

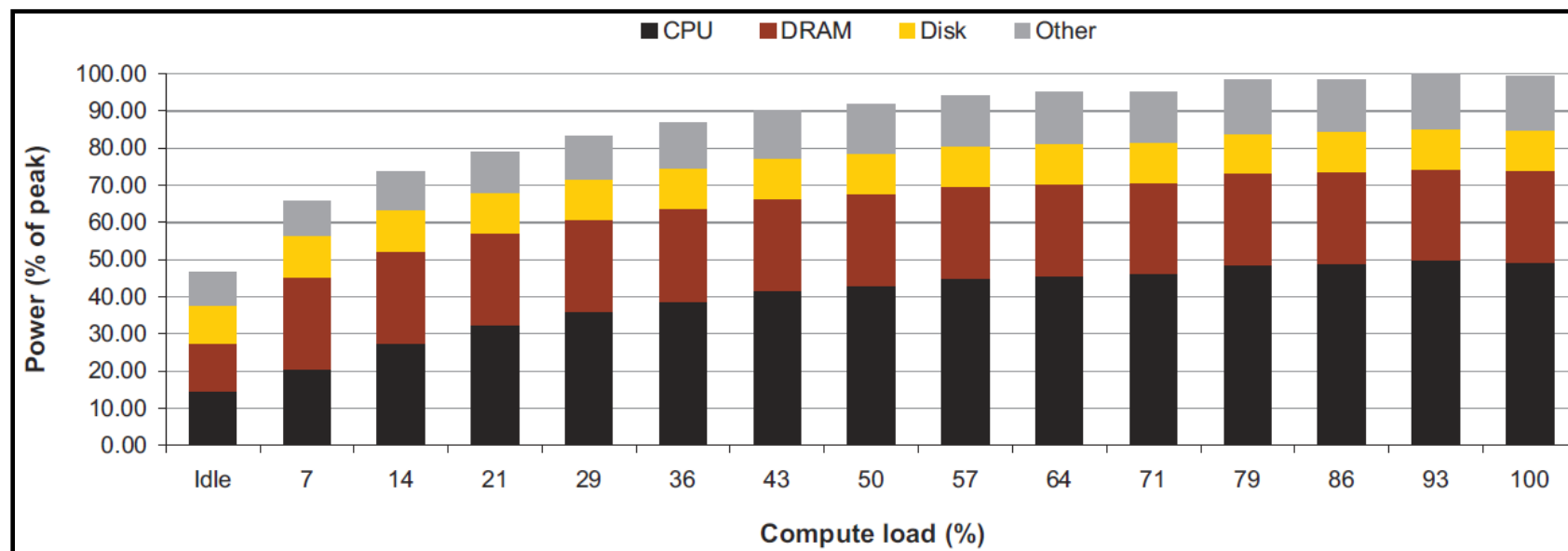
Rethinking DRAM Power Modes for Energy Proportionality

Krishna Malladi¹, Ian Shaeffer², Liji Gopalakrishnan²,
David Lo¹, Benjamin Lee³, Mark Horowitz¹

Stanford University¹, Rambus Inc², Duke University³

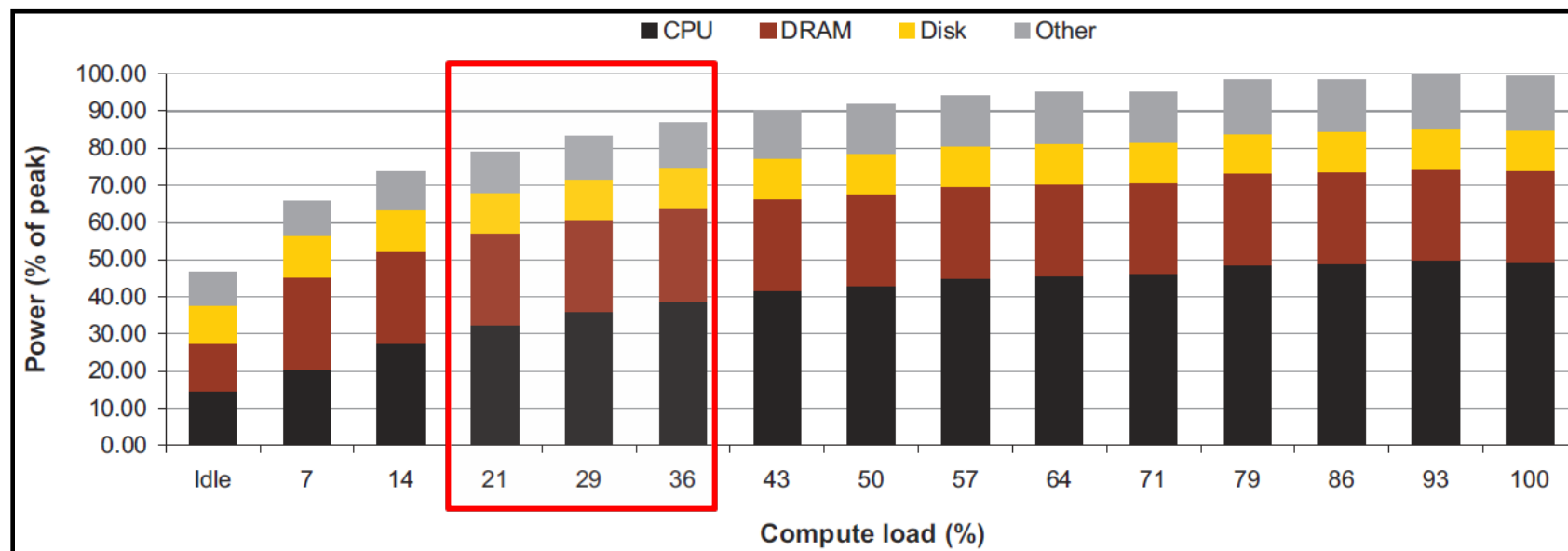
ktej@stanford.edu

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 \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

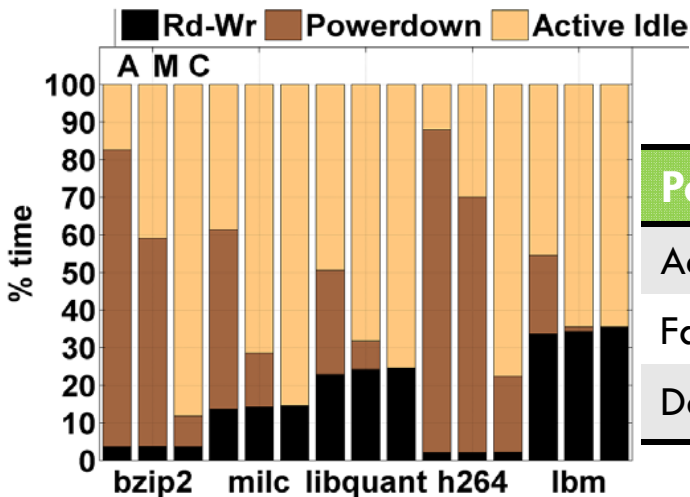
Outline

- 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



Power Mode	DIMM Idle Power (W)	Exit Latency (ns)
Active Idle	5.36	0
Fast Powerdown	2.79	20
Deep Powerdown	0.92	768

