Processor

The word *processor* has several meanings, when related to information technologies; however most of them are related to a chip, which is embedded in computer hardware. Examples are:

- **Central Processing Unit (CPU)**
- **Graphics Processing Unit (GPU)**

Physics Processing Unit (PPU) Central Processing Unit

The computer CPU is responsible for handling all instructions and calculation it receives from other hardware components in the computer and software programs running on the computer.

First CPUs were complicated circuits with many elements (relays, vacuum tubes, discrete transistors)

Example: ENIAC contained 17,458 vacuum tubes, 7,200 crystal diodes, 1,500 relays, 70,000 resistors, 10,000 capacitors, and 5 million hand-soldered joints. The total power consumption was 160 kW. In one second, the ENIAC could perform 5 000 additions, 357 multiplications, or 38 divisions. Btw. the weight of the whole contraption was about 27 tons).

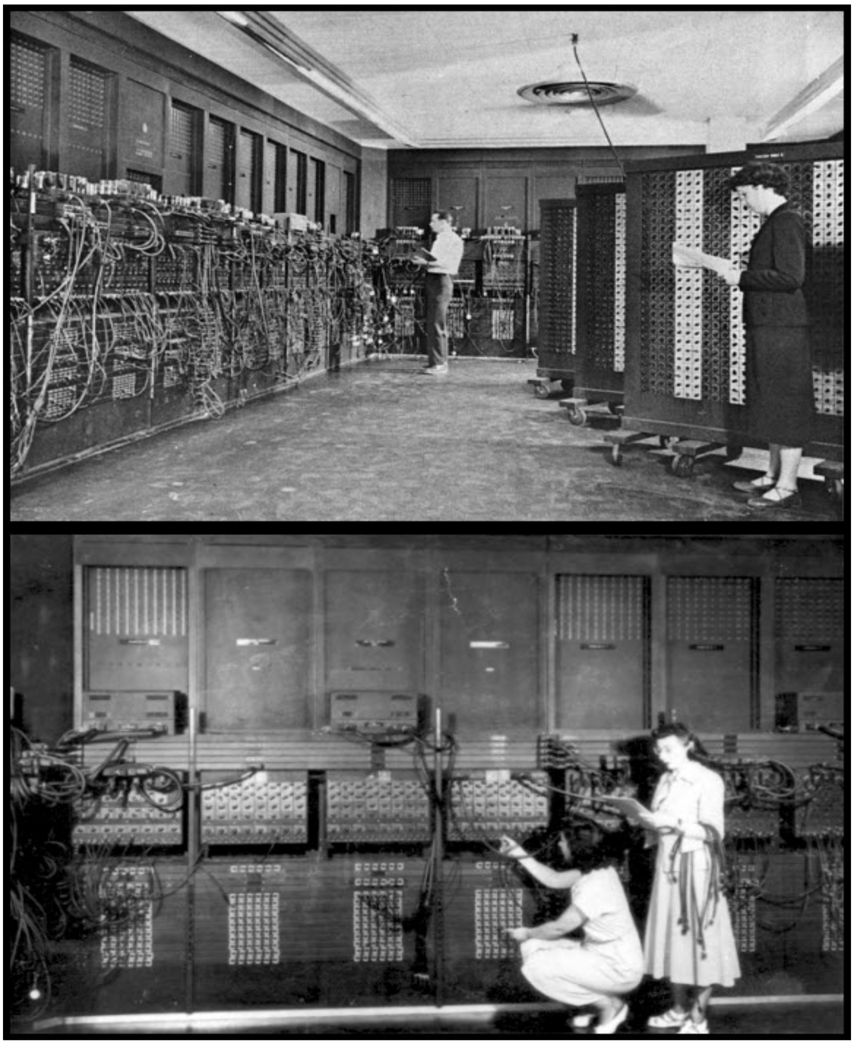


Figure 1: ENIAC and Its Programming (source: http://royal.pingdom.com/2009/12/11/retro-delight-gallery-of-early-computers-1940s-1960s/)

In 1958 Jack Kilby (Texas Instruments) demonstrated the first **INTEGRATED CIRCUIT** – a circuit made of one piece of semiconducting material (he was awarded the Nobel Prize for that invention in 2000).

The **first one-chip CPU** – **Intel 4004 (1970)**; its successor was extremely successful **Intel 8008** (its parameters could be considered a bit funny nowadays – e.g. clock rate 2 MHz and 6000 integrated transistors - however it was a real breakthrough).

