Tapping ZettaRAM™ for Low-Power Memory Systems

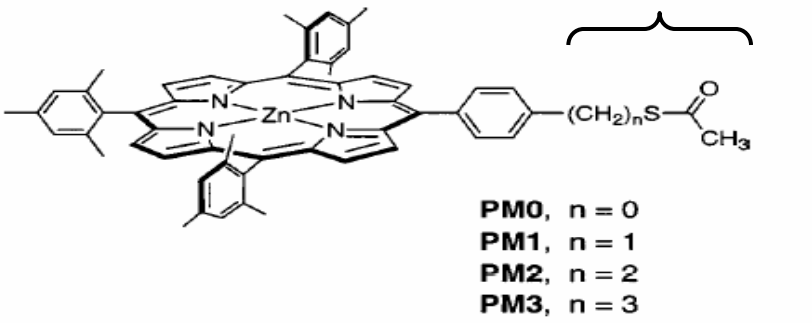
Ravi K. Venkatesan, Ahmed S. AL-Zawawi, Eric Rotenberg

Center for Embedded Systems Research (CESR)
Department of Electrical & Computer Engineering
North Carolina State University

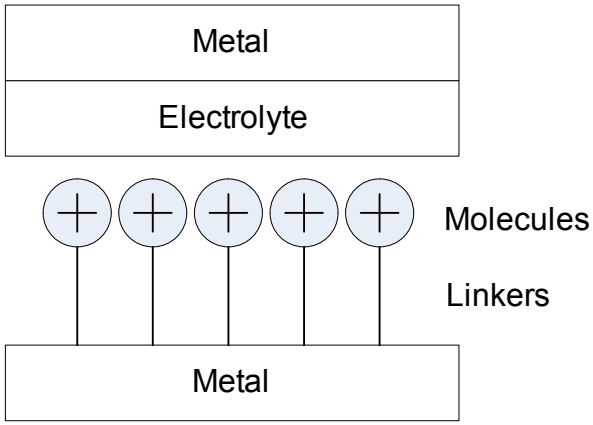
ZettaRAM

- New memory from ZettaCore™
  - Genesis in DARPA Moletronics
  - Molecule stores 1 charge
- Long-term
  - 1 molecule = 1 bit
- Near-Term
  - Use molecules in aggregate
  - Replace DRAM capacitor

Porphyrin Molecule



ZettaRAM Molecular Capacitor



Benefits

- Manufacturing
  - Self–assembly
  - High charge density
  - Precise control of molecules' attributes

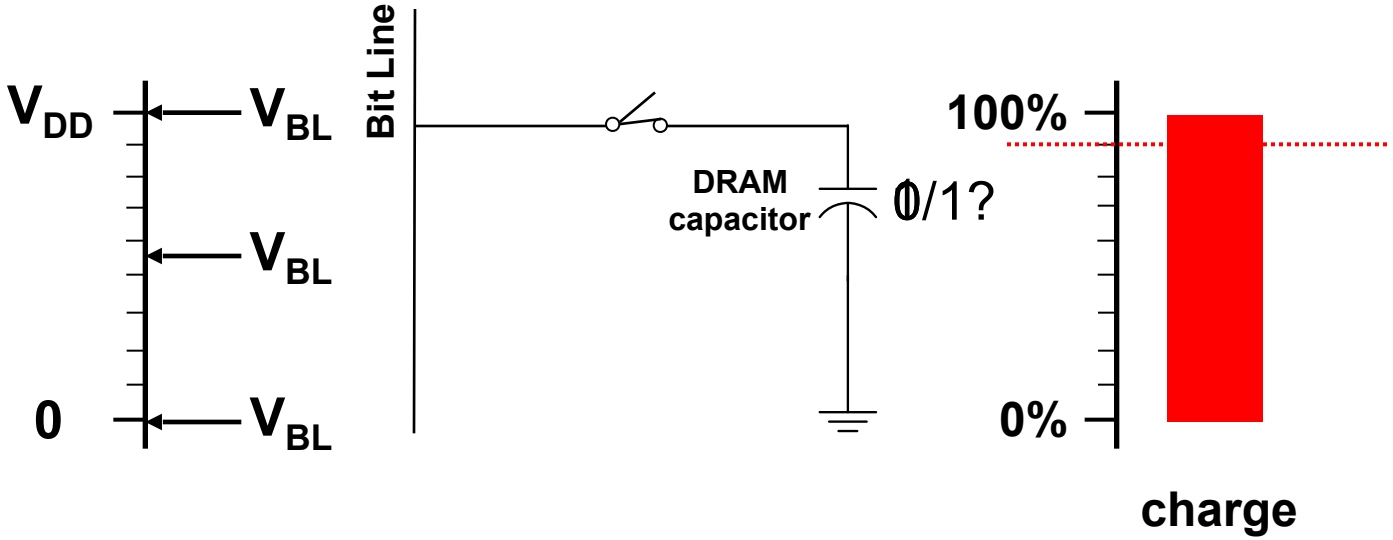
cost-effectively scale DRAM density

flexibly control density, performance, energy

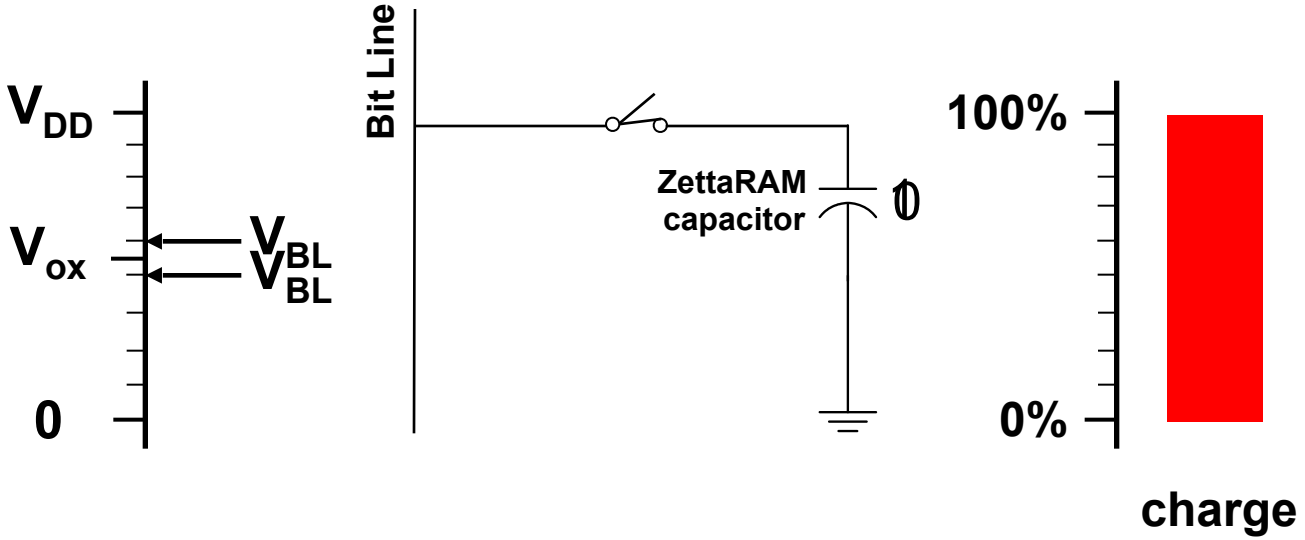
- Other unusual properties?

