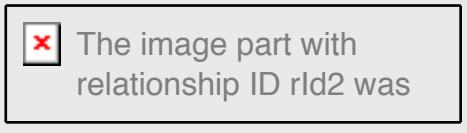
ADVANCED MEMORY CONTROLLERS

Mahdi Nazm Bojnordi

Assistant Professor

School of Computing

University of Utah



CS/ECE 6810: Computer Architecture

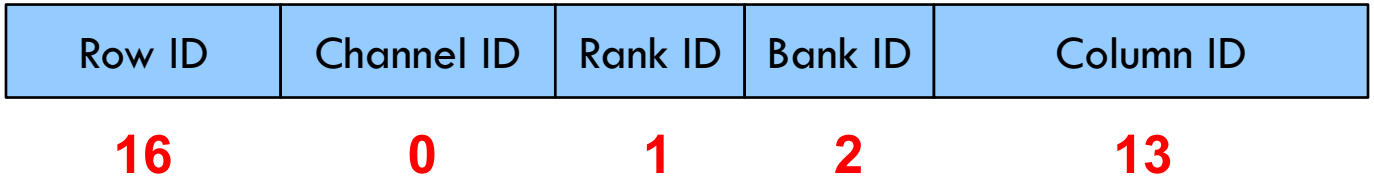
Overview: DRAM Control Tasks

- Refresh management
  - Periodically replenish the DRAM cells (burst vs. distributed)
- Address mapping
  - Distribute the requests to destination banks (load balancing)
- Request scheduling
  - Generate a sequence of commands for memory requests
    - Reduce overheads by eliminating unnecessary commands
- Power management
  - Keep the power consumption under a cap
- Error detection/correction
  - Detect and recover corrupted data

Address Mapping

- A memory request

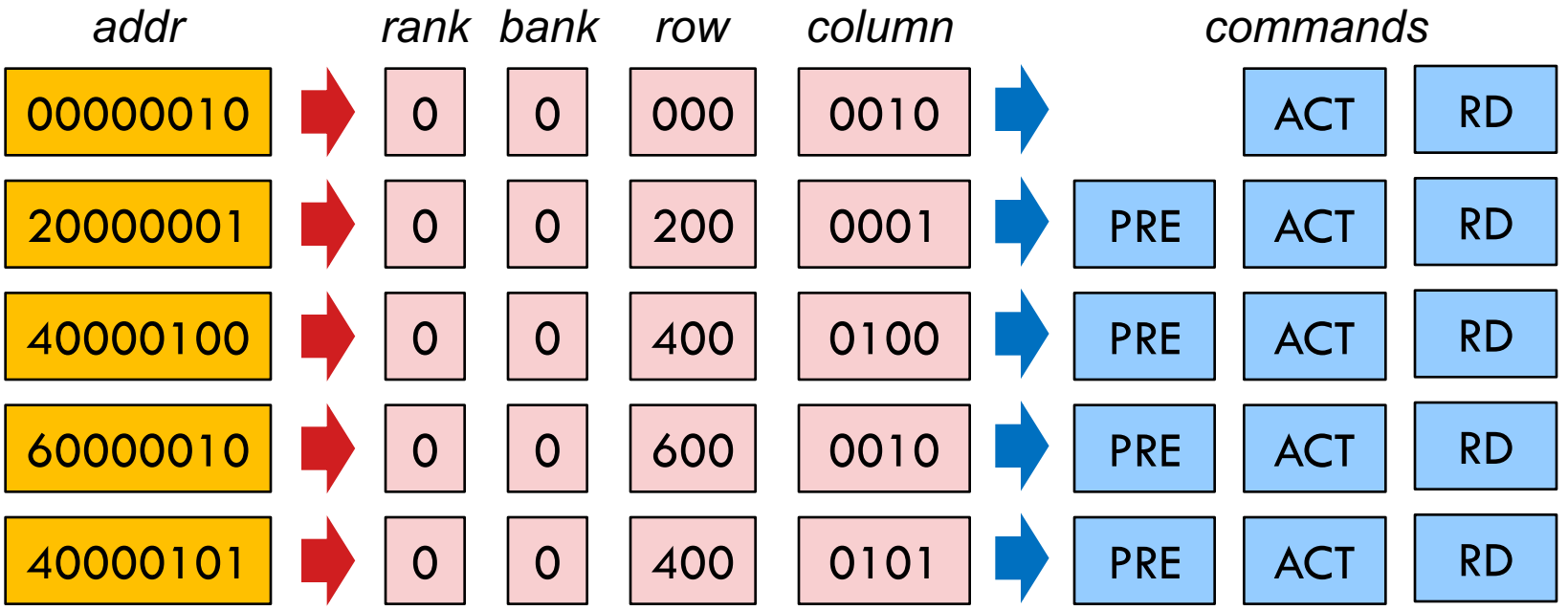
- Address is used to find the location in memory
  - Channel, rank, bank, row, and column IDs
- Example physical address format



