

뉴로모픽 포토닉스 기술 동향

Trends in Neuromorphic Photonics Technology

권용환 (Y.H. Kwon, yhkwon@etri.re.kr)
 김기수 (K.S. Kim, kimks1136@etri.re.kr)
 백용순 (Y.S. Baek, yongb@etri.re.kr)

광융합부품연구실 책임연구원
 광융합부품연구실 책임연구원/실장
 광무선원천연구본부 책임연구원/본부장

ABSTRACT

The existing Von Neumann architecture places limits to data processing in AI, a booming technology. To address this issue, research is being conducted on computing architectures and artificial neural networks that simulate neurons and synapses, which are the hardware of the human brain. With high-speed, high-throughput data communication infrastructures, photonic solutions today are a mature industrial reality. In particular, due to the recent outstanding achievements of artificial neural networks, there is considerable interest in improving their speed and energy efficiency by exploiting photonic-based neuromorphic hardware instead of electronic-based hardware. This paper covers recent photonic neuromorphic studies and a classification of existing solutions (categorized into multilayer perceptrons, convolutional neural networks, spiking neural networks, and reservoir computing).

KEYWORDS 뉴로모픽, 인공지능, 포토닉스, 마이크로 링 공진기

1. 서론

최근 들어 인공지능 기술은 상당히 많은 수의 응용 분야, 즉 이미지 분류, 음성 인식 및 언어 번역, 의사 결정, 웹 검색, 소셜 네트워크의 콘텐츠 필터링, 전자 상거래 웹 사이트의 권장 사항 등에 활용되고 있다[1]. 인공 신경망(ANN: Artificial Neural

Network)은 이러한 대규모의 데이터를 처리하고 대량의 정보를 신속하게 결합 및 분석하는 데 유용하며, 이를 위해 여러 가지 신경망 아키텍처 및 이를 구현하기 위한 하드웨어 기술이 연구되고 있다.

빅데이터 분석에 대한 현재 접근 방식은 소프트웨어를 기반으로 해서 직렬화된 중앙 집중식 폰 노이만 머신에서 실행되고 있으며, 최근 인공지능의

* DOI: <https://doi.org/10.22648/ETRI.2020.J.350404>

* 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 하였음[2018-0-00219, 직독식 수질복합센서 및 초분광영상 기반 시공간 복합 인공지능 녹조 예측 기술].



급격한 발전에는 GPU(Graphic Processing Unit)의 성능 개선이 크게 기여했다. 이러한 기존의 컴퓨팅 방식은 무어의 법칙으로 널리 알려진 디지털 전자 회로의 기하급수적인 확장으로 인해 지속되고 있다. 지난 60년 동안 마이크로프로세서의 트랜지스터 밀도, 클럭 속도 및 전력 효율은 18개월마다 약 두 배로 증가하였다. 그러나 최근 10년간 마이크로프로세서의 에너지 효율은 정점을 보여주고 있고, 보다 더 많은 연산량을 요구하는 인공지능 데이터 처리를 위한 하드웨어 확보를 위해 다음과 같이 연구가 진행되고 있다[2].

첫 번째는 기존의 폰 노이만 구조하에서 성능을 개선하는 것으로 다양한 종류의 인공지능 가속기가 개발되고 있으며, 이는 기존 프로세서 내부 구조의 병렬화 증대, 프로세서 자체 메모리 증대, 같은 데이터 재사용, 계산 정확성의 다소 희생 등의 관점에서 진행되고 있다.

구글의 딥러닝 특화 반도체인 TPU(Tensor Processing Unit)의 경우 연산 정밀도를 희생하여 32비트가 아닌 8/16비트의 연산기를 기본으로 하고 있으며, 연산 결과를 연산 회로 내에서 직접 전달하여 메모리의 읽기를 대폭 줄이고 동시에 전력 소비를 줄이는 대규모 행렬 연산 파이프라인을 실현하였다.

