

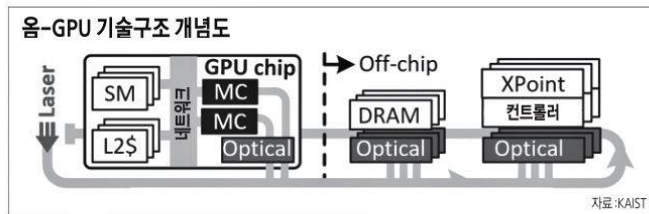
GPU메모리 용량 확 키우는 기술 나왔다

정명수 KAIST교수팀 기술 개발
이중 메모리 시스템 채택해 구현

국내 연구진이 엔비디아 등 소수의 글로벌 IT기업이 개발하고 있는 GPU(그래픽처리장치) 메모리 용량과 성능을 대폭 향상시킨 기술을 개발했다. AI(인공지능), 빅데이터, 클라우드 컴퓨팅 등 대용량, 고대역폭 데이터를 처리하는 고성능 가속기 메모리 시스템에 쓰일 것으로 기대된다.

KAIST는 정명수 교수 연구팀이 '3D XPoint 메모리(DRAM에 비해 용량은 크지만 데이터 전송 속도가 느린 메모리)'와 DRAM 메모리(데이터 전송 속도는 빠르나 용량이 작은 메모리)를 통합한 이중 메모리 시스템에서 광네트워크로 통신하는 '옴-지피유(Ohm-GPU)' 기술을 개발했다고 2일 밝혔다.

GPU는 다수의 연산 장치로 구성돼 있어 연산 속도가 매우 빠르지만, DRAM을 단독 사용하기 때문에 낮은 메모리 용량과 좁은 데이터 전송 대역폭 등으로 인해 연산 성능을 충분히 발휘



하지 못한다.

이 때문에 DRAM을 XPoint 메모리로 대체해 8배 많은 데이터 용량을 얻을 수 있는 데 반해, 읽기와 쓰기 성능은 각각 4배, 6배 떨어지는 한계가 있다.

아울러, 대역폭을 증가시키기 위해 HBM(3D로 DRAM을 쌓은 메모리) 기술을 적용하고 있으나, 단일 면적 내 장착할 수 있는 전기체널(구리선) 개수의 한계로 인해 GPU 메모리 시스템이 요구하는 대역폭을 만족시키기 어렵다.

연구팀은 대용량 XPoint와 고성능의 DRAM을 통합한 이중 메모리 시스템을 채택해 메모리 용량을 높였다. 또 단일 광체널(광섬유)로 서로 다른 파장의 다중 광신호를 전달할 수 있는 광

네트워크를 활용해 메모리 고대역폭을 구현했다.

'옴-지피'로 명명된 이 기술은 XPoint와 DRAM을 하나의 통합된 공간으로 구현해 대용량 메모리 공간을 제공하면서, XPoint와 DRAM 간 데이터 교체 처리를 통해 속도를 높였다. 두 메모리 간의 데이터 이동으로 생기는 오버헤드 역시 메모리 접근의 광 파장을 달리하고, 메모리 컨트롤러의 개입을 최소화해 완화했다고 연구팀은 설명했다.

기존 DRAM을 단독으로 사용하는 전기 네트워크 기반의 GPU 메모리 시스템에 비해 다양한 그래프 처리와 실행 등에서 1.8배 이상의 높은 성능 향상을 보였다.

이준기기자 bongchu@

KAIST, 연산 성능 181% 향상 ‘GPU 메모리 시스템’ 개발

광 네트워크로 메모리 대역폭 넓혀 고성능 가속기 메모리 시장 선점 기여

한국과학기술원(KAIST·총장 이광형)은 정명수 전기 및 전자공학부 교수팀(컴퓨터 아키텍처 및 운영체제 연구실)이 3D 크로스포인트(XPoint), D램 메모리를 통합한 이중 메모리 시스템에서 광 네트워크로 통신하는 ‘옴-GPU(Ohm-GPU)’ 기술개발에 성공했다고 2일 밝혔다.

기존 D램을 단독 사용한 전기 네트워크 기반 GPU 메모리 시스템 대비 181% 이상 성능 향상을 얻었다. XPoint는 D램 대비 용량이 크지만 데이터 전송 속도가 느린 메모리다.

기존 GPU는 다수 연산 장치로 구성돼 연산 속

도가 빠르지만 D램 단독 사용 메모리 시스템의 낮은 메모리 용량과 좁은 데이터 전송 대역폭 탓에 연산 성능을 충분히 활용하지 못한다.

D램을 XPoint로 대체해 용량을 늘리는 방법이 있으나 읽기·쓰기 성능이 낮아진다. D램을 적층해 대역폭을 증가시키는 HBM 기술도 GPU 메모리 시스템이 요구하는 고대역폭을 만족하기 어렵다.

옴-GPU 기술은 대용량 XPoint와 고성능 D램을 통합한 이중 메모리 시스템을 채택, 기존 메모리 시스템과 동일한 성능을 가지면서 메모리 용량을 증가시켰다. 또 단일 광 채널(광섬유)로 다중 광신호를 전달할 수 있는 장점을 활용해 메모리 대역폭을 대폭 넓혔다.

GPU 내부에 있는 메모리 컨트롤러 및 인터페

이스를 수정해 이중 메모리의 모든 메모리 요청을 광신호로 처리한다. 이중 메모리 간 데이터 이동 대기 시간은 연산 메모리 접근과 데이터 이동 메모리 접근 광 파장을 다르게 설정하고 메모리 컨트롤러 개입을 최소화하는 등 방법으로 완화했다.

옴-GPU 기술은 기존 D램을 단독으로 사용하는 전기 네트워크 기반 GPU 메모리 시스템 대비 181% 성능 향상을 달성했다. 고성능 가속기 메모리 시스템을 대체할 수 있을 것으로 기대된다.

정명수 교수는 “GPU 메모리 시스템 기술은 일부 해외 우수 기업이 주도하지만, 이번 연구성과를 기반으로 GPU 및 GPU와 유사한 모든 고성능 가속기 메모리 시스템 관련 시장에서 우위를 선점할 가능성을 열었다”고 강조했다.

대전=김영준기자 kyj85@etnews.com