- A 4GB channel, 2 ranks, 4 banks/rank, 8KB page

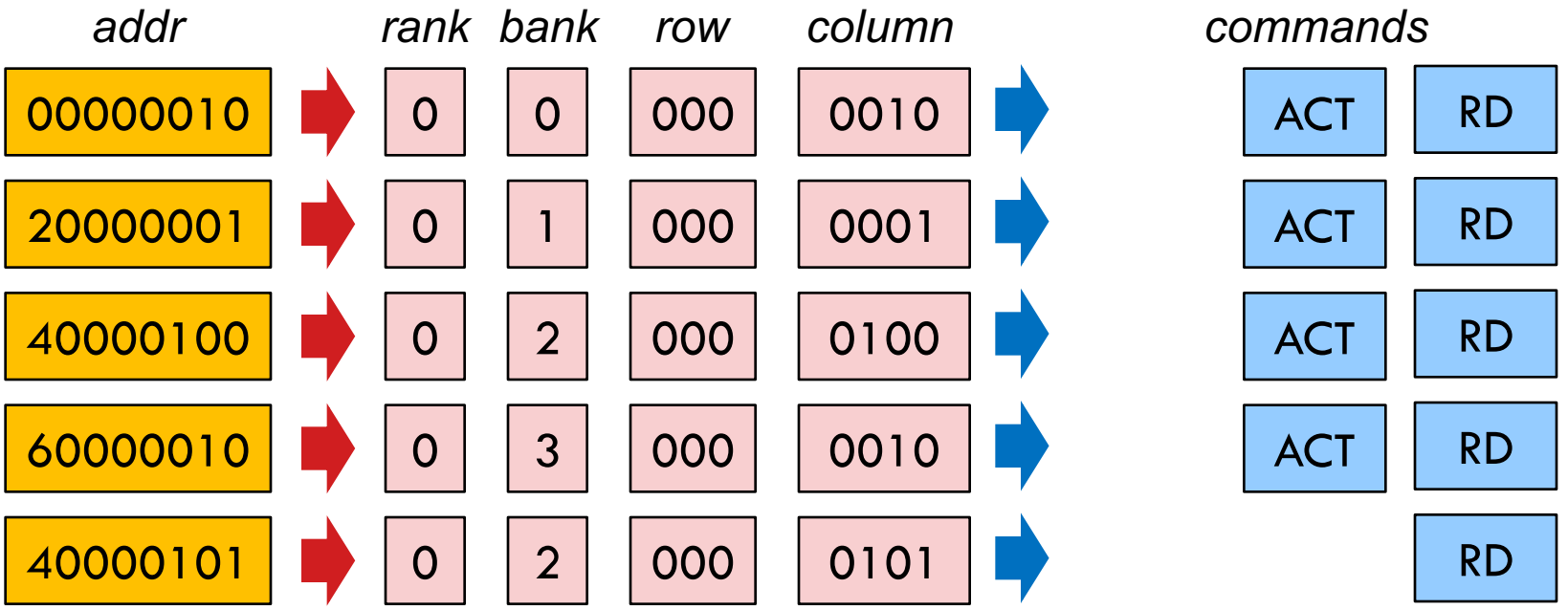
Example Problem

- Start with empty row buffers, find the total number of commands if all the request are served in order
  - Address= row(12):channel(0):rank(1):bank(3):column(16)



Example Problem

- Find the total number of commands using the following address mapping scheme
  - Address= bank(3):rank(1):channel(0):row(12):column(16)