다음으로는 인간 뇌의 하드웨어인 뉴런과 시냅스를 모사하는 컴퓨팅 아키텍처와 인공 신경망에 대한 연구가 이루어지고 있으며, 이를 구현하는 하드웨어에 대한 새로운 시도가 전 세계적으로 산·학·연 협력으로 진행 중이다.

현재 인간 뇌를 모사하는 뉴로모픽 시스템을 하드웨어로 구현하는 기술은 기존 실리콘 트랜지스터를 활용하여 뉴런을 모사하는 방법과 멤리스터 등 신개념의 뉴로모픽 소자를 활용하여 구현하는 방법으로 나눌 수 있다.

기존 실리콘 트랜지스터를 활용하는 최근의 사례로는 4,096개의 코어(1코어=256뉴런+256시냅스)가 집적되고 기존 마이크로프로세서의 1만분의 1의 전력을 소모하는 IBM의 TrueNorth 칩, 128개의 코어(1코어=1,024뉴런+4,096시냅스)가 집적되어 지도학습 대신 실시간으로 유입되는 정보를 받아들여 스스로 학습하는 Intel의 Loihi 칩 등이 있다. 하지만 복잡한 회로 구조 때문에 칩 면적이 크며, 전력 소모가 여전히 많고, CMOS 스케일링 문제는 동일하게 가지고 있어 이러한 기존 실리콘 트랜지스터 기반의 뉴로모픽 시스템은 미래 인공지능을 위한 궁극적인 해결책은 아닐 가능성이 높다.

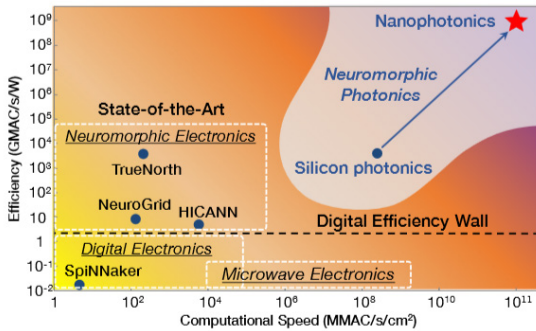
멤리스터 등 신개념 소자는 다양한 메커니즘 기반의 비휘발성/휘발성의 저항 변화 특성을 지니고 있으며, 여러 단계의 시냅스 강도를 표현할 수 있는 소자들로 금속 이온 이동 기반 소자, 산소정공 이동 기반 소자, 상 변화 기반 소자 등 최근 다양한 물질과 구조들을 활용한 연구결과가 발표되고 있다[3].

하지만 새로운 뉴로모픽 소자들은 아직 제대로 된 신경망 시스템 동작을 지원하기에는 많은 기술적 난관들을 가지고 있으며, 향후 확장 가능하고 균일한 특성을 가진 소자 성능 확보와 더불어 복잡한 생물학적 역학과정 이해, 신경망의 구조 및 동작원리에 대해 많은 연구가 필요하다.

II. 뉴로모픽 포토닉스 기술 개념

포토닉스 기술 기반으로 뉴로모픽 시스템을 구성하고자 하는 연구가 최근 급증하고 있다. 뉴런의 핵심 기능인 가중치(Weight)는 필터와 감쇄기(Attenuator) 등으로 구현하고, 활성화 기능(Activation Function)은 기본적인 광의 비선형성을 이용한 다양한 방향의 연구가 진행 중이다[4].

최근 몇 년 동안 폰 노이만 방식의 마이크로프로



출처 By T. F. D. Lima et al. (Licensed under CC BY-NC-ND).

그림 1 뉴로모픽 하드웨어 플랫폼 비교

세서의 이슈는 프로세서와 메모리 간 데이터 전송에 있어서의 전력 소모 및 대역폭의 문제였다. 이를 해결하기 위해 디지털 컴퓨터의 특정 부품을 대체 및 성능을 향상시키기 위한 아날로그 전자 회로가 등장하였다. 앞 절에서 소개하였던 다양한 멤리스터 소자는 크로스바 어레이 형태를 통해 스파이킹 신경망(SNN: Spiking Neural Network)을 구성하여 인간의 뉴런을 보다 더 잘 모사하는 새로운 가능성의 단초를 보여주고 있다. 하지만 멤리스터의 주요 단점으로 저항 기반 동작의 고유한 전력 손실, 어레이의 IR 강하, 재현성 및 확장성 확보의 어려움, 정확한 시뮬레이션 통과 공정 표준의 부재 등이 제시되고 있다.

