 전력 소비를 최소화한다. 실제로 NPU를 사용한 AI 작업 처리는 CPU나 GPU 대비 10배 이상의 전력 효율을 보일 수 있다.



[그림 8] Heterogeneous Computing 기술이 적용된 Qualcomm 사의 AI Engine [8]

인메모리 컴퓨팅(In-Memory Computing)은 데이터 이동으로 인한 에너지 소비와 지연 시간을 줄이기 위해 메모리 내에서 직접 연산을 수행하는 혁신적인 컴퓨팅 패러다임이다. 전통적인 폰 노이만 구조에서는 데이터를 메모리에서 프로세서로 이동시켜 연산을 수행하고 결과를 다시 메모리에 저장하는 과정이 필요하다. 이 과정에서 상당한 에너지가 소비되고 지연이 발생한다. 인메모리 컴퓨팅은 이러한 데이터 이동을 최소화하여 에너지 효율성을 크게 높인다. 특히 대규모 병렬 처리가 필요한 AI 연산에서 그 효과가 두드러진다. 새로운 메모리 기술들이 인메모리



[그림 9] 전통적인 폰 노이만 구조의 컴퓨팅(위)과 인메모리 컴퓨팅(아래) 구조 [9]

컴퓨팅을 위한 플랫폼으로도 주목받고 있다. 예를 들어, ReRAM(Resistive RAM)은 저항 상태를 이용해 데이터를 저장하고 연산할 수 있어, 매우 효율적인 행렬 연산이 가능하다. MRAM(Magnetoresistive RAM)은 비휘발성과 고속 동작의 장점을 살려 캐시 메모리 대체와 동시에 연산 기능을 제공할 수 있다. 이러한 기술들은 특히 엣지 디바이스에서의 AI 처리에 혁신을 가져올 것으로 기대된다.

엣지 AI는 클라우드에 의존하지 않고 로컬 디바이스에서 AI 연산을 수행하는 것을 의미한다. 이는 빠른 응답 시간, 개인정보 보호, 네트워크 부하 감소 등의 장점이 있지만, 제한된 리소스 내에서 효율적인 AI 처리가 필요하다.

- 쿼타이제이션, 프루닝 등을 통한 AI 모델 경량화/최적화: AI 모델의 크기와 연산량을 줄이는 것은 엣지 디바이스에서의 효율적인 실행을 위해 필수적이다. 쿼타이제이션(Quantization)은 모델의 파라미터를 더 적은 비트 수로 표현하는 기법이다. 예를 들어, 32비트 부동소수점을 8비트 정수로 변환하면 모델 크기를 75% 줄일 수 있으며, 연산 속도도 크게 향상된다. 프루닝(Pruning)은 모델에서 중요도가 낮은 연결이나 뉴런을 제거하는 기법이다. 이를 통해 모델 크기를 대폭 줄이면서도 성능 저하를 최소화할 수 있다. 이러한 기법들을 통해 일부 모델은 원래 크기의 10% 이하로 축소되면서도 90% 이상의 정확도를 유지할 수 있다.
- 하드웨어-소프트웨어 코디자인: 엣지 AI의 성능을 극대화하기 위해서는 소프트웨어(AI 모델)와 하드웨어를 함께 고려한 설계가 필요하다. 특정 AI 모델의 연산 패턴에 최적화된 하드웨어 가속기를 설계한다. 예를 들어, 컨볼루션 신경망(CNN)에 특화된 가속기는 행렬 연산을 효율적으로 처리할 수 있는 구조를 갖는다. 반대로, 하드웨어의 특성을 고려하여 AI 모델을 설계하거나 수정한다. 예를 들어, 특정 하드웨어에서 효율적으로 실행될 수 있는 연산 유형을 선호하여 모델을 구성한다. 이러한 코디자인 접근 방식은 텐서플로우 라이트(TensorFlow Lite)나 NVIDIA의 TensorRT와 같은 프레임워크를 통해 지원되고 있으며, 이를 통해 엣지 디바이스에서 최대 수십 배의 성능 향상을 얻을 수 있다.

이러한 시스템 레벨 최적화 기술들은 서로 유기적으로 결합되어 모바일 엣지 컴퓨팅 환경에서 최상의 성능과 전력 효율을 달성하는 데 기여한다. 앞으로도 하드웨어와 소프트웨어의 경계를 넘나드는 혁신적인 최적화 기술들이 계속해서 등장할 것으로 예상된다.

V

차세대 컴퓨팅 패러다임을 위한 반도체 기술 개발 방향

현재 우리는 기존의 컴퓨팅 패러다임의 한계에 직면하고 있다. 데이터의 폭발적 증가, AI의 보편화, IoT 기기의 확산 등으로 인해 새로운 컴퓨팅 요구사항이 계속해서 등장하고 있다. 이러한 도전과제를 해결하기 위해 혁신적인 컴퓨팅 패러다임이 제안되고 있으며, 이를 실현하기 위한 첨단 반도체 기술의 개발이 필수적이다. 특히 엣지 컴퓨팅과 AI의 융합은 차세대 컴퓨팅의 핵심 트렌드로 부상하고 있으며, 이를 효과적으로 지원할 수 있는 반도체 기술의 중요성이 더욱 커지고 있다.

1. 엣지 AI를 위한 반도체 기술

엣지 AI는 클라우드에 의존하지 않고 로컬 디바이스에서 AI 연산을 수행하는 기술을 말한다. 이는 실시간 처리, 개인정보 보호, 네트워크 부하 감소 등의 장점을 제공하지만, 제한된 리소스 내에서 고성능 AI 처리를 실현해야 하는 도전과제를 안고 있다. 이를 위해 다음과 같은 혁신적인 반도체 기술들이 개발되고 있다.

