

차세대 고성능 AI/ML 시스템을 위한 광 네트워크 기술

김대업*, 이준기
한국전자통신연구원
*artkdu@etri.re.kr

Optical network technology for next-generation high-performance AI/ML systems

Dae-Ub Kim and Joon Ki Lee
Electronics and Telecommunications Research Institute (ETRI).

요약

생성형 인공지능 기술의 발달로 인해 발생하는 전력 소비 증가에 대응하고 고성능 AI/ML에 필수적인 CPU, 가속기, 메모리의 연결을 실현하기 위한 광네트워크 적용기술에 대하여 논의하고, 제안된 차세대 고성능 AI 실현을 위한 이중 컴퓨팅 자원 클러스터 형상에 적용된 광네트워크 연결 기술을 통해 최소한의 지연으로 전기적 거리 제한을 극복하였다.

I. 서론

현재 데이터 센터의 통신 영역에서 생성형 인공지능의 등장으로 기존 서버들 간의 통신 기술이 한계에 도달하여, 차세대 고성능 AI/ML을 구현하기 위한 새로운 형상의 통신 기술이 부각되고 있다.

미국에서 발간한 연구 보고서에 따르면 에너지 소비에 따른 CO₂ 배출량이 항공기는 2, 인간의 생애주기 11, 전체 미국인 36, 미국 전체 차량 126에 비해서 AI 모델에 따른 배출량이 626으로 극도로 증가할 것이라고 예측하고 있다.[1] 그래서 AI/ML에 필수적인 가속기, GPU와 CPU 제조 업체들인 NVIDIA, INTEL, AMD 등은 저전력 반도체 기술을 위해 노력하고 있지만, 실제 컴퓨팅 자원을 연결하는 대역폭은 수 Tbps 이상으로 증가하고 있어 전기적인 스위치를 통한 연결은 소용 전력량은 엄청나게 증가될 수 있다. 고성능 AI/ML과 Computing 시스템 고도화의 진행은 하나의 GPU의 추론 성능이 10년 만에 1000배로 증가하였고 이에 따라서 소비 전력도 급격히 증가되고 있지만, 생성형 AI는 더 고성능의 가속기 솔루션을 요구하고 있으며, 이는 하나의 가속기 성능으로 만족하지 못하고 여러 개의 가속기를 연결하여 병렬 처리를 요구하고 있다.

이는 가속기 연결을 통한 AI Fabric의 확장과 가속기 간의 연결을 위한 전기적 스위치의 처리량(Throughput)의 증가가 필요하다. NVIDIA는 현재 고성능 AI Fabric을 구성하기 위해서 두 개의 연결 네트워크 도메인을 구성하고 있다. 그림 1에서처럼 첫번째 네트워크는 256개의 가속기 노드 연결을 위한 NVLink 프로토콜을 활용하는 네트워크 도메인이며, 두 번째는 100G Infiniband 프로토콜을 활용하는 네트워크 도메인이다. 하지만 이러한 구조는 기존 이더넷이나 Infiniband 프로토콜을 활용하는 기존 방식은 AI/ML의 학습 데이터를 취합하는 Host CPU와 가속기 간의 데이터 전달이 증가함에 따라 컴퓨팅 성능 향상에 제약 사항이 되고, 가속기 노드 사이의 전기적 스위치의

사용은 전력 효율적으로 바람직하지 않아 새로운 기술을 요구하고 있다.