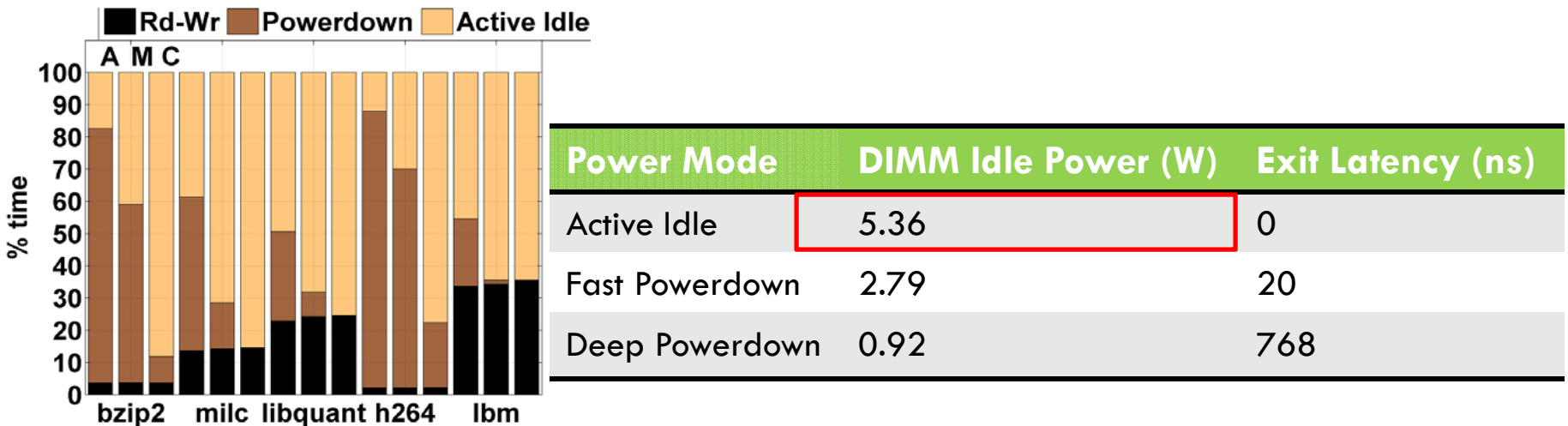
■ DDR3 optimized for high bandwidth

- 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

DDR3 Energy & Powermodes



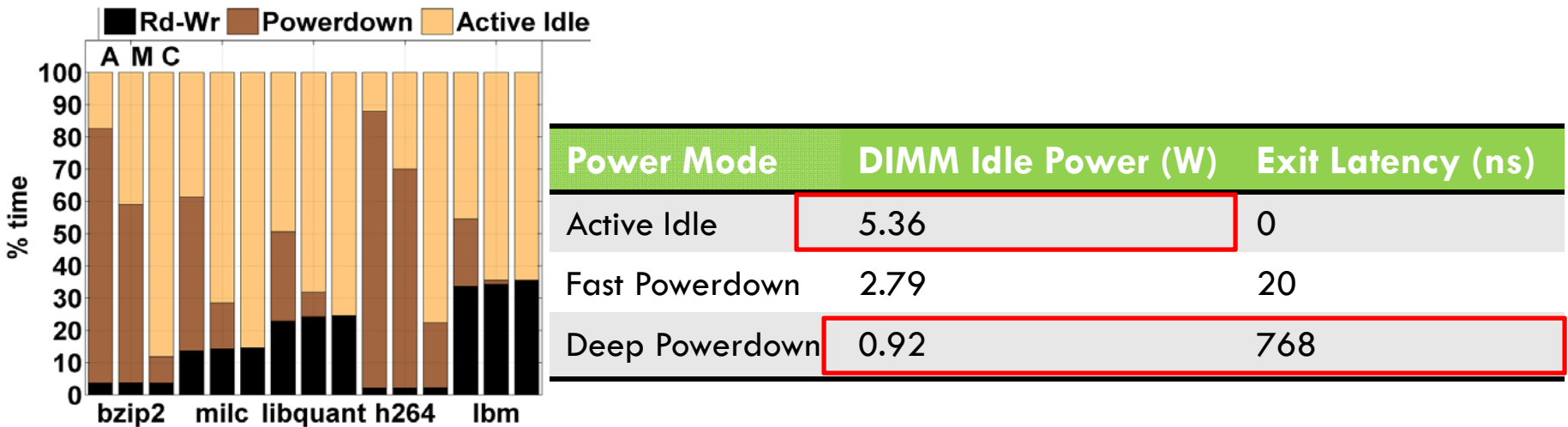
■ DDR3 optimized for high bandwidth

- 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

DDR3 Energy & Powermodes



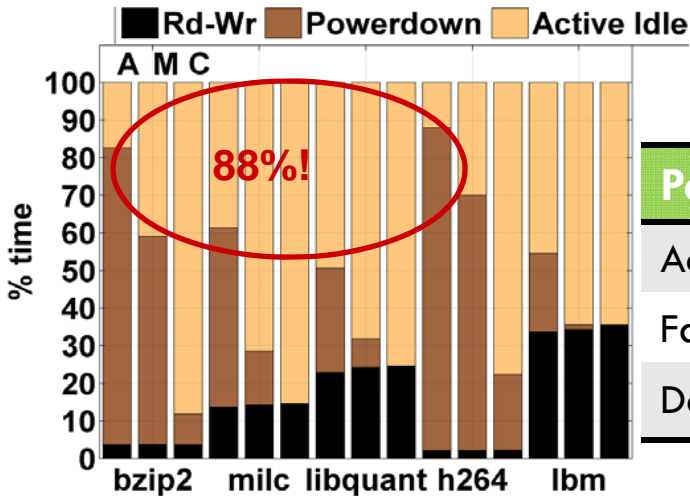
■ DDR3 optimized for high bandwidth

- 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

DDR3 Energy & Powermodes



Power Mode	DIMM Idle Power (W)	Exit Latency (ns)
Active Idle	5.36	0
Fast Powerdown	2.79	20
Deep Powerdown	0.92	768

■ DDR3 optimized for high bandwidth

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

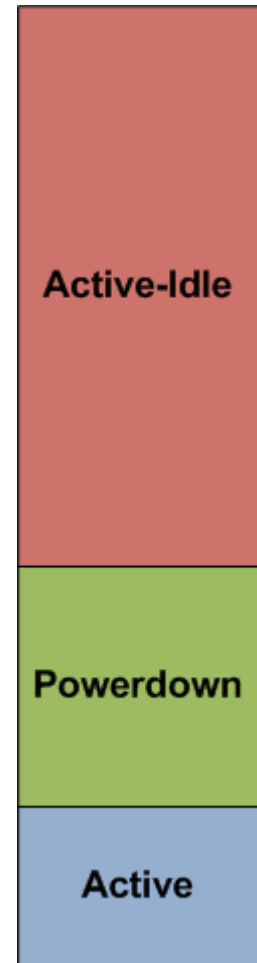
■ Hard to powerdown to deep states

- Long impractical wakeup time to power up interface
- Insufficient idleness in workloads → Significant active-idle time

Path to Energy-Proportionality

Path to Energy-Proportionality

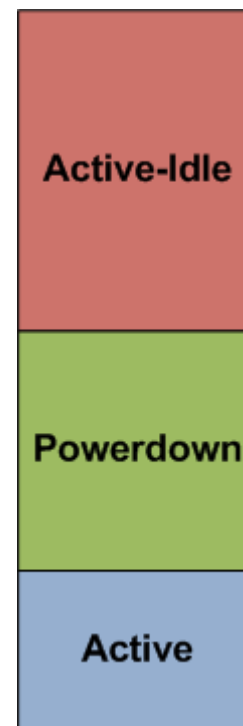
DRAM Power



Path to Energy-Proportionality

DRAM Power

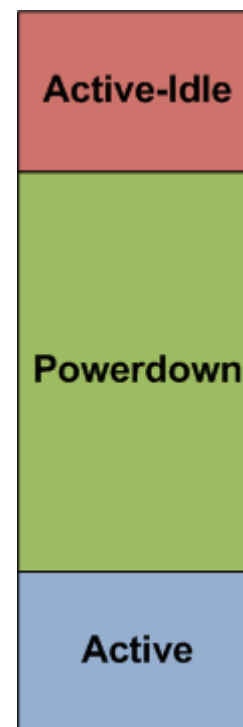
- Reduce active-idle power



Path to Energy-Proportionality

DRAM Power

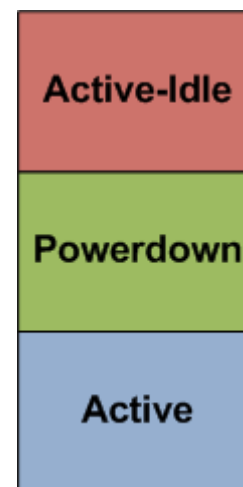
- Reduce active-idle power
- Reduce time in active-idle
- Increase time in power-down