[초저전력 AI 프로세서]

엣지 디바이스의 제한된 전력 환경에서 고성능 AI 처리를 실현하기 위해 초저전력 AI 프로세서의 개발이 필수적이다. 전력 효율적인 AI 연산 구조의 대표적인 예로, 인메모리 컴퓨팅을 들 수 있다. 이는 데이터 이동을 최소화하여 에너지 효율성을 크게 높이는 기술이다. 예를 들어, IBM의 아날로그 AI 칩은 메모리 어레이 내에서 직접 행렬 연산을 수행하여 기존 GPU 대비 100배 이상의 에너지 효율을 달성했다. 뉴로모픽 컴퓨팅 구조도 주목받고 있다. 인간 뇌의 뉴런과 시냅스를 모방한 이 구조는 스파이킹 신경망을 효율적으로 처리할 수 있어, 특정 AI 작업에서 기존 구조 대비 1000배 이상의 에너지 효율을 보일 수 있다.

동적 정밀도 조절 기능을 갖춘 유연한 연산 유닛도 생각해 볼 수 있다. AI 연산의 특성상 모든 단계에서 고정밀도 연산이 필요하지 않다는 점을 활용한다. 예를 들어, NVIDIA의 Turing GPU 아키텍처는 INT4에서 FP32까지 다양한 정밀도를 지원하며, 작업에 따라 동적으로 정밀도를 조절할 수 있다. 구글의 TPU(Tensor Processing Unit)도 bfloat16이라는 절반 정밀도 부동소수점 형식을 도입하여 연산 효율성을 크게 높였다. 하지만, 정밀도를 낮추면 전력 효율성은 높아지지

만, 정확도가 떨어질 수 있어, 이 둘 사이의 최적 균형점을 찾는 것이 중요하다. 또한, 빠르게 발전하는 다양한 AI 모델들(예: CNN, RNN, Transformer 등)을 효율적으로 지원할 수 있는 유연한 아키텍처 설계가 필요하다.

[온디바이스 학습 기술]

엣지 디바이스에서 직접 학습을 수행하는 기술은 개인화된 AI 서비스 제공과 프라이버시 보호에 큰 도움이 된다. 온디바이스 학습을 위해서는 기존의 추론 중심 하드웨어와는 다른 지속적 학습이 가능한 하드웨어 구조 접근이 필요하다. 예를 들어, 퀄컴의 Snapdragon Neural Processing Engine은 온디바이스에서의 전이학습과 증분학습을 지원한다. 뉴로모픽 칩은 그 특성상 지속적 학습에 적합한 구조를 가지고 있어, Intel의 Loihi 칩과 같은 뉴로모픽 프로세서가 이 분야에서 주목받고 있다.

프라이버시 보호를 위한 안전한 학습 메커니즘도 필요하다. 연합 학습(Federated Learning)과 같은 기술을 하드웨어 레벨에서 지원하여 개인 데이터를 디바이스 밖으로 유출시키지 않고도 모델을 개선할 수 있다. 또한, 차등 프라이버시(Differential Privacy) 기술을 하드웨어에 내장하여 학습 과정에서 개인 정보 보호를 강화할 수 있다. 하지만, 제한된 메모리와 연산 능력을 가진 엣지 디바이스에서 효과적인 학습을 수행하기 위해서는 모델 압축, 증분 학습 등의 기술도 필요할 것이다. 덧붙여, 배터리 수명을 고려한 에너지 효율적인 학습 알고리즘과 하드웨어 설계가 요구된다.

[센서-AI 통합 기술]

AI와 센서 기술의 통합은 엣지 AI의 성능과 효율성을 크게 향상시킬 수 있는 접근 방식이다. 센서와 AI 프로세서를 하나의 칩에 통합함으로써 데이터 이동에 따른 지연과 에너지 소비를 최소화할 수 있다. 예를 들어, Sony의 IMX500 이미지 센서는 AI 처리 기능을 내장하여 카메라 모듈 자체에서 객체 인식 등의 작업을 수행할 수 있다. 이러한 통합은 특히 웨어러블 기기나 IoT 센서와 같은 초소형, 초저전력 디바이스에서 큰 이점을 제공한다. 또 다른 예를 들어보면, 센서에서 받아들인 아날로그 신호를 디지털로 변환하기 전에 아날로그 도메인에서 초기 처리를 수행하면 에너지 효율성을 크게 높일 수 있다. 예를 들어, 음성 인식 시스템에서 아날로그 필터를 사용해 초기 특징 추출을 수행하면, 디지털 처리에 비해 훨씬 적은 에너지로 유사한 성능을 달성할 수 있다.

2. 뉴로모픽 컴퓨팅을 위한 반도체 기술

뉴로모픽 컴퓨팅은 인간 뇌의 구조와 기능을 모방한 새로운 컴퓨팅 패러다임이다. 이 기술은 기존의 폰 노이만 아키텍처의 한계를 극복하고, 특히 패턴 인식, 의사 결정, 적응적 학습과 같은 인지 작업에서 뛰어난 성능과 에너지 효율성을 제공할 수 있다. 뉴로모픽 컴퓨팅을 실현하기 위해서는 다음과 같은 혁신적인 반도체 기술들이 필요하다.

