Computer Organization

Agenda

1 Basic Structure of Computers

2 Instruction Formats

3 Central Processing Unit

4 Types of Memories

5 Input /Output organization

6 Types of Computers and IBM’s AS 400

7 BUS Architectures

Basic Structure of Computers

Module 1

Objectives

At the end of this module, you will be able to:

- Identify user's, software developers, computer designers view
- List different components of computer
- Discuss layered model of computer
- Describe CPU configuration registers, buses
- Review electronic switch, register, clocked circuit

Duration: 2 hrs

Computer Architecture

User view:

A computer is a machine to help the user in a variety of tasks.

- Computational
- Non- computational e.g. email
- Edutainment – Games, educational software

Software Developer's view:

A platform on which a software package can be developed and tested.

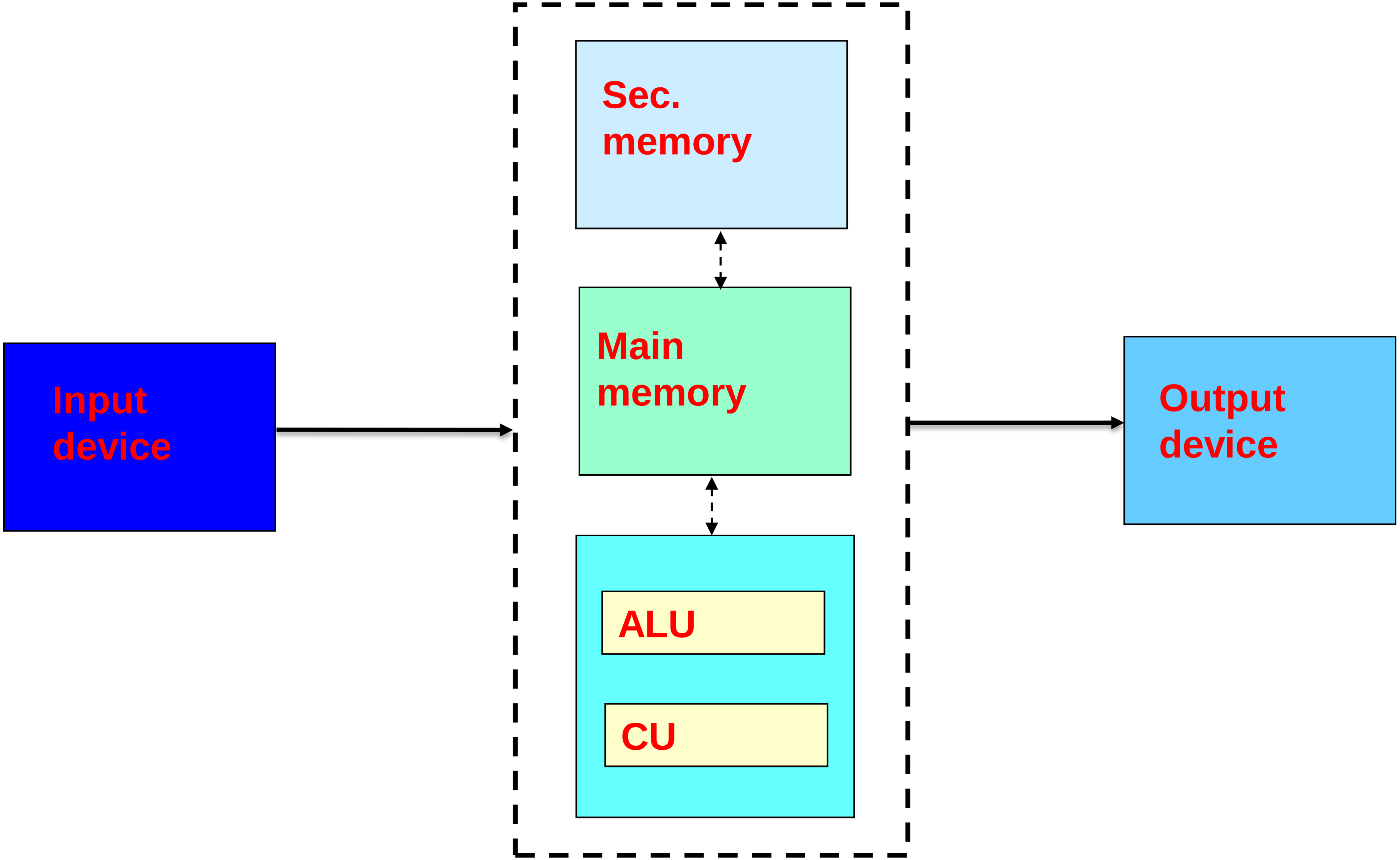
- General purpose, e.g.. text editor, compiler, etc.
- Special purpose, i.e.. targeted to specific applications

Computer Designer's view:

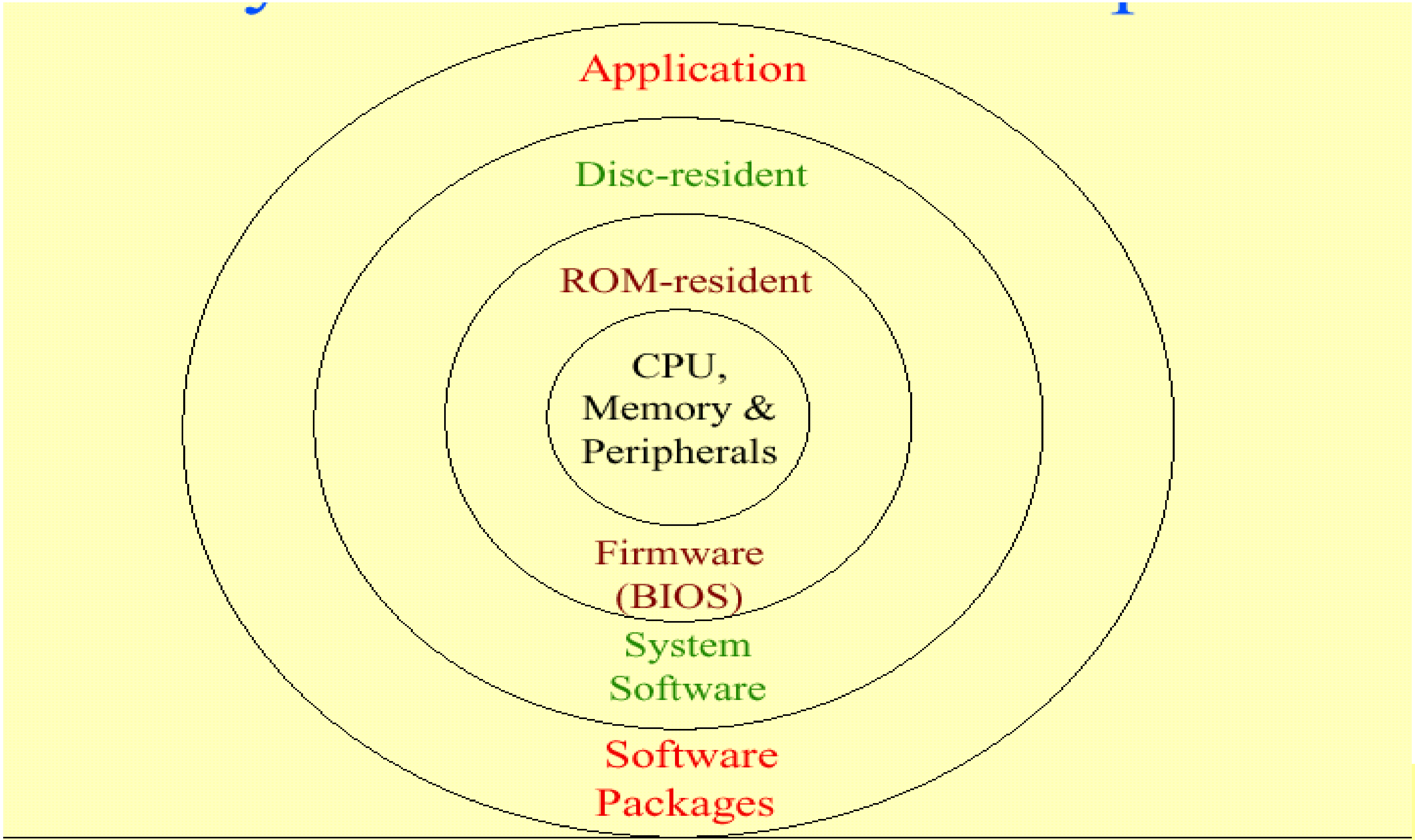
A programmable digital system, consisting of:

- System hardware able to support the software to be run.
- Software packages providing the necessary support in terms of utilities and programming languages.

A typical computer



A layered model of a computer



Application

Disc-resident

ROM-resident

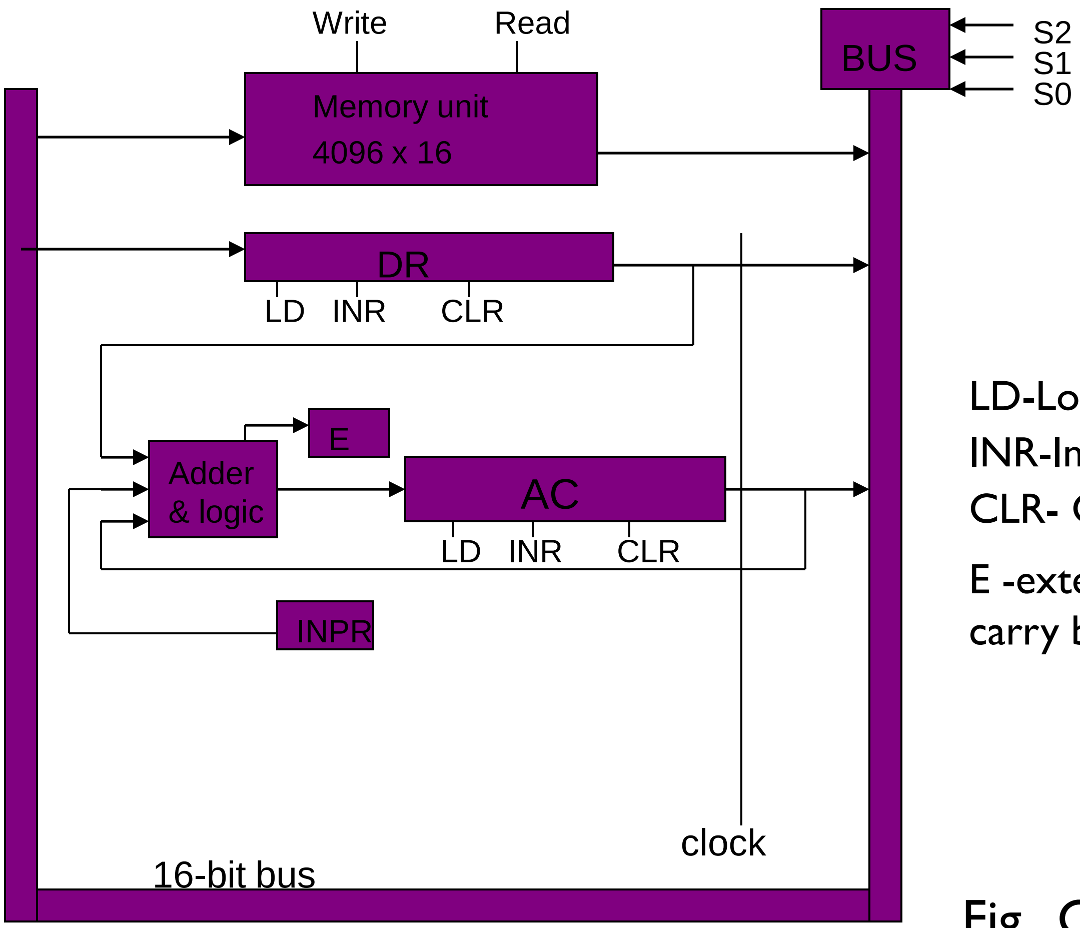
CPU,
Memory &
Peripherals

Firmware
(BIOS)

System
Software

Software
Packages

Configuration of Registers and Buses



LD-Load

INR-Increment

CLR- Clear

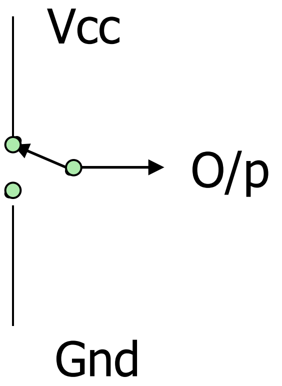
E -extended AC bit for carry bit

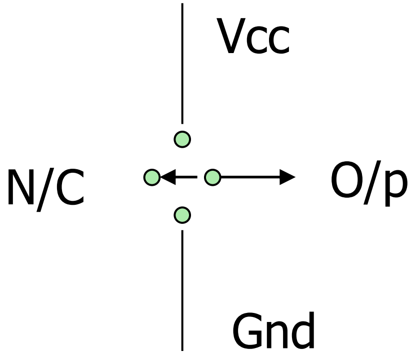
Fig. Computer Registers

Configuration of Registers and Buses

CPU consists of

- Combinational logic
  - Address, MUX, D-MUX, Coders, Decoders etc.
- Sequential logic
  - Counters, Registers, Shifters, etc
- All the logic devices' out-puts have tri-state logic





Summary

In this module, we discussed:

- User, Software developer’s and Hardware designer’s views
- Typical computer parts – CPU(ALU,CU), IO, Memories (main and secondary)
- Layered architecture: innermost- hardware, outermost-applications
- CPU configuration registers, buses
- CPU logic: Combinational, sequential, tri-state logic

Instruction Formats

Module 2

Objectives

At the end of this module , you will be able to:

- Describe Instruction formats
- Explain control unit of a computer

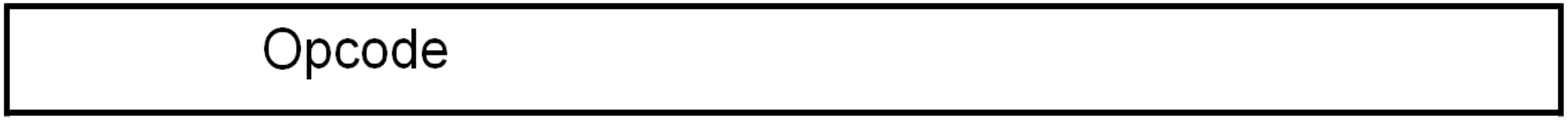
Duration: 1hr

Instruction Format

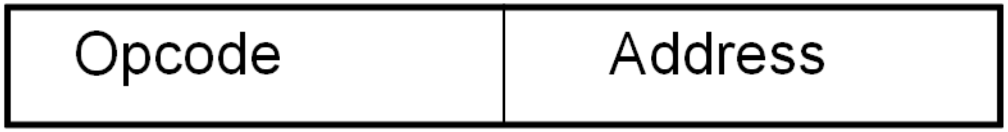
- Program consists of a sequence of instructions.
- Instructions are binary codes that specify some action.
- Many instructions contain or specify data/address used by them.
- Instruction specified by field/s :
  - Opcode (operation code)- specifies the operation
  - Operand- specifies the data or address of the data

Instruction Format (Contd.).

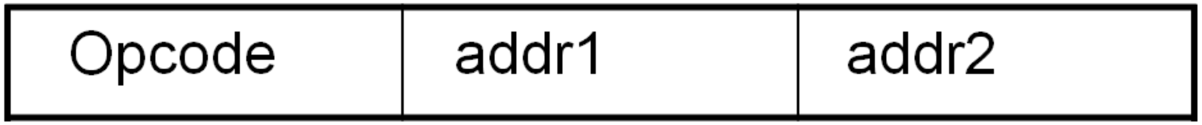
(a) Zero-operand instruction



(a) One-operand instruction



(b) Two-operand instruction



Instructions in a Simple Computer

Consider a basic computer having three instruction formats, each of 16-bits. (Opcode 3 bits)

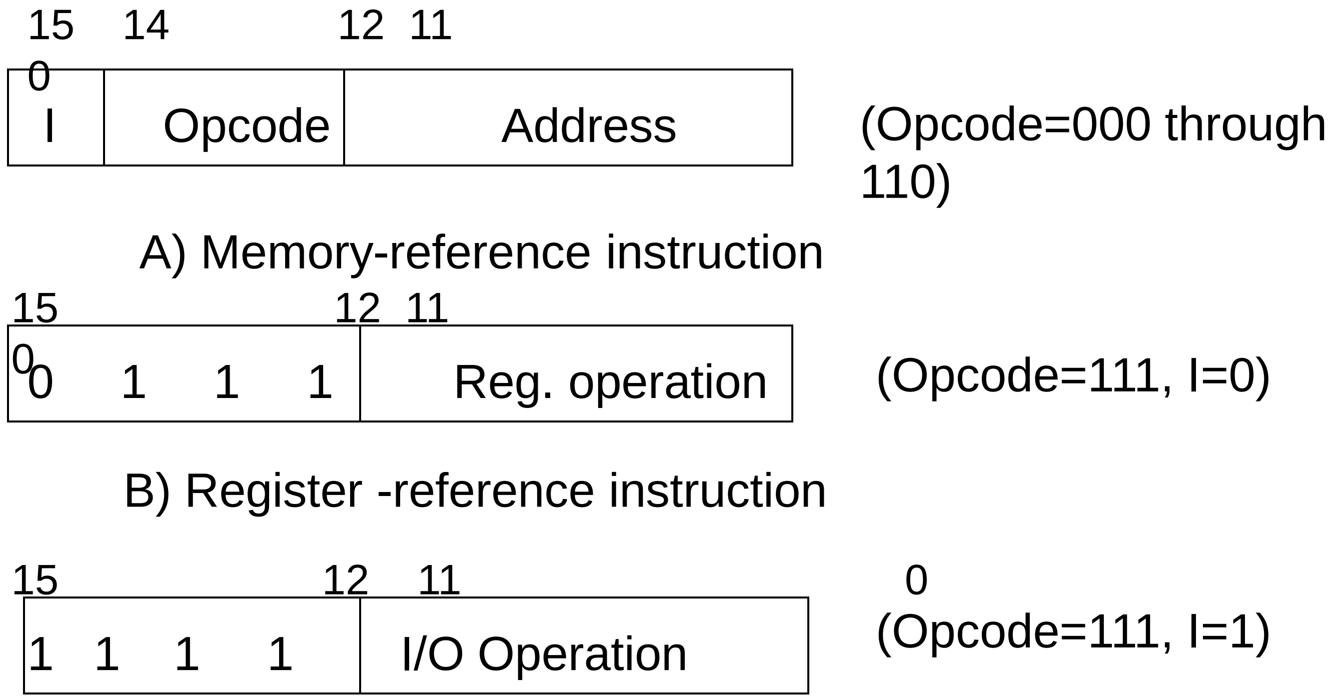
1. *Memory-reference instruction:* uses 12 bits to specify an address & one bit for addressing mode I, I is 0 for direct & I is 1 for indirect addressing.

2. *The register-reference instruction:* are recognized by 111 in operation code with 0 in left most bit. Remaining 12 bits specify an operation or test on AC.

3. *Input-output instruction:* does not need to refer memory & is recognized by 1 in the left most bit & 111 in other position of the opcode.

Instructions in a Simple Computer (Contd.).

- Basic computer instruction format:



15 14 12 11

0

I Opcode Address

(Opcode=000 through 110)

A) Memory-reference instruction

15 12 11

0

0 1 1 1 Reg. operation

(Opcode=111, I=0)

B) Register -reference instruction

15 12 11

0

1 1 1 1 I/O Operation

(Opcode=111, I=1)

C) Input-output instruction

Instructions in a Simple Computer (Contd.).

- Instruction register bits are connected in the following way:

  Bit 15 is transferred to flip flop represented by I.

  Bits 0-11 are applied to control logic gates.

  Opcode bits 12-14 are sent to a decoder (3:8).

- 4-bit sequence counter (SC) in the control unit counts in binary from 0 to 15 which are decoded into 16 timing signals,T0 through T15.

- The SC can be incremented or cleared synchronously.

Control Unit of a Simple Computer

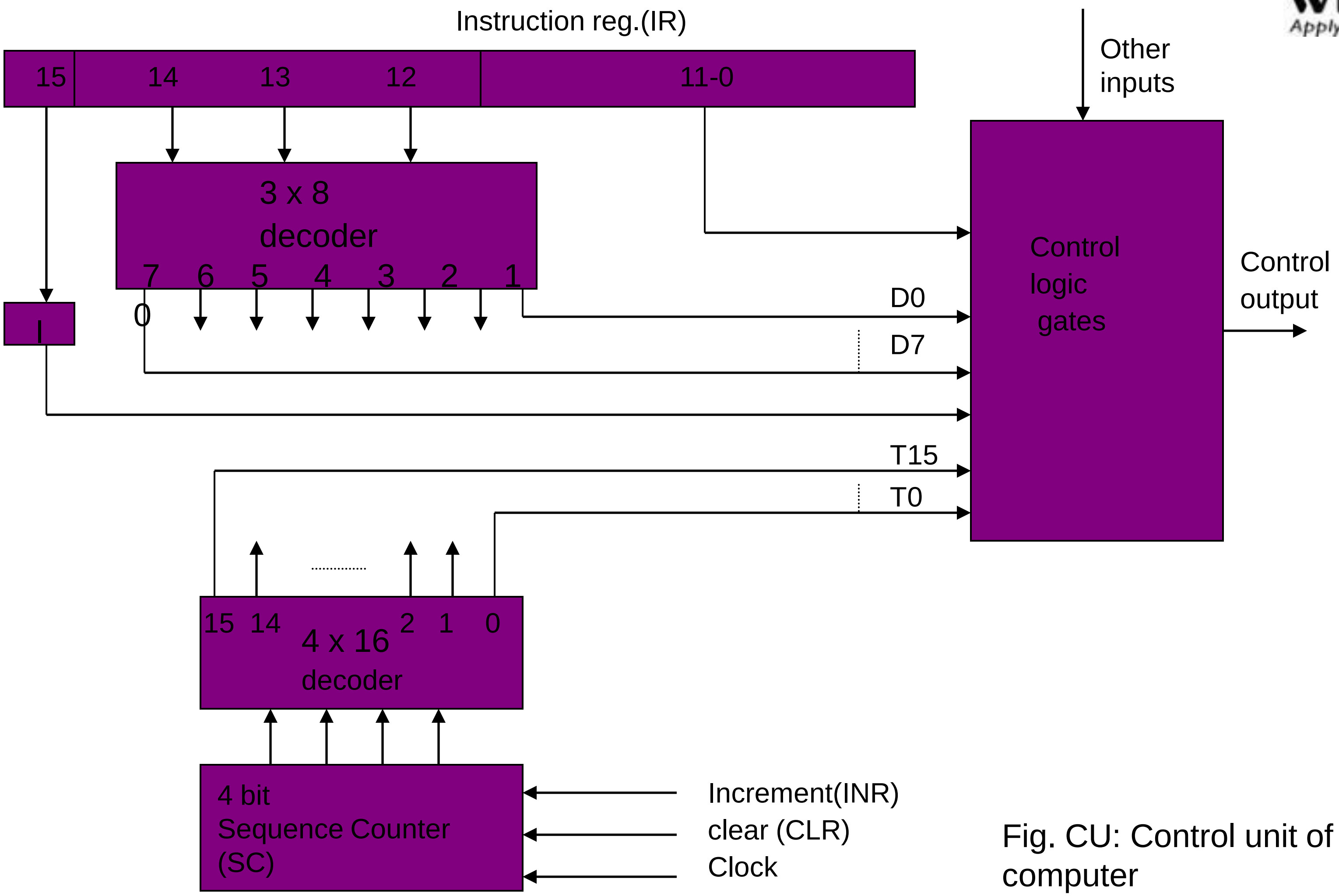


Fig. CU: Control unit of Simple computer

Summary

In this module, we discussed:

- Instruction formats
- Zero, one two operand instructions
- Simple control unit of a computer

Central Processing Unit

Module 3

Objectives

At the end of this module, you will be able to:

- Describe building blocks of CPU registers (MAR,PC,IR etc)
- Define Register Transfer Language
- Illustrate bus and Memory transfer data transfer
- Explain Micro-operations – Arithmetic and logic
- Know what is hardwired and micro programmed control

Duration: 1 hr

The Central Processing Unit (CPU)

- Building blocks of a CPU
- Register Transfer
- Bus and Memory Transfer
- Arithmetic, Logic and Shift Micro-operations
- Arithmetic, Logic Shift unit
- Control Unit

Basic building blocks of a CPU

- Memory Address Register (MAR) - This register holds the address of the memory location to be accessed. The desired address is loaded into MAR through a common Bus.