- Architectural opportunities?

DRAM



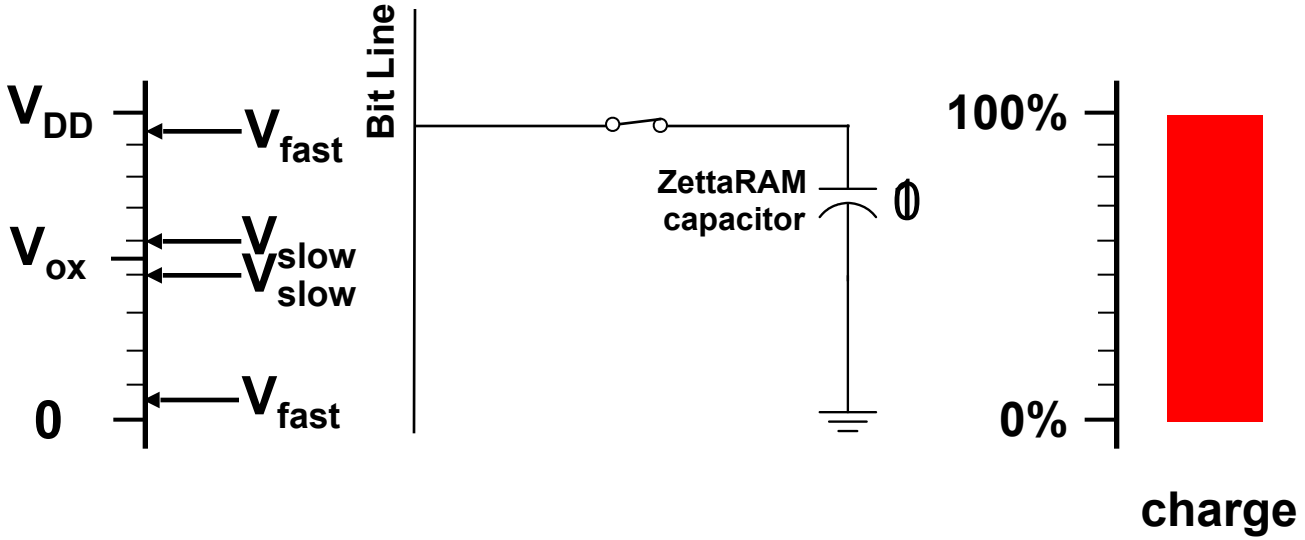
ZettaRAM



$$E_{bitline} = C_{BL} \cdot V_{BL} \cdot \Delta V_{BL}$$

Very low-power memory

But…

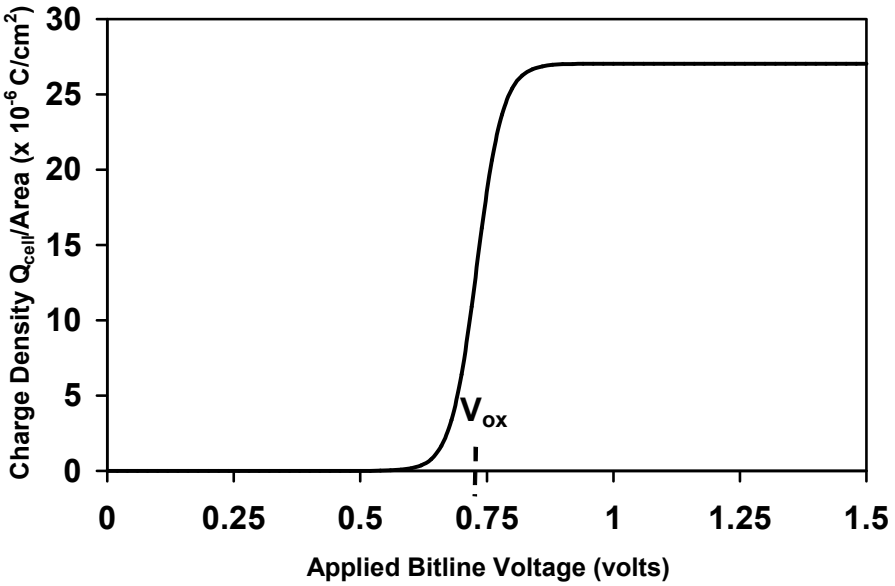


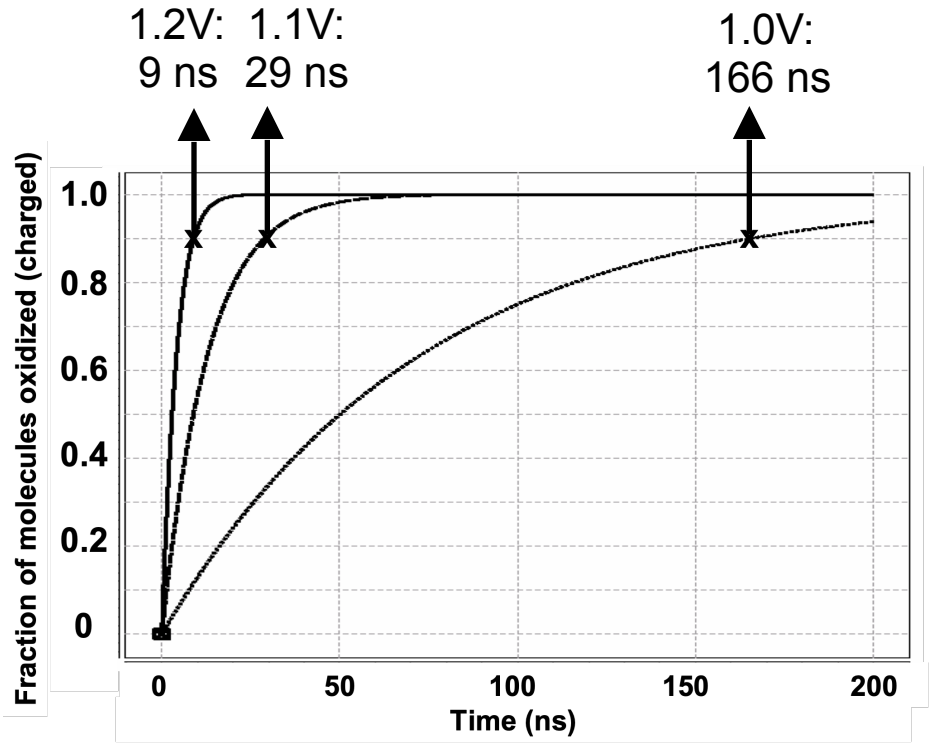
Speed-energy tradeoff

SPICE Results

➢ SPICE model based on Butler-Volmer equation

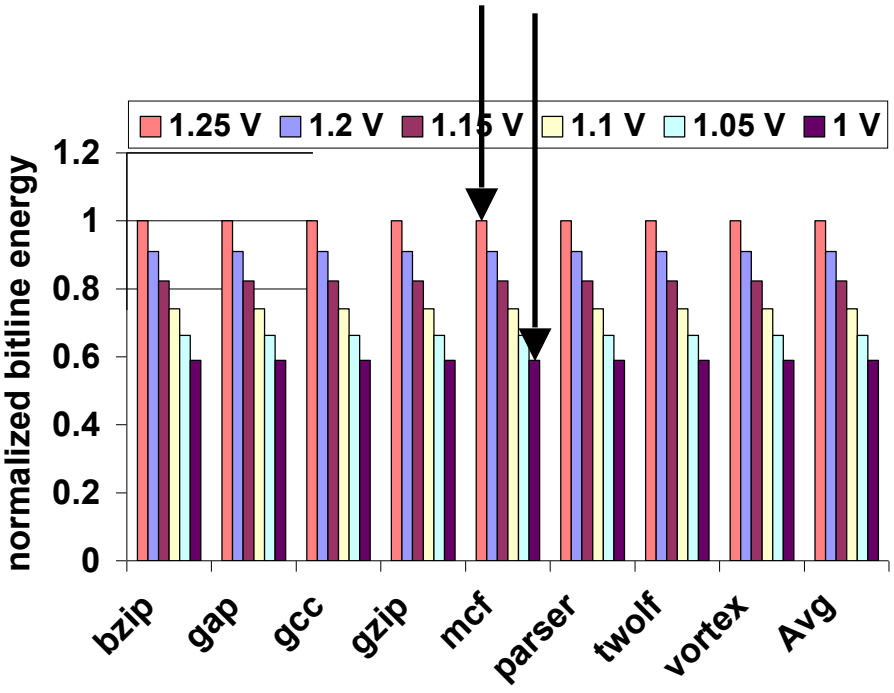
$$I \propto e^{\pm k(V_{BL} - V_{ox})}$$