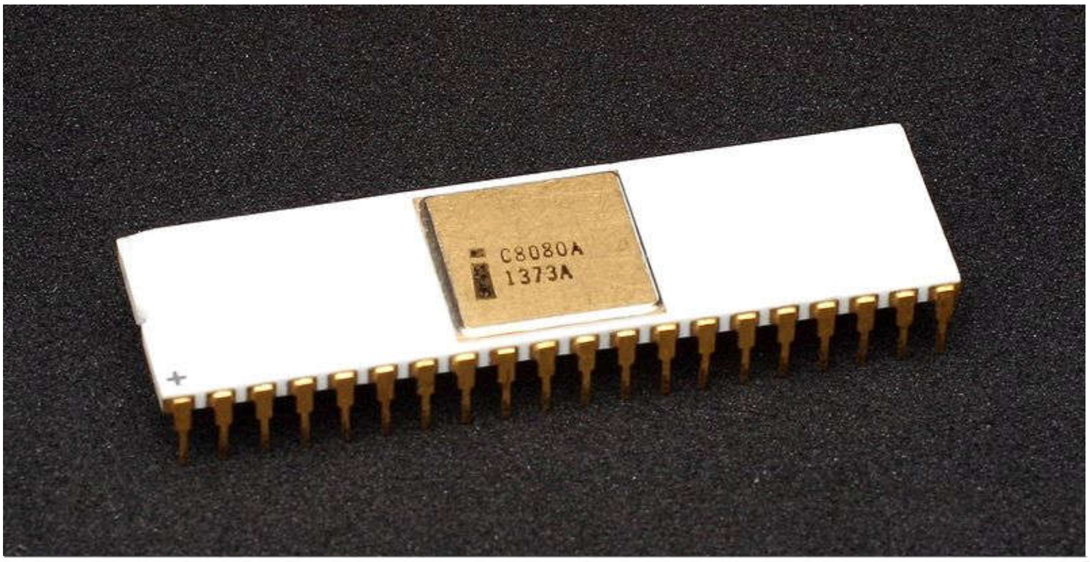


Figure 2: Intel 8008

Present

- Two major producers of CPUs for desktops, laptops and servers: **Intel** and **AMD** for the **x86 architecture** (i.e. CPUs with the backward compatibility with Intel 8086 processor) – Intel Core i7, i5, i3, Celeron; AMD Ryzen
- **ARM architecture** – major share in smartphones, tablets, smart TVs and other devices, where the power efficiency is crucial.

Other architectures either failed in the competition with x86 architecture, or they are used in more specific systems (Intel 8051) or router, servers and supercomputers (Power Architecture, Intel Itanium, MIPS).

Physical Properties of CPU

CPU – the main chip in the computer

Chip

- a pattern of interconnected switches and gates on the surface of a semiconducting crystal (usually Silicon)
- Switches and gates are made of
    1. N-type silicon areas,
    2. P-type silicon areas,
    3. insulating zones,
    4. conductive lines, which connect semiconductive parts – usually copper
- **Chip layout** – the placement of areas and the conductive lines

- The layout is transformed to a series of masks – each represents one layer of the chip
- The masks are *photolitographically* projected onto a **die** – rectangular zone of a **wafer.**

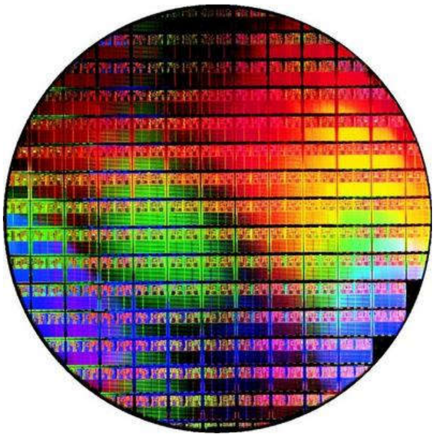
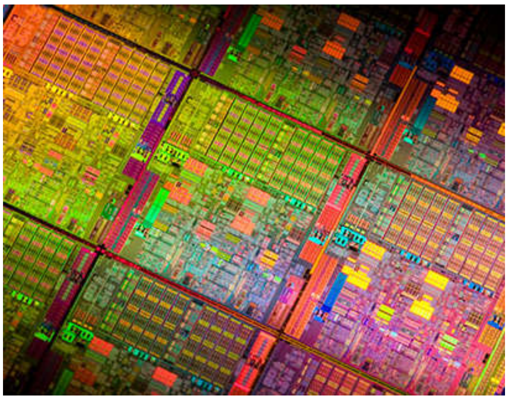



Figure 3: Silicon Wafer and Die

The chip main element – **transistors** (Intel Core Core i5-5257U: 1.3 billion transistors).

Moore's Law – a long-term trend in the CPU manufacturing process: the number of transistors, which can be inexpensively placed on a chip is doubled every 1.5 to 2 years (but it does not mean the performance is doubled as well!!! The overall performance depends on multiple factors).

