

## chapter 2

# AI와 융합을 위한 차세대 디바이스 기술 동향



임태규 || 삼성전자 선임연구원

박재웅 || 현대오트모터 대리

## I. 서론

차세대 디바이스는 웨어러블, IoT(Internet of Things), AR(Augmented Reality), AI(Artificial Intelligence) 디바이스 등 다양한 형태로 발전하고 있다. 특히, AI는 딥러닝(Deep Learning) 모델을 기반으로 이미지, 자연어 등 시퀀셜(Sequential)한 데이터 처리에서 혁신적인 성능 향상을 이루어내 다양한 산업 분야로 확산되고 있다.

차세대 AI 디바이스는 딥러닝 모델을 학습시키고 추론하기 위해서 강력한 GPU(Graphic Processing Unit) 기반의 컴퓨팅 파워를 필요로 한다. 이를 위해 실제 서비스 환경에서는 강력한 컴퓨팅 파워를 가진 서버를 중심으로 한 클라우드 인프라가 요구된다. 하지만 이러한 클라우드 인프라는 몇 가지 문제점을 갖고 있다. 첫째, 수많은 IoT 디바이스에서 나오는 빅데이터가 네트워크 트래픽을 폭발적으로 증가시킨다는 점이다. 이는 필연적으로 네트워크 환경을 악화시키고 성능 저하를 유발한다. 둘째, 엔드포인트(Endpoint)에 있는 디바이스에서 퍼스트 마일(First Mile)에 있는 클라우드 서버까지 데이터를 송/수신하기 위한 네트워크의 물리적 거리로 인한 성능 저하 문제이다. 라스트 마일>Last Mile)에서

\* 본 내용은 임태규 선임연구원(☎ 070-4872-6200, lswsjw@gmail.com))에게 문의하시기 바랍니다.

\*\* 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

처리되는 데이터와 퍼스트 마일에서 처리되는 데이터는 동일한 처리속도를 갖더라도, 응답속도에서 차이가 발생할 수밖에 없다. 실시간 처리를 요구하는 자율주행 등의 미션 크리티컬한(Mission Critical) 도메인에서는 이 차이가 매우 큰 영향을 미친다. 셋째, 수집되는 데이터에 포함된 사용자의 개인정보 보호 문제이다. 사용자의 개인정보가 클라우드로 전송되면 개인정보 침해에 대해 더 많은 위협에 노출되며, 개인정보 보호를 위한 많은 추가적 조치가 필요하다.

이러한 문제점을 극복하기 위해 5G 네트워크 인프라 확충, 개인정보 보호 강화 조치(GDPR, 개인정보보호법 등)가 제시되고 있으나, 이는 근본적인 해결책이 되기는 어렵다. 학습 및 추론을 클라우드에서만 수행하지 않고, 디바이스에서 직접 수행한다면 네트워크 트래픽을 감소시킬 수 있고, 클라우드까지 데이터가 왕복하는 시간을 줄여 초저지연을 실현하기 용이하며, 개인정보 및 민감정보에 대한 처리 및 보호도 가능하다. 하지만 현재의 디바이스로는 직접 학습과 추론을 수행하는 데 한계가 있어, 차세대 디바이스를 통해 이를 실현하기 위한 다양한 연구들이 수행되고 있다.

본 고에서는 차세대 디바이스와 AI의 융합기술 및 현재 산업계의 기술 현황을 소개하고 발전 방향을 제시하고자 한다. II장에서 차세대 디바이스와 AI의 융합기술에 대해 딥러닝 알고리즘 경량화, 지능형 디바이스 부품 2가지 측면에서 소개한다. 또한, III장에서 AI 융합을 위한 차세대 디바이스 산업의 국내외 현황을 소개한다. IV장에서는 결론으로 차세대 디바이스와 AI의 성공적 융합을 위한 방안을 제시한다.

## II. 차세대 디바이스와 AI 융합기술

디바이스에서 직접 학습/추론을 수행하기 위해서는 알고리즘 개선과 HW 성능 개선 등 다각적인 접근이 필요하다. 딥러닝 알고리즘은 경량화를 통해 제한된 디바이스 환경에서 동작이 가능하도록 개선되어야 하며, HW는 저전력 및 소형화를 통해서 다양한 IoT 기기의 폼팩터에서 성능을 낼 수 있어야 한다. 또한, 병렬화가 가능한 딥러닝 알고리즘의 특성을 이용한 뉴로모픽 기술을 통해서 HW 구조를 개선하고 SW 성능을 향상시키는 혁신적인 방법도 고려해야 한다.

