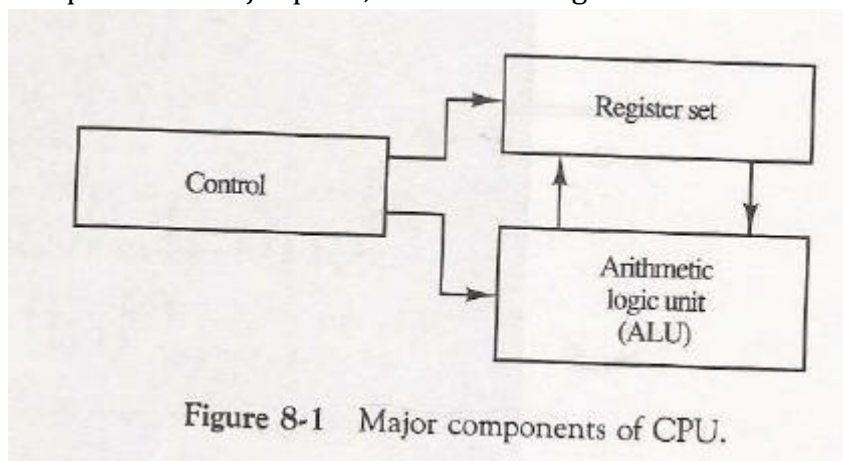


Unit V:

- Central Processing Unit
 - General Register Organization
 - Stack Organization
 - Addressing Modes
 - Data transfer and Manipulation
 - Program Control
 - CISC and RISC
- Parallel Processing
 - Pipeline
 - General Consideration
- Input – Output Organization
 - Peripheral devices
 - I/O Interface
- Memory Organization
 - Memory Hierarchy
 - Main Memory
 - Auxiliary Memory

Central processing unit**Introduction:**

- The main part of the computer that performs the bulk of data-processing operations is called the central processing unit and is referred to as the CPU.
- The CPU is made up of three major parts, as shown in Fig. 8-1.



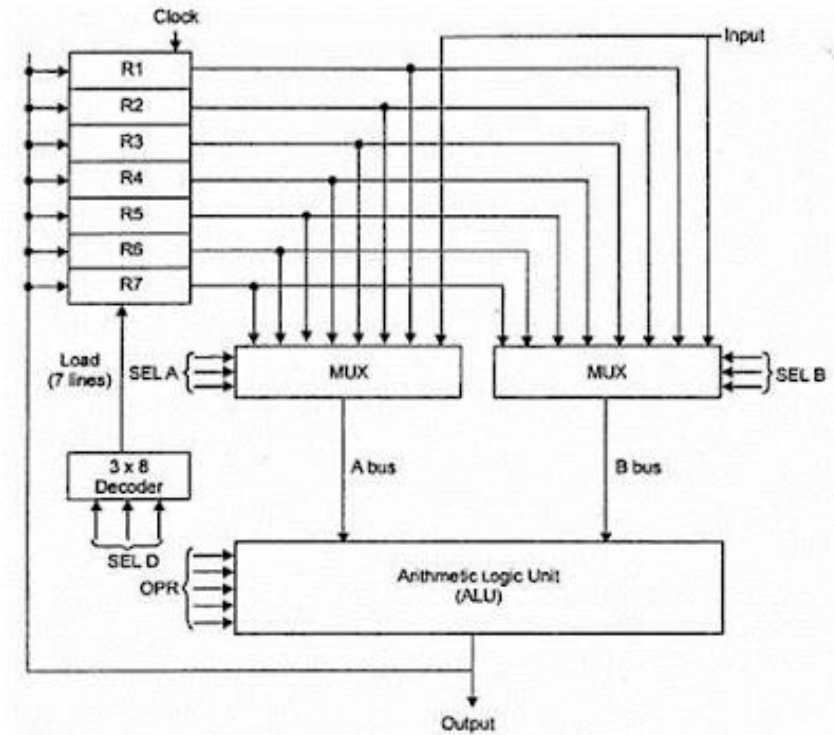
- The register set stores intermediate data used during the execution of the instructions.
- The arithmetic logic unit (ALU) performs the required micro operations for executing the instructions.
- The control unit supervises the transfer of information among the registers and instructs the ALU as to which operation to perform.

General Register organization

- Generally CPU has seven general registers. Register organization show how registers are selected and how data flow between register and ALU. A decoder is used to select a particular register.

- The output of each register is connected to two multiplexers to form the two buses A and B. The selection lines in each multiplexer select the input data for the particular bus.
- The A and B buses form the two inputs of an ALU. The operation select lines decide the micro operation to be performed by ALU. The result of the micro operation is available at the output bus. The output bus connected to the inputs of all registers, thus by selecting a destination register it is possible to store the result in it.

A bus organization for seven CPU register:



EXAMPLE:

- To perform the operation $R3 = R1 + R2$ We have to provide following binary selection variable to the select inputs.
 1. **SEL A : 001** -To place the contents of R1 into bus A.
 2. **SEL B : 010** - to place the contents of R2 into bus B
 3. **SEL OPR : 10010** - to perform the arithmetic addition A+B
 4. **SEL REG or SEL D : 011** - to place the result available on output bus in R3.

Register and multiplexer input selection code

Binary code	SELA	SELB	SEL-D or SEL-REG
000	Input	Input	---
001	R1	R1	R1
010	R2	R2	R2
011	R3	R3	R3
100	R4	R4	R4

101	R5	R5	R5
110	R6	R6	R6
111	R7	R7	R7

Operation with symbol

Operation code	selection	Operation	symbol
0000		Transfer A	TSFA
0001		Increment A	INC A
0010		A+B	ADD
0011		A-B	SUB
0100		Decrement A	DEC
0101		A AND B	AND
0110		A OR B	OR
0111		A XOR B	XOR
1000		Complement A	COMA
1001		Shift right A	SHR
1010		Shift left A	SHL

Control word:

- The combined value of a binary selection inputs specifies the control word.
- It consists of four fields SELA, SELB and SELD or SELREG contains three bit each and SELOPR field contains four bits thus the total bits in the control word are 13-bits.

SELA	SELB	SELREG OR SELD	SELOPR
-------------	-------------	-----------------------	---------------

Format of control word:

- The three bit of SELA select a source registers of the a input of the ALU.
- The three bits of SELB select a source registers of the b input of the ALU.
- The three bits of SELED or SELREG select a destination register using the decoder.
- The four bits of SELOPR select the operation to be performed by ALU.

CONTROL WORD FOR OPERATION $R2 = R1+R3$:

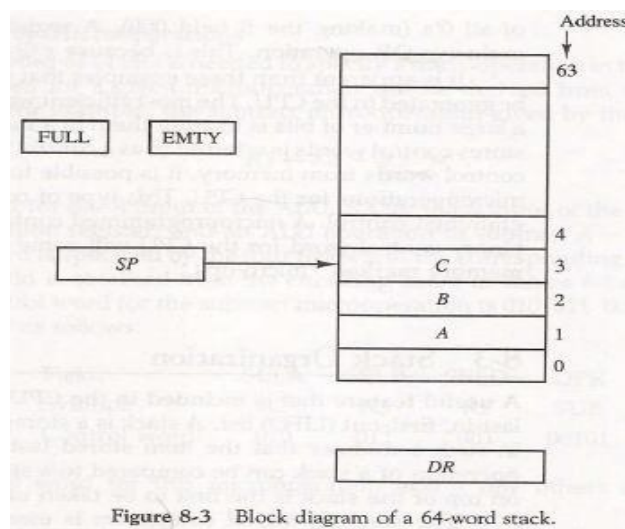
SEL A	SEL B	SEL- D OR SEL- REG	SEL OPR
001	011	010	0010

MICROOPERATION	SEL A	SEL B	SEL D OR SEL REG	SELOPR	CONTROL WORD			
$R2 = R1+R3$	R1	R3	R2	ADD	001	011	010	0010

- A stack or last-in first-out (LIFO) is useful feature that is included in the CPU of most computers.
- Stack:
 - A stack is a storage device that stores information in such a manner that the item stored last is the first item retrieved.
- The operation of a stack can be compared to a stack of trays. The last tray placed on top of the stack is the first to be taken off.
- In the computer stack is a memory unit with an address register that can count the address only.
- The register that holds the address for the stack is called a stack pointer (SP). It always points at the top item in the stack.
- The two operations that are performed on stack are the insertion and deletion.
- The operation of insertion is called *PUSH*.
- The operation of deletion is called *POP*.
- These operations are simulated by incrementing and decrementing the stack pointer register (SP).

