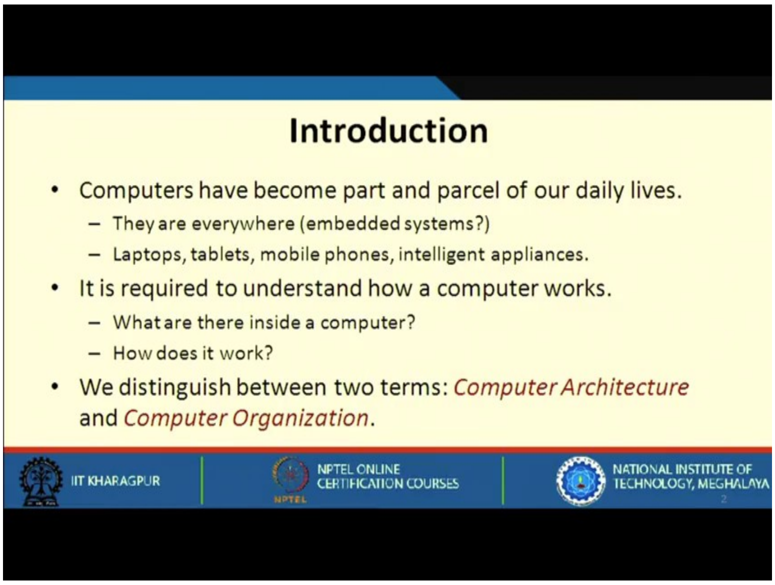
Computer Architecture and Organization
Prof. Kamalika Datta
Department of Computer Science and Engineering
National Institute of Technology, Meghalaya

Lecture - 01
Evaluation of Computer System

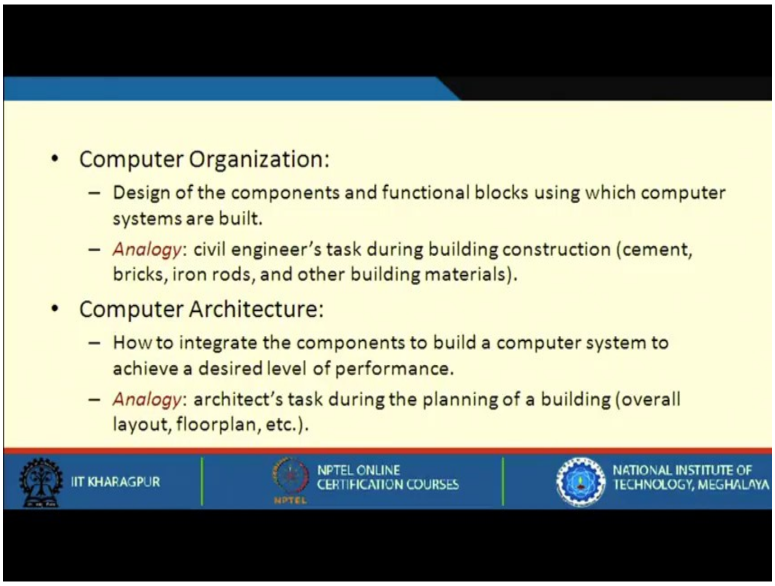
I welcome you all to the MOOC course on Computer Architecture and Organization. In this particular course, we expect to cover various aspects of computer design where you will be seeing how we can make a computer faster, how a computer actually works, how the information data are stored there and various other aspects. The lectures will span over 12 weeks where we will cover the instruction set architecture, processor design, arithmetic and logic unit design, memory unit design, input-output system design and then we will also cover parallel processing and pipeline.

(Refer Slide Time: 01:28)



To start with this course, I will come first to evolution of computer system. So, we all know that computer has become a part and parcel of our daily lives. We cannot disagree to this fact. We see everywhere computers that is some kind of processing unit. When you think about a laptop which we use in our daily use, tablets, mobile phones which are used by one and all today and intelligent appliances like off course your smart phone is one apart from that you have smart watch, and various other appliances. So, computer has become a part and parcel of our life. So, we need to understand how a computer actually works. So, what is there inside a computer? So, we in this particular course we will be seeing all these various aspects where the two terms computer architecture and computer organization will be taken care.

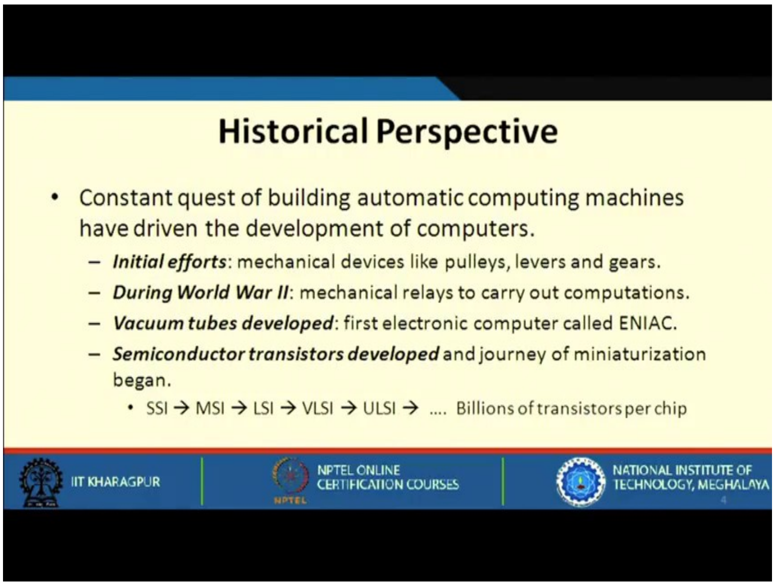
(Refer Slide Time: 02:40)



Now coming to what is computer architecture, and what is computer organization, the title of the course. So, here computer architecture consists of those attributes of the system that are visible to the programmer. By this what we mean is how the various components of a computer system are integrated to achieve the desired level of performance. In an analogy, I can say that you think of an architect who does who plans the entire design of your house, but it is ultimately the civil engineers who actually does the exact building like what kind of construction will be taken care of, how the construction will be taken care of, how much percentage of cement, bricks will be there , will be taken care by a civil engineer. So, in that respect the design of components and functional blocks using which computer systems are built comes to the organizational part.