- Program Counter (PC) - This register keeps track of the program flow. Its count is incremented after every fetch of Op Code or Operand from the Program Memory.

- Stack Pointer (SP) - This is an up-down counter which holds the address of the top of the Stack. Its count is incremented or decremented after every Push or Pop operation of the Stack.

- Instruction Register (IR) - The Op Code fetched from the program memory is loaded here, and its output is fed to the Instruction Decoder for generating the Control Signals.

- Temporary Registers - These registers are accessible only to the system, and cannot be accessed by the user.

Basic building blocks of a CPU (Contd.).

- Programmable ALU - The arithmetic and logic functions of the ALU are selectable by a multi-bit control input (SAL).

- Register array - This consists of general-purpose registers selectable by a multi-bit control input (SRG).

- Instruction Decoder - This block generates the sequence of control signals necessary for making each of the functional blocks, described above, to work according to the Op Code held in IR.
- The internal hardware organization of a digital computer is defined by the following:
  - ✓ The set of registers it contains & their functions.
  - ✓ The sequence of micro-operations performed on the binary information stored in the registers.
  - ✓ The control that initiates the sequence of micro-operations.

Register Transfer

Descriptive explanation of the micro operations becomes lengthy and ambiguous.

Use Register transfer language (RTL):

Symbolic notation used to describe the micro-operations on the contents of the registers.

Ex: MAR-register holds the address of a memory location.

PC - Program counter, IR - Instruction register.

Advantages:

- Helps to express in symbolic form - Concise, Precise.
- Tool for describing the internal organization.
- Facilitate design process.

Register Transfer (Contd.).

Control functions:

- Transfer/manipulation of data occurs only under a predetermined condition which is generated using control signals.

Ex: If (P=1) then (R2   R1)

where P is the control ← signal generated in control section.

Above statement can also be represented as:

P: R2   R1, where P is the control function (0 or 1).

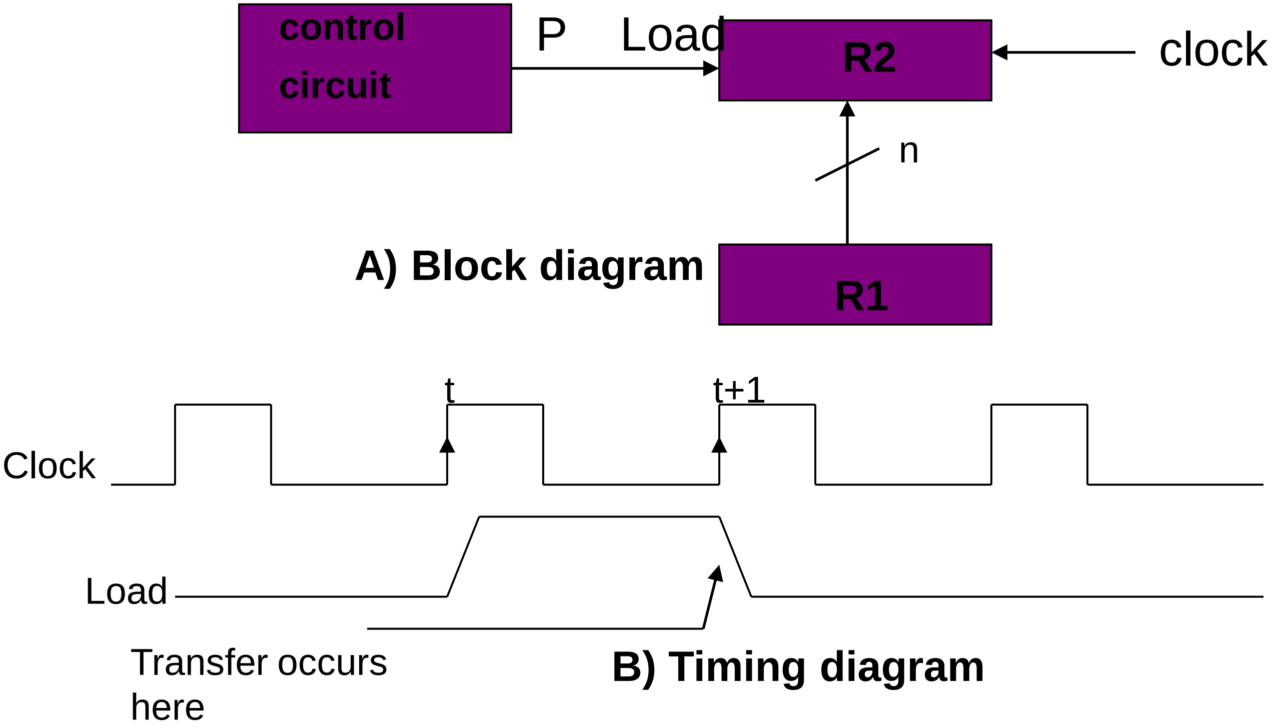
←

Ex:T: R2   R1, R1   R2 .

←   ←

Register Transfer (Contd.).

- Every register transfer implies a hardware construction for implementing it.
- Consider transfer from R1 to R2 when P=1. i.e. $(P: R2 \leftarrow R1)$



A) Block diagram

B) Timing diagram

If P becomes 0 transfer stops; **else** transfer takes place with each new clock pulse.

Bus and Memory Transfer

- Common bus: Efficient scheme for transferring information between multiple registers.

  $\text{Bus} \leftarrow C, \quad R1 \leftarrow \text{Bus} \quad \Longrightarrow \quad R1 \leftarrow C$

A common bus system can be constructed using multiplexers.

- Memory read: The read operation can be depicted as:

  Read: $DR \leftarrow M[AR]$

  AR is address register , DR is data register and M is memory word.

- Memory write**:** The write operation can be depicted as:

  Write : $M[AR] \leftarrow R1$

Micro-operations

- Micro-operations are elementary operations performed within defined clock cycles, on the data stored in the registers present in the various modules.

  Ex: Operations like movement of data within the registers of the CPU such as

  shift, clear, load etc,

  arithmetic operations,

  logic operations like AND, OR, NOT, XOR, etc.,

- Each instructions given to a computer consists of sequence of these micro-operations within the CPU.
- Micro-operations are controlled by the Control unit that generates the control signal for execution of operations.

Types of Micro-operations

The micro-operations are classified into four categories:

1. Register transfer micro-operations: transfer binary information from one register to another. (No change of information)

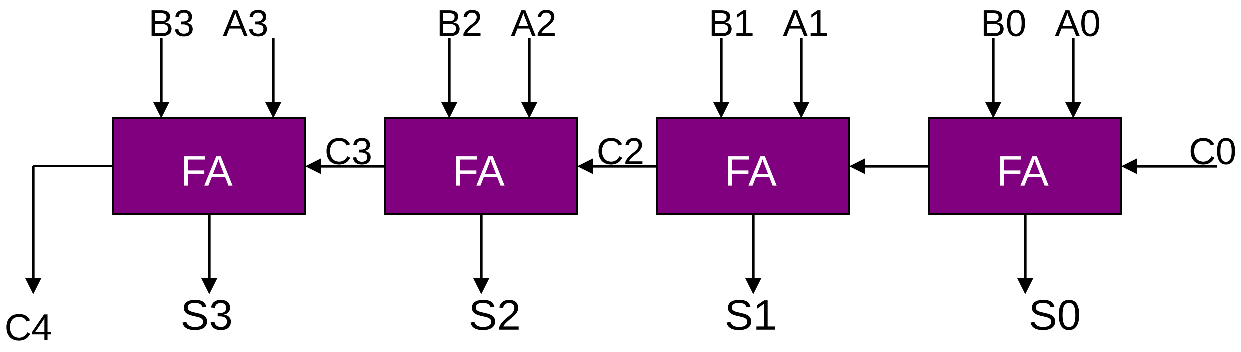
2. Arithmetic micro-operations: perform arithmetic operations on numeric data stored in registers.

3. Logic micro-operations: perform bit manipulation operations on non-numeric data stored in registers.

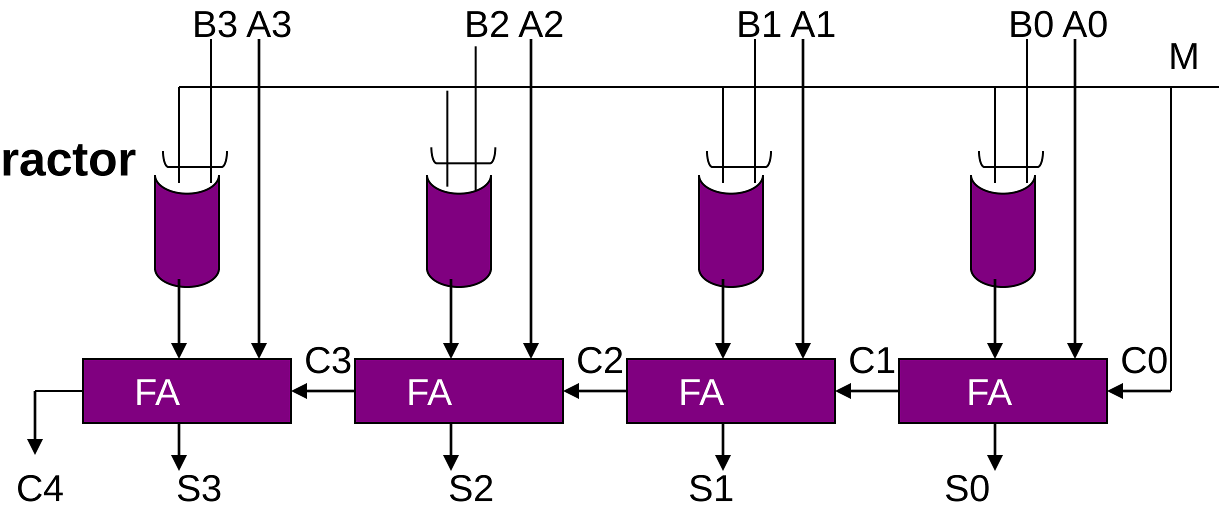
4. Shift micro-operations: perform shift operations on data stored in registers.

Arithmetic Micro-operations

4-Bit binary adder.



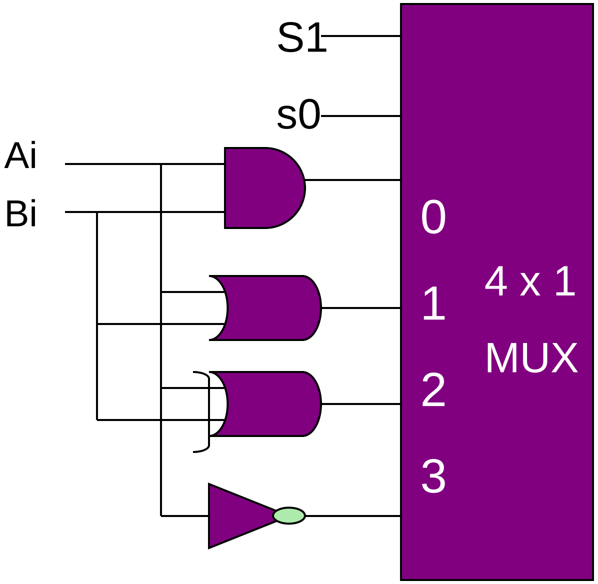
4-bit adder-subtractor



Logic Micro-operations

- One stage of logic circuit: Implementation requires the logic gates be inserted for each bit or pair of bits in the registers to perform the required logic function.

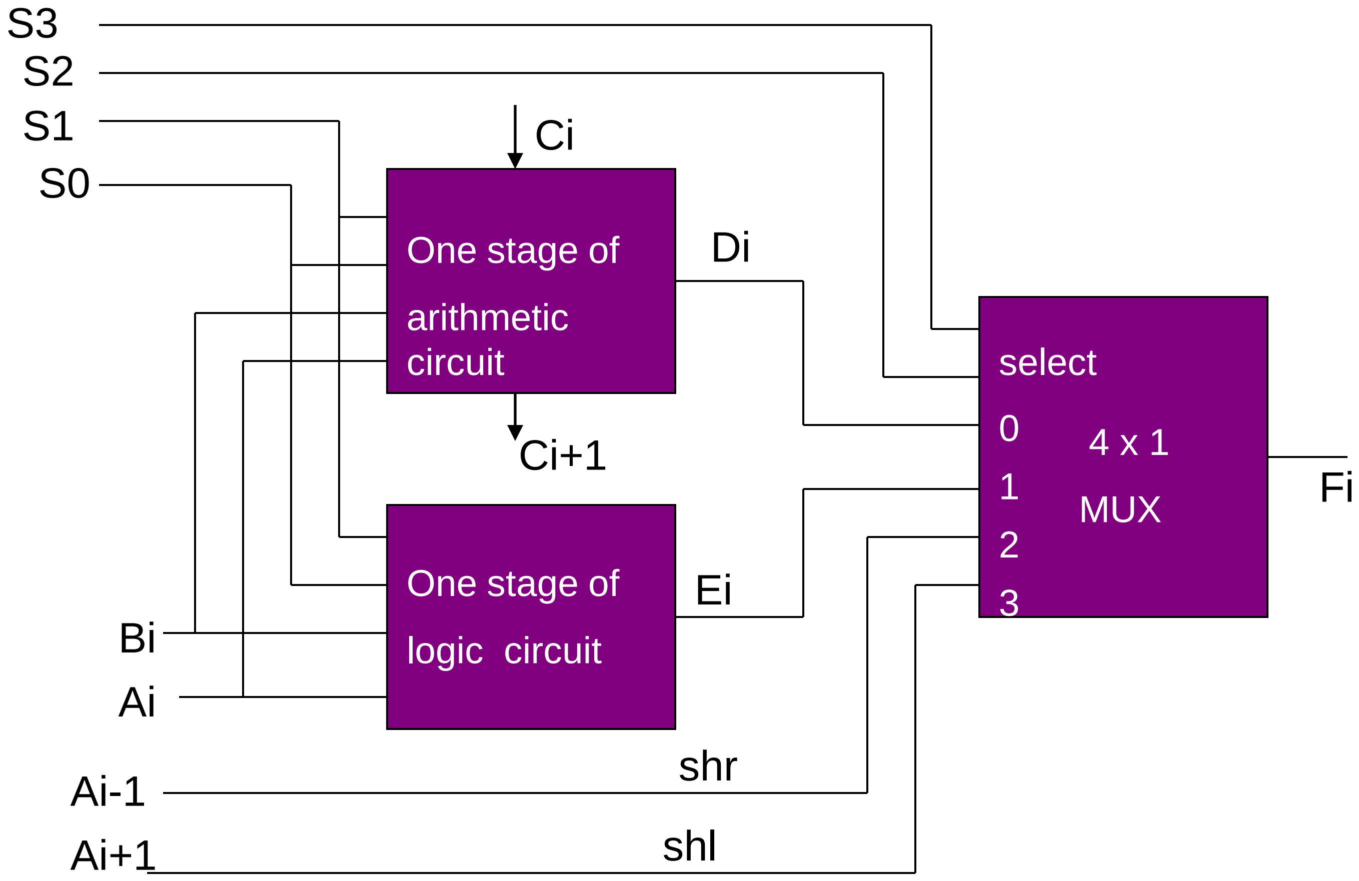
A) Logic diagram



B) Function table

S1 S0	Output	Operation
0 0	E= A.B	AND
0 1	E=A+B	OR
1 0	E=A ⊕ B	XOR
1 1	E = A’	NOT

Arithmetic Logic Shift Unit



A)One stage of arithmetic logic shift unit

CPU Organization

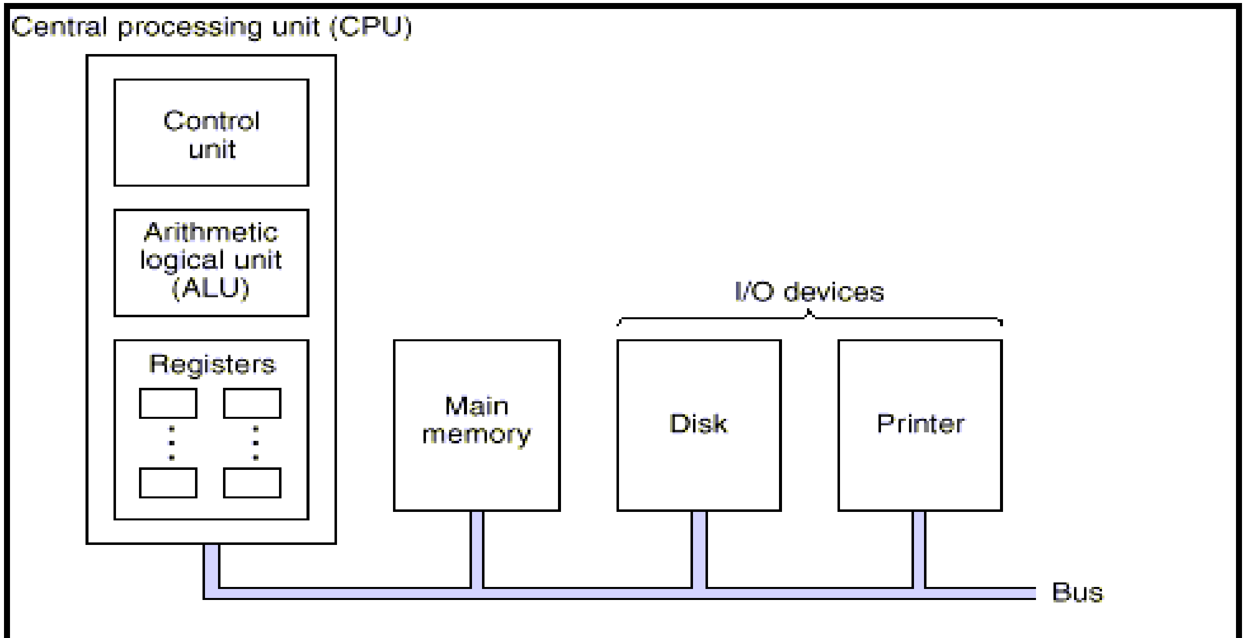
- Central role in the computer system and does the bulk of data processing operations.

- Executes programs (instructions) stored in the main memory by fetching the instructions, examining & executing the instructions one by one.

- Supervises other system components, peripherals

- Has three parts: Register set, ALU, CU.

CPU Organization (Contd.).



The organization of a simple computer with one CPU and two I/O devices.

CPU Organization (Contd.).

Instruction execution:

CPU steps in executing the instruction:

1. Fetch the next instruction from the memory into the instruction register (IR).
2. Update the Program Counter (PC).
3. Determine the type of instruction just fetched (decode).
4. If the instruction uses data in memory, determine their location
5. Fetch the data, if any, into the internal CPU registers.
6. Execute the instruction.
7. Store the result(s) in the proper place.
8. Go to step 1 to begin executing the next instruction.

The above sequence is referred as fetch-decode-execute cycle

CPU Organization (Contd.).

Primary function of CPU: Main function of a processor is to execute sequences of instructions stored in a memory called the main memory.

The sequence of operations involved in processing an instruction constitutes an **instruction cycle**, which can be divided as two phases:

1. **Fetch cycle:** In the fetch cycle the instruction is obtained from the main memory

2. **Execution cycle:** The execution cycle includes decoding the instruction, fetching any required operands, & performing the operation specified by the instruction's opcode.

CPU Organization (Contd.).

Secondary function of CPU:

In addition to executing the program, the CPU supervises the other system components,

For Ex., the CPU directly or indirectly controls the IO operations such as data transfer between IO devices & main memory.

These operations require CPU attention infrequently & such an IO request for CPU is called an interrupt.

CPU Organization (Contd.).

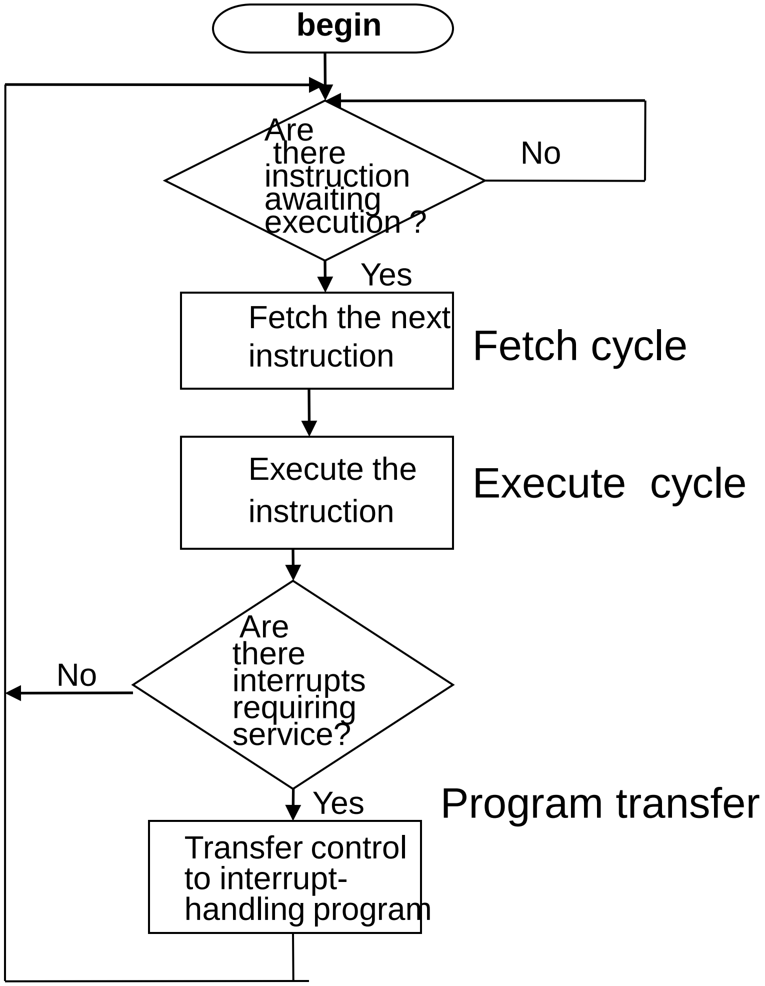

begin
Are there instruction awaiting execution ?
No
Yes
Fetch the next instruction
Fetch cycle
Execute the instruction
Execute cycle
Are there interrupts requiring service?
No
Yes
Transfer control to interrupt-handling program
Program transfer

Fig.A. Overview of CPU behavior.

