

Memory Organization

5.1 The Memory Hierarchy

One of the most important considerations in understanding the performance capabilities of a modern processor are the memory hierarchy. Not all memory is created equal, and some types are far less efficient and thus cheaper than others. To deal with this disparity, today's computer systems use a combination of memory types to provide the best performance at the best cost. This approach is called hierarchical memory. As a rule, the faster memory is, the more expensive it is per bit of storage. By using a hierarchy of memories, each with different access speeds and storage capacities, a computer system can exhibit performance above what would be possible without a combination of the various types.

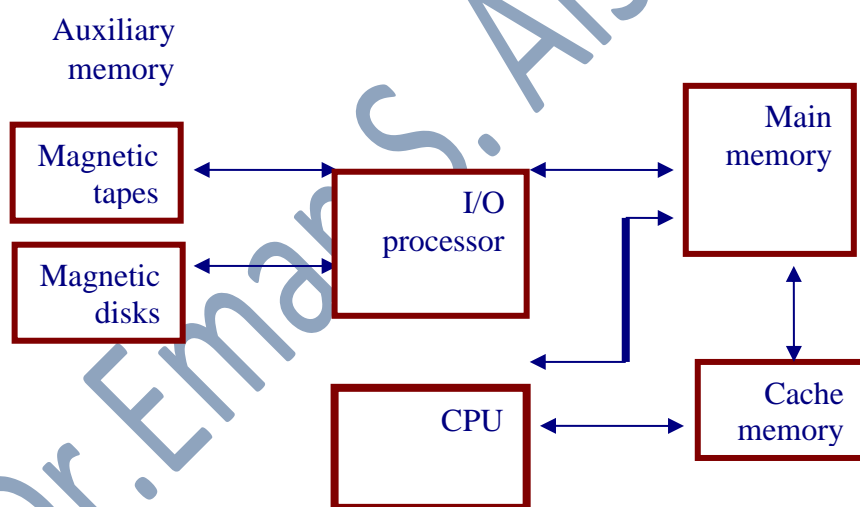


Fig 1.1 Memory hierarchy in computer system

The memory hierarchy can be characterized by a number of parameters. Among these parameters are the (**access type, capacity, cycle time, latency, bandwidth, and cost**).

- . The term access refers to the action that physically takes place during a read or writes operation.
- . The capacity of a memory level is usually measured in bytes.

- . The cycle time is defined as the time elapsed from the start of a read operation to the start of a subsequent read.
- . The latency is defined as the time interval between the request for information and the access to the first bit of that information.

The bandwidth provides a measure of the number of bits per second that can be accessed.