## 1. 딥러닝 알고리즘 경량화

딥러닝이 강력한 컴퓨팅 파워를 필요로 하는 이유는 모델의 높은 복잡성(Complexity) 때문이다. 2014년에 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 2위를 차지한 VGG-19 모델은 19개의 레이어를 가지며,  $3 \times 3$  필터를 16개 가지고 있다. 채널의 개수까지 고려하면 어마어마하게 많은 웨이트를 갖게 된다. ResNet-152의 경우 레이어가 152계층에 달한다. 이처럼 복잡한 모델을 차세대 디바이스에서 프로세싱하기 위해 모델의 구조 개선이 필요한데, 이를 위해 약간의 성능저하가 발생하더라도 제한된 리소스 안에서 학습 및 추론이 가능하도록 파라미터를 줄이고 모델을 단순화하는 연구들이 진행되었다. SqueezeNet은 Fire Module을 통해 필터 크기를  $1 \times 1$ 로 줄이고, 이로 인한 정확도 손실을 보완하는 레이어를 추가했다. MobileNet은 Depthwise Convolution과 Pointwise Convolution을 구분하는 DSC(Depthwise Separable Convolution) 기법을 적용했고, ShuffleNet은 Pointwise Convolution의 추가적인 개선을 위해 Group Convolution에 셔플(Shuffle) 기법을 더했다. 각 모델별 특징은 [표 1]과 같다.

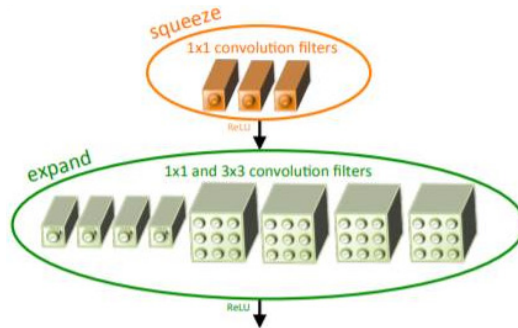
[표 1] 딥러닝 알고리즘 경량화

모델	적용 기술	설명
SqueezeNet	Fire Module	- 1/9 parameter 감소
MobileNet	DSC	- 약 1/9 parameter 감소
ShuffleNet	Group Convolution+Shuffling	- 13배 성능 향상

〈자료〉 국제학술대회 및 논문지 저자 조사 및 편집

### 가. SqueezeNet

SqueezeNet은 더 작은 모델 사이즈로 동일한 성능을 내기 위한 모델이다. 50배 작은 파라미터로 AlexNet과 비슷한 성능을 내며, 모델 사이즈는 0.5MB에 불과하다. SqueezeNet은  $3 \times 3$  필터를  $1 \times 1$  필터로 대체한 Fire Module을 사용한다([그림 1] 참조). 이를 통해 학습할 Weight 수를 1/9로 줄일 수 있다. Fire Module은 스쿼즈 레이어(Squeeze Layer)와 익스펜드 레이어(Expand Layer)의 2계층으로 구성된다. 스쿼즈 레이어는  $1 \times 1$  필터를 사용하여 입력 데이터의 채널을 하나로 합쳐서 익스펜드 레이어에 전달한다. 익스펜드 레이어는  $1 \times 1$  필터와  $3 \times 3$  필터를 사용하여 적은 파라미터로 정확도를 높인다.



〈자료〉 "SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and < 0.5MB model size"

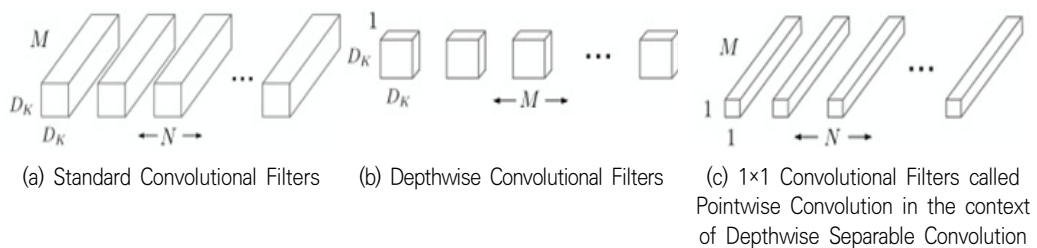
[그림 1] Fire Module

전체적으로 SqueezeNet은 8개의 fire module을 사용하여 모델을 구성한다. 입력 데이터는 콘벌루션 필터(Convolution Filter)를 거쳐 8개의 Fire Module에서 처리되고, 마지막 콘벌루션 필터를 지나 Global Avg Pool 처리된다[1].

#### 나. MobileNet

MobileNet은 파라미터를 줄이기 위해 DSC 기법을 사용한다. 먼저 각 채널별로 Depthwise Convolution을 수행하고, 이후  $1 \times 1$  필터를 이용하여 Pointwise Convolution을 수행한다([그림 2] 참조). 기존 모델이  $D_k \times D_k \times M \times D_f \times D_f \times N$ 개의 파라미터를 가진다면, MobileNet은 Depthwise Conv에서  $D_k \times D_k \times M \times D_f \times D_f \times 1 + D_f \times D_f \times M \times N$ 개의 파라미터를 갖는다. 일반적으로 커널 크기가 3이고, 채널 사이즈는 32 이상이므로, 파라미터는 약 1/8~ 1/9 가량 감소한다.