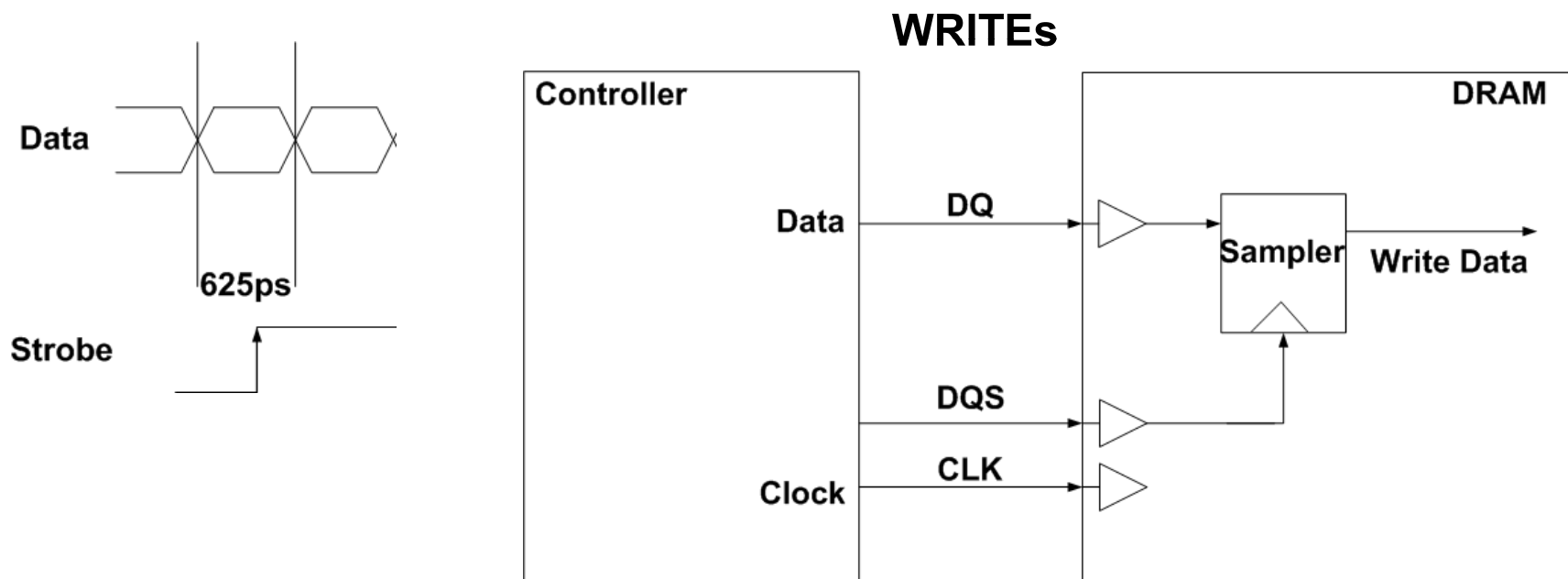
Path to Energy-Proportionality

DRAM Power

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

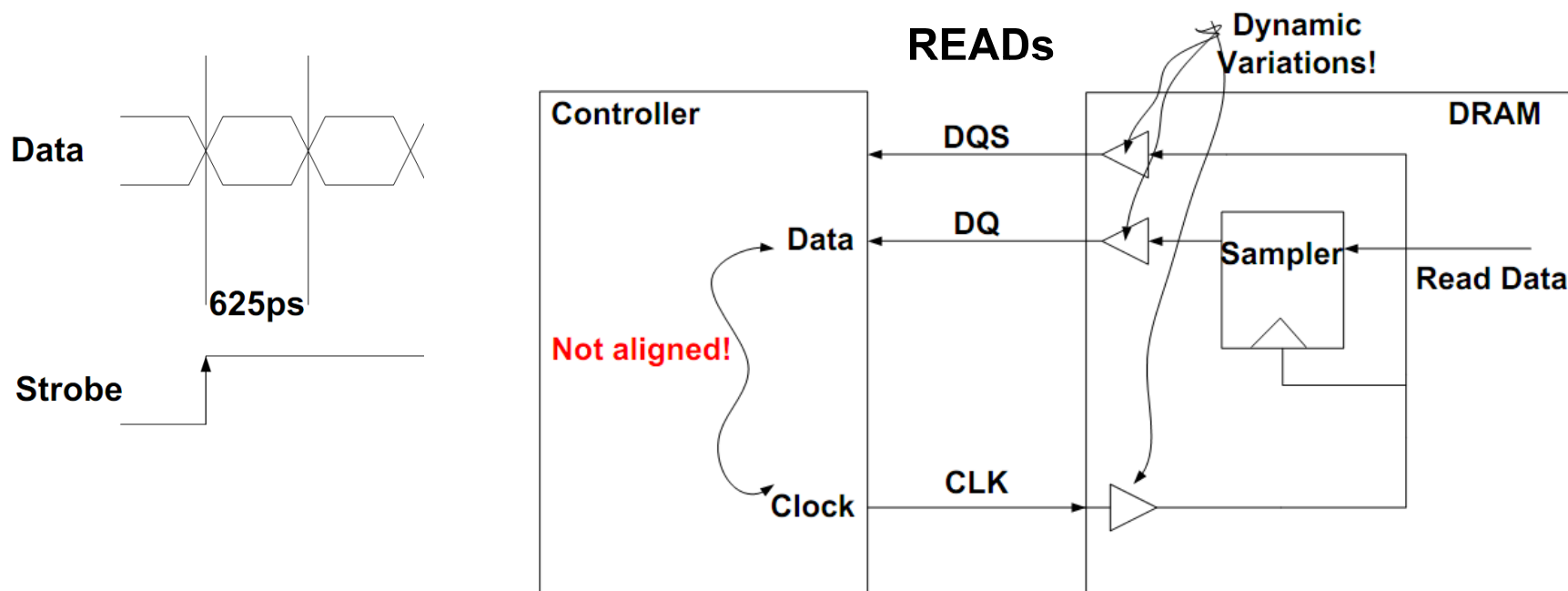


DRAM Interfaces



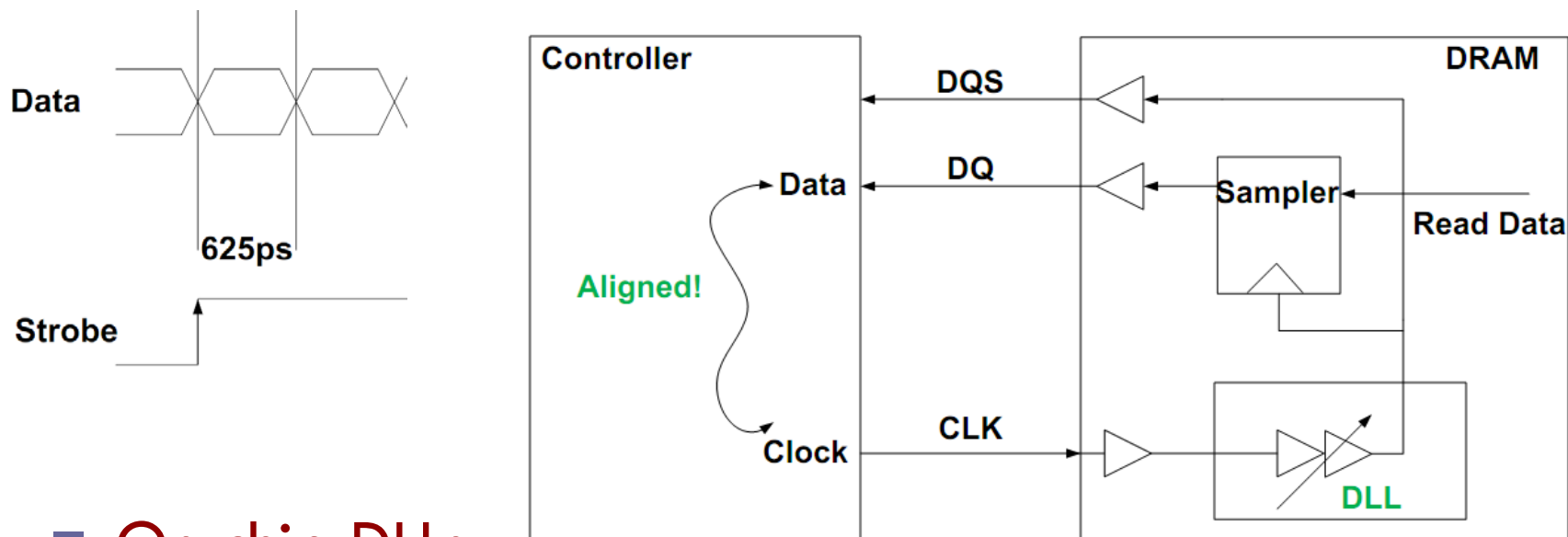
- Bits are short
 - Sampling window is only 625ps
- Data (DQ) and Clock (CLK) signals forwarded to DRAM
- Write data aligned to Clock edges

DRAM Interfaces



- **Dynamic chip variations affect Reads**
 - PVT variations → Misaligned DQS and CLK signals
 - Non-deterministic Read timing → 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 → poor powermodes

Live with Slow-Powerup

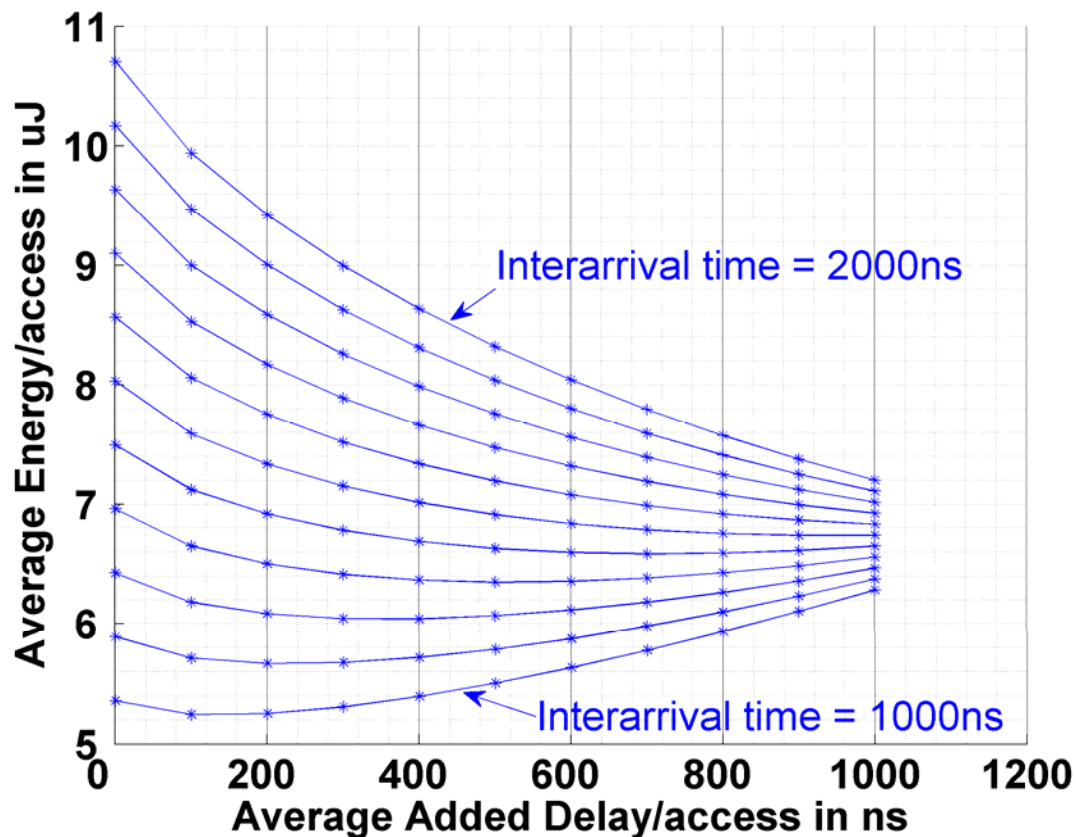
■ S/W mechanisms

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

■ H/W mechanisms

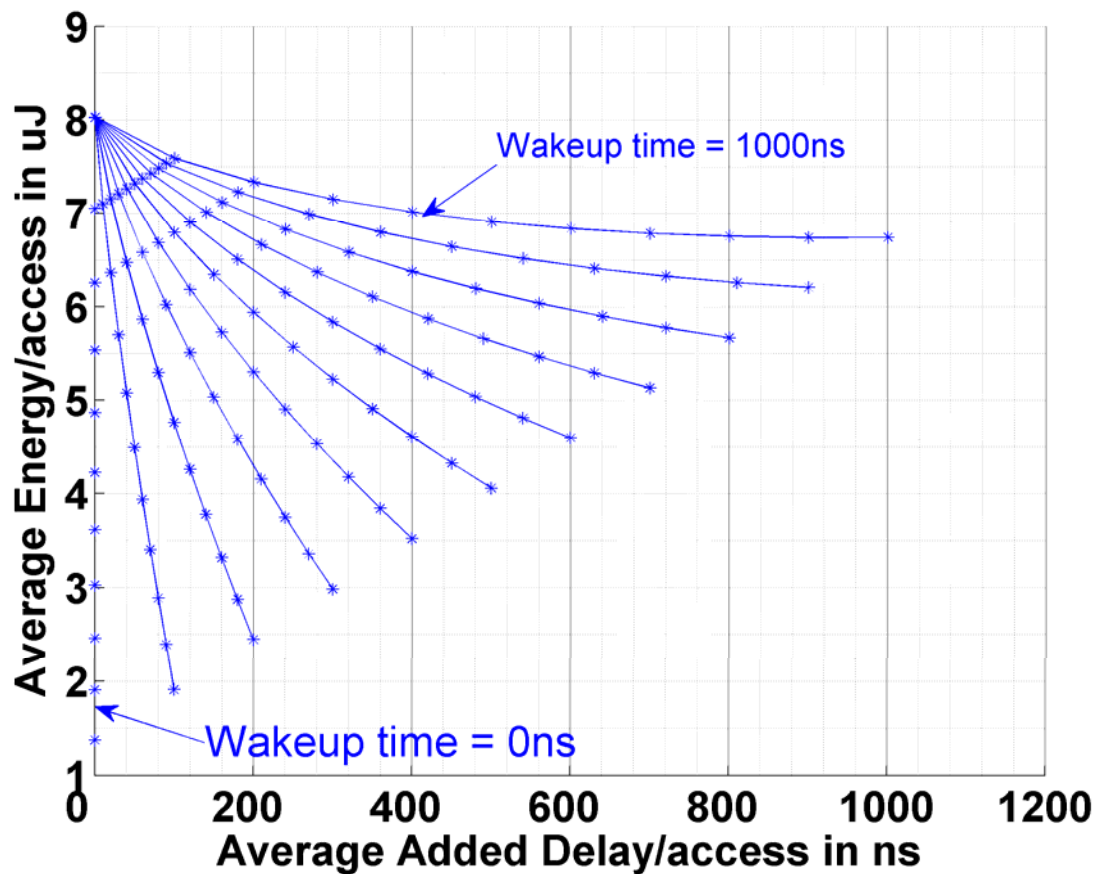
- Statically Disable DLLs in BIOS → Statically lowers bandwidth
 - Worse performance
- Use current deep powermodes
 - Long memory wake-up latency

