

# Network-aware Data Management for Disaggregated Memory Systems

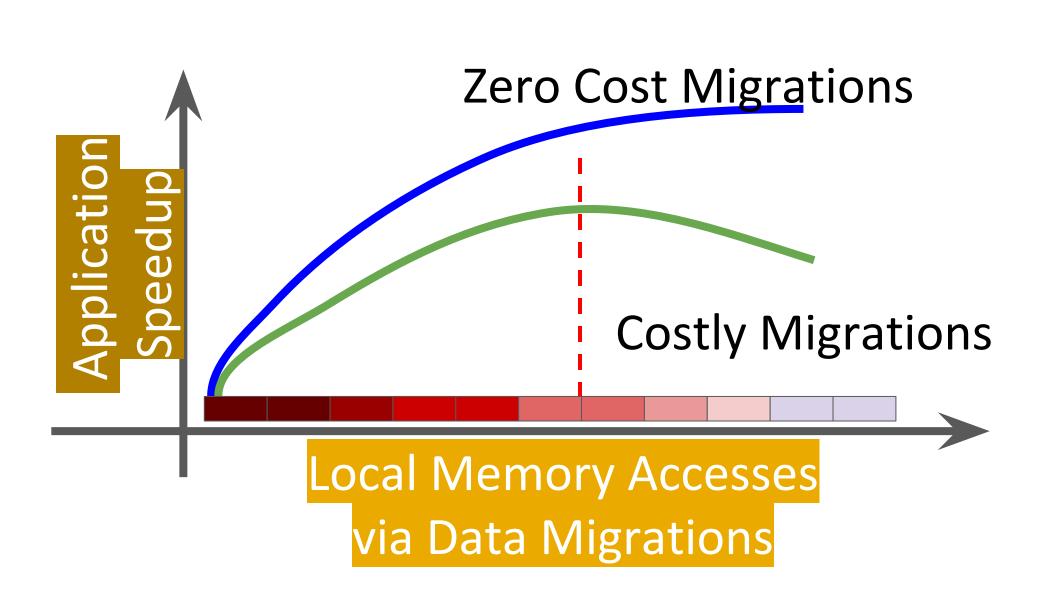
Scan me

Thaleia Dimitra Doudali and Ada Gavrilovska Georgia Institute of Technology

### 1. Problem Space

### Application in-memory data cold data hot data → access frequency → Data Management Software Dynamic Data Tiering hot data Data rest of **Local Memory** the data Migrations HBM DRAM NVM DRAM CPU FPGA -NVM High Speed DRAM Compute Unit Interconnect NVM CPU GPU **Remote Memory** DRAM NVM (Shared or Dedicated) **Local Memory** Disaggregated Memory System

## 2. Challenges



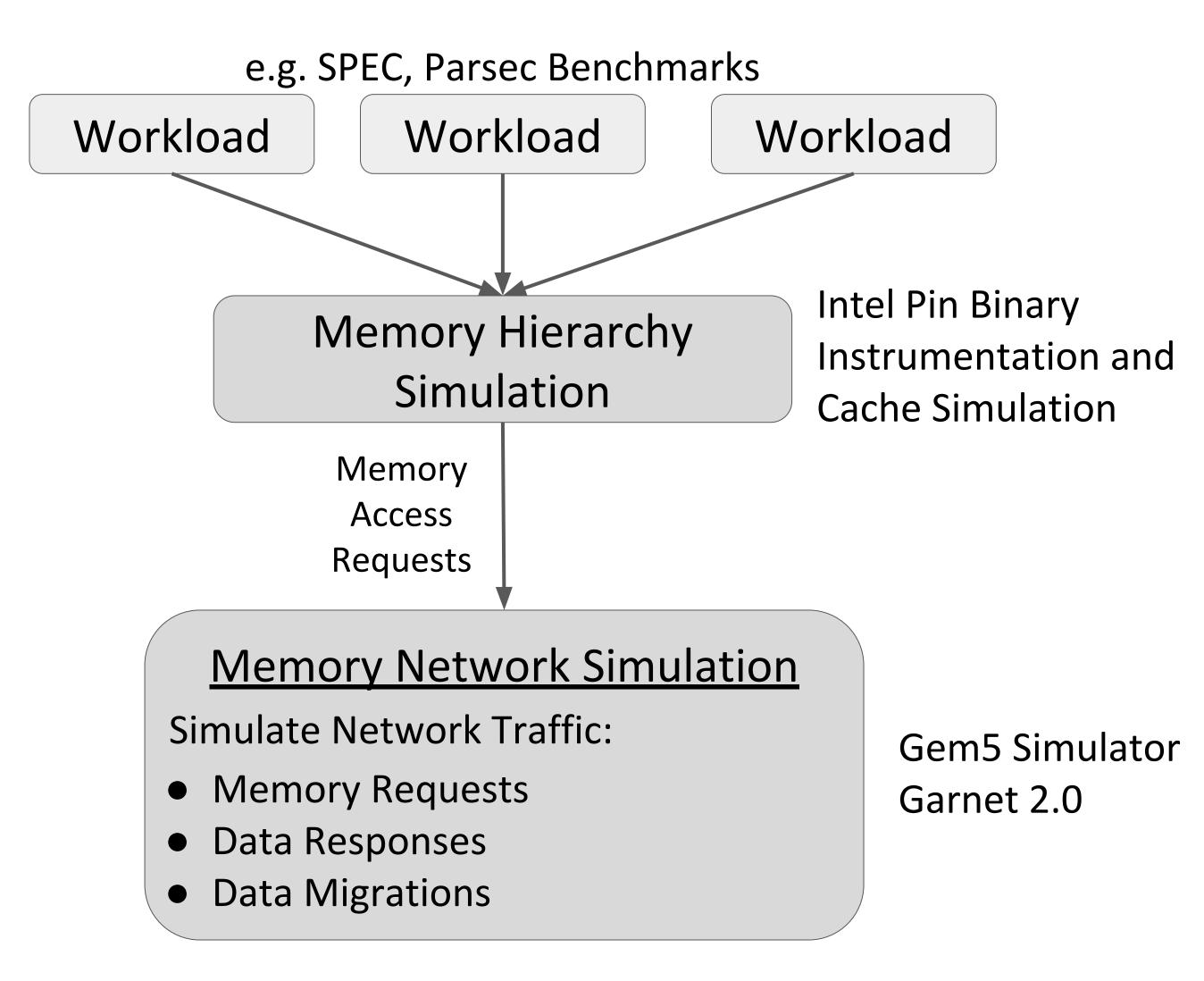
Time to migrate the pages > Time to access them from remote memory.

- 1. How to predict which application data to migrate? [Kleio¹ HPDC '19 best paper award finalist]
- 2. How to predict the performance curve for zero cost migrations? [Mnemo<sup>2</sup> HPBDC workshop of IPDPS '19]
- 3. How to predict the performance curve for costly migrations? [Work In Progress]

### 3. Solution

# Data Management Software Dynamic scheduling of data migrations based on memory and network load Explicit management of data migration traffic Remote Memory Remote Memory Access Traffic High Speed Interconnect

# 4. Implementation



### **Evaluation Metrics:**

- Average Data Response Time
- Bandwidth Utilization

Goal
Optimize Application Performance

Maximize Local Memory Accesses

Managed bursts of Data Migrations