



MEMORY HIERARCHY

OBJECTIVE

- Study about the various types of memory according to size, cost and speed.





Memory Hierarchy

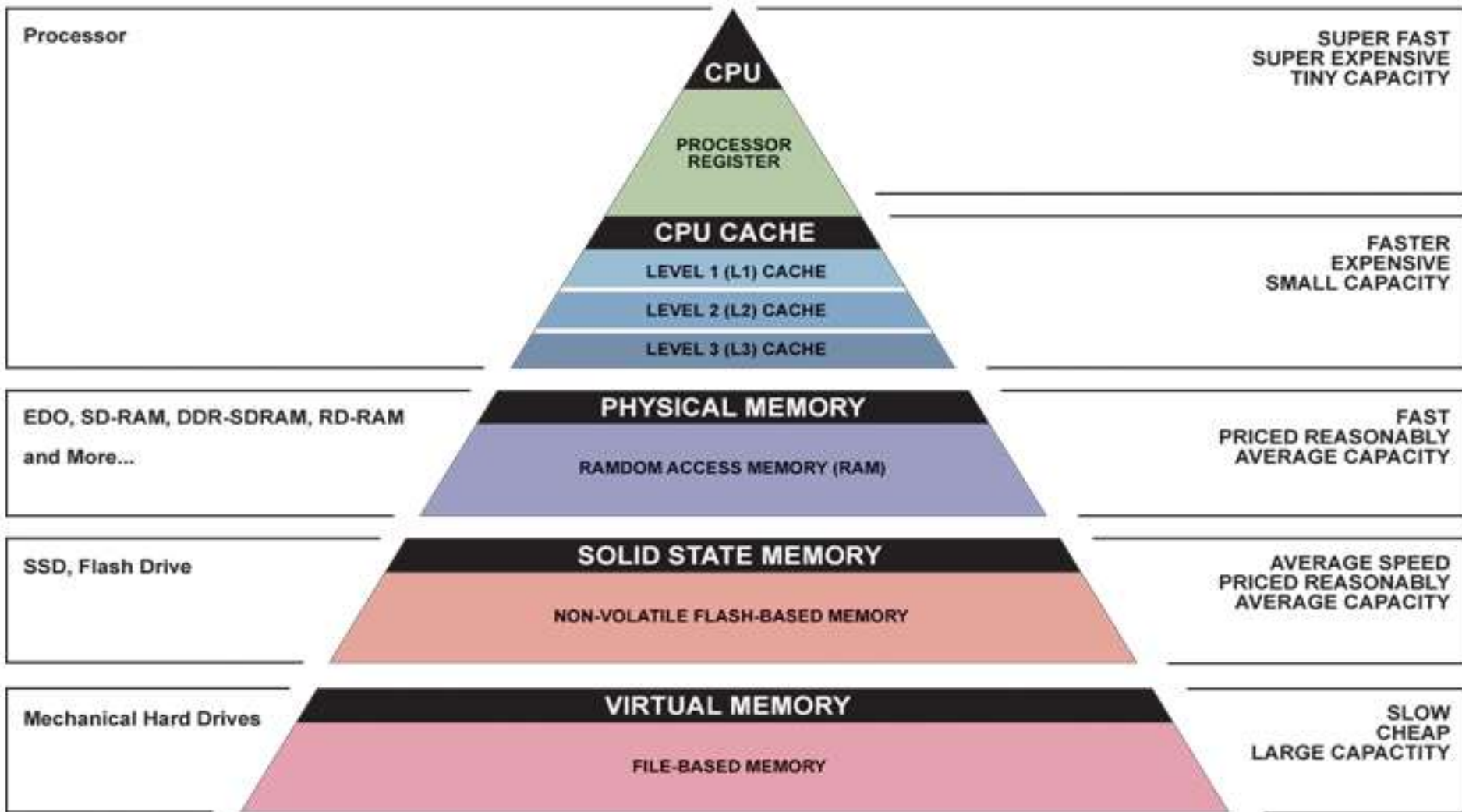
- The memory unit is an essential component in any digital computer since it is needed for storing programs and data
- Not all accumulated information is needed by the CPU at the same time
- Therefore, it is more economical to use low-cost storage devices to serve as a backup for storing the information that is not currently used by CPU