그림 1과 같이 뉴로모픽 소자의 컴퓨팅 속도와 에너지 효율의 관계를 보면 기존 실리콘 트랜지스터를 사용하여 뉴런을 모사하는 경우에 폰 노이만 방식의 디지털 회로에서 예측하는 에너지 효율 벽 (Energy efficiency wall: ~ 10 GMAC/s/W 또는 100 pJ/MAC, 여기서 MAC은 Multiply-accumulate operation)을 뛰어넘는 좋은 결과를 보여주고 있다[4]. 포토닉스는 기존 전자 솔루션과 비교하여 ANN에서 훈련 및 추론 작업을 진행할 때 계산 속도 및 에너지 효율 측면에서 각각 $\sim 10^3$ 배 이상의 개선 가능성을 기대하고 있다. 계산 속도의 개선은 고유한 넓은 대역폭에 기인하며, 에너지 효율의 경우 신경망을 구현하는 행렬 곱에서 주로 에너지를 소모하는 부분인 시냅스(Synaptic Computations)의 인터커넥트 구동전력이 CMOS에 비해 포토닉스에서는 매우 적은 전력이 소모되는 것이 원인으로 제시된다.

Shastri 등은 표 1에서와 같이 여러 변수에 대하여 전자 및 포토닉스 솔루션의 성능을 비교하였다[4]. 기존 전자 솔루션의 경우 속도 측면에서는 HICANN 칩이 가장 우수한데, 이는 많은 전력 소모를 감수하면서 빠른 속도로 구동하는 것을 목표로 했기 때문이며, 다른 전자 솔루션(TrueNorth, Neurogrid, SpiNNaker)의 경우 이의 1/1,000 정도의 MAC 속도를 보여주고 있다. 포토닉스 솔루션

표 1 전자 및 광 뉴로모픽 소자의 특성 비교

| Chip | MAC rate/processor | Energy/MAC(pJ) | Processor fanin | Area/MAC(μm^2) | Synapse precision(bit) |
|---|--------------------|-----------------|-----------------|-----------------------------|------------------------|
| Photonic Hybrid III-V/Si | 20GHz | 0,26 | 108 | 205 | 5,1 |
| Sub- λ Photonics (future trend) | 200GHz | 0,0007 | ~ 200 | 20 | 8 |
| HICANN | 22,4MHz | 198,4 | 224 | 780 | 4 |
| TrueNorth | 2,5kHz | 0,27 | 256 | 4,9 | 5 |
| Neurogrid | 40,1kHz | 119 | 4,096 | 7,1 | 13 |
| SpiNNaker | 3,2kHz | 6×10^5 | 320 | 217 | 16 |

출처 Reprinted from T. F. D. Lima et al., Progress in neuromorphic photonics, Nanophotonics 2017; 6(3): 577-599, CC BY-NC-ND