Register Stack:

- A stack can be placed in a portion of a large memory or it can be organized as a collection of a finite number of memory words or registers.



- The below figure shows the organization of a 64-word register stack.
- The stack pointer register SP contains a binary number whose value is equal to the address of the word is currently on top of the stack. Three items are placed in the stack: A, B, C, in that order.
- In above figure C is on top of the stack so that the content of SP is 3.
- For removing the top item, the stack is popped by reading the memory word at address 3 and decrementing the content of stack SP.
- Now the top of the stack is B, so that the content of SP is 2.
- Similarly for inserting the new item, the stack is pushed by incrementing SP and writing a word in the next- higher location in the stack.
- In a 64-word stack, the stack pointer contains 6 bits because $2^6 = 64$.
- Since SP has only six bits, it cannot exceed a number greater than 63 (111111 in binary).
- When 63 is incremented by 1, the result is 0 since $111111 + 1 = 1000000$ in binary, but SP can accommodate only the six least significant bits.

- Then the one-bit register FULL is set to 1, when the stack is full.
- Similarly when 000000 is decremented by 1, the result is 111111, and then the one-bit register EMTY is set 1 when the stack is empty of items.
- DR is the data register that holds the binary data to be written into or read out of the stack.

PUSH:

- Initially, SP is cleared to 0, EMTY is set to 1, and FULL is cleared to 0, so that SP points to the word at address 0 and the stack is marked empty and not full.
- If the stack is not full (if FULL = 0), a new item is inserted with a push operation.
- The push operation is implemented with the following sequence of microoperations:

$SP \leftarrow SP + 1$	Increment stack pointer
$M[SP] \leftarrow DR$	Write item on top of the stack
If ($SP = 0$) then ($FULL \leftarrow 1$)	Check if stack is full
$EMTY \leftarrow 0$	Mark the stack not empty

- The stack pointer is incremented so that it points to the address of next-higher word.
- A memory write operation inserts the word from DR the top of the stack.
- The first item stored in the stack is at address 1.
- The last item is stored at address 0.
- If SP reaches 0, the stack is full of items, so FULL is to 1.
- This condition is reached if the top item prior to the last push way location 63 and, after incrementing SP, the last item is stored in location 0.
- Once an item is stored in location 0, there are no more empty registers in the stack, so the EMTY is cleared to 0.

POP:

- A new item is deleted from the stack if the stack is not empty (if EMTY = 0).
- The pop operation consists of the following sequence of min operations:

$DR \leftarrow M[SP]$	Read item from the top of stack
$SP \leftarrow SP - 1$	Decrement stack pointer
If ($SP = 0$) then ($EMTY \leftarrow 1$)	Check if stack is empty
$FULL \leftarrow 0$	Mark the stack not full

- The top item is read from the stack into DR.
- The stack pointer is then decremented. If its value reaches zero, the stack is empty, so EMTY is set 1.
- This condition is reached if the item read was in location 1. Once this it is read out, SP is decremented and reaches the value 0, which is the initial value of SP.
- If a pop operation reads the item from location 0 and then is decremented, SP changes to 111111, which is equivalent to decimal 63 in above configuration, the word in address 0 receives the last item in the stack.

Memory Stack:

- In the above discussion a stack can exist as a stand-alone unit. But in the CPU implementation of a stack is done by assigning a portion of memory to a stack operation and using a processor register as stack pointer.

- The below figure shows a portion computer memory partitioned into three segments: program, data, and stack.

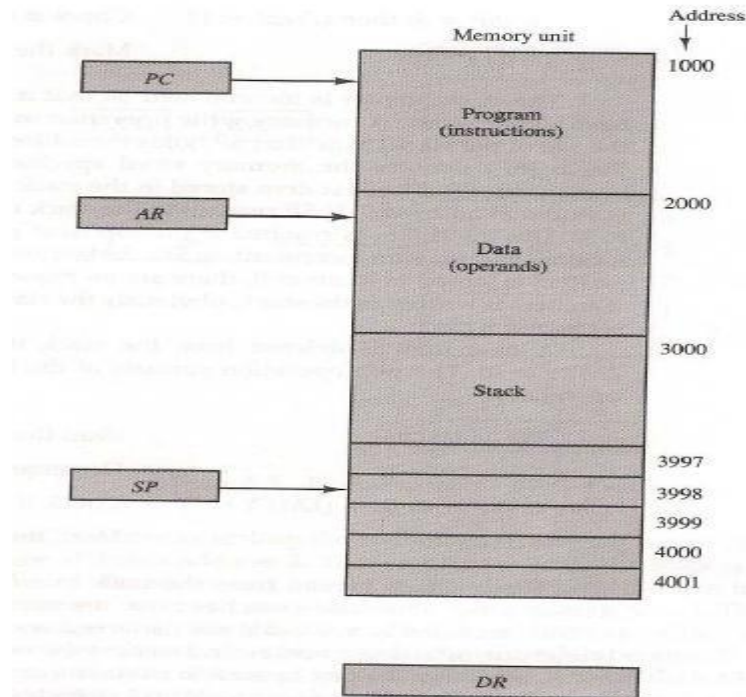


Figure 8-4 Computer memory with program, data, and stack segments.

- The program counter *PC* points at the address of the next instruction in program.
- The address register *AR* points at an array of data.
- The stack pointer *SP* points at the top of the stack.
- The three registers are connected to a common address bus, and either one can provide an address for memory.
 - *PC* is used during the fetch phase to read an instruction.
 - *AR* is used during the exec phase to read an operand.
 - *SP* is used to push or pop items into or from stack.
- As shown in Fig. 8-4, the initial value of *SP* is 4001 and the stack grows with decreasing addresses.
- Thus the first item stored in the stack is at address 4000, the second item is stored at address 3999, and the last address that can be used for the stack is 3000.
- No provisions are available for stack limit checks.
- The items in the stack communicate with a data register *DR*. A new item is inserted with the push operation as follows:

$$SP \leftarrow SP-1$$

$$M [SP] \leftarrow DR$$

- The stack pointer is decremented so that it points at the address of the next word.
- A memory write operation inserts the word from *DR* into the top of stack. A new item is deleted with a pop operation as follows:
 - $DR \leftarrow M [SP]$
 - $SP \leftarrow SP+1$
- The top item is read from the stack into *DR*. The stack pointer is then decremented to point at the next item in the stack.
- Most computers do not provide hardware to check for stack overflow (full stack) or underflow (empty stack).