Memory Hierarchy

- Computer Memory Hierarchy is a pyramid structure that is commonly used to illustrate the significant differences among memory types.
- The memory unit that directly communicate with CPU is called the *main memory*
- Devices that provide backup storage are called *auxiliary memory*
- The memory hierarchy system consists of all storage devices employed in a computer system from the slow by high-capacity *auxiliary* memory to a relatively faster main memory, to an even smaller and faster *cache* memory



Figure



▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

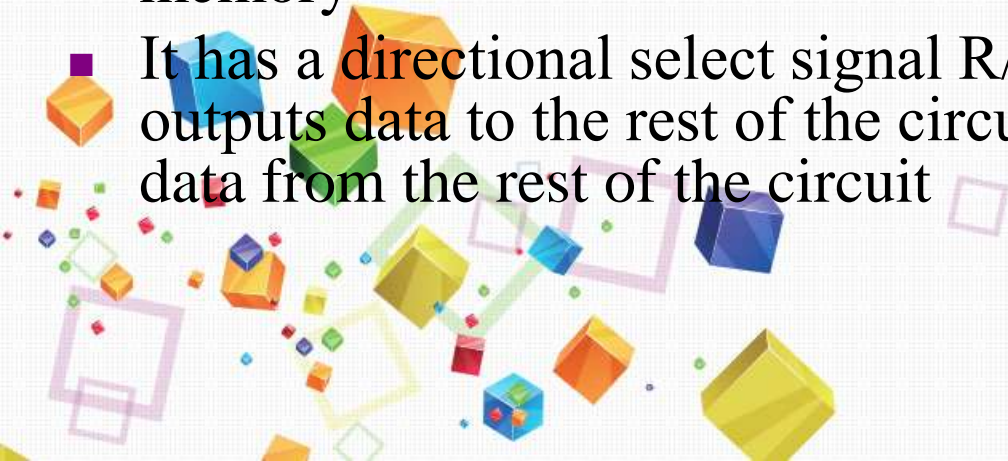
Main memory

- Most of the main memory in a general purpose computer is made up of RAM integrated circuits chips, but a portion of the memory may be constructed with ROM chips
- RAM– Random Access memory
 - Integrated RAM are available in two possible operating modes, *Static and Dynamic*
- ROM– Read Only memory



RAM

- A RAM chip is better suited for communication with the CPU if it has one or more control inputs that select the chip when needed
- Read/write memory, that initially doesn't contain any data
- The computing system that it is used in usually stores data at various locations to retrieve it latter from these locations
- Its data pins are bidirectional (data can flow into or out of the chip via these pins), as opposite to those of ROM that are output only
- It loses its data once the power is removed, so it is a volatile memory
- It has a directional select signal R/W' ; When $R/W'=1$, the chip outputs data to the rest of the circuit; when $R/W' = 0$ it inputs data from the rest of the circuit