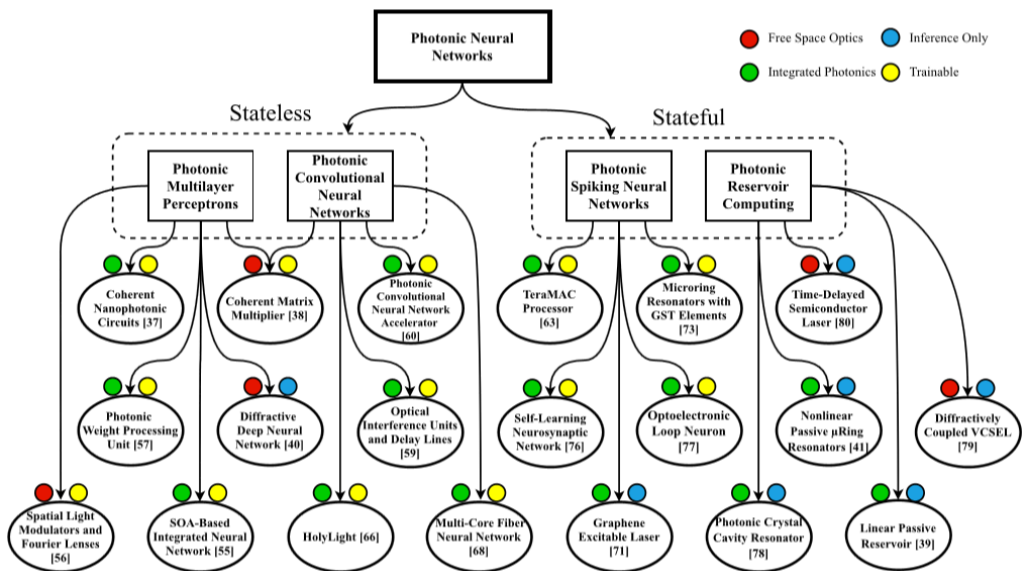
의 경우 높은 대역폭과 낮은 대기 시간의 장점으로 HICANN에 비해서도 $\sim 10^3$ 배 정도, 나머지 전자 솔루션에 대해서는 $\sim 10^6$ 배 정도의 매우 빠른 컴퓨팅 속도 개선을 기대할 수 있다. 포토닉스 솔루션에서 시냅스 기능으로 주로 활용되는 수동 필터는 누설 전류가 낮고 전력 소모가 거의 없으며, 전력은 주로 레이저에 의해 정적으로 소비되기 때문에 기존 클럭 속도에 따라 전력 소비가 동적으로 증가하고 인터커넥트 전력 소모가 과다한 전자 솔루션에 비해 높은 에너지 효율을 달성할 수 있다. 현재 수준으로도 매우 낮은 전력소모를 목표로 개발된 TrueNorth 칩과 비슷한 수준의 에너지 효율 ($\sim 0.2\text{pJ}/\text{MAC}$)을 확보하고 있으며, 다파장 구조를 활용할 때 향후 1,000배 이상의 에너지 효율성 확보가 기대된다. MAC당 면적 지수는 포토닉 소자의 경우 광의 회절 한계를 넘어서 축소할 수 없기 때문에 중요한 비교 지표이다. 가장 작은 MAC당 면적지수를 가지는 TrueNorth에 비해서 40배 정도

큰 상황이며 다파장 구조를 활용할 때 이를 1/10 정도 크기로 줄여서 대부분의 전자 솔루션과 동일한 정도의 사이즈를 가질 것으로 생각된다.

포토닉스 소자를 뉴로모픽 시스템에 적용하는 경우 이와 같이 에너지 효율성, 대역폭 및 대기 시간 측면에서 궁극적으로 매우 유리한 장점이 기대되어 학계 및 산업계에서 활발하게 연구가 진행되고 있다. 실제 산업적인 측면에서도 최근의 포토닉스 분야 중 실리콘 포토닉스 분야는 오랫동안의 실리콘 파운드리 기술에 의해 확보한 신뢰성과 비용 효율성을 모두 갖춘 대량 생산 플랫폼을 구축하여 새로운 뉴로모픽 시스템 동작을 지원하는 유력한 대안으로 제시되고 있다.

