

Memory Hierarchy

CPS 5401 Fall 2014

<http://svmoore.pbworks.com/>

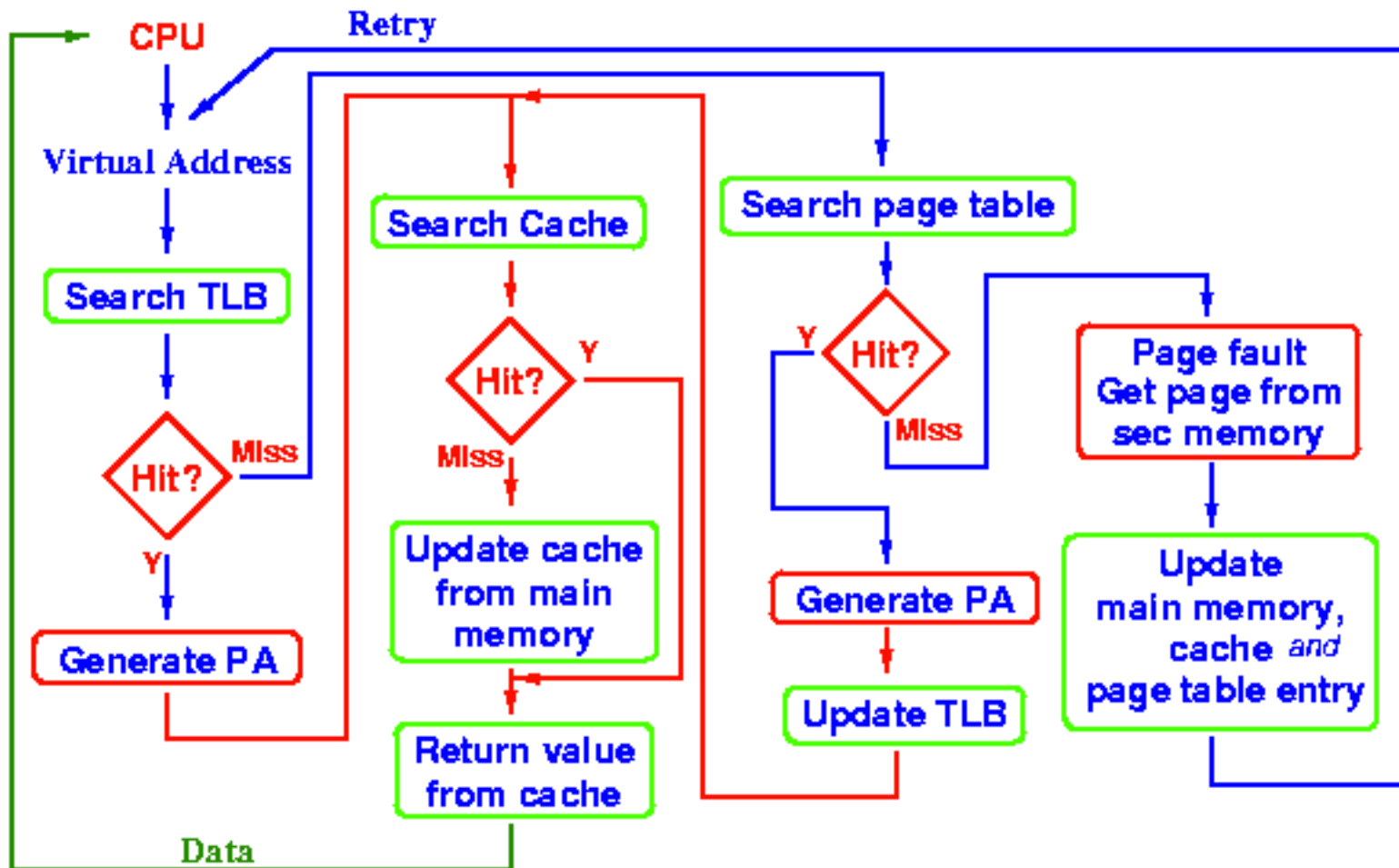
Shirley Moore, Instructor

September 3 Class

Learning Outcomes

- After successfully completing this lesson, you will be able to:
 - Describe how memory references are processed by the hardware and the operating system.
 - Calculate how memory blocks are mapped to cache locations for a given cache configuration.
 - Describe how a memory address is used to identify a cache block containing the data.
 - Explain how the block replacement policy can affect code performance.