[대규모 뉴로모픽 시스템]

대규모 뉴로모픽 시스템 구현을 위해서는 수백만 개 이상의 뉴런과 시냅스로 구성된 칩 설계가 필수적이다. IBM의 TrueNorth나 Intel의 Loihi 2와 같은 칩들이 이미 백만 단위의 뉴런을 집적하는데 성공했으며, 이를 통해 복잡한 인지 작업 수행이 가능해지고 있다. 또한, 뇌의 복잡한 연결 구조를 모방하기 위해 3D 집적 기술이 활용되고 있다. 삼성전자 등에서 연구 중인 3D 적층 기술은 뉴런 간의 연결 밀도를 크게 높일 수 있어, 더욱 복잡한 신경망 구현이 가능해질 전망이다. 하지만 이러한 대규모 시스템 구현에는 전력 관리와 수율 향상이라는 중요한 과제가 남아 있다. 수백만 개의 뉴런이 동시에 활성화될 때 발생하는 열을 효과적으로 관리해야 하며, 복잡한 3D 구조로 인한 제조 과정에서의 수율 문제도 해결해야 한다.

[온라인 학습 가능한 아키텍처]

온라인 학습 가능한 아키텍처는 뉴로모픽 시스템의 핵심 특징 중 하나이다. STDP와 같은 생물학적 학습 규칙을 하드웨어로 구현함으로써 실시간 학습과 적응이 가능해진다. BrainScaleS 프로젝트의 뉴로모픽 칩이 이러한 기술을 선보였으며, 이 외에도 다양한 생물학적 학습 규칙들이 하드웨어로 구현되고 있다. 또한, 시냅스 가중치를 효율적으로 저장하고 업데이트하기 위해 ReRAM이나 PCM과 같은 비휘발성 메모리 기술이 활용되고 있다. 이러한 기술들은 생물학적 시냅스의 특성을 더 잘 모방할 수 있어 주목받고 있다. 그러나 온라인 학습 시스템에서는 학습의 안정성과 수렴성을 보장하는 것이 큰 과제로 남아있으며, 이를 해결하기 위해 생물학적 뇌에서 영감을 받은 다양한 조절 메커니즘들이 연구되고 있다.

[범용 뉴로모픽 프로세서]

뉴로모픽 컴퓨팅이 널리 활용되기 위해서는 범용 뉴로모픽 프로세서의 개발이 필요하다. 다양한 신경망 모델을 지원할 수 있는 유연한 구조와 기존 컴퓨팅 시스템과의 효율적인 인터페이스가 중

요한 요소이다. Intel의 Loihi 2 칩은 이러한 유연성을 목표로 설계되었으며, IBM의 TrueNorth 시스템은 기존 컴퓨터 시스템과의 원활한 통신을 위한 인터페이스를 제공한다. 그러나 이러한 범용 뉴로모픽 프로세서의 개발에는 여전히 많은 과제가 남아 있다. 효과적인 프로그래밍 모델과 개발 도구의 개발이 필요하며, 뉴로모픽 컴퓨팅의 장점을 극대화할 수 있는 새로운 응용 분야를 발굴하는 것도 중요한 과제이다.

이러한 뉴로모픽 컴퓨팅을 위한 반도체 기술들은 아직 초기 단계에 있지만 빠르게 발전하고 있다. 이 기술들이 성숙해짐에 따라 우리는 더욱 지능적이고 에너지 효율적인 컴퓨팅 시스템을 구현할 수 있을 것이며, 이는 AI와 컴퓨팅 기술의 새로운 지평을 열 것으로 기대된다. 향후 이러한 기술들이 로봇틱스, 자율주행, 실시간 영상 처리 등 다양한 분야에서 혁신을 이끌어낼 것으로 전망된다.

3. 그 외, 새로운 컴퓨팅 패러다임 탐색

새로운 컴퓨팅 패러다임의 탐색은 현재 컴퓨팅 기술의 한계를 극복하고 미래의 복잡한 문제들을 해결하기 위한 중요한 연구 분야입니다. 이 중에서 특히 주목받고 있는 세 가지 혁신적인 접근 방식은 DNA 컴퓨팅, 광학 컴퓨팅, 그리고 바이오 인스파이어드 컴퓨팅입니다.

DNA 컴퓨팅은 생명의 기본 정보 저장 매체인 DNA 분자를 이용해 정보를 저장하고 처리하는 혁신적인 기술이다. 이 기술은 DNA의 높은 정보 밀도와 병렬 처리 능력을 활용하여 특정 문제들을 효율적으로 해결할 수 있는 잠재력을 가지고 있다. 또한, DNA의 생체 적합성은 의료 분야에서의 응용 가능성을 크게 높인다. 예를 들어, 체내에서 특정 조건을 감지하고 반응하는 나노 크기의 컴퓨팅 장치를 만들 수 있을 것이다. 그러나 이 기술이 실용화되기 위해서는 DNA의 읽기와 쓰기 속도를 크게 향상시키고 오류율을 낮추는 것이 중요한 과제로 남아 있다.

광학 컴퓨팅은 전자 대신 빛을 이용해 정보를 처리하는 기술로, 초고속 연산과 낮은 전력 소비를 동시에 실현할 수 있는 잠재력을 가지고 있다. 빛의 속도로 정보를 전달하고 처리할 수 있기 때문에, 특히 대규모 데이터 처리나 복잡한 시뮬레이션 등에서 큰 이점을 제공할 수 있다. 현재는 광-전자 하이브리드 시스템 개발에 많은 노력이 집중되고 있으며, 이를 통해 기존 전자 시스템과의 호환성을 유지하면서도 광학 컴퓨팅의 이점을 활용할 수 있을 것으로 기대된다. 그러나 광학 컴퓨팅 기술이 널리 채택되기 위해서는 시스템의 소형화와 집적도 향상이 필요하다. 나노포토닉스 기술의 발전이 이러한 과제 해결에 핵심적인 역할을 할 것으로 예상된다.