With Wakeup = 1 μ sec



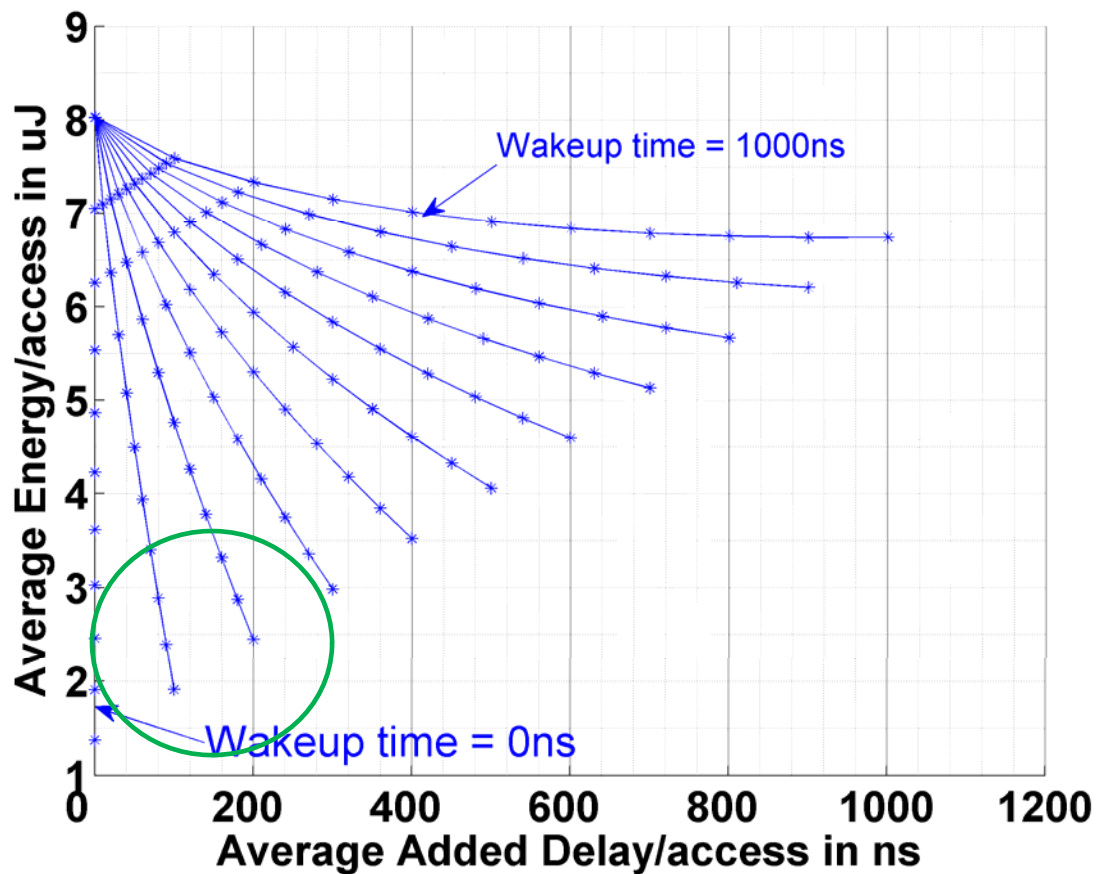
- E-D curves flat
- Can't win with long wakeups

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

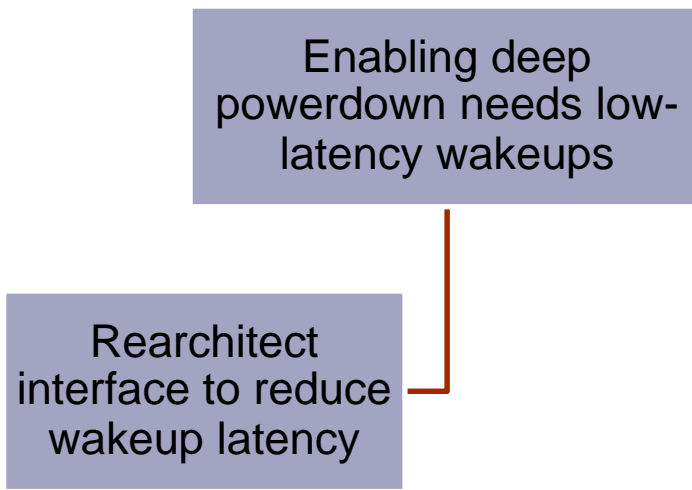
Fast DRAM Wakeups

Enabling deep
powerdown needs low-
latency wakeups

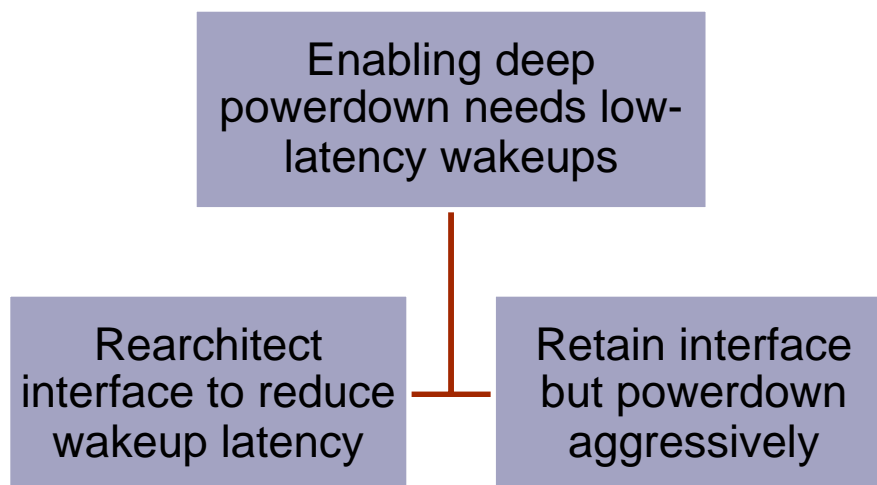
Fast DRAM Wakeups

Enabling deep
powerdown needs low-
latency wakeups

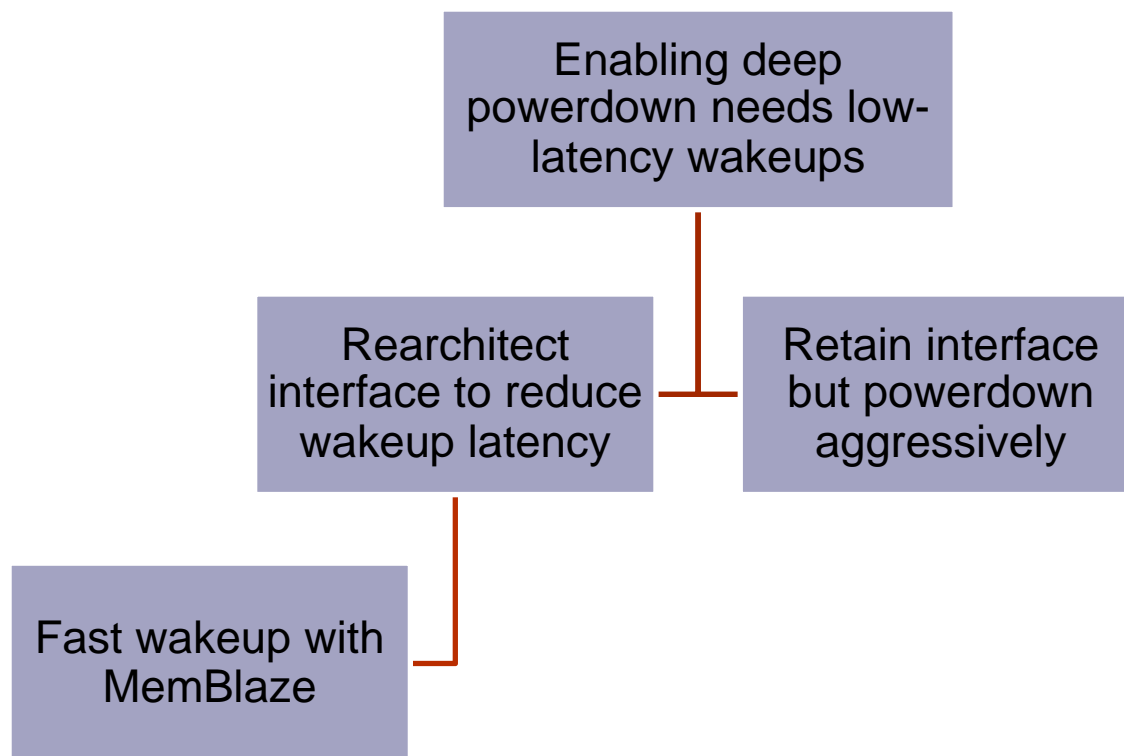
Rearchitect
interface to reduce
wakeup latency

A diagram consisting of two light blue rectangular boxes. The top box contains the text "Enabling deep powerdown needs low-latency wakeups". A thin brown line extends downwards from the bottom center of this box, then turns 90 degrees to the left, connecting to the right side of the bottom box. The bottom box contains the text "Rearchitect interface to reduce wakeup latency".

