NMTSim: Transaction-Command based Simulator for New Memory Technology Devices

Peng Gu¹, Benjamin S. Lim²
Wenqin Huangfu¹, Krishna T. Malladi², Andrew Chang², Yuan Xie¹

¹University of California, Santa Barbara
²Samsung Semiconductor
New Non-volatile Memory Technologies

Higher Capacity, Lower Cost/Bit

Non-persistent  Persistent

Load/Store Interface

Block Interface

SRAM cache  DRAM  New Memory Technology  NAND SSD  HDD

$10^0$  $10^1$  $10^2$  $10^3$  $10^4$  $10^5$  $10^6$

Media Access Time (ns)

Intel Optane™
New Non-volatile Memory Technologies

Key benefits:
• Significant performance improvement
• Potential to augment main memory

Challenges:
• Traditional DDR interface uses deterministic timing. How to tolerate different media access latency?

<table>
<thead>
<tr>
<th></th>
<th>Read Latency</th>
<th>Write latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDD</td>
<td>3-5 ms</td>
<td>3-5 ms</td>
</tr>
<tr>
<td>NAND FLASH</td>
<td>15-35 us</td>
<td>200-50 us</td>
</tr>
</tbody>
</table>

"Emerging NVM: A Survey on Architectural Integration and Research Challenges"
Traditional DRAM simulator

- **Deterministic timing**
  - HOST MC
  - DRAM Device
  - t0 \rightarrow \text{READ} \rightarrow t1
  - t1 \rightarrow \text{DATA}\

- **Indeterministic timing**
  - HOST MC
  - DRAM Device
  - t0 \rightarrow \text{READ} \rightarrow t1
  - t1 \rightarrow \text{DATA}\

**New Memory Devices**
- STT-RAM
- ReRAM
- FeRAM
- PCM
NMTSim: Transaction Command Simulator

Indeterministic timing

- Require data
- Receive read ready signal
- Respond SEND command
- Receive data

HOST MC

XREAD/XADR

Media Controller

READ

DATA

New Memory Devices

STT-RAM

ReRAM

FeRAM

PCM
NMTSim validation

Baseline (DDR-T/RDIMM) | NMTSim (DRAM) | Latency Error (norm. to Baseline)
--- | --- | ---
All read | 2.8% latency error for DRAM device
2read+1write

Baseline (DDR-T/Optane) | NMTSim (NVM) | Latency Error (norm. to Baseline)
--- | --- | ---
All read | 3.4% latency error for NVM device
2read+1write
NMTSim optimizations

Command Issue Priority:
- p1: XREAD > SEND
- p2: SEND > XREAD

Takeaway:
- Reduce response queuing latency is important in latency reduction at high memory rates
NMTSim optimizations

- e1: No early notification
- e2: Early notification enabled

Takeaway:
- Enabling early notification helps reduce latency significantly
NMTSim: Transaction-Command based Simulator for New Memory Technology Devices

Thank you!