CPU Organization (Contd.).

A simple Accumulator - based CPU.

Here the CPU register called the Accumulator (AC) plays the central role, being used to store an input or an output operand (result) in the execution of most instructions.

'Hence the name'.

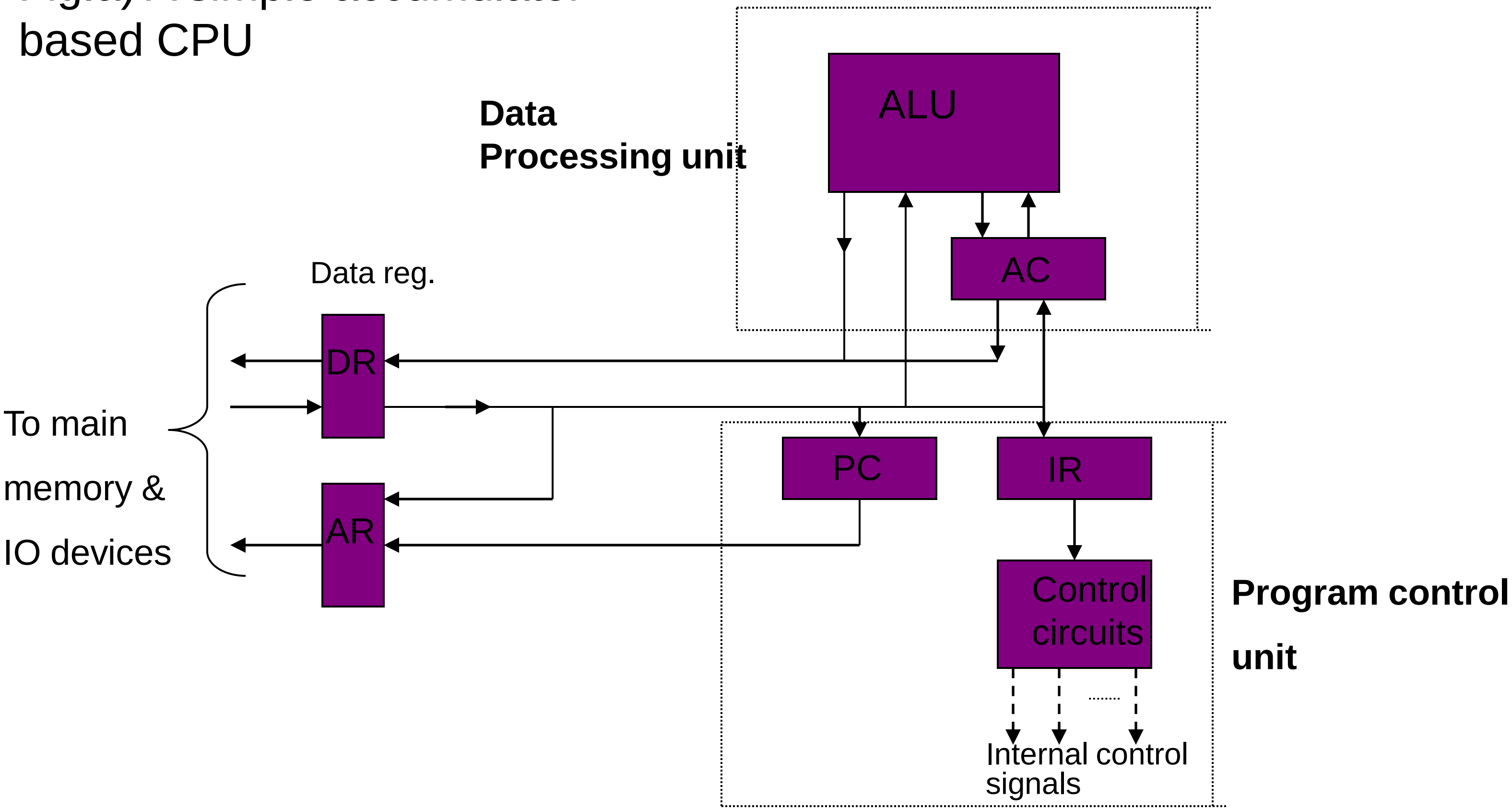
Refer Fig a. A simple accumulator based CPU.

Fig b. Operation of the CPU.

(On the next two pages)

CPU Organization (Contd.).

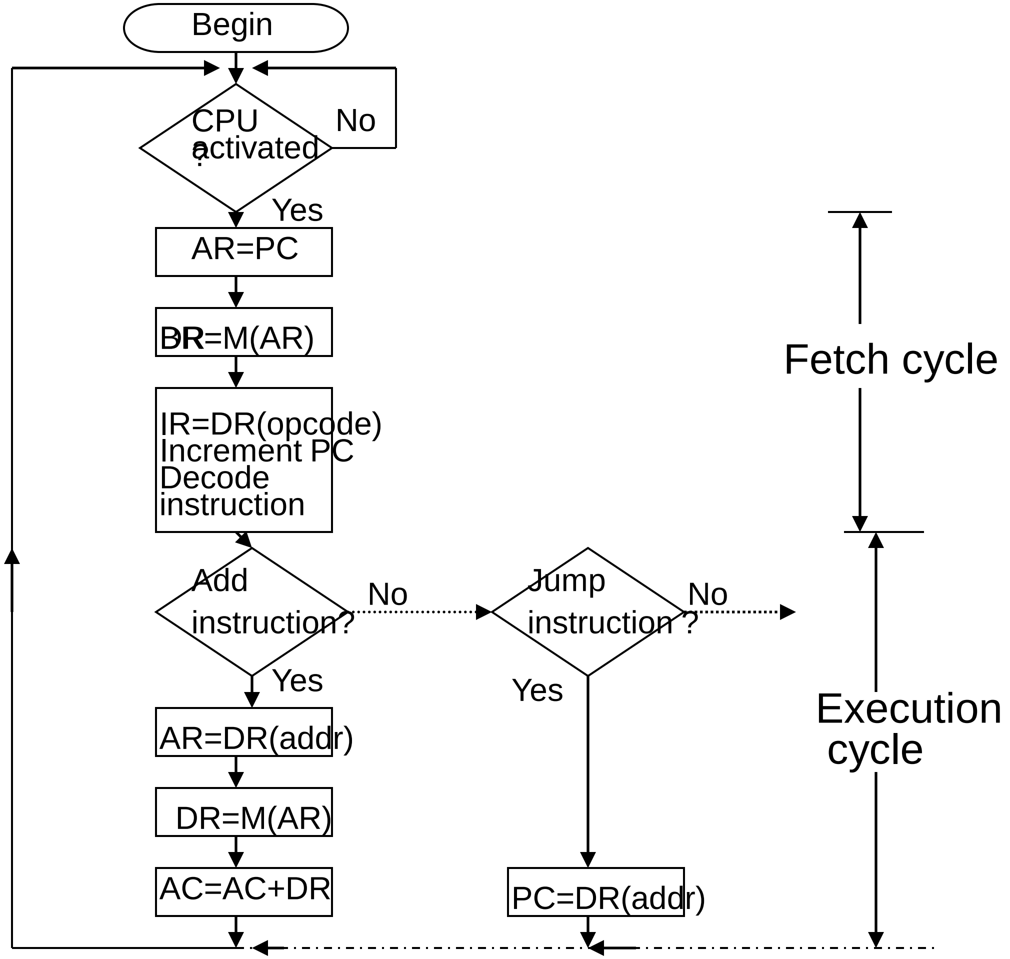
Fig.a) A simple accumulator based CPU



CPU Organization (Contd.).

Operation of the CPU

Fig.b)



Control Unit

Digital system → data processing unit.
→ control unit.

- Data processing unit: It is a network of functional units that perform certain operation on data.
- Control unit: Issues control signals or instruction to the data processing unit.
- Two major types of control unit:
  - 1.Hardwired control: The control logic is implemented with gates, flip-flops, decoders & other digital circuits.
  - 2.Microprogrammed control: The control memory is programmed to initiate the required sequence of micro-operations.

Control unit (Contd.).

A typical hardwired control unit:

(1)The block diagram consist of two decoders, sequence counter & a number of control logic gates. Ref. Fig. in the notes page.

(2)Timing signals: The sequence counter (SC) can be incremented to provide sequence of timing signals.

Ex: at time T4, SC is cleared to 0 if decode D3 output is active.

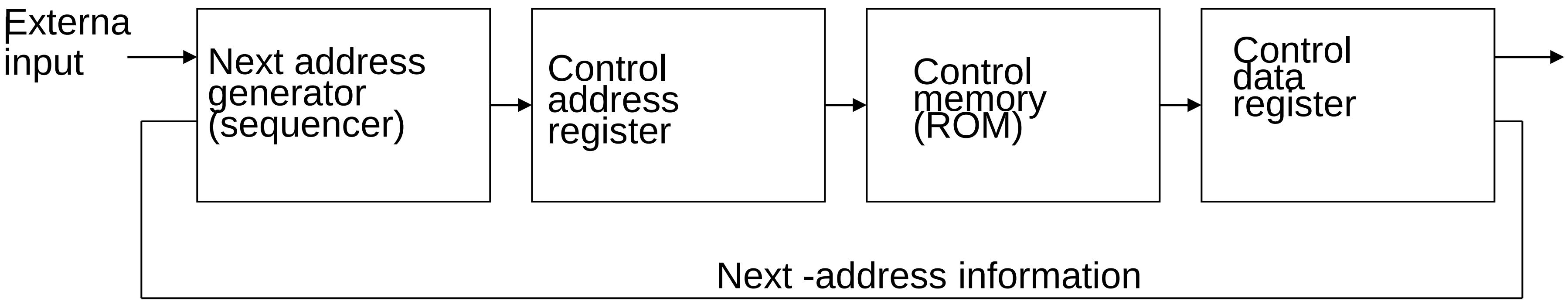
This can be given by statement,

$D_3T_4: SC \leftarrow 0$

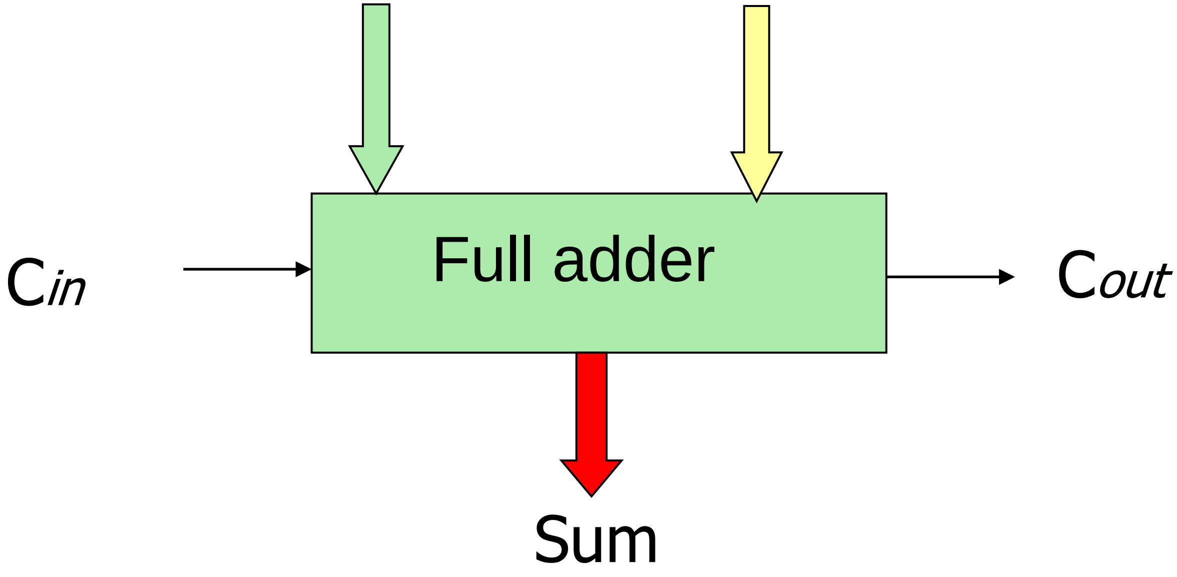
Control unit (Contd.).

Micro-programmed control:

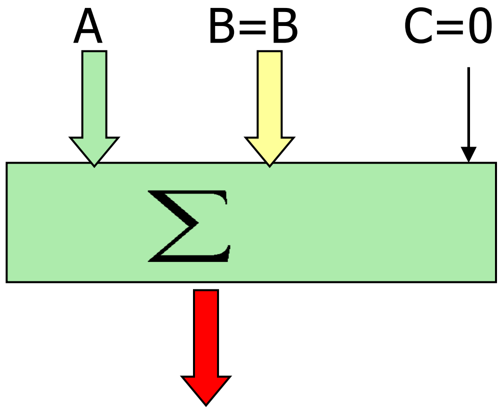
- Control memory: - control word.
  - -microinstruction.
  - -microprogram.
  - -control memory.
  - -control address register.
  - -sequencer.
  - -pipeline register & hardwired control.



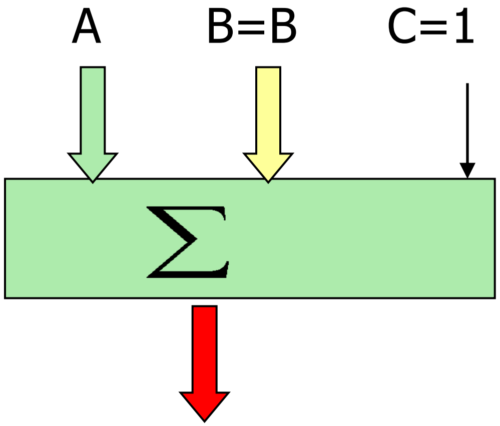
Control unit (Contd.).



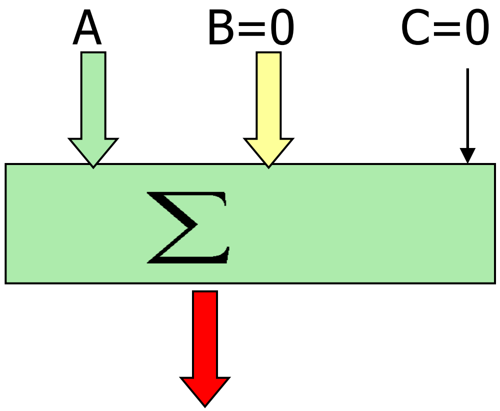
Control unit (Contd.).



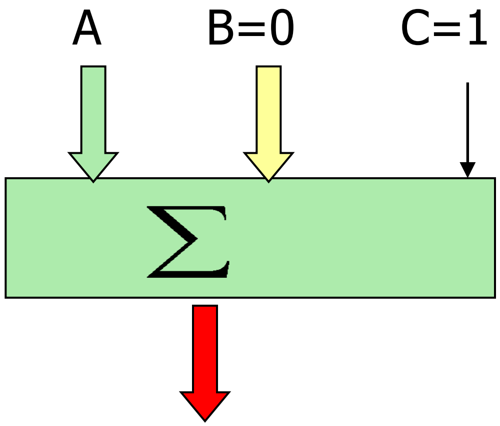
OUT=A+B



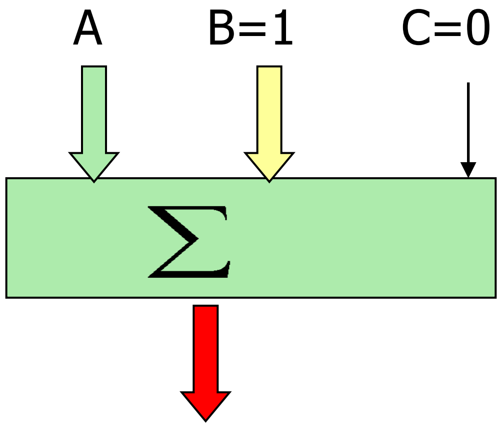
OUT=A+B+1



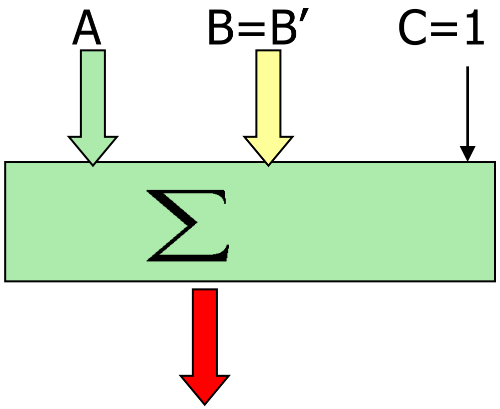
OUT=A



OUT=A+1



OUT=A-1



OUT=A-B

Control unit (Contd.).

- Data
- Address
- Control

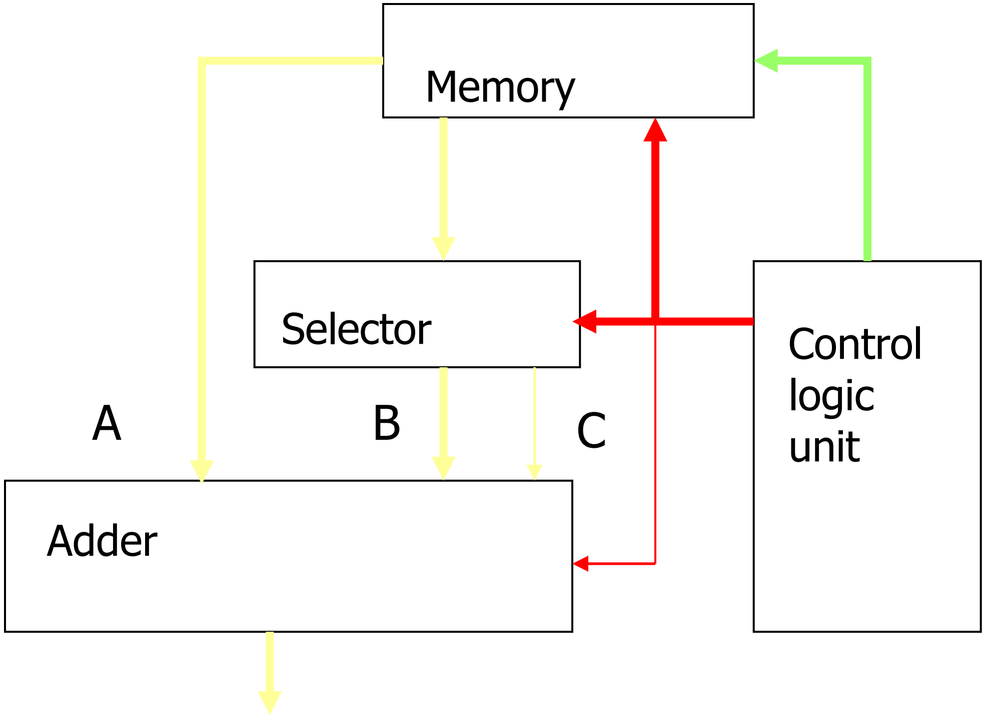


Fig. Computer hardware configuration

Control unit (Contd.).

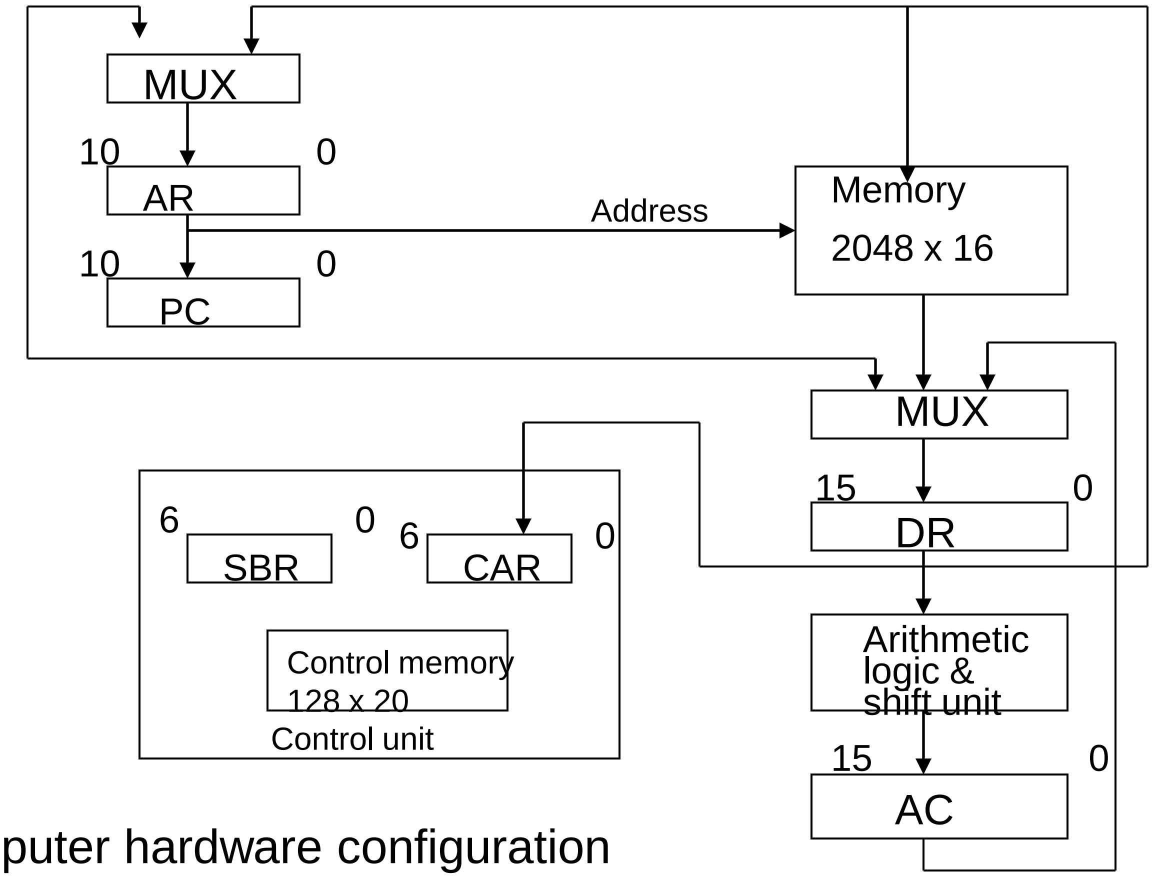


Fig. Computer hardware configuration

Summary

In this module, we discussed:

- Building blocks of CPU registers (MAR,PC,IR etc)
- Register Transfer Language
- Bus and Memory transfer
- Micro-operations – Arithmetic and logic
- CPU organization
- Control unit – Hardwired and microprogrammed
- Simple Control Unit

Memory Organization

Module 4

Objectives

At the end of this module, you will be able to:

- List Memory types Memories
- Discuss Magnetic. Optic (CDROM) and semiconductor memories, ROMs and RAMs
- Know cache memories, Flash memories

Duration: 1 hr

Memory Organization

1.Internal processor memory: This consists of a small set of high-speed registers used as a working memory for temporary storage of instructions & data.

