UNIT – I
BASIC STRUCTURE OF A COMPUTER SYSTEM
PART: A

1. **Define Computer Architecture.**

Computer Architecture is defined as the functional operation of the individual hardware unit in a computer system and the flow of information to control the hardware units.

2. **What are the components of a computer system? (Nov/Dec 2017) (April/May 2017)**

The components of a computer system are

- Input Devices
- Output Devices
- Memory
- CPU or processor
- Network

3. **What are the addressing modes? (Nov/Dec 2017)**

- Immediate addressing mode
- Register addressing mode
- Base or displacement addressing mode
- PC-relative addressing mode
- Direct addressing mode

4. **State the need for indirect addressing mode. Give an example. (April/May 2017)**

In direct addressing mode, the length of the address field is usually less than the word length, thus limiting the address range. To overcome this, in indirect addressing the address field refer to the address of a word in memory, which in turn contains a full-length address of the operand.

Example: MOV AX, [BX]

5. **What is an instruction register? (Nov/Dec 2016)**

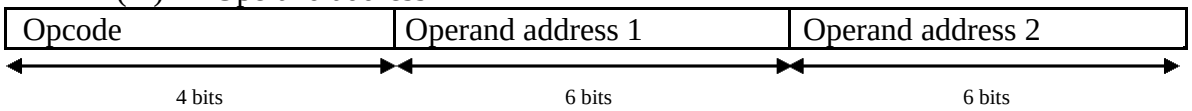
IR is the part of a CPU's control unit that holds the instruction currently being executed or decoded. The output of the IR is available to control circuits which generate the timing signals that control the various processing elements involved in executing the instruction.

6. **Give the formula for CPU execution time for a program. (Nov/Dec 2015)(Nov/Dec 2016)**

CPU execution time for a program = CPU clock cycle for a program / Clock Rate

7. **How to represent instruction in a computer system? (May/June 2016)**

(i) Opcode
(ii) Operand address 1
(iii) Operand address 2



8. **Distinguish between auto increment and auto decrement addressing mode. (May/June 2016)**

- Auto Increment: The effective address of the operand is the contents of a register specified in the instruction. After accessing the operand, the contents of this register are automatically incremented to point to the next item in the list.
- Auto Decrement: The contents of a register specified in the instruction are first automatically decremented and are then used as the effective address of the operand.

9. What is instruction set architecture? (Nov/Dec 2015)

The Instruction Set Architecture (ISA) is the part of the processor that is visible to the programmer or compiler writer. The ISA serves as the boundary between software and hardware.

The ISA of a processor can be described using 5 categories:

- Operand Storage in the CPU
- Number of explicit named operands
- Operand location
- Operations
- Type and size of operands

10. List the eight great ideas invented by computer architects. (April/May 2015)

- Design for Moore's law
- Use abstraction to simplify designs
- Make the common case fast
- Performance via parallelism
- Performance via pipelining
- Hierarchy of memories
- Dependability via redundancy
- Performance via predication

11. Distinguish pipelining from parallelism. (April/May 2015)

- Parallelism is simply multiple operations happening at the same time.
- Pipelining is a particular arrangement of functions, so that different portions of an operation flow through a particular set of sub-functions, with the sub-functions happening in parallel.

12. State Amdahl's Law. (Nov/Dec 2014)

Amdahl's law states that in parallelization, if P is the proportion of a system or program that can be made parallel, and 1-P is the proportion that remains serial, then the maximum speedup that can be achieved using N number of processors is 1/((1-P)+(P/N).

13. Brief about relative addressing mode with an example. (Nov/Dec 2014)

The PC-relative addressing mode is used to load a register with a value stored in program memory which is a short distance away from the current instruction. It can be seen as a special case of the "base plus offset" addressing mode, one that selects the program counter (PC) as the "base register".

Example: JNZ BACK

14. What are the functions of control unit?

The control unit co-ordinates and controls the activities among the functional units. The basic function of control unit is to fetch the instructions stored in the main memory, identify the operations, the devices involved in it and generate control signals to execute the desired operations.