바이오 인스파이어드 컴퓨팅은 생물학적 시스템, 특히 뇌의 구조와 기능에서 영감을 받아 새로운 컴퓨팅 아키텍처를 개발하는 접근 방식이다. 이 기술은 뉴로모픽 컴퓨팅을 포함하여 더 넓은 범위의 생물학적 원리를 컴퓨팅에 적용하는 것을 목표로 한다. 자기 조직화 능력과 높은 적응성은 이러한 시스템의 핵심 특징으로, 복잡하고 동적인 환경에서 효과적으로 작동할 수 있는 컴퓨팅 시스템을 만드는 것이 목표이다. 예를 들어, 자가 수리 능력을 갖춘 컴퓨터 칩이나 환경 변화에 자동으로 적응하는 로봇 시스템 등을 구현할 수 있을 것이다. 그러나 이러한 시스템을 실현하기 위해서는 생물학적 시스템의 뛰어난 효율성과 강건성을 인공적인 시스템에서 구현하는 것이 큰 과제로 남아있다.

이러한 새로운 컴퓨팅 패러다임들은 각각 고유한 장점과 도전 과제를 가지고 있다. DNA 컴퓨팅은 생체 내 컴퓨팅과 초고밀도 정보 저장의 가능성을 제시하고, 광학 컴퓨팅은 초고속, 저전력 연산의 미래를 보여주며, 바이오 인스파이어드 컴퓨팅은 적응성과 효율성이 뛰어난 시스템의 비전을 제시한다. 이러한 기술들이 성숙해짐에 따라, 우리는 현재의 실리콘 기반 컴퓨팅의 한계를 뛰어넘는 혁신적인 컴퓨팅 솔루션을 보게 될 것이다. 이는 인공지능, 의료기술, 환경 모니터링, 우주 탐사 등 다양한 분야에서 획기적인 발전을 이끌어낼 수 있을 것으로 기대된다.

결론적으로, 반도체 기술의 미래는 도전과 기회가 공존하고 있다. 기술적 한계를 극복하고 새로운 패러다임을 창출하는 과정에서, 우리는 더욱 스마트하고 지속가능한 세상을 만들어갈 수 있을 것이다. 이를 위해서는 장기적인 비전을 가지고 꾸준히 투자하며, 글로벌 협력을 강화하는 동시에 각국의 기술 주도권을 확보하기 위한 전략적 접근이 필요할 것이다.



참고 문헌

- [1] ASML
- [2] Yuan, G., Wang, N., Huang, S., & Liu, J. (2016, August). A brief overview of atomic layer deposition and etching in the semiconductor processing. In 2016 17th International Conference on Electronic Packaging Technology (ICEPT) (pp. 1365-1368). IEEE.
- [3] 삼성전자
- [4] J. Park, Chiplets and Heterogeneous Packaging Are Changing System Design and Analysis, Accessed on July 7, 2023.
- [5] Veloso, A., Huynh-Bao, T., Matagne, P., Jang, D., Eneman, G., Horiguchi, N., & Ryckaert, J. (2020). Nanowire & nanosheet FETs for ultra-scaled, high-density logic and memory applications. *Solid-State Electronics*, 168, 107736.
- [6] Lee, J. C., Kim, J., Kim, K. W., Ku, Y. J., Kim, D. S., Jeong, C., ... & Lee, S. H. (2016, October). High bandwidth memory (HBM) with TSV technique. In 2016 International SoC Design Conference (ISOCC) (pp. 181-182). IEEE.
- [7] Cheng, K., & Khakifirooz, A. (2016). Fully depleted SOI (FDSOI) technology. *Science China Information Sciences*, 59, 1-15.
- [8] Qualcomm
- [9] Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R., & Eleftheriou, E. (2020). Memory devices and applications for in-memory computing. *Nature nanotechnology*, 15(7), 529-544.



용어정리집

I. 반도체 기술 개요 및 동향

【무어의 법칙 (Moore's Law)】

- 18~24개월 정도의 주기로, 반도체 집적회로의 성능이 향상되는 경험 법칙. 흔히, 반도체 칩에 들어가는 소자 수가 2년마다 2배씩 증가해, 칩 성능이 향상된다고 알려져 있음. 하지만, 반도체 칩의 복잡도가 2년마다 2배씩 증가하는 경험 법칙.

【초미세 공정 (3nm 이하 제조 공정)】

- 반도체 소자의 크기를 나노미터 스케일로 아주 작게 만드는 제조 공정 기술.

【EUV 리소그래피 (Extreme Ultraviolet Lithography)】

- 13.5nm 파장의 빛을 이용해 미세한 회로 패턴을 웨이퍼 위에 전사하는 포토 공정 기술.

【GAA(Gate-All-Around) 트랜지스터】

- 트랜지스터의 채널 내에 흐르는 전류를 제어하는 게이트가, 채널 모든 면을 감싸는 3차원 구조 반도체 트랜지스터.

【PPAC (Power, Performance, Area, Cost)】

- 반도체 소자 및 설계에서 중요한 네 가지 요소로, 전력(Power), 성능(Performance), 면적(Area), 비용(Cost)의 줄임말. 반도체 칩 설계와 제조 공정에서 이 요소들을 최적화하는 것이 경쟁력의 핵심입니다.

【4F²】

- 메모리 셀 하나의 면적을 나타내는 표현으로, F는 패턴 크기를 의미합니다. 미세 패턴 간의 pitch를 2F로 하기 때문에, 최소의 면적은 2F와 2F의 곱인 4F²이 됨.