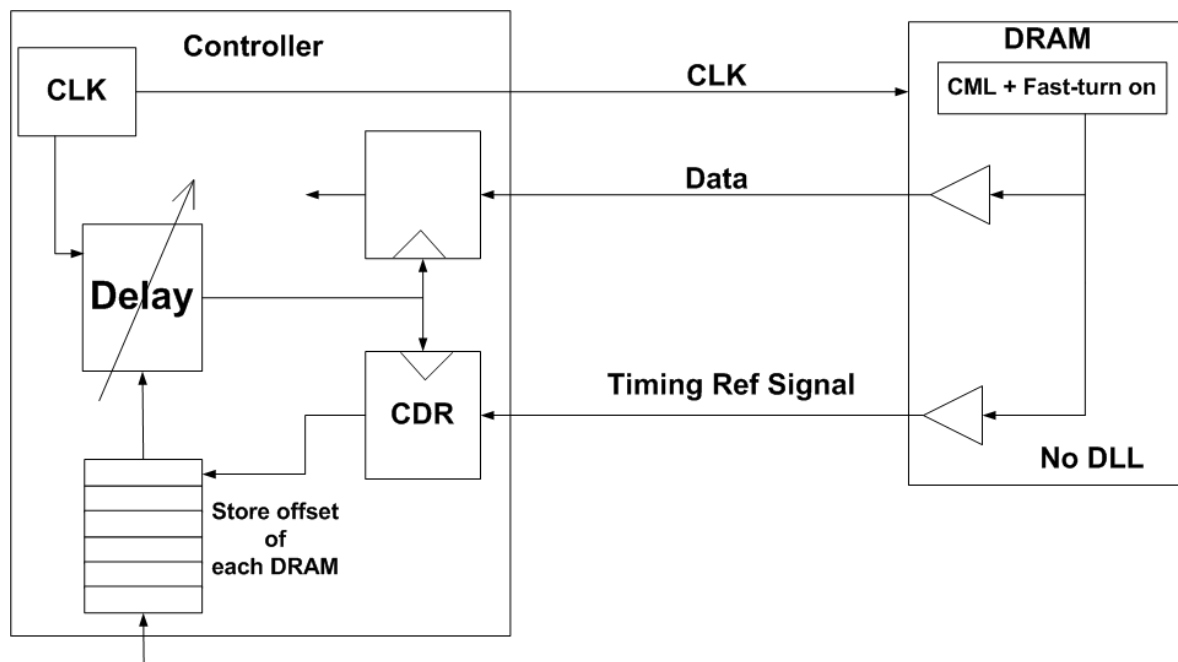
Fast DRAM Wakeups



Fast DRAM Wakeups



Fast Wakeup with MemBlaze



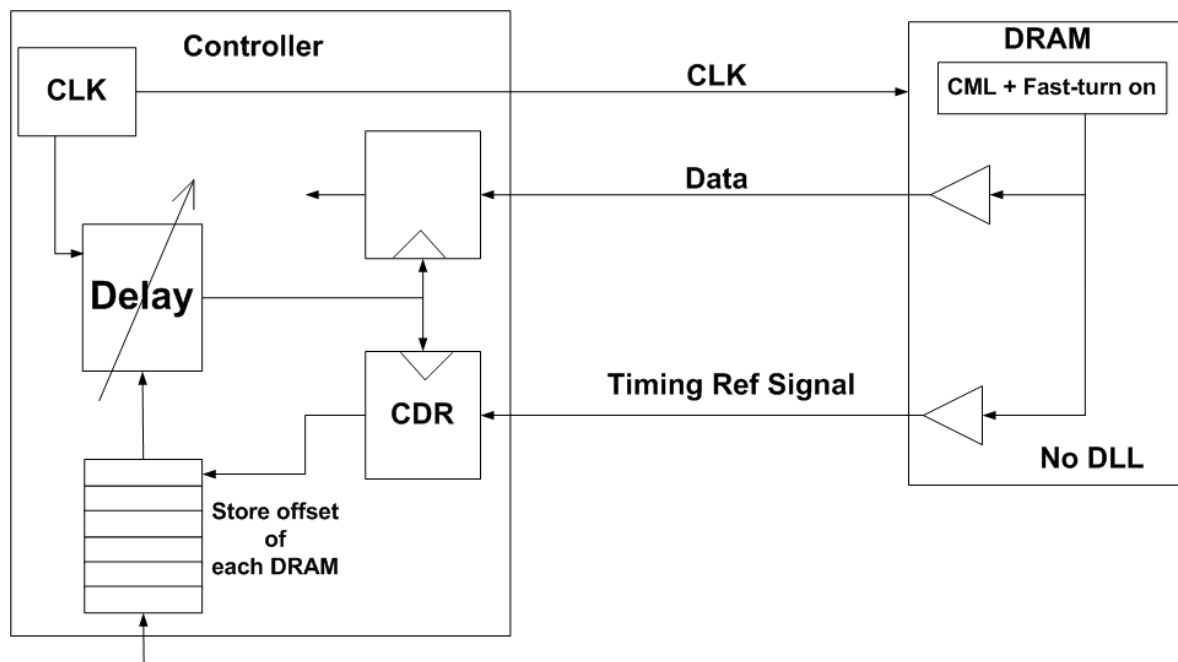
■ No DLL

- 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

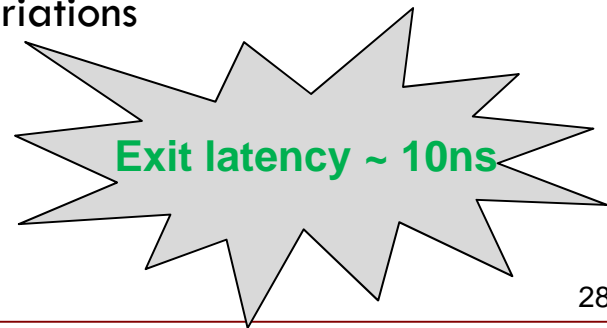


■ No DLL

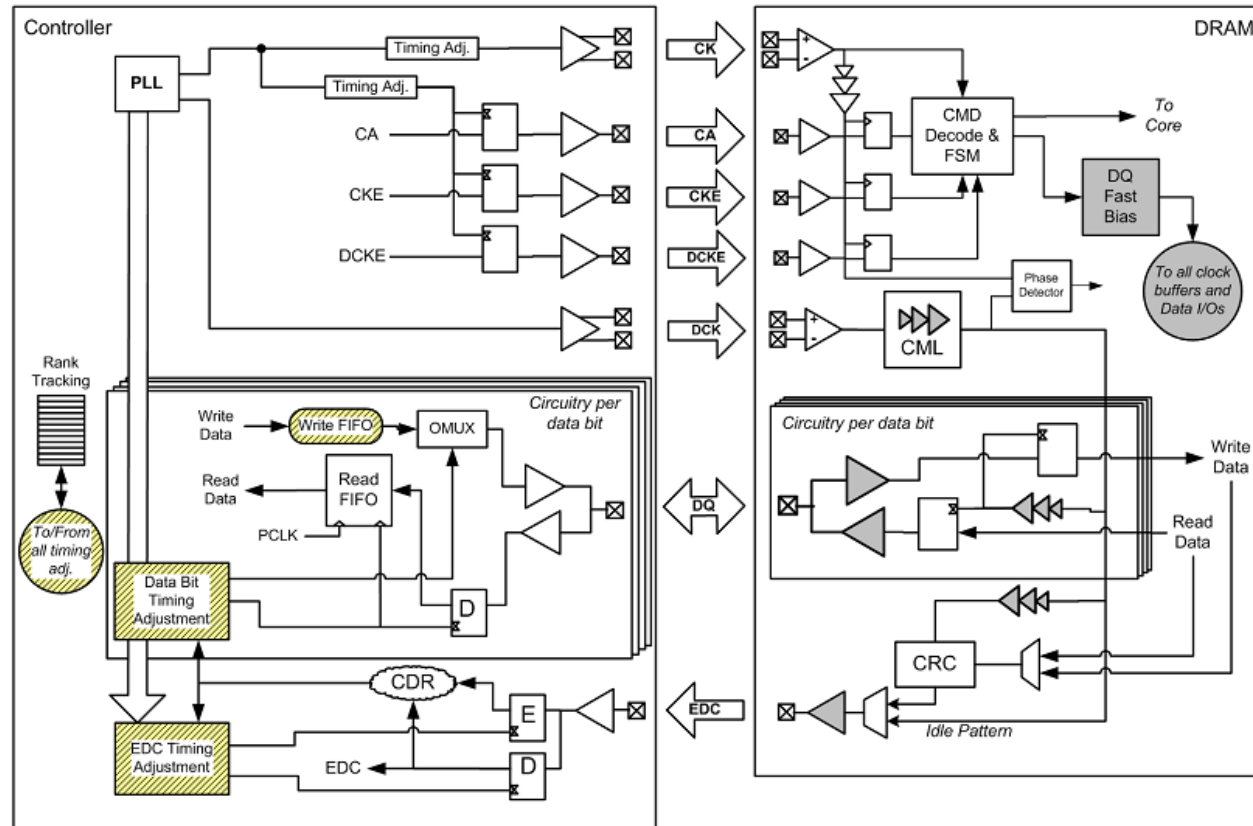
- Periodic Timing reference signal stores DRAM offset in controller
- Current-mode logic (CML) clocking has fewer variations

■ Fast turn-on of datapath

- Capacitive boosting quickly restores bias values

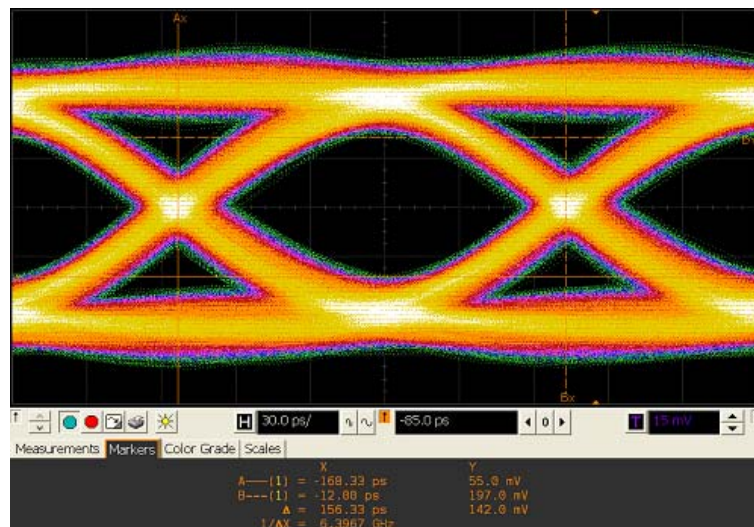
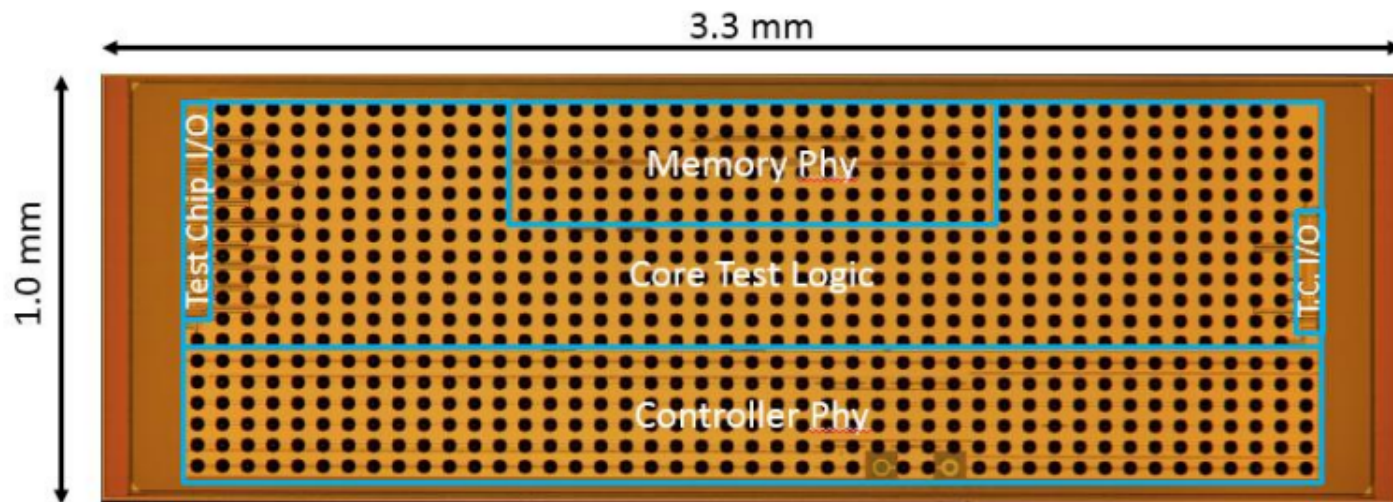