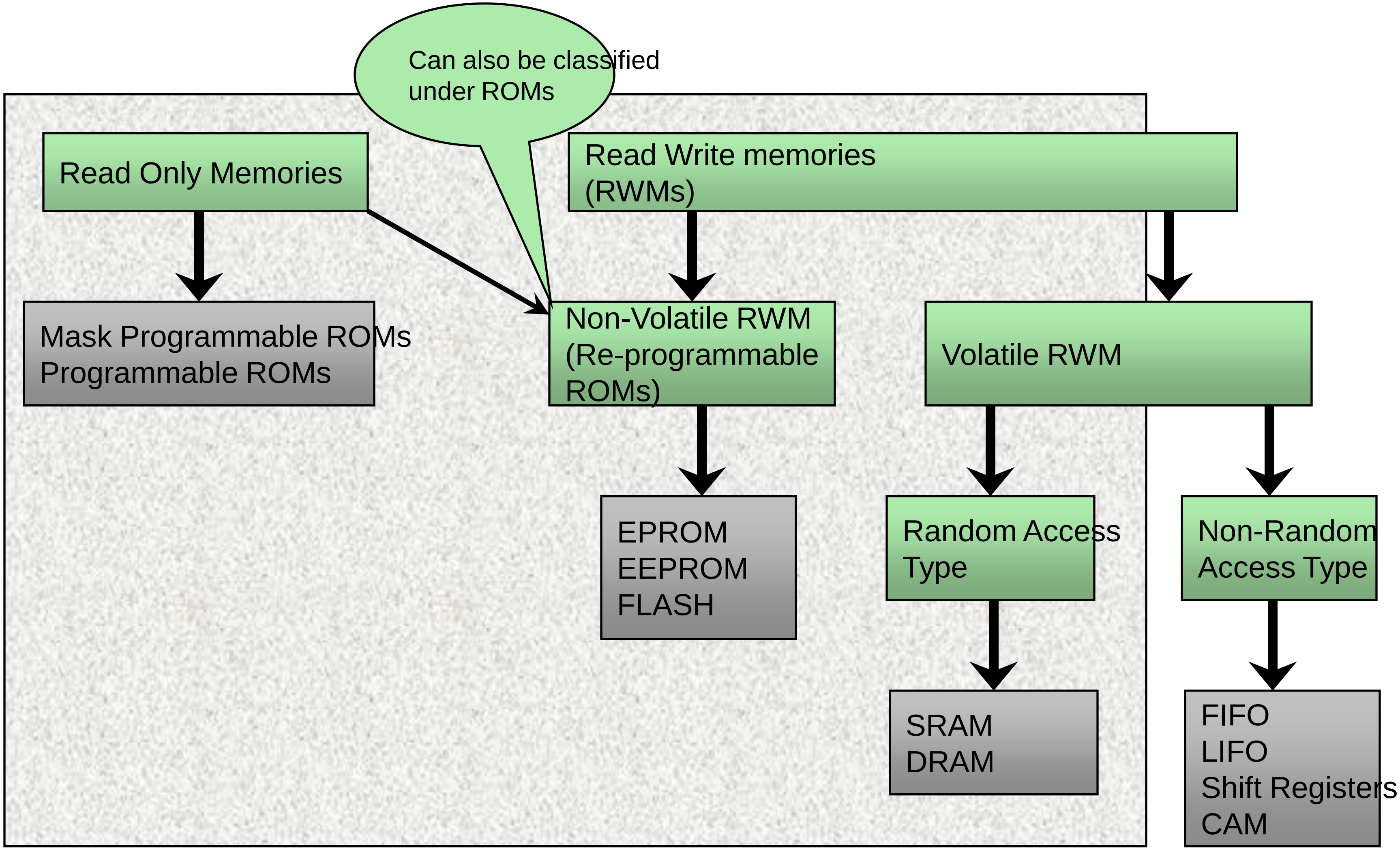
2.Cache memory: It is a very high speed memory. It allows the CPU to access information quickly.

3.Main memory: The memory unit directly communicates with CPU. This is a relatively large fast memory used for program & data storage during computer operation.

4.Auxiliary memory: Devices that provide backup storage are called auxiliary memory.

Ex: Hard disks, CD-ROMs, magnetic tapes.

Memory Organization (Contd.).



Can also be classified under ROMs

Read Only Memories

Read Write memories (RWMs)

Mask Programmable ROMs
Programmable ROMs

Non-Volatile RWM (Re-programmable ROMs)

Volatile RWM

EPROM
EEPROM
FLASH

Random Access Type

Non-Random Access Type

SRAM
DRAM

FIFO
LIFO
Shift Registers
CAM

Memory Organization (Contd.).

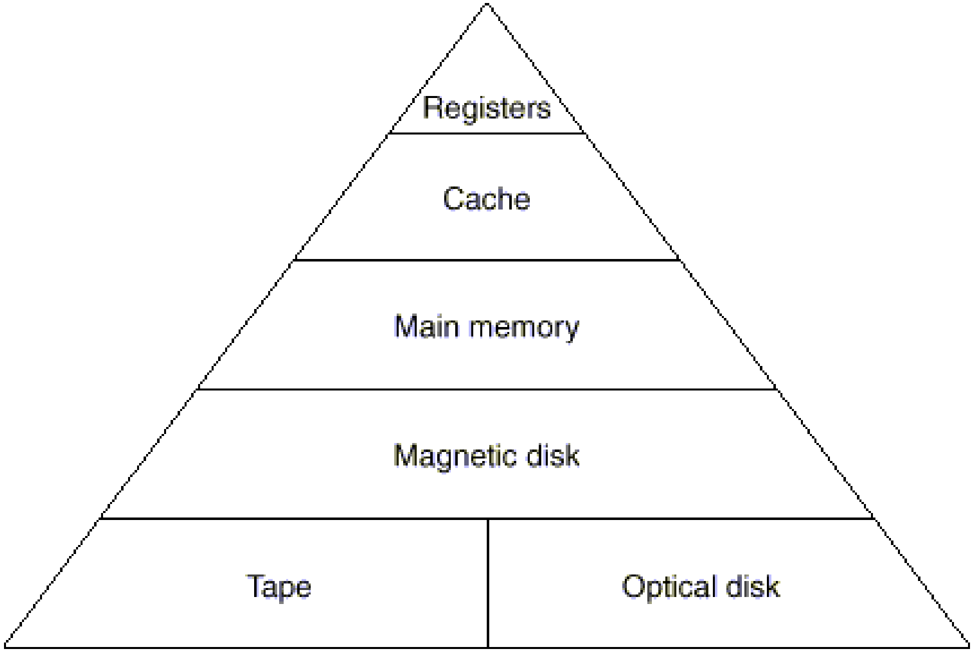


Fig.UIV.1 A five-level memory hierarchy.

Fig. Five levels of memory hierarchy

Memory Organization (Contd.).

Memory hierarchy in a computer

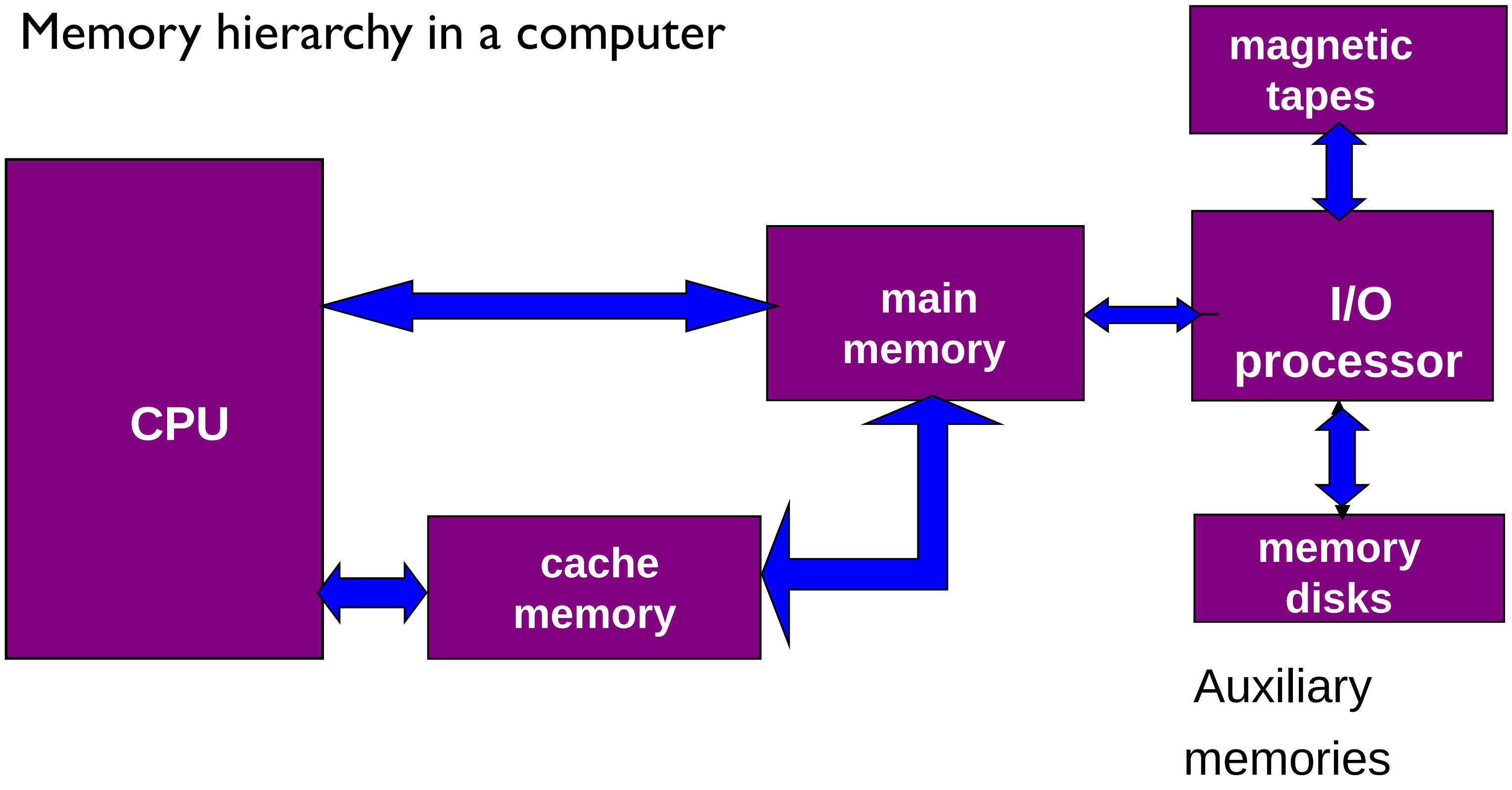


Fig. Memory hierarchy.

Memory Organization (Contd.).

Access mode: An important property of a memory device is the order or sequence in which information can be accessed.

- If location can be accessed in any order & the access time is independent of the location being accessed, then the memory is called a random-access memory (RAM).

  Ex: Ferrite-core & semiconductor memories.

- Memories where storage location can be accessed only in certain predetermined sequences are called serial-access memories.

  Ex: Magnetic-tape, Magnetic-bubble, & optical memories employ serial methods.

Memory Organization (Contd.).

Storage mechanism: The physical processes involved in storage are sometimes inherently unstable, so the stored information will be lost unless proper action will be not taken.

- Three important memory characteristics that can destroy the information.

  i) Destructive readout

  ii) dynamic storage

  iii) volatility.

Memory Organization (Contd.).

- Destructive Read Only (DRO): Memories having the property that the method of reading a memory location destroys the stored information.

- Non Destructive (NDRO): Memories in which reading does not affect the stored data.

- Static Memory: Memories that does not require periodic refreshing

- Dynamic Memory: Memories that require periodic refreshing
  - Refreshing: In some memory, over a period of time, the stored information (in form of charges) tends to leak (discharge), unless the charge is restored by a periodic refreshing.

Memory Organization (Contd.).

- t_A is the time between the receipt of a read request by the memory & the delivery of the requested information to its external output terminals.

- Cycle time(t_M) is time needed to complete any read or write operation in the memory.

- But in DRO & dynamic memories, it is required to refresh the memory state.

- Data-transfer rate : The maximum amount of information that can be transferred to or from the memory, every second is $1/t_A$,this quantity is called the data-transfer rate or bandwidth b .

Memory Organization (Contd.).

- Volatile memory: Memories in which the stored information remains valid as long as the power is applied to the unit are called volatile memories.

  Ex : RAMs.

- Non-volatile memory: Memories that can store the information even when the applied power is removed

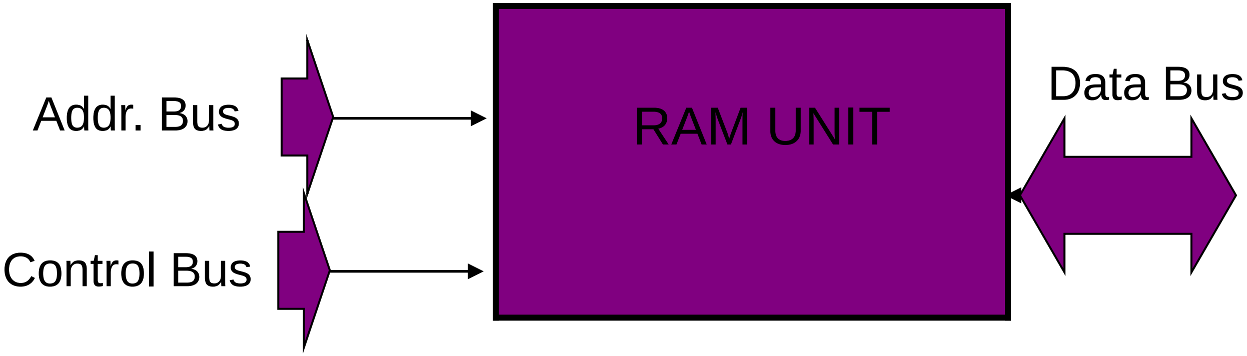
Ex: ROMs, magnetic tapes, magnetic disks, optical disks.

Most semiconductor memories are volatile & most of the magnetic memories are nonvolatile.

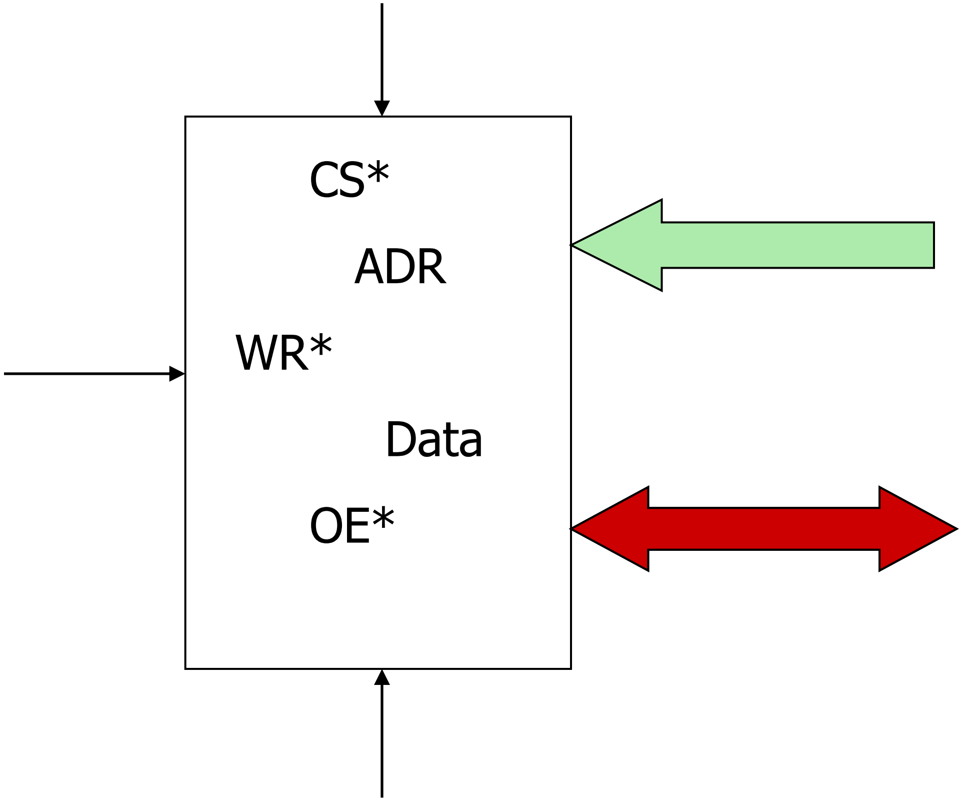
Memory Org (Contd.).

RAM Organization: A general approach to reducing the access circuitry cost in RAM is by using matrix, or array, organization.

- One-dimensional or 2-dimensional memory organization is used.
- The semiconductor & ferrite core can be used for memory array organization.
- Fig. below shows a RAM unit with input-outputs:



Memory Organization (Contd.).



Memory Organization (Contd.).

The figure shows the internals of a RAM unit.

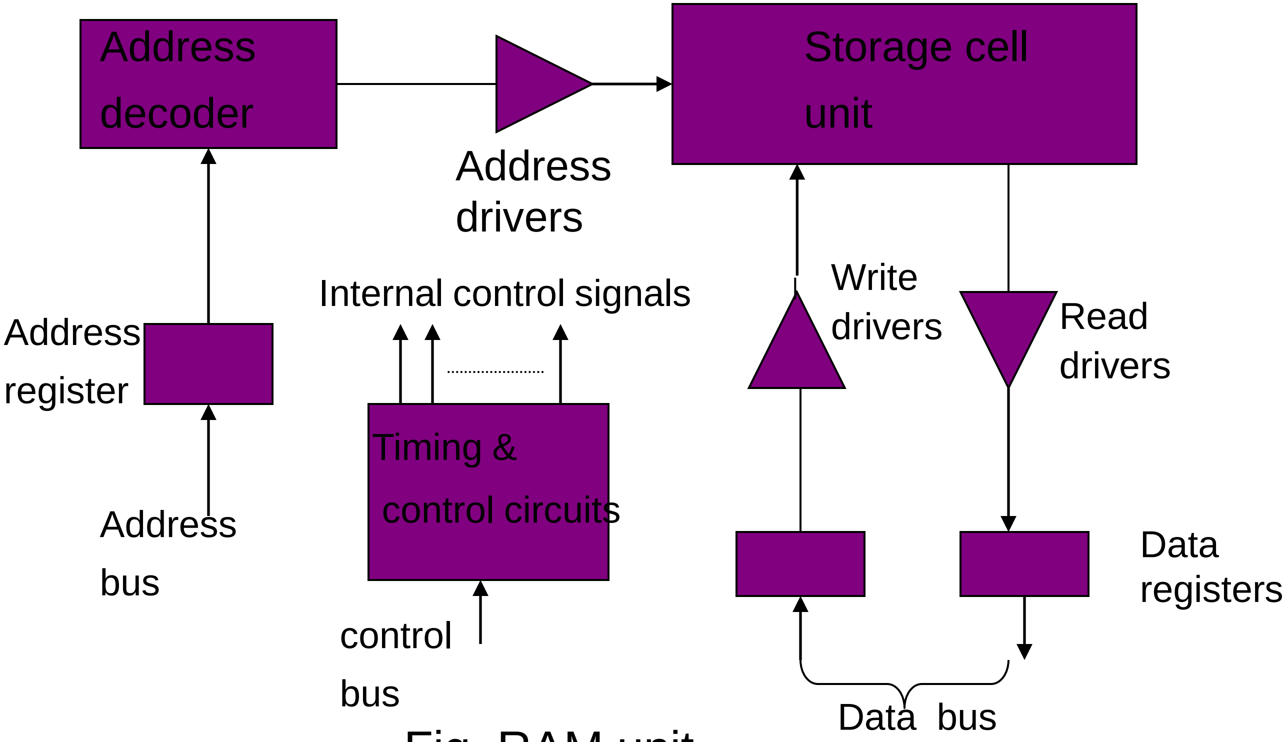


Fig. RAM unit

Memory Organization (Contd.).

A 128x8 RAM chip:

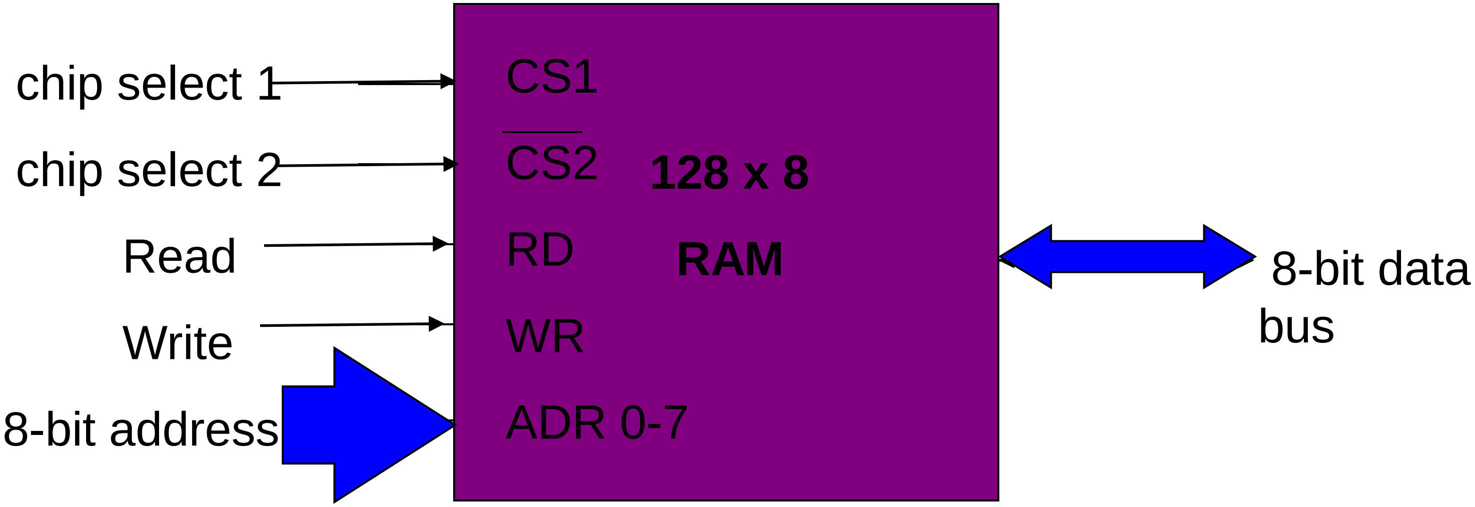


Fig. RAM

Memory Org (Contd.).