Row Buffer Management Policies

- Open-page policy
  - After access, keep page in DRAM row buffer
  - If access to different page, must close old one first
    - Good if lots of locality
- Close-page policy
  - After access, immediately close page in DRAM row buffer
  - If access to different page, old one already closed
    - Good if no locality (random access)

Command Scheduling

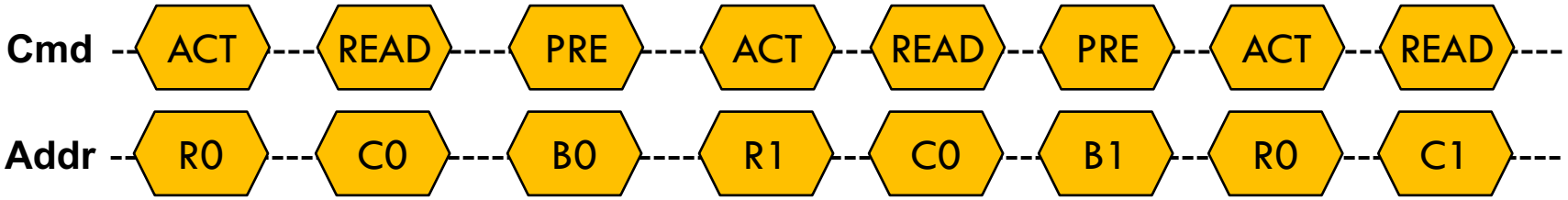
- Write buffering
  - Writes can wait until reads are done
- Controller queues DRAM commands
  - Usually into per-bank queues
  - Allows easily reordering ops. meant for same bank
- Two common policies
  - First-Come-First-Served (FCFS)
    - In order request scheduling
  - First-Ready First-Come-First-Served (FR-FCFS)
    - Out of order request scheduling

Command Scheduling

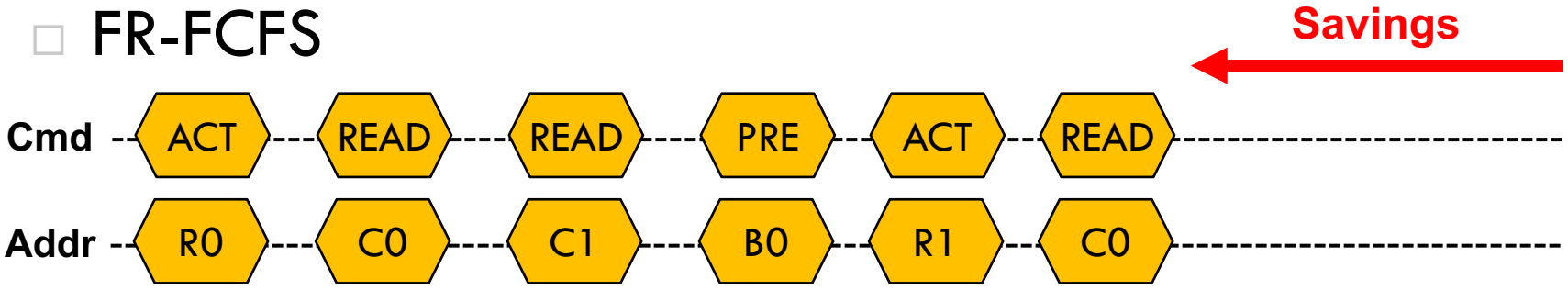
- First-Come-First-Served
  - Oldest request first
- First-Ready First-Come-First-Served
  - Prioritize column accesses over row changes
  - Skip over older conflicting requests
  - Find row hits (on queued requests)
    - Find oldest
    - If no conflicts with in-progress request → good
    - Otherwise (if conflicts), try next oldest