MemBlaze DRAM + Controller



- 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

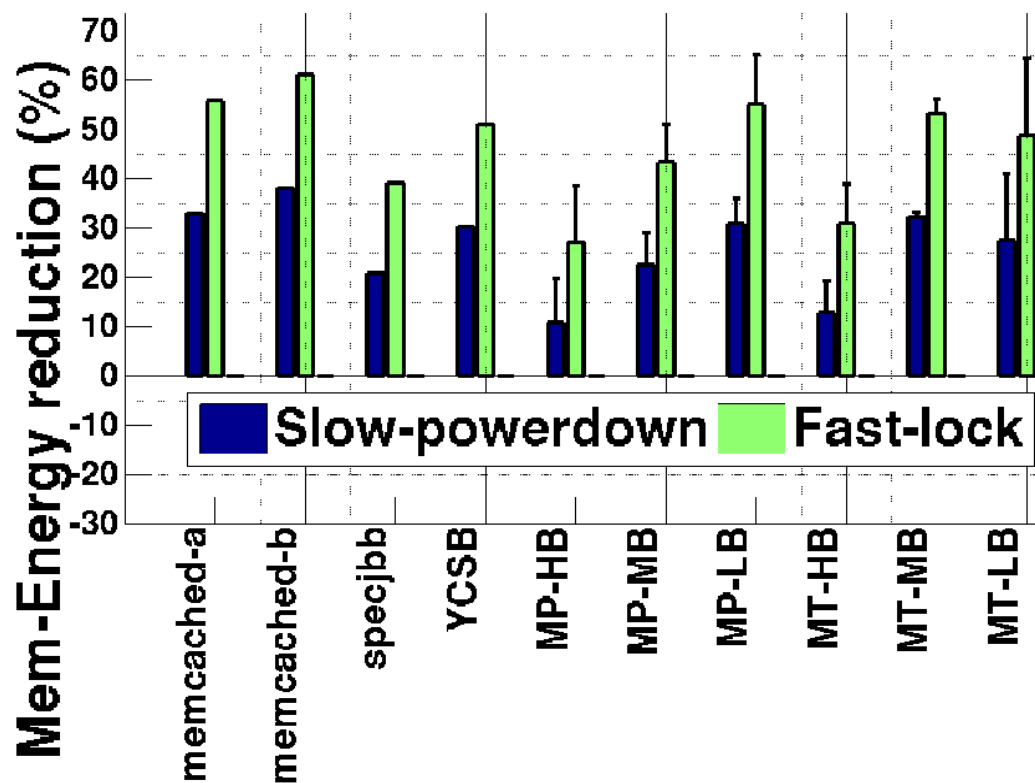
■ Multiprogrammed (MP) and Multithreaded (MT)

- SPECCPU 2006, SPECOMP 2001, PARSEC
- High BW (HB), Medium BW (MB), Low BW (LB)

■ Architecture

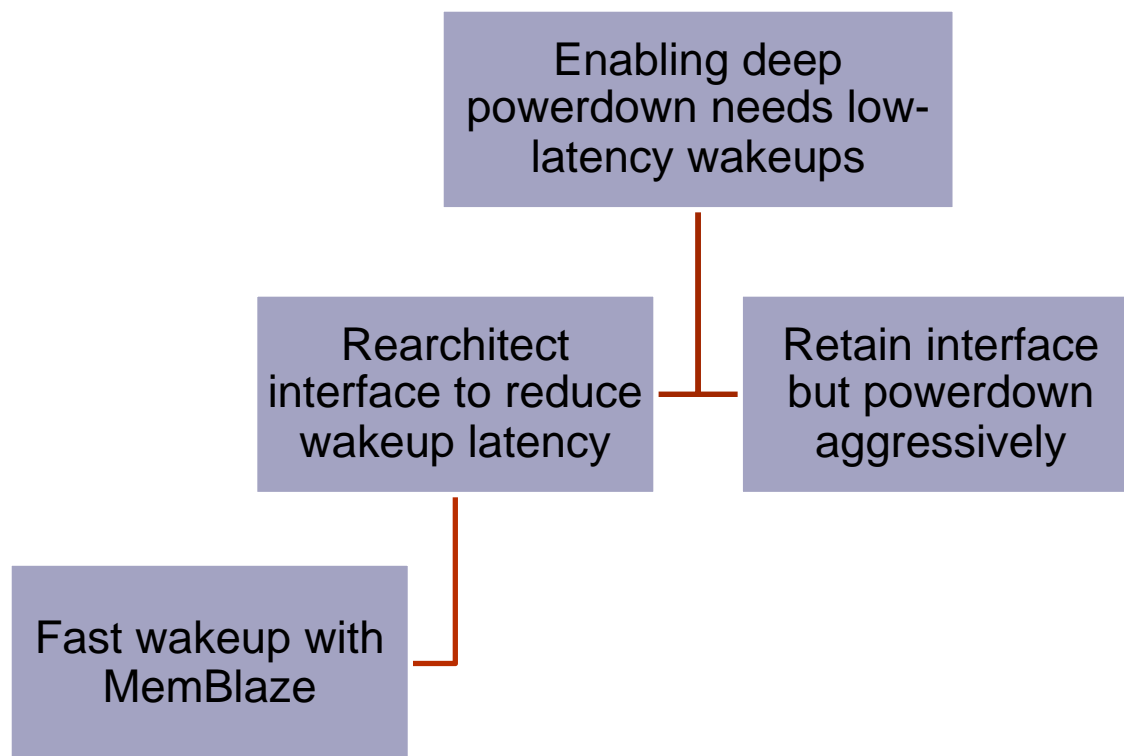
- 8 OoO Nehalem cores at 3GHz, 8MB shared L3 cache
- 32 GB DRAM, 2Gb DDR3-1333 chips
- Fast powerdown baseline, 15 cycle powerdown timer