- The stack limits can be checked by using processor registers:
 - one to hold the upper limit (3000 in this case)
 - Other to hold the lower limit (4001 in this case).
- After a push operation, *SP* compared with the upper-limit register and after a pop operation, *SP* is compared with the lower-limit register.
- The two micro operations needed for either the push or pop are
 - (1) An access to memory through *SP*
 - (2) Updating *SP*
- The advantage of a memory stack is that the CPU can refer to it without having specify an address, since the address is always available and automatically updated in the stack pointer.

Reverse Polish Notation:

- A stack organization is very effective for evaluating arithmetic expressions.
- The common arithmetic expressions are written in *infix notation*, with each operator written *between* the operands.
- Consider the simple arithmetic expression.

A*B+C*D
- For evaluating the above expression it is necessary to compute the product A*B, store this product result while computing C*D, and then sum the two products.
- For doing this type of infix notation, it is necessary to scan back and forth along the expression to determine the next operation to be performed.
- The Polish mathematician Lukasiewicz showed that arithmetic expression can be represented in *prefix notation*.
- This representation, often referred to as *Polish notation*, places the operator before the operands. So it is also called as *prefix notation*.
- The *Postfix notation*, referred to as *reverse Polish notation (RPN)*, places the operator after the operands.
- The following examples demonstrate the three representations
 - Eg: A+B ----- > Infix notation
 - +AB -----> Prefix or Polish notation
 - AB+ -----> Post or reverse Polish notation
- The reverse Polish notation is in a form suitable for stack manipulation. The expression

A*B+C*D

 Is written in reverse polish notation as

AB* CD* +

And it is evaluated as follows

- ✓ Scan the expression from left to right.
- ✓ When operator is reached, perform the operation with the two operands found on the left side of the operator.
- ✓ Remove the two operands and the operator and replace them by the number obtained from the result of the operation.
- ✓ Continue to scan the expression and repeat the procedure for every operation encountered until there are no more operators.
- For the expression above it find the operator * after A and B. So it perform the operation A*B and replace A, B and * with the result.
- The next operator is a * and it previous two operands are C and D, so it perform the operation C*D and places the result in places C, D and *.
- The next operator is + and the two operands to be added are the two products, so we add the two quantities to obtain the result.
- The conversion from infix notation to reverse Polish notation must take into consideration the operational hierarchy adopted for infix notation.

- This hierarchy dictates that we first perform all arithmetic inside inner parentheses, then inside outer parentheses, and do multiplication and division operations before addition and subtraction operations.

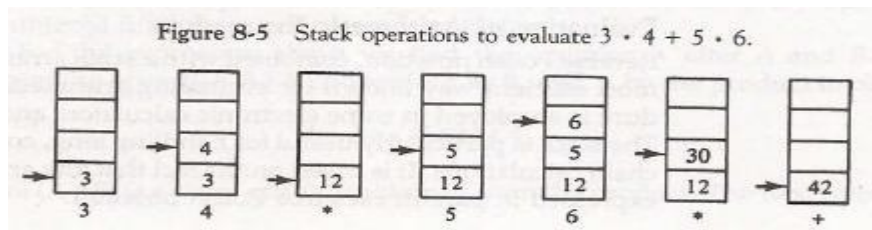
Evaluation of Arithmetic Expressions:

- Reverse Polish notation, combined with a stack arrangement of registers, is the most efficient way known for evaluating arithmetic expressions.
- This procedure is employed in some electronic calculators and also in some computer.
- The following numerical example may clarify this procedure. Consider the arithmetic expression

$$(3*4) + (5*6)$$

In reverse polish notation, it is expressed as 34 * 56* +

- Now consider the stack operations shown in Fig. 8-5.



- Each box represents one stack operation and the arrow always points to the top of the stack.
- Scanning the expression from left to right, we encounter two operands.
- First the number 3 is pushed into the stack, then the number 4.
- The next symbol is the multiplication operator *.
- This causes a multiplication of the two top most items the stack.
- The stack is then popped and the product is placed on top of the stack, replacing the two original operands.
- Next we encounter the two operands 5 and 6, so they are pushed into the stack.
- The stack operation results from the next * replaces these two numbers by their product.
- The last operation causes an arithmetic addition of the two topmost numbers in the stack to produce the final result of 42.

Instruction Formats:

- The format of an instruction is usually depicted in a rectangular box symbolizing the bits of the instruction as they appear in memory words or in a control register.
- The bits of the instruction are divided into groups called fields.
- The most common fields found in instruction formats are:
 1. An operation code field that specifies the operation to be perform
 2. An address field that designates a memory address or a processor register.
 3. A mode field that specifies the way the operand or the effective address is determined.
- Computers may have instructions of several different lengths containing varying number of addresses.
- The number of address fields in the instruct format of a computer depends on the internal organization of its registers.

- Most computers fall into one of three types of CPU organizations:
 1. Single accumulator organization.
 2. General register organization.
 3. Stack organization.

Single Accumulator Organization:

- ✓ In an accumulator type organization all the operations are performed with an implied accumulator register.
- ✓ The instruction format in this type of computer uses one address field.
- ✓ For example, the instruction that specifies an arithmetic addition defined by an assembly language instruction as
 - **ADD X**
- ✓ Where X is the address of the operand. The ADD instruction in this case results in the operation $AC \leftarrow AC + M[X]$. AC is the accumulator register and M[X] symbolizes the memory word located at address X.

General register organization:

- ✓ The instruction format in this type of computer needs three register address fields.
- ✓ Thus the instruction for an arithmetic addition may be written in an assembly language as
 - **ADD R1, R2, R3**to denote the operation $R1 \leftarrow R2 + R3$. The number of address fields in the instruction can be reduced from three to two if the destination register is the same as one of the source registers.
- ✓ Thus the instruction **ADD R1, R2** would denote the operation $R1 \leftarrow R1 + R2$. Only register addresses for R1 and R2 need be specified in this instruction.
- ✓ General register-type computers employ two or three address fields in their instruction format.
- ✓ Each address field may specify a processor register or a memory word.
- ✓ An instruction symbolized by **ADD R1, X** would specify the operation $R1 \leftarrow R1 + M[X]$.
- ✓ It has two address fields, one for register R1 and the other for the memory address X.

Stack organization:

- ✓ The stack-organized CPU has PUSH and POP instructions which require an address field.
- ✓ Thus the instruction **PUSH X** will push the word at address X to the top of the stack.
- ✓ The stack pointer is updated automatically.
- ✓ Operation-type instructions do not need an address field in stack-organized computers.
- ✓ This is because the operation is performed on the two items that are on top of the stack.
- ✓ The instruction **ADD** in a stack computer consists of an operation code only with no address field.
- ✓ This operation has the effect of popping the two top numbers from the stack, adding the numbers, and pushing the sum into the stack.
- ✓ There is no need to specify operands with an address field since all operands are implied to be in the stack.

- Most computers fall into one of the three types of organizations.
- Some computers combine features from more than one organizational structure.
- The influence of the number of addresses on computer programs, we will evaluate the arithmetic statement

$$X = (A+B) * (C+D)$$

- Using zero, one, two, or three address instructions and using the symbols ADD, SUB, MUL and DIV for four arithmetic operations; MOV for the transfer type operations; and LOAD and STORE for transfer to and from memory and AC register.
- Assuming that the operands are in memory addresses A, B, C, and D and the result must be stored in memory address X and also the CPU has general purpose registers R1, R2, R3 and R4.

Three Address Instructions:

- ✓ Three-address instruction formats can use each address field to specify either a processor register or a memory operand.
- ✓ The program assembly language that evaluates $X = (A+B) * (C+D)$ is shown below, together with comments that explain the register transfer operation of each instruction.

```
ADD    R1, A, B    R1 ← M[A] + M[B]
ADD    R2, C, D    R2 ← M[C] + M[D]
MUL    X, R1, R2   M[X] ← R1 * R2
```

- ✓ The symbol M [A] denotes the operand at memory address symbolized by A.
- ✓ The advantage of the three-address format is that it results in short programs when evaluating arithmetic expressions.
- ✓ The disadvantage is that the binary-coded instructions require too many bits to specify three addresses.

