Presentation guide for unit 5
Memory systems

The objective of this unit is to present the concept of memory hierarchy in a computer:

1. Introduction to the hierarchy of memory
2. Cache memory
3. Virtual memory

1. Introduction to the memory hierarchy

This section describes the different types of devices used in a computer as a memory device: semiconductor memory, magnetic memory and optical memory. It discusses how to obtain the number of memory accesses a program generates during its execution and the impact it has on the computer's performance. Next, it is described the memory hierarchy concept and it is analyzed the structure of the static and dynamic RAM memories.

2. Cache memory

The cache is a small fast static memory (SRAM), which is located between the processor and the main memory and allows to accelerate the execution of programs. This unit studies the basic operation of cache memory and shows how it improves the program execution tuime. The average time to access a memory system using cache memory is presented.

Next, the cache structure and its internal design is studied by analyzing its size, the matching function, cache replacement algorithms and the writing policy. Special attention is given to the matching function, which is the algorithm that determines where a particular block of the main memory can be stored in the cache. It is the mechanism that allows you to know which block of main memory is located on a cache line. The direct, associative and set associative matching functions are described.

3. Virtual memory

This unit begins by illustrating the process of loading and running a program. A process is called a running program and memory image is the set of memory addresses that a program occupies during its execution. It also describes the process of relocating a program into memory, both software and hardware.

In a system without virtual memory, the memory image of a program must completely reside in the computer's main memory in order to be executed. In a system with virtual memory, process memory images reside in a virtual memory map or virtual address space that is physically supported by two levels of the memory hierarchy: main memory and disk. During the execution of a program, there will be parts of a process memory image that will be in main memory and others that may reside on disk.

In this topic all the basic foundations of virtual memory are studied and with special detail the mechanism known as pagination. In a paginated virtual memory system, the virtual address space of a process is divided into equally sized pieces called pages. The main memory is divided into equal-sized chunks called page frames and the disk area that supports virtual memory is divided into equal-sized chunks called swap pages. The tasks that the memory management unit

(MMU) performs to translate the virtual addresses generated by the processor to the physical addresses where the data the processor is referencing is actually stored are described. It also describes the tasks to be performed by the operating system during a page failure, an interruption that occurs when trying to access an address that does not reside in main memory.

Page tables are described, which are the structures used by computers to know in which parts of memory or disk a certain part of the virtual memory image of a process is stored. Next, the process of translating virtual addresses to physical ones is described and different page table structures are presented:

- Single-level page tables.
- Two-level page tables.
- Inverted page tables

This section also introduces the concept of TLB (Translation Lookaside Buffer) which is a small cache of more recently used page table entries used to speed up the process of finding a page in memory. Finally, the general operation of a system using virtual memory and cache is described.

Material

As material associated with this unit is included the theory material and a collection of exercises proposed and resolved on the aspects covered in the subject. Other resources provide a link to various cache simulators that illustrate the basic operation of a cache.

Recommended bibliography

- "Problemas resueltos de estructuras de computadores" (GARCIA CARBALLEIRA, Félix et al.).
- "Computer organization and design. The hardware/software interface" (PATTERSON, David, et al).
- "Computer Organization and Architecture" (STALLINGS, William).