III. 뉴로모픽 포토닉스 기술 분류

최근 뉴로모픽 포토닉스 분야에서 진행된 다양한 연구들을 그림 2와 같이 분류하였다. 먼저 다층



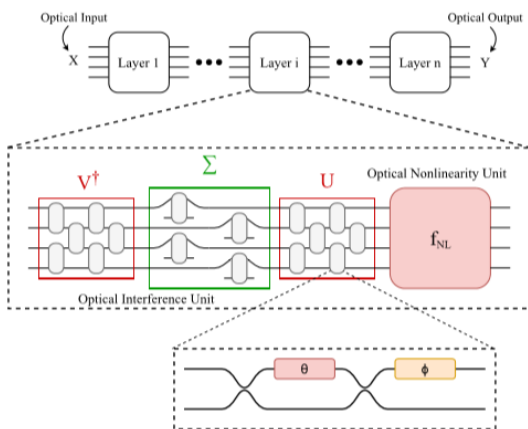
출처 By L. D. Marinis et al. (Licensed under a Creative Commons Attribution 4.0 License).

그림 2 뉴로모픽 포토닉스 연구 분류(광 다층 신경망, 컨볼루션 신경망, 스파이킹 신경망, 축적 컴퓨팅)

신경망(Multilayer Perceptrons) 및 컨볼루션 신경망(CNN: Convolutional Neural Networks)과 같이 무상태(stateless, i.e., without memory) 신경망과 스파이킹 신경망과 축적 컴퓨팅(Reservoir Computing)과 같은 상태를 저장하는(Stateful) 신경망 용도로 나누어 볼 수 있다[1]. 그리고 더 세부적으로는 자유 공간 광학과 집적 광학 소자를 이용하는 구조와 추론만 진행되는 경우와 훈련과 추론을 동시에 진행하는 연구로 분류해 보았다.

1. 광 다층 신경망

광 다층 신경망 구조는 수 개의 뉴런 층으로 구성되며, 각 뉴런은 이전 층과 다음 층의 모든 뉴런과 완전히 연결되어 있다. 첫 번째 레이어는 입력 신호를 수신하는 입력 레이어이며, 최종 레이어는 추론 작업의 결과를 제공하는 출력 레이어이다. 일반적으로 몇 개의 숨겨진 층이 입력과 출력 층 사이에 배치되며, 각 층에서 정보는 시냅스 가중치와 이전 층의 결과의 선형 조합(즉, 행렬 곱셈)을 통해



출처 By L. D. Marinis et al. (Licensed under a Creative Commons Attribution 4.0 License).

그림 3 마하 켄더 간섭계로 구성된 광학 간섭 유닛 및 비선형 유닛 개념도

전과된 후 각 뉴런이 비선형 기능을 수행한다.

한 예로서 코히어런트 광을 처리하는 나노 광자 회로에 기초한 DNN(Deep Neural Network)이 제안되었다[5]. 그림 3과 같이 본 구조의 계층은 광학 매트릭스 곱셈을 수행하는 광학 간섭 유닛(OIU: Optical Interference Unit)과 비선형 활성화 기능을 구현하는 광학 비선형 유닛(ONU: Optical Nonlinear Unit)의 두 부분으로 구성된다. OIU는 56개의 프로그래밍 가능한 마하 켄더 간섭계(MZI: Mach-Zehnder Interferometers)로 구성된 일련의 배열로 이루어지며, 이때 MZI는 노드에서 신호의 분기비를 조정할 수 있다. 온칩 교육 절차가 제안되었지만 실험적인 학습 훈련은 실제로 컴퓨터에서 수행되었고, MNIST 숫자 인식 작업에서 95%의 정확도를 보여주었다. 이 MZI 기반 접근법에 대한 최신 연구 중 하나로 광 신경망에서 모드 믹싱 및 광학 비선형성과 같은 양자 광학 기능을 활용하는 양자 광 신경망에 대한 제안이 있었다.

2. 컨볼루션 신경망

컨볼루션 신경망은 다차원 배열의 형태로 설계되어 주로 이미지와 비디오 데이터를 처리하는 데 사용된다. 첫 번째 단계는 컨볼루션 계층으로 공간 필터링의 기능을 수행하는데 이전 레이어의 주요 특징을 추출하고, 다음 단계인 풀링 계층에서는 유사한 특징을 병합하여 연산량을 줄여주는 역할을 진행한다. 컨볼루션 계층들은 각 뉴런은 이전 레이어의 소수의 뉴런에만 연결이 되는, 즉 성긴(Sparse) 토폴로지를 통해 연결되고, 두 컨볼루션 레이어 사이에는 일반적으로 정류 선형 유닛(ReLU: Rectified Linear Unit)이 포함되어 있다.