Logical Structure of the CPU

There are 3 blocks:

1. **Arithmetic-logic unit**
2. **Control unit**
3. **Register unit**

Register unit contains **REGISTERS** – storage units of the predefined size, which can hold data or instructions.

Registers:

1. GENERAL-PURPOSE REGISTERS – no specific use, temporary storage for data processed by instructions
2. SPECIAL-PURPOSE REGISTERS – its content represents specific value, which can be modified by privileged instructions or not at all (can be read only) – e.g. program counter, instruction register, flag register (it indicates situation like overflow, resulting zero etc.)

Control Unit

The Control Unit has three main jobs:

1. It CONTROLS AND MONITORS THE HARDWARE ATTACHED to the system to make sure that the commands given to it by the application software are used.
   For example, if you send something to print, the control unit will keep a check that the instructions are sent to the printer correctly.
2. It CONTROLS THE INPUT AND OUTPUT of data so that the signals go to the right place at the right time

3. IT CONTROLS THE FLOW OF DATA WITHIN THE CPU - which is the **FETCH-DECODE-EXECUTE CYCLE**.

Arithmetic-logic Unit

It performs basic arithmetic and logic operations, which are done by circuits – properly interconnected gates. Typical operations are addition, subtraction, comparison, or negation.

ALU calculations are typically limited to the integer domain. Floating point operations are handled by the **FLOATING-POINT UNIT** – a part of the CPU, which has optimised design for the floating-point calculations. Its speed is measured in **FLOPS** – floating-point operations per second.

Instructions, Machine Language, Assembly Language

INSTRUCTION – elementary operation, which can be processed by the CPU.

INSTRUCTION = OPCODE + OPERANDS

OPCODE – specification of the activity the CPU is to undergo.

MICROCODE – hardware implementation of the instruction. It can vary for the same instruction (e.g. Intel and AMD CPUs have compatible instructions; however, their implementations – the microcode-are very different).

INSTRUCTION SET – definition of the instructions, data types, registers, memory access etc.

Two types of instruction sets (mentioned in the previous document):

- **CISC**
- **RISC**

CPU Core

A core is the **basic computation unit of the CPU** - it can **run a single program context** (or multiple ones if it supports HW threads such as hyperthreading on Intel CPUs), maintaining the correct program state, registers, and correct execution order, and performing the operations through ALUs.

Modern CPU have multiple core, so it is capable of parallel processing – it can run multiple programs (or its parts) at the same time – it improves the computer performance. Single core CPU can emulate that behaviour by switching the programs; however, it will be slower than the true parallel execution.

CPU Clock Rate

It is the fundamental **rate in cycles per second at which a computer performs its most basic operations** such as adding two numbers or transferring a value from one register to another. Current CPUs have clock rates measured in GHz.

The clock rate helps to identify the performance of the CPU, but only when compared with CPUs of the same family (e.g. Core i3 processors), but it is almost useless for comparison of different CPU types.

Cache Memory

The cache is a special type of the primary memory, which can be accessed much faster than RAM. CPU copies a block of RAM, which contains the currently processed instructions and data. When the required data/instruction is not in the cache (*cache miss*), it loads another block of RAM with the required data/instruction.