Two Address Instructions:

- ✓ Two-address instructions formats use each address field can specify either a processor register or memory word.
- ✓ The program to evaluate $X = (A+B) * (C+D)$ is as follows

```
MOV    R1, A    R1 ← M[A]
ADD    R1, B    R1 ← R1 + M[B]
MOV    R2, C    R2 ← M[C]
ADD    R2, D    R2 ← R2 + M[D]
MUL    R1, R2   R1 ← R1 * R2
MOV    X, R1    M[X] ← R1
```

- ✓ The MOV instruction moves or transfers the operands to and from memory and processor registers.
- ✓ The first symbol listed in an instruction is assumed be both a source and the destination where the result of the operation transferred.

One Address Instructions:

- ✓ One-address instructions use an implied accumulator (AC) register for all data manipulation.
- ✓ For multiplication and division there is a need for a second register. But for the basic discussion we will neglect the second register and assume that the AC contains the result of all operations.
- ✓ The program to evaluate $X=(A+B) * (C+D)$ is

LOAD	A	AC ← M[A]
ADD	B	AC ← AC + M[B]
STORE	T	M[T] ← AC
LOAD	C	AC ← M[C]
ADD	D	AC ← AC + M[D]
MUL	T	AC ← AC * M[T]
STORE	X	M[X] ← AC

- ✓ All operations are done between the AC register and a memory operand.
- ✓ T is the address of a temporary memory location required for storing the intermediate result.

Zero Address Instructions:

- ✓ A stack-organized computer does not use an address field for the instructions ADD and MUL.
- ✓ The PUSH and POP instructions, however, need an address field to specify the operand that communicates with the stack.
- ✓ The following program shows how $X = (A+B) * (C+D)$ will be written for a stack-organized computer.

(TOS stands for top of stack).

PUSH	A	TOS ← A
PUSH	B	TOS ← B
ADD		TOS ← (A + B)
PUSH	C	TOS ← C
PUSH	D	TOS ← D
ADD		TOS ← (C + D)
MUL		TOS ← (C + D) * (A + B)
POP	X	M[X] ← TOS

- ✓ To evaluate arithmetic expressions in a stack computer, it is necessary to convert the expression into reverse Polish notation.
- ✓ The name "zero-address" is given to this type of computer because of the absence of an address field in the computational instructions.

RISC Instructions:

- ✓ The instruction set of a typical RISC processor is use only load and store instructions for communicating between memory and CPU.
- ✓ All other instructions are executed within the registers of CPU without referring to memory.
- ✓ LOAD and STORE instructions that have one memory and one register address, and computational type instructions that have three addresses with all three specifying processor registers.
- ✓ The following is a program to evaluate $X=(A+B)*(C+D)$

LOAD	R1, A	R1 ← M[A]
LOAD	R2, B	R2 ← M[B]
LOAD	R3, C	R3 ← M[C]
LOAD	R4, D	R4 ← M[D]
ADD	R1, R1, R2	R1 ← R1 + R2
ADD	R3, R3, R2	R3 ← R3 + R4
MUL	R1, R1, R3	R1 ← R1 * R3
STORE	X, R1	M[X] ← R1

- ✓ The load instructions transfer the operands from memory to CPU register.
- ✓ The add and multiply operations are executed with data in the register without accessing memory.
- ✓ The result of the computations is then stored memory with a store in instruction.

Addressing Modes:

- The way the operands are chosen during program execution is dependent on the addressing mode of the instruction.
- Computers use addressing mode techniques for the purpose of accommodating one or both of the following provisions:
 - To give programming versatility to the user by providing such facilities as pointers to memory, counters for loop control, indexing of data, and program relocation.
 - To reduce the number of bits in the addressing field of the instruction
- Most addressing modes modify the address field of the instruction; there are two modes that need no address field at all. These are *implied* and *immediate* modes.

Implied Mode:

- ✓ In this mode the operands are specified implicitly in the definition of the instruction.
- ✓ For example, the instruction "complement accumulator" is an implied-mode instruction because the operand in the accumulator register is implied in the definition of the instruction.
- ✓ All register reference instructions that use an accumulator are implied mode instructions.
- ✓ Zero address in a stack organization computer is implied mode instructions.

Immediate Mode:

- ✓ In this mode the operand is specified in the instruction itself.
- ✓ In other words an immediate-mode instruction has an operand rather than an address field.
- ✓ Immediate-mode instructions are useful for initializing registers to a constant value.
 - The address field of an instruction may specify either a memory word or a processor register.
 - When the address specifies a processor register, the instruction is said to be in the register mode.

Register Mode:

- ✓ In this mode the operands are in registers that reside within the CPU.
- ✓ The particular register is selected from a register field in the instruction.

Register Indirect Mode:

- ✓ In this mode the instruction specifies a register in CPU whose contents give the address of the operand in memory.
- ✓ In other words, the selected register contains the address of the operand rather than the operand itself.
- ✓ The advantage of a register indirect mode instruction is that the address field of the instruction uses few bits to select a register than would have been required to specify a memory address directly.

Auto-increment or Auto-Decrement Mode:

- This is similar to the register indirect mode except that the register is incremented or decremented after (or before) its value is used to access memory.

- The address field of an instruction is used by the control unit in the CPU to obtain the operand from memory.
- Sometimes the value given in the address field is the address of the operand, but sometimes it is just an address from which the address of the operand is calculated.
- The basic two mode of addressing used in CPU are *direct* and *indirect* address mode.

Direct Address Mode:

- ✓ In this mode the effective address is equal to the address part of the instruction.
- ✓ The operand resides in memory and its address is given directly by the address field of the instruction.
- ✓ In a branch-type instruction the address field specifies the actual branch address.

Indirect Address Mode:

- ✓ In this mode the address field of the instruction gives the address where the effective address is stored in memory.
- ✓ Control fetches the instruction from memory and uses its address part to access memory again to read the effective address.
- A few addressing modes require that the address field of the instruction be added to the content of a specific register in the CPU.
- The effective address in these modes is obtained from the following computation:
Effective address = address part of instruction + content of CPU register
- The CPU register used in the computation may be the program counter, an index register, or a base register.
- We have a different addressing mode which is used for a different application.

Relative Address Mode:

- ✓ In this mode the content of the program counter is added to the address part of the instruction in order to obtain the effective address.

Indexed Addressing Mode:

- ✓ In this mode the content of an index register is added to the address part of the instruction to obtain the effective address.
- ✓ An index register is a special CPU register that contains an index value.

Base Register Addressing Mode:

- ✓ In this mode the content of a base register is added to the address part of the instruction to obtain the effective address.
- ✓ This is similar to the indexed addressing mode except that the register is now called a base register instead of an index register.

Data Transfer and Manipulation:

- Most computer instructions can be classified into three categories:
 1. *Data transfer instructions*
 2. *Data manipulation instructions*
 3. *Program control instructions*

Data Transfer Instructions:

- Data transfer instructions move data from one place in the computer to another without changing the data content.
- The most common transfers are between memory and processor registers, between processor registers and input or output, and between the processor registers themselves.
- Table 8-5 gives a list of eight data transfer instructions used in many computers.