15. Define throughput and throughput rate.

Throughput is defined as the total amount of work done in a given time. Throughput rate is defined as the rate at which the total amount of work done at a given time.

16. What do you mean by response time?

Response time or execution time is the total time required for the computer to complete a task including disk accesses, memory accesses, I/O activities, operating system overhead etc.

17. State the advantages of multiprocessor system.

- Improves cost / performance ratio of the system.
- Tasks are divided among the modules.
- If fault occurs in one processer, a second processor can take the responsibility of performing the task which improves the reliability of the system.

18. List various instruction formats with example.

- Three –address instruction – ADD A,B,C
- Two –address instruction – ADD A,B
- One address instruction – ADD A
- Zero address instruction - CMA

19. List the various elements of instruction.

- Operation Code
- Source / Destination operand
- Source / operand address
- Destination operand address
- Next instruction address

20. List the various technologies used in computers.

- Vacuum Tubes
- Transistors
- Integrated Circuits
- VLSI (Very Large Scale Integrated Circuits)
- ULSI (Very Large Scale Integrated Circuits)

PART: B

1. Explain various instruction formats and illustrate the same with an example. **(Nov/Dec 2017)**
2. Explain with an example about the operations and operands of the computer hardware. **(Nov/Dec 2014)**
3. Explain in detail the various components of computer system with neat diagram. **(Nov/Dec 2016) (May/ June 2016) (Nov/Dec 2015) (Nov/Dec 2014)**
4. Explain the different types of addressing modes with suitable examples. **(Nov/Dec 2016) (May/ June 2016) (Nov/Dec 2015) (April/May 2015) (April/May 2017)**
5. Discuss about the various techniques to represent instructions in a computer system. **(April/May 2015)**
6. State the CPU performance equation and discuss the factors that affect the performance. **(Nov/Dec 2014)**
7. Assume a two address format specified as source, destination. Examine the following sequence of instructions and explain the addressing modes used and the operation done in every instruction..
    a. Move (R5)+, R0
    b. Add (R5)+, R0
    c. Move R0, (R5)
    d. Move 16(R5), R3
    e. Add #40,R5

8. Consider the computer with three instruction classes and CPI measurements are given below. Instruction counts for each instruction class for the same program from two different compilers are given. Assume that the compiler's clock rate is 4GHz. Which code sequence will execute faster according to execution time? Which code sequence executes the most instructions? What is the CPI for each sequence? **(Nov/Dec 2014)**

Code from	CPI for this Instruction Class		
	A	B	C
CPI	1	2	3
Code from	Instruction Count for each class		
	A	B	C
Compiler 1	2	1	2
Compiler 2	4	1	1

9. Explain the eight ideas invented for computer design.
10. Explain the important measure of the performance of a computer and derive the basic performance equation. **(April/May 2017)**
11. What is the disadvantage of ripple carry addition and how it is overcome in carry look ahead adder and draw the logic circuit CLA. **(Nov/Dec 2016)**

UNIT: II
ARITHMETIC FOR COMPUTERS
PART:A

1. **What are the rules to perform addition on floating point numbers? (April/May 2017)**
   - Rewrite the smaller number such that its exponent matches with the exponent of the larger number.
   - Add the mantissas
   - Put the result in Normalized Form
   - Round the result

2. **Subtract $(11010)_2$-$(10000)_2$ using 1's complement and 2's complement method. (April/May 2017)**

1's complement method:

The 1's complement of 10000 is 01111

	1	1	0	1	0	+
	0	1	1	1	1	
1	0	1	0	0	1	+
Add carry					1	
	0	1	0	1	0	

$(11010)_2$-$(10000)_2$ using 1's complement method is $(01010)_2$

2's complement method:

	1	0	0	0	0	
1's complement	0	1	1	1	1	
					1	
2's complement	1	0	0	0	0	
	1	1	0	1	0	+
	1	0	0	0	0	
1	0	1	0	1	0	

Ignore carry, so $(11010)_2$-$(10000)_2$ using 2's complement method is $(01010)_2$