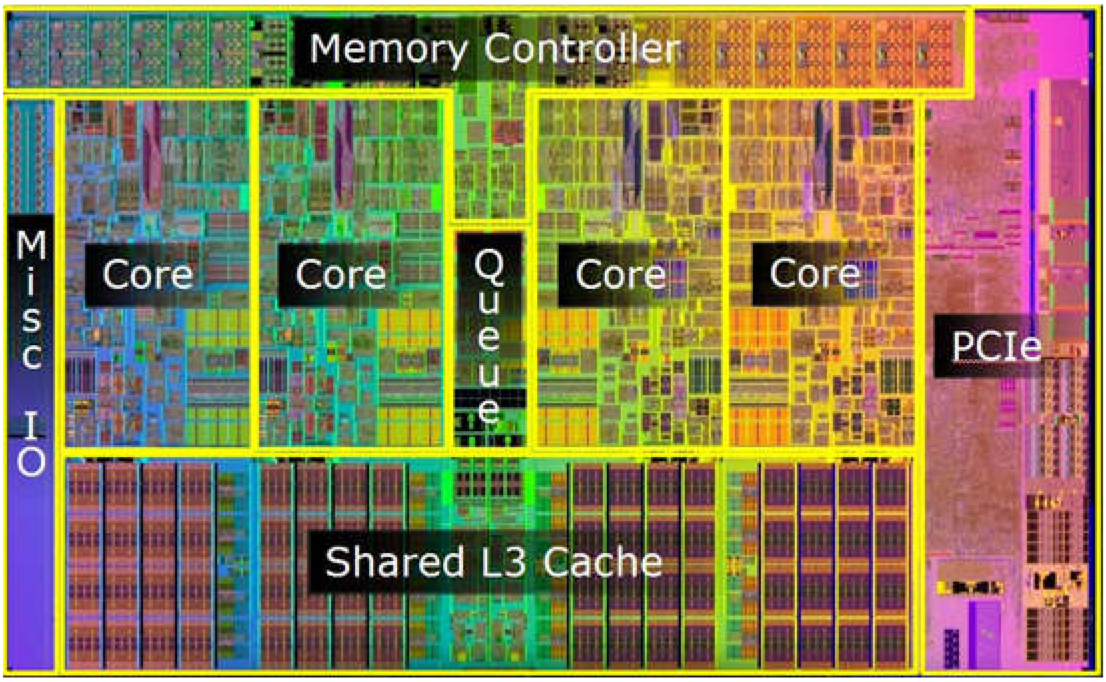
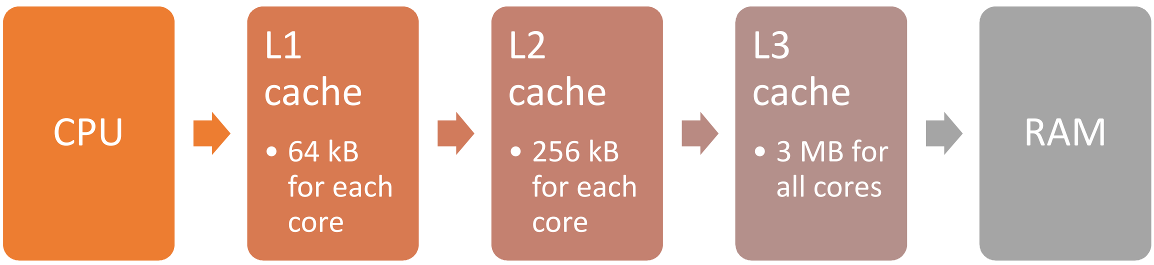


Figure 4: Layout of the Core i5 CPU (source: http://www.pcstats.com/articleview.cfm?articleid=2581&page=4)

There are usually 2 or 3 levels of the cache:

- L1 is the smallest and fastest one; there are usually 2 – one for data, one for instructions
- L2
- L3 – usually shared among cores



Memory Controller

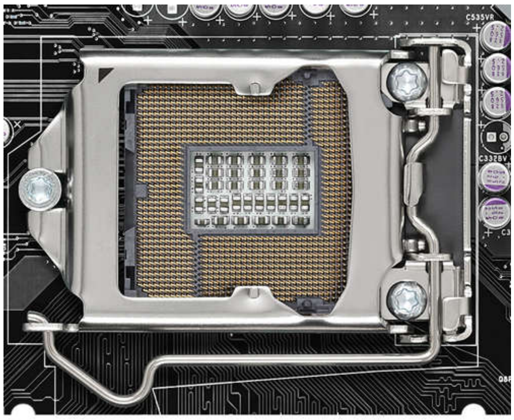
A digital circuit, which manages the cooperation between the CPU and RAM. The circuits handle writing and reading operations as well as the refreshing of DRAM (it will be explained in the next document).

SoC – System-on-Chip

A chip, which integrates all vital parts of the computer system – e.g. CPU + GPU + memory + USB controller + network controller (Ethernet or WiFi).

These can be found in tablets, smartphones etc. – e.g. Qualcomm Snapdragon, Samsung Exynos, MediaTek Helio.

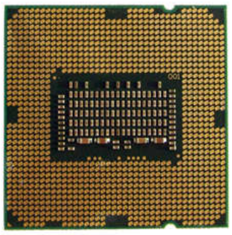
Socket



Socket – the physical connection of the CPU and a motherboard. Different producers use different sockets and a new CPU generation often introduces a new socket (e.g. AMD: socket AM3, AM4; Intel: Socket 1150, Socket 1151).

TDP

Power Consumption – referred by TDP – how much energy is required by the CPU cooler to dissipate the CPU heat to keep it at its highest allowed temperature.



GPU

- Graphics Processing Unit

A microprocessor dedicated to graphics rendering.

It is usually on a graphics card (**dedicated graphics card**) or sometimes right on the PC's mother board or even in the CPU (**integrated graphics card**). It has on-board implementation of floating point arithmetic, graphics algorithms and 2D and 3D primitives. Their execution is done entirely by the GPU without taking clock cycles from the main processor, so it offloads the CPU. The result is a much faster rendition of complex graphics in applications such as games, maps, cartography, CAD modeling; accelerated video etc.