【칩렛 (Chiplet)】

- 여러 기능의 반도체 칩을 하나의 큰 다이(die)로 만들지 않고, 여러 개의 작은 칩(칩렛)으로 나누어 개별적으로 제작한 후 서로 연결하여 하나의 칩을 만드는 방식. 이 방식은 제조 공정을 효율적으로 관리하고 설계를 유연하게 할 수 있는 장점이 있음.

【mini-GVC (global value chain)】

- 미니 글로벌 가치 사슬(Global Value Chain)의 준말로, 반도체 등 첨단 산업에서 가치 사슬의 복잡성을 줄이고 특정 지역 내에서 공급망을 최적화하는 전략을 의미합니다.

【FinFET, GAAFET (Gate-All-Around FET), CFET (Complementary FET)】

- FinFET: 물고기 지느러미 (Fin)의 모양과 유사하게 좁고 긴 채널 영역을 가진 3차원 반도체 소자.
- AAFET : 트랜지스터의 채널 내에 흐르는 전류를 제어하는 게이트가, 채널 모든 면을 감싸는 3차원 구조 반도체 트랜지스터.
- CFET: 기존 3차원 트랜지스터를 위아래로 적층한 반도체 소자. n-type 트랜지스터와 p-type 트랜지스터를 위아래로 적층해서, complementary 트랜지스터로 명명됨.

【Cu-to-Cu and Oxide-to-Oxide】

- 반도체 칩의 인터커넥트(interconnect)에서 사용되는 접합 방식으로, Cu-to-Cu는 구리 대 구리 접합을, Oxide-to-Oxide는 산화물 대 산화물 접합을 뜻합니다. 각각 고속 데이터 전송과 안정적인 전기적 특성 제공을 위해 활용됩니다.

【GaN, SiC, 그래핀, MoS2】

- GaN (Gallium Nitride): 전력 소모가 적고 효율이 높아 전력 반도체에 주로 쓰이는 반도체 재료입니다.
- SiC (Silicon Carbide): 고온과 고전압에서도 안정적인 성능을 발휘하는 소재로, 전기차와 전력 반도체에 많이 사용됩니다.
- 그래핀 (Graphene): 매우 얇고 강하며 전기 전도성이 우수한 소재로, 다양한 차세대 전자기기 및 반도체 소자로 연구되고 있습니다.
- MoS2 (Molybdenum Disulfide): 2차원 물질로, 그래핀보다 넓은 밴드갭을 가지고 있어 디지털 소자에 유용하게 사용됩니다.

【클럭 신호 (Clock Signal)】

- 전자 회로에서 동작 타이밍을 제어하기 위한 신호입니다. 클럭 신호는 시스템이 규칙적으로 동작하도록 해 프로세서와 같은 디지털 회로의 동기화를 유지합니다.

【3D NAND, DRAM, SRAM, MRAM】

- 3D NAND: 셀을 수직으로 적층하여 집적도를 높인 플래시 메모리입니다. 저장 공간을 더 많이 확보할 수 있어 고용량 메모리에 적합합니다.
- RAM: 동적 랜덤 액세스 메모리로, 데이터 저장이 휘발성이라 전원이 끊기면 데이터가 사라지지만 NAND보다 빠른 속도로 데이터를 읽고 쓰기에 적합합니다.
- SRAM: 정적 랜덤 액세스 메모리로, 전원이 켜져 있는 동안 데이터를 유지하며 가장 빠른 접근 속도를 가집니다. 일반적으로 캐시 메모리로 사용됩니다.
- MRAM: 자기 저장형 랜덤 액세스 메모리로, 자성 물질을 이용해 데이터를 저장하는 비휘발성 메모리입니다.

【M1, M2 등의 칩】

- 애플의 M1, M2 칩은 자사의 제품을 위해 개발한 고성능 ARM 기반의 SoC(System on a Chip)로, CPU와 GPU, 뉴럴 엔진 등을 하나의 칩에 통합하여 높은 효율성과 성능을 제공합니다.

II. 반도체 제조 공정 혁신**【반도체 웨이퍼】**

- 반도체 칩의 제조에 사용되는 얇고 평평한 원형의 기판입니다. 일반적으로 규소(Si)를 사용하여 만들어지며, 이 위에 리소그래피, 증착, 식각 등의 공정을 거쳐 트랜지스터 및 집적회로가 형성됩니다.

【리소그래피 (Lithography)】

- 빛을 이용해 반도체 웨이퍼 표면에 미세한 회로 패턴을 인쇄하는 공정입니다. 포토레지스트라는 감광 소재를 이용해 빛을 쬐면 원하는 회로 패턴이 형성됩니다. 반도체 미세 공정의 핵심 기술로, 회로 선폭을 줄이기 위해 EUV와 같은 기술이 개발되었습니다.

【High-NA EUV (High Numerical Aperture Extreme Ultraviolet Lithography)】

- 기존 EUV 장비보다 개구수 (Numerical Aperture; 광학에서 시스템이 빛을 받아들이거나 내보내는 입사각의 특징을 지닌 무차원 수)를 더 키워서, 보다 더 정밀한 포토 공정을 가능하게 하는 노광 장비 기술.

【블록 공중합체 (Directed Self-Assembly, DSA)】

- 자가 조립(Self-Assembly) 특성을 이용해 나노미터 크기의 패턴을 형성하는 기술로, 특정 조건에서 서로 다른 두 종류의 고분자 블록이 규칙적으로 배열됩니다. 리소그래피와 결합하여 미세 공정에서의 패턴 형성에 활용될 수 있습니다.