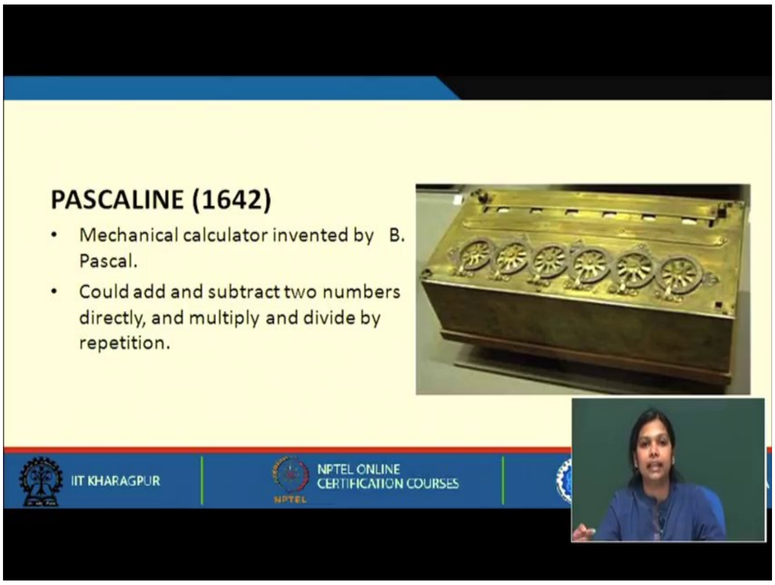
So, I will take a very small example like you have in your computer some functional blocks like your processor unit; inside processor unit we will be seeing that we have many other components like registers, ALU and other units. I will just take a small example let say I will have an adder, but what kind of adder I will be having that is to the discretion of the computer organization, whether we will have a carry save adder or a carry look ahead adder or anything else. So, these are the two different aspects of computer organization and computer architecture that we will be seeing in this particular course.

(Refer Slide Time: 04:49)



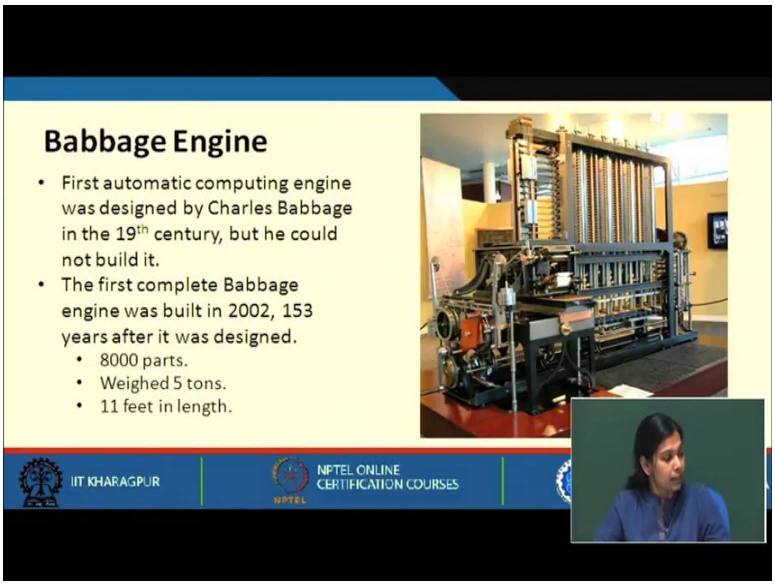
Coming to the historical perspective, how computers have evolved over the years. So, whenever there is a need for doing certain things then only something comes up like a constant quest of building automatic computing machines have driven the development of computers. So, in the initial efforts, some mechanical devices like pulleys, levers, gears were built. During World War 2, mechanical relays were used to perform some kind of computation like using small relays people design circuits to carry out the operations. Then comes vacuum tubes using which the first electronic computer called ENIAC was developed. And from then semiconductor transistors were developed when semiconductor transistors came into picture then the journey of miniaturization started. First with small scale integration then people moved with medium scale integration then large scale integration then to very large scale integration and now the era of ultra large scale integration where we stand today.

(Refer Slide Time: 06:22)



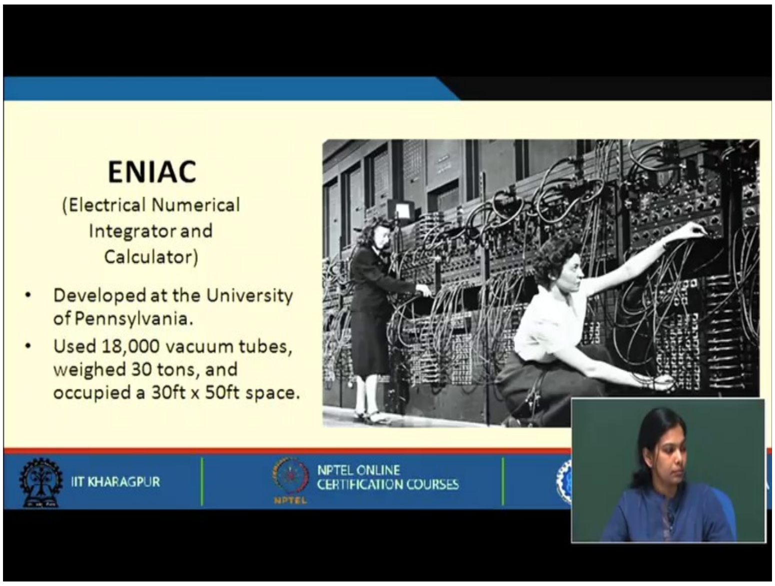
So, this is the first mechanical calculator that was invented by B Pascal. So, this particular calculator could add only two numbers, it can only add two numbers or it can only subtract two numbers. And if you wanted to do multiplication and division, it could have been done by repeated addition or repeated subtraction.

