Rethinking DRAM Power Modes for Energy Proportionality

Krishna Malladi\textsuperscript{1}, Ian Shaeffer\textsuperscript{2}, Liji Gopalakrishnan\textsuperscript{2}, David Lo\textsuperscript{1}, Benjamin Lee\textsuperscript{3}, Mark Horowitz\textsuperscript{1}

Stanford University\textsuperscript{1}, Rambus Inc\textsuperscript{2}, Duke University\textsuperscript{3}

ktej@stanford.edu
Main Memory in Datacenters

- Server power main energy bottleneck in datacenters
  - PUE of ~1.1 \(\Rightarrow\) the rest of the system is energy efficient
- Significant main memory (DRAM) power
  - 25-40% of server power across all utilization points
  - Low dynamic range \(\Rightarrow\) No energy proportionality
Main Memory in Datacenters

- Server power main energy bottleneck in datacenters
  - PUE of ~1.1 ➞ the rest of the system is energy efficient
- Significant main memory (DRAM) power
  - 25-40% of server power across all utilization points
  - Low dynamic range ➞ No energy proportionality
Outline

- Inefficiencies of DRAM interfaces

- Energy-proportionality via fast DRAM interfaces
  - MemBlaze
  - MemCorrect
  - MemDrowsy
Outline

- Inefficiencies of DRAM interfaces

- Energy-proportionality via fast DRAM interfaces
  - MemBlaze
  - MemCorrect
  - MemDrowsy
**DDR3 Energy & Powermodes**

- DDR3 optimized for high bandwidth
  - High speed interface with DLLs, CLKs, ODTs
  - Very high static power in active-idle

- Hard to powerdown to deep states
  - Long impractical wakeup time to power up interface
  - Insufficient idleness in workloads ➔ Significant active-idle time

<table>
<thead>
<tr>
<th>Power Mode</th>
<th>DIMM Idle Power (W)</th>
<th>Exit Latency (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Idle</td>
<td>5.36</td>
<td>0</td>
</tr>
<tr>
<td>Fast Powerdown</td>
<td>2.79</td>
<td>20</td>
</tr>
<tr>
<td>Deep Powerdown</td>
<td>0.92</td>
<td>768</td>
</tr>
</tbody>
</table>
**DDR3 Energy & Powermodes**

- **DDR3 optimized for high bandwidth**
  - High speed interface with DLLs, CLks, ODTs
  - Very high static power in active-idle

- **Hard to power down to deep states**
  - Long impractical wakeup time to power up interface
  - Insufficient idleness in workloads → Significant active-idle time

<table>
<thead>
<tr>
<th>Power Mode</th>
<th>DIMM Idle Power (W)</th>
<th>Exit Latency (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Idle</td>
<td>5.36</td>
<td>0</td>
</tr>
<tr>
<td>Fast Powerdown</td>
<td>2.79</td>
<td>20</td>
</tr>
<tr>
<td>Deep Powerdown</td>
<td>0.92</td>
<td>768</td>
</tr>
</tbody>
</table>

---

- **Power Mode DIMM Idle Power (W) Exit Latency (ns)**
- **% time**
- **Rd-Wr**
- **Powerdown**
- **Active Idle**
**DDR3 Energy & Powermodes**

- **DDR3 optimized for high bandwidth**
  - High speed interface with DLLs, CLKs, ODTs
  - Very high static power in active-idle

- **Hard to powerdown to deep states**
  - Long impractical wakeup time to power up interface
  - Insufficient idleness in workloads → Significant active-idle time

<table>
<thead>
<tr>
<th>Power Mode</th>
<th>DIMM Idle Power (W)</th>
<th>Exit Latency (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Idle</td>
<td>5.36</td>
<td>0</td>
</tr>
<tr>
<td>Fast Powerdown</td>
<td>2.79</td>
<td>20</td>
</tr>
<tr>
<td>Deep Powerdown</td>
<td>0.92</td>
<td>768</td>
</tr>
</tbody>
</table>
**DDR3 Energy & Powermodes**

