Toward High-Fidelity Heterogeneous Memory System Modeling in gem5

Maryam Babaie, Ayaz Akram, Jason Lowe-Power

DArchR Research Group, CS Department, University of California Davis

{mbabaie, yazakram, jlowepower}@ucdavis.edu

ModSim Workshop
Seattle, WA, USA
Aug. 2022

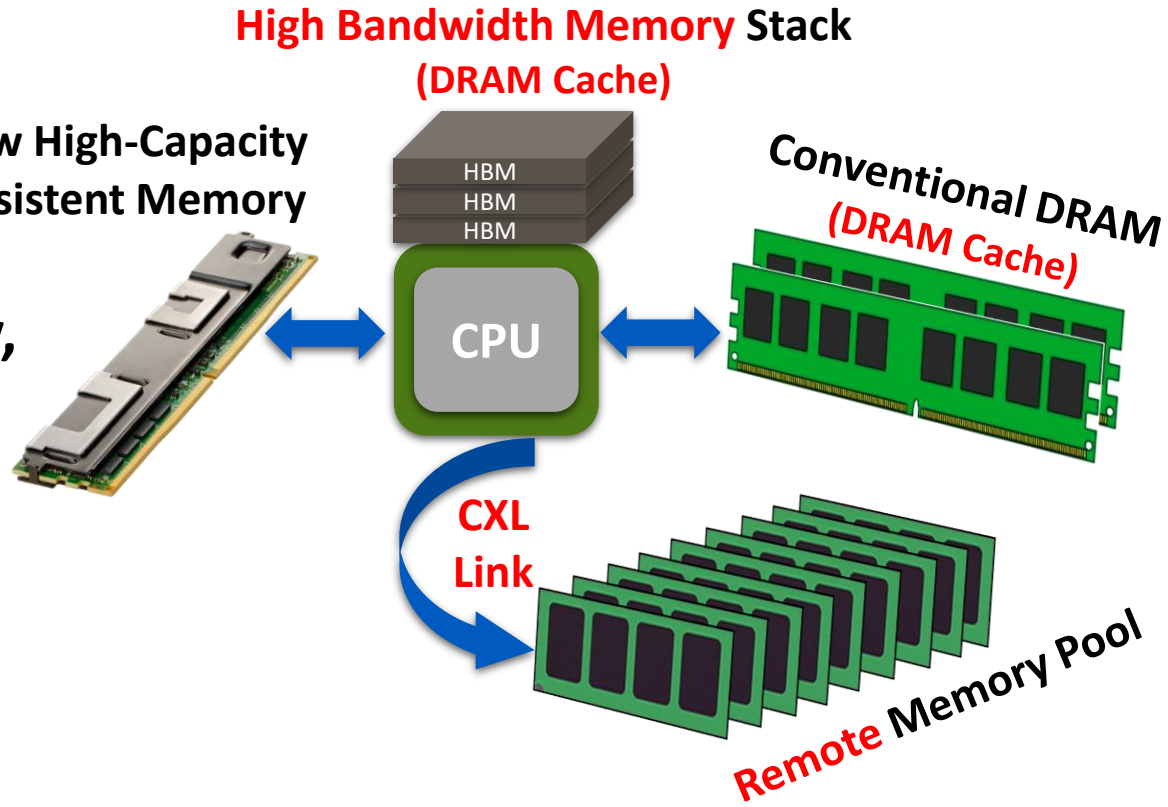
UC DAVIS
UNIVERSITY OF CALIFORNIA

Motivation

HPC systems rely on **heterogeneous memory,**
e.g., *Knights Landing, Cascade Lake, Sapphire Rapids.*

We need to rethink memory management.

☹ No accurate model available for community.