3. **Subtract $(11011)_2$-$(10011)_2$using 2's complement. (Nov/Dec 2017)**

	1	0	0	1	1	
1's complement	0	1	1	0	0	
					1	
2's complement	0	1	1	0	1	
	1	1	0	1	1	+
	0	1	1	0	1	
1	0	1	0	0	0	

Ignore carry, so $(11011)_2$-$(10011)_2$using 2's complement is $(01000)_2$

4. **Divide $(1001010)_2 \div (1000)_2$. (Nov/Dec 2017)**

```
          1001
      ____________
1000 |
     | 1001010
       1000
       ----
         1010
         1000
         ----
            10
```

Quotient: $(1001)_2$
Remainder: $(10)_2$

5. **What is a guard bit and what are the ways to truncate the guard bits? (Nov/Dec 2016)** Guard bits are extra bits which are processed during the intermediate steps to yield maximum accuracy in the final results.

Ways to truncate the guard bits:

- Chopping
- Von Neumann rounding
- Rounding

6. **What is arithmetic overflow? (Nov/Dec 2016)**

Overflow in addition:

Overflow occurs in addition when adding two numbers that have the same sign. Overflow in subtraction:

Overflow occurs in subtraction

- When a negative number is subtracted from a positive number and a negative result is generated.
- When a positive number is subtracted from a negative number and a positive result is generated.

7. **Define ALU. (May/June 2016)**

ALU is Arithmetic and Logic Unit. It is responsible for performing arithmetic operations such as addition, subtraction, multiplication, division and logical operations such as AND, OR, NOT etc.

8. **What is sub word parallelism? (May/June 2016)(April/May 2015)**

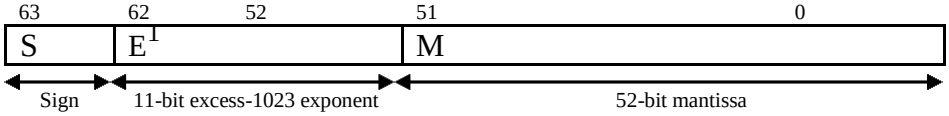
Subword Parallelism is a technique that enables the full use of word-oriented data paths when dealing with lower-precision data. It is a form of low-cost, small-scale SIMD parallelism.

9. What are the overflow/underflow conditions for floating point addition and subtraction? (Nov/Dec 2015)(April/May 2015)

In a single precision floating point number, if the number requires an exponent less than -126 or in a double precision floating point number if the number requires an exponent less than -1022 to represent its normalized form then underflow occurs.

In a single precision floating point number, if the number requires an exponent greater than +127 or in a double precision floating point number if the number requires an exponent greater than +1023 to represent its normalized form then overflow occurs.

10. State the representation of double precision floating point number. (Nov/Dec 2015)



11. Define Little Endian arrangement. (Nov/Dec 2014)

The least significant byte (the "little end") of the data is placed at the byte with the lowest address. The rest of the data is placed in order of the next three bytes in memory.

12. What is DMA? (Nov/Dec 2014)

Direct memory access (DMA) is a method that allows an Input/Output (I/O) device to send or receive data directly to or from the main memory, bypassing the CPU to speed up memory operations. The process is managed by a chip known as a DMA controller (DMAC).

13. Discuss the role of Booth's algorithm in the design of fast multipliers.

To speed-up the multiplication process in the Booth's algorithm a technique called bit-pair recording is used. It is also called modified Booth's algorithm. It halves the maximum number of summands. In this technique, the Booth-recorded multiplier bits are grouped in pairs. Then each pair is represented by its equivalent single bit multiplier reducing total number of multiplier bits to half.

14. Define floating point single and double precision standard.

The 32-bit standard representation (single-precision representation) occupies a single 32-bit word. The 32-bits are divided into three fields as shown below:

(Field 1) Sign 1 – bit
(Field 2) Exponent 8 – bits
(Field 3) Mantissa 23 - bits

The 64–bit standard representation (double – precision representation**)** occupies two 32-bit words. The 64-bits are divided into three fields as shown below:

(Field 1) Sign 1 – bit
(Field 2) Exponent 11 – bits
(Field 3) Mantissa 52 – bits

15. What is 2's complement of numbers?

The 2's complement is the binary number that results when we add 1 to the 1's complement. It is given as 2' complement = 1's complement + 1.