(입력:  $D_f \times D_f$ , 입력채널:  $M$ , 커널:  $D_k \times D_k$ , 출력채널:  $N$ )



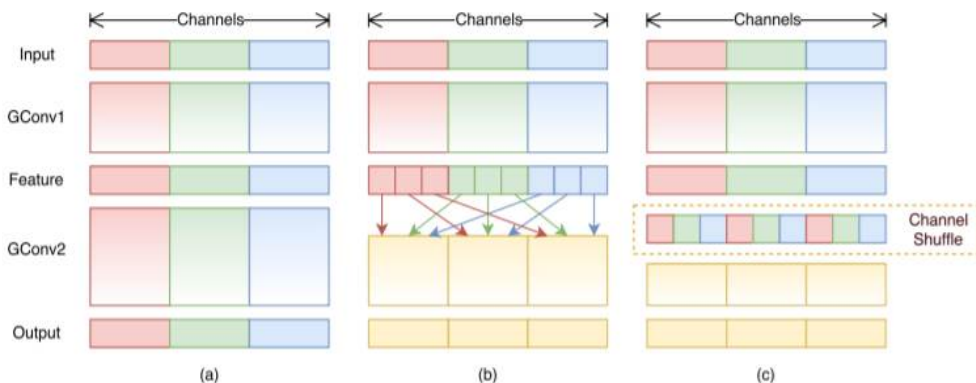
〈자료〉 "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications"

[그림 2] MobileNet의 DSC 적용

이 외에도 ResNet에서 소개된 Linear Bottleneck을 적용하여 추가적인 성능 향상을 이끌어낸 MobileNet V2 등 다양한 변형 모델들이 연구되고 있다[2],[3].

#### 다. ShuffleNet

MobileNet에서 DSC를 기반으로 파라미터를 줄였지만,  $1 \times 1$  컨벌루션은 여전히 많은 파라미터를 갖는다. ShuffleNet에서는 이 Pointwise convolution을 개선하기 위해 Group Convolution을 적용한다. Group Convolution에서는 채널을 몇 개의 그룹으로 나누어, 각 그룹별로 컨벌루션을 수행하고, 그 결과를 컨벌루션한다(그림 3 참조). 이러한 Group Convolution을 통해 파라미터를 줄일 수 있지만, 정보가 그룹 단위로만 연결되어 그룹 간에 연관성을 반영하지 못하는 문제가 발생한다. 따라서 ShuffleNet에서는  $1 \times 1$  컨벌루션을 또 다시 그룹 수만큼 나누어서 셔플한다. 이러한 과정을 통해 그룹 간에 정보를 공유하게 되어 Group Convolution의 단점을 극복하고, 적은 파라미터로 높은 성능을 낼 수 있다[4].



〈자료〉 "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices"

[그림 3] ShuffleNet 구조

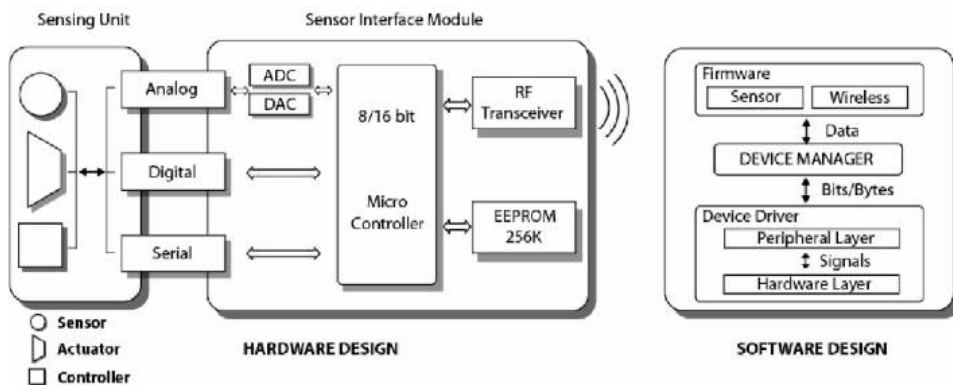
## 2. 지능형 디바이스 부품

지능형 ICT(Information & Communication Technology) 부품이란 컴퓨터 기능을 할 수 있는 전자기기에 지능화를 부여할 수 있는 기기 부품 및 기술이다. 스마트 센서 및 센서 플랫폼, 지능형 반도체, 지능형 배터리, 자립형 전원 기술 등을 포괄하고 있다[5].

차세대 디바이스가 스스로 학습하고 판단하는 자율성과 인공지능이 요구하는 대용량 데이터 처리, 집약적 알고리즘을 수행하기 위해서는 인간과 유사한 구조로 구성된 부품 및 기술이 반드시 필요하며 그 부분을 지능형 ICT 부품이 해결할 수 있을 것으로 기대된다.