TABLE 8-5 Typical Data Transfer Instructions

Name	Mnemonic
Load	LD
Store	ST
Move	MOV
Exchange	XCH
Input	IN
Output	OUT
Push	PUSH
Pop	POP

- The **load** instruction has been used mostly to designate a transfer from memory to a processor register, usually an accumulator.
- The **store** instruction designates a transfer from a processor register into memory.
- The **move** instruction has been used in computers with multiple CPU registers to designate a transfer from one register to another and also between CPU registers and memory or between two memory words.
- The **exchange** instruction swaps information between two registers or a register and a memory word.
- The **input** and **output** instructions transfer data among processor registers and input or output terminals.
- The **push** and **pop** instructions transfer data between processor registers and a memory stack.
- Different computers use different mnemonics symbols for differentiate the addressing modes.
- As an example, consider the *load to accumulator* instruction when used with eight different addressing modes.
- Table 8-6 shows the recommended assembly language convention and actual transfer accomplished in each case
- *ADR* stands for an address.
- *NBA* a number or operand.
- X is an index register.
- The @ character symbolizes an indirect addressing.
- R1 is a processor register.
- AC is the accumulator register.
- The \$ character before an address makes the address relative to the program counter *PC*.
- The # character precedes the operand in an immediate-mode instruction.

Data Manipulation Instructions:

- Data manipulation instructions perform operations on data and provide the computational capabilities for the computer.
- The data manipulation instructions in a typical computer are usually divided into three basic types:
 1. Arithmetic instructions
 2. Logical and bit manipulation instructions
 3. Shift instructions

Arithmetic instructions

- ✓ The four basic arithmetic operations are addition, subtraction, multiplication and division.
- ✓ Most computers provide instructions for all four operations.
- ✓ Some small computers have only addition and possibly subtraction instructions. The multiplication and division must then be generated by mean software subroutines.
- ✓ A list of typical arithmetic instructions is given in Table 8-7.

TABLE 8-7 Typical Arithmetic Instructions

Name	Mnemonic
Increment	INC
Decrement	DEC
Add	ADD
Subtract	SUB
Multiply	MUL
Divide	DIV
Add with carry	ADDC
Subtract with borrow	SUBB
Negate (2's complement)	NEG

- ✓ The increment instruction adds 1 to the value stored in a register or memory word.
- ✓ A number with all 1's, when incremented, produces a number with all 0's.
- ✓ The decrement instruction subtracts 1 from a value stored in a register or memory word.
- ✓ A number with all 0's, when decremented, produces number with all 1's.
- ✓ The add, subtract, multiply, and divide instructions may be use different types of data.
- ✓ The data type assumed to be in processor register during the execution of these arithmetic operations is defined by an operation code.
- ✓ An arithmetic instruction may specify fixed-point or floating-point data, binary or decimal data, single-precision or double-precision data.
- ✓ The mnemonics for three add instructions that specify different data types are shown below.
 - ADDI Add two binary integer numbers
 - ADDF Add two floating-point numbers
 - ADDD Add two decimal numbers in BCD
- ✓ A special carry flip-flop is used to store the carry from an operation.
- ✓ The instruction "add carry" performs the addition on two operands plus the value of the carry the previous computation.
- ✓ Similarly, the "subtract with borrow" instruction subtracts two words and borrow which may have resulted from a previous subtract operation.

- ✓ The negate instruction forms the 2's complement number, effectively reversing the sign of an integer when represented in signed-2's complement form.

Logical and bit manipulation instructions

- ✓ Logical instructions perform binary operations on strings of bits stored in registers.
- ✓ They are useful for manipulating individual bits or a group of bits that represent binary-coded information.
- ✓ The logical instructions consider each bit of the operand separately and treat it as a Boolean variable.
- ✓ By proper application of the logical instructions it is possible to change bit values, to clear a group of bits, or to insert new bit values into operands stored in register memory words.
- ✓ Some typical logical and bit manipulation instructions are listed in Table 8-8.

TABLE 8-8 Typical Logical and Bit Manipulation Instructions

Name	Mnemonic
Clear	CLR
Complement	COM
AND	AND
OR	OR
Exclusive-OR	XOR
Clear carry	CLRC
Set carry	SETC
Complement carry	COMC
Enable interrupt	EI
Disable interrupt	DI

- ✓ The clear instruction causes the specified operand to be replaced by 0's.
- ✓ The complement instruction produces the 1's complement by inverting all bits of the operand.
- ✓ The AND, OR, and XOR instructions produce the corresponding logical operations on individual bits of the operands.
- ✓ The logical instructions can also be used to perform bit manipulation operations.
- ✓ There are three bit manipulation operations possible: a selected bit can be cleared to 0, or can be set to 1, or can be complemented.
 - The AND instruction is used to clear a bit or a selected group of bits of an operand.
 - The OR instruction is used to set a bit or a selected group of bits of an operand.
 - Similarly, the XOR instruction is used to selectively complement bits of an operand.
- ✓ Other bit manipulation instructions included in the above table perform operations on individual bits such as a carry can be cleared, set, or complemented.
- ✓ Another example is a flip-flop that controls the interrupt facility and is either enabled or disabled by means of bit manipulation instructions.

Shift Instructions

- ✓ Shifts are operations in which the bits of a word are moved to the left or right.
- ✓ The bit shifted in at the end of the word determines the type of shift used.
- ✓ Shift instructions may specify logical shifts, arithmetic shifts, or rotate-type operations.
- ✓ In either case the shift may be to the right or to the left.
- ✓ Table 8-9 lists four types of shift instructions.

TABLE 8-9 Typical Shift Instructions

Name	Mnemonic
Logical shift right	SHR
Logical shift left	SHL
Arithmetic shift right	SHRA
Arithmetic shift left	SHLA
Rotate right	ROR
Rotate left	ROL
Rotate right through carry	RORC
Rotate left through carry	ROLC

- ✓ The logical shift inset to the end bit position.
- ✓ The end position is the leftmost bit position for shift rights the rightmost bit position for the shift left.
- ✓ Arithmetic shifts usually conform to the rules for signed-2's complement numbers.
- ✓ The arithmetic shift-right instruction must preserve the sign bit in the leftmost position.
- ✓ The sign bit is shifted to the right together with the rest of the number, but the sign bit itself remains unchanged.
- ✓ This is a shift-right operation with the end bit remaining the same.
- ✓ The arithmetic shift-left instruction inserts 0 to the end position and is identical to the logical shift-instruction.
- ✓ The rotate instructions produce a circular shift. Bits shifted out at one of the word are not lost as in a logical shift but are circulated back into the other end.
- ✓ The rotate through carry instruction treats a carry bit as an extension of the register whose word is being rotated.
- ✓ Thus a rotate-left through *carry* instruction transfers the carry bit into the rightmost bit position of the register, transfers the leftmost bit position into the carry, and at the same time, shift the entire register to the left.

Program Control:

- Program control instructions specify conditions for altering the content of the program counter.
- The change in value of the program counter as a result of the execution of a program control instruction causes a break in the sequence of instruction execution.
- This instruction provides control over the flow of program execution and a capability for branching to different program segments.
- Some typical program control instructions are listed in Table 8.10.

TABLE 8-10 Typical Program Control Instructions

Name	Mnemonic
Branch	BR
Jump	JMP
Skip	SKP
Call	CALL
Return	RET
Compare (by subtraction)	CMP
Test (by ANDing)	TST

- Branch and jump instructions may be conditional or unconditional.
- An unconditional branch instruction causes a branch to the specified address without

any conditions.