FCFS vs. FR-FCFS

- Single bank memory
  - READ(B0,R0,C0) READ(B0,R1,C0) READ(B0,R0,C1)
- FCFS



- FR-FCFS

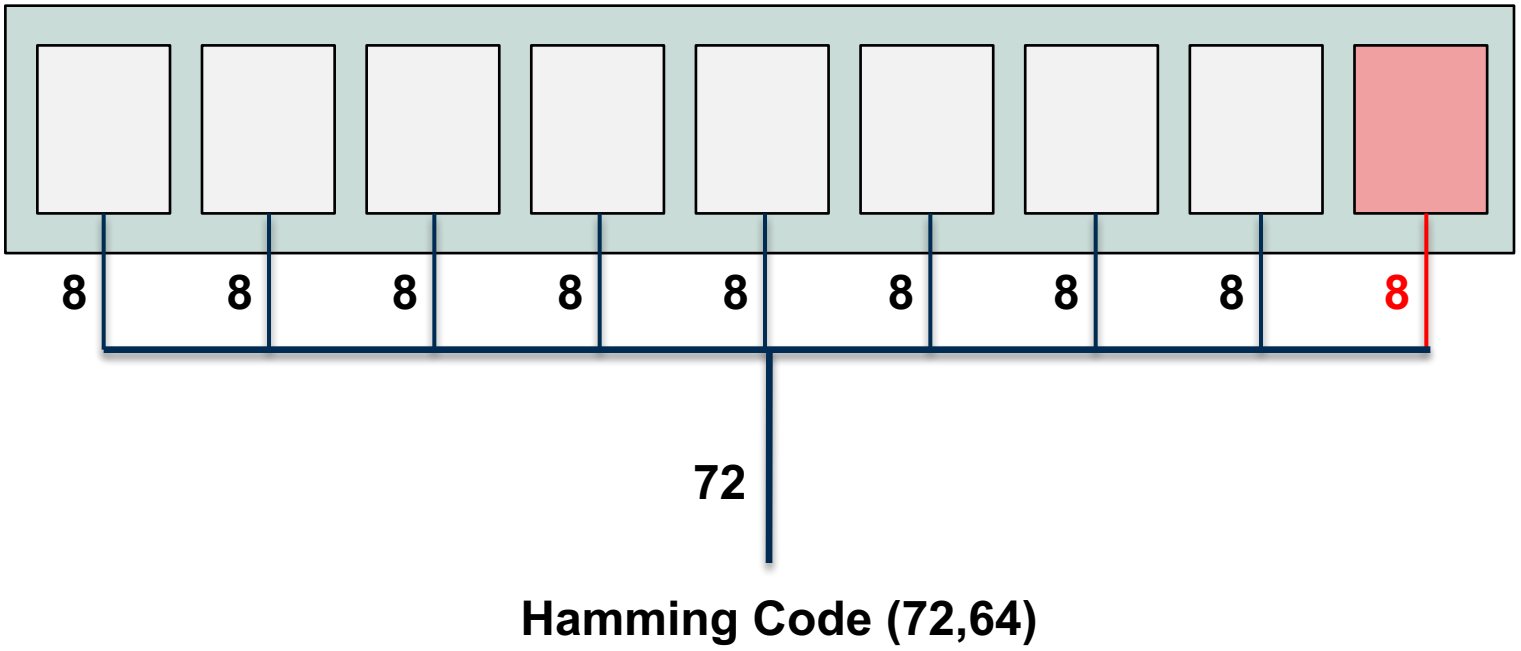


Error Detection/Correction

- Data in memory may be corrupted
  - Many reasons: leakage, alpha particles, hard errors.
- Can errors be detected?
  - Error detection codes: additional parity bits
- Can errors be corrected?
  - Error correction codes: ECC bits are added to data
- Single-Error Correction, Double-Error Detection
  - Commonly used in memory systems

ECC DIMM

- An additional DRAM chip is used for storing SECDED ECC bits for error correction