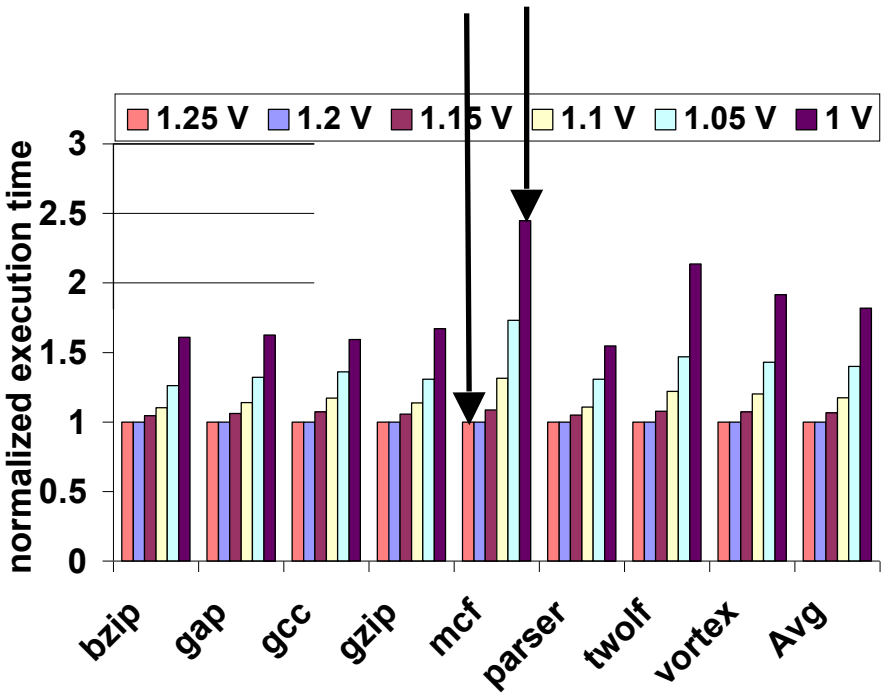


ZettaRAM : Performance-Energy Tradeoff

Bitline energy w.r.t. DRAM



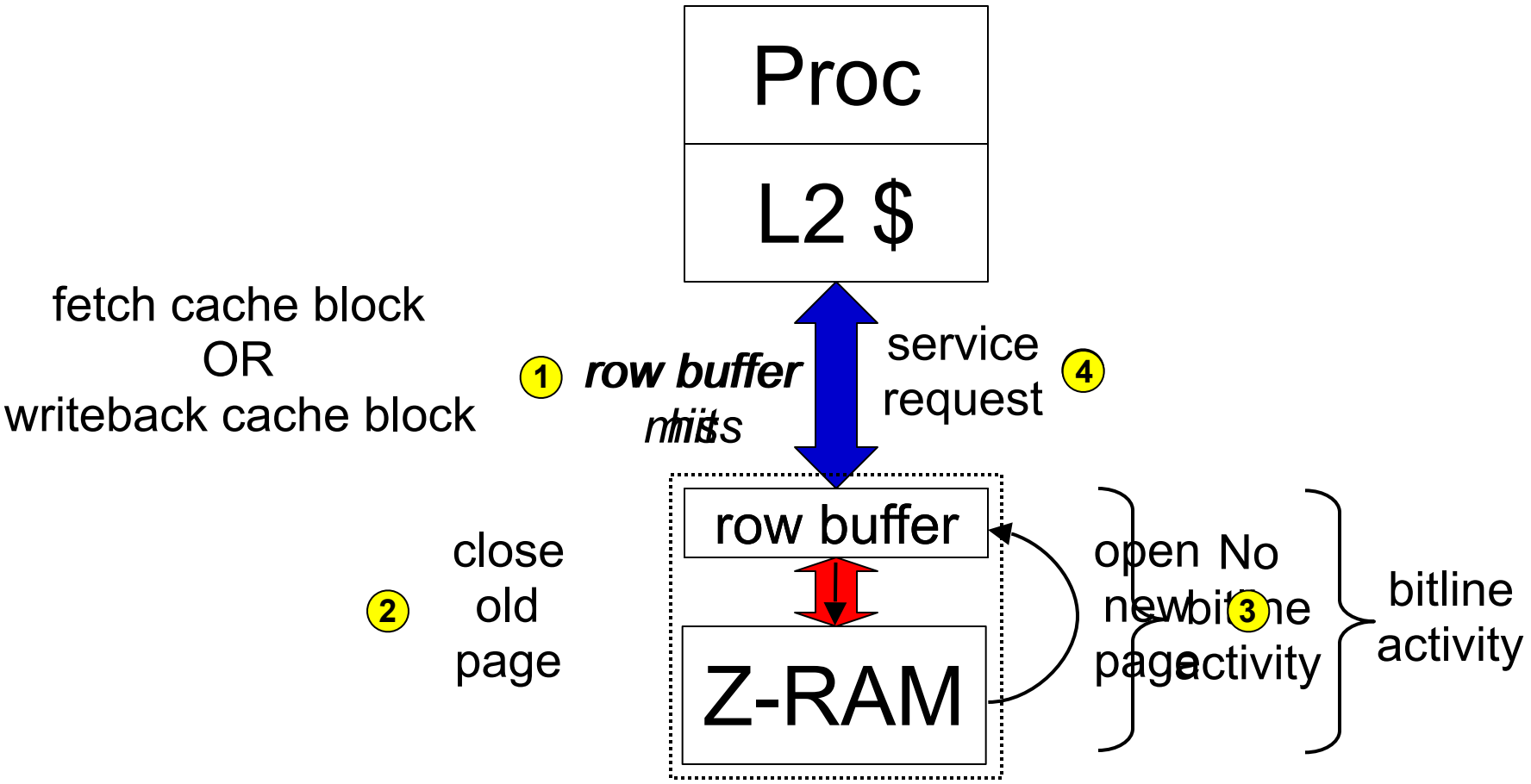
Execution time w.r.t. DRAM



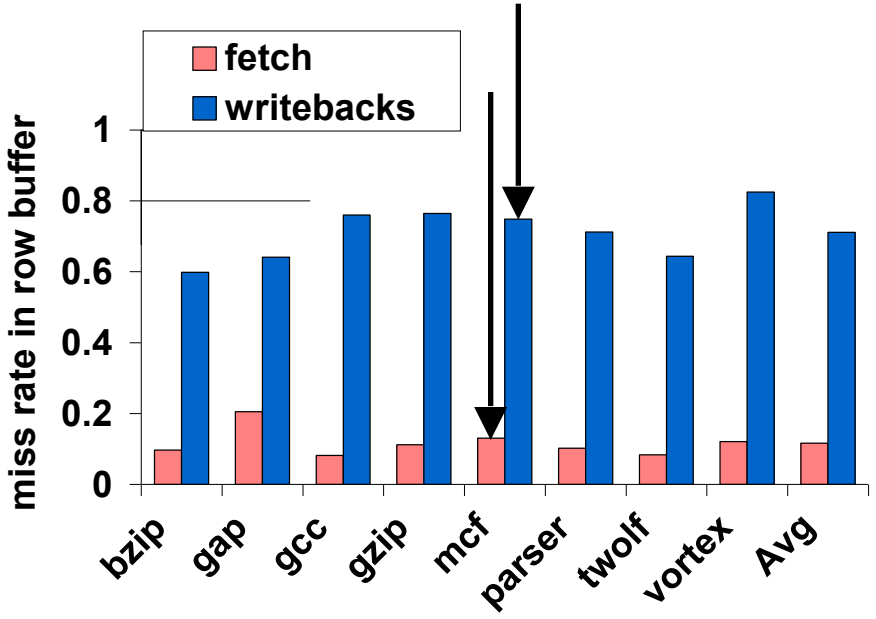
Goal