- . The cost of a memory level is usually specified as costs per megabytes

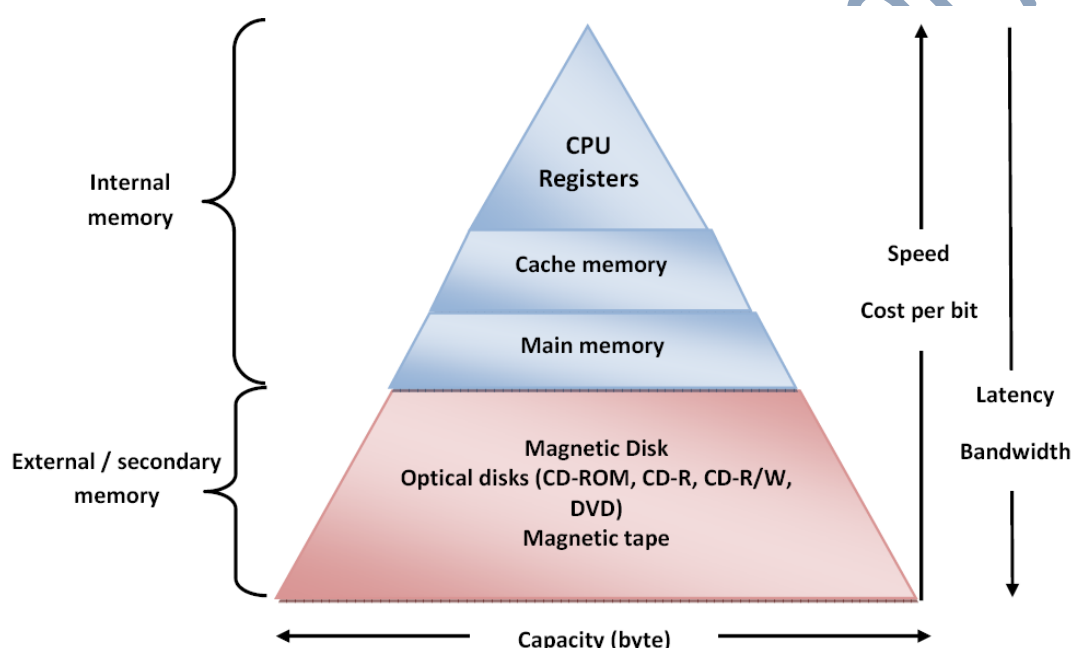


Fig 5.2 the parameters of memory hierarchy

The term random access refers to the fact that any access to any memory location takes the same fixed amount of time regardless of the actual memory location and/or the sequence of accesses that takes place. For example, if a write operation to memory location 100 takes 15 ns and if this operation is followed by a read operation to memory location 3000, then the latter operation will also take 15 ns.

This is to be compared to sequential access in which if access to location 100 takes 500 ns, and if a consecutive access to location 101 takes 505 ns, then it is expected that an access to location 300 may take 1500 ns. This is because the memory has to cycle through locations 100 to 300, with each location requiring 5 ns.

5.1.1 RAM and ROM Chips.

A RAM chip is better suited for communication with the CPU if it has one or more control inputs that select the chip only when needed. Another common feature is a bi-directional data bus that allows the transfer of data either from memory to CPU during a read operation or from CPU to memory during a write operation. So if the memory is 128 words of eight bits (one byte) per word. This requires a 7-bit address and an 8-bit bi-directional data bus. The read and write inputs specify the memory operation and the two chips select (CS) control inputs are for enabling the chip only when it is selected by the microprocessor. The availability of more than one control input to select the chip facilitates the decoding of the address lines when multiple chips are used in the microcomputer. The read and write inputs are sometimes combined into one line labeled R/W. When the chip is selected, the two binary states in this line specify the two operations of read or write. The unit is in operation only when $CS1 = 1$ and $CS2 = 0$. The bar on top of the second select variable indicates that this input is enabled when it is equal to 0. If the chip select inputs are not enabled, or if they are enabled but the read or write inputs are not enabled, the memory is inhibited and its data bus is in a high-impedance state.

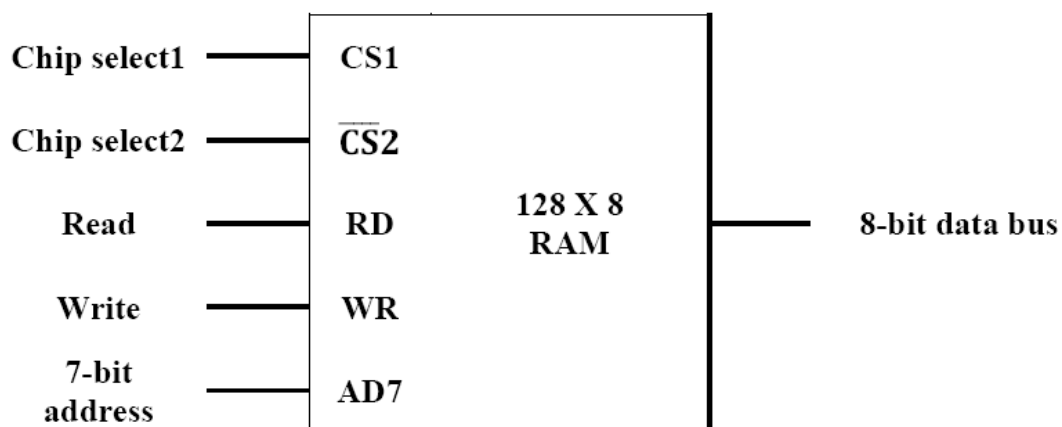
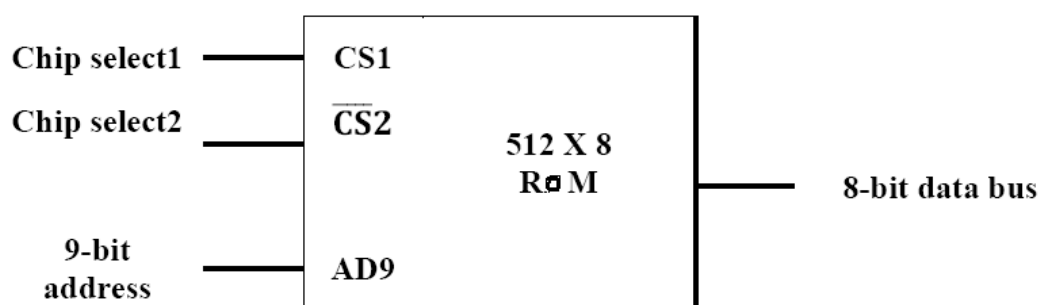


