Computer Architecture (part 1)

Part 1

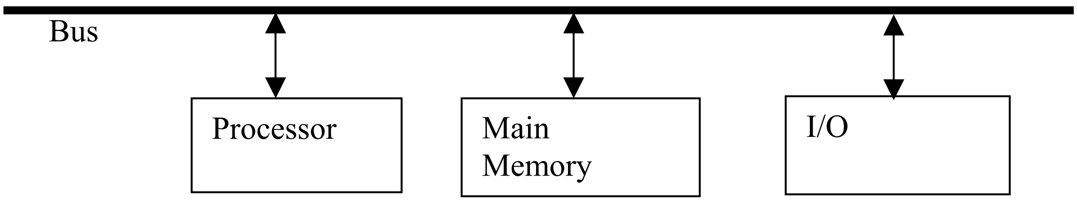
Internal Hardware and External Components of a Computer

Three-box model

A computer system can be considered as consisting of three boxes:

a) the processor, commonly known as the central processing unit (CPU e.g. Pentium
b) storage, known as main memory or simply memory
c) input and output (I/O) electronics for communication with other devices.

These three boxes need interconnecting. This is usually done by wiring known as a bus



The three-box model

This architecture was first proposed in 1945 by John von Neumann, which is why it is known as the von Neumann architecture. It is still used in the modern PC.

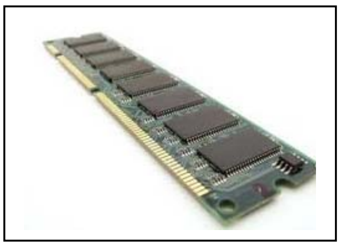
Processor

The processor is the brain in the system. It is responsible for executing programs and supervising the functioning of the other parts of the system. The processor is perhaps the most complex element in the system. However, it is simply a big finite state machine (FSM). Single-chip processors are known as microprocessors, e.g. Pentium.

A chip, or integrated circuit (IC), is encased in a plastic package. A typical microprocessor contains some 100,000,000 or more transistors. The number of transistors that can be integrated into the chip for minimum component costs doubles about every 18-24 months.

CPU is used to describe just the processor plus main memory. Since the invention of the microprocessor, CPU has come to mean the microprocessor.

Main memory



In principle the main memory, or immediate access store (IAS), also known simply as store, is the simplest part of the computer. In a von Neumann computer it is used as a store for program instructions and data. Main memory is available in one or more integrated circuits, or memory chips. Memory chips with the capacity to store 512 KB (512 x 1024 bytes) are often used in PC systems. Each byte goes into a separate area of the chip, called a memory location. Main memory is made up of RAM chips but can also include one or more ROM chips.

RAM
Random access memory (RAM) is used for memory that is readable and veritable. It can be implemented in many technologies (e.g. SRAM and SDRAM). RAM is used by the processor to store (write) data and load ,read) it back. The contents of RAM are volatile - they are forgotten when the power is turned off.

ROM
Read-only memory (ROM) doesn't forget when the power is turned Off - the contents are non-volatile. ROM provides random access like RAM but it cannot be written to once it is set up. It is used to hold fixed programs such as the bootstrap program in a PC.

EEPROM
Electrically erasable programmable read-only memory (EEPROM) does allow the contents to be altered. EEPROM chips require considerable time to alter a location - writing takes over 100 times longer than reading. Flash memory is faster at writing than other types of EEPROM because blocks of bytes are written in one go rather than single bytes. Flash memory EEPROM chips are widely used for non-volatile storage n consumer applications such as SIM cards in mobile phones, memory cards in digital cameras and memory in MP3 players, memory sticks and PDAs.

Bus
The three boxes are connected by a bus called the **system bus,** or **external bus**, which acts as a connecting tube between the three components. This connecting tube can transmit a single binary word between the processor and the memory or I/O component, and send an address to the memory or I/O component.