(Refer Slide Time: 06:58)



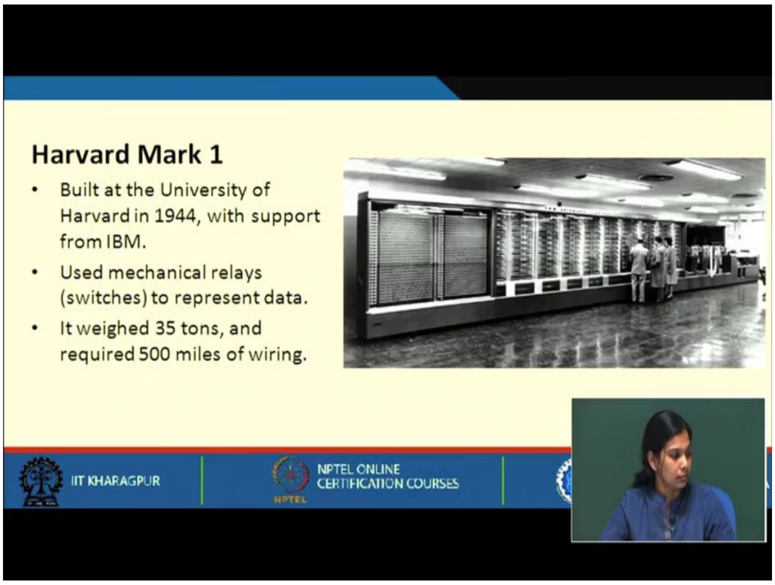
Then the Babbage engine came this was the first automatic computing engine designed by father of computer Charles Babbage in the 19th century, but he could not build that only he designed that. Later in 2002 that is 153 years after its design, it was built and it consists of 8,000 parts and it weighted 5 tons and it was 11 feet in length. So, you can imagine how large it is.

(Refer Slide Time: 07:43)



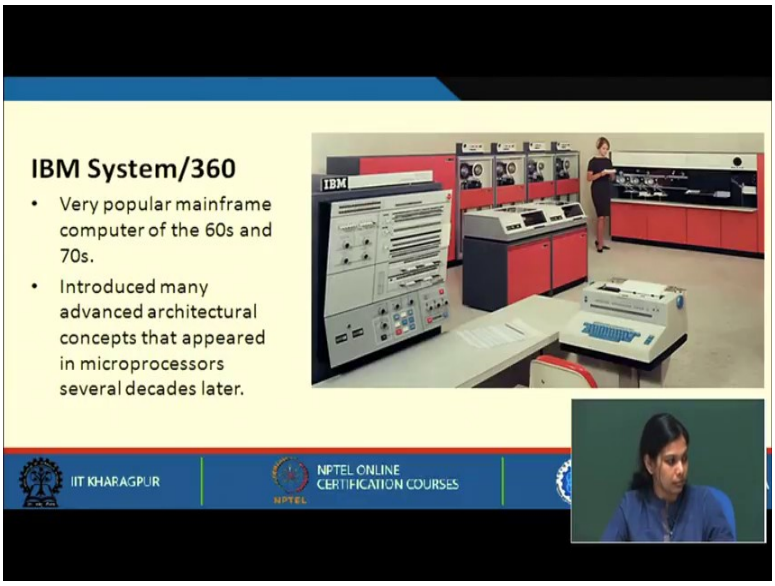
The first electronic computer was built which is called ENIAC - Electrical Numerical Integrated and Calculator. It was developed by University of Pennsylvania, and it uses 18,000 vacuum tubes and weighted 30 tons, and it also occupied a very large space which is 30 feet cross 50 feet. So, what is a vacuum tube? Vacuum tube is a device that controls electric current between electrodes in an evacuated container in a tube. So, using those vacuum tubes the first computer ENIAC was built. It also dissipated a huge amount of heat.

(Refer Slide Time: 08:44)



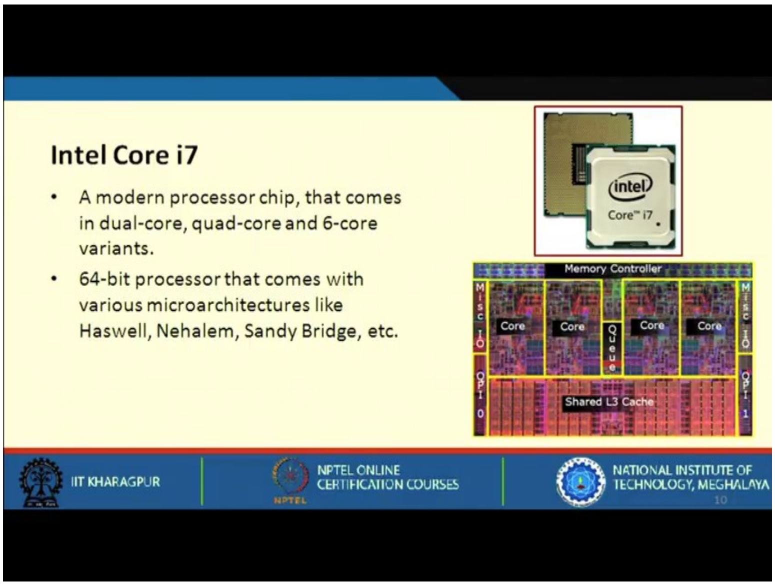
Next was Harvard mark 1. This was built at the University of Harvard in 1944, with support from IBM and it uses mechanical relays and off course, some electric signals were also used to work with the relays to represent the data. And it also weighted 35 tons, and required 500 miles of wiring. So, these are the computers which were built in the early stages.

(Refer Slide Time: 09:20)



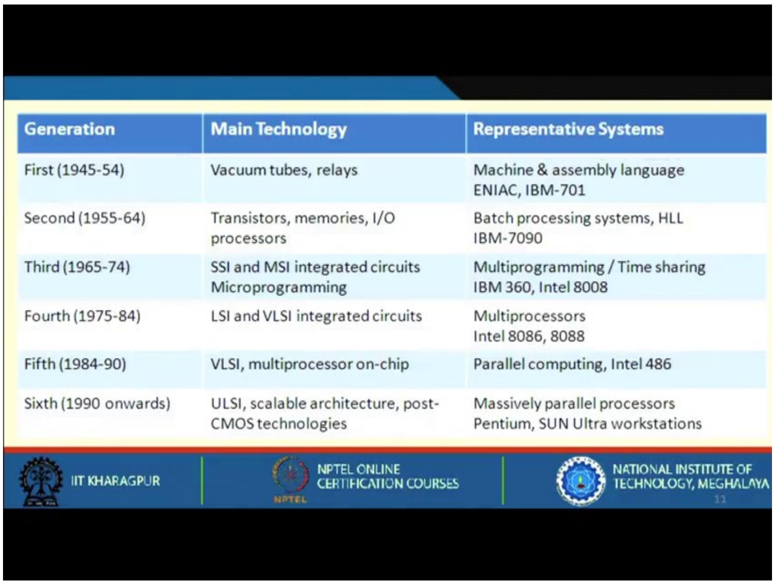
Then comes in 60s and 70s where the popular mainframe computer came into picture. So, this popular computer IBM system 360 was introduced; it introduces many advanced architectural concepts that we use today, but you can see from the picture that how big it was. But as I said some of the architectural aspects or concepts that were used in that computer appeared in today's microprocessor several decades later.