We extend *gem5* with a heterogeneous memory system model for design space exploration of future HPC systems. We add support for:

1. DRAM Cache 2. HBM interface and controller 3. Modular memory controller design

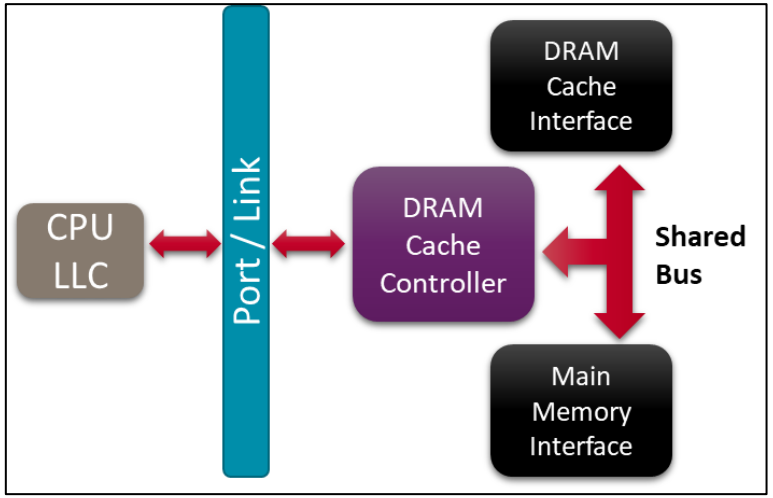
gem5's DRAM Cache Support

A. Unified DRAM Cache/Memory Controller

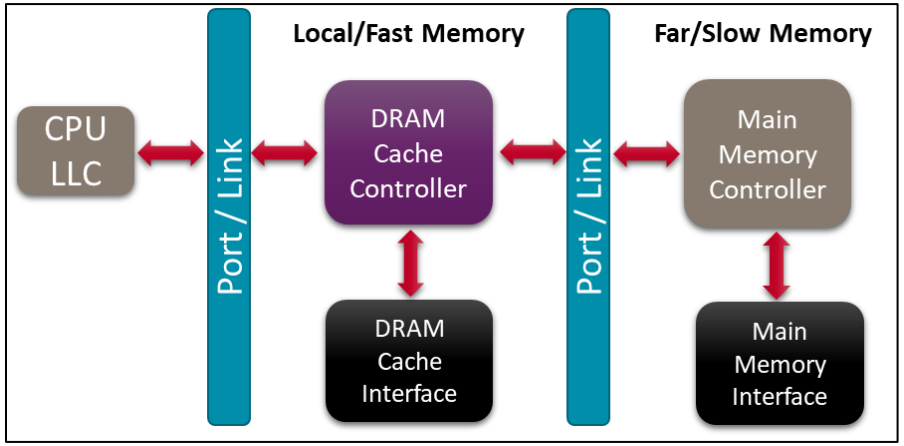
B. Disaggregated DRAM Cache Controller

Goal: Enable <u>cycle-level</u> analysis of <u>DRAM cache designs</u>

- **Cache** VS **Backing-store** Combination
- **Fast/Local** VS **Slow/Far** Memory Nodes
- Analysis of **CXL** Connections
- **Data Management** analysis



(A)



(B)

