Recitation 7

Coming Up

- HW2 due tonight @ 8:00 PM
  - Don't forget to zip your pdf report along with your solution
- HW3 will be released on Thursday
- Coding Exam 2 next Tuesday @ 5:00pm

Exams

- Coding Exam 1 Curve Released
- Midterm 1 Curve Released

- Coding exams are best 2 out of 3
- Single 5 point question
  - Strict outputs required
- Curve will apply to the remaining 2 as well

Caches

- What primary three types are there?
- Direct Mapped
- Associative (i.e., *fully* associative)
- N-way Associative
- Let’s look at them in this order:
  - Direct
  - Associative
  - N-way

Direct Mapped

- This is a one-way associative, but why is that?
- By rules of N-way, only do first step
- I.e., map an element to cache line by partial address
- How might this help or hinder performance?

Associative Mapping

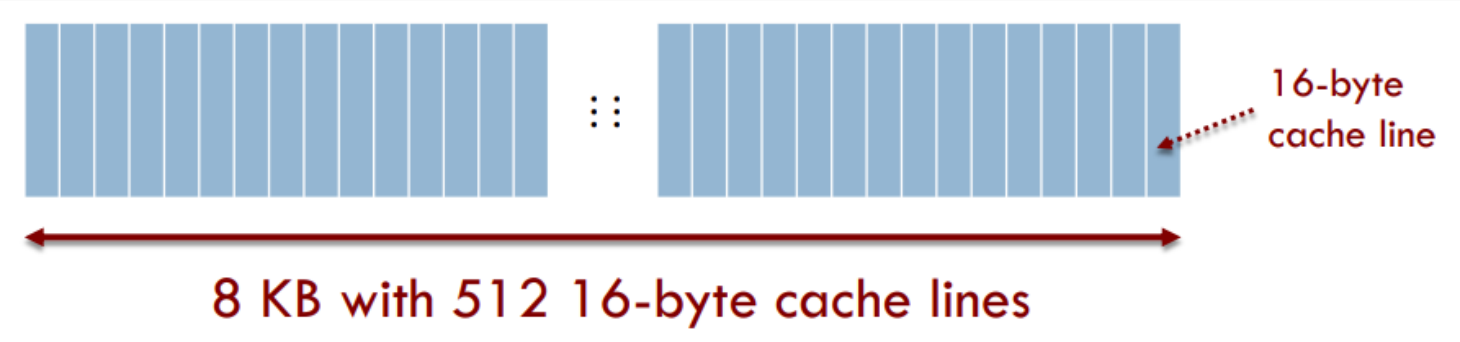
- "*Associative*" implies a relational grouping
  - Mathematics and Computer Science
  - Sometimes thought of as "key-value" pairs
- Here the association is a time and a value read from memory
- Goal is to store a value from memory in the cache
- Time is something like a counter (always incrementing)
- When reading a value from memory
  - The value is stored in cache associated with its time counter
- If no cache available:
  - Oldest value is evicted (smallest time counter)
  - New value takes it place (associated with current time counter)

N-way Mapping

- Same as *Associative Mapping* **BUT**
- Not over the entire set of cache lines
- Instead, **cache lines subdivided into sets of size N**
- Two step process
  - First, determine which cache line set a value goes to
    - Based on something like partial address value
  - Second, use association (time counter) to evict one of N cache lines

Cache Lines

- A **cache line** is the smallest unit of data transferred between main memory and cache
- This is what memory is mapping to



Completion Activity

- Worksheet posted on Canvas covering caching techniques.