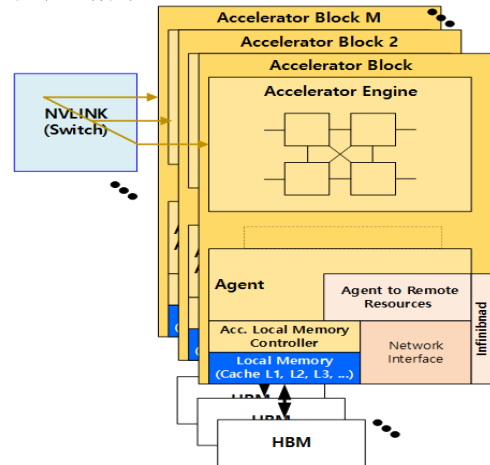


그림 1. 현재 고성능 AI 가속기 클러스터 형상

본 논문에서 이러한 요구사항에 따른 새로운 컴퓨팅 자원 연결 프로토콜을 논의하고, 전기적 인터페이스와 스위칭에 사용되는 전력 효율을 높이기 위한 해결방안을 논의한다.

II. 본론

AI/ML의 컴퓨팅 동작 특성을 보면 데이터 입출력 인터페이스와 프로세서를 통해 여러 형태로 수집되는 데이터를 기반으로 가속기를 통해 반복적인 학습을 진행하여 인공지능이 예측 동작하도록 되어있다. 그래서 그림 2와 같이 현재 INTEL, NVIDIA, AMD 뿐만 아니라 삼성, 하이닉스 등의 메모리 업체 등이 참여한 새로운 형태의 컴퓨팅 자원 연결 기술에 대한 연구가 시작되었다.

컴퓨터 고속 링크(Compute express link, CXL) 기술은 2019년에 Intel의 주도로 시작된 산업표준으로 그림 2과 같이 현재는 수백개의 기업체가 참여하고 있는 오픈형 상호 연결 표준으로, Host 프로세서와 CXL 장치

간에 저지연 고대역폭 접근을 제공하기 위하여 제안된 표준이다.

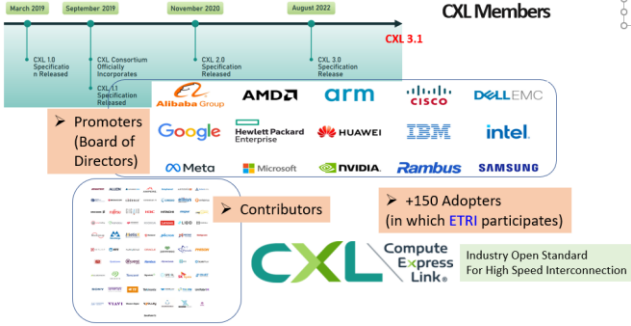


그림 2. Compute express link (CXL) 연구 현황

가속기와 메모리는 CXL 장치로 분류되며, Host CPU 와 가속기 사이에 각 각의 메모리에 서로 접근할 수 있도록 하여 AL/ML 을 포함하는 이종 컴퓨팅의 전체 성능을 향상시킬 수 있는 구조를 제공한다. 최근 데이터 센터에서 적용하기 위해 활발하게 연구되고 있고, 관련 초기 시제품이 출시되고 있다. 빅데이터와 AI 등에서 필요한 고집적 워크로드의 연산을 수행하기 위해서 GPU 나 FPGA 와 같은 가속기의 도입 확대가 필요하고 CXL 에서는 이런 가속기들이 CXL 장치로서 Host 와 협업할 수 있도록 하였다. CXL 은 기존의 non-coherent (비간섭) IO 장치, 메모리 장치, 가속기 등의 주변 장치를 상호 연결하는 데 사용할 수 있는 고성능 I/O 버스 아키텍처 기술이다. [2-3]