Fig 5.3 Typical RAM chip

When CS1 = 1 and CS2 = 0, the memory can be placed in a write or read mode. When the WR input is enabled, the memory stores a byte from the data bus into a location specified by the address input lines. When the RD input is enabled, the content of the selected byte is placed into the data bus. The RD and WR signals control the memory operation as well as the bus buffers associated with the bi-directional data bus. A ROM chip is organized externally in a similar manner. However, since a ROM can only read, the data bus can only be in an output mode. The block diagram of a ROM chip is shown in Figure 5.4. The diagram specifies a 512-byte ROM, while the RAM has only 128 bytes. The nine address lines in the ROM chip specify any one of the 512 bytes stored in it. The two chip select inputs must be CS1 = 1 and CS2 = 0 for the unit to operate. Otherwise, the data bus is in a high-impedance state.

**Fig 5.4 Typical ROM chip**

5.1.2 Memory Address Map.

The designer of a computer system must calculate the amount of memory required for the particular application and assign it to either RAM or ROM. The interconnection between memory and processor is then established from knowledge of the size of memory needed and the type of RAM and ROM chips available. The addressing of memory can be established by means of a table that specifies the memory address assigned to each chip. The table, called a memory address map, is a pictorial representation of assigned address space for each chip in the system. To demonstrate with a particular example, assume that a computer system needs 1024 bytes of memory (512 bytes of RAM by using 128 bytes RAM chips and 512 bytes of ROM by using a 512 ROM

chips).

So to solve this problem we make the sequence of solution below:

1. Number of RAM chips = total capacity of RAM / capacity for single chip

$$= 512 \text{ bytes} / 128 \text{ bytes}$$

$$= 4 \text{ chips.}$$

Number of ROM chips = total capacity of ROM / capacity for single chip

$$= 512 \text{ bytes} / 512 \text{ bytes}$$

$$= 1 \text{ chip.}$$

Component	Dec Address	Hex Address	Address Bus															
			16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
RAM 1	000	0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	127	007F	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	
RAM 2	128	0080	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	255	00FF	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	
RAM 3	256	0100	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
	383	017F	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	
RAM 4	384	0180	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
	511	01FF	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	
ROM	512	0200	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
	1023	03FF	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	

Table 5.1 memory address map

- The component column specifies whether a RAM or a ROM chip is used.
- The decimal address column assigns a range of decimal equivalent addresses for each chip.
- The hexadecimal address column assigns a range of hexadecimal equivalent addresses for each chip.
- The address bus lines are listed in the third column. Although there are 16 lines in the address bus.
- The RAM chips have 128 bytes and need seven address lines. The ROM chip has 512 bytes and needs 9 address lines.

5.1.3 Memory Connection to CPU.

RAM and ROM chips are connected to a CPU through the data and address buses. The low-order lines in the address bus select the byte within the chips and other lines in the address bus select a particular chip through its chip select inputs. The connection of memory chips to the CPU is shown in Figure 5.3. This configuration gives a memory capacity of 512 bytes of RAM and 512 bytes of ROM. Each RAM receives

the seven low-order bits of the address bus to select one of 128 possible bytes. The particular RAM chip selected is determined from lines 8 and 9 in the address bus. This is done through a 2x4 decoder whose outputs go to the CS1 inputs in each RAM chip. Thus, when address lines 8 and 9 are equal to 00, the first RAM chip is selected. When 01, the second RAM chip is selected, and so on. The RD and WR outputs from the microprocessor are applied to the inputs of each RAM chip. The selection between RAM and ROM is achieved through bus line 10.