Alternative use – GPUs are designed to support massive parallelism (simultaneous processing of tasks), so it can be used for tasks like deciphering of encrypted data or simulations of complex systems (environmental simulations, medical simulations).

- Producers
  - **AMD** – ***Radeon*** (games), ***FirePro*** (professional design, CAD)
  - **NVIDIA** – ***GeForce*** (games), ***Quadro*** (professional design, CAD)

Memory

Primary and Secondary Memory

Modern electronic computers generally possess several distinct types of **memory**, each of which stores information for subsequent use. The vast majority of computer memory can be placed into one of two categories: **primary memory** and **secondary memory**.

Primary memory

- Often called **main memory**, constitutes that device, or group of devices, that holds instructions and **data** for rapid and direct **access** by the computer's **central processing unit** (**CPU**).
- Primary memory is synonymous with **random-access memory (RAM)**. As a computer performs its calculations, it is continuously reading and writing (i.e. storing and retrieving) information to and from RAM. For instance, instructions and data are retrieved from RAM for processing by the CPU, and the results are returned to RAM.

- RAM is a **volatile** form of information storage, meaning that when electrical power is terminated any data that it contains is lost.
- There are other semiconductor memory devices accessed by the CPU that are generally considered as being distinct from primary memory (i.e. different from RAM). These memory units include:
  - **Cache Memory**
  - **Read-Only Memory (ROM)**
  - **Programmable Read Only Memory (PROM)**

- Modern RAM is made of **semiconductor** circuitry, which replaced the magnetic core memory widely used in computers in the 1960s.