Figure shows the RAM in 2-D structure

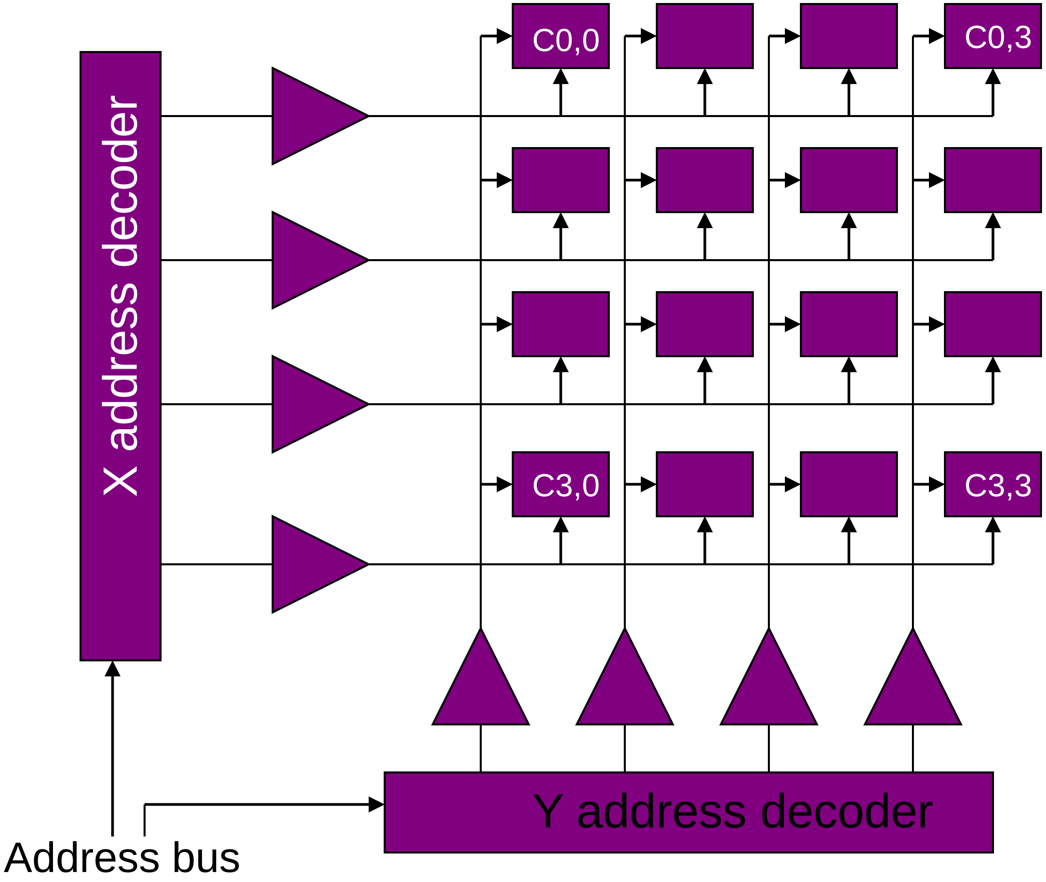


Fig. 2-D scheme

Memory organization (Contd.).

Random Access: Access time is same for all locations, i.e., access can be made randomly irrespective of the locations.

- SRAM: Static Random Access Memory
  - Low density, high power, expensive, fast
  - Static: content last "forever"(until power down)
- DRAM: Dynamic Random Access Memory
  - High density, low power, cheap, slow
  - Dynamic: needs to be "refreshed" regularly

Memory Organization (Contd.).

ROM: Memories whose contents cannot be altered are called read-only memories.

- A ROM is non-erasable storage devices.
- The ROMs are widely used for storing the control programs such as micro-programs.

PROMs: ROMs whose contents can be changed(usually off-line & with some difficulty) are called Programmable Read-Only Memories (PROMs).

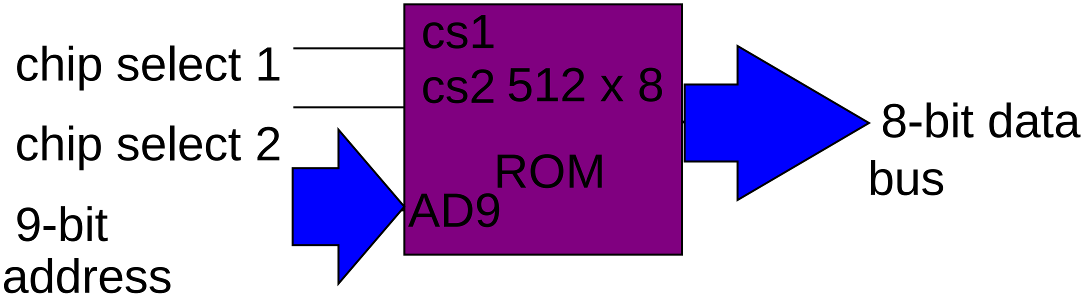
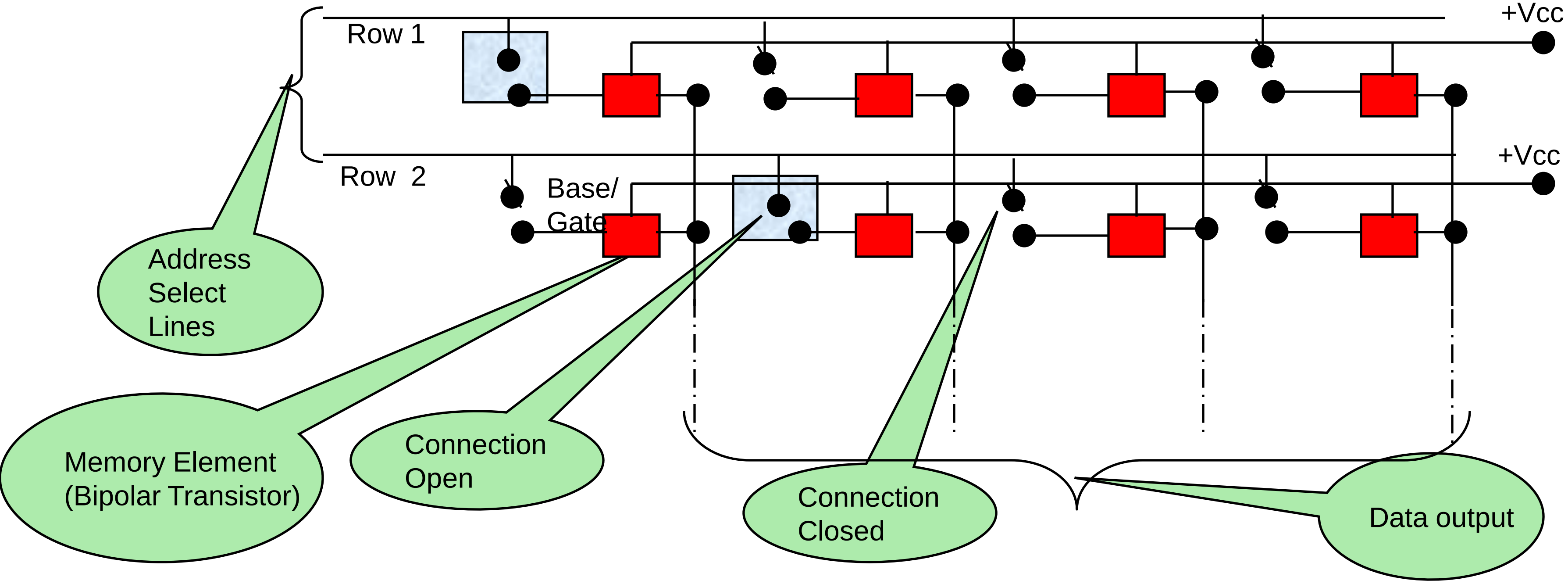


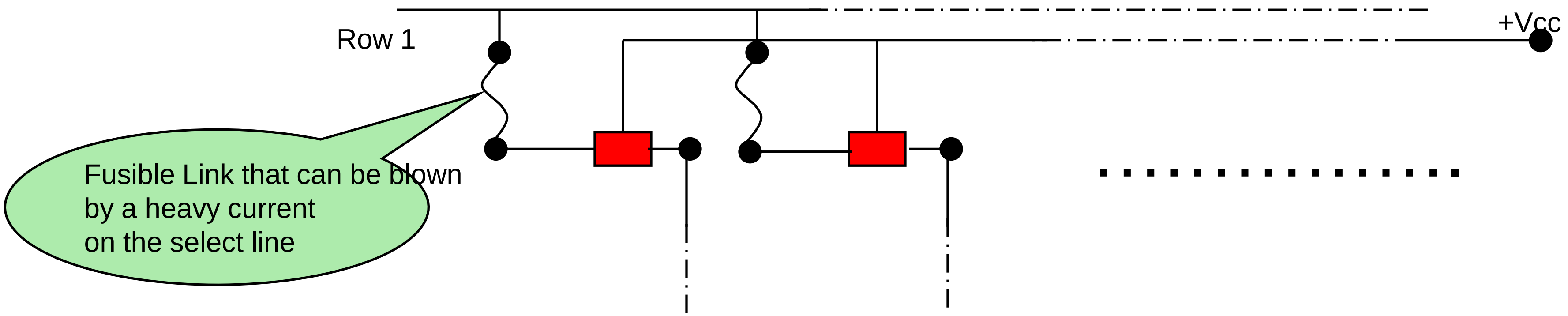
Fig. ROM Chip

Memory organization ROMs



Memory organization PROMs

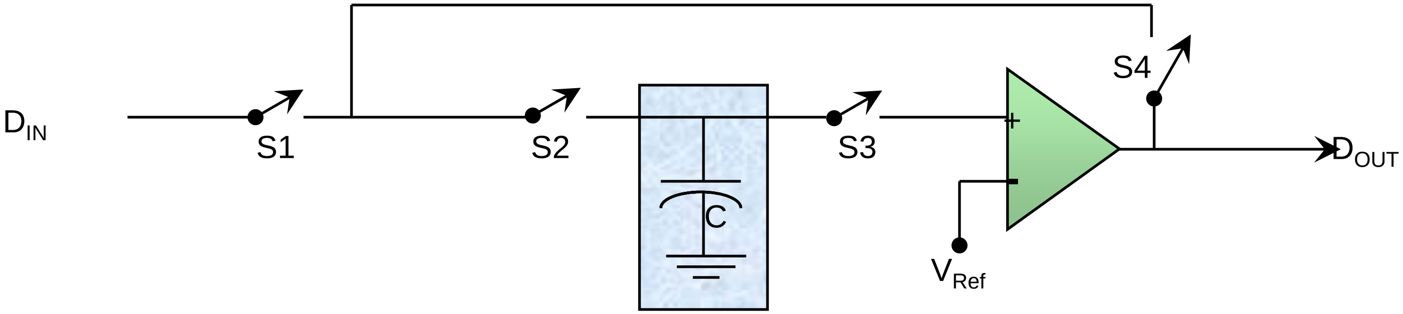
- One-time user-Programmable ROM
- Suitable for lower volume applications
- Internal structure similar to ROM
- Has fusible links for each memory elements that can be permanently blown depending on the bit to be stored in the element
- Can be *downloaded* with user data using a special equipment called PROM *Programmer* .
- Lacks Re-programmability


Row 1
+Vcc
Fusible Link that can be blown by a heavy current on the select line

Memory organization DRAM

- MOS Capacitor based RAMs
  - Data stored as a charge on the capacitor
  - Hence has the tendency to discharge and lose the data
  - Needs periodic *refreshing* of the charge
- Density is 4 times that of SRAM due to the simple Cell structure
  - Consume less power
  - Cost is considerably lower than SRAM
- Address inputs need to be handled in a complex manner than the one in SRAM
  - Requires complex timing control in order to take of care of the refreshing
  - Slower than SRAMs in access speed

Memory organization DRAM (Contd.).



Write Operation - S1, S2 Closed

S3, S4 Open

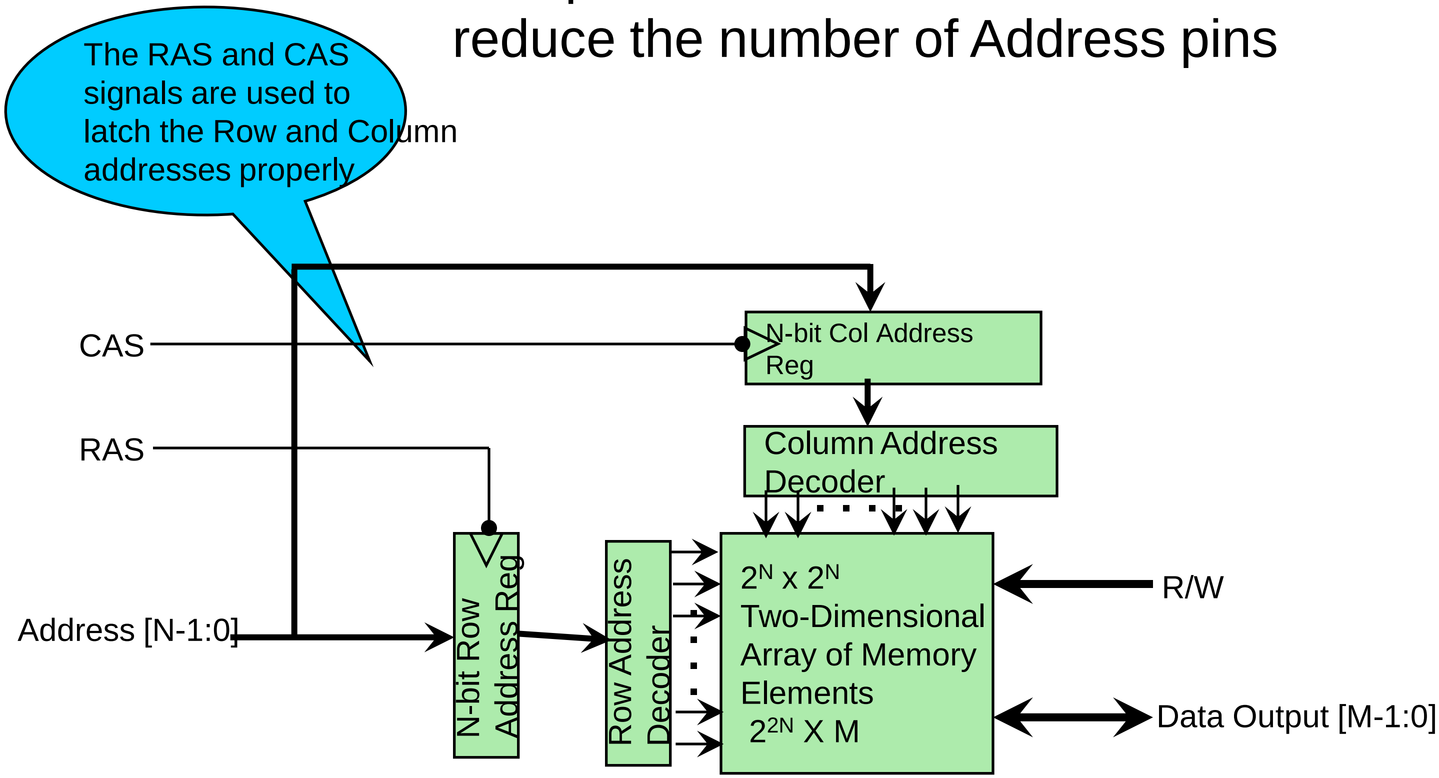
Read Operation - S2, S3, S4 Closed

S1 Open

Read with refresh

Memory organization DRAM (Contd.).

Multiplexed Row and Column Addressing to reduce the number of Address pins


The RAS and CAS signals are used to latch the Row and Column addresses properly
CAS
RAS
Address [N-1:0]
N-bit Col Address Reg
Column Address Decoder
N-bit Row Address Reg
Row Address Decoder
2^N x 2^N Two-Dimensional Array of Memory Elements 2^{2N} X M
R/W
Data Output [M-1:0]

Memory organization DRAM

- Refreshing is an essential event in DRAM
  - Needs to be performed at regular intervals in order to ensure Data integrity
- DRAM Controller on board/on chip is an essential part of the system
  - Controls refresh operations of DRAM
- One of the first things the init code must do is : -
  - Initialize the DRAM controller
- Consult board/chip designer to determine correct initialization sequence.

Memory organization (Contd.).

Auxiliary memory:

Magnetic Disks: It is a circular plate constructed of metal or plastic coated with magnetized material.

- bits are stored in magnetized concentric surfaces is called tracks.
- tracks divided in to sectors.
    - disks may have multiple read/write head.
    - disks attached permanently called hard disks.
    - removable are floppy disks.

Memory Organization (Contd.).

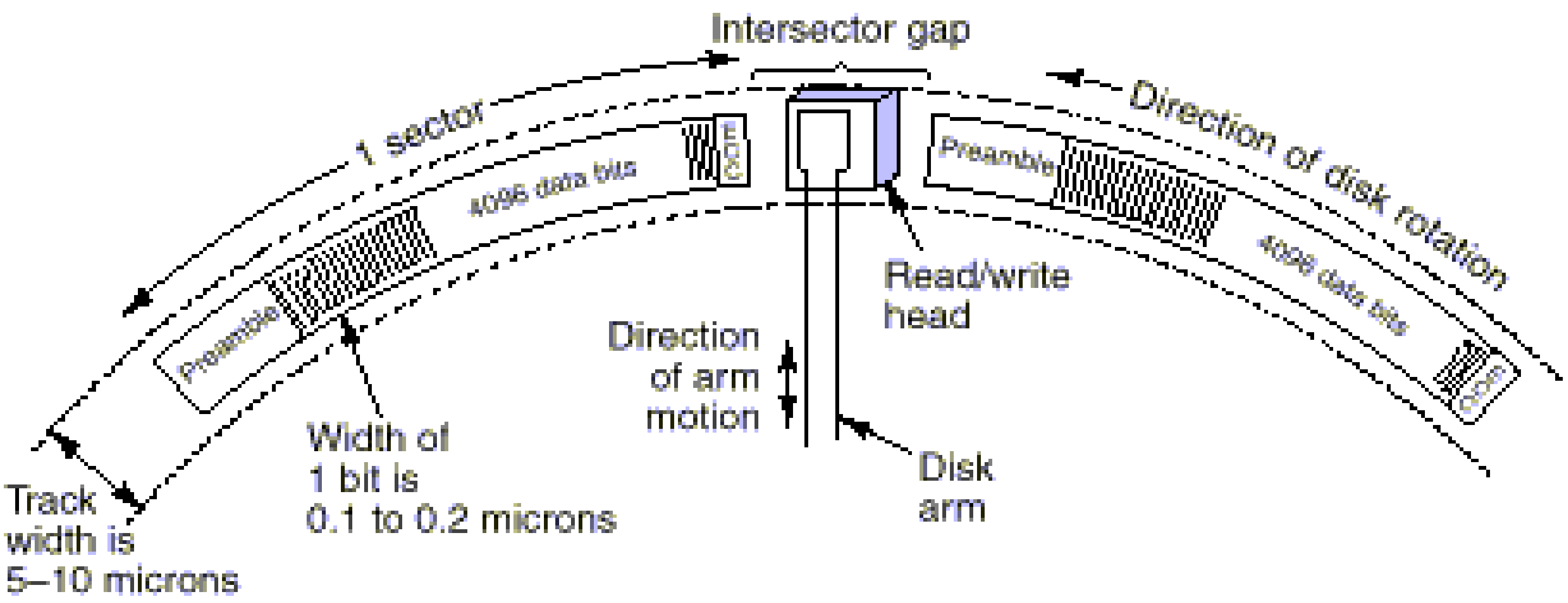


Fig. A portion of a disk track. Two sectors are illustrated.

Memory Organization (Contd.).