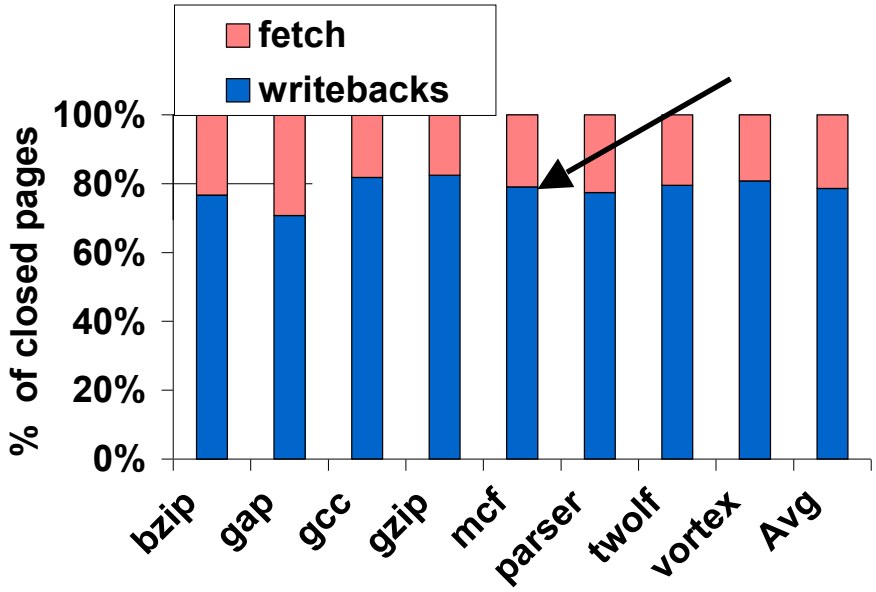
- ZettaRAM offers novel performance-energy tradeoff
  - DRAM inflexible
- Architectural techniques to manage main memory
  - Tap the energy-savings potential
  - Minimize performance degradation