MemBlaze Evaluation

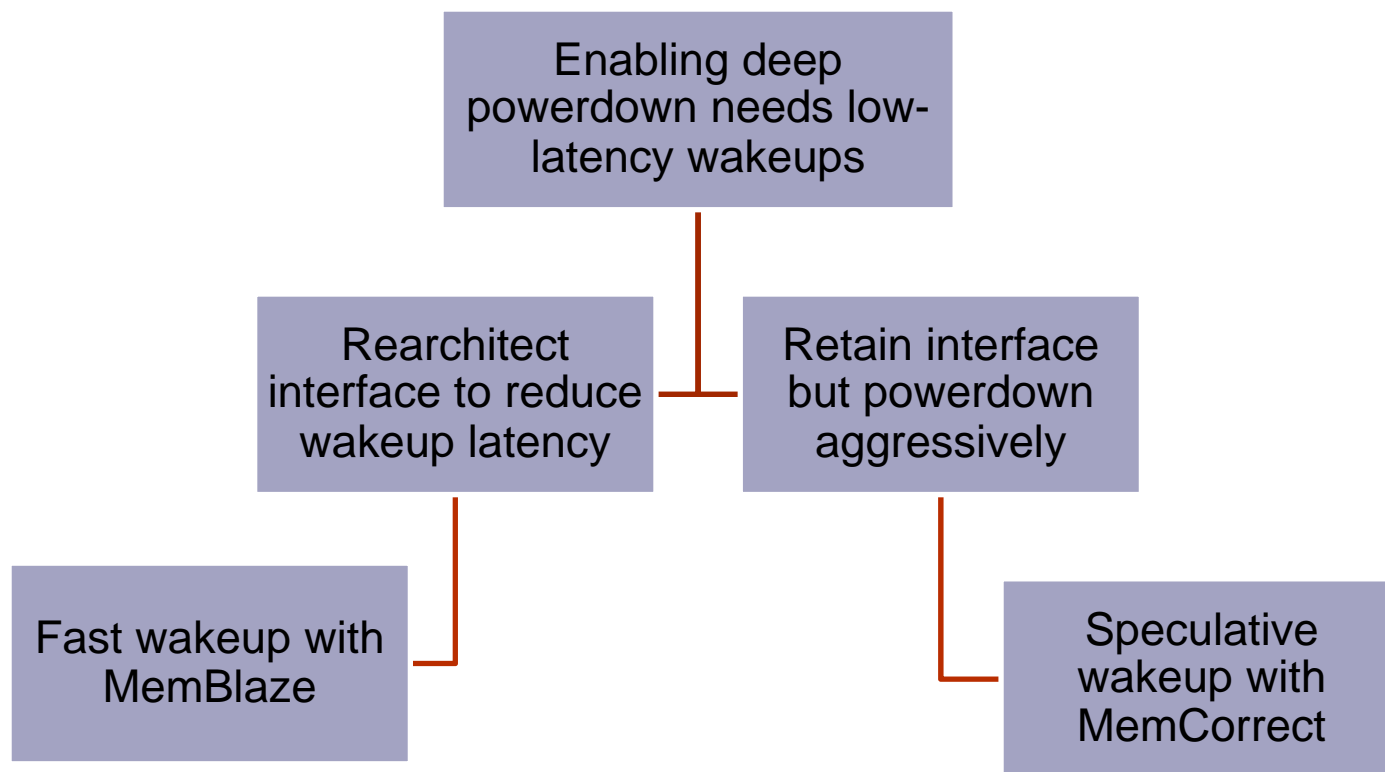


- 66% lower memory energy with MemBlaze fastlock
- No performance penalty

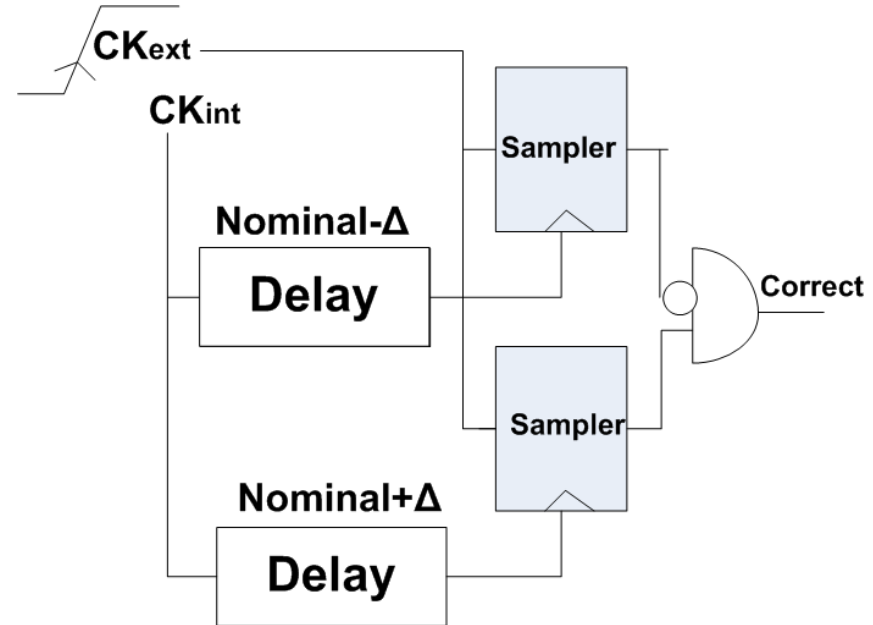
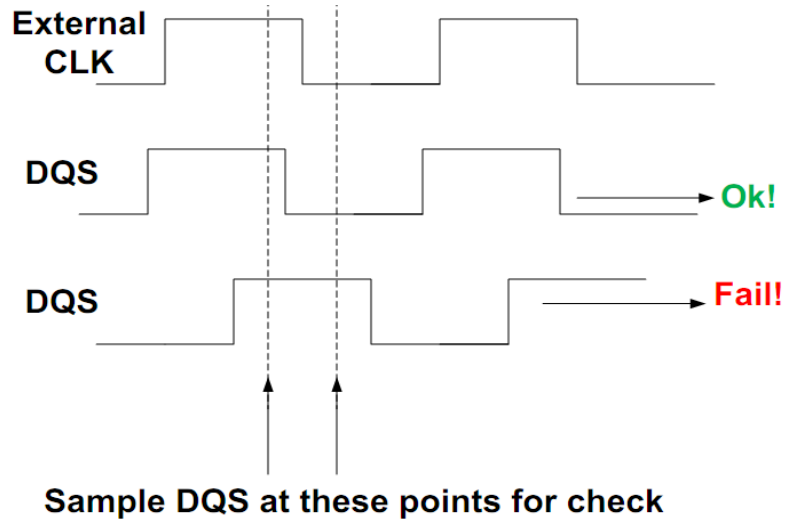
Fast DRAM Wakeups



Fast DRAM Wakeups



Speculative Wakeup with MemCorrect



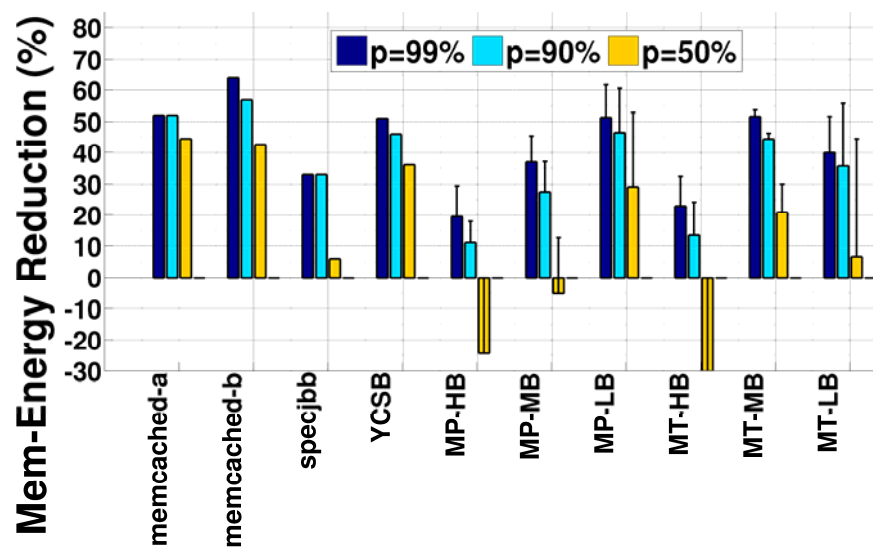
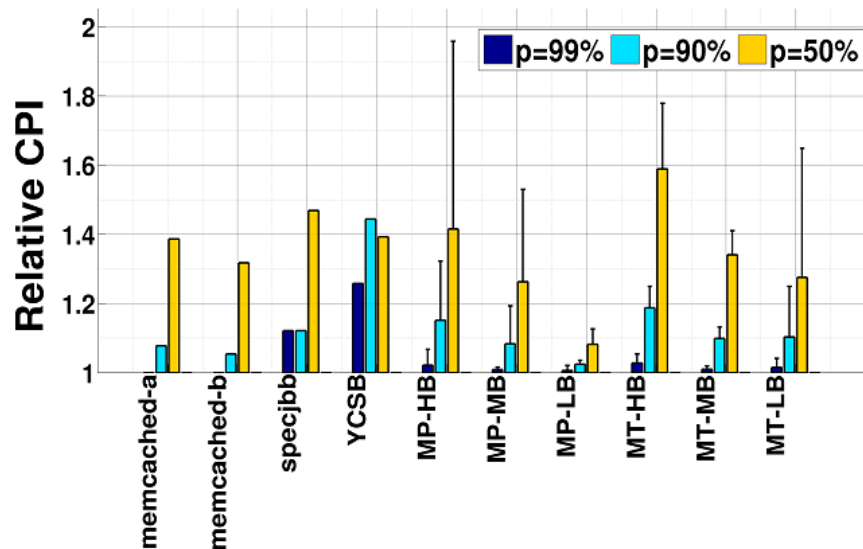
■ Fast wakeup

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

■ Error Detection/Correction

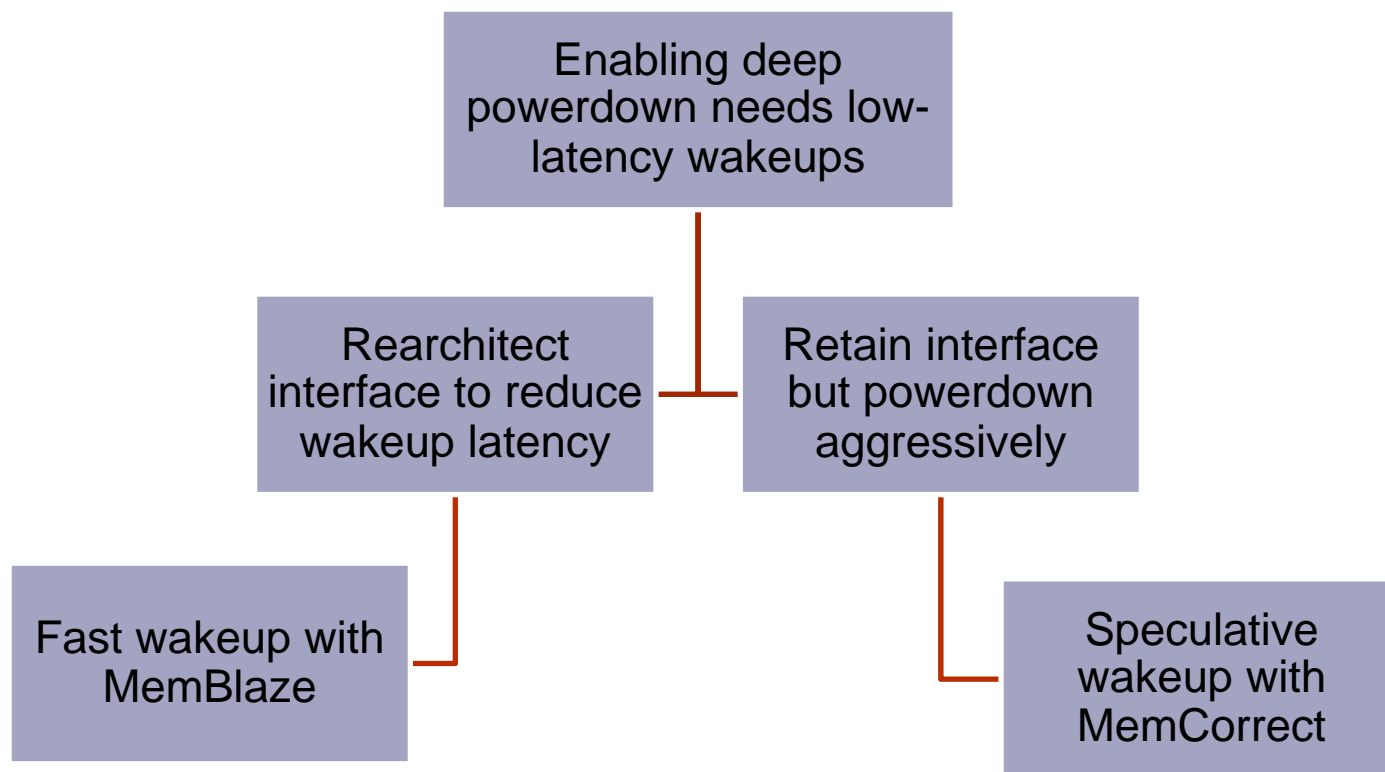
- Detector fires if power-down period accumulated large skew
- Corrector waits for recalibration before transfer