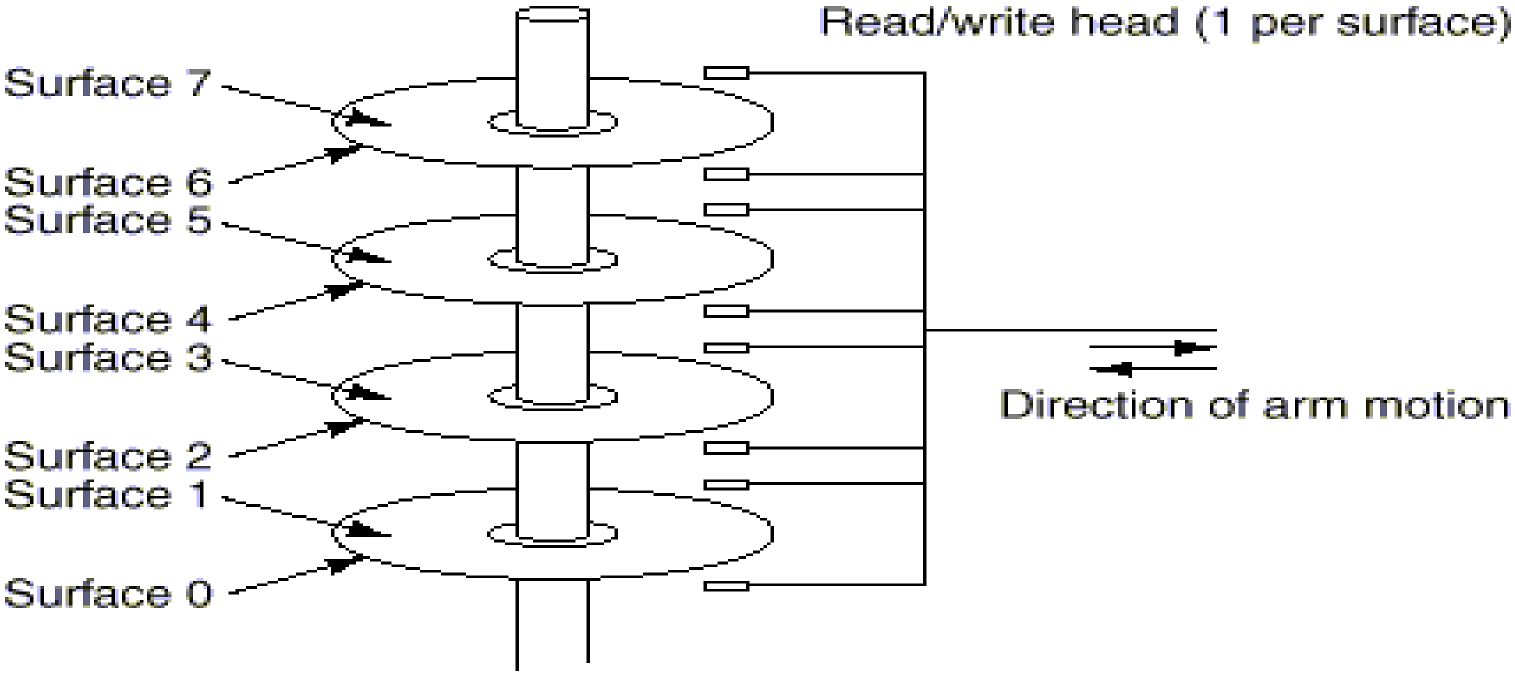
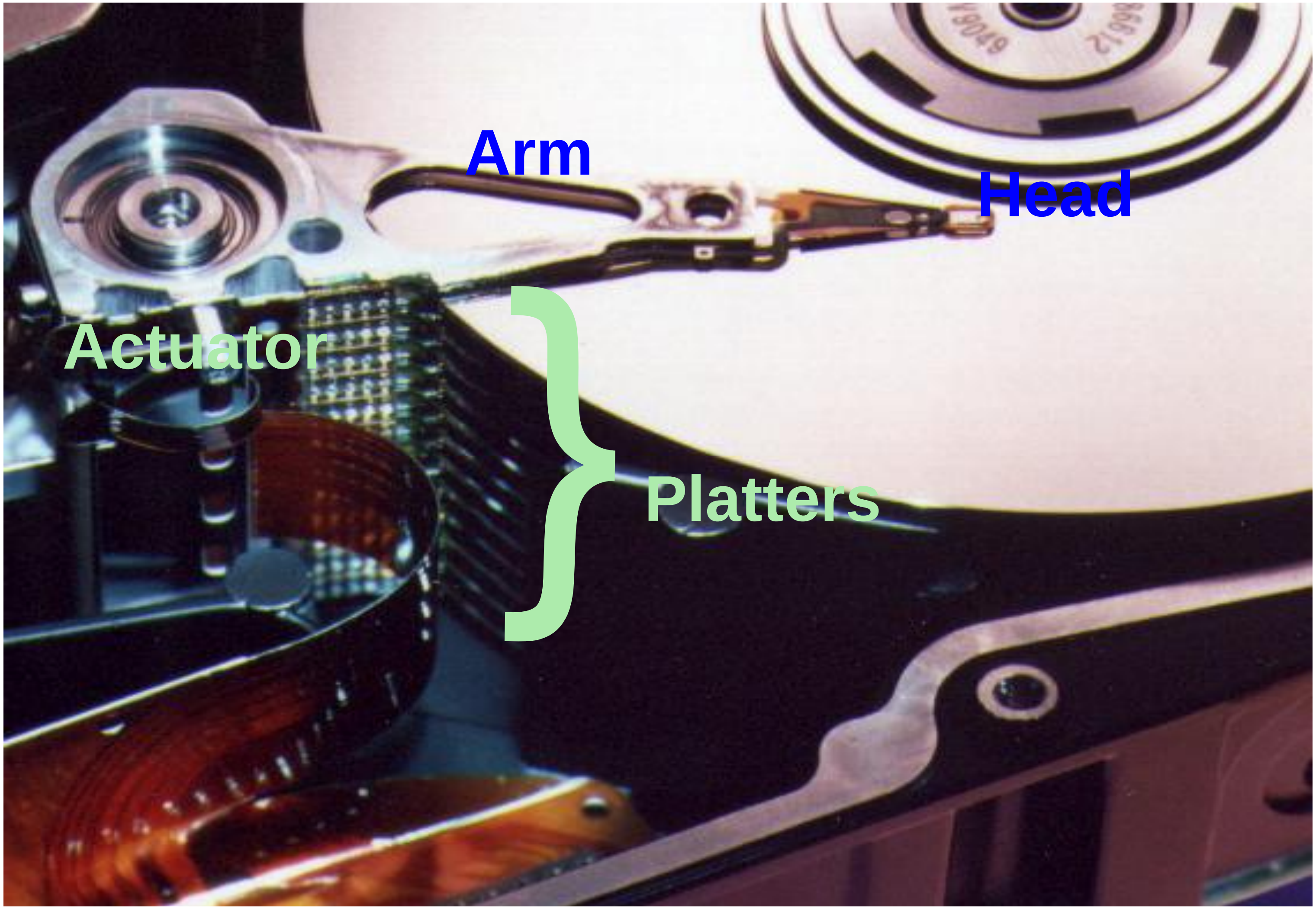


Fig. A disk with four platters.

Photo of Disk Head, Arm, Actuator



Memory Organization - CD ROM

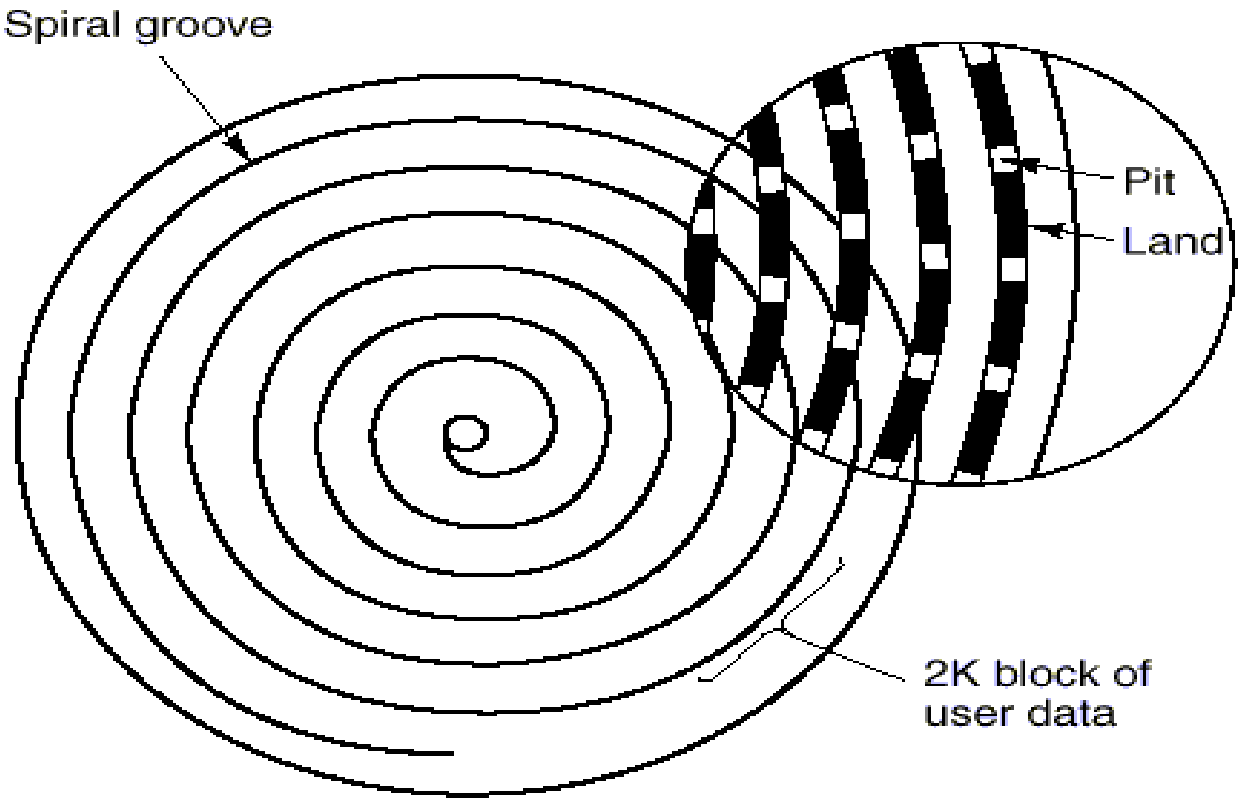


Fig. Recording structure of a compact-disc or CD ROM

Memory Organization - Magnetic Tape

2. Magnetic tape: The tape is a strip of plastic coated with a magnetic recording medium. Bits are recorded as magnetic spots on the tape.

Fuji film DDS-4 tapes

Cache Memory

Definitions

- *Dictionary meaning*: a. A hiding place used especially for storing provisions. ***b.*** A place for concealment and safekeeping, as of valuables. ***c.*** A store of goods or valuables concealed in a hiding place

- *Computer science meaning:* A fast storage buffer in the central processing unit of a computer.

Cache Memory (Contd.).

- Cache: High-speed speed memory, logically placed between CPU and main memory.
- Works on the locality of reference
- At a given instant of time, memory is accessed within a small neighborhood.
- This block of memory can be stored in a small high-speed memory rather than in normal RAM
- This high-speed memory is the *cache memory*

Cache Operation

- When the CPU needs to access memory, it checks the cache first.

- If the data is found in cache (*cache hit*), it is read.

- Else (*cache miss*) data is accessed from main memory.
  - A block of data from the required location is then transferred to the cache memory.

- Cache performance is measured in *hit ratio.*

Cache swapping

- Whenever cache miss occurs, a block of data must be transferred from main memory to the cache.
- A block of data already residing in cache should be written back to the main memory
- It requires a mapping algorithm to implement the above tasks.
- Moreover we need a *search* algorithm to check if the data is in cache
  - We use associative memory to search a cache memory.

Cache Memory Mapping

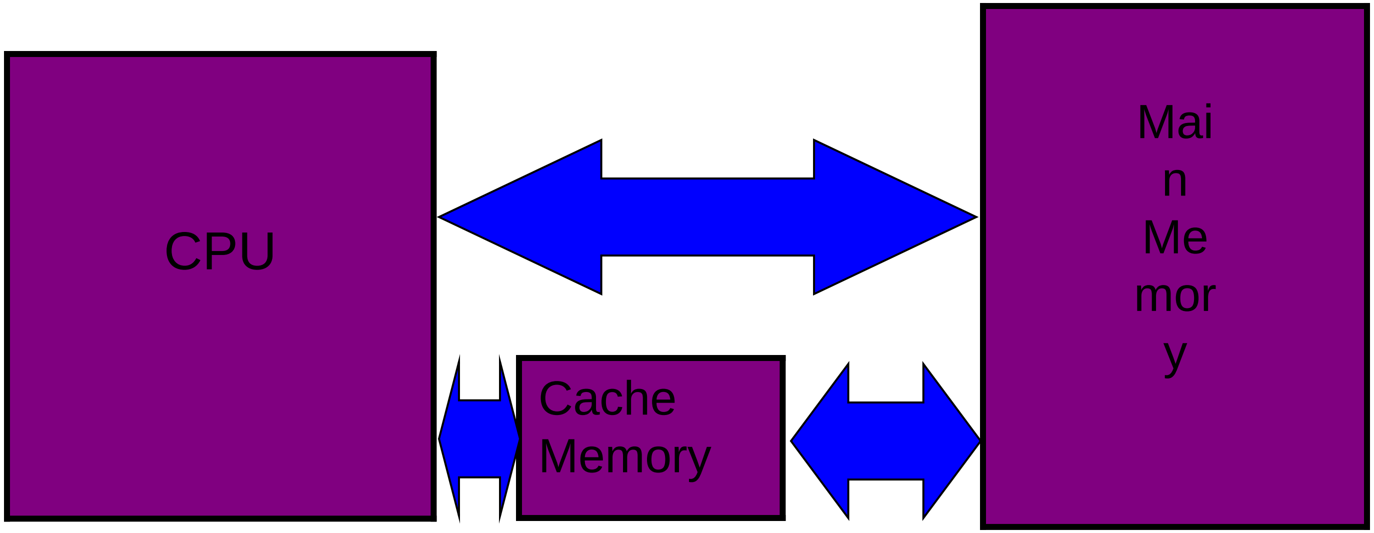
Mapping : The transformation of data from main memory to cache is referred as mapping process.

Three types of mapping:

1. Associative mapping.
2. Direct mapping.
3. Set-associative mapping.

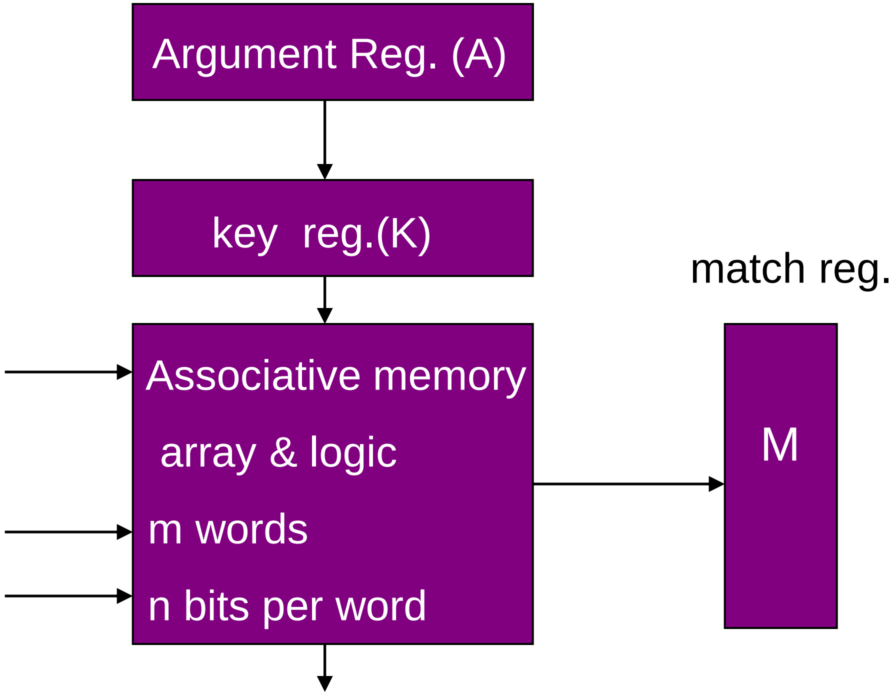
Cache Memory (Contd.).

- Example of cache memory. The cache memory is logically closer to CPU - implies faster access.



Cache Memory - Mapping

- Hardware part: Block diagram of associative memory used in Associative mapping



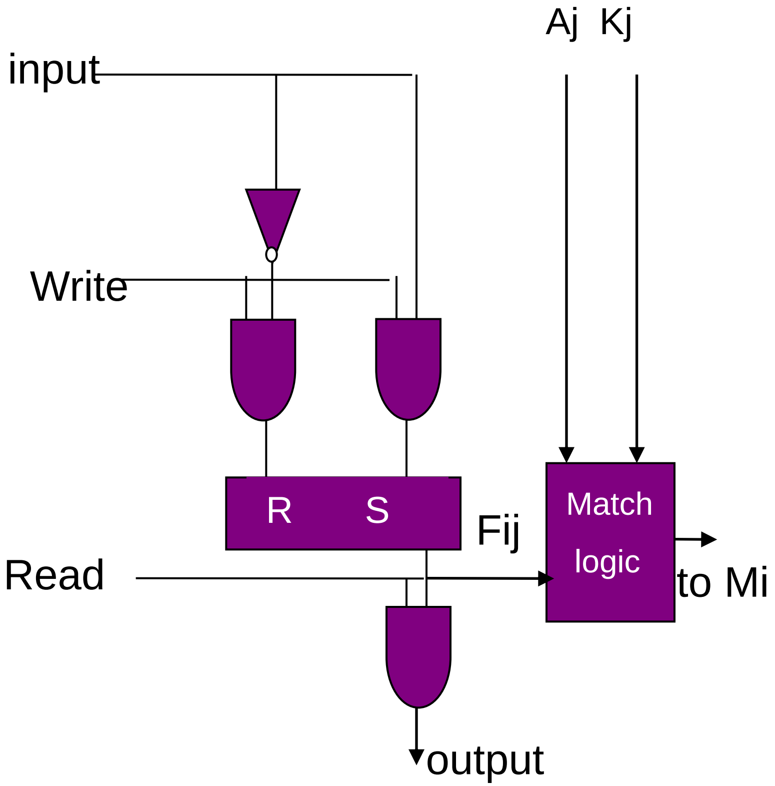


Fig. Match logic

Replacement algorithm

- Whenever a cache miss occurs, we have to replace them.
- We have two popular algorithms
- Least Recently Used (LRU)
- Least Frequently Used (LFU)

Writing to the Cache

- While writing in to the cache, both cache and main memory are to be updated.
- Two popular techniques are proposed
- *Write through*: Whenever there is write operation, main memory is also updated
- *Write back*: Write only when swapping is done.
  - This requires a special bit in cache to indicate the memory is changed
  - This special bit is called *dirty bit*

Flash Memory

- Latest of the re-programmable ROM family
  - Named so because of its rapid erase and write times
  - Typical erase and write times – 10ms

- Posses both bulk erase and sector erasing features
  - Doesn’t have the Byte-by-Byte re-programming feature

- Compromise between
  - EPROMs’ low cost and high densities
  - EEPROMs’ high re-programming speeds

Flash Memory (Contd.).

- Comes in two flavors
  - NOR flash : Similar in read - operation to SRAM
  - NAND flash: Read block wise
- NOR more popular
- Consists of segments
  - Segments need not be of uniform size

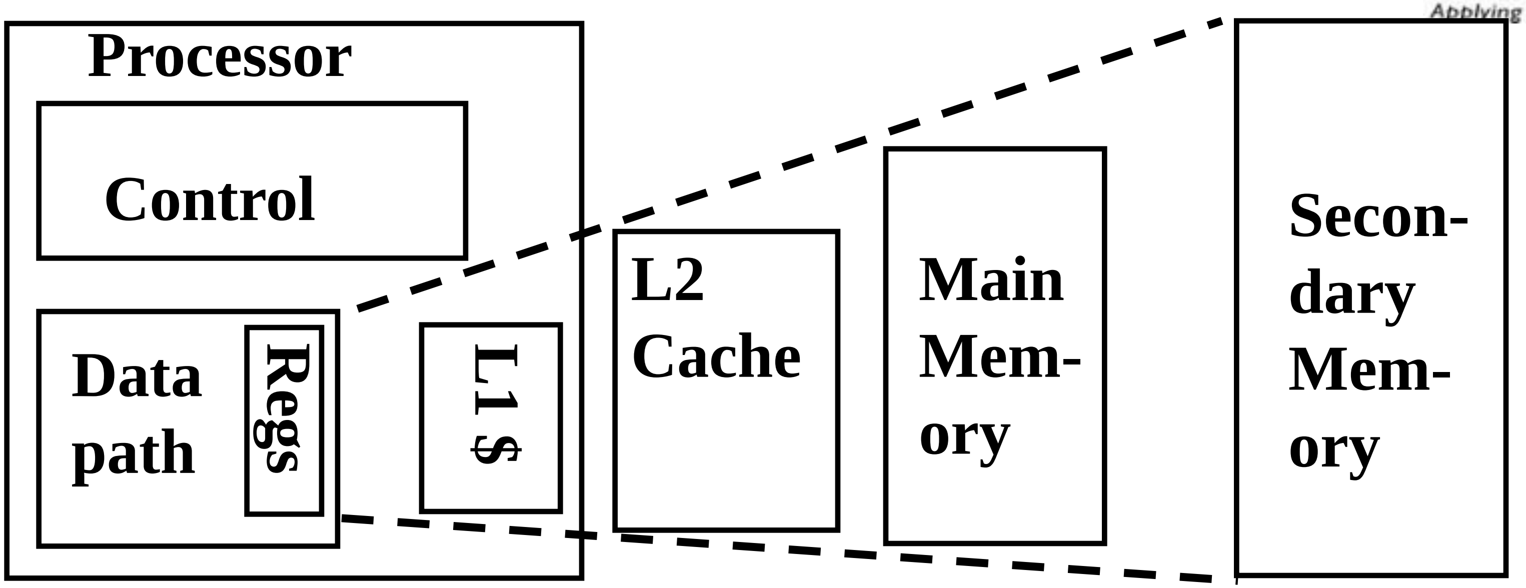
Flash – To be noted

Erase before write – otherwise ?

Nice about Flash

- Non Volatile
- Faster reads
- Lock
  - Protection
- High density

Current Memory Hierarchy



Speed(ns):	0.5ns	2ns	6ns	100ns	10,000, 00ns
Size (MB):	0.0005	0.05	1-4	100-1000	100,000
Cost($/MB):	--	$100	$30	$1	$0.05

Summary

In this module, we discussed:

- Memory types Memories and pyramid
- Random and sequential memories
- Magnetic. Optic (CDROM) and semiconductor memories
- ROMs and RAMs
- DRAMs and refreshing DRAMs
- Cache memories
  - Operations
  - Hit, miss, hit ratio
  - Mapping and associative memory
- Flash memories

Input Output Organization

Module 5

Objectives

At the end of this module, you will be able to:

- Discuss Peripherals, memory and IO mapping
- List Modes of Data transfer
- Know IO Processors

Duration: 1 hr

I/O Organization

Peripheral devices:

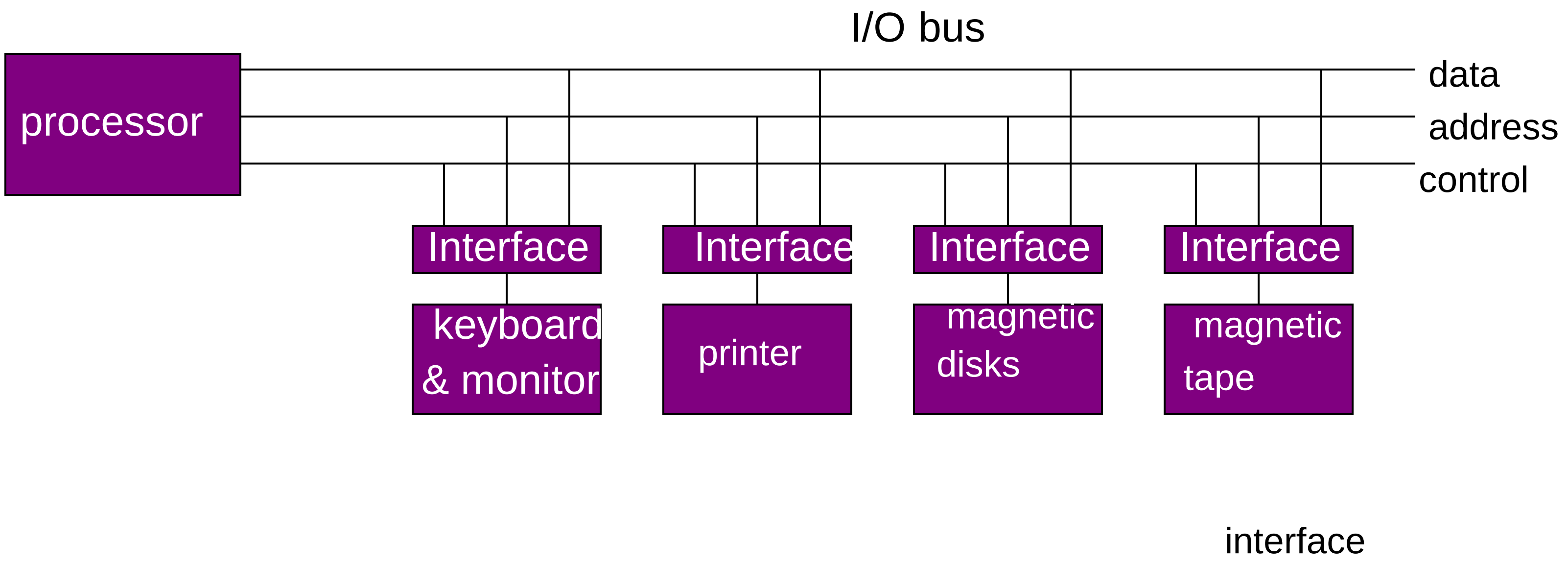
- I/O: The input-output subsystem of computer termed as I/O, provides efficient mode of communication between central system & outside environment.