【원자층 식각 (Atomic Layer Etching, ALE)】

- 매우 얇은 층을 원자 단위로 정밀하게 제거하는 식각 기술입니다. 기존 식각 공정보다 정밀하고 균일한 두께 조절이 가능해 고성능 반도체 제조에 적합합니다. 미세 공정에서 특히 중요한 식각 기술로 자리 잡고 있습니다.

【원자층 증착 (Atomic Layer Deposition, ALD)】

- 원자 단위로 얇은 박막을 균일하게 쌓는 증착 기술입니다. 공정의 두께를 원자 단위로 제어할 수 있어 정밀한 두께 조절이 필요할 때 사용되며, 고성능 반도체와 디스플레이 제조 공정 등에 활용됩니다.

【TSV (Through-Silicon Via)】

- 반도체 칩 내부에서 수직으로 신호를 주고받는 통로.

【저유전률 (Low-k Dielectric)】

- 유전율(k 값)이 낮은 절연체를 의미하며, 반도체 칩 내 전력 소모를 줄이고 전송 신호 간 간섭을 줄이기 위해 사용되며, 고성능 집적회로 설계에 중요한 소재입니다.

Ⅲ. 차세대 반도체 아키텍처

【나노와이어 (Nano-wire)】

- 직경이 나노미터 수준인 가늘고 긴 도선으로, 반도체 소자, 센서 등에서 활용됨.

【나노시트 (Nano-sheet)】

- 두께가 나노미터 수준인 얇은 판 형태의 물질로, 고성능 반도체에 사용됨.

【선택적 에피택시 성장 (Selective Epitaxial Growth, SEG)】

- 특정 영역에서 결정 성장을 유도하여 소자 특성을 향상시키는 기술.

【고유전율 게이트 절연막 (High-k Dielectric)】

- 전통적인 실리콘 산화막보다 높은 유전율을 가진 절연층으로, 누설 전류를 줄임.

【다층 박막 증착】

- 박막을 여러 층으로 증착하여 원하는 전기적, 물리적 특성을 얻는 공정.

【패터닝 기술】

- 특정 형태나 패턴을 반도체 웨이퍼에 형성하는 기술.

【Channel Release 공정】

- 채널의 캐리어 이동성을 향상시키기 위해 반도체에서 채널을 분리하는 공정.

【희생층 (Sacrificial Layer)】

- 제조 과정에서 임시로 사용되며 최종 단계에서 제거되는 층.

【등방성 식각 기술】

- 모든 방향으로 균일하게 식각하여 특정 형태를 만드는 기술.

【소스/드레인 최적화 기술】

- 소자 성능 향상을 위해 소스와 드레인의 전기적 특성을 개선하는 기술.

【저전력 모바일 AP (Application Processor)】

- 모바일 기기에 사용되는 애플리케이션 프로세서로, 낮은 전력 소모와 높은 성능을 제공.

【저잡음 증폭기 (LNA, Low Noise Amplifier)】

- 신호의 품질을 유지하면서 잡음을 최소화하여 약한 신호를 증폭하는 장치.

【NMOS, PMOS】

- NMOS: 전자의 이동을 기반으로 작동하는 트랜지스터.
- MOS: 정공의 이동을 기반으로 작동하는 트랜지스터.

【Parasitic Capacitance (기생 커패시턴스)】

- 회로에서 의도치 않게 발생하는 커패시턴스로 신호 속도와 전력 효율에 영향을 미침.

【인터커넥트 길이】

- 반도체 칩 내에서 소자 간 신호 연결의 길이. 짧을수록 성능이 향상됨.

【Backside Power Distribution Network】

- 칩 뒷면을 통해 전력을 공급하여 성능과 효율을 향상시키는 전력 분배 기술.

【신호 무결성 (Signal Integrity)】

- 신호 전송 시 왜곡 없이 정확하게 전달되는 능력.

【이종 집적 기술】

- 서로 다른 특성을 가진 소자를 하나의 시스템으로 통합하는 기술.

【시스템 온 칩 (SoC, System on Chip)】

- CPU, GPU, 메모리, I/O(입/출력) 등을 하나의 칩에 통합한 시스템.

【Metrology and Inspection 기술】

- 반도체 제조 과정에서 소자의 크기와 품질을 측정하고 검사하는 기술.

【엣지 컴퓨팅】

- 데이터를 생성한 디바이스 근처에서 처리하여 지연을 줄이고 실시간성을 높이는 기술.

【모놀리식 3D 집적】

- 동일한 기판 위에 여러 층의 회로를 수직으로 쌓아 올리는 방식.

【웨이퍼 본딩 기술】

- 두 개 이상의 웨이퍼를 접합하여 하나의 칩으로 만드는 기술.

【고대역폭 메모리 (HBM, High Bandwidth Memory)】

- 대규모 데이터를 빠르게 처리할 수 있는 고성능 메모리 기술.

【Via 홀 형성】

- 반도체 기판에 전기적 연결을 위해 미세한 구멍을 만드는 공정.

【Via 충전】

- Via 홀에 금속을 채워 전기적 연결을 완성하는 공정.

【절연막 증착】

- 전기적 절연을 위한 절연층을 형성하는 공정.

【배리어/시드 레이어 형성】

- 금속 증착 전 확산 방지(배리어)와 초기 증착(시드)을 돕는 층 형성.

【CMP (Chemical Mechanical Polishing)】

- 반도체 표면을 화학적, 기계적으로 연마하여 평탄화하는 공정.
- AI 가속기 (GPU, TPU): 인공지능 연산을 빠르게 처리하는 장치.

