CAPE COMPUTER SCIENCE UNIT ONE
MODULE I

(MAIN) MEMORY

Main memory is also known as primary memory or simply memory.
There are two main types of main memory :

1. **ROM**: Read Only Memory, a special type of memory that is built onto the motherboard and preprogrammed by the computer manufacturers. i.e. it is hard-wired on a chip. It contains the data and instructions necessary for the computer to start up itself and readies the system for operation. This is called booting, so booting is carried out by the code prewritten into the ROM microchip (in the BIOS). BIOS is firmware. *Firmware is software that is built into a piece of hardware by the manufacturer.*

BIOS = Basic Input Output System : a program stored on ROM which is run on power on of the machine.

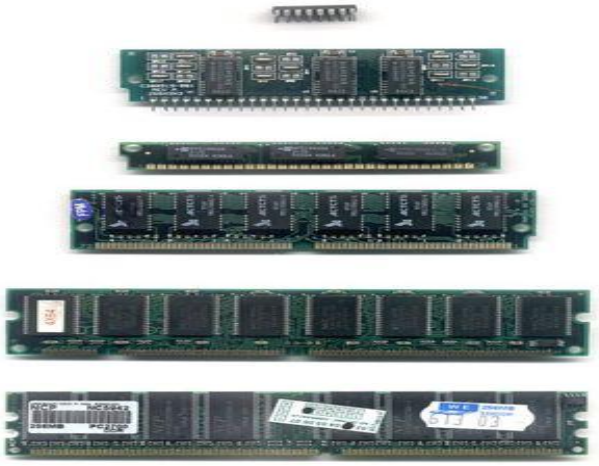
There are instances though when the ROM doesn't have any code in it. These are usually built for specialized computers though, not generalized machines like PC's. In this case the machine comes with a PROM (Programmable Read Only Memory) and the user gets to load the contents with what they want. This can only be done once.

In special cases you will also find EPROM (erasable programmable read only memory) and EEPROM (electronically erasable programmable read only memory) for highly technical environments in which they want to be able to scratch their machines and start fresh. The highly technical user in those cases gets to erase the contents of ROM and reprogram it if and when they desire. EPROM can be erased by exposure to ultraviolet light while EEPROM can be erased by high voltages.

Generally though the contents of ROM are permanent. It is not affected by loss of power to the computer and is therefore regarded as involatile.

2. **RAM** is Random Access Memory. RAM is high-speed electronic memory containing all currently active data and programs. After a computer is turned on, its RAM becomes active and ready. After booting the operating system is loaded into a special part of RAM. The operating system then takes over the running of the machine. The first program in RAM therefore is always the operating system.

RAM is removable and upgradeable and is located near to the CPU. This proximity to the CPU and its great sped mean that it is also called immediate access memory. RAM provides the processor with temporary storage for running programs and data waiting for the processor or just processed. All programs and data must be transferred to RAM from an input device or a storage device before they can be executed or operated on. RAM space is premium so when the program/data is no longer needed, the space is reallocated to the next program waiting.



RAM is volatile, it is made of semiconductor material (CMOS - complimentary metal oxide semiconductor) which only works when it has an electricity supply, so any data in RAM is lost when the machine is turned off.

New microcomputers now typically come with no less than 1GB of RAM. The RAM chip is about 1 inch high and 4 to 6 inches long.

RAM

- stores data inputted into the computer
- holds the operating system
- holds running applications
- stores data about to be outputted