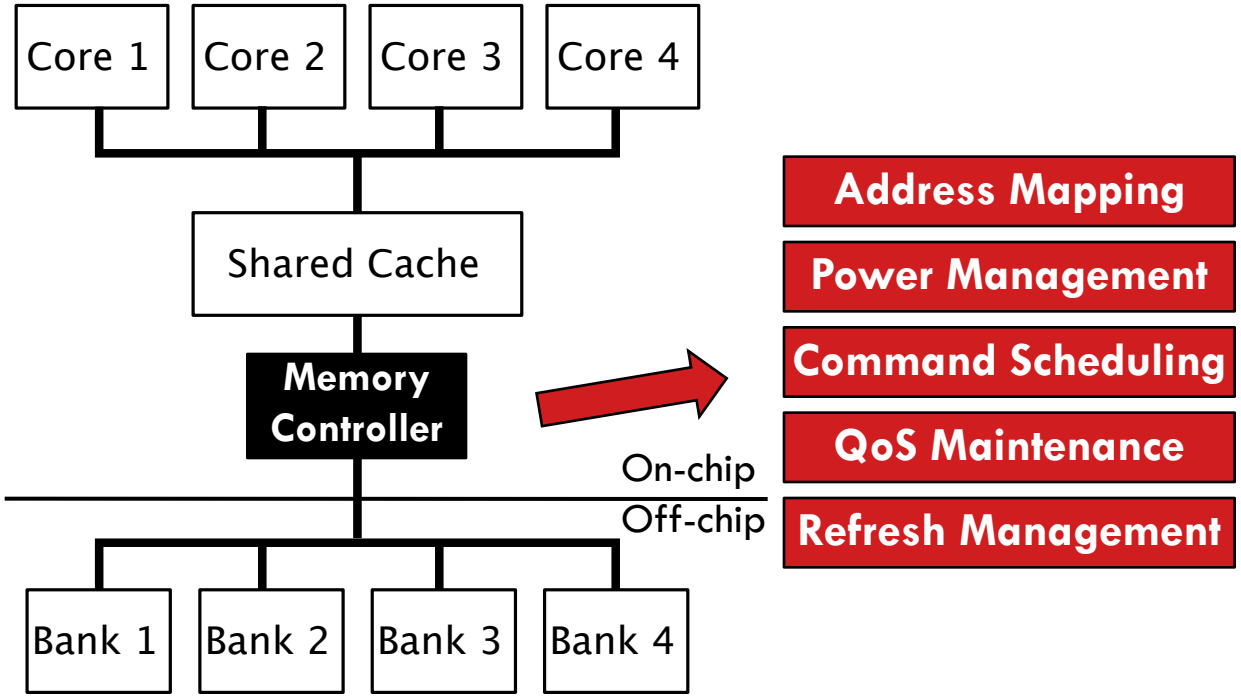
Limitations to Existing Memory Controllers