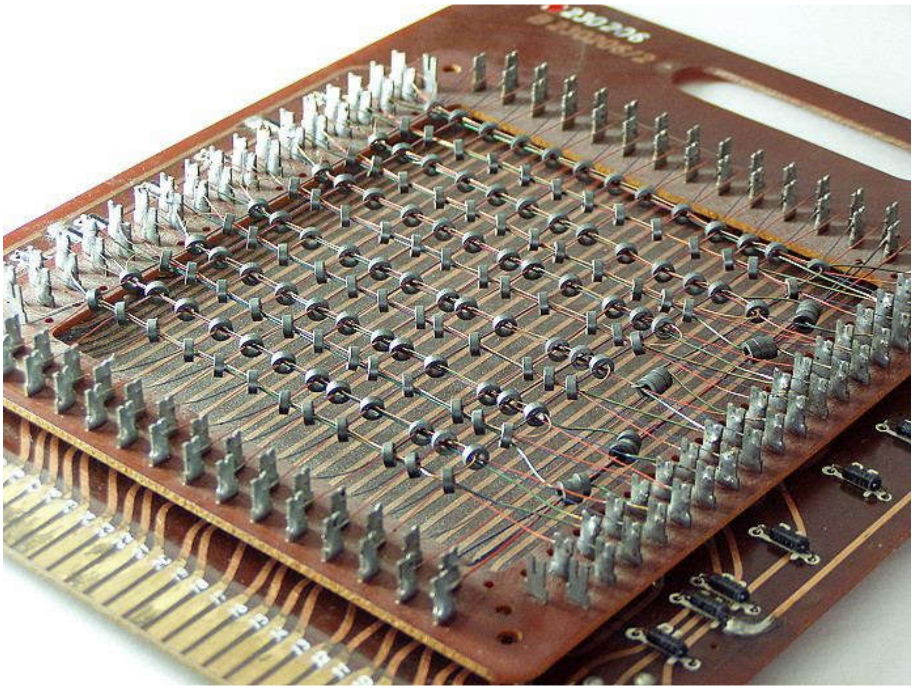


Figure 5: Magnetic Core Memory

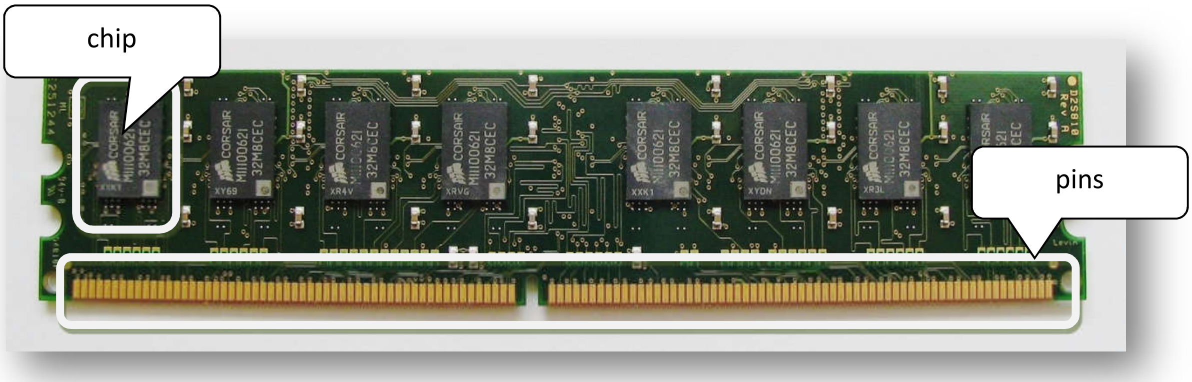
The main type – **DRAM = DYNAMIC RAM**

- Each memory cell holds one bit of information and is made up of two parts:
    1. a **transistor,** and
    2. a **capacitor**.
       The capacitor holds the bit of information -- a 0 (discharged) or a 1 (charged). The transistor helps to keep or modify the value held in the capacitor. However, the capacitor keeps the charge for short time only, so it has to be refreshed – done by the memory controller.
- Advantage: simple and cheap
- Disadvantage: ***dynamic*** - its content must be refreshed millions of times per second, otherwise it is lost – it slows down the memory and causes delays.

- Types
    - SDRAM
    - DDR
    - DDR2
    - DDR3
    - DDR4

Packaging

- Memory modules – several chips of DRAM on a single circuit board with standardized connectors (pins).



- Numbers of pins depends on the DRAM type and the computer type
  - desktop: SDRAM – 168, DDR – 184, DDR2 and DDR3 – 240 (DIMM modules)
  - notebook: DDR and DDR2 – 200, DDR3 – 204 (SO-DIMM modules)

Performance

1. **Clock Rate** (in MHz) - the highest frequency the module is able to cooperate with (e.g. DDR2-800 = DDRAM2 with the top frequency 800 MHz – *in fact the half, i.e. 400 MHz*).
2. **Timing**[1] - Timings measure the time the memory chip delays doing internal operations. In general – the longer the worse.
   a. **CAS Latency**
   b. **RAS to CAS Latency**
   c. **RAS Precharge**
   d. **Active to Precharge Delay**

Figure 6: A memory module with the 5-5-5-15 timing

Static RAM

Bits are stored in the latching circuits – there is no need to refresh them. It is faster than DRAM, but chips are more complicated → expensive with less capacity.

Use:

---

[1] The chipset accesses the ROW of the memory matrix by putting an address on the memory's address pins and activating the RAS signal.
There is a waiting period for a few clock cycles (this is called the RAS-to-CAS Delay).
The column address is put on the pins and the CAS signal is activated to access the correct COLUMN of the memory matrix.
There is a wait a few clock cycles - this is known as CAS latency - and then the data appears on the pins of the RAM.
The data at this time is clocked and read, usually one or two clock cycles.

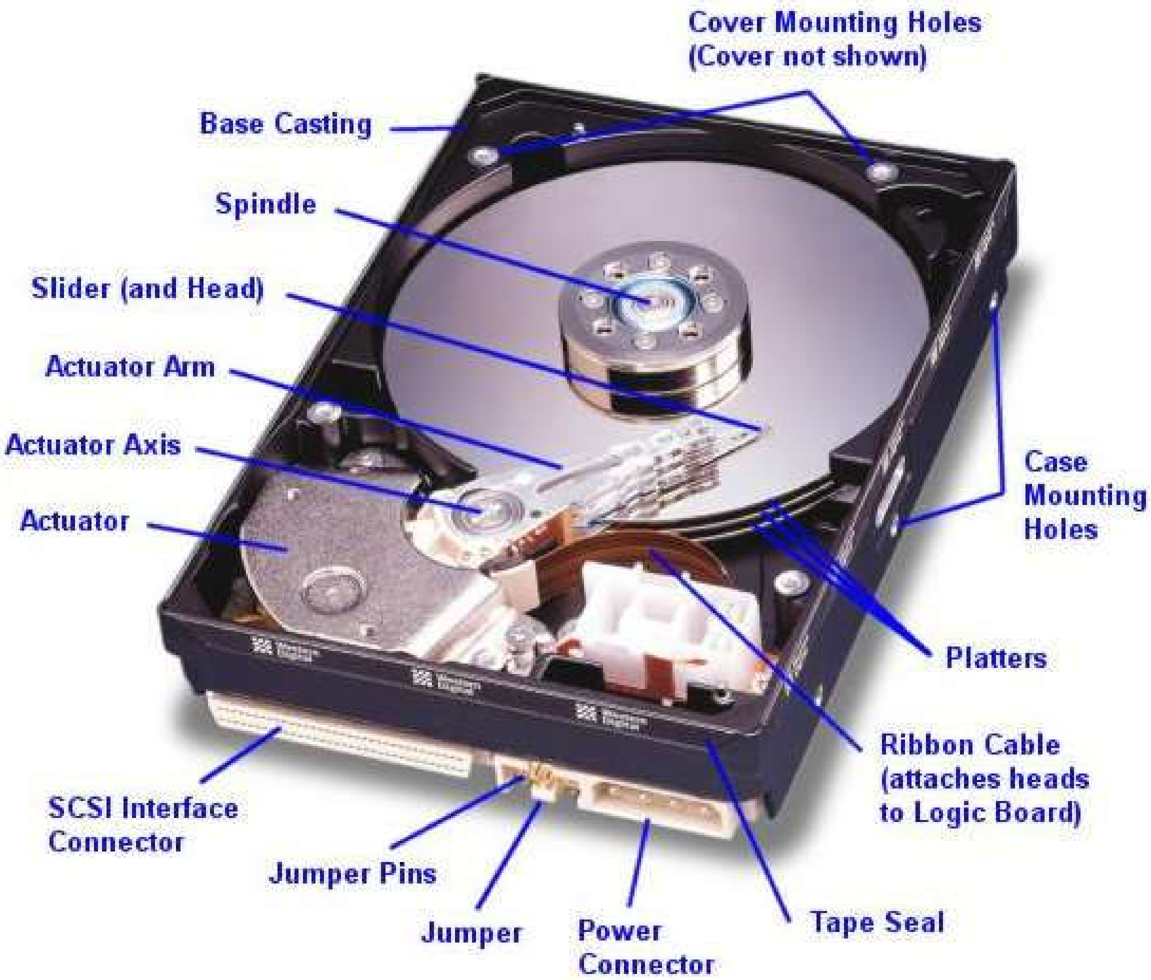
- Cache in CPU
- Registers in CPU
- HDD buffers

