027 메모리 연산 / PIM

Processing In Memory

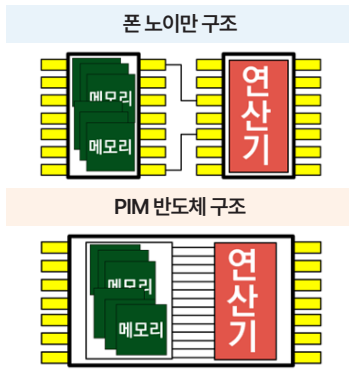
메모리 내부에서 연산을 직접 수행해 처리 효율을 높이는 컴퓨팅 기술

- 데이터 이동 없이 메모리 자체에서 계산을 수행하는 컴퓨팅 구조
- 대규모 작업에서 지연을 줄이고 새로운 시스템 설계를 가능하게 하는 방식

PIM의 개념

PIM은 CPU와 메모리 간 반복적인 데이터 이동으로 발생하는 비효율을 줄이기 위해 등장한 기술로, 연산 기능을 메모리 내부 또는 매우 가까운 위치에 배치하는 구조를 의미합니다. 기존 컴퓨팅 방식에서는 모든 연산이 CPU에서 이루어져 대규모 데이터 처리 시 병목이 쉽게 발생했습니다. PIM은 이러한 구조적 한계를 개선하기 위한 새로운 접근으로, 특히 데이터 접근 요구가 빈번한 작업 환경에서 주목받고 있습니다. AI 모델의 연산량 증가와 메모리 중심 처리 요구가 높아지면서, 메모리 자체에 연산 기능을 통합하는 방식이 차세대 컴퓨팅 구조의 변화 방향으로 논의되고 있습니다.

PIM의 작동 방식



PIM은 메모리 모듈 내부에 간단한 연산 유닛을 포함하거나, 메모리와 매우 가까운 영역에 연산 장치를 배치해 동작합니다. 이를 통해 데이터가 먼 연산 장치로 이동하지 않고, 메모리 내부에서 필요한 계산을 수행할 수 있습니다. 이는 연산장치와 메모리가 분리된 폰 노이만 구조에서 발생하는 데이터 이동 병목을 줄여줍니다. D램 기반 PIM 방식은 메모리 셀의 구조를 활용해 간단한 연산을 병렬적으로 처리할 수 있도록 하며, 일부 구조에서는 행렬 곱셈과 같은 특정 연산을 메모리에서 직접 수행하도록 설계됩니다. 이러한 방식은 복잡한 알고리즘을 모두 메모리에서 처리한다기보다, 대량의 반복 연산이나 특정 패턴의 계산을 메모리 근처에서 빠르게 처리하도록 최적화된 구조에 가깝습니다.

PIM의 활용

PIM은 메모리 접근이 많은 작업에서 구조적 이점을 제공하여, 그래프 탐색, 추천 시스템, 벡터 검색처럼 데이터 위치 정보를 반복적으로 조회해야 하는 작업에서 응답 지연을 줄이고 처리 흐름을 단순화할 수 있습니다. 또한 PIM은 기존의 CPU·GPU 중심 구조를 대체하기보다는, 메모리 중심 처리가 필요한 특정 영역을 보완하는 기술로서 가치가 있습니다. 이를 통해 제약이 컸던 메모리 기반 알고리즘을 실용적으로 구현할 수 있는 기반이 마련되며, 향후 컴퓨팅 아키텍처 설계에서 새로운 선택지를 제공한다는 점에서 의미가 있습니다.