【UCle (Universal Chiplet Interconnect Express)】

- 칩 간 데이터를 고속으로 주고받기 위한 표준 인터페이스.

【이미지 센서】

- 광신호를 전기 신호로 변환하여 디지털 이미지를 생성하는 소자.

【RF/아날로그 소자】

- 무선 주파수와 아날로그 신호를 처리하는 소자.

【뉴로모픽 칩】

- 인간 뇌의 뉴런 구조를 모방하여 설계된 칩으로, AI 응용에 사용됨.

【광전자 소자】

- 빛과 전기 신호 간 상호작용을 기반으로 동작하는 소자.

【MEMS (Micro-Electro-Mechanical Systems)】

- 미세 전기 및 기계 부품을 통합한 시스템.

【FOWLP (Fan-Out Wafer Level Packaging)】

- 웨이퍼 레벨에서 소자를 패키징하여 성능과 효율을 높이는 기술.

【TMDC (Transition Metal Dichalcogenides)】

- 이차원 반도체 물질군으로 MoS₂ 등이 포함됨.

【MoS₂ (이황화몰리브덴)】

- 뛰어난 전자적, 광학적 특성을 가진 이차원 반도체 물질.

【위상학적 절연체 (Topological Insulator)】

- 내부는 절연체이나 표면에서는 전기가 흐르는 특이한 물질.

【자기조립 (Self-assembly) 기술】

- 분자나 나노 구조가 스스로 조립되어 특정 구조를 형성하는 기술.

【다이아몬드 기반 열전도층 삽입 기술】

- 열전도율이 높은 다이아몬드층을 소자에 삽입하여 열 방출을 개선하는 기술.

【마이크로채널 냉각 시스템】

- 미세 채널을 통해 냉각제를 순환시켜 효과적으로 열을 제거하는 기술.

【Bio-inspired 3D 집적 기술】

- 생물학적 구조에서 영감을 받아 설계된 3D 집적 기술.

【3D 뉴럴 네트워크】

- 3차원 구조로 이루어진 신경망으로, 데이터 처리와 학습 효율을 극대화.

【DNA 오리가미 기술】

- DNA를 접어 나노 구조물을 형성하는 기술로, 나노기기 설계에 활용.

【나노스케일 3D 패터닝 기술】

- 나노미터 단위에서 3D 구조를 형성하는 패터닝 기술.

【3D FPGA 구조】

- 3D로 집적된 FPGA로, 성능과 전력 효율이 향상된 구조.

【포스트 무어 시대】

- 무어의 법칙이 한계에 도달한 시대로, 새로운 컴퓨팅 기술이 요구되는 시대.

【에너지 하베스팅 기능】

- 주변의 에너지를 수집하여 전력으로 변환하는 기술.

【픽셀 어레이】

- 이미지 센서에서 빛을 전기 신호로 변환하는 픽셀들의 배열.

【신호 처리 회로】

- 신호를 원하는 형태로 변환하거나 분석하는 전자 회로.

IV. 모바일 엣지 컴퓨팅을 위한 저전력 반도체 기술

【eDRAM (embedded DRAM)】

- CPU나 GPU 등 프로세서 내부에 임베디드(내장)된 DRAM으로, 외부 메모리에 비해 접근 속도가 빠르고 대역폭이 넓습니다. eDRAM은 캐시 메모리로 사용되며, 전력 소모를 줄이고 메모리 전송 지연 시간을 줄일 수 있습니다.

【MRAM (Magnetoresistive RAM)】

- 자성 물질을 사용해 데이터를 저장하는 비휘발성 메모리로, 전원이 꺼져도 데이터가 유지됩니다. MRAM은 높은 쓰기 및 읽기 속도와 긴 수명을 제공하며, 전력 효율성이 높아 차세대 메모리로 주목받고 있습니다.

【FDSOI (Fully Depleted Silicon On Insulator)】

- 절연체 위에 얇은 실리콘을 얹어 반도체 소자를 제작하는 공정으로, 기존 벌크 실리콘 공정보다 누설 전류를 줄이고 전력 효율을 높일 수 있습니다. FDSOI는 주로 저전력 고성능 소자에 사용됩니다.

【유전상수 (Dielectric Constant, k)】

- 물질이 전기장을 저장할 수 있는 능력을 나타내는 상수입니다. 반도체 칩에서 저유전상수 (low-k) 물질을 사용할 경우, 칩 내 전력 소모와 신호 간섭을 줄일 수 있어 고성능 집적회로 설계에 유리합니다.

【NPU (Neural Processing Unit)】

- 인공지능 및 머신러닝 작업을 효율적으로 처리하기 위한 전용 프로세서입니다. 뉴럴 네트워크의 계산에 최적화되어 있으며, CPU나 GPU에 비해 전력 소모가 적고 빠른 속도로 딥러닝 연산을 처리합니다.

【퀀타이제이션 (Quantization)】

- 딥러닝 모델의 가중치와 활성화 값을 낮은 비트로 변환해 모델의 크기를 줄이고 계산 속도를 높이는 최적화 기술입니다. 퀀타이제이션을 통해 메모리 사용량과 전력 소모를 줄일 수 있으며, 특히 모바일 기기와 엣지 디바이스에서 유용하게 사용됩니다.

【프루닝 (Pruning)】

- 딥러닝 모델의 불필요한 뉴런이나 연결을 제거하여 모델을 경량화하고 추론 속도를 높이는 기술입니다. 프루닝을 통해 모델의 복잡성을 줄이면서도 정확도를 크게 손상시키지 않도록 합니다.