- The conditional branch instruction specifies a condition such as branch if positive or branch if zero.
- The skip instruction does not need an address field and is therefore a zero-address instruction.
- A conditional skip instruction will skip the next instruction if the condition is met. This is accomplished by incrementing program counter.
- The call and return instructions are used in conjunction with subroutines.
- The compare instruction forms a subtraction between two operands, but the result of the operation not retained. However, certain status bit conditions are set as a result of operation.
- Similarly, the test instruction performs the logical AND of two operands and updates certain status bits without retaining the result or changing the operands.

Status Bit Conditions:

- The ALU circuit in the CPU have status register for storing the status bit conditions.
- Status bits are also called *condition-code* bits or *flag* bits.

Figure 8-8 shows block diagram of an 8-bit ALU with a 4-bit status register.

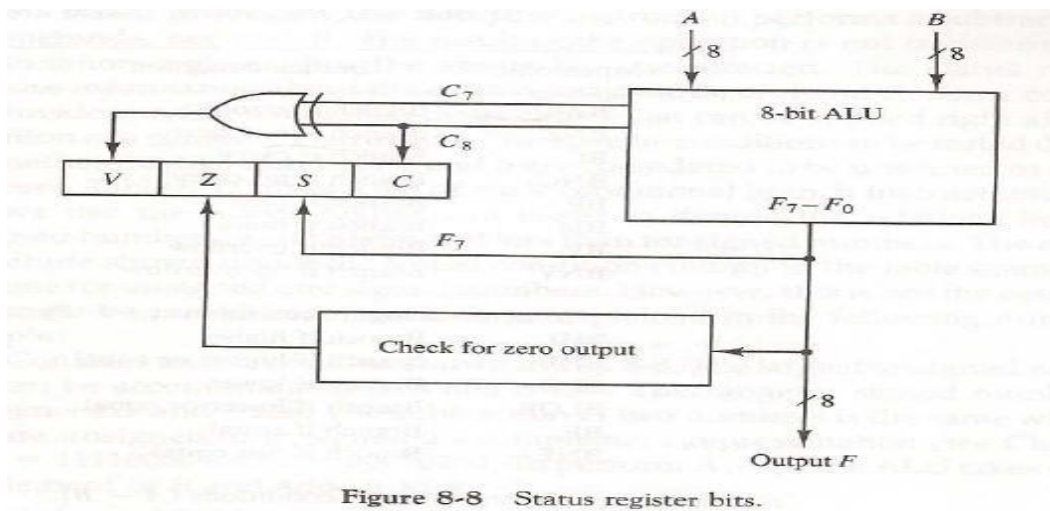


Figure 8-8 Status register bits.

- The four status bits are symbolized by C, S, Z, and V. The bits are set or cleared as a result of an operation performed in the ALU.
 - Bit C (carry) is set to 1 if the end carry C_8 is 1. It is cleared to 0 if the carry is 0.
 - S (sign) is set to 1 if the highest-order bit F_7 is 1. It is set to 0 if the bit is 0.
 - Bit Z (zero) is set to 1 if the output of the ALU contains all 0's. It is clear to 0 otherwise. In other words, $Z = 1$ if the output is zero and $Z = 0$ if the output is not zero.
 - Bit V (overflow) is set to 1 if the exclusive-OR of the last two carries equal to 1, and cleared to 0 otherwise.
- The above status bits are used in conditional jump and branch instructions.

Subroutine Call and Return:

- A subroutine is self contained sequence of instructions that performs a given computational task.
- The most common names used are call subroutine, jump to subroutine, branch to subroutine, or branch and save return address.

- A subroutine is executed by performing two operations
 - The address of the next instruction available in the program counter (the return address) is stored in a temporary location so the subroutine knows where to return
 - Control is transferred to the beginning of the subroutine.
- The last instruction of every subroutine, commonly called *return from subroutine*, transfers the return address from the temporary location in the program counter.
- Different computers use a different temporary location for storing the return address.
- The most efficient way is to store the return address in a memory stack.
- The advantage of using a stack for the return address is that when a succession of subroutines is called, the sequential return addresses can be pushed into the stack.
- A subroutine call is implemented with the following microoperations:

$SP \leftarrow SP - 1$	Decrement stack pointer
$M[SP] \leftarrow PC$	Push content of <i>PC</i> onto the stack
$PC \leftarrow \text{effective address}$	Transfer control to the subroutine

- The instruction that returns from the last subroutine is implemented by the microoperations:

$PC \leftarrow M[SP]$	Pop stack and transfer to <i>PC</i>
$SP \leftarrow SP + 1$	Increment stack pointer

Program Interrupt:

- Program interrupt refers to the transfer of program control from a currently running program to another service program as a result of an external or internal generated request.
- The interrupt procedure is similar to a subroutine call except for three variations:
 - The interrupt is initiated by an internal or external signal.
 - Address of the interrupt service program is determined by the hardware.
 - An interrupt procedure usually stores all the information rather than storing only PC content.

Types of interrupts:

- ✓ There are three major types of interrupts that cause a break in the normal execution of a program.
- ✓ They can be classified as
 - **External interrupts:**
 - These come from input—output (I/O) devices, from a timing device, from a circuit monitoring the power supply, or from any other external source.
 - Ex: I/O device requesting transfer of data, I/O device finished transfer of data, elapsed time of an event, or power failure.
 - **Internal interrupts:**
 - These arise from illegal or erroneous use of an instruction or data.

- Internal interrupts are also called *traps*.
- Ex: interrupts caused by internal error conditions are register overflow, attempt to divide by zero, an invalid operation code, stack overflow, and protection violation.
- ✓ Internal and external interrupts are initiated nals that occur in hardware of CPU.
 - **Software interrupts**
 - A software interrupt is initiated by executing an instruction.
 - Software interrupt is a special call instruction that behaves like an interrupt rather than a subroutine call.

Reduced Instruction Set Computer:

- A computer with large number instructions is classified as a *complex instruction set computer*, abbreviated as CISC.
- The computer which having the fewer instructions is classified as a *reduced instruction set computer*, abbreviated as RISC.

CISC Characteristics:

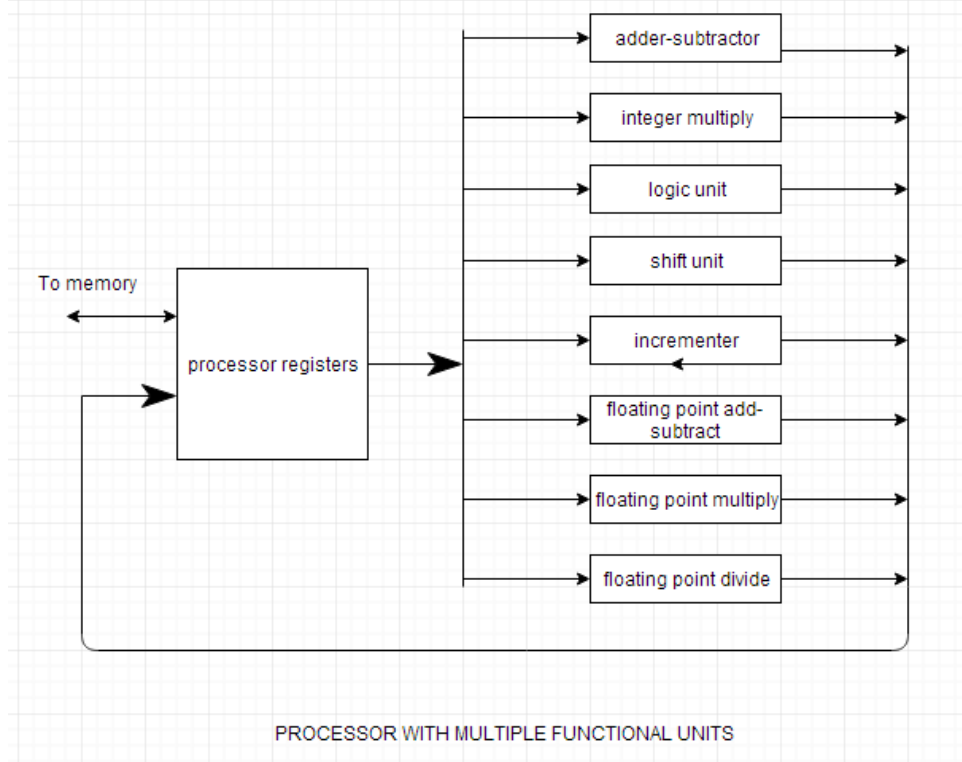
- ✓ A large number of instructions--typically from 100 to 250 instructions.
- ✓ Some instructions that perform specialized tasks and are used infrequently.
- ✓ A large variety of addressing modes—typically from 5 to 20 differ modes.
- ✓ Variable-length instruction formats
- ✓ Instructions that manipulate operands in memory