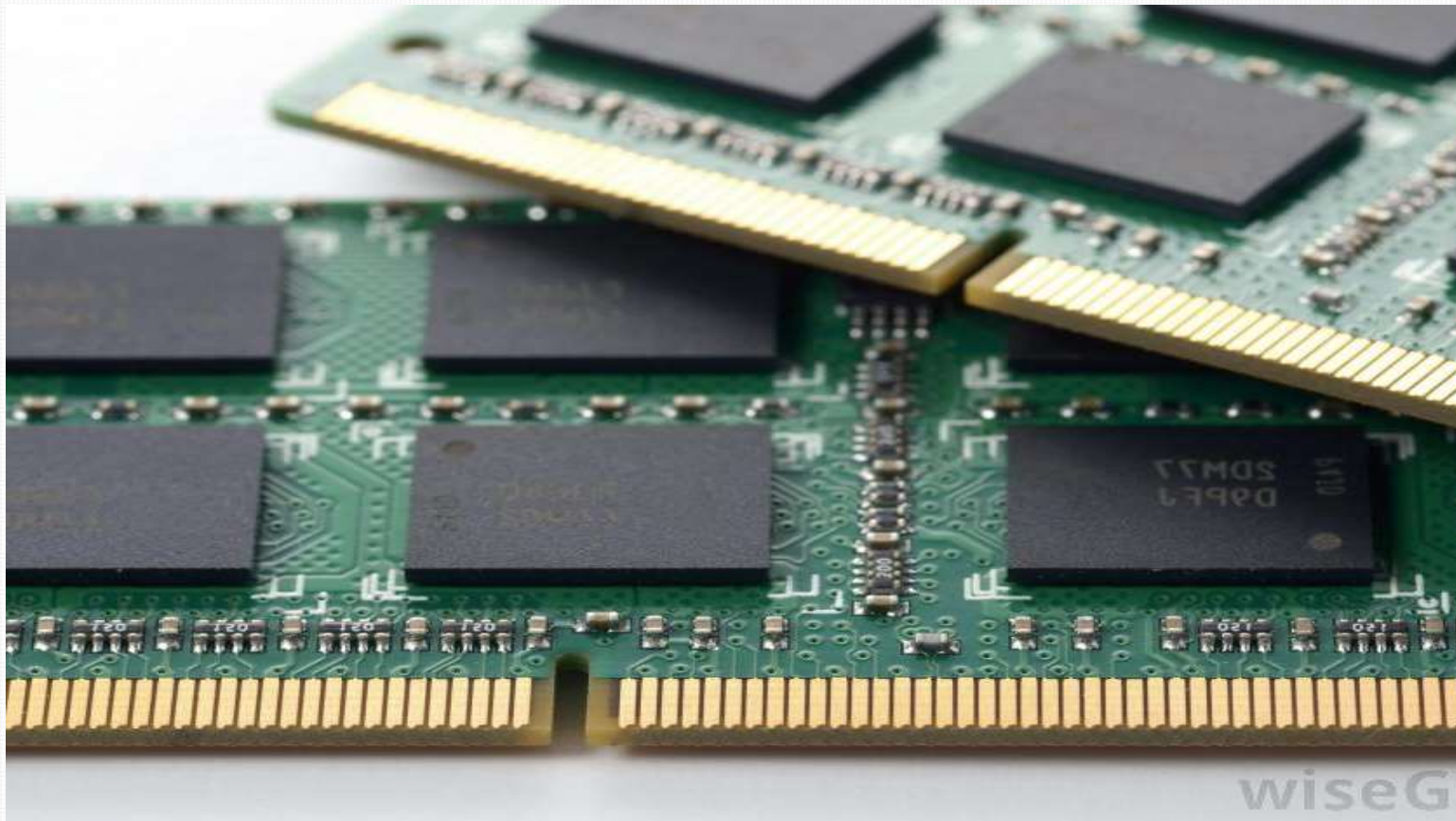


FIG: RAM



Random-Access Memory Types

Static RAM (**SRAM**)

- Each cell stores bit with a six-transistor (Diode) circuit.
- Retains value indefinitely, as long as it is kept powered.
- Relatively insensitive to disturbances such as electrical noise.
- Faster and more expensive than DRAM.

Dynamic RAM (**DRAM**)

- Each cell stores bit with a capacitor and transistor.
- Value must be refreshed every 10-100 ms.
- Sensitive to disturbances.
- Slower and cheaper than SRAM.