(Refer Slide Time: 10:03)



Now, we stand, where the modern processor chips comes in dual-core, quad-core and also in 6-core variants. So, you can see that how many core today's computer has, where each core is a 64-bit processor that comes with various micro architectures. These are the various micro architectures within an I7, where they are called Haskell, Haswell, Nehalem, Sandy Bridge etc.

(Refer Slide Time: 10:52)

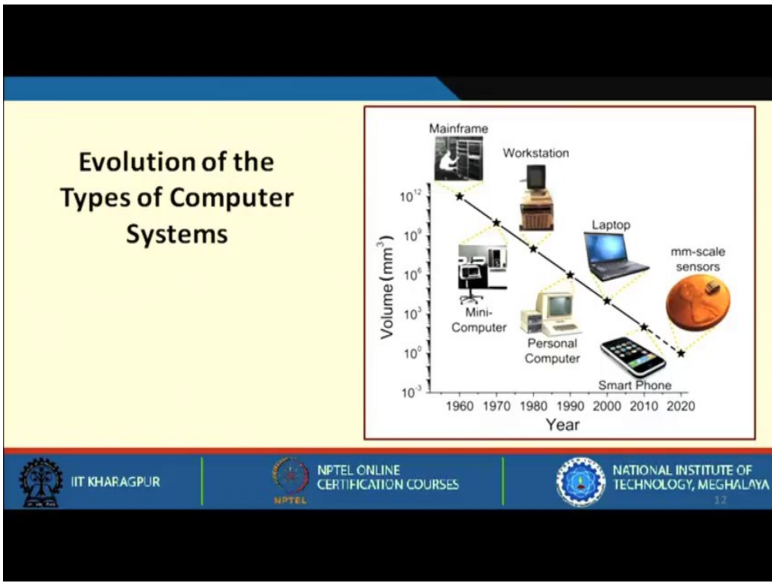


Generation	Main Technology	Representative Systems
First (1945-54)	Vacuum tubes, relays	Machine & assembly language ENIAC, IBM-701
Second (1955-64)	Transistors, memories, I/O processors	Batch processing systems, HLL IBM-7090
Third (1965-74)	SSI and MSI integrated circuits Microprogramming	Multiprogramming / Time sharing IBM 360, Intel 8008
Fourth (1975-84)	LSI and VLSI integrated circuits	Multiprocessors Intel 8086, 8088
Fifth (1984-90)	VLSI, multiprocessor on-chip	Parallel computing, Intel 486
Sixth (1990 onwards)	ULSI, scalable architecture, post-CMOS technologies	Massively parallel processors Pentium, SUN Ultra workstations

Now, coming to the generation of computers. So, broadly we can classify the growth of computer, how computer has evolved into some generations. So, these actually represent some features of operating system and representative machines. So, between the years 1945 to 54, vacuum tubes and relays were used; and the representative system uses machine language. So, if you want to enter some data into the computer for processing, you have to enter in machine language or either in assembly language, no high level language were used to enter the data. Then comes the second generation where transistors, memories, IO processors are the main technologies. Here batch processing is performed, and you could enter the data in high level language.

Then comes the third generation, where we were into semiconductor industry where small-scale integration and medium-scale integration integrated circuits were used. And in that IBM 360, Intel 8008, where the first one to make the step into this computer market. Then in the fourth and fifth generation, this large scale integration and very large scale integration came into picture and we could see multi processors. Now, what has happened is like the space has become larger because the components that we use to build those basic blocks has become smaller. So, in a same space where earlier we could keep small number of components, now we could keep more number of components into the same space. And hence multiprocessors comes into picture. And now we are in the era of ultra large scale integration with many scalable architectures and post-CMOS technologies. Now, we also have massively parallel processors like Pentium, SUN Ultra workstation etc.

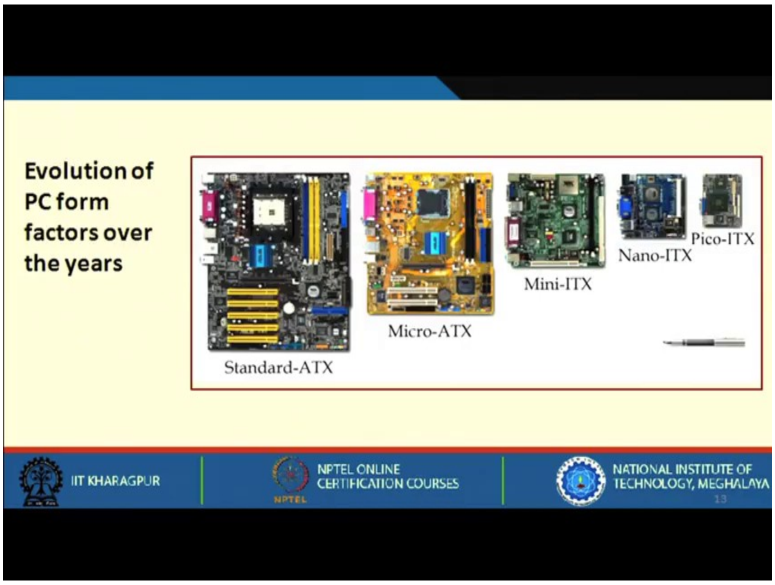
(Refer Slide Time: 13:34)



Now, you see that evolution of the types of computer. So, initially mainframes came in the year 60s and late 70 in between 60s and 70s. Then came mini computers then came workstations, and finally we see personal computers. And then further we see laptops and smart phones. So, every smart phone today is having some processing capability that is why we call it smart phone with some smart features into it. And we also have now millimeter scale sensors and these are not only sensors that it will sense something rather these sensors are having some intelligent capability like some processing capability which can process those data that are collected from the sensors and also it has the communication feature, so that it can communicate to others if required. So, this millimeter scale sensors have both the capabilities of processing as well as communication. So, we can see how the evolution of types of computer systems have come up.