- Modern memory controllers are performance-critical and complex

Multiple performance objectives ✗

Application-specific optimizations ✗

Patches and in-field updates ✗



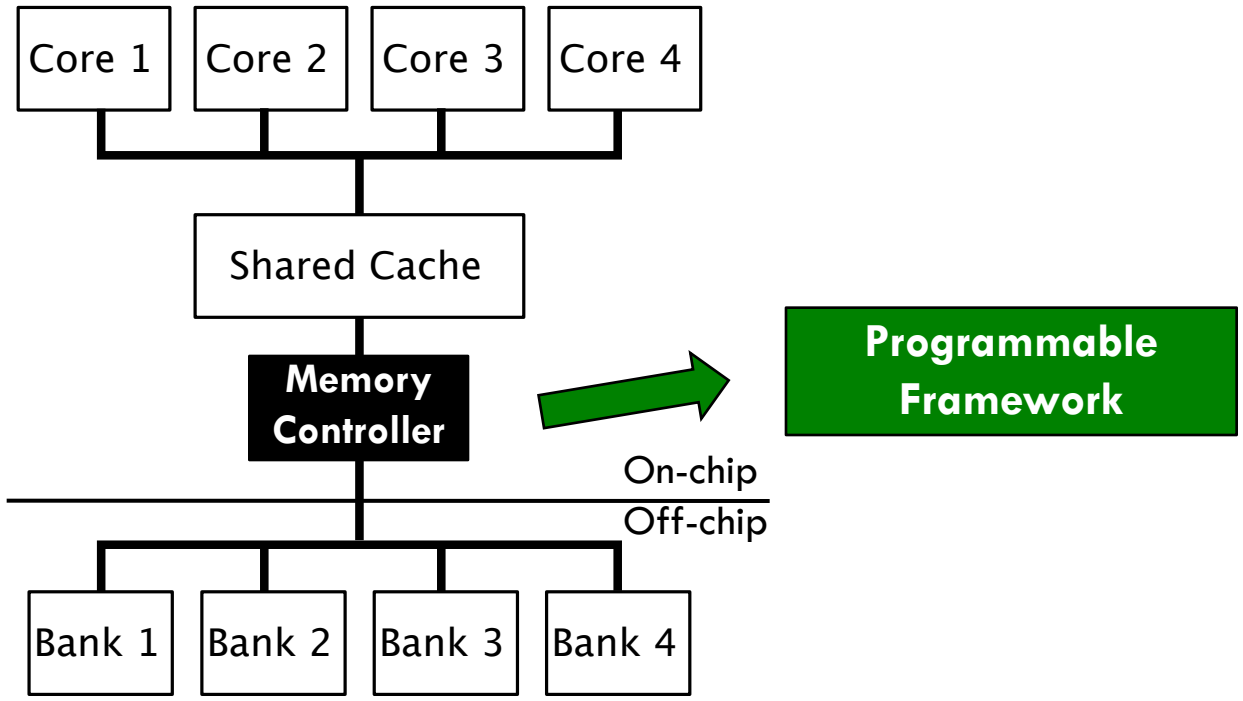
Programmable Memory Controllers

- Programmability can make a memory controller higher-performance and more flexible