#### 가. 스마트 센서 및 센서 플랫폼

측정 대상으로부터 각종 정보를 감지하여 전기적 신호로 변환시켜주는 센서에 제어, 판단, 저장, 통신 기능이 결합되어 고기능, 고부가가치를 제공하는 센서 및 시스템이다([그림 4 참조]). 1990년대에 나노 및 반도체 기술이 도입되면서 소형화되었으며 2012년 이후 MCU(Micro Controller Unit)가 센서에 내장되면서 지능화 기능을 갖춘 스마트센서로 진화되었다[6]. 구현기술로는 멀티모달 복합 센싱 기술, MEMS(Micro Electromechanical System) 기술, 반도체 SoC(System on Chip) 기술, 스마트 트랜스듀서(Smart transducer) 기술 등이 있다.



〈자료〉 Wireless Industrial Monitoring and Control Using a Smart Sensor Platform

[그림 4] 스마트센서 시스템 아키텍처

#### 나. 지능형 반도체

인식·추론·학습·판단 등 인공지능 서비스에 최적화(지능화, 저전력, 안정화)된 소프트웨어와 시스템반도체가 융합된 반도체이다[7]. [표 2]는 지능형 반도체 기술의 유형별 세부 기술이다. 수많은 신경세포로 이루어져 모든 정보가 모여 처리되는 인간의 뇌와 같은 역할을 하는 지능형 반도체는 그 명성에 걸맞게 소자부터 설계, 공정, 장비 등 ICT 산업 전 부분의 복합적인 기술 개발이 필요하다. 수천 수만 개의 코어를 집적하여 컴퓨팅 파워

[표 2] 지능형 반도체 기술

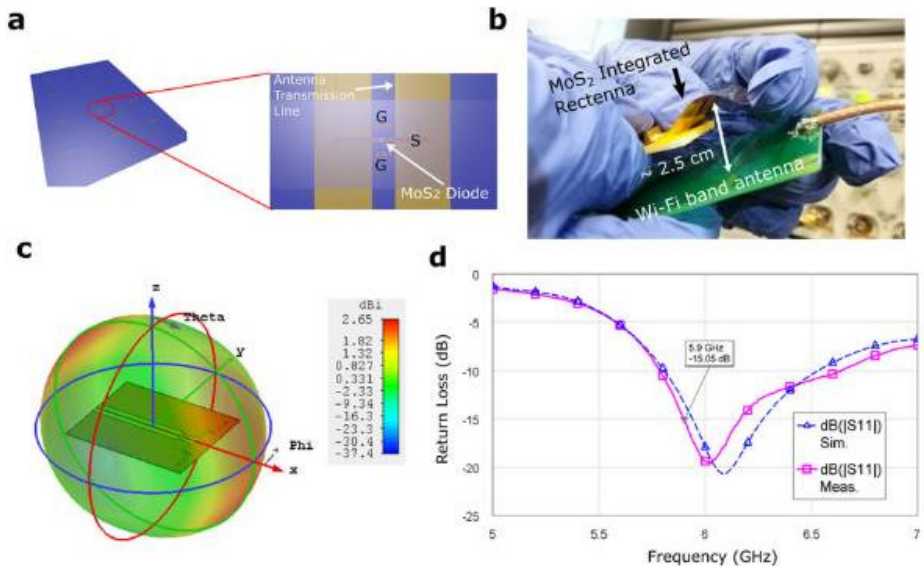
중분류	소분류	세부 기술
지능형 반도체	지능화 기술	초병렬 매니코어 프로세서
		딥러닝 뉴럴넷 코어
		초고속 온칩 프로세서-메모리
		재구성형 인공지능 회로설계
	저전력 기술	저전력 프로세서 아키텍처
		저전력 차세대 메모리 아키텍처
		저전력 무선통신 설계
		저전력 회로IP 설계
		저전력 소자 기술
	고신뢰 기술	기능안정성 설계
		정보보안 반도체 기술
		고신뢰 소자 설계
		차세대 제조 공정 장비
	프로세싱 SW	컴파일러 및 프로그래밍 언어
		프로세싱 API
		시스템 SW

〈자료〉 한국과학기술기획평가원, “차세대지능형반도체 기술개발사업 예비타당성조사 보고서”, 2019. 6.

를 제공하는 지능화 기술, 사용자의 지속성과 낮은 유지비용을 위한 저전력 기술, 제품의 기능 안정성 및 정보보호를 위한 고신뢰 기술 등이 있다.