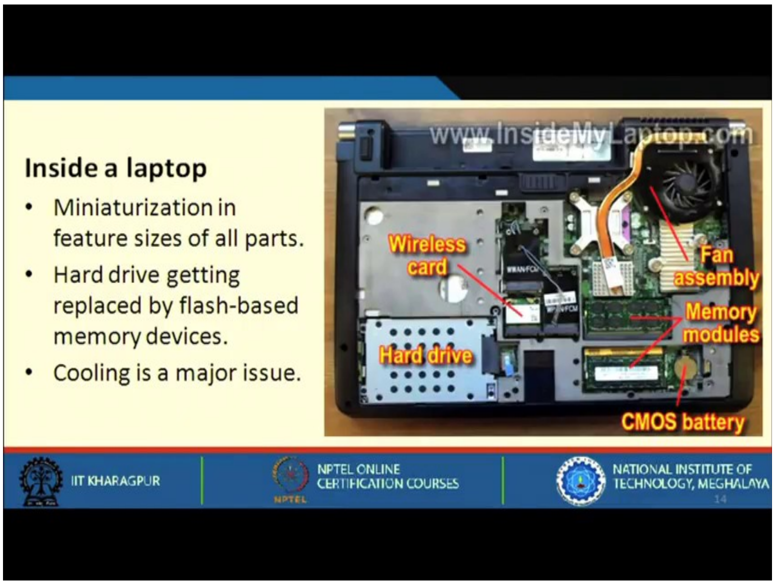
And now where we stand in the future, we will see large scale IOT based systems where these sensors itself are very small, where they will have not only the sensing capability, but also the processing capability.

(Refer Slide Time: 15:25)



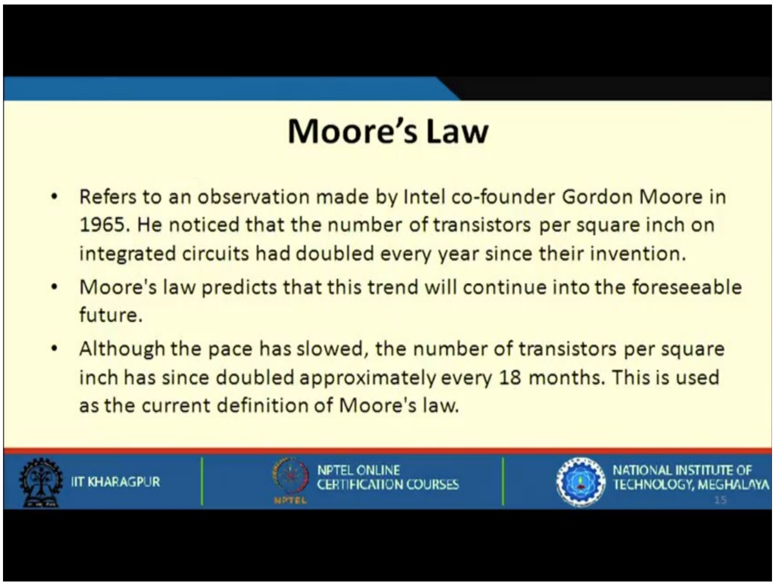
Now, you see this evolution of PC form factors over the years. This is the first standard ATX where you can see the motherboard slot. You can see this is a motherboard, you can see the processor, you can see the slots for memories for all other components. Now, how that miniaturization has taken place. First it was standard-ATX, then micro-ATX then goes mini-ITX, now nano-ITX now we are in pico-ITX. So, we are in the era of miniaturization. And now with miniaturization of course, we are able to do some good thing for performance, but at the same time power issue is one of the important factors that we have to also handle which we cannot deny.

(Refer Slide Time: 16:37)



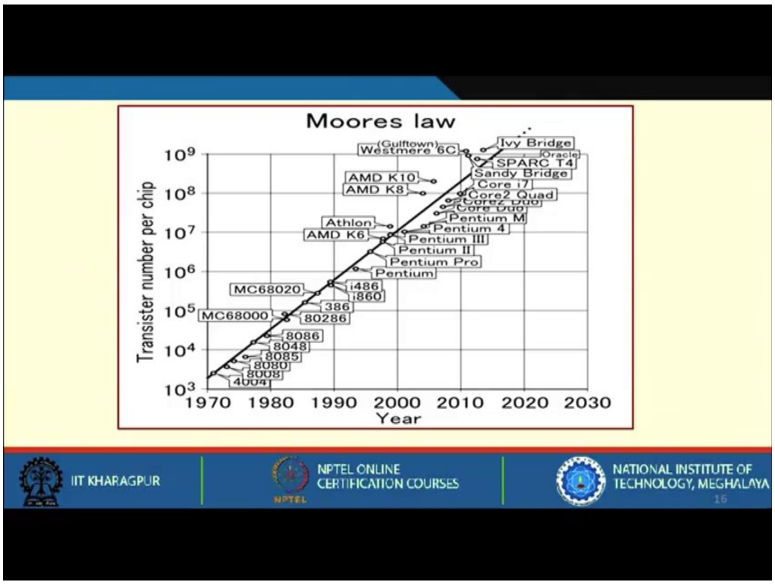
Now, inside a laptop if you see what all components we have. So, we have shrunked each of the components. So, hard drive is now getting replaced by flash-based memory devices and because of miniaturization cooling is a major issue that has come up.

(Refer Slide Time: 17:23)



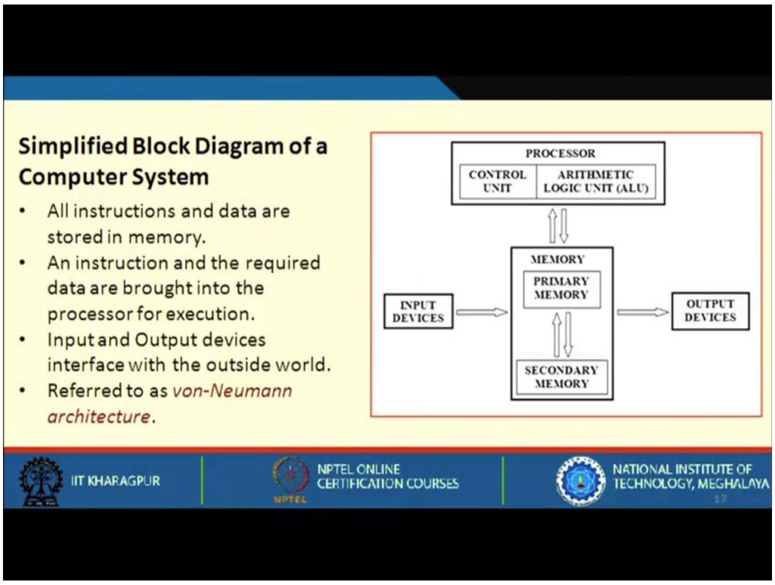
Now, this is the famous Moore's law that refers to an observation made by Intel co-founder Gordon Moore in 1965. What he noticed is that the number of transistors per square inch on integrated circuits had doubled every year since invention. And this law predicts that this trend will continue into the foreseeable picture, but now Moore's law stand here where it says that the number of transistor that can be placed in an integrated circuit gets doubled in every 18 months. And this has held over a period.