Secondary memory

- Also called **auxiliary memory** or **mass storage**, consists of devices not directly accessible by the CPU.
- Reasons, why secondary memory is used:
    1. First, most secondary storage devices are capable of **CONTAINING MUCH MORE INFORMATION THAN** is feasible for **PRIMARY MEMORY** (hence the use of the term "*mass storage*" as a synonym for secondary memory).
    2. A second feature of **SECONDARY MEMORY** is that it **IS NON-VOLATILE**. This means that data is stored with or without electrical power being supplied to the device, as opposed to RAM, which can retain its data only as long as electrical power is present.
- Medias widely used for secondary storage are:
    - **Hard Drives**
    - **Floppy Disks**
    - **Punch Cards**
    - **Optical Disks**
        - **CD**
        - **DVD**
        - **Blu-ray**
    - **Flash memory-based Devices**
        - **Flash Disk**
        - **SSD drives**
        - **Memory Cards**
    - **Tapes**
- The **input and output** of these devices is much slower than for the semiconductor devices that provide the computer's primary memory.

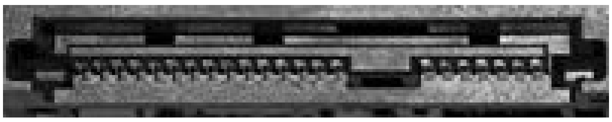
Hard Disks

- A Hard Disks Drive (**HDD**) is a non-volatile secondary storage device, which stores digitally encoded data on the **magnetic surface** of it.

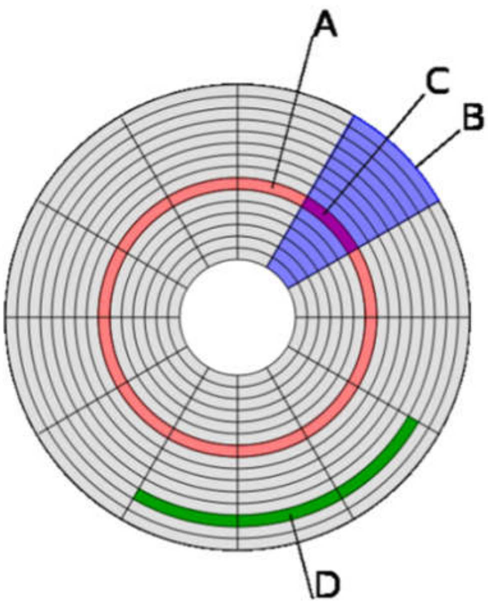


- Typical capacity: hundreds of GB to several TB(2017 – 12 TB)
- Size: 3,5" – in desktop computers, 2,5" – in laptops; all sizes also in external drives
- Performance
    1. **Transfer Rate** – how many MB can be loaded from/saved to the HDD (nowadays 80-130 MB/s)
    2. **Seek Time** – time required to move the read/write head to the place, where the required data is stored
- Performance is influenced also by the rotation of plates – the speed can be 5400 rpm (revolutions per minute) - laptops, 7200 rpm - desktops, 10 000 rpm, or 15 000 rpm (servers/high performance computers).