FIG :SRAM



FIG:DRAM



ROM

- ROM is used for storing programs that are **PERMENTLY** resident in the computer and for tables of constants that do not change in value once the production of the computer is completed.
- The ROM portion of main memory is needed for storing an initial program called *bootstrap loader*, witch is to start the computer software operating when power is turned off



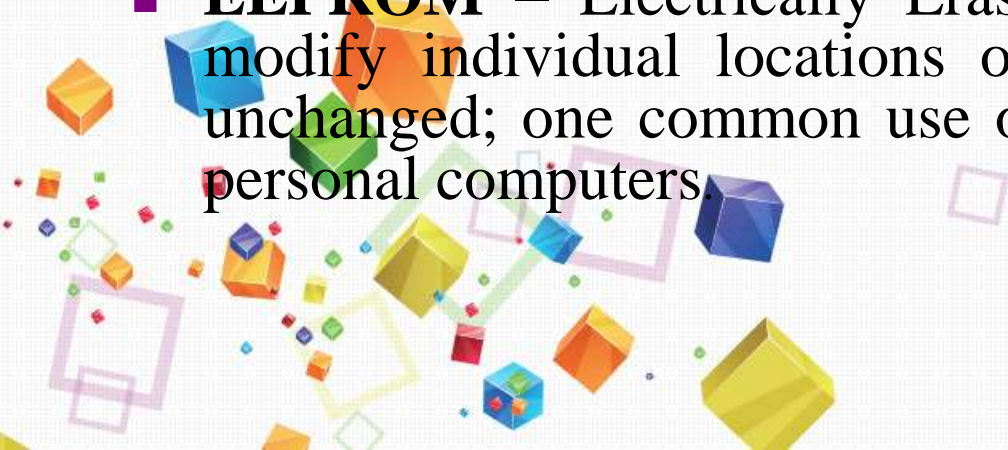


FIG: ROM



ROM Types

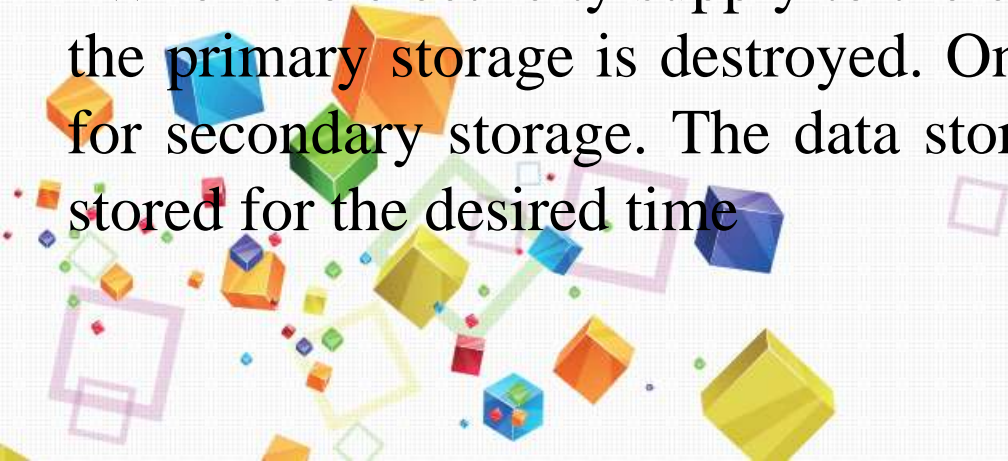
- **Masked ROM** – programmed with its data when the chip is fabricated
- **PROM** – programmable ROM, by the user using a standard PROM programmer, by burning some special type of fuses. Once programmed will not be possible to program it again
- **EPROM** – erasable ROM; the chip can be erased and chip reprogrammed; programming process consists in charging some internal capacitors; the UV light (method of erase) makes those capacitors to leak their charge, thus resetting the chip
- **EEPROM** – Electrically Erasable PROM; it is possible to modify individual locations of the memory, leaving others unchanged; one common use of the EEPROM is in BIOS of personal computers.



Auxiliary Memory

- The main memory construction is costly. Therefore, it has to be limited in size. The main memory is used to store only those instructions and data which are to be used immediately. However, a computer has to store a large amount of information. The bulk of information is stored in the auxiliary memory. This is also called backing storage or secondary storage. They include hard disk, floppy disks, CD-ROM, USB flash drives, etc.