- Peripherals: I/O devices attached to the computer are also called peripherals. Common peripherals are keyboards, display unit & printers.

I/O organization (Contd.).

- I/O Interface: Due to difference in signal, speed, mode & word format b/w peripherals & CPU , I/O interface are required.

- I/O Bus & Interface Modules:



I/O Organization (Contd.).

- **I/O command:** Processor selects and issues the function code to the interface through control lines.

  Selected interface responds to the function code and executes it. These codes are called I/O commands.

- Four types of I/O commands:

  1. Control command: It is issued to activate or initialize the I/O before any data transfers.

     Ex: To rewind the tape, to start the tape moving in the forward direction.

  2. Status: The command is used to test the various status conditions in the interface & peripheral.

     Ex: The OS might try to check the status of the peripheral before a transfer is initiated such that interface can recover the errors through status register.

I/O Organization (Contd.).

3. Output data: These command causes the interface to respond by transferring data from the bus into one of its registers.

   Ex: While sending data to tape, processor checks the correct position of the tape by status command & than processor issues a data output command.

4. Input data: It is opposite of the data output command.

Memory & I/O Mapping

- Memory Mapping
  - Assignment of addresses to memory registers in various memory chips in a system
  - To ensure only one memory device is activated during each cycle
    - What happens if more than one location is selected for the same address?
  - To transfer data to/from the device, uP does a memory write/read
  - Devices need direct access to the memory bus

I/O Mapping

IO Mapping

- The way IO is connected to the CPU

- Memory Mapped IO
  - Memory and IO Ports share the same set of addresses
  - CPU Communicates with IO device the same way as it does with external memory

- IO Mapped IO
  - IO has a separate address
  - CPU employs IO Instructions

Modes of Data Transfer

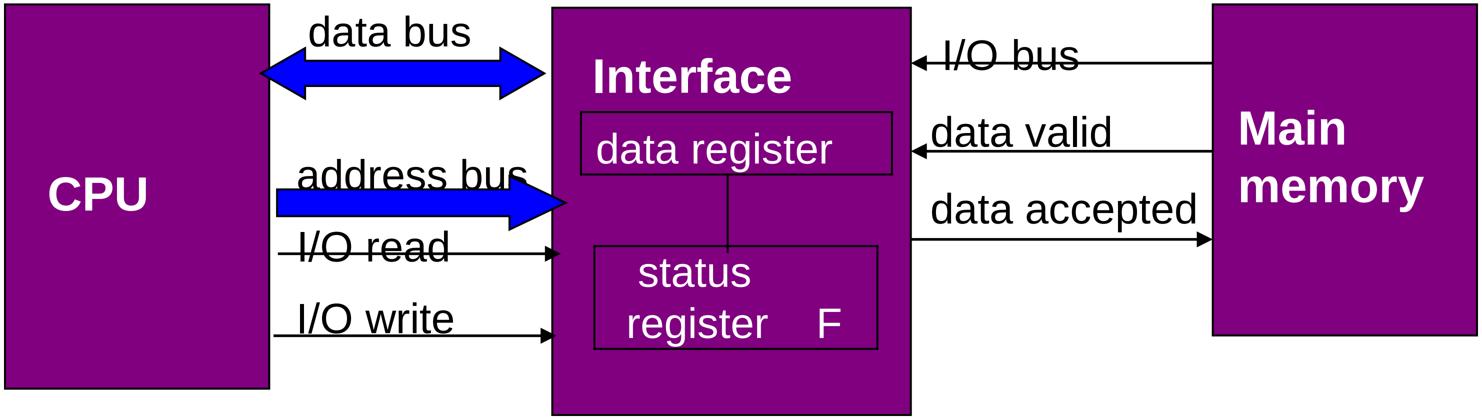
Data transfer to & from peripherals may be handled in one of the three possible modes.

1. Programmed I/O.
2. Interrupt-initiated I/O.
3. Direct Memory Access (DMA).

1. Programmed I/O:
   - Programmed I/O operations are the results of I/O instructions written in the computer program.
   - Each data item transfer is initiated by an instruction in the program.
   - Here the CPU stays in a program loop until the I/O unit indicate that it is ready for data transfer.
   - This is time consuming process since processor is busy needlessly.

Modes of Data Transfer (Contd.).

2. Interrupt- initiated I/O:
   - This uses interrupt facility and some special commands to inform the interface to issue an interrupt request signal when the data are available from the device.
   - Meanwhile, CPU can proceed to execute another program and the interface will be monitoring the device.



Data transfer from I/O devices to CPU.

Interrupts

- A signal informing a program that an event has occurred
- Normal Program execution temporarily suspended by
  - Some external signal
  - Special instruction in the program
- Calls a procedure which services the interrupt
- Execution returned to the interrupted program

Interrupts (Contd.).

Interrupt Sources

- Hardware Interrupts
  - Non Maskable Interrupt NMI input pin
  - Interrupt INTR input pin
  - Edge triggered/level triggered interrupts
  - Vectored/non-vectored interrupts
- Software Interrupts
  - Execution of the interrupt instruction
  - Conditional Interrupts e.g. Divide by Zero Interrupt.

Modes of Data Transfer

3. Direct Memory Access (DMA)

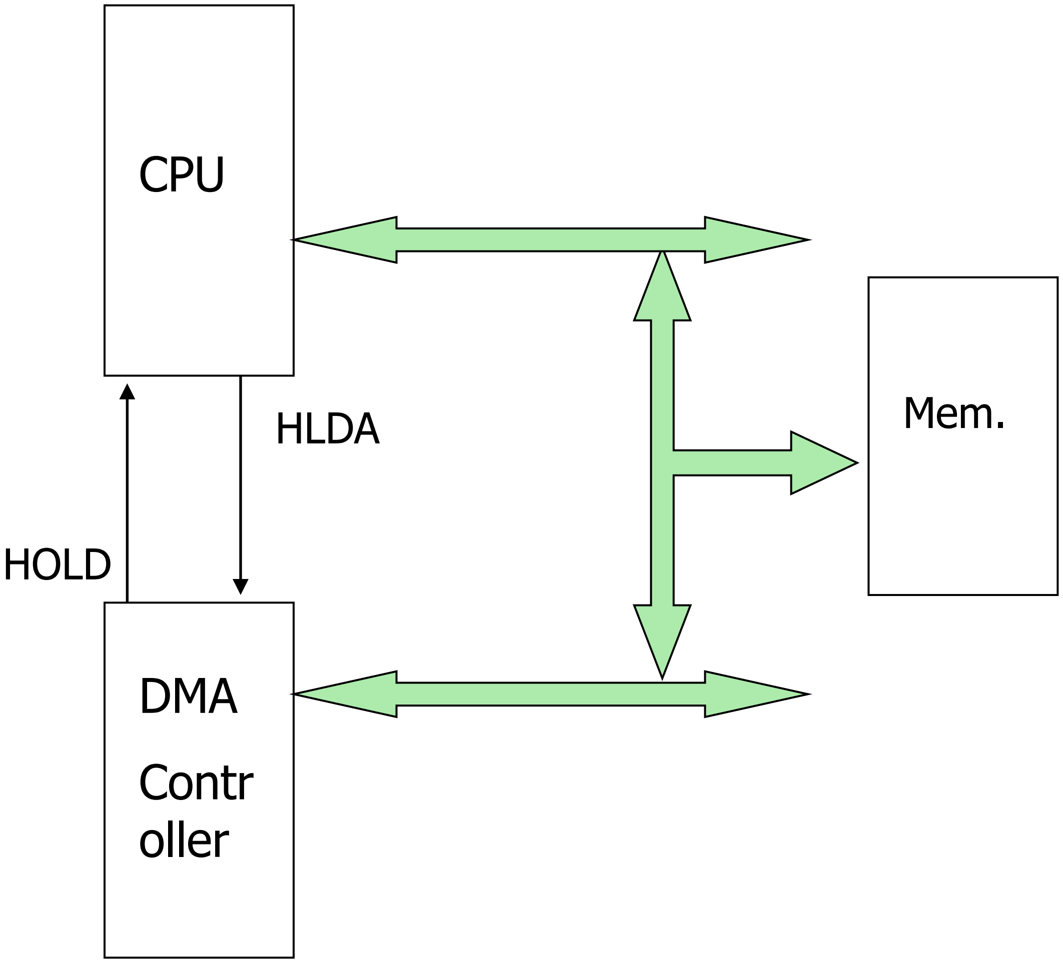
- The transfer of the data between a slow storage devices such as magnetic disks & main memory is often limited by the access time of these memory devices.
- Removing CPU from the path & allowing the peripheral device manage the memory buses directly will improve the speed of transfer and saves the CPU time.

This transfer technique is called Direct Memory Access (DMA)

During DMA transfer,

- CPU is idle & has no control of the memory buses.
- DMA controller manage the transfer directly between the I/O device & memory.

DMA Operations



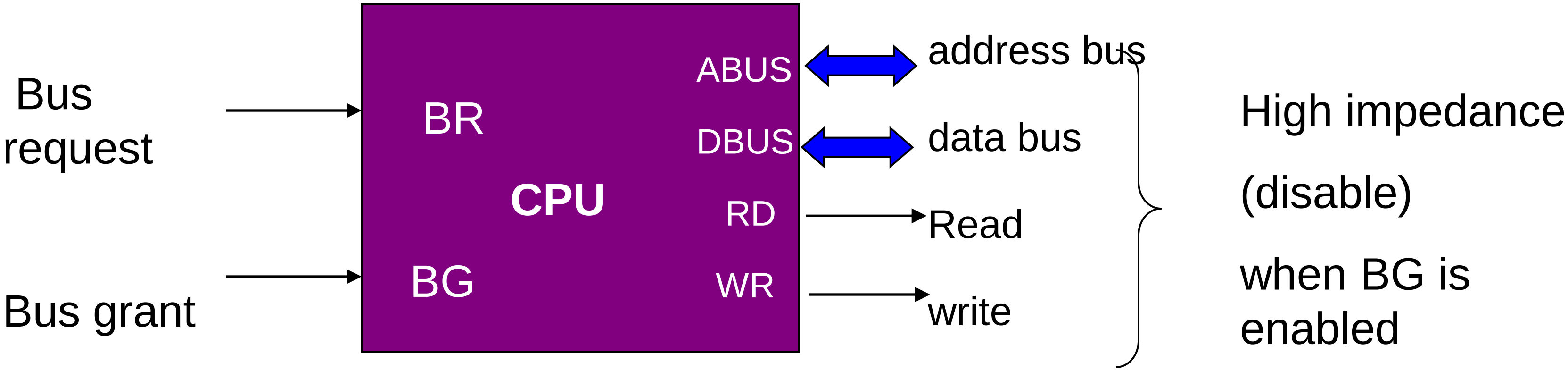
DMA Operations (Contd.).

- Two signals in the CPU that facilitate DMA transfer:
  1. Bus request: The bus request(BR) input is used by the DMA controller to request the CPU to relinquish control of the buses.
  2. Bus grant: The CPU activates the bus grant (BG) output to inform the external DMA that the buses are in the high impedance state. Then DMA controller can take the hold on memory buses without intervention of the CPU.

- Types of DMA Transfer:
  1. Burst transfer: It is one method of transfer, where a block sequence of number of words is transferred in a continuous burst while DMA controller is master over memory buses. Burst transfer is best suited for fast devices like disks.
  2. Cycle stealing : Another technique where DMA controller allowed to transfer one data word at a time after which it must return control of the buses to CPU.

DMA (Contd.).

- DMA Controller: Needs the usual circuits of an interface to communicate with the CPU & I/O device. In addition, it needs an address register, a word register & a set of Address lines.
- Address register & address lines are used for direct communication with the memory.
- The word count register specifies the number of words that must be transferred.



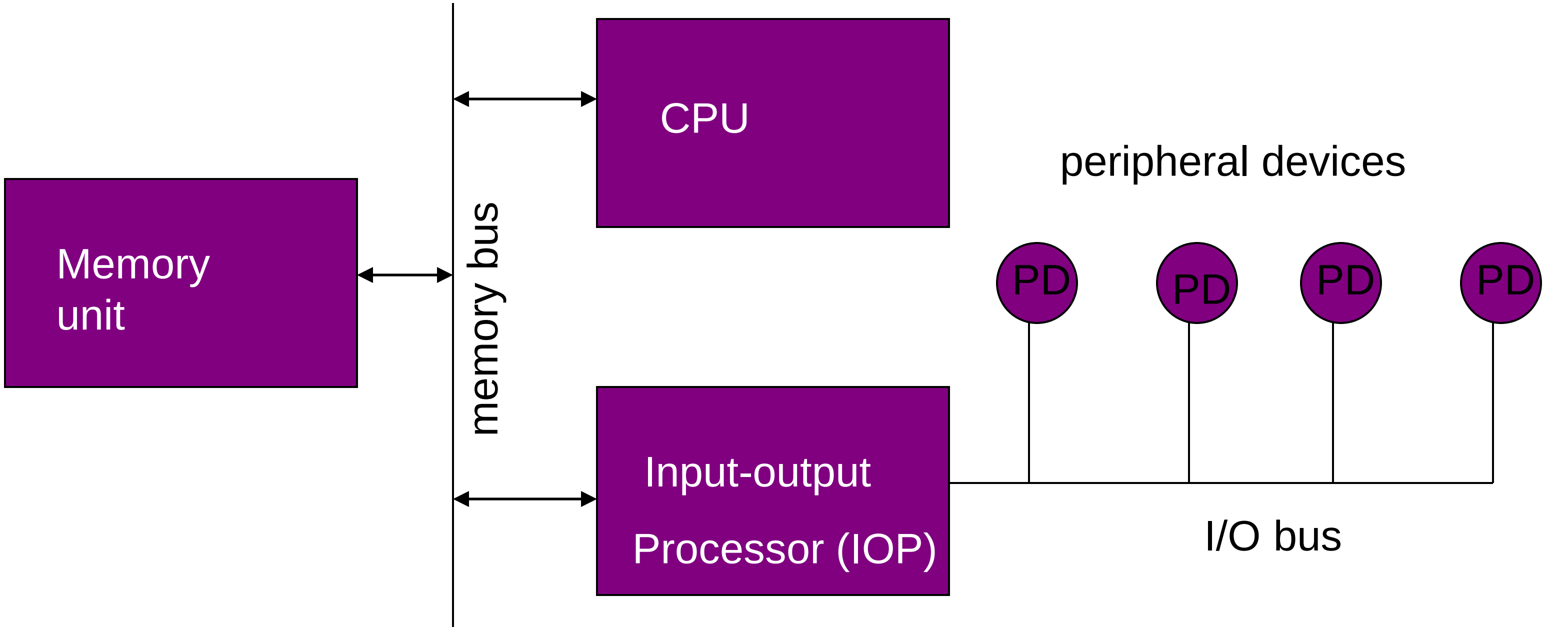
Input -Output Processor

- The computer may incorporate one or more external processors & assign them the tasks of communicating directly with I/O devices, instead of each interface communicate with the CPU.

- The processor that communicate with remote terminals over telephone & other media are called data communication processor (DCP).

Input -Output Processor (Contd.).

- Block diagram of a computer with IOP.



Input -Output Processor (Contd.).

- Commands :
  - Instructions that are read from memory by an IOP are called commands, to distinguish them from instructions that are read by the CPU.
- CPU-IOP Communication:

Memory unit acts as a message center where each processor leaves information for the other.

Summary

In this module, we discussed:

- Peripherals, memory and IO mapping
- Modes of Data transfer
  - Programmed IO
  - Interrupt driven
    - Types of interrupts
  - DMA
- IO Processors

Types of Computers

Module 6

Objectives

At the end of this module, you will be able to:

- List types of Computer
- Discuss IBM's AS-400

Duration: 1 hr

Types of Computers

Computers can be generally classified by size and power as follows (though there is considerable overlap):

- Personal computer: A small, single-user computer based on a microprocessor.
- Workstation: A powerful, single/multi user computer. A workstation is like a personal computer, but it has a more powerful microprocessor and, in general, a higher-quality monitor.
- Minicomputer: A multi-user computer capable of supporting up to hundreds of users simultaneously.
- Mainframe: A powerful multi-user computer capable of supporting many hundreds or thousands of users simultaneously.
- Supercomputer: An extremely fast computer that can perform hundreds of millions of instructions per second.

Supercomputer

- Supercomputer, a broad term referring to the fastest computers currently available.

- They are very expensive.

- Employed for specialized applications requiring immense amounts of mathematical calculations (number crunching).

  Ex: Weather forecasting, scientific simulations, (animated) graphics, fluid dynamic calculations, nuclear energy research, electronic design, and analysis of geological data (e.g. in petrochemical prospecting).

- In India the C-DAC and SERC at IISc., Bangalore have developed many versions of Supercomputer, latest being the PARAM 10000.

Mainframe

- Mainframe was a term originally referring to the cabinet containing the central processor unit or "main frame" of a room-filling Stone Age batch machine.
- Now-a-days a Mainframe is a very large and expensive computer capable of supporting hundreds, or even thousands, of users simultaneously.
- Difference between a supercomputer and a mainframe is:
  - supercomputer channels all its power into executing a few programs as fast as possible.
  - Mainframe uses its power to execute many programs concurrently. Speed is not the major criteria

Minicomputer

- It is a midsize computer.
- A minicomputer is a multiprocessing system capable of supporting up to 200 users simultaneously.
- The distinction between large minicomputers and small mainframes has blurred, however, as has the distinction between small minicomputers and workstations

Workstations

- A type of computer used for applications that require a moderate amount of computing power and relatively high quality graphics capabilities.
- Generally comes with a large, high-resolution graphics screen, large amount of RAM, built-in network support, and a graphical user interface.
- Most have a mass storage device such as a disk drive. but a special type of workstation, called a diskless workstation, comes without a disk drive.
- Commonly used OSs are UNIX (or UNIX flavors) and Windows NT.
- Workstations are typically linked together to form a local-area network, although they can also be used as stand-alone systems.

Personal Computer

- A small, relatively inexpensive computer designed for an individual user.
- All are based on the microprocessor technology that enables manufacturers to put an entire CPU in one chip.
- In businesses they are mainly used for word processing, accounting, desktop publishing, and for running spreadsheet and database management applications.
- At home, the most popular use for personal computers is for playing games and for surfing the Internet.

Notebook Computer

- An extremely lightweight personal computer, typically weigh less than 6 pounds and are small enough to fit easily in a briefcase.
- Uses a variety of techniques, known as flat-panel technologies, to produce a lightweight and non-bulky display screen.
- Modern notebook computers are nearly equivalent to desktop computers in computing power. They have the same CPUs, memory capacity, and disk drives.
- They are expensive. Notebook computers cost about twice as much as equivalent regular-sized computers.
- Notebook computers come with battery packs that enable you to run them without plugging them in. Batteries need to be recharged every few hours.

PDA

- Short for Personal Digital Assistant, a handheld device that combines computing, telephone/fax, and networking features.
  - A typical PDA can function as a cellular phone, fax sender, and personal organizer.

- Most PDAs are pen-based, using a stylus rather than a keyboard for input, incorporating handwriting recognition features.
  - Some also react to voice input by using voice recognition technologies.

- The field of PDA was pioneered by Apple Computer, with first PDA, the Newton MessagePad in 1993.
  - Modest success in the marketplace, due to their high price and limited applications. However, they may eventually become common gadgets.
  - PDAs are also called palmtops, hand-held computers and pocket computers.

IBM AS-400

- http://en.wikipedia.org/wiki/IBM_System_i

Summary

In this module, we discussed:

- Types of Computer:
  - PC, Workstation, Mini-computer, main frame, Super computer
  - Note book computers, PDAs
- IBM AS-400

Bus Architectures

Module 7

Objectives

At the end of this module, you will be able to:

- Discuss Bus - address, data and control buses
- List different Bus architecture

Duration: 1 hr

Bus

- Definition -- Signals and protocol
- Use -- Co-exist
- Example -- ISA, PCI, VME, RS-232, USB, Firewire, Ethernet
- Bus considerations
- Basic Signals

Bus (Contd.).

- Bus is a set of (common) wires that interconnect components in a computer system.
- All the devices *share* the bus.
- *Address bus* carries address information from the CPU to the device
- *Data bus* carries data between CPU and the device
- *Control bus* carries control information between CPU and device.

Bus (Contd.).

- Address and Data bus are standardized. Variations occur in Control bus.