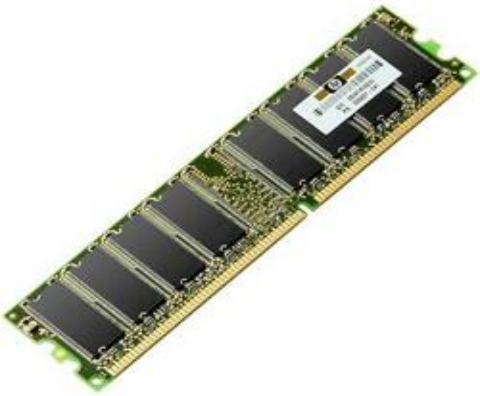
Types of RAM

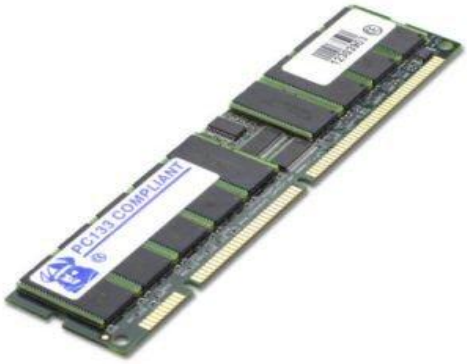
The 2 most common types of RAM are SRAM (Static RAM) and DRAM (Dynamic RAM). DRAM requires that its contents be flushed out every 4 ms, so it is slow, but it consumes little power.
SRAM is significantly faster as no refreshing is necessary but it uses lots of power and generates lots of heat.

A derivative of DRAM called SDRAM (Synchronous DRAM) is beginning to take off. Still another type of DRAM named DDR SDRAM (double data rate SDRAM) is also used by newer motherboards.

RAM can also be classified according to whether it is SIMM or DIMM. This refers to the physical makeup of the memory card.
SIMM – single in line memory module
DIMM – dual in-line memory module

SIMM and DIMM differ in the number of pins and notches they have for connection to the motherboard. The DIMM has two notches and the SIMM has only one. They are therefore not interchangeable. A particular motherboard is built to either use SIMM or DIMM. SIMMs are less expensive but DIMMs have a 64-bit data path (twice that of the SIMM).





Cache A high speed holding area for frequently used program instructions and data. Cache is about 10 times faster than RAM so it takes less time to access the data thus speeding up the processes. Cache may be a special chip (cache memory chip) that sits between the processor and RAM or more generally it may be a part of the processor. Cache which is inside the processor is called primary cache. If the cache is outside the processor on the motherboard it is termed secondary cache

Naturally cache is expensive but most computers will have a little cache memory to speed up the time it takes to retrieve data for use by the CPU. Cache is simply another layer of memory between RAM and the CPU designed to improve throughput. Cache is checked first; if the data is not there then RAM is next checked. Thus the role of cache is to speed up the accessing of data for the processor so that it can operate as near as possible to its maximum speed.
Usually the operating system will store the most frequently used data and instructions in cache.
If cache is too large it gets slower and so defeats its purpose of caching in the first place.

RAM, ROM and its derivatives are very fast. The semiconductor has no moving parts (the electrons move). Any piece of equipment with no moving parts will be faster and more durable than a mechanical device. As a result though, they are expensive so there's a limit to how much can be included in one machine.

Data Storage Review

The memory of all digital computers are bistable devices. Meaning it can be set to only one of two states (0 or 1). This is a bit.

Data Storage

The amount of data and instructions that can be stored in the memory of a computer is measured in bytes. A byte is 8 bits. A byte stores one character i.e. a number, a letter, a symbol, a punctuation mark or a blank space. What is a nibble? 4 bits

1024 bytes = 1 KB

1024 x 1024 bytes = 1024 KB = 1 MB

1024 x 1024 x 1024 bytes = 1024 x 1024 KB = 1024 MB = 1GB

1024 x 1024 x 1024 x 1024 bytes = 1024 x 1024 x 1024 KB = 1024 x 1024 MB
= 1024 GB = 1 TB

bits → bytes → characters → fields → records → files

So data is stored in files in secondary storage. Information about the file must also be stored in a directory. On disk this directory is called the FAT or the MFT (master file table).

Data is transferred to and from storage devices in blocks. A block is a minimum of 128 bytes and can go up to tens of thousands of bytes. This is determined by the bandwidth of the external data bus. The bandwidth varies from machine to machine.

Terms

Access time: time elapsing between the request for transfer of data from a disk storage device to RAM and the instant this operation is complete.

Memory consists of a series of sequential locations each a set number of bits wide. The size of each location is the word size of the machine (CPU). The wider the slots the faster the machine as more data is transferred to the CPU on each transfer because that CPU can handle more data at a time. Typical word sizes are 16 bits (older machines), 32 bits, 64 bits, 128 bits.

The CPU is naturally very very fast. How fast it actually operates depends on the amount of cache and RAM it is working with.

Word - the maximum amount of data that the CPU can process at any one time

Word size – the capacity of the CPUs data bus/amount of data transferrable/ 'processible'

Typical sizes for RAM in microcomputers – 512MB, 1GB, 2GB, 4GB