Source of Bitline Activity



Row Buffer Miss Rates



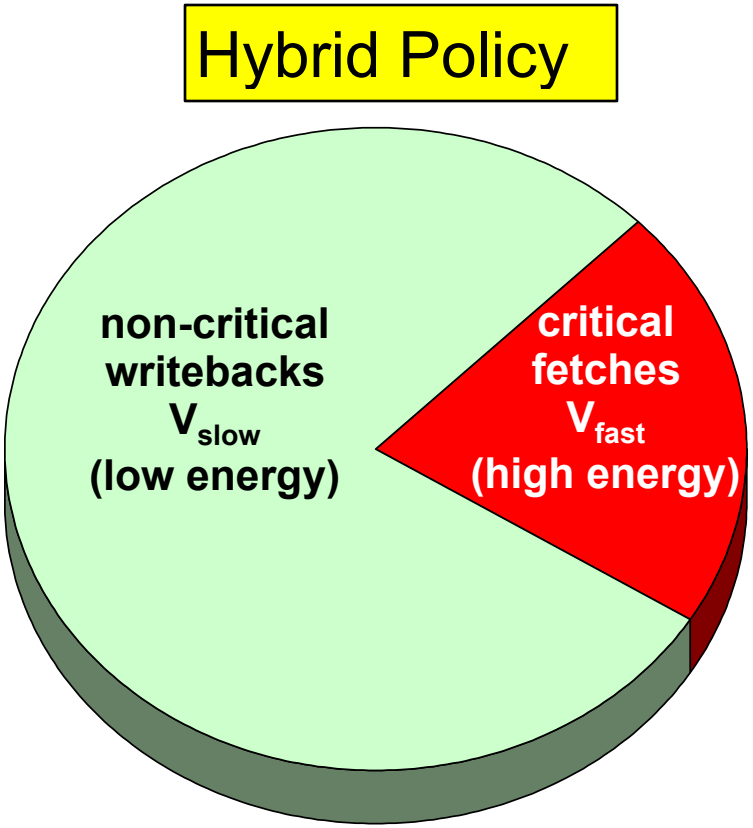


➢ L2 writebacks account for 80% of bitline activity
- Most of energy consumption

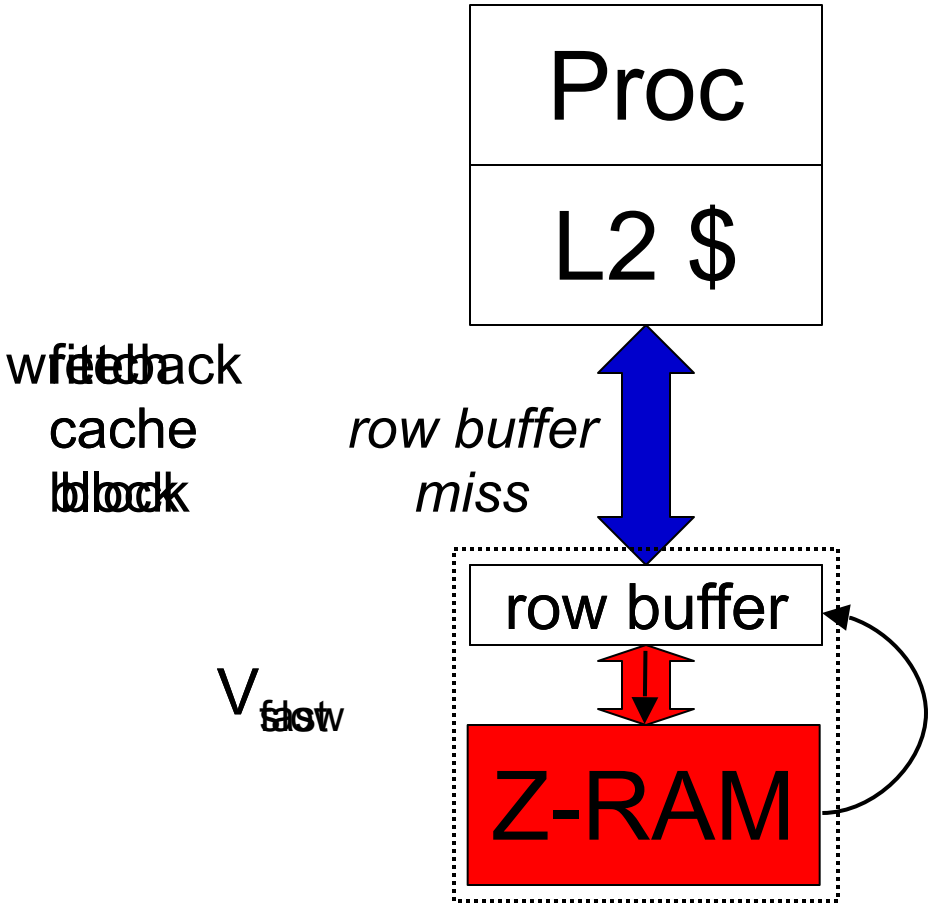
Ideal Combination of Factors

Target writebacks

- Most of energy savings potential
- Not performance-critical

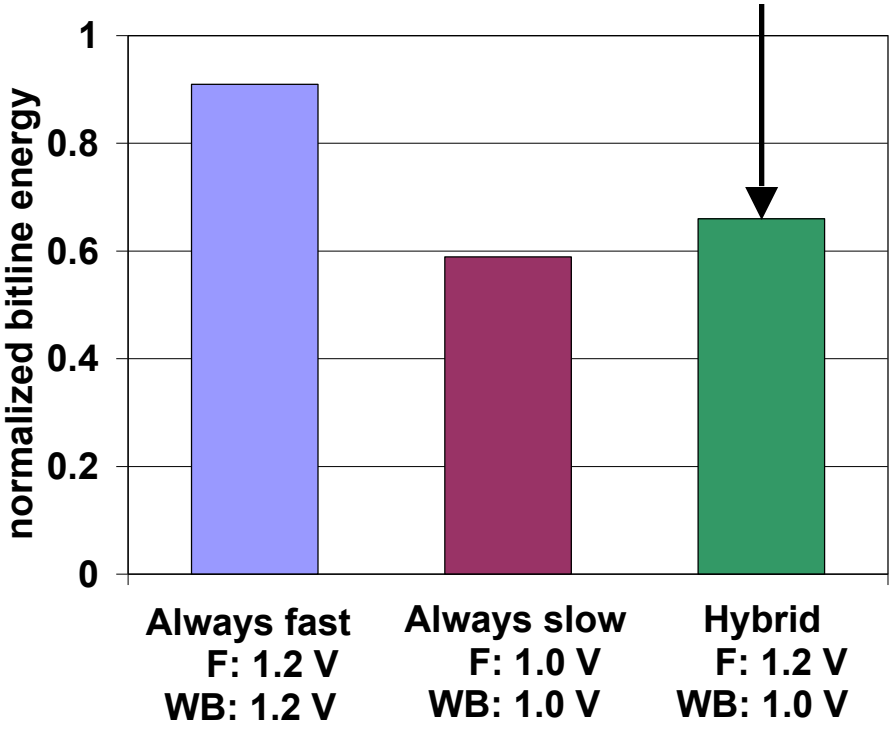


Hybrid Policy

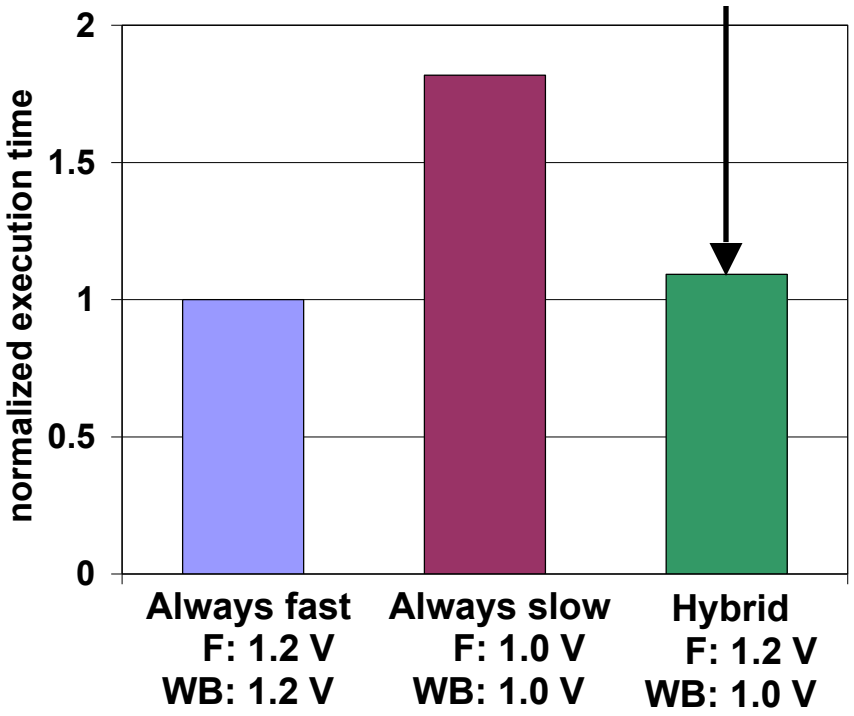