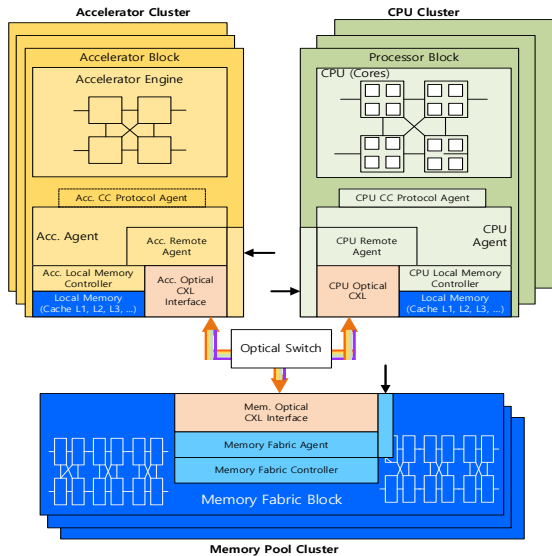


그림 3. 차세대 고성능 AI 실현을 위한 이종 컴퓨팅 자원 클러스터 형상

본 논문에서 그림 3 과 같은 우리가 개발하는 차세대 고성능 AI 실현을 위한 이종 컴퓨팅 자원 클러스터 형상을 설명하고, 프로세서, 가속기, 메모리의 연결 대역폭은 현재도 수십 Tbps 이상이고, 여러 개의 ASIC 칩을 연결하기 위해서 전달 거리와 입출력 대역폭 확장도 큰 도전으로 인식되고 있어, 기술 발전을 위해 적용 가능한 기술을 논의하고자 한다. 우리는 그림 4 와 같이 현재의 광 트랜시버를 실리콘 포토닉스 기술을 활용하여 전력 소모량을 줄이고 소규모 형태로 칩에 직접 연결이 가능한 Linear Pluggable Optics (LPO)과 Co-package Optics (CPO) 형태로 개발하였다.

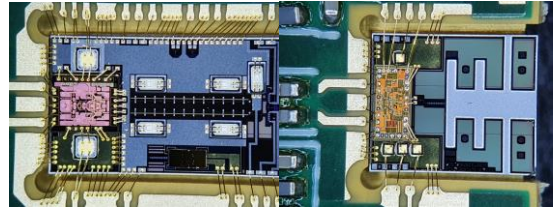


그림 4. SiPh 기반의 광 연결 개발

Optically Networked CXL (ON-CXL) 프로토콜과 에플레이터를 통하여 컴퓨팅 자원 간의 연결 프로토콜을 기존 이더넷이나 PCIe 와는 다르게 데이터의 저장과 복사를 최소화하고, OS 의 소프트웨어 관여도 최소화하여, 최소한의 지연과 성능유지를 통해서 전기적 거리 제한을 극복하였다. [4]

III. 결론

본 논문에서 차세대 고성능 AI/ML 을 구현하기 위해서는 증가하는 전력 소비에 대한 문제를 설명하고, 기존 기술과 다른 새로운 형상의 컴퓨팅 자원 연결 기술로서 CXL 을 소개하였다. CXL 프로토콜을 기반으로 우리는 증가하는 전력 소비에 대응하고 고성능 AI/ML 에 필수적인 CPU, 가속기, 메모리의 원격연결을 실현하기 위한 광네트워크 기술의 적용과 해당 차세대 고성능 AI 실현을 위한 이종 컴퓨팅 자원 클러스터 형상을 제시하였다. 제안된 이종 컴퓨팅 자원 클러스터 형상에 적용된 Optically networked compute express link (ON-CXL) 기술은 광네트워크 연결 기술을 통해 최소한의 지연으로 전기적 거리 제한을 극복하였다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2019-0-00002, 광 클라우드 네트워킹 핵심원천 기술 개발)

참 고 문 헌

- [1] An American life has a larger carbon footprint than a human life because the U.S. is widely regarded as one of the top carbon dioxide emitters in the world, Data a compiled Oct 9, 2019
- [2] S. Ryu, et. al., " System Optimization of Data Analytics Platforms using Compute Express Link (CXL) Memory," IEEE BigComp, Feb. 2023. pp.09-12
- [3] D. D. Sharma and I. Agarwal, "Compute Express Link 3.0", Aug. 2022. Miles E. Smid, "From DES to AES," 2000, (<http://www.nist.gov/aes>).
- [4] D.U. Kim, et al., " Optically Networked Heterogeneous Data-Centric Computing System with Silicon Photonics Transceivers," OFC, March 2024, M1G.2.