Data Interfaces – main types

SATA – latest type of interface	
ATA or Parallel ATA= PATA – old type of interface	
SAS – Serial Attached SCSI (read SKAZI) – high performance computers – compatible with SATA (partially)	 

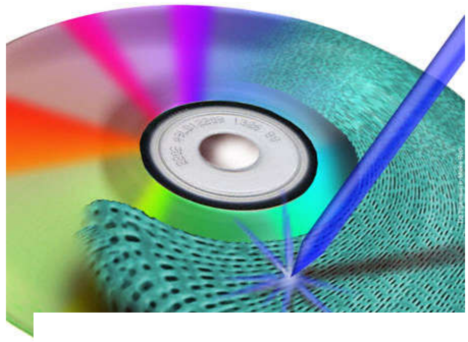
HDD Geometry



Hard Drive Structure:

A = track
B = sector
C = sector of a track
D = cluster

Optical Disc Storage



Optical Disc Drives use laser light to read and write information to optical discs. There are drives that only read (ROM) while on the other hand some drives are also there that performs both reading and writing (RW, RAM). These devices include CD, DVD and Blue-ray Disc.

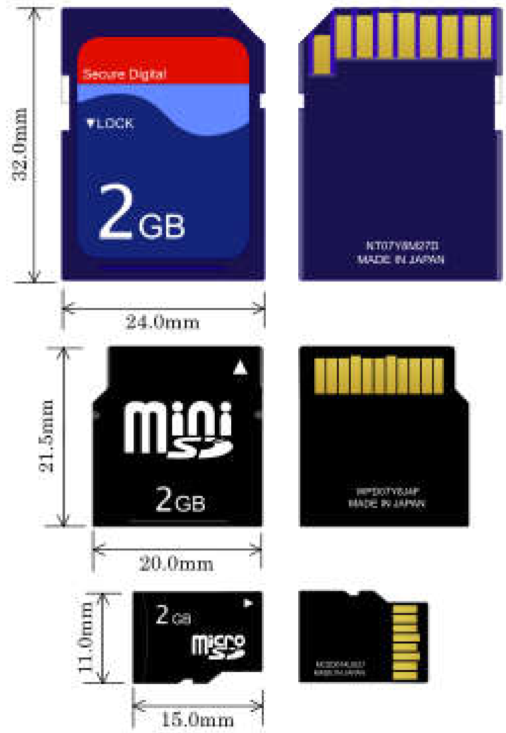
Flash Memory-based Devices

- **FLASH MEMORY** = a non-volatile computer storage that can be electrically erased and reprogrammed.
- Primarily used in memory cards and USB flash drives.
- A specific type of EEPROM - **E**lectrically **E**rasable **P**rogrammable **R**ead-**O**nly **M**emory – which can be erased in blocks, unlike the original EEPROM, which must be erased as a whole.

Flash Drive

- Flash memory + USB interface + controller
- It replaced floppy disks
- Small, large capacity, non-volatile, relatively immune to mechanical damage
- Sometimes contains preloaded software or an operating system (Linux, Google OS)
- Capacity (2015): 16 GB to 128 GB

Memory Cards



- Mainly used in portable devices as the replaceable storage (mobile phones, MP3 players, digital cameras, camcorders …)
- There are more than 20 different types (SD, MMC, CF)
- Nowadays (2015) the most common type is SD card (limit 2 GB) and its improved versions – SDHC (32 GB), SDXC (2 TB) – in 3 form factors (SD, SD mini, SD micro)

SSD Drive

- SSD = Solid-state Drive
- Several types – the main one based on NAND flash non-volatile memory, so it can be used as the hard disk replacement
- **Advantages** of SSD in comparison with the HDD
  - Significantly faster
  - No mechanical/moving parts – mechanically more immune than HDD
  - Silent
  - File fragmentation has lower performance impact than in HDD
- **Disadvantages**
  - Relatively expensive (Febuary 2015 - SSD: 100 € - 240 GB; HDD: 100 € - 2 TB; January 2017 - 100 € - 256 GB; HDD: 100 € - 2 TB)
  - Each modification of the memory blocks results in small damage of the cells, which hold data – SSDs have finite amount of data that can be written.

Figure 7:SSD Drive Uncovered (source: http://www.dslrfilmnoob.com/wp-content/uploads/Samsong-840-ssd.jpg)

Formats

- 2.5" case – like HDD
- M2 card
- PCIe card

Hybrid Drives

A combination of the SSD and HDD – either the drive moves the most frequently used to its SSD (not controlled by the user), or the SSD forms an independent drive – Western Digital Black2 120 GB SSD + 1000 GB HDD.