Multiple performance objectives ✓

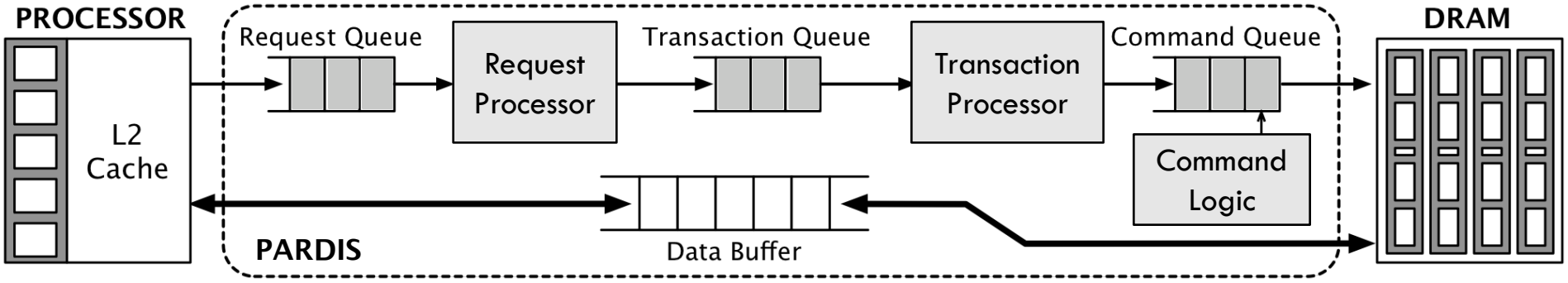
Application-specific optimizations ✓

Patches and in-field updates ✓



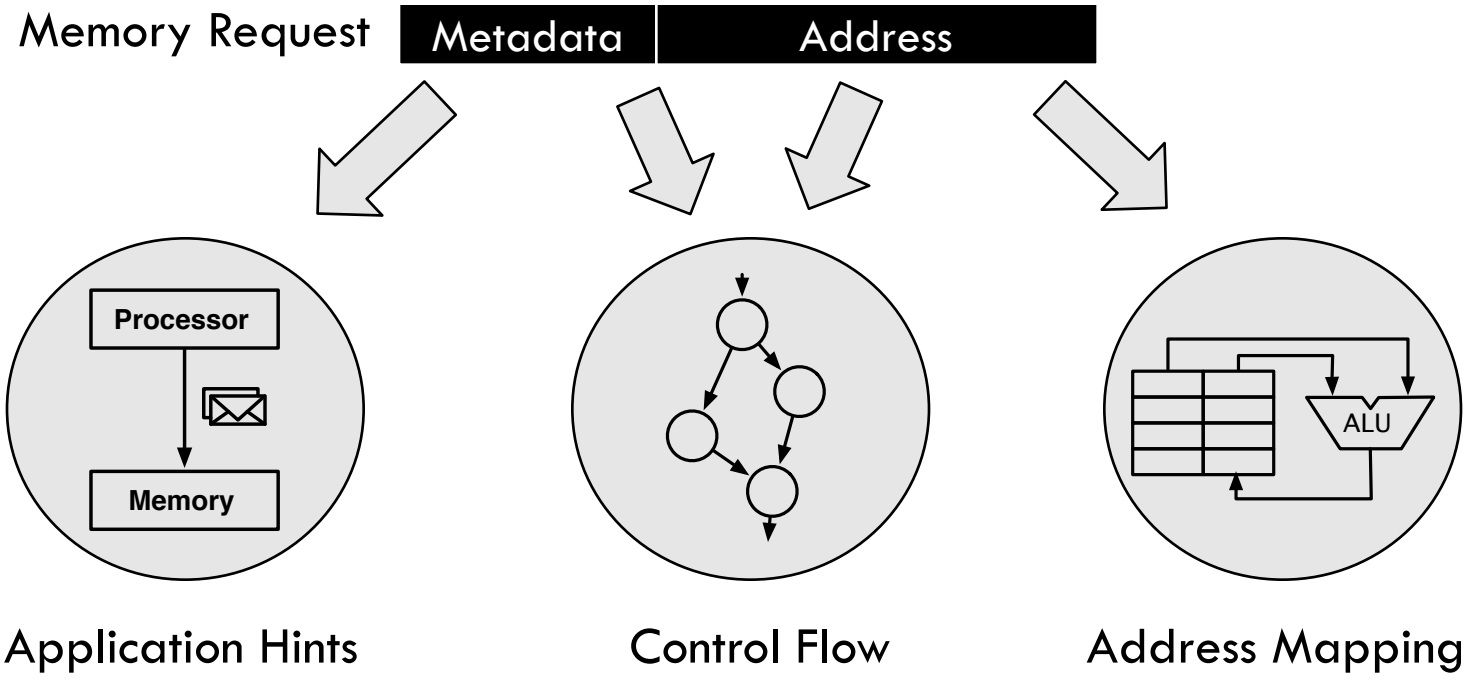
Design Overview

- Key idea: Judicious division of labor between specialized hardware and firmware
  - Request and transaction processing in firmware
  - Configurable timing validation in hardware



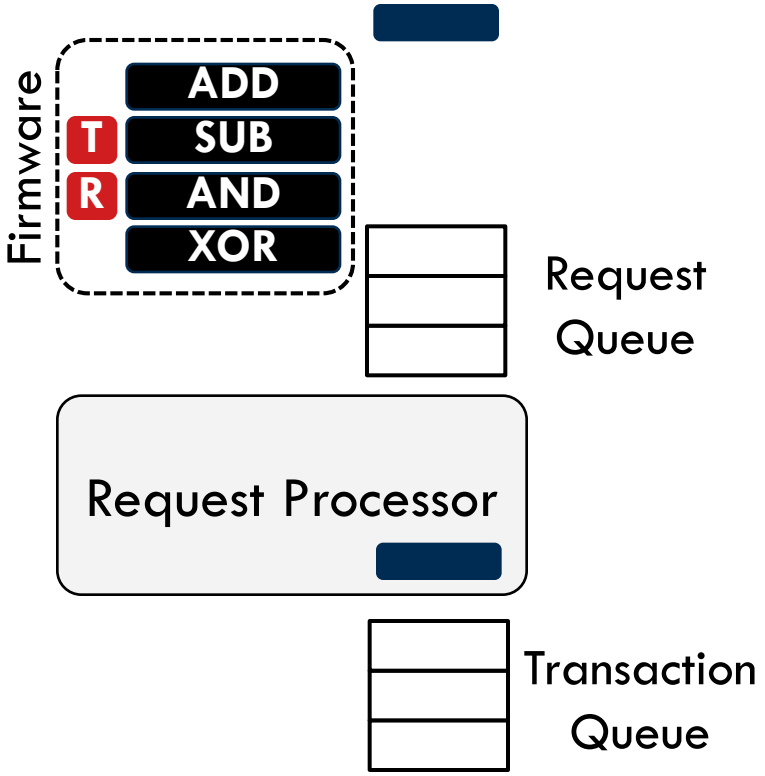
Request Processing

- A RISC ISA for operating on memory requests



Request Processing

- Queue management with instruction flags
  - R flag enqueues a request
  - T flag dequeues a transaction
- An instruction can be annotated with both R and T flags if needed



Implementation

- Two five-stage pipelines and one configurable timing validation circuit