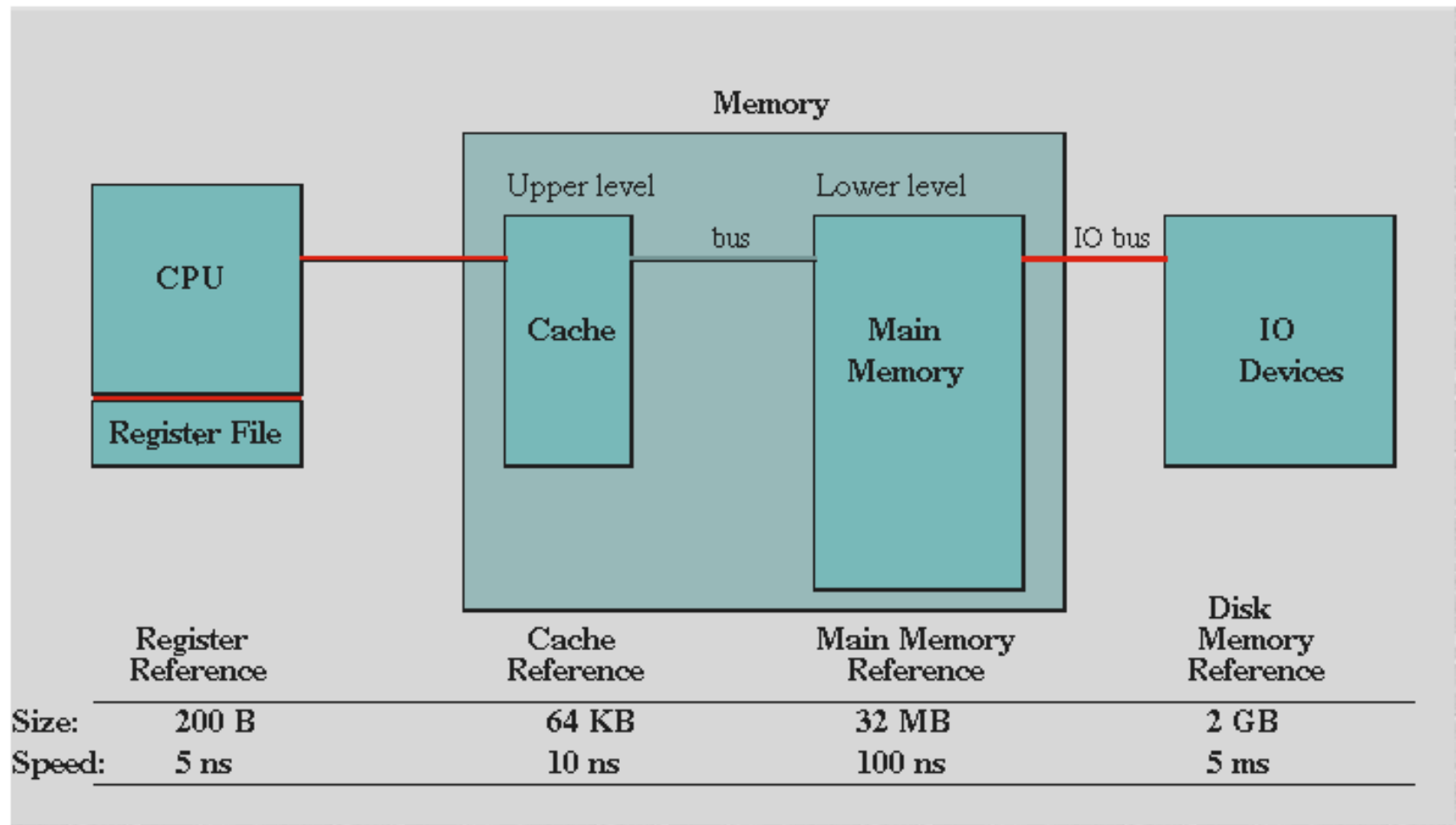
Processing of Memory References



Virtual Memory

- Each process's address space is partitioned in to fixed size *pages* (e.g., 4Kbytes).
- Pages are loaded from secondary storage (i.e., disk) into memory as needed.
- The memory manager keeps track of all pages currently loaded by means of a page table.
- A *page fault* occurs when a process requests a page that is not currently loaded into main memory.
- When a page fault occurs and the main memory is full, the memory manager has to decide which page to evict to make room for the new page.
 - An approximation to Least Recently Used (LRU) is usually used.
- A *translation lookaside buffer* (TLB) is a cache for the page table.
 - A *TLB miss* occurs if the desired mapping is not in the TLB.

Simplified Memory Hierarchy



Cache Associativity

- Direct mapped
- Fully associative
- Set-associative

- See

http://www.cs.iastate.edu/~prabhu/Tutorial/CACHE/bl_place.html

- Try it!

http://www.cs.iastate.edu/~prabhu/Tutorial/CACHE/bl_place_applet.html

Block Identification

- http://www.cs.iastate.edu/~prabhu/Tutorial/CACHE/bl_ident.html
- http://www.cs.iastate.edu/~prabhu/Tutorial/CACHE/bl_ident_applet.html

Block Replacement

- http://www.cs.iastate.edu/~prabhu/Tutorial/CACHE/bl_replace.html

Types of Cache Misses

- Compulsory
- Capacity
- Conflict
- Cache tutorial at UMass
 - <http://www.ecs.umass.edu/ece/koren/architecture/Cache/tutorial.html>