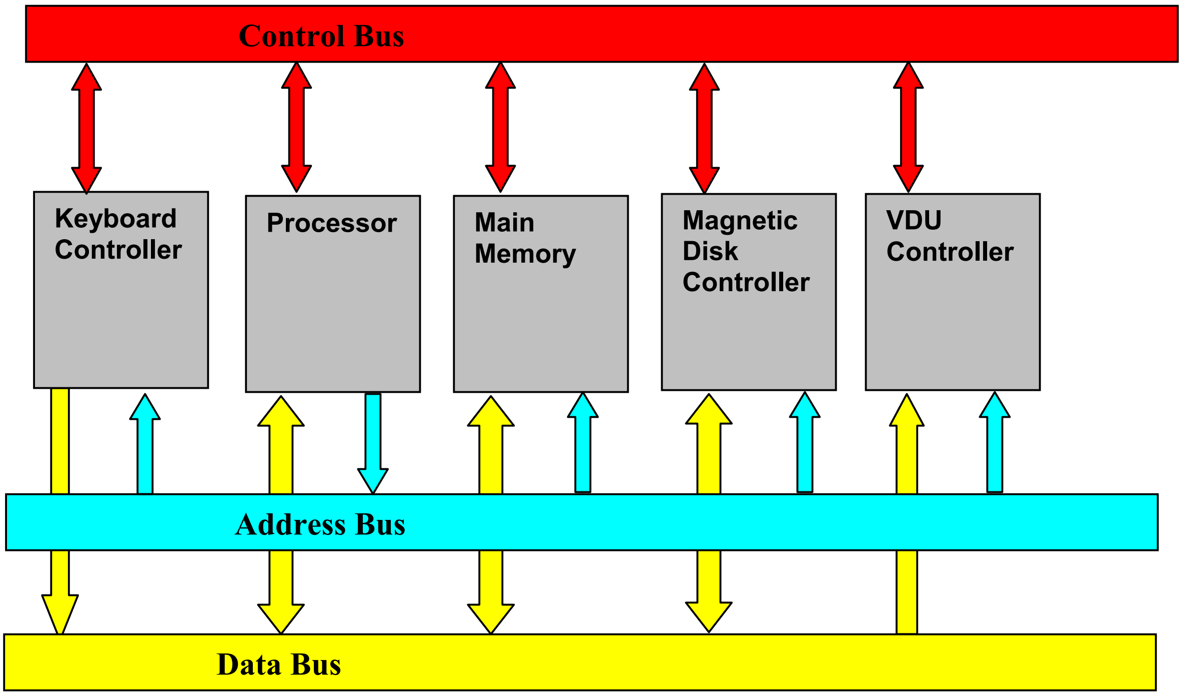
In reality, a computer bus is a set of parallel wires that connect independent components of a computer system in order to pass signals between them. Some signals represent data, some represent an address and others control information. For this reason, the system bus is split into three separate buses:

Data bus: a bidirectional bus, typically consisting of 32 wires, used to transport data between the three components of the three-box model.
Address bus: a unidirectional bus, typically consisting of 32 wires, used to address memory and I/O locations.
Control bus: a bidirectional bus, typically consisting of 8 wires, used to transport control signals between the three components of the three-box model.

I/O
To be useful, the CPU needs to communicate with devices known as peripherals. The peripherals are connected to the CPU through I/O controllers. The diagram below shows a more detailed view. The control bus carries control signals such as these:

Clock signal: for timing purposes
Reset Signal: used to initialise components
Memory read: used to assert that the memory location currently in use is being read from
Memory write: used to assert that the memory location currently in use is being written to

I/O: used to indicate that the processor wishes to use an I/O controller not main memory when I/O and main memory share memory addresses.



Expanded 3 box model

Peripherals

A computer device that is not part of the CPU (processor and main memory) is called a peripheral or peripheral device. The system bus is not connected directly to I/O devices; it is connected to an I/O controller then the I/O controller is connected to the I/O device.

I/O controllers and I/O devices

The processor communicates through mechanisms and devices that are grouped into I/O controllers and I/O devices. I/O devices are peripherals, such as keyboards, visual display units, magnetic disks, CD-R and DVD-R drives, network cards and speakers. I/O controllers include the keyboard controller and the magnetic disk controller. Peripheral devices cannot be connected directly to the processor. Each peripheral operates in a different way and it would not be sensible to design processors to directly control every possible peripheral. If processors controlled peripherals directly, the invention of a new type of peripheral would require the processor to be redesigned. Instead, the processor controls and communicates with a peripheral device through an I/O controller or device controller. Some I/O controllers can input and output bits, e.g. a floppy disk controller. Other controllers operate in one direction only, either as an input controller, e.g. a keyboard controller, or as an output controller, e.g. a VDU controller.

The controller is an electronic circuit consisting of three parts:

a) electronics that interface the controller to the system bus
b) a set of data, command and status registers

c) electronics appropriate for sending control signals to the device connected to the computer.

1/0 port

The processor exchanges data with a peripheral device through a part of an I/O controller called an I/O port. An I/O port is simply a set of data, command and status registers. Registers are storage locations that can be written to and read from.

Secondary storage

Secondary storage or backing store is permanent storage memory, not directly connected to the processor. Therefore secondary storage is a type of peripheral. A magnetic hard disk is a typical example of secondary storage. Secondary storage is connected to the processor and main memory through an I/O controller. Data is transferred from the secondary storage device to the I/O controller then across the system bus to the main memory or from main memory across the system bus to the I/O controller then to secondary storage. Table 1 shows some examples of secondary storage.

Storage	**Secondary storage**
Magnetic hard disk	No (if internal in a PC system)
CD-ROM CD-R, CD-RW	Yes
DVD-ROM, DVD-R, DVD-R	Yes
Memory stick (USB)	Yes

Exercise 1

a) Name the three boxes in the three-box model.
b) What is a computer bus?
c) Name the three sub-buses that make up the system bus.
d) What is an I/O controller?
e) Give a definition for
i) main memory
ii) secondary storage
iii) peripheral.