16. What is half adder?

The logic circuit which performs addition of two binary bits is called a half-adder.

17. What is full adder?

The circuit which performs addition of three bits (two significant bits and a previous carry) is a full-adder.

18. Give the full – adder circuit truth table.

INPUTS			OUTPUTS	
A	B	C_{IN}	SUM	CARRY
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

19. List the two division algorithms.

The two division algorithms are:

a. Restoring division algorithm
b. Non restoring division algorithm

20. What is a Carry Look-ahead adder?

A carry look-ahead adder is a type of adder in digital logic. It improves speed by reducing the amount of time required to determine the carry bits.

PART: B

1. Illustrate division algorithm with an example. **(Nov/Dec 2017)**
2. (a) Add the numbers (0.5)10 and (0.4375)10 using floating point addition.
   (b) Multiply the numbers (0.5)10 and (0.4375)10 using floating point multiplication. **(Nov/Dec 2017)**
3. Explain Booth's multiplication algorithm with suitable example. **(Nov/Dec 2016) (May/June 2016) (Nov/Dec 2015) (April/May 2015) (April/May 2017)**
4. Discuss in detail about division algorithm in detail with diagram and examples. **(Nov/Dec 2016) (Nov/Dec 2015) (April/May 2017)**
5. Explain briefly about floating point addition and subtraction algorithms.**(May/June 2016) (April/May 2015) (April/May 2017)**
6. Multiply the following pair of signed numbers using Booths's bit pair recoding of the multiplier A=+13(Multiplicand) and B=-6 (Multiplier) **(Nov/Dec 2014)**
7. Briefly explain carry lookaheadadder. **(Nov/Dec 2014)**
8. Divide $(12)_{10}$ by $(3)_{10}$ using the restoring and non-restoring division algorithm with step by step intermediate results and explain.**(Nov/Dec 2014)**
9. What is meant by sub word parallelism? Explain. **(April/May 2017)**

UNIT III
PROCESSOR AND CONTROL UNIT

PART A

1. Mention the various types of pipelining. (Nov/Dec 2017) The various types of pipelining are

- Instruction pipeline
- Operation pipeline
- Multi-issue pipeline

2. **Mention the various phase in executing an instruction. (Nov/Dec 2017)** The various phases in executing an instruction are
    - Fetch
    - Decode
    - Execute
    - Memory Access
    - Write Back

3. **What is meant by pipeline bubble? (Nov/Dec 2016)**

    Pipeline bubble or pipeline stall is a delay in execution of an instruction which occurs in an instruction pipeline inorder to resolve a hazard. A bubble is represented in the execution stage as a NOP instruction, which has no effect other than to stall the instructions being executed in the pipeline.

4. **What is a data path? (Nov/Dec 2016)**

    A datapath is a collection of functional units such as arithmetic logic units or multipliers that perform dataprocessing operations, registers, and buses. It composes the central processing unit (CPU) along with the control unit.

5. **What are the advantages of pipelining?(May/June 2016)**

    The cycle time of the processor is reduced and it increases the instruction throughput. If pipelining is used, the CPU arithmetic logic unit can be designed faster.

6. **What is exception? (May/June 2016)(Nov/Dec 2014)**

    Exceptions are internally generated unscheduled events that disrupt program execution and they are used to detect overflow. Examples for exception are arithmetic overflow, invoking the operating system from user program and using an undefined instruction.

7. **What is a hazard? What are its types?(Nov/Dec 2015)**

    Any situation that prevents the next instruction in the instruction stream from executing during its designated cycle is called a hazard. Various types of hazard are: Structural hazard, Data hazard and Control hazard.