Hybrid Policy

Bitline energy w.r.t. DRAM



Execution time w.r.t. DRAM

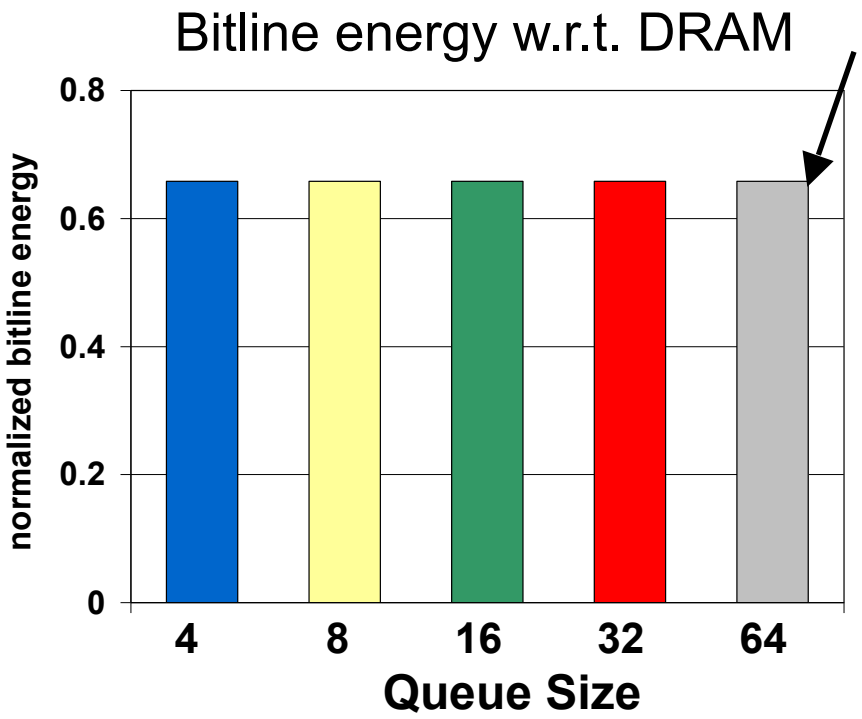


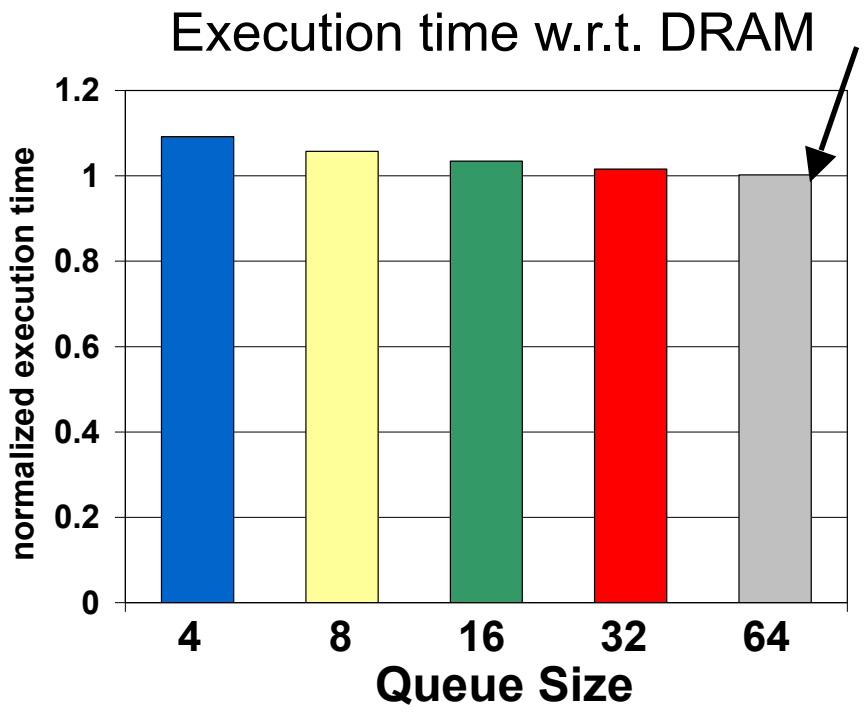
Eliminating Residual Slowdown

- Residual 10% slowdown
  - Delayed writebacks occupy queues in memory controller
  - Eventually stalls processor (indirectly)

- Tolerating delayed writebacks
  - Large buffers with access reordering
  - L2-cache eager writeback [Lee et al., MICRO33, 2000]