- **DDR3** optimized for high bandwidth
  - High speed interface with DLLs, CLKs, ODTs
  - Very high static power in active-idle

- **Hard to powerdown to deep states**
  - Long impractical wakeup time to power up interface
  - Insufficient idleness in workloads → Significant active-idle time

<table>
<thead>
<tr>
<th>Power Mode</th>
<th>DIMM Idle Power (W)</th>
<th>Exit Latency (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Idle</td>
<td>5.36</td>
<td>0</td>
</tr>
<tr>
<td>Fast Powerdown</td>
<td>2.79</td>
<td>20</td>
</tr>
<tr>
<td>Deep Powerdown</td>
<td>0.92</td>
<td>768</td>
</tr>
</tbody>
</table>

88%!
Path to Energy-Proportionality
Path to Energy-Proportionality

- Active
- Powerdown
- Active-Idle
- DRAM Power
Path to Energy-Propportionality

- Reduce active-idle power
Path to Energy-Proportionality

- Reduce active-idle power
- Reduce time in active-idle
- Increase time in power-down
Path to Energy-Proportionality

- Reduce active-idle power
- Reduce time in active-idle
- Increase time in power-down
- Reduce power-down power
DRAM Interfaces

- Bits are short
  - Sampling window is only 625ps
- Data (DQ) and Clock (CLK) signals forwarded to DRAM
- Write data aligned to Clock edges
Dynamic chip variations affect Reads

- PVT variations $\rightarrow$ Misaligned DQS and CLK signals
- Non-deterministic Read timing $\rightarrow$ Incorrect sampling
**DRAM Interfaces**

- On-chip DLLs
  - Adjust delay to match chip temperature, voltage variations
  - Align DQS, DQ to CLK
- Power hungry, long settling time \(\rightarrow\) poor power modes
Live with Slow-Powerup

- S/W mechanisms
  - Batch requests (or) subset ranks (or) Predict idleness
    - Degrades application performance
    - Degraded device density

- H/W mechanisms
  - Statically Disable DLLs in BIOS → Statically lowers bandwidth
    - Worse performance
  - Use current deep powermodes
    - Long memory wake-up latency
With Wakeup = 1u sec

- E-D curves flat
- Can’t win with long wakeups
Faster Wakeups

Powerups should be much smaller

- 100ns
Faster Wakeups

Powerups should be much smaller

- 100ns
Outline

- Inefficiencies of DRAM interfaces

- Energy-proportionality via fast DRAM interfaces
  - MemBlaze
  - MemCorrect
  - MemDrowsy
Fast DRAM Wakeups

Enabling deep powerdown needs low-latency wakeups
Fast DRAM Wakeups

- Enabling deep powerdown needs low-latency wakeups
- Rearchitect interface to reduce wakeup latency
- Retain interface but powerdown aggressively
- Speculative wakeup with MemCorrect
- Lazy wakeup with MemDrowsy
Fast DRAM Wakeups

- Enabling deep powerdown needs low-latency wakeups
- Rearchitect interface to reduce wakeup latency
- Retain interface but powerdown aggressively

Speculative wakeup with MemCorrect
Lazy wakeup with MemDrowsy

Fast wakeup with MemBlaze
Fast DRAM Wakeups

- Enabling deep powerdown needs low-latency wakeups
- Rearchitect interface to reduce wakeup latency
- Retain interface but powerdown aggressively
- Fast wakeup with MemBlaze

Extra information:
- Speculative wakeup with MemCorrect
- Lazy wakeup with MemDrowsy
Fast Wakeup with MemBlaze

- No DLL
  - Periodic Timing reference signal stores DRAM offset in controller
  - Current-mode logic (CML) clocking has fewer variations

- Fast turn-on of datapath
  - Capacitive boosting quickly restores bias values
Fast Wakeup with MemBlaze