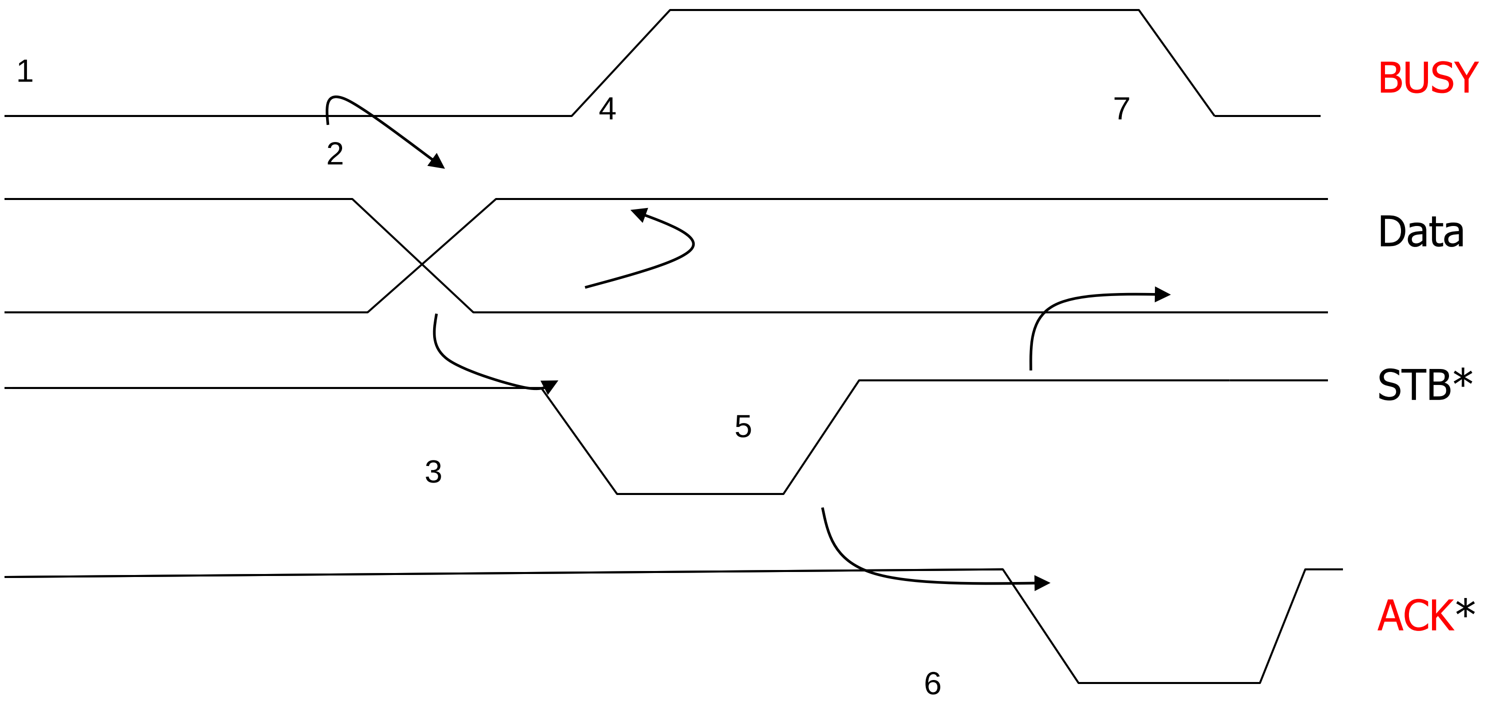
- Control bus should have the signals like
  - Read, write, Select
  - Request and grant
  - Interrupts
  - Timing synchronization signals.
  - Clock information.
- Data transfer over a bus can either be synchronous or asynchronous

Centronics

Used for parallel printers

- 8 data bits
- 3 control signals
- BUSY indicates the printer is busy and can't accept further data
- STROBE indicates the CPU has put a byte on the data bus
- ACKnowledge indicates printer has accepted the byte on the data bus.

Centronics Bus



Serial Data Communication: RS-232

- RS-232C is a standard interface
  - for serial transmission of data between computers and other devices
  - Describes both the physical interface and the transmission protocol.

- Standards and interfaces usually restrict RS-232 to <=20kbps and line lengths of < =15m.
  - However, in practice, RS-232 is far more robust than the traditional specified limits of 20kbps over a 15m line would imply.

- Successors to RS-232 : RS-422 and RS-423.
  - backward compatible so that RS-232 devices can connect to an RS-422 port.

USB

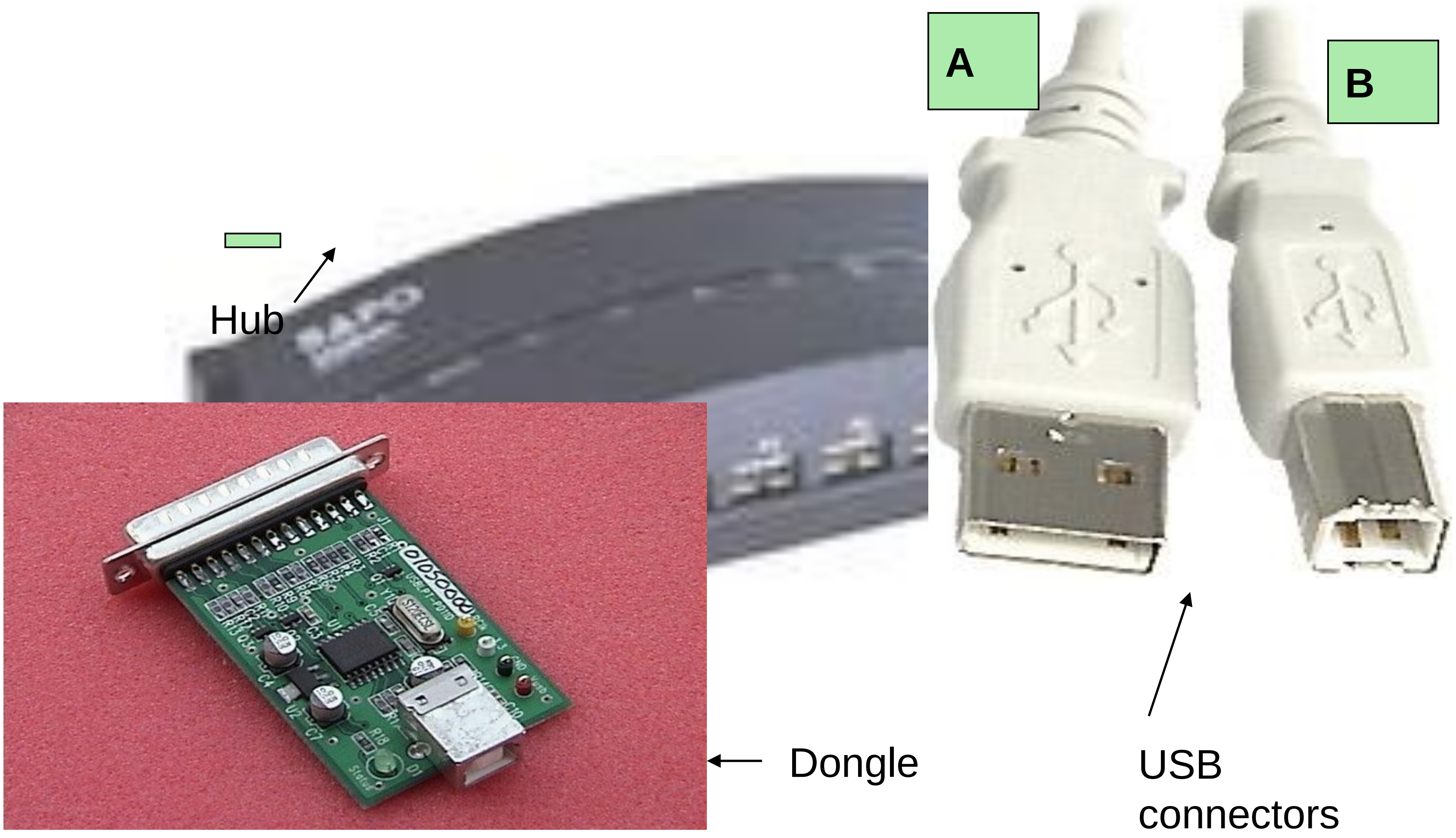
- Up to 127 devices
- Data transfer speed 12Mbps
- Cable length 5 meters.
- Bi-phase NRZI (Non-Return to Zero Inverted) data
- Bit-stuffing sync

Why USB ?

- Low Cost
  - Low-cost cables and connecters
  - Optimized for integration in device and host hardware
  - Low-cost sub channel at 1.5 Mb/Sec

- Ease of use
  - Single model for cable and connectors
  - Hot attach and Detach support
  - Dynamically attachable and configurable peripherals

USB Device View



A

B

Hub

Dongle

USB connectors

USB Architecture

USB Device Types :

- Host
  - The master device in a USB system
  - Initiates all data transfer
  - Manages device attach / disconnect
- Device or Function
  - The slave device in a USB system
  - Respond only to request from the host
  - Cannot initiate data transfers
- Hub
  - Intermediary between Host and Device

USB Host

- Host PC is in charge of bus.
- The host has to know that devices are on the bus and capabilities of each.
- On power up hubs make the host aware of all the devices attached by the system
- The host assigns a unique USB address to the device and then determines if the newly attached USB device is a Hub a or a function .
- When USB device has been removed from the hub ports, the hub disable the port and provide an indication of device removal to the host.

USB Host (Contd.).

- The host manages the flow of data on the bus.
- Multiple peripherals may want to transmit data at the same time.
- The host controller handles this by dividing the data path into 1 millisecond frames and giving each transmission as a portion of the frame.
- Host also has error checking duties and adds error checking bits during data transfer

USB Protocol

USB is polled bus.

- For each transaction, host controller sends USB packet on schedule basis. The packet which describes type and direction is called *token packet*. USB device selects itself by decoding its address.

- In a given transaction, data is transferred from host to a device or from a device to host

- The source of transaction then sends a data packet or indicates it has no data to transfer. Destination responds with hand shake.
  - USB data transfer model between source and destination (on host & endpoint)is referred to as a pipe.

SCSI

- Small Computer System Interface.
- A common parallel bus for a number of devices.
- Each device is addressed.
- Most of the SCSI devices are plug and play
- Data length up to 16 bits and data transfer rates up to 160 MBps.

PCI

- PCI is an extension of the earlier EISA bus
- Support plug and Play (PnP).
- Can support multiple bus and protocol conversion
- Supports bus masters (switches)

PCI Bus

- Basic terminology
  - Peripheral Component Interconnect
  - Bussed signals / Point-to-Point signals
  - Signal types
  - Turn-around cycle
  - Decoding
    - Positive
    - Subtractive

- Master / Slave
  - Arbitration
- Bridges
  - Host to PCI
  - PCI to PCI

PCI Connector Mother board & add-on card

PCI Address Space

- Memory Space
- I/O Space
- Configuration Space
  - Pre-defined Header Region
  - Device Dependent Region

Configuration Space Header

Device ID	Vendor ID	00h
Status	Command	04h
Class code	Rev ID	08h
Predefined		0Ch
Base Address Registers(6)		10h
Reserved		
Expansion ROM Base Adc		30h
Reserved/Predefined		3Ch

PCI Commands

- Memory Read
- Memory Read line
- Memory Read Multiple
- Memory Write
- Memory Write and Invalidate
- Dual Address Cycle
- I/O Read
- I/O Write
- Configuration Read
- Configuration Write
- Special Cycle
- Interrupt Acknowledge

Bus Arbitration

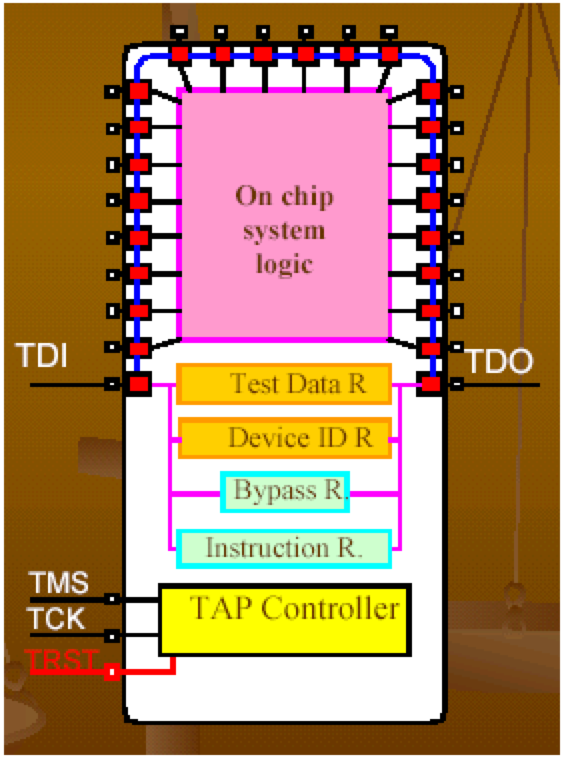
- More than one device may request for the bus – *bus contention*

- A master has to resolve these requests and grant to one of the device at a time.

- This is known as *Bus Arbitration.*

JTAG

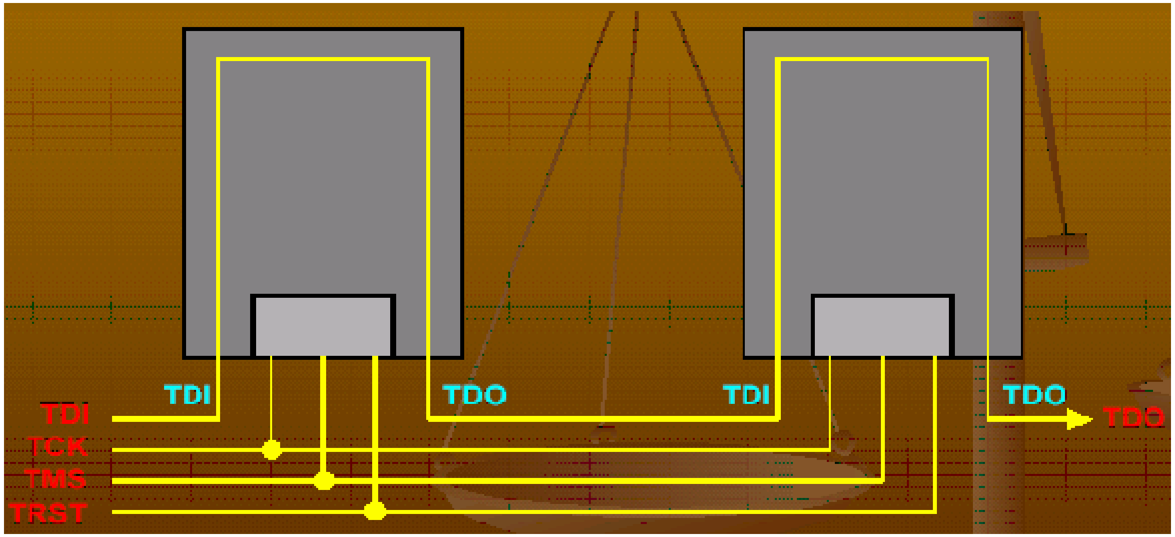
- Acronym for "Joint Test Action Group"
- A methodology for testing ICs
- Methodology specified by IEEE standard 1149.1



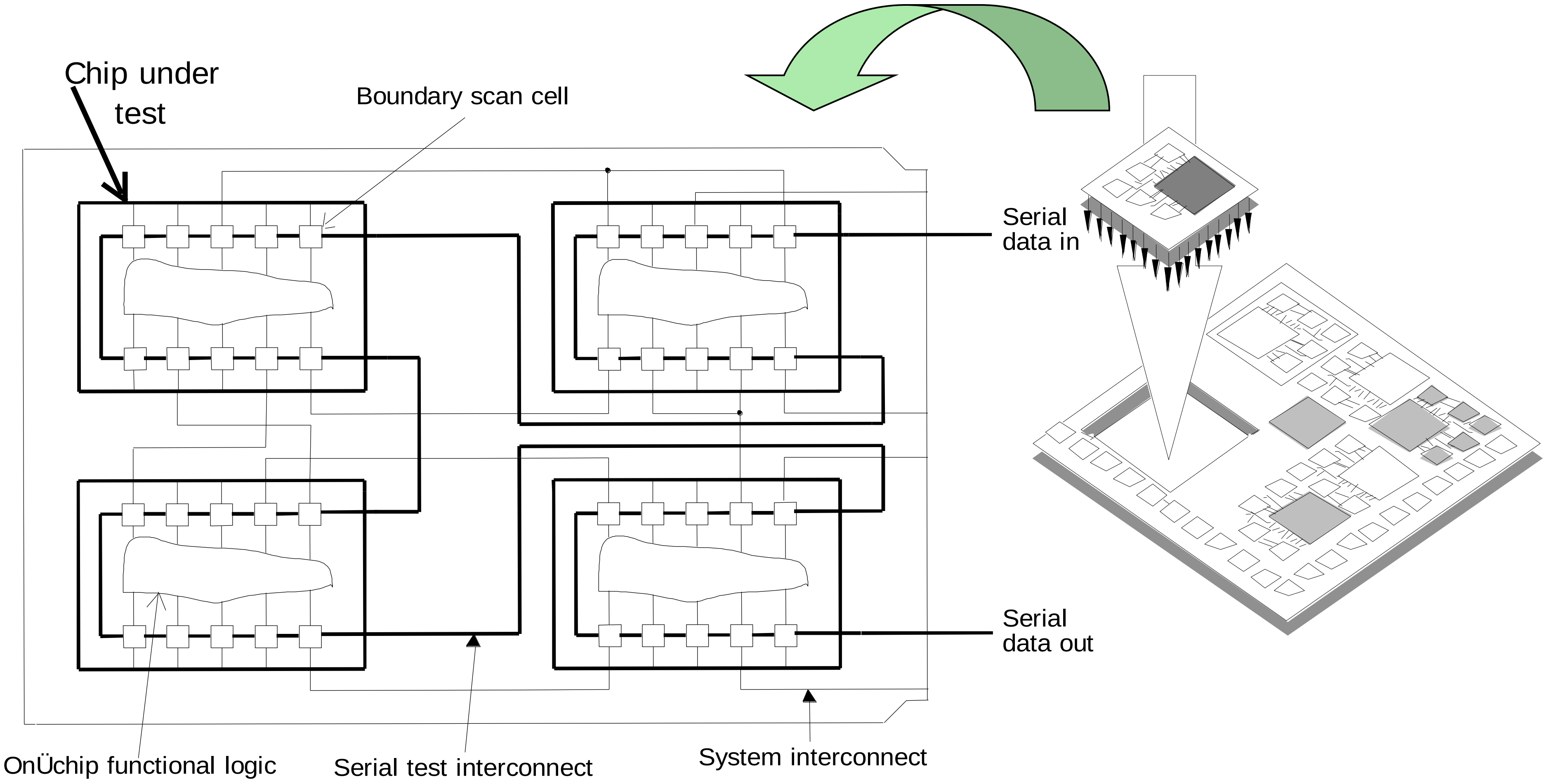
JTAG Signals/ Pins

Four or five signals dedicated to test

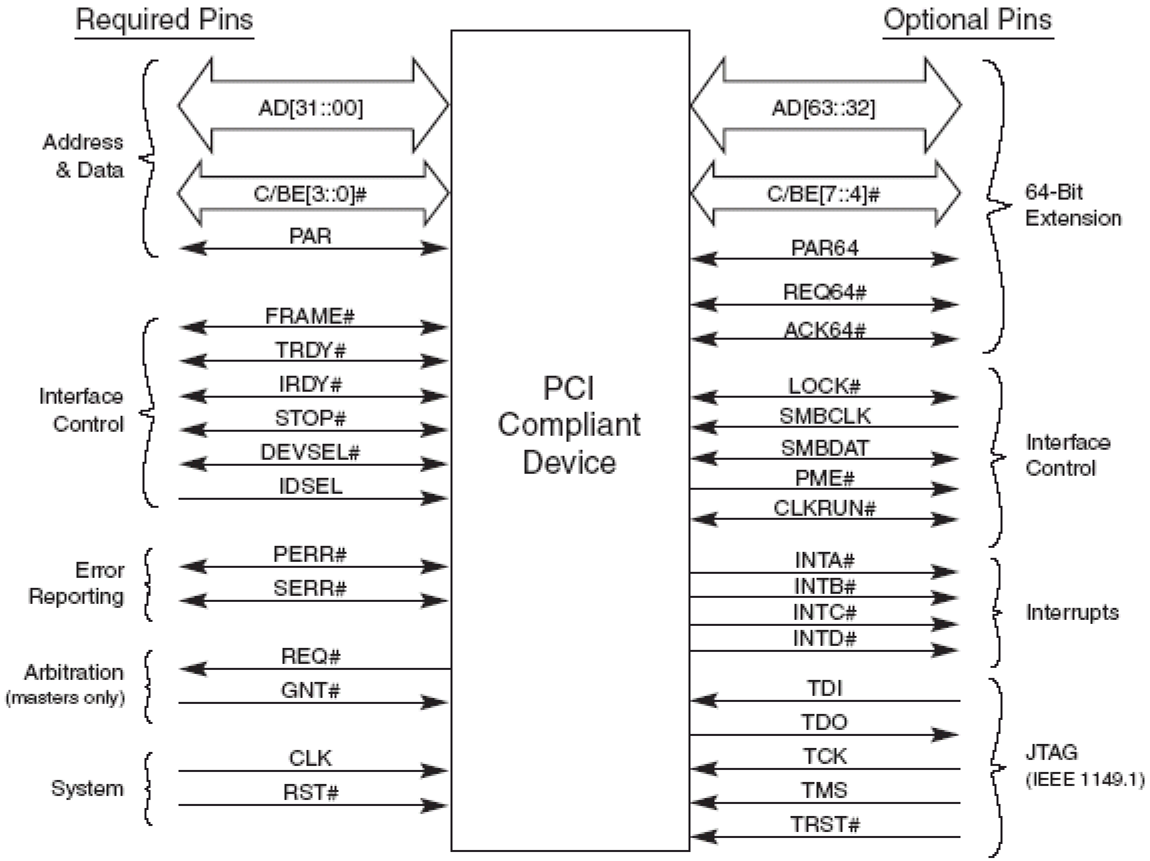
- Data signals daisy chained (TDI/TDO)
- Control Signals are shared (TMS/TCK/TRST)



How Does it Work



PCI Signals


Required Pins
Optional Pins
Address & Data
AD[31::00]
C/BE[3::0]#
PAR
Interface Control
FRAME#
TRDY#
IRDY#
STOP#
DEVSEL#
IDSEL
Error Reporting
PERR#
SERR#
Arbitration (masters only)
REQ#
GNT#
System
CLK
RST#
PCI Compliant Device
AD[63::32]
C/BE[7::4]#
PAR64
REQ64#
ACK64#
64-Bit Extension
LOCK#
SMBCLK
SMBDAT
PME#
CLKRUN#
Interface Control
INTA#
INTB#
INTC#
INTD#
Interrupts
TDI
TDO
TCK
TMS
TRST#
JTAG (IEEE 1149.1)

Summary

In this module, we discussed:

- Bus - address, data and control buses
- Centronics bus and timing diagram
- Serial bus – RS232 and SCSI
- USB
  - Host, client and hub
- PCI bus
  - Configuration space address
    - Programmability and PnP
  - Commands, protocol fundamentals
  - Arbitration
- JTAG

References

1. Carl Hamacher, Hill V and Vranesic, Zvonko G. *Computer Organization*, Ed 4. New Delhi: McGraw Hill, 2001.
2. Mano, Morris, M. *Computer System Architecture*. Ed 3. New Delhi: Prentice-Hall India, 1992.

Thank You