(Refer Slide Time: 18:17)



Now, you can see this diagram which how Moore's law works, this is the number of transistor per chip and this is over the years how it has grown. So, this straight line actually indicates that Moore's law holds. And starting from 4004 we are now in ivy bridge, sandy bridge, and various architectural advancement has taken place and number of transistor that can be put in a chip has doubled in every 18 months. And as long as it is a straight line, we can say that Moore's law holds.

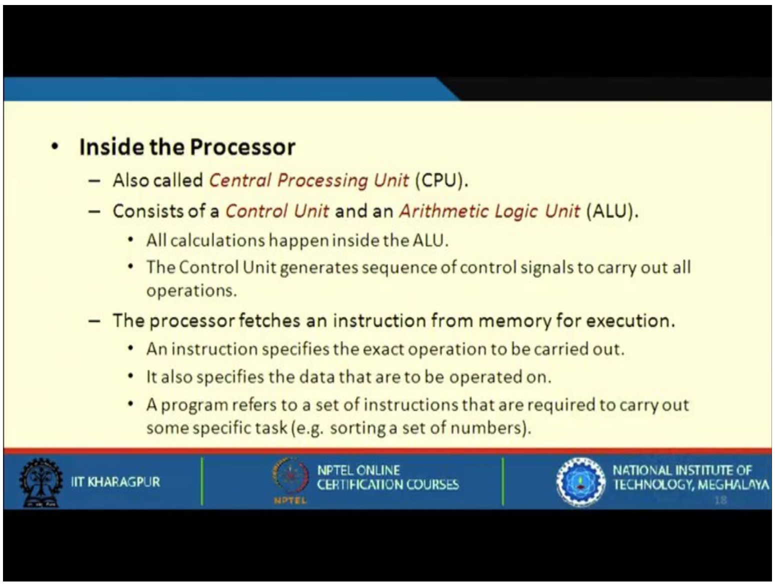
(Refer Slide Time: 19:16)



Now, coming to the simplified block diagram of a computer system. So, in this diagram what you can see is that we have a processor. The processor is having control unit and arithmetic logic unit. We have a memory; memory is divided into primary memory and secondary memory. And you have input devices and output devices. In this architecture course we will be taking each and every aspect of this design like we will be seeing the processor unit design; in the processor unit design there are two parts we will be seeing control unit design and arithmetic logic unit design. We will also look into the memory, memory is broadly divided into primary memory and secondary memory. What we have in primary memory we will be seeing, what we have in secondary memory we will be seeing. And of course, how the inputs are given to the computer and how we get the outputs from the computer that also we will be look into.

So, all instructions and data are stored in memory. Whatever instruction or data that we want to execute that is stored in memory. And every time an instruction and data is required it is brought from the memory to the processor for execution, and input output devices are connected to it. So, if input is required input is taken from an input device processing takes place in the processor and the output is provided in the output device. This typical architecture where we store both program and data in the memory alongside we call it von-Neumann architecture we will be seeing this in more detail little later.

(Refer Slide Time: 21:31)

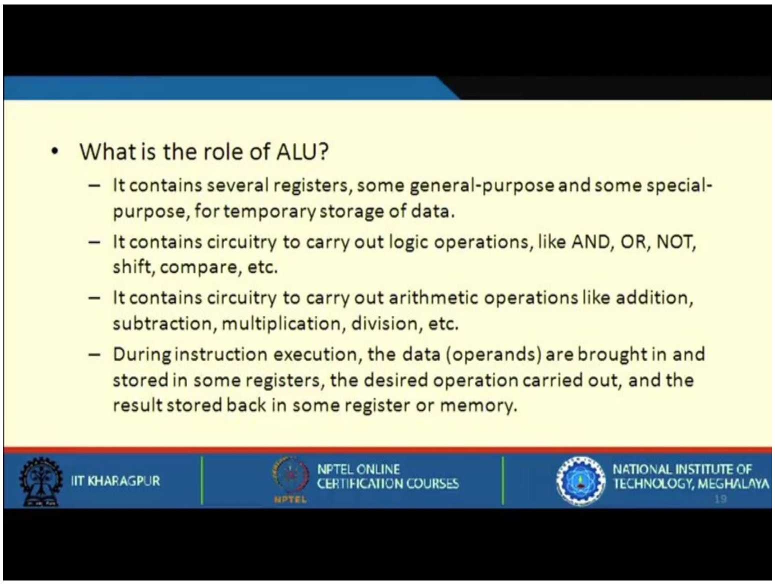


Now, let us see what is there inside a processor. A processor is also called a Central Processing Unit; it consists of a Control Unit and it consists of an Arithmetic Logic Unit. All the calculations happens inside the CPU. So, any arithmetic computation, like we need to add two numbers, we need to divide two numbers or any arithmetic operation that are performed inside the ALU. The control unit basically generates the sequence of control signals to carry out all operations. So, all the operations that are performed, it is the control unit that generates those sequence of control signals i.e, when I say that we will execute an instruction. So, you have to instruct the computer that, execute this particular instruction. So, giving the computer some kind of instruction, some kind of control signals that, yes now you have to execute this now you have to execute that and finally, you have to store the result or you have to display the result. So, all these steps that we are instructing to a computer is generated by the control unit.

In a processor an instruction actually specify the exact operation that is to be performed. So, the instruction will tell you ADD A, B. So, A, B are some operands that perform the required operation, like ADD. So, you have to instruct that this is an operation that has to be performed on some operands. It also specifies the data that are to be operated on.

And now let us see what is a program. I have talked about a single instruction. Now, a program is a set of instruction, like I need to add ten numbers. So, adding ten numbers I have to write a set of some instructions; those set of some instructions are written to perform that particular task. A program is a set of instructions that constitute a program.