【TensorRT】

- NVIDIA가 개발한 딥러닝 모델 최적화 및 추론 라이브러리입니다. TensorRT는 모델의 속도를 높이고 메모리 사용량을 줄여 실시간 AI 응용 프로그램의 성능을 향상시키는 데에 사용됩니다.

V. 차세대 컴퓨팅 패러다임을 위한 반도체 기술 개발 방향**【TPU (Tensor Processing Unit)】**

- 구글이 개발한 AI와 머신러닝 작업 전용 프로세서로, 텐서 연산을 효율적으로 처리하도록 설계되었습니다. TPU는 구글 클라우드와 AI 연구에서 고속 추론과 학습을 지원하기 위해 널리 사용됩니다.

【Federated Learning (연합 학습)】

- 데이터를 중앙 서버로 수집하지 않고, 분산된 여러 기기에서 개별적으로 학습한 모델을 결합하여 하나의 글로벌 모델을 생성하는 학습 방식입니다. 개인정보 보호가 중요한 환경에서 사용되며, 대표적으로 스마트폰에서의 개인화 서비스에 활용됩니다.

【Differential Privacy (차등 프라이버시)】

- 데이터를 분석할 때 개인의 프라이버시를 보호하기 위해 고안된 기술입니다. 데이터에 노이즈를 추가하여 개별 데이터가 유출되지 않도록 하면서도 유용한 통계 결과를 제공할 수 있습니다. 연합 학습 및 데이터 분석에서 개인정보 보호를 강화하는 데 사용됩니다.

【Transfer Learning (전이 학습)】

- 기존에 학습된 모델의 지식을 새로운 문제에 적용하는 학습 방법입니다. 보통 많은 데이터를 필요로 하는 대규모 모델을 사전 학습한 후, 이를 특정 작업에 맞게 미세 조정합니다. 이로 인해 학습 시간과 데이터 요구량을 줄일 수 있습니다.

【Incremental Learning (증분 학습)】

- 기존에 학습된 모델에 새로운 데이터를 점진적으로 추가하면서 학습하는 방식입니다. 처음부터 다시 학습할 필요 없이 점진적으로 성능을 개선할 수 있어, 특히 지속적으로 변화하는 데이터에 대응하는 데 유용합니다.

【Spike-Timing-Dependent Plasticity (STDP)】

- 뉴런 간의 연결 강도가 신호 전달 시간에 따라 변하는 생물학적 학습 메커니즘입니다. 이 메커니즘은 뉴로모픽 컴퓨팅에서 시냅스 강도를 조절하는 방법으로 사용되며, 뉴런 간의 상호작용을 모사하여 효율적인 신경망 모델을 개발하는 데 기여합니다.

【Resistive Random-Access Memory (ReRAM)】

- 전류에 의해 저항값이 변화하는 특성을 이용해 데이터를 저장하는 비휘발성 메모리입니다. ReRAM은 빠른 속도, 높은 집적도, 낮은 전력 소모를 제공하며, 기존 플래시 메모리의 대체 기술로 주목받고 있습니다.

【Phase-Change Memory (PCM)】

- 물질의 상변화(Phase Change)를 이용하여 데이터를 저장하는 비휘발성 메모리입니다. 열을 가해 물질의 상태를 결정해 데이터를 기록하고, 읽을 때는 전기 저항을 통해 상태를 감지합니다. PCM은 빠른 읽기 속도와 높은 내구성을 갖춰 차세대 메모리로 연구되고 있습니다.

【Tensor Processing Unit (TPU)】

- 구글이 딥러닝 연산에 최적화하여 설계한 프로세서로, 텐서 연산을 매우 빠르게 수행합니다. TPU는 고성능의 AI 학습 및 추론을 위해 설계되었으며, 주로 클라우드 서비스와 연구 환경에서 사용됩니다.

【bfloat16】

- 딥러닝에서 활용되는 16비트 부동소수점 데이터 형식으로, 기존 32비트보다 메모리 효율이 높으면서도 수학적 정밀도를 유지하도록 설계되었습니다. bfloat16은 특히 텐서 연산의 속도를 높이는 데 유용하며, TPU 및 기타 딥러닝 하드웨어 가속기에서 지원됩니다.

NIA

Digital Insight 2024

반도체 기술발전과

미래컴퓨팅 기술의 진화(상)

▮ 발 행 : 2024.11.27.

▮ 발행인 : 황중성

▮ 발행처 : 한국지능정보사회진흥원(NIA)인공지능정책본부 미래전략팀

▮ 기획 및 문의 : 정현영 주임연구원(hyeon@nia.or.kr)

▮ 작 성 : 고려대학교 신창환 교수(cshin@korea.ac.kr)

- NIA 「Digital Insight 2024」는 AX(AI Transformation)·DX(Digital Transformation) 시대를 맞이해 다가오는 미래를 준비하고, 미래 지능화 시대를 선제적으로 대응하기 위해 한국지능정보사회진흥원(NIA, 이하 한국지능정보원)에서 발간하는 보고서입니다.
- 본보고서는 방송통신발전기금으로 수행한 정보통신·방송 연구개발 사업의 결과물이므로, 보고서의 내용을 발표할 때는 반드시 과학기술정보통신부 정보통신·방송 연구개발 사업의 연구 결과임을 밝혀야 합니다.
- NIA의 승인 없이 본 보고서의 무단전재나 복제를 금하며, 인용하실 때는 반드시 NIA 「Digital Insight 2024」 라고 밝혀주시기 바랍니다. 보고서 내용에 대한 문의나 제안은 위의 연락처로 해주시기 바랍니다.
- 본 보고서의 내용은 한국지능정보원(NIA)의 공식 견해와 다를 수 있습니다.