The RAMs are selected when the bit in this line 0, and the ROM when the bit is 1. The other chip select input in the ROM is connected to the RD control line for the ROM chip to be enabled only during a read operation.

Address bus lines 1 to 9 are applied to the input address of ROM without going through the decoder. This assigns addresses 0 to 511 to RAM and 512 to 1024 to ROM. The data bus of the ROM has only an output capability, whereas the data bus

connected to the RAMs can transfer information in both directions.

The example just shown give an indication of the interconnection complexity that can exist between memory chips and the CPU.

The more chips that are connected, the more external decoders are required for selection among the chips. The designer must establish a memory map that assign addresses to the various chips from which the required connections are determined.

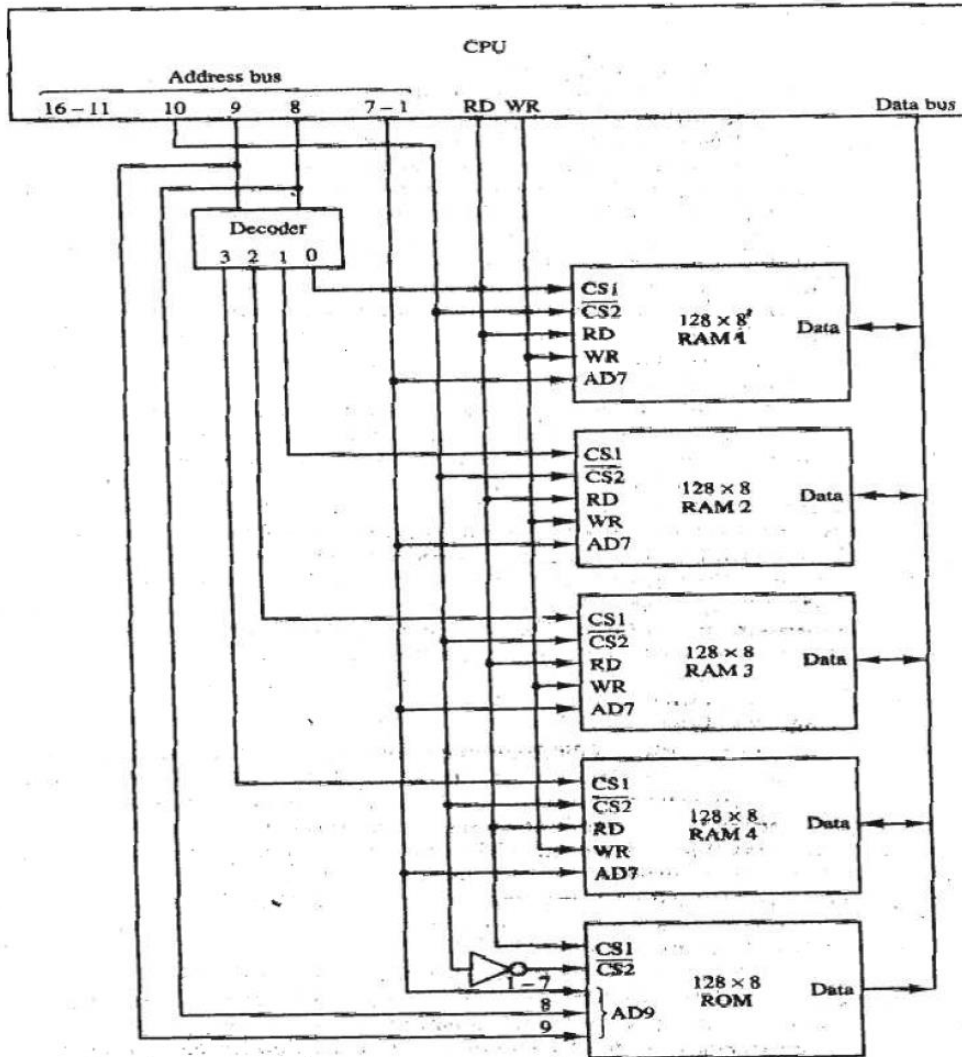


Fig 5.5 Memory chips connection to CPU.

Dr.Eng