#### 다. 자립형 전원

충전이 필요한 배터리나 전력망 없이 스스로 전원을 공급하여 차세대 디바이스의 지속적인 사용을 가능하게 하는 에너지 공급 기술이다. 대표적인 자립형 전원 기술로는 에너지 하베스팅(Energy Harvesting) 기술이 있다. 에너지 하베스팅 기술은 태양에너지, 인체 운동에너지 등 다양한 형태의 주변 에너지를 수집하는 기술이다. 대표적으로 기계적 외력이 가해지면 전압이 발생하는 압전소자를 이용한 압전 에너지 하베스팅, 마찰 전기의 극성이 다른 두 개의 물질 사이의 접촉으로 표면 전하의 이동이 발생하는 것을 이용한 마찰력 에너지 하베스팅이 있다. 최근에는 무선 통신의 발전에 따라 [그림 5]처럼 일상생활에서 노출되어 있는 전파(Wifi 등)를 활용한 전파 에너지 하베스팅과 같은 기술도 개발되었다.



\* Wi-Fi 대역(5.9GHz)에서 전자기 방사 에너지를 수집하는 완전 유연한 MoS<sub>2</sub> 안테나(네이처지 캡처)  
 <자료> Two-dimensional MoS<sub>2</sub>-enabled flexible rectenna for Wi-Fi-band wireless energy harvesting

[그림 5] Wi-Fi 대역(5.9GHz)에서 전자기 방사 에너지를 수집하는 MoS<sub>2</sub> 렉테나

### III. AI 융합을 위한 차세대 디바이스 산업 현황

#### 1. 국내 산업 현황

국내에서는 삼성전자, 네패스 등의 기업이 On-Device 학습 및 추론을 위한 AI 프로세서 기술을 선도하고 있고, ETRI, 한국뇌연구원 등은 차세대 디바이스를 통해 사용자에게 인터페이스를 제공하기 위한 인터랙션 기술을 개발하고 있다([표 3] 참조).

[표 3] AI 융합을 위한 국내 차세대 디바이스 기술 현황

구분	제조사	설명
AI 프로세서	삼성전자	- Exynos 9820 - AP에 NPU 탑재 - 10배 인력 확충, IT 전 분야 확대
	네패스	- NM500 - NPU 세계 최초 상용화 - Quick AI 플랫폼에 탑재, 분석도구 제공

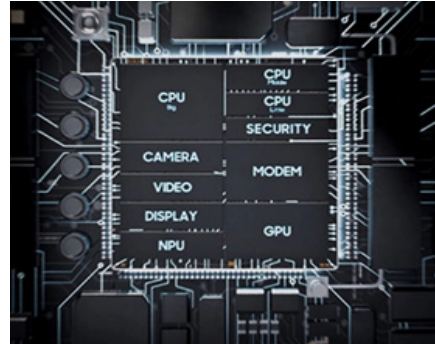


구분	제조사	설명
HCI	ETRI	- 모션 융합 기반의 정적 및 동적 손 동작 인식 기술 - 기술이전 설명회 개최
	한국뇌연구원	- BCI 기술 기반 - 뇌파 드론 체험

〈자료〉 국제학술대회 및 논문지 저자 조사 및 편집

#### 가. 삼성전자

SoC에 NPU(Neural Processing Unit)를 탑재한 AP(Application Processor)인 Exynos 9(9820)를 출시했다(그림 6) 참조). NPU는 병렬컴퓨팅을 통해 수천 개 이상의 연산을 동시에 효율적으로 처리할 수 있어, 딥러닝 알고리즘 수행에 최적화된 핵심 기술이다. NPU를 탑재한 AP를 이용하여 서버를 거치지 않고, 모바일 기기에서 자체적으로 On-Device AI를 구현할 수 있게 되었다. 향후 삼성전자는 모바일 기기에 그치지 않고, 전장 분야의 IVI(In-Vehicle Infotainment)와 ADAS(Advanced Driver Assistance System), 데이터센터, IoT 등 IT 전 분야로 확대할 계획이며, NPU 분야 인력을 2,000명 규모로 10배 이상 확대하고 전사적 역량을 집중한다는 전략이다[8].



〈자료〉 <https://www.gsmarena.com>

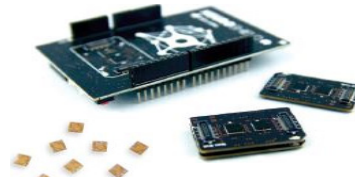
[그림 6] NPU가 탑재된 삼성의 Exynos 9

#### 나. 네패스

단독 보유중인 PLP(Panel Level Package) 기술을 기반으로 최초로 상용 NPU인 NM500을 출시했다(그림 7) 참조). NM500은 반도체 칩에 576개의 인공뉴런을 집적하



QuickAI HDK



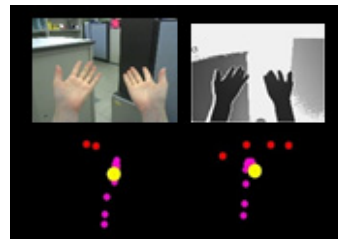
네패스 NM500

[그림 7] 네패스 NM500

여 고속/병렬연산을 처리할 수 있다. 이외에도 에너지 효율적인 소형 폼팩터 구성요소로 NeuroMem 기술을 구현하였으며, 자체 개발한 SW 분석도구인 Knowledge Studio도 제공하고 있다. NM500은 미국 QuickLogic사에서 발표한 인공지능 플랫폼인 QuickAI에 탑재될 예정이다[9],[10].