Functional characteristics of a processor

The processor controls the system bus, and the system bus communicates between the boxes in the three-box model. A processor performs operations on data by executing program instructions. A program changes the contents of main memory by sending single binary words back and forth through the system bus. Program instructions and data are fetched from main memory and the results are written to main memory.

Only one operation happens at a time, so the control signals and the data bus can be shared over the whole memory. The processor places a numeric code on the address bus to specify which location is currently being used, or addressed. A processor spends a lot of time calculating addresses, fetching addresses from memory along the data bus and sending addresses to memory along the address bus. This is just so that useful data can be fetched or stored in memory.

In the von Neumann computer, a large proportion of traffic on the system bus is not useful data but merely the addresses of this data.

Addressable memory
The processor needs to be able to distinguish the parts of main memory that store individual bytes, so each main memory location has a main memory address, like a house address. It is a unique numeric code. When a processor needs to select a main memory location, it puts the unique address corresponding to this location onto the address bus. The processor then asserts, over the control bus, whether it wishes to read from this location or write to this location.

Finally, the processor uses the data bus to transfer a byte or bytes between itself and the addressed memory location. All three components of the system bus - address bus, control bus and data bus - are needed when the processor uses a single memory location. A given memory location gains exclusive access of the system bus at a particular instant but it must share access with all the other memory locations at all other times, so the system bus is a shared bus.

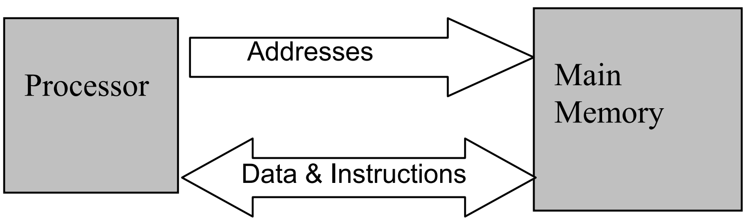
Stored program concept
The stored program concept was proposed by John von Neumann and Alan Turing in separate publications in 1945:

a) A program must be resident in main memory to be executed.
b) Machine code instructions are fetched, one after another, from main memory in sequence and are executed, one at a time, in the processor

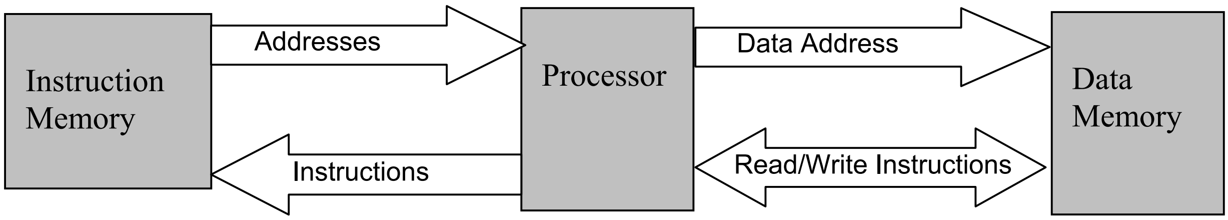
The processor is instructed to perform arithmetic and logical operation such as ADD, SUBTRACT, AND and OR. These computer instructions are represented by numbers called machine code instructions and are stored in the same way as data. Thus a bit pattern such as 01000110 might represent the number 46 (hexadecimal) or the letter F as data but it could also be used to tell a processor to perform an addition. Thus a single memory location holds values which can be interpreted as data or as instructions by the processor in a von Neumann computer.

Types of stored program computer

***Von Neumann** stored program computer*
The von Neumann stored program computer is a serial machine. Serial means that instructions and data are fetched one after another, one at a time. The von Neumann computer has a single memory shared between program instructions and data. Data and instructions travel along a shared data bus (see below).



Harvard stored program computer
The Harvard stored program computer has separate instruction and data memories



The term 'Harvard architecture' is normally used for stored program computers that use separate instruction and data buses. Instructions are fetched serially from instruction memory and executed in the processor. When an instruction needs data, it is fetched from data memory.

Von Neumann versus Harvard
In the von Neumann architecture, instruction fetches and data references share the same bus, so they compete for resources. In a Harvard architecture there is no competition, so instruction fetches and reading and writing of data can take place in parallel. This increases the overall processing speed.

Microcontroller
A microcontroller is a complete computer (processor, memory and I/O) on a single chip. The most common high-end (32-bit) microcontroller computers on a chip are ARM microcontrollers.

Microcontrollers are typically used as embedded computers where a cheap, compact yet programmable system is needed. An embedded computer is a special-purpose computer system built into another device, e.g. an MP3 player, a car or a muter, and designed to perform a few dedicated functions. In contrast, a general-purpose computer, such as a personal computer, can be programmed and reprogrammed to do many different tasks. Microcontrollers often use the Harvard architecture.

Exercise 2
a) Describe the role of each of the following buses when the processor uses a main memory location for a reading operation:
(i) address bus, (ii) control bus, (iii) data bus.
b) Explain the stored program concept.
c) Give a definition for
(i) microcontroller (ii) embedded computer.