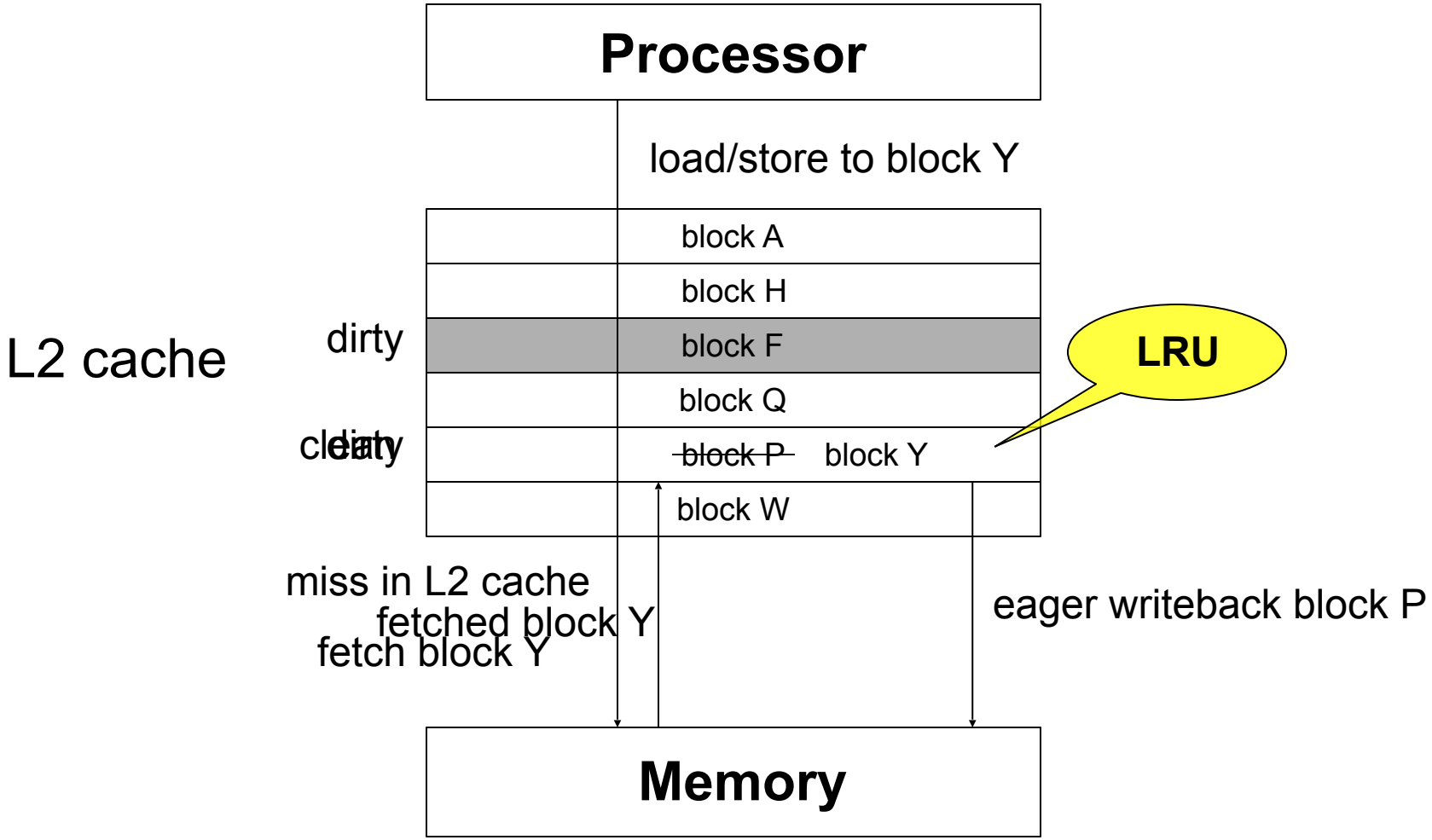
Large Buffers with Access Reordering



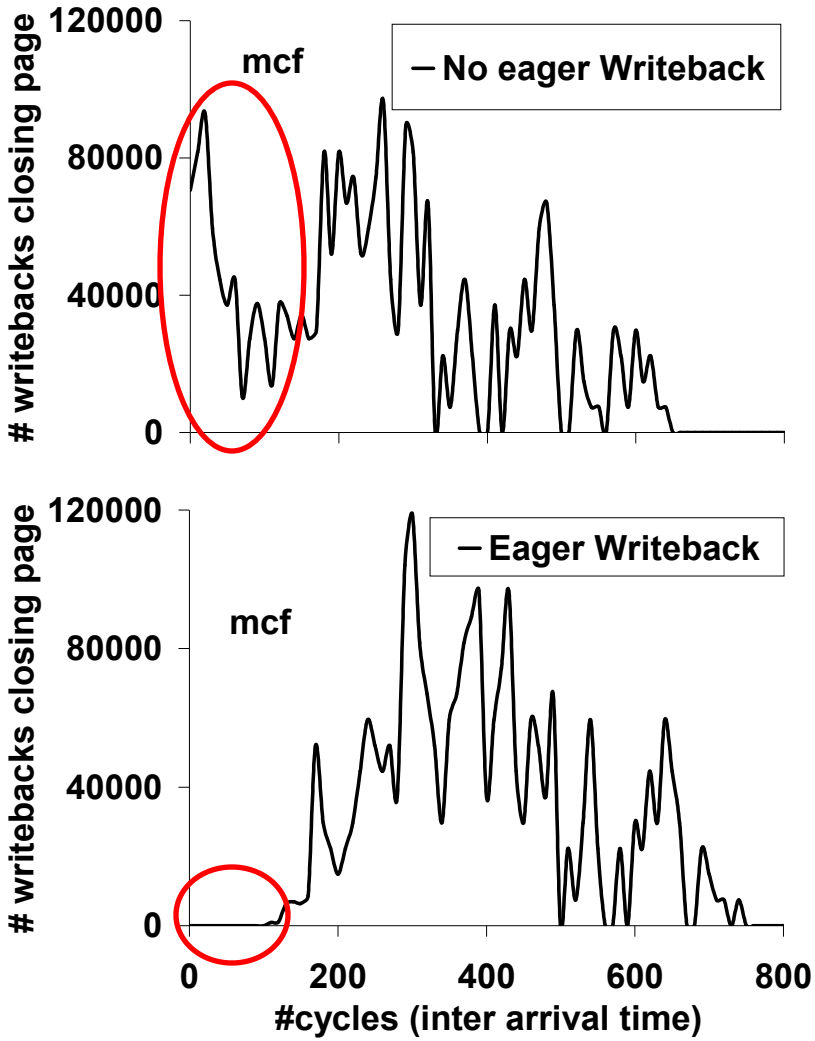


- Enlarging queue increases
  - Cost: each entry contains cache block
  - Complexity: fetches that bypass writebacks must first search queue for read-after-write hazards

L2 Cache Eager Writeback



De-clustering L2 writeback requests

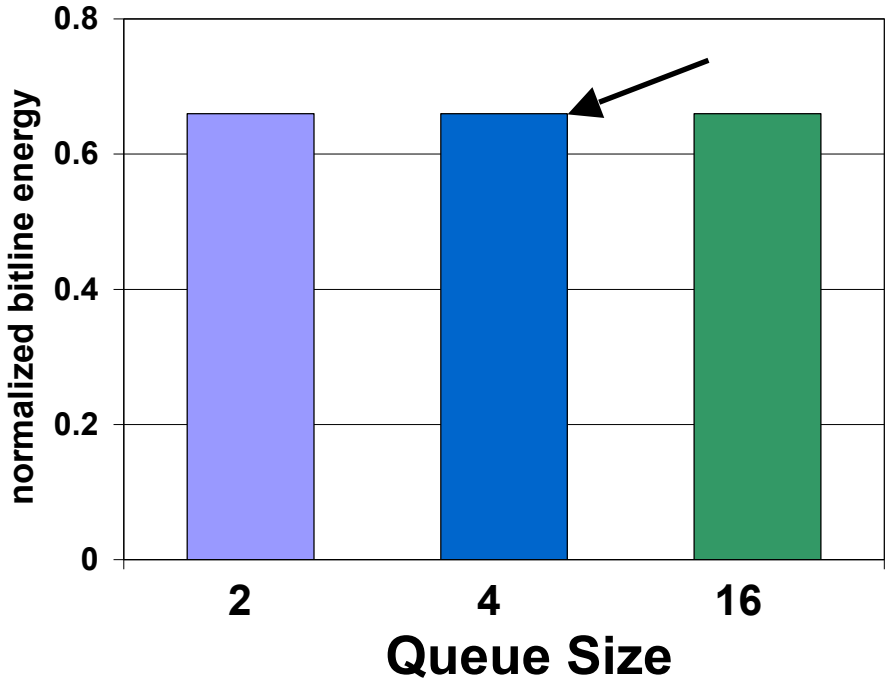


Hybrid policy with
4 queue entries

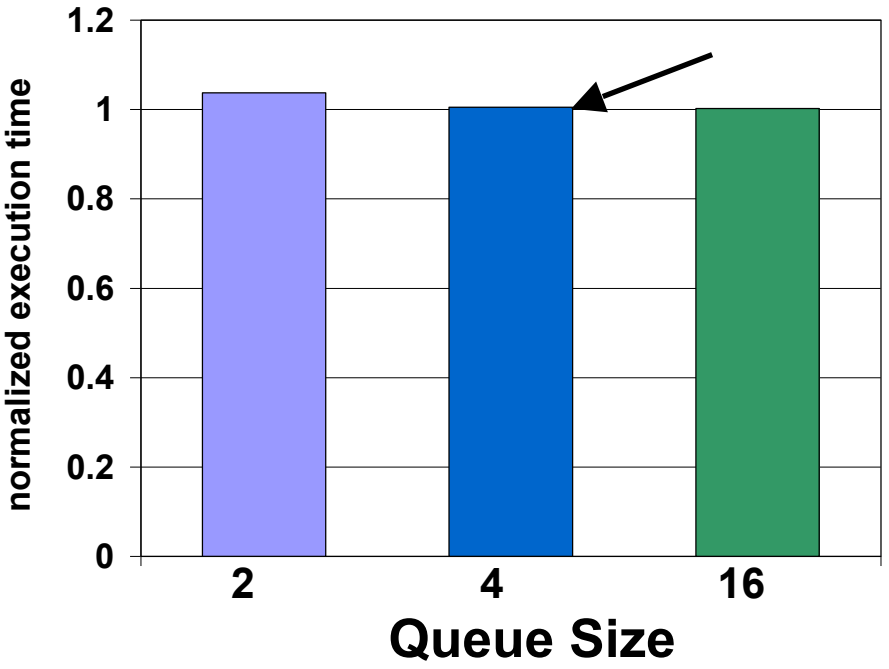
Hybrid policy with
4 queue entries
in conjunction with
eager writeback

Hybrid Policy + Eager Writeback in L2 cache

Bitline energy w.r.t. DRAM



Execution time w.r.t. DRAM



Conclusions

- Molecular capacitor unique
  - Functional with arbitrarily small voltage swings
  - Speed is voltage dependent
- Intelligently managing ZettaRAM
  - Hybrid policy
  - Eager writeback synergistic with ZettaRAM
- Results
  - 34% energy savings (out of 41% with uniformly slow writes)
  - Less than 1% performance degradation

Thank You

Questions ?!?