#### 다. ETRI

모션 융합 기반의 정적 및 동적 손동작 인식 기술을 개발하여 라이브러리 형태로 제공하고 있으며, 가상 객체와 상호작용을 위한 인터랙티브 콘텐츠도 바이너리 형태로 함께 제공하고 있다([그림 8] 참조). 이를 통해 인공지능이 결합된 가상현실 콘텐츠에서 다양한 상호작용을 지원하며, 이에 대한 기술이전설명회를 개최하는 등 사업화도 지원하고 있다[11].



〈자료〉 ETRI

[그림 8] 모션 융합 기반의 정적 및 동적 손 동작 인식 기술

#### 라. 한국뇌연구원

BCI(Brain-Computer Interface) 기술을 기반으로, 인간의 두뇌를 컴퓨터와 직접 연결하여 뇌파만으로 드론을 조종할 수 있는 “뇌파 드론”을 공개하고 체험 기회를 제공하였다[12].

#### 마. 정부

1조 원 예산 규모의 지능형반도체 연구개발(R&D) 사업의 예비타당성조사를 통과시킴으로써 핵심 기초·원천기술 개발을 추진할 계획이다[13]. 센서산업 클러스터 조성을 통해 센서 기술 경쟁력을 확보하고, 단일 제품 중심이 아닌 복합센서, 연계 솔루션 개발 등의 차별화를 위해 2019년 2월 대전시에서는 스마트센서 클러스터를 구축했다[14].

## 2. 해외 산업 현황

AI융합산업을 주도적으로 이끌고 있는 미국을 중심으로 각 국가의 기업 및 선도기관들

업체명	데이터센터 서버용				입력형	에지 디바이스용			
	GPU	FPGA	ASIC	뉴로 모픽		GPU	FPGA	ASIC	뉴로 모픽
반도체	intel		○	△				●	
	NVIDIA	●				●			
	AMD	○							
	XILINX		●						
	QUALCOMM				○				
IT	Google		●					○	
	Microsoft							○	
	amazon.com							△	
	IBM								●
	Apple				●				
기타	Hewlett Packard			△					
	TESLA							△	

●: 선도자, ○: 시장 추격자, △: 제품 개발 중

〈자료〉 과학기술&ICT 정책 기술동향, 과학기술정보통신부 및 한국과학기술기획평가원, 제 116호, 2018, p.6.

[그림 9] 미국 주요 대기업의 지능형 반도체 개발 현황

이 치열한 경쟁을 통해 AI 융합 차세대 디바이스 산업을 발전시키고 있다. 초기에는 인텔과 엔비디아와 같은 제조기업들이 시장을 선점하고 있었으나 제조 사업을 영위하지 않는 기업들도 제품 및 서비스 경쟁력 강화를 위해 자체 연구와 개발을 진행하고 있어 경쟁이 치열해지는 상황이다([그림 9] 참조)[15].

## 가. NVIDIA

GPGPU(General-Purpose computing on Graphics Processing Units)로 지능형 반도체 선두주자 역할을 하고 있다. 고속 GPU 상호연결용 NVLINK와 데이터센터용 볼타(Volta), 75W의 소형 PCIe 폼팩터로 제공되는 Tesla T4 GPU 등으로 AI 기술이 적용되는 산업 전반에 영향을 미치고 있다. 2019년 초 데이터센터용 서버와 스토리지, 하이퍼 컨버전스 인프라용 고속 네트워크 칩 등에서 높은 기술력을 보유하고 있는 이스라엘 기반 네트워크 칩 전문기업인 Mellanox



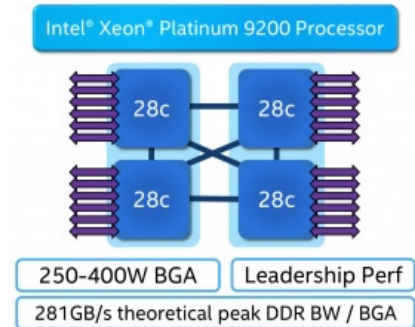
Tesla P4	
GPU	1 NVIDIA Pascal™ GPU
NVIDIA® CUDA® Cores	2,560
Memory Size	8 GB GDDR5
H.264 1080p30 Streams	24
vGPU Profiles	1 GB, 2 GB, 4 GB, 8 GB
Form Factor	PCIe 3.0 single slot (low profile) for rack servers
Power	75 W
Thermal Solution	Passive

[그림 10] nVidia Tesla P4

의 인수 계획을 발표하는 등 적극적인 M&A 전략을 펼치고 있다[16].