RISC Characteristics:

- ✓ Relatively few instructions
- ✓ Relatively few addressing modes
- ✓ Memory access limited to load and store instructions
- ✓ All operations done within the registers of the CPU
- ✓ Fixed-length, easily decoded instruction format
- ✓ Single-cycle instruction execution
- ✓ Hardwired rather than micro programmed control
- ✓ A relatively large number of registers in the processor unit
- ✓ Efficient instruction pipeline

Parallel Processing:

- Instead of processing each instruction sequentially, a parallel processing system provides concurrent data processing to increase the execution time.
- In this the system may have two or more ALU's and should be able to execute two or more instructions at the same time. The purpose of parallel processing is to speed up the computer processing capability and increase its throughput.
- **Throughput** is the number of instructions that can be executed in a unit of time.
- Parallel processing can be viewed from various levels of complexity. At the lowest level, we distinguish between parallel and serial operations by the type of registers used. At the higher level of complexity, parallel processing can be achieved by using multiple functional units that perform many operations simultaneously.



Data Transfer Modes of a Computer System

- According to the data transfer mode, computer can be divided into 4 major groups:
 1. SISD
 2. SIMD
 3. MISD
 4. MIMD

SISD (Single Instruction Stream, Single Data Stream)

- It represents the organization of a single computer containing a control unit, processor unit and a memory unit.
- Instructions are executed sequentially. It can be achieved by pipelining or multiple functional units.

SIMD (Single Instruction Stream, Multiple Data Stream)

- It represents an organization that includes multiple processing units under the control of a common control unit.
- All processors receive the same instruction from control unit but operate on different parts of the data.
- They are highly specialized computers. They are basically used for numerical problems that are expressed in the form of vector or matrix. But they are not suitable for other types of computations

MISD (Multiple Instruction Stream, Single Data Stream)

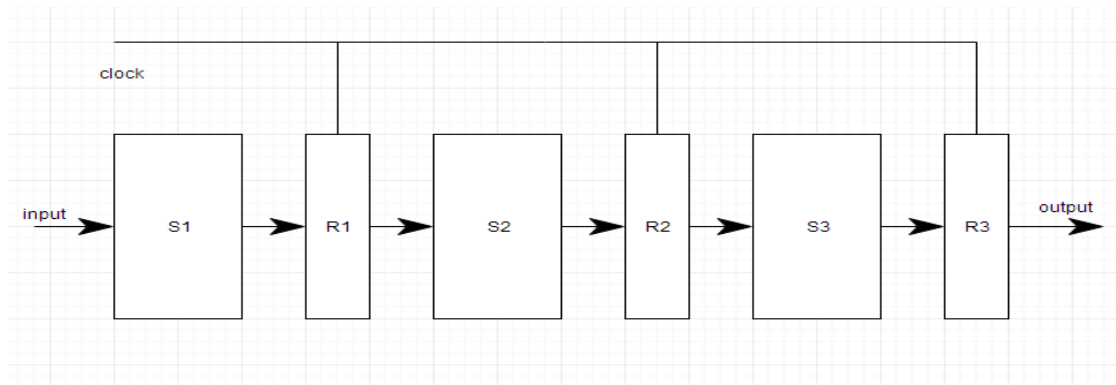
- It consists of a single computer containing multiple processors connected with multiple control units and a common memory unit.
- It is capable of processing several instructions over single data stream simultaneously.
- MISD structure is only of theoretical interest since no practical system has been constructed using this organization.

MIMD (Multiple Instruction Stream, Multiple Data Stream)

- It represents the organization which is capable of processing several programs at same time.
- It is the organization of a single computer containing multiple processors connected with multiple control units and a shared memory unit.
- The shared memory unit contains multiple modules to communicate with all processors simultaneously.
- Multiprocessors and multicomputer are the examples of MIMD.
- It fulfills the demand of large scale computations.

Pipelining:

- Pipelining is the process of accumulating instruction from the processor through a pipeline. It allows storing and executing instructions in an orderly process. It is also known as **pipeline processing**.
- Pipelining is a technique where multiple instructions are overlapped during execution. Pipeline is divided into stages and these stages are connected with one another to form a pipe like structure. Instructions enter from one end and exit from another end.
- Pipelining increases the overall instruction throughput.
- In pipeline system, each segment consists of an input register followed by a combinational circuit. The register is used to hold data and combinational circuit performs operations on it. The output of combinational circuit is applied to the input register of the next segment.



- Pipeline system is like the modern day assembly line setup in factories. For example in a car manufacturing industry, huge assembly lines are setup and at each point, there are robotic arms to perform a certain task, and then the car moves on ahead to the next arm.

Types of Pipeline

It is divided into 2 categories:

1. Arithmetic Pipeline
2. Instruction Pipeline

Arithmetic Pipeline

- Arithmetic pipelines are usually found in most of the computers. They are used for floating point operations, multiplication of fixed point numbers etc. For example: The input to the Floating Point Adder pipeline is:

$$X=A*2^a$$

$$Y=B*2^b$$

- Here A and B are mantissas (significant digit of floating point numbers), while **a** and **b** are exponents.
- The floating point addition and subtraction is done in 4 parts:
 - Compare the exponents.
 - Align the mantissas.
 - Add or subtract mantissas
 - Produce the result.
 - Registers are used for storing the intermediate results between the above operations.

Instruction Pipeline

- In this a stream of instructions can be executed by overlapping *fetch*, *decode* and *execute* phases of an instruction cycle. This type of technique is used to increase the throughput of the computer system.
- An instruction pipeline reads instruction from the memory while previous instructions are being executed in other segments of the pipeline. Thus we can execute multiple instructions simultaneously. The pipeline will be more efficient if the instruction cycle is divided into segments of equal duration.

General Considerations:

- There are some factors that cause the pipeline to deviate its normal performance. Some of these factors are given below:

1. Timing Variations

- All stages cannot take same amount of time. This problem generally occurs in instruction processing where different instructions have different operand requirements and thus different processing time.

2. Data Hazards

- When several instructions are in partial execution, and if they reference same data then the problem arises. We must ensure that next instruction does not attempt to access data before the current instruction, because this will lead to incorrect results.

3. Branching

- In order to fetch and execute the next instruction, we must know what that instruction is. If the present instruction is a conditional branch, and its result will lead us to the next instruction, then the next instruction may not be known until the current one is processed.

4. Interrupts

- Interrupts set unwanted instruction into the instruction stream. Interrupts effect the execution of instruction.

5. Data Dependency

- It arises when an instruction depends upon the result of a previous instruction but this result is not yet available.