- No DLL
  - Periodic Timing reference signal stores DRAM offset in controller
  - Current-mode logic (CML) clocking has fewer variations
- Fast turn-on of datapath
  - Capacitive boosting quickly restores bias values

Exit latency ~ 10ns
MemBlaze DRAM + Controller

- Integrated into DRAMs. Fabricated and tested
- More details in the paper
Silicon Results
Methodology

- Workloads
  - Memcached
    - Key/value pairs with 100B and 10KB values
    - Zipf popularity distribution with exponential inter-arrival times
  - Yahoo! Cloud Benchmark (YCSB), SPECjbb
  - Multiprogrammed (MP) and Multithreaded (MT)
    - SPECCPU 2006, SPECOMP 2001, PARSEC
    - High BW (HB), Medium BW (MB), Low BW (LB)

- Architecture
  - 8 OoO Nehalem cores at 3GHz, 8MB shared L3 cache
  - 32 GB DRAM, 2Gb DDR3-1333 chips
  - Fast powerdown baseline, 15 cycle powerdown timer
MemBlaze Evaluation

- 66% lower memory energy with MemBlaze fastlock
- No performance penalty
Fast DRAM Wakeups

- Enabling deep powerdown needs low-latency wakeups
- Rearchitect interface to reduce wakeup latency
- Retain interface but powerdown aggressively
- Fast wakeup with MemBlaze

33
Fast DRAM Wakeups

Enabling deep powerdown needs low-latency wakeups

Rearchitect interface to reduce wakeup latency

Retain interface but powerdown aggressively

Fast wakeup with MemBlaze

Speculative wakeup with MemCorrect
Speculative Wakeup with MemCorrect

- Fast wakeup
  - Use deep power-down, which powers-off DLL, CLK
  - Transfer speculatively before the long DLL recalibration

- Error Detection/Correction
  - Detector fires if power-down period accumulated large skew
  - Corrector waits for recalibration before transfer
MemCorrect Evaluation

- Vary probability of correct timing ($p$)
- 40% energy savings (esp. for datacenters)
- Small $p$ $\Rightarrow$ Recalibration latency exposed
  - Degrades performance for high-BW apps
  - Increases energy/bit
Fast DRAM Wakeups

Enabling deep powerdown needs low-latency wakeups

Rearchitect interface to reduce wakeup latency

Retain interface but powerdown aggressively

Fast wakeup with MemBlaze

Speculative wakeup with MemCorrect
Fast DRAM Wakeups

- Enabling deep powerdown needs low-latency wakeups
  - Rearchitect interface to reduce wakeup latency
  - Retain interface but powerdown aggressively

  - Fast wakeup with MemBlaze
  - Speculative wakeup with MemCorrect
  - Lazy wakeup with MemDrowsy
Lazy Wakeup with MemDrowsy

- Fast wakeup
  - Wakeup from deep-powerdown
  - Transfer at lower rate before DLL recalibration completes

- Reduced Sampling Rate
  - Lower data rate for READs during calibration time (≈ 700ns)
  - Transfer each bit multiple times → Wider sampling window
  - Eliminates timing uncertainty
MemDrowsy Evaluation

- Vary sampling reduction rate (Z)
- 40% energy savings for datacenter apps
- High Z harms both performance and energy/bit
  - Energy per bit increases from wake-ups, higher bus activity
  - Z=2 more realistic
MemCorrect + MemDrowsy

- Combine MemCorrect and MemDrowsy
- If error detected, halve sampling rate instead of backoff
- ≤10% performance penalty
- 50% energy/bit savings
Conclusion

- DDR3 is energy-disproportional
  - DRAMs dissipate high static power
- DDR3 interfaces are efficiency bottlenecks
  - High active-idle power
  - Long wake-ups from power modes
- Re-architect interfaces with MemBlaze
- Or use MemCorrect + MemDrowsy
  - Provide fast wake-up from power modes
  - Energy efficiency improves by 40-70%
  - Performance impact is ≤ 10%
Thank you for your attention!

Questions?

ktej@stanford.edu