#### 나. Intel

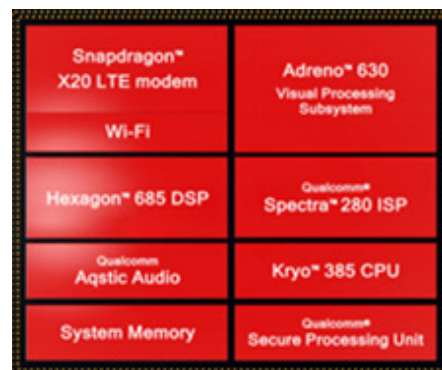
적극적으로 반도체 인수합병을 추진하고 다양한 시스템 반도체 제품 포트폴리오를 보유하고 있었으나 최근 NVIDIA에게 지능형 반도체 산업 경쟁에서는 밀리는 추세이다. 가장 대표적인 제품군은 Xeon 시리즈이다. 2019년 56개 코어와 12개 메모리 채널을 탑재한 “인텔 제온 플래티넘 9200” 프로세서를 포함하여 50종 이상의 제온 프로세서를 선보였다([그림 11] 참조)[17].



[그림 11] 인텔 제온 플래티넘 9200

#### 다. Qualcomm

2013년 자사의 Spiking Neural Networks (SNNs)를 모방한 뇌처럼 학습하는 프로세서 (Neural Processing Unit: NPU)인 Zeroth를 공개하였으며[18], 현재는 Snapdragon의 AI로 통일되어 4세대 멀티코어 웰컴 AI 엔진이 탑재된 Qualcomm Snapdragon 855+ Mobile Platform을 출시하는 등 AI 기술 집약을 꾀하고 있다([그림 12] 참조)[19].

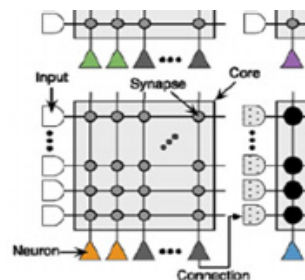


(자료) www.anandtech.com

[그림 12] Snapdragon 845 Block Diagram

#### 라. IBM

DARPA SyNAPSE 프로젝트를 통해 포유류의 뇌를 모방한 전자두뇌시스템 개발을 추진하여 SRAM 어레이를 시냅스 모방 소자로 활용한 트루노스(TrueNorth) 칩을 개발하였다([그림 13] 참조)[20].

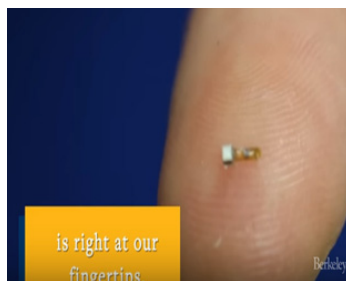


(자료) A Noise Filtering Algorithm for Event-Based Asynchronous Change Detection Image Sensors on TrueNorth and Its Implementation on TrueNorth

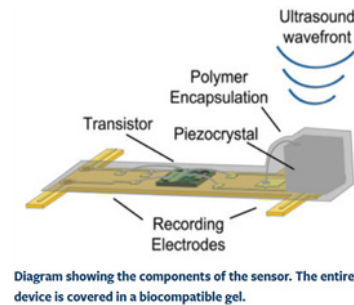
[그림 13] IBM TrueNorth Core Architecture

마. University of California, Berkeley

UC Berkeley 연구진은 인트라바디 나노 커넥터를 위한 모래 알갱이 크기의 인체 삽입용 센서인 Neural Dust를 개발하였다([그림 14] 참조)[21]. 인트라바디 나노 커넥터 기술이란 인체 내 신경계 및 내분비계를 활용하여 신호를 전달 및 통신하는 인체 친화적 인터페이스 기술이다([그림 15] 참조). 인체 내부 신경, 근육, 기관 등을 실시간으로 모니터링할 수 있으며 장애를 치료하고 면역계를 자극하거나 신경 및 근육을 자극할 수 있다.