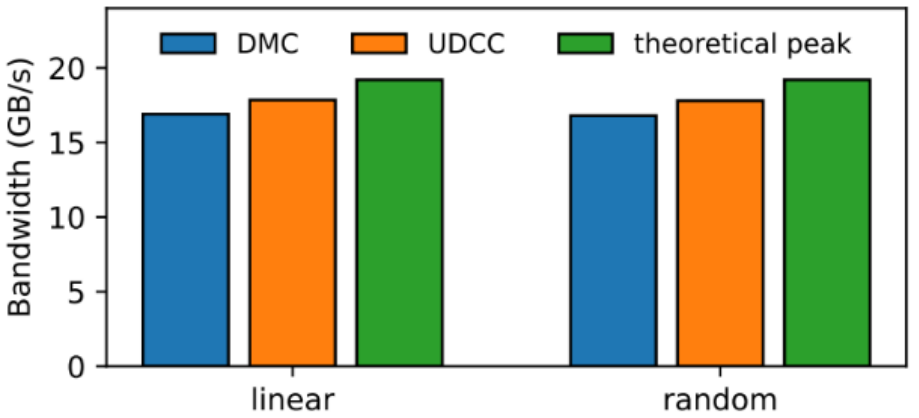
UCDAVIS

gem5's DRAM Cache Validation

Performance Validation

1. Read bandwidth compares well with the theoretical peak.

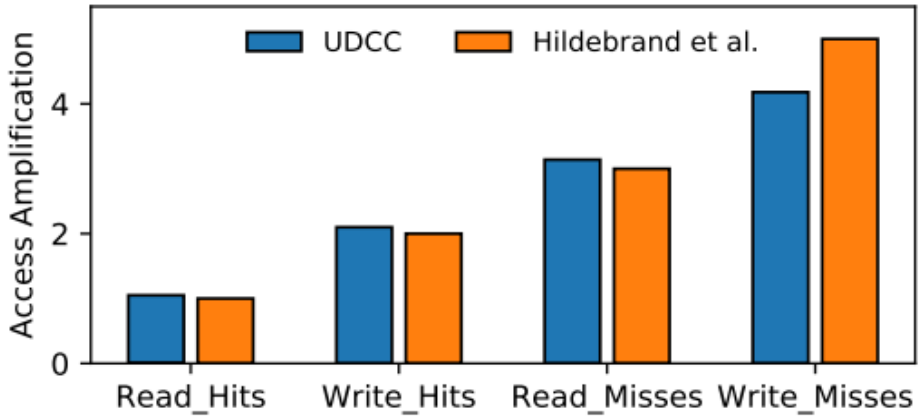
2. Access amplification is same as in real hardware [1].



DMC: Default Memory Controller of gem5
UDCC: Unified DRAM Cache/Memory Controller

Functional Validation

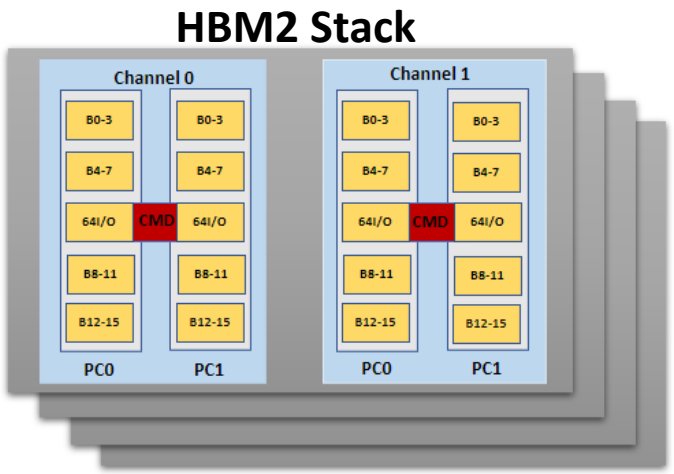
- Successfully tested Linux kernel booting
- Tested GAPBS and NPB benchmark suites