- When the electricity supply to the computer is off, all data stored in the primary storage is destroyed. On the other hand, this is not true for secondary storage. The data stored in secondary storage can be stored for the desired time





Memory Card Reader



**USB Flash
Memory**



**Media
Devices**



External Optical Drives



FIG:AUXILLARY MEMORY DEVICES

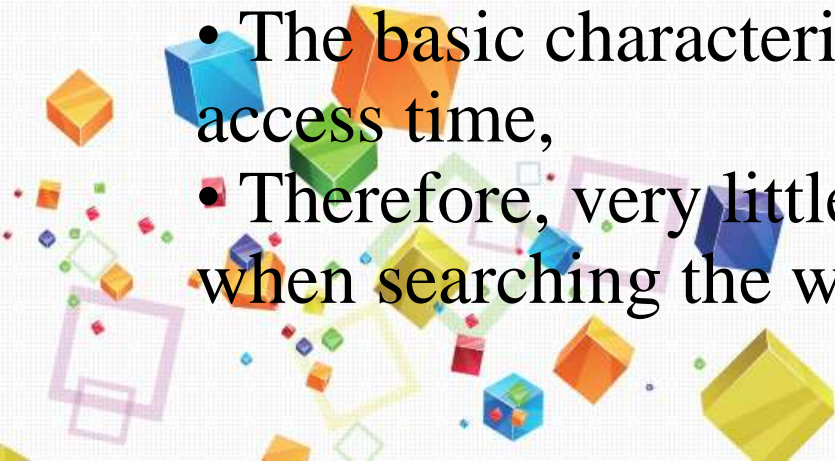
Cache memory

- If the active portions of the program and data are placed in a fast small memory, the average memory access time can be reduced,
- Thus reducing the total execution time of the program
- Such a fast small memory is referred to as cache memory
- The cache is the fastest component in the memory hierarchy and approaches the speed of CPU component



Cache memory cont.....

- When CPU needs to access memory, the cache is examined
- If the word is found in the cache, it is read from the fast memory
- If the word addressed by the CPU is not found in the cache, the main memory is accessed to read the word
- The basic characteristic of cache memory is its fast access time,
- Therefore, very little or no time must be wasted when searching the words in the cache



L1 ,L2 and L3 cache

- **L1 cache (2KB - 64KB)**
- L1 cache (also known as primary cache or Level 1 cache) is the top most cache in the hierarchy of cache levels of a CPU. It is the fastest cache in the hierarchy. It has a smaller size and a smaller delay (zero wait-state) because it is usually built in to the chip. SRAM (Static Random Access Memory) is used for the implementation of L1.



- **L2 cache (256KB - 512KB)**
- L2 cache (also known as secondary cache or Level 2 cache) is the cache that is next to L1 in the cache hierarchy. L2 is usually accessed only if the data looking for is not found in L1. L2 is typically implemented using a DRAM (Dynamic Random Access Memory). Most times, L2 is soldered on to the motherboard very close to the chip (but not on the chip itself), but some processors like Pentium Pro deviated from this standard



- L3 cache (1MB -8MB)
- Level 3 Cache - With each cache miss, it proceeds to the next level cache. This is the largest among the all the cache, even though it is slower, its still faster than the RAM.



Difference between L1, L2 and L3 cache

- All processors rely on L1 cache, this is usually located on the die of the processor and is very fast memory (and expensive). L2 cache is slower, bigger and cheaper than L1 cache. Older processors used L2 cache on the motherboard, nowadays it tends to be built in to the processor. L3 cache is slower, bigger and cheaper than L2 cache. Again this can be on chip or on the motherboard.



Virtual memory

- *Virtual memory* is imaginary memory: it gives you the illusion of a memory arrangement that's not physically there.



Thank you....

BY...

SREELAKSHMI KV