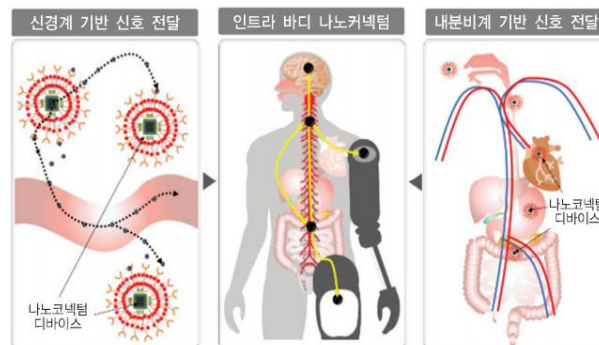
Neural Dust Sensor 외형



Neural Dust Sensor 구성도

〈자료〉 Berkeley News Youtube

[그림 14] Neural Dust Sensor의 외형 및 구성도



〈자료〉 김혜진 외 3명, 신경 인터페이스 기반 초감각 디바이스 기술 동향, 한국전자통신연구원, 전자통신동향분석 제33권 제 6호, 2018 12, p76

[그림 15] 인트라바디 나노 커넥터 개념도

## IV. 결론

차세대 디바이스는 시간이 지날수록 더욱 AI 기술과의 융합을 가속화할 것이다. On-

Device AI 환경을 갖추고, 다양한 HCI(Human-Computer Interaction) 기술을 기반으로 사용자에게 인공지능 서비스를 제공하는 창구를 늘여갈 것이다. AI와 차세대 디바이스의 융합을 선도하기 위해서는 정부, 기업, 학계의 전방위적인 협력이 필요하다.

정부는 다양한 AI 기반의 국가 프로젝트를 주도하고, 중장기 로드맵에 기반한 투자지원을 통해 AI 기술력을 확보하고 전문인력을 양성하여 연구기반을 조성해야 한다. 기업은 이미 보유하고 있는 강력한 반도체 인프라와 지능형 반도체 설비기술을 더욱 고도화하고, 메모리 분야를 넘어서 시스템 칩 기술 분야도 선도하기 위한 연구 개발에 나서야 한다. 학계에서는 소수의 연구소 단위의 개별적 연구환경을 벗어나 융합연구를 통해 경쟁력을 제고하고, 인공지능 핵심 SW 기술에 대한 기반기술 연구를 통해 국제 경쟁력을 갖추도록 해야 할 것이다.

#### [ 참고문헌 ]

- [1] F.N. Iandola et al., "SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and < 0.5MB model size," arXiv:1602.07360, 2016.
- [2] A.G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 2017.
- [3] M. Sandler et al., "MobileNet V2: Inverted Residuals and Linear Bottlenecks," arXiv:1801.04381, 2018.
- [4] X. Zhang et al., "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," arXiv:1707.01083, 2017.
- [5] 임태범 외 17명, ICT 디바이스 지능형 디바이스, 한국정보통신기술협회(TTA), ICT 표준화 전략맵 종합보고서 4, 2017-0-00059 ICT 표준화 체계, 2017, p.417.
- [6] 이미혜, 센서산업 현황 및 경쟁력, 한국수출입은행, 2018 ISSUE STUDY, VOL.2018-중점-05 (2019.1.23), 2019, p.1.
- [7] 김정언 외 3명, 지능형반도체 기술개발을 위한 기획연구, 정보통신정책연구원, 정책연구 18-29, 2018, p.20.
- [8] 인공지능신문, "삼성전자, 차세대 SoC 적용, 업계 최고 성능의 On-Device AI 구현한다", 2019. 6. 18.
- [9] 지디넷, "네패스 인공지능 칩, 美 쿼로직 '퀵AI'에 탑재", 2018. 5. 8.
- [10] Global Vision, "NM500 neuromorphic chip is available," 2017. 12. 21.
- [11] 김기홍, "모션 융합 기반의 정적 및 동적 손 동작 인식 기술", ETRI, 2019. 2.
- [12] 카이스트 신문, "BCI, 생각만으로 기계를 조종하다", 2019. 5. 14.
- [13] 한국과학기술기획평가원, "차세대지능형반도체 기술개발사업", 2019. 6.
- [14] 전자신문, "대전시, 스마트센서 산업 키운다", 2019. 7.

- [15] 과학기술정보통신부 및 한국과학기술기획평가원, 과학기술&ICT 정책 기술동향, 제 116호, 2018, p.5.
- [16] ZDNet Korea, 엔비디아 “연내 멜라노스 인수 완료, AI 기술력 높일 것”, 2019. 7. 4.
- [17] 전자신문, “인텔, 2세대 데이터센터 CPU로 ‘데이터센터릭’ 시대 겨냥”, 2019. 6.
- [18] iTNEWS, “퀄컴의 엣지인공지능칩(Edge AI Chip)-NPU 개발 동향”, 2019. 6.
- [19] KIPOST, “퀄컴, 5G 게이밍 스마트폰용 애플리케이션프로세서(AP) 내놔”, 2019. 7.
- [20] 신재호 외 2명, “차세대지능형반도체 기술개발사업”, 한국과학기술기획평가원, 2018년도 예비타당성 조사 보고서, 2019. 6, p.109.
- [21] iTNEWS, “BCI/BMI용 먼지 크기 인체 이식형 무선센서”, 2018. 8.
- [22] 김혜진 외 3명, “신경 인터페이스 기반 초감각 디바이스 기술 동향”, 한국전자통신연구원, 전자통신 동향분석 제33권 제 6호, 2018. 12.