A. Mehrabian 등은 마이크로 링 공진기(MRR: Micro-Ring Resonator) 기반으로 컨볼루션 신경망

을 위한 광 가속기를 제안하였다[6]. MRR은 소자 크기가 작은 장점을 가지고 있으며, 노드에 위치하여 MAC 연산을 수행하고 BW(Broadcast-and-Weight) 프로토콜을 구현하도록 설계되었다. BW 프로토콜은 파장분할다중화(WDM: Wavelength Division Multiplexing) 방법을 사용하여 광 뉴런 프로세서 간에 확장 가능한(Scalable) 상호 연결을 제공할 수 있어 향후 큰 배열의 제작도 가능할 전망이다. 다중 파장을 사용하는 BW 프로토콜은 원칙적으로 신경망 유형과 무관하여 다층 및 스파이킹 신경망에서도 이용 가능하며, 뉴런 간에 대규모 병렬 통신을 가능하게 한다. 가중치는 노드의 드롭 계수를 0%에서 100%로 조정함으로써 변경할 수 있다. 본 연구는 전자 솔루션에 비해 $\sim 10^3$ 배 정도 빠른 속도를, 그리고 에너지는 0.75배 소비되는 것을 보고하였다.

3. 스파이킹 신경망

스파이킹 신경망에서 정보는 이벤트 또는 스파이크로 인코딩되며, 스파이크가 발생하는 시간은 아날로그 속성을, 그리고 이때 진폭은 디지털 속성을 가지는 하이브리드 인코딩 형태이다. 스파이킹 신경망은 기존 방식에 비해 학습방법에 대한 자료가 충분하지 않고 스파이크의 시간 정보를 잃지 않은 상태로 경사하강법을 사용할 수 없는 등 효율적인 구현에 있어 제약이 있지만, 사람이 가지고 있는 뉴런의 생물학적 메커니즘을 보다 근사적으로 모방함으로써 궁극적인 에너지 효율성 확보를 목표로 많은 연구가 수행되고 있다.

초기에는 초고속 벌크 광학 부품을 사용하여 큰 광섬유 기반 시스템을 구축하여 스파이킹 신경망의 개념을 구현하였으며, 이후 확장성, 에너지 효율성, 하드웨어 비용 절감 및 환경 변화에 대한 강

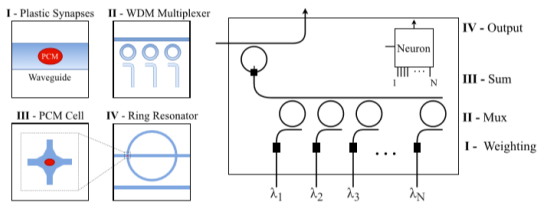
인함을 확보하기 위해 주로 집적 광학 부품에 대한 연구가 진행되었다.

그래핀 레이저는 스파이크 정보 처리를 위한 스파이킹 뉴런으로 제안되었다. 중간에 삽입된 그래핀 층은 비선형 활성화 기능을 수행하기 위해 광포화 흡수체로 사용되었으며, 시간적 패턴 인식을 시연하기 위해 광섬유 기반의 프로토타입으로 구현된 바 있다. 좀 더 작은 크기의 부품 구현을 위해 분포피드백형(DBF: Distributed Feedback) 레이저를 기반으로 하는 스파이킹 뉴런의 개념이 소개되었다. 본 연구에서는 각각 억제(Inhibitory) 및 흥분(Excitatory) 자극을 제공하는 두 개의 광 검출기를 사용하였으며, 본 시스템은 BW 프로토콜과 호환되었고, 최대 10^{12} MAC/s의 처리속도를 보여주었다[7].

