HW #11 Summary of Chapter 12.

12-1 Memory Hierarchy

The cache is the lowest level of the hierarchy, it small but fast. The next level is the main memory, it serves CPU instruction (mainly). Then, the top level is the hard drive, it only accessed when instructions or operand fetch are not found in main memory.

12-2 Locality of Reference

In order to achieving the appearance of a large, fast memory, we need to deal with the foundation of this assumption: locality of reference. → The relative time that operands and instructions are accessed and the locations of them in the main memory.

In the processing of accessing instructions and operands, temporal and spatial localities are occured often. Those presence of the significant temporal and spatial localities lead us to the design of caches and virtual memory.

For CPU, the instruction that was fetched once will tend to be used again will be in the cache for faster access. The cache also fetch nearby instructions into its SRAM for faster access.

A similar relationship is built between main memory and the hard drive in order to reduce the access to the main memory and improve system speed: the virtural memory. Words are read from and written to the drive in blocks referred to as pages. The virtual memory is an automatic arrangement that fetch the pages which contain instruction to the main memory.

12-3 Cache Memory

The cache control places word data on the bus to the CPU, Complete the fectch operation. Cache hit is memory word is fetched. Cache miss is when a tag is missmatched More cache hit makes cache effective.

Direct mapping is only one specific location in cache can Contain the word from a particular main memory location. To avoid the issues of direct mapping, the fully associative mapping is introduced. The fully associative mapping provides maximum flexibility and good performance but it still has cost and performance issues.

The set-associative mapping is a better solution than direct or fully associative mapping.

It is also possible to extend the depth of the memory hierarchy by adding additional levels of cache. Cache. The often used method is 2 level cache. The design of 2 level cache is more complex that a single level cache

12-4

Virtual memory is the answer for the quest of pursuing fast and large memory. The connection is made between the hard drive and main memory. The hard drive provides most of the virtual address space and later those addresses are translated into physical addresses when we permit the software to map the virtual addresses. The virtual addresses are divided as pages, translate and fectched into main memory as needed.

The memory management unit provides a special hardware Called translation lookaside Buffer (TLB) to boost the performance of the virtual memory translation process.