Test 3 Study Guide

## Definitions:

* The Principal Of Locality
  + A program tends to access data that forms a physical cluster in memory. Multiple accesses may be made within the same block.
* Temporal Locality
  + If an item is referenced, it will tend to be referenced again soon.
* Spatial Locality
  + If an item is referenced, items where memory addresses are close by will tend to be referenced soon.
* Memory Hierarchy
  + A structure that uses multiple levels of memory, in order to decrease memory access time.
* Cache Memory
  + The level of the memory hierarchy between the processor and the main memory used to take advantage of temporal and spatial locality.
* Memory Rule
  + As the distance from the processor increases, the size of the memory as well as the access time increases. The access time for main memory may be up to 100 times slower than cache.
* Basic Memory Structure

Cache  
Small, Fast, Expensive

Main Memory  
Large, Slow, Inexpensive

Processor

Blocks

Words

* Cache Block
  + The minimum amount of information transferred between cache and main memory.
* Hit Rate
  + The probability that a requested memory item is in the cache. (0 <= p <= 1)
* Cache Hit
  + If the requested word is in the cache, read or write operations are performed directly in the cache without access memory
* Cache Miss
  + If the requested word is not in the cache, a block of words containing the requested word is brought to cache.
* Miss Rate
  + The probability that a requested word is not in the cache. (1 – HitRate)
* Average Memory Access Time (AMAT)
  + Formula
  + Ex. 5.1 (AMAT)
  + Ex. 5.2 (Speedup)
  + Ex. 5.15 (AMAT)
  + Ex. 5.16 (AMAT)
  + Ex. 5.3
  + Ex. 5.4
* Cache/Memory Properties
  + Increase Cache Size
    - More blocks are kept in the cache. The chance of a miss is reduced. Large cache is slower and expensive
  + Increase Block Size
    - More data is available, which reduces chance for a miss, but fewer blocks in a cache increase the chance for a miss.
  + Increase Hit Rate
    - Hit rate increases with cache size, but hit rate depends on block size.

Types Of Cache

* Direct-Mapped
  + A cache structure in which each cache location is mapped to exactly one location in main memory.
  + References (5.18, 5.21, Ex. 5.5)
* Fully-Associative
  + A cache structure in which every location is mapped to one location in cache.
  + References (5.19, 5.21, Ex. 5.6)
* Set-Associative
  + A cache structure that has a fixed number of locations where each block can be placed (a set).
  + References (Ex. 5.7, Ex. 5.8, Ex. 5.9)
* Cache Coherence (5.22, 5.25)
  + The process of ensuring that the copy of a data in cache is identical to the copy in memory
* Write Through
  + Write to memory and cache simultaneously always.
    - Drawback – writing to main memory is ~100 times slower than to cache
* Write Back
  + Writes only to cache, but sets a “dirty bit” in the block where write is performed. When a block with a “dirty bit” is flagged “on”, the data is first written to main memory before being overwritten.
* Virtual Memory
  + Implementation of memory where main memory acts as a “Cache” for secondary storage.
* Virtual Page
  + A block of data in main memory used for virtual memory implementation.
* Main Memory Unit (MMU)
  + Translates a virtual address to a physical address. Complete address table is large and kept in main memory.
* Translation Look aside Buffer (TLB)
  + A small cache of the address translation table.

A flow chart showing the steps needed when the CPU issues a virtual address to get data.

Physical Address

Virtual or Logical Address

MMU  
Memory Management Unit

Data

Cache

Processor

Physical Addresses

Data

Main Memory

Disk

DMA (Direct Memory Access)

* Page Fault
  + An event that occurs when an accessed page is not present in memory

Processor  
Cache  
MMU (TMB)

Main Memory

Disk

Pages

(Write-back same as in Cache)

All Data, organized in pages (~4kb), accessed by physical address

Cached pages, Page table  
  
Page Fault: A required page is not found in main memory

Cache Miss: A required block is not found in cache.  
  
TLB miss: A required virtual address is not found in TMB.

# From the Professor:

## Topics

* Pipelining Basics
* Pipelined Processor Design
* Dependencies and Hazards
* Memory Hierarchy

## Pipelining Basics

## Pipelined Processor Design

## Dependencies and Hazards

## Memory Hierarchy