최근 들어 위상변화재료(PCM: Phase Change Materials)를 광소자에 적용하여 비휘발성(Non-volatile) 광메모리를 구현하는 선도적인 연구에 관심이 증대되고 있다[8]. 전자 메모리 소자 중 P(Phase-change)-RAM 기술의 위상변화재료인 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST)를 광도파로에 적용하여 다층(multi-level) 다비트의 비휘발성 광메모리의 가능성을 보여준 바 있는데, 이를 뉴로모픽 분야에 적용하여 빠른 쓰기 속도 및 저전력 구동 특성을 확보하는 우수한 뉴런 특성을 보여주고 있다[9].

제안된 구도의 단위 뉴런은 양극성의 적분 및 발사(Integrate-and-fire) 기능을 가지는데, 이때 적분 유닛은 PCM이 적용된 두 개의 MRR 소자로 구성되며 합쳐진 MRR 출력이 상황에 따라 발사 유닛을 활성화하도록 구성된다. 물리적으로는 PCM 소재의 MRR 적용 시에 일어나는 흡수 및 굴절률 변화 현상을 이용하여 시냅스의 가중치와 뉴런의 integrate-and-fire 기능을 수행할 수 있다.

이러한 구도가 실용성을 확보하기 위해서는 매우 큰 규모의 어레이 구현이 필요한데 이런 측면



출처 By J. Feldmann et al. (Licensed under a Creative Commons Attribution 4.0 License).

그림 4 위상변화재료를 마이크로 링 공진기에 적용한 포토닉 뉴런의 개념도

에서 그림 4와 같이 확장성을 가지면서 all-optical로 동작 가능한 PCM 기반 연구가 보고되었다[10]. 본 연구에서는 단위 소자, 단위 뉴런, 단위 레이어들의 계위를 보여주고 이들이 통합되어 하나의 뉴로모픽 포토닉 칩을 구성할 수 있음을 제안하였고, 입출력 펄스를 피드백 라인을 통해서 겹치게 하여 스파이크 생성에 기여한 모든 입력의 시냅스 가중치는 증가하는 반면, 다른 입력은 감소하도록 하는 비지도학습의 가능성을 제시하였다.

4. 축적 컴퓨팅

축적 컴퓨팅에서 레이어는 입력, 저장소 및 판독 레이어의 세 계층으로 구성된다. 이중 가장 복잡한 연결을 제공하는 저장소 부분이 희박하고(Sparsely) 무작위로 연결된 고정 네트워크로 구현되는 점이 특이하며, 이는 연결 복잡성을 크게 줄일 수 있다는 것을 의미하고 포토닉스 기반으로 비교적 쉽게 구현이 가능하다. 실제 동작은 입력 및 판독 레이어에 따라 달라지므로 특정 저장소를 다른 응용 분야에 사용할 수 있고, 훈련은 저장소 영역은 고정하고 일반적으로 저장소와 판독 레이어 사이에서 발생한다.

축적 컴퓨팅은 최근 잠재적 유용성을 인정받아 많은 연구가 진행되고 있으며, 포토닉스 분야에서

도 포토닉스를 활용한 저장소의 물리적인 구현과 수동 포토닉 소자를 시뮬레이션하여 통신망에서의 비선형 분산 보상 등의 응용에 적용하는 등 많은 연구가 이루어지고 있다[11].

IV. 결론

기존 폰 노이만 구조하에서 폭증하는 AI를 위한 데이터 처리에 한계를 보이고 있고, 이를 해결하기 위한 방법으로 인간 뇌의 하드웨어인 뉴런과 시냅스를 모사하는 컴퓨팅 아키텍처와 인공 신경망에 대한 연구가 진행되고 있다. 이의 코어 기술로 뉴로모픽 칩들에 대한 관심이 급증하고 있는데, IBM 및 인텔 등에서는 기존 CMOS 기반으로 하드웨어를 구성하여 상용시제품 수준의 칩 기술을 보여주고 있다.