Advantages of Pipelining

1. The cycle time of the processor is reduced.
2. It increases the throughput of the system
3. It makes the system reliable.

Disadvantages of Pipelining

1. The design of pipelined processor is complex and costly to manufacture.
2. The instruction latency is more.

Input Output Organization:

Input Output Subsystem

- The I/O subsystem of a computer provides an efficient mode of communication between the central system and the outside environment.
- It handles all the input-output operations of the computer system.

Peripheral Devices

- Input or output devices that are connected to computer are called **peripheral devices**. These devices are designed to read information into or out of the memory unit upon command from the CPU and are considered to be the part of computer system. These devices are also called **peripherals**.
- For example: *Keyboards, display units* and *printers* are common peripheral devices.
- There are three types of peripherals:
 1. **Input peripherals:** Allows user input, from the outside world to the computer. Example: Keyboard, Mouse etc.
 2. **Output peripherals:** Allows information output, from the computer to the outside world. Example: Printer, Monitor etc
 3. **Input-Output peripherals:** Allows both input(from outised world to computer) as well as, output(from computer to the outside world). Example: Touch screen etc.

Interfaces

- Interface is a shared boundary between two separate components of the computer system which can be used to attach two or more components to the system for communication purposes.
- There are two types of interface:
 1. CPU Interface
 2. I/O Interface

Input-Output Interface

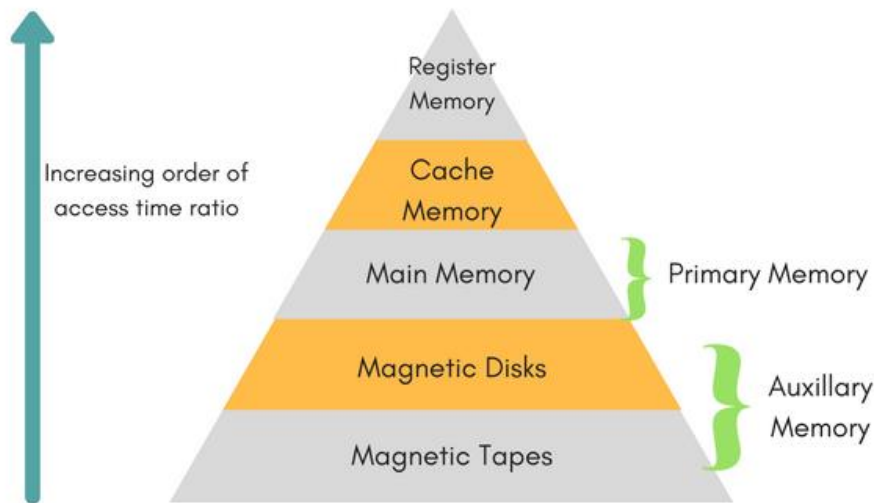
- Peripherals connected to a computer need special communication links for interfacing with CPU.
- In computer system, there are special hardware components between the CPU and peripherals to control or manage the input-output transfers. These components are called **input-output interface units** because they provide communication links between processor bus and peripherals.

- They provide a method for transferring information between internal system and input-output devices.

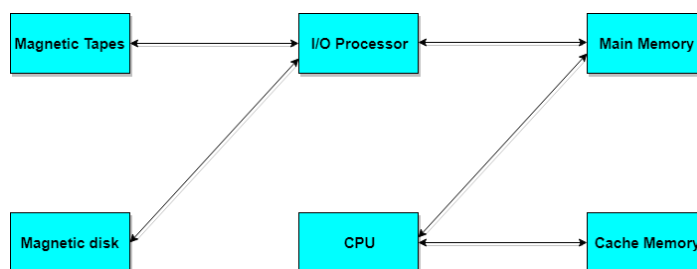
Memory Organization:

- A memory unit is the collection of storage units or devices together. The memory unit stores the binary information in the form of bits. Generally, memory/storage is classified into 2 categories:
 - **Volatile Memory:** This loses its data, when power is switched off.
 - **Non-Volatile Memory:** This is a permanent storage and does not lose any data when power is switched off.

Memory Hierarchy:



- The total memory capacity of a computer can be visualized by hierarchy of components. The memory hierarchy system consists of all storage devices contained in a computer system from the slow Auxiliary Memory to fast Main Memory and to smaller Cache memory.
- **Auxiliary memory** access time is generally **1000 times** that of the main memory, hence it is at the bottom of the hierarchy.
- The **main memory** occupies the central position because it is equipped to communicate directly with the CPU and with auxiliary memory devices through Input/output processor (I/O).
- When the program not residing in main memory is needed by the CPU, they are brought in from auxiliary memory. Programs not currently needed in main memory are transferred into auxiliary memory to provide space in main memory for other programs that are currently in use.
- The **cache memory** is used to store program data which is currently being executed in the CPU. Approximate access time ratio between cache memory and main memory is about **1 to 7~10**



Memory Access Methods

- Each memory type, is a collection of numerous memory locations. To access data from any memory, first it must be located and then the data is read from the memory location. Following are the methods to access information from memory locations:
 1. **Random Access:** Main memories are random access memories, in which each memory location has a unique address. Using this unique address any memory location can be reached in the same amount of time in any order.
 2. **Sequential Access:** This method allows memory access in a sequence or in order.
 3. **Direct Access:** In this mode, information is stored in tracks, with each track having a separate read/write head.

Main Memory

- The memory unit that communicates directly within the CPU, Auxiliary memory and Cache memory, is called main memory. It is the central storage unit of the computer system. It is a large and fast memory used to store data during computer operations. Main memory is made up of **RAM** and **ROM**, with RAM integrated circuit chips holding the major share.
- RAM: Random Access Memory
 - **DRAM:** Dynamic RAM, is made of capacitors and transistors, and must be refreshed every 10~100 ms. It is slower and cheaper than SRAM.
 - **SRAM:** Static RAM, has a six transistor circuit in each cell and retains data, until powered off.
 - **NVRAM:** Non-Volatile RAM, retains its data, even when turned off. Example: Flash memory.
- ROM: Read Only Memory, is non-volatile and is more like a permanent storage for information. It also stores the **bootstrap loader** program, to load and start the operating system when computer is turned on. **PROM** (Programmable ROM), **EPROM**(Erasable PROM) and **EEPROM**(Electrically Erasable PROM) are some commonly used ROMs.

Auxiliary Memory

- Devices that provide backup storage are called auxiliary memory.
- **For example:** Magnetic disks and tapes are commonly used auxiliary devices.
- Other devices used as auxiliary memory are magnetic drums, magnetic bubble memory and optical disks.
- It is not directly accessible to the CPU, and is accessed using the Input/Output channels.

Cache Memory

- The data or contents of the main memory that are used again and again by CPU, are stored in the cache memory so that we can easily access that data in shorter time.
- Whenever the CPU needs to access memory, it first checks the cache memory. If the data is not found in cache memory then the CPU moves onto the main memory. It also transfers block of recent data into the cache and keeps on deleting the old data in cache to accommodate the new one.

Hit Ratio

- The performance of cache memory is measured in terms of a quantity called **hit ratio**. When the CPU refers to memory and finds the word in cache it is said to produce a **hit**. If the word is not found in cache, it is in main memory then it counts as a **miss**.
- The ratio of the number of hits to the total CPU references to memory is called hit ratio.
- Hit Ratio = Hit/(Hit + Miss)

Associative Memory

- It is also known as **content addressable memory (CAM)**.
- It is a memory chip in which each bit position can be compared.
- In this the content is compared in each bit cell which allows very fast table lookup.
- Since the entire chip can be compared, contents are randomly stored without considering addressing scheme.
- These chips have less storage capacity than regular memory chips.