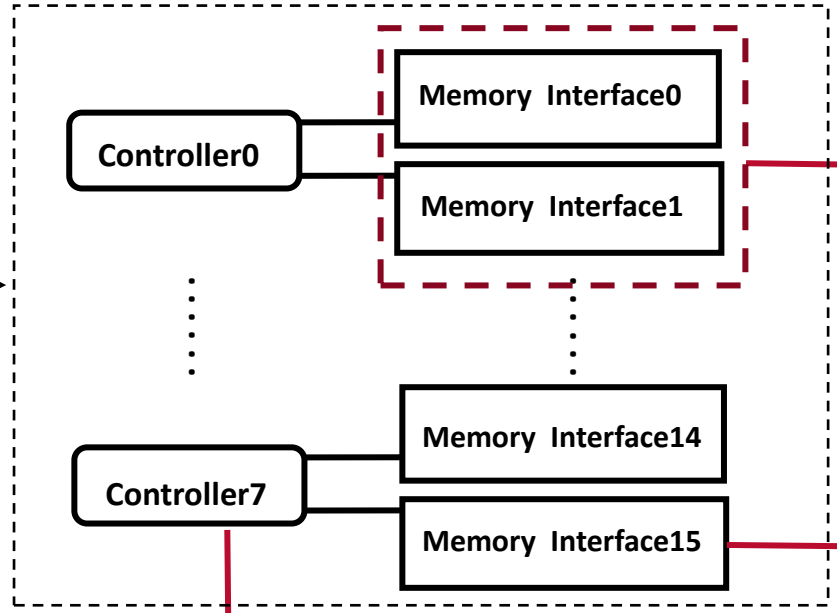


1 M. Hildebrand, J. T. Angeles, J. Lowe-Power, and V. Akella, "A case against hardware managed dram caches for nvram based systems," in 2021 ISPASS.

UCDAVIS

HBM2 Stack Modeling in gem5





Physical channel contains two pseudo channels (PCs)
- Have separate data bus
- Share command bus

A new HBM controller
- Controls a single HBM physical Ch
- Able to model pseudo channels
- 8 controllers model a 4H stack

A validated HBM2 memory interface^

HBM Controller

* Parallel accesses to 2 DRAM interfaces (PCs) if command (cmd) bandwidth allows

Partitioned/shared queues

HBM2 Interface

Configured using JEDEC specs. and other tools *

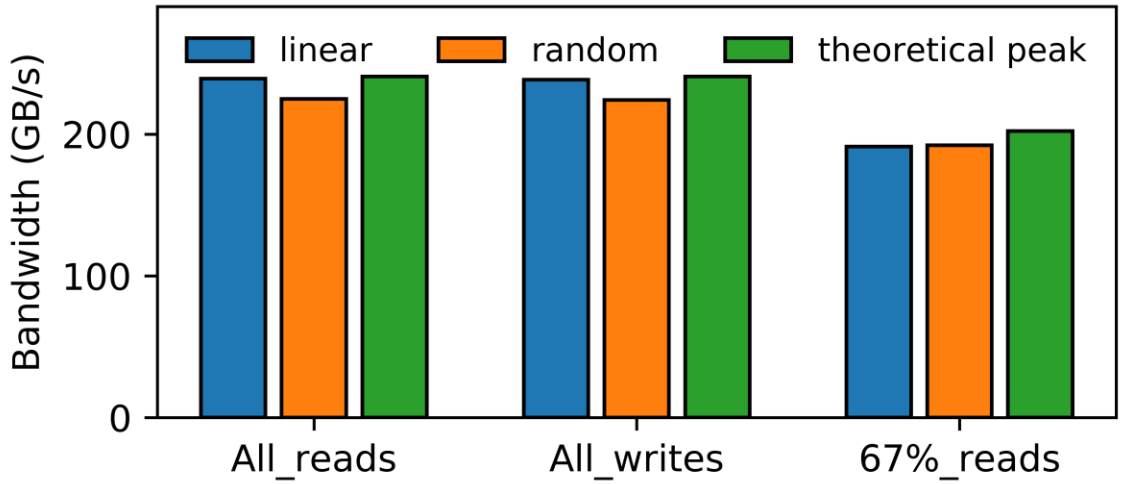
Supports asynchronous Rd/WR latencies

*Our contribution in gem5. ^memory interface is gem5 abstraction of a memory device model.

UCDAVIS

HBM2 Model Performance Evaluation

Results match the theoretical peak



Refactoring of gem5 memory controller

Baseline Memory Controller

- **Heterogenous Memory Controller**
- **HBM Controller**
- **DRAM Cache Controller**
- **Your Controller**

Thank you and see you in the poster session ☺

Questions:

Maryam Babaie
mbabaie@ucdavis.edu

Ayaz Akram
yazakram@ucdavis.edu

UCDAVIS