MemCorrect Evaluation

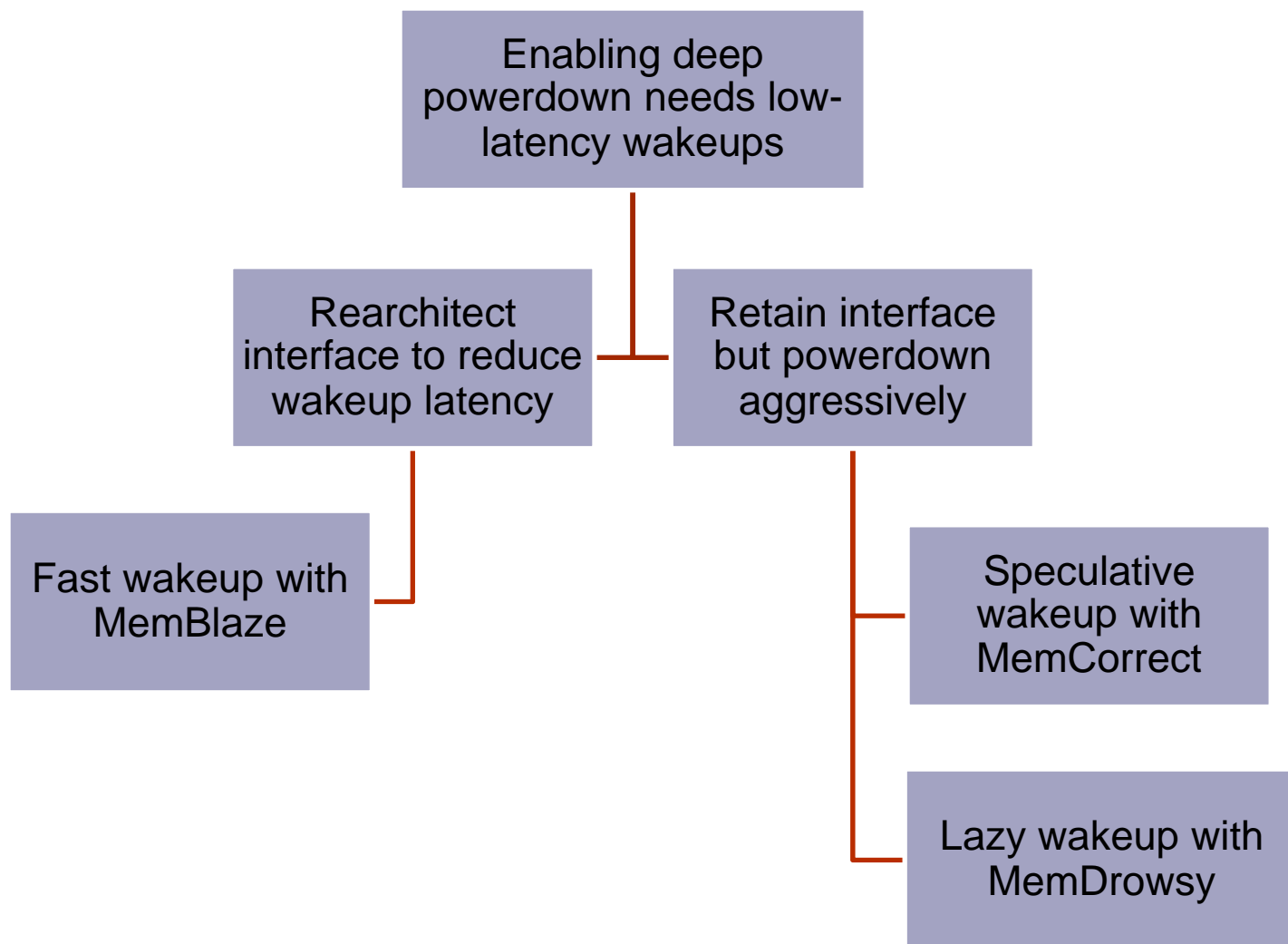


- Vary probability of correct timing (p)
- 40% energy savings (esp. for datacenters)
- Small $p \rightarrow$ Recalibration latency exposed
 - Degrades performance for high-BW apps
 - Increases energy/bit

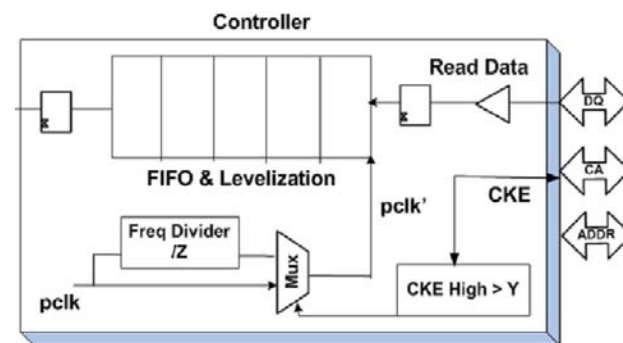
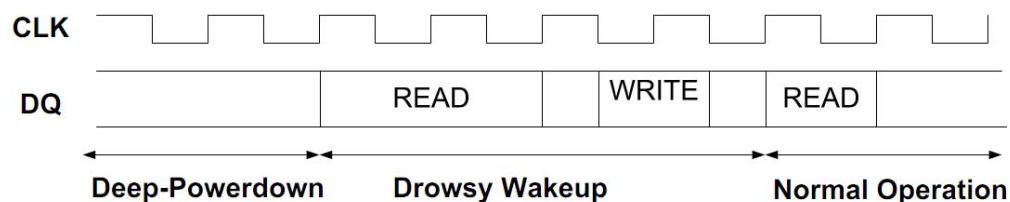
Fast DRAM Wakeups



Fast DRAM Wakeups



Lazy Wakeup with MemDrowsy



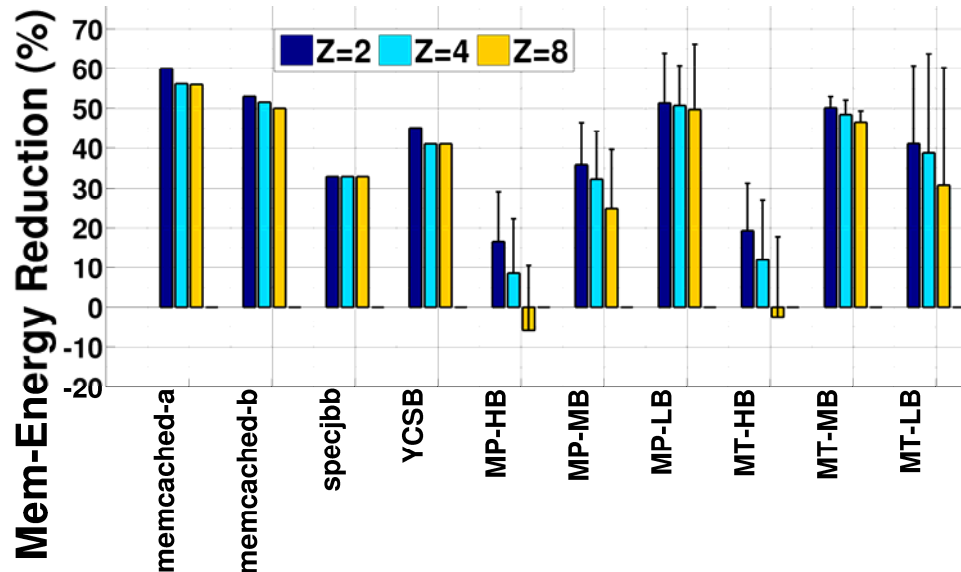
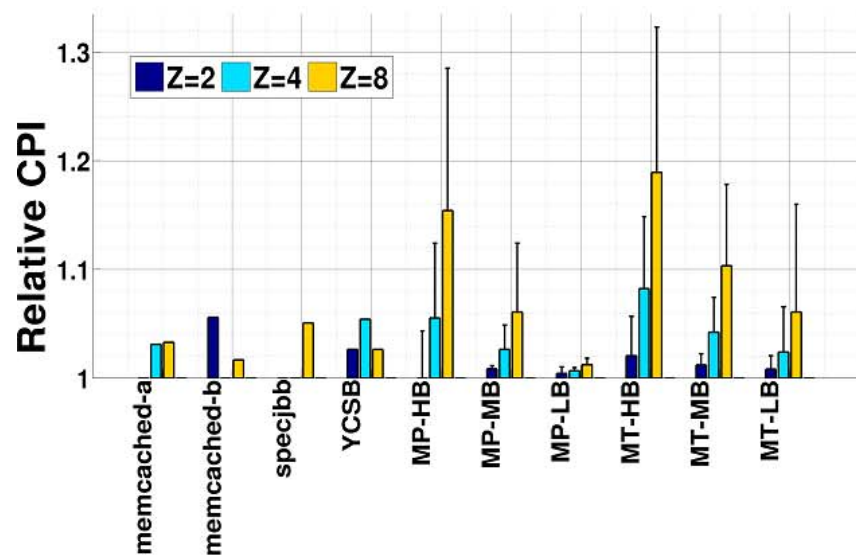
■ Fast wakeup

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

■ Reduced Sampling Rate

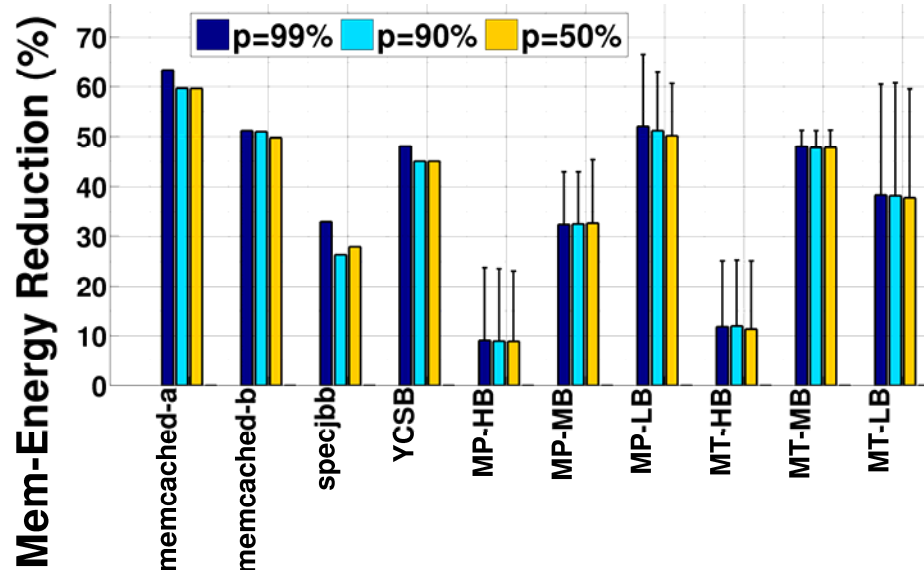
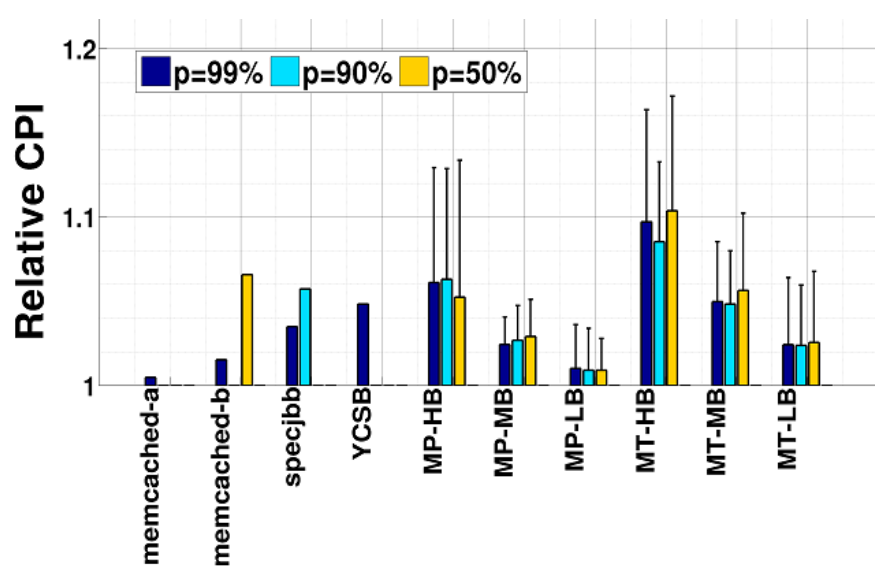
- Lower data rate for READs during calibration time ($\sim 700\text{ns}$)
- Transfer each bit multiple times \rightarrow Wider sampling window
- Eliminates timing uncertainty

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
- $\leq 10\%$ 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 $\leq 10\%$

Thank you for your attention!

Questions?

ktej@stanford.edu