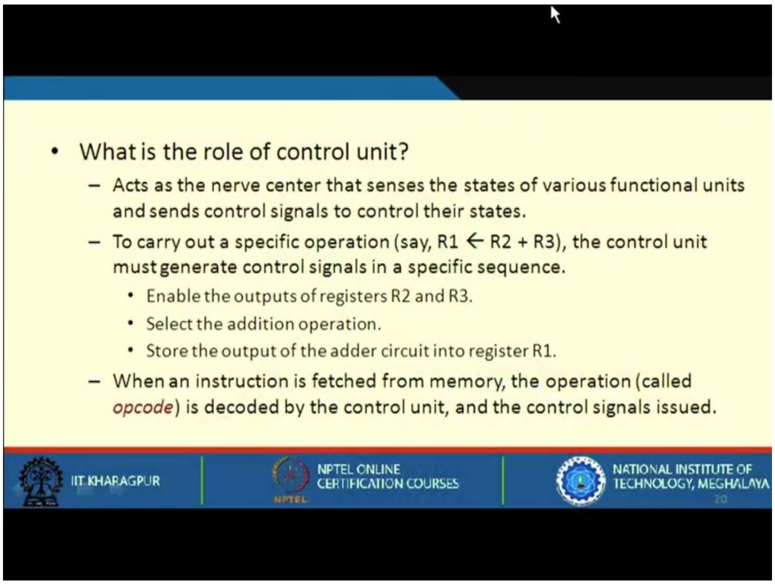
(Refer Slide Time: 24:35)



Now, what is the role of ALU? ALU or the processor unit consists of several registers; some registers are called general purpose registers, and some are called special purpose registers, and some of them are temporary storage. So, what are these registers? Registers are some storage unit, and these registers are used to store data, then again we compute some operation and again store back that results into it. We store data for computation; and after the computation is performed, we also store back the data.

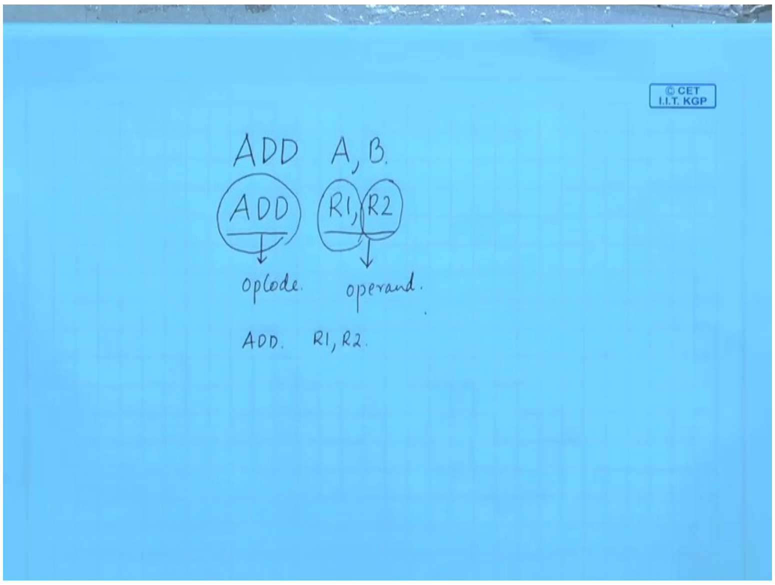
It contains circuitry to carry out the logic operations. So, basically when we say we are adding two numbers, we are subtracting two numbers, we are doing some kind of operation, some kind of logic operation is performed to carry out that particular operation. So, it also contains circuitry to carry out arithmetic operations like addition, subtraction, multiplication, division. We will be seeing in more detail every aspect of ALU in a separate lecture unit. Now, during execution, the data or the operands are brought in and stored in some register, the desired operation is carried out and the result is stored back in some register or memory. So, what does it means that during an instruction execution the instructions and data are stored in some memory locations. So, we have to bring the data from those memory locations into some of the registers, perform the operation and then we store back the data into either register or into those memory locations.

(Refer Slide Time: 26:41)



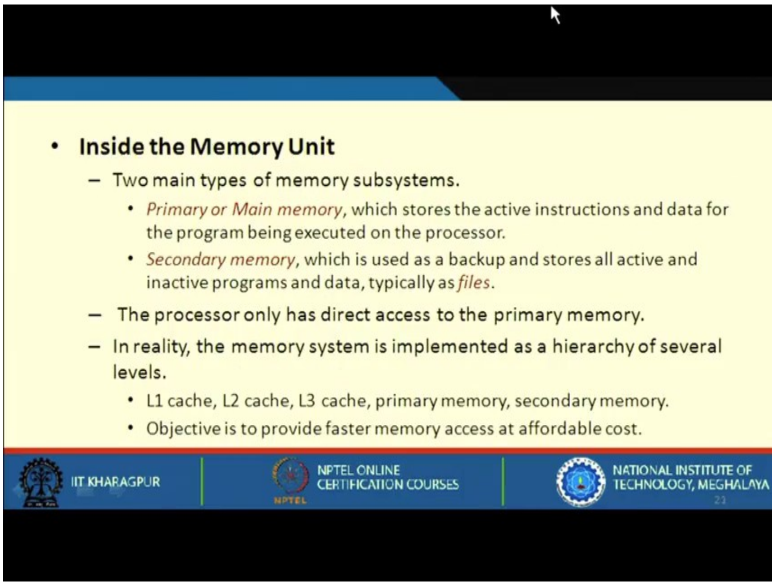
Finally what is the role of control unit? As I said earlier it generates the signals that is necessary to perform the task. So, it acts as a nerve center that senses the states of various functional units and sent control signals to control the states. Suppose you have to carry out an operation where you have to add R2 and R3 and store back in R1. So, what you need to do is that you need to enable the output of R2 and R3 such that the outputs of R2 and R3 are available in a place where you can do the operation. After the operation is performed it is stored in the register R3. So, you have to store the output in a circuit into the of the adder circuit into register R1.

(Refer Slide Time: 28:04)



Let us say this is an instruction. ADD A,B: this is an instruction. ADD R1,R2 is also an instruction. So, this instruction consists of two parts: the first part we call it Opcode and next part is the operand. Opcode specifies what operation we will be doing and operand is on which we will be doing the operation. So, in this particular case the operation that we will be doing is ADD and on which we will be operating are R1 and R2, these are the two registers where we will be doing the operation.

