Memory

Young W. Lim

May 16, 2016

Based on

Computer System Design : System-on-Chip
by M.J. Flynn and W. Luk

Scrtchpad and Cache Memory

- small size memory can access fater
- frequently used instructions / data must be kept in a separate small size memory
- Scratchpad Memory
  - A programmer manages directly
  - usually data only
- Cache Memory
  - A dedicated hardware manages
  - instruction / data

Cache Memory Principles

- Spatial Locality
  - neighbor location of a previous access will be accessed again
- Temporal Locality
  - a sequence of n location references will be accessed again
- Sequentiality
  - next address location of a previous access will accessed again

Cache Parameters

- Physical Word
  - unit of transfer between processor and cache
- Block Size or Line
  - the basic unit of transfer between cache and memory
  - n physical words
- Access Time for a cache hit
  - depends on the cache size and organization
- Access Time for a cache miss
  - depends on the memory and bus
- Time for computing the real address from virtual address
  - depends on the address translation hardware
- Number of processor requests per cycle

Cache Organization

- Fetch on demand
  - used in simple computers
  - new memory locality only when a miss occurs
- Prefetch
  - anticipate the locality
  - used in I-caches
- 3 types of cache organizations
  - Fully Associative Mapping
  - Direct Mapping
  - Set Associative Mapping

Three Types of Cache Organizations

- Fully Associative Mapping
  - the address of a request is compared with those of all entries in the directory
  - if there exists a match (a directory hit) the correponding data is accessed in the cache
  - otherwise a miss occurs
- Direct Mapping
  - the lower address bits access the directory
  - multiple addresses share the same lower address
  - the higher address bits must be compared to the directory address
  - accessing the cache array can be performed in parallel with accessing the directory
- Set Associative Mapping
  - the combination of Fully Associative and Direct Mapping
  - the lower address bits access the directory
  - 2/4/8 complete line addresses in the directory
  - each address corresponds to a location in a subcache
  - these subarrays can be accessed simultaneously
  - together with the cache directory

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Memory

- Instruction Set

Reference

[1] M.J. Flynn and W. Luk, “Computer System Design : System-on-Chip”, Wiley, 2011