소자 수준에서 뉴로모픽 개념을 구현하는 연구로 상변환 소자, 강유전체 분극 반전 소자, 스핀 소자 등에 대한 연구가 활발하게 진행되고 있으며, 특히 기본적으로 고속, 병렬성의 장점을 가지며 향후 에너지 효율 및 확장성을 확보할 수 있는 포토닉스 기반의 뉴로모픽 소자에 대한 관심이 최근 급증하고 있다.

앞 절에서 전술한 것처럼 실리콘 포토닉스 기술은 실리콘 파운드리를 활용한 대량 생산 플랫폼을 구축하여 초고속 광통신용 트랜시버로 수조 원 이상의 시장을 이미 형성하였다. 초고속 광통신뿐만 아니라 컴퓨터 내에서 프로세서와 메모리 사이의 고속 병렬 인터페이스로 많은 연구가 진행된 바 있으나 기본적으로 전기 신호가 광 신호로 변환되는 과정들이 필요하여 실제 컴퓨터 내부의 버스로는 활용이 제한되었다.

뉴로모픽 포토닉스 연구는 컴퓨터 내부의 프로세서 안에서 기본적으로 광신호 기반의 프로세싱

을 진행하므로 향후 자연스럽게 기존 광을 이용한 고속 인터페이스 연구와 연결점을 찾을 수 있을 것으로 예상된다.

최근 휴대용 기기, 스마트 디바이스, 네트워크 노드 및 IoT 기기가 요구하는 엣지 컴퓨팅의 핵심은 초저전력을 소모하면서 지능적인 판단을 수행하는 AI 반도체이다. 뉴로모픽 포토닉스 기술은 차세대 AI 신경망인 스파이킹 신경망, 축적 컴퓨팅 등의 핵심 뉴로모픽 칩 및 프로세서 기술의 유력한 대안으로 제시되고 있으며, 향후 국내외에서 보다 많은 연구가 이루어질 것으로 예측된다.

약어 정리

| | |
|-----|-------------------------------|
| ANN | Artificial Neural Network |
| BW | Broadcast and Weight |
| CNN | Convolutional Neural Network |
| GPU | Graphic Processing Unit |
| MAC | Multiply Accumulate Operation |
| MRR | Micro Ring Resonator |
| MZI | Mach-Zehnder Interferometer |
| PCM | Phase Change Materials |

| | |
|-----|------------------------|
| SNN | Spiking Neural Network |
| TPU | Tensor Processing Unit |

참고문헌

- [1] L. D. Marinis et al., "Photonic Neural Networks: A Survey," *IEEE Access*, vol. 7, 2019, pp. 175827-175841.
- [2] M. A. Zidan et al., "The future of electronics based on memristive systems," *Nature Electron.* vol. 1, 2018, pp. 22-29.
- [3] 문승언, "차세대 뉴로모픽 하드웨어 기술동향," *전자통신동향분석*, 제33권 제6호, 2018, pp. 58-68.
- [4] B. J. Shastri et al., "Principles of Neuromorphic Photonics," *arXiv: 1801.00016*, 2018.
- [5] Y. Shen et al., "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, 2017, pp. 441-447.
- [6] A. Mehrabian et al., "PCNNA: A Photonic Convolutional Neural Network Accelerator," *arXiv:1807.08792*, 2018.
- [7] M. A. Nahmias et al., "A TeraMAC neuromorphic photonic processor," in *Proc. IEEE Photon. Conf. (IPC)*, 2018, pp. 1-2.
- [8] C. Rios et al., "Integrated all-photonic non-volatile multi-level memory," *Nature Photon.*, vol. 9, 2015, pp. 725-733.
- [9] I. Chakraborty et al., "Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons," *Scientific Reports*, vol. 8, 2018, pp. 1-9.
- [10] J. Feldmann et al., "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, 2019, pp. 208-215.
- [11] A. Katumba et al., "Neuromorphic Computing Based on Silicon Photonics and Reservoir Computing," *IEEE J. Select. Topics Quantum Electron.*, vol. 24, 2018, p. 8300310.