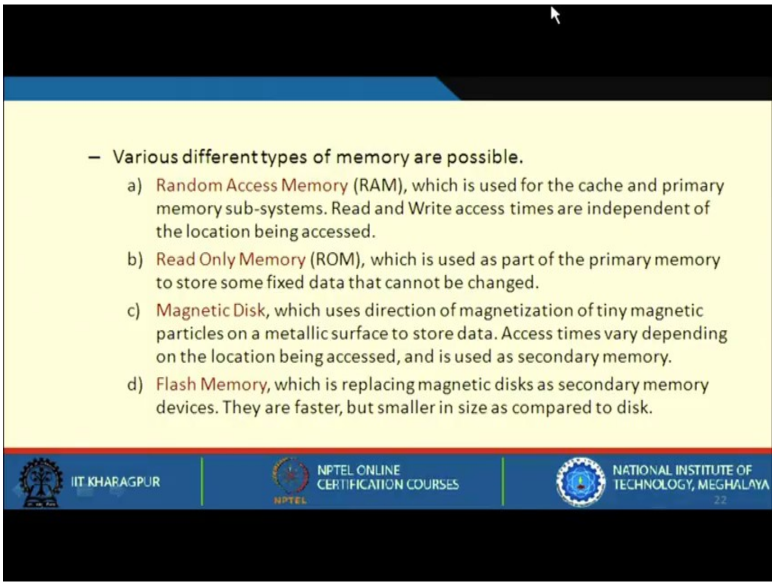
(Refer Slide Time: 29:10)



Now, consider the memory unit. There are two types of memory subsystems, two main types. So, primary or main memory stores the active instructions and data for program being executed on the processor. And the secondary memory is used as backup and stores all active and inactive programs and data typically the files. Now the processor can only have a direct access to primary memory. As I said the programs and data is stored in your primary memory and whenever it is required the processor ask it from your primary memory and not from your secondary memory.

And in reality, the memory system is implemented as a hierarchy of several levels. So, we have L1 cache, we have L2 cache, we have L3 cache, primary memory and secondary memory. What is the objective of all these things to make the processing faster? So, we will be seeing all these aspects in course of time, but for now you must know that in the memory unit we store instructions and data. And for processing of those instructions on data, you have to bring those instructions and data to the processors to execute it.

(Refer Slide Time: 30:59)

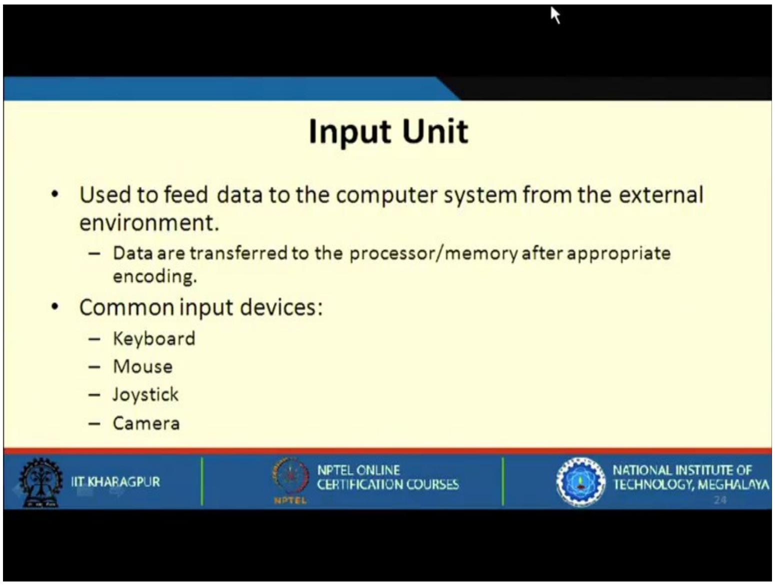


Now, we have various different types of memory, Random Access Memory, Read Only Memory, we have Magnetic Disk, we have Flash Memories. Random Access Memory is used for cache and primary memory and Read and Write access times are independent of the location being accessed. This means, you either access location 1 or you access the last location or the middle location, the access time will be same. Read Only Memories are used as a part of primary memory to store some fixed data that is not required to be changed. Magnetic Disk uses direction of magnetization of tiny magnetic particles on a metallic surface to store the data and the access time vary depending on the location being accessed, and these are used in secondary memory. Now, Flash Memories are coming into market, which is replacing this magnetic disk as secondary memory. They are much faster and smaller in size, and they do not have any movable part.

(Refer Slide Time: 32:28)

Now, these are the pictures showing these. The first picture shows the RAM primary memory. Next one is the ROM. This is a hard disk. If you open a hard disk it looks like this, this one is the SSD and so on.

(Refer Slide Time: 32:53)

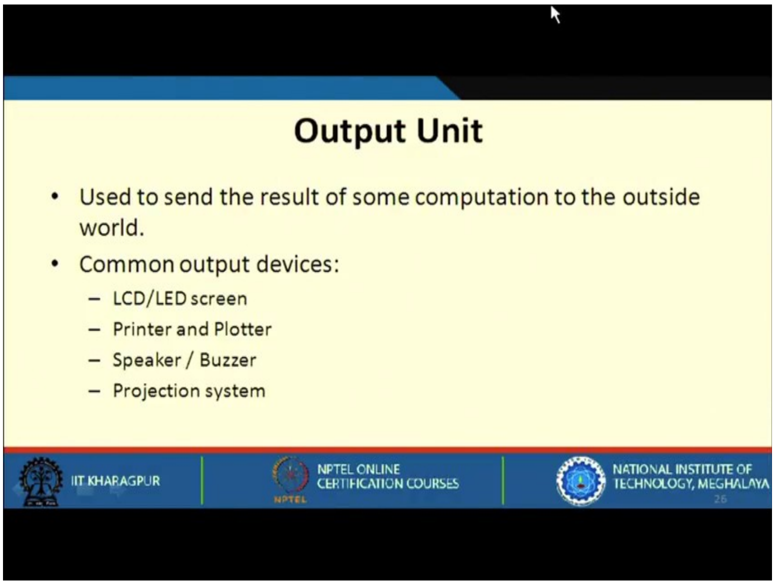


Coming to the input unit, it is used to feed data to the computer system. The commonly used devices are keyboards, mouse, joystick and camera.

(Refer Slide Time: 33:03)

These are the relevant pictures that are shown.

(Refer Slide Time: 33:15)



And the output unit is used to send the result of some computation to outside world like printer is used to print the data, LCD screen or LED screen is used to see the output on the screen, you have speakers, you have projection system that are also used as an output unit.

(Refer Slide Time: 33:42)

So, these are some of the relevant pictures showing the output units. So, we have come to the end of lecture 1 where we have seen how computer systems have evolved over the years and what are the main functional components of a computer system, and how these components are required and how we can execute a particular instruction.

Thank you.