8. **What is meant by branch prediction?(Nov/Dec 2015)**

    Branch prediction is a method of resolving a branch hazard that assumes a given outcome for the branch and proceeds from that assumption rather than waiting to ascertain the actual outcome. Two branch prediction strategies are:
    - Static branch prediction
    - Dynamic branch prediction

9. **What is a branch prediction buffer?(April/May 2015)**
    - Branch prediction buffer also called branch history table.
    - It is a small memory that is indexed by the lower portion of the address of the branch instruction.
    - It contains one or more bits indicating whether the branch was recently taken or not.

10. What are R-Type instructions? (April/May 2015)

R (Register) Format

This divides the instruction into six fields as follows:

6	5	5	5	5	6
op	**rs**	**rt**	**rd**	**shamt**	**funct**

Where,

op = opcode

rs = identifier of first source register

rt = identifier of second source register

rd = identifier of destination register

shamt = shift amount indicating how many bits the contents of a register must be shifted left or right (only used in shift instructions)

funct = distinguishes among R-type instructions as all R-type instructions have op = 0.

11. Name the control signals required to perform arithmetic operations. (April/May 2017)

ALU control lines	Function
0000	AND
0001	OR
0010	add
0110	subtract
0111	set on less than
1100	NOR

12. Define data hazard. Give an example for data hazard. (April/May 2017)

Data hazards occur when the pipeline changes the order of read/write accesses to operands so that the order differs from the order seen by sequentially executing instructions on the un-pipelined machine.

Example: ADD **R1**, R2, R3
SUB R4, **R1**, R5
AND R6, **R1**, R7
OR R8, **R1**, R9
XOR R10, **R1**, R11

13. What is the need for speculation? (Nov/Dec 2014)

Speculation is the technique which is needed to keep the instruction execution at high rate by using prediction based on program structure and profile. There are two types of speculated execution of instructions:

- Compiler speculation.
- Hardware based speculation.

14. What are the instructions set available in MIPS architecture?

- Memory reference instruction set
- Arithmetic logical instruction set
- Branch and jump instruction

15. What is meant by program counter?

Program counter is a register containing the address of the instruction in the program being executed.

16. **What are the units needed to implement MIPS load and store instructions?** Four units that are needed to implement MIPS load and store instruction are
    - Register file
    - ALU
    - Data memory unit
    - Sign extension unit

17. **What are the ways in which pipelining can be implemented?**
    - Single cycle implementation
    - Multiple cycle implementation

18. **How data hazards are resolved?**
    - Forwarding method is used to resolve the data hazards. It is also called bypassing.
    - Forwarding is a method of resolving data hazard by retrieving the missing data element from internal buffers rather than waiting for it to arrive from registers or memory.

19. **How control hazards are resolved?**
    - Control hazard or branch hazard can be resolved using branch prediction method.
    - Branch prediction is a method of resolving a branch hazard that assumes a given outcome for the branch and proceeds from that assumption rather than waiting to ascertain the actual outcome.

20. **What is meant by dynamic branch prediction?**
    Dynamic branch prediction is a prediction of branches at runtime using runtime information.

PART: B

1. Explain in detail the operation of the data path and its control. **(Nov/Dec 2017)(Nov/Dec 2014)**
2. Explain the pipeline hazard in detail. **(Nov/Dec 2017)(May/June 2016) (April/May 2015) (Nov/Dec 2014)(April/May 2017)**
3. Why is branch prediction algorithm needed? Differentiate between the static and dynamic techniques. **(Nov/Dec 2016)**
4. Explain how the instruction pipeline works? What are the various situations where an instruction pipeline can stall? **(Nov/Dec 2016) (May/June 2016)(Nov/Dec 2015) (April/May 2017)**
5. Explain the basic MIPS implementation with necessary multiplexer and control lines. **(Nov/Dec 2015)**
6. Explain in detail how exceptions are handled in MIPS architecture. **(April/May 2015)**

UNIT IV
PARALLELISM

PART A

1. **Define strong scaling and weak scaling.** **(Nov/Dec 2017)(April/May 2015)**

   In strong scaling, speed up is achieved on a multiprocessor without increasing the size of the problem. In weak scaling, speed up is achieved on a multiprocessor by increasing the size of the problem proportional to the increase in the number of processors.

2. **Difference between fine-grained multi-threading and coarse grained multi- threading.** **(Nov/Dec 2017) (May/June 2016)**

Fine grained multithreading	Coarse grained multithreading
Fine grained multithreading is a version of hardware multithreading that implies switching threads after every instruction.	Coarse grained multithreading is a version of hardware multithreading that implies switching between thread only after significant events such as last level cache miss.

3. **What is instruction level parallelism?** **(Nov/Dec 2016)(May/June 2016)(Nov/Dec 2015)(April/May 2017)**

   Instruction level parallelism is a kind of parallelism that executes the instruction in parallel way in order to reduce the execution time. It uses multiple functional units and multiple processors.

4. **What is multi-threading?** **(Nov/Dec 2016)(Nov/Dec 2014)**

   Multi-threading is a process which divides the instruction stream into several smaller streams called threads such that the threads can be executed in parallel. Multi-threading may be

   - Implicit multithreading
   - Explicit multithreading

5. **Distinguish implicit multi-threading and explicit multi-threading.**(April/May 2017)

   Implicit multi-threading is concurrent execution of multiple threads extracted from single sequential program. Explicit multi-threading concurrently execute instructions from different explicit threads. It interleaves instructions from different threads on shared pipelines or parallel execution on parallel pipelines.

6. **What is Flynn's classification?** **(Nov/Dec 2014)**

   Flynn's classification was first proposed by Michael J. Flynn in 1972. Flynns introduced the concept of instruction and data streams for categorizing computers. The four categories are:

   - Single Instruction Stream – Single Data Stream (SISD)
   - Single Instruction Stream – Multiple Data Stream (SIMD)
   - Multiple Instruction Stream – Single Data Stream (MISD)
   - Multiple Instruction Stream – Multiple Data Stream (MIMD)

7. **Define super scalar processor.** **(Nov/Dec 2015)**

   A superscalar processor is a CPU that implements a form of parallelism called instruction-level parallelism within a single processor. It can execute more than one instruction during a clock cycle by simultaneously dispatching multiple instructions to different execution units on the processor. It allows for more throughput.

8. Compare UMA and NUMA multiprocessors. (April/May 2015)

UMA multiprocessors	NUMA multiprocessors
Uniform Memory Access multiprocessor	Non-Uniform Memory Access multiprocessor
A multiprocessor in which latency to any word in main memory is about the same no matter which processor requests the access.	A type of single address space multiprocessor in which some memory accesses are much faster than others depending on which processor asks for which word.
It shares the physical memory uniformly.	Here the main memory is divided and attached to different microprocessors or to different micro controllers on the same chip.

9. What is meant by task level parallelism?

Task level parallelism is also called as process level parallelism. Task level parallelism utilizes multiple processors by running independent programs simultaneously.

10. What is a multicore microprocessor?

A multiprocessor contains multiple processors (cores) in a single integrated circuit which is called multicore multiprocessor. It is a single computing component that contains two or more distinct cores in the same physical package.

11. What are the challenges includes in parallel programming?

Parallel programming challenges includes scheduling, partitioning the work into parallel pieces, balancing the load, time to synchronize and overhead for communication.

12. What are the advantages of SIMD?

- Cost of the control unit over dozens of execution unit.
- It has reduced instruction band width and space.
- It needs only one copy of the code that is being executed simultaneously.

13. What is MISD?

Multiple Instruction-stream Single Data-stream (MISD) processor is a stream processor that perform a series of computations on a single data stream in a pipelined fashion.

14. What is MIMD?

Multiple Instruction-stream Multiple Data-stream (MIMD) is a multiprocessor, which has more than one CPU and the ability to execute several programs simultaneously.

15. What is the basic principle of vector architecture?

Basic principle of vector architecture is to collect data elements from memory put the data into a large set of registers, operate on them sequentially in registers using pipeline execution units and then write the results back to memory.

16. Write the advantages and disadvantages of fine grained multithreading Advantage

It can hide the throughput losses that arise from both short and long stalls because instruction from other threads can be executed when one thread stalls.

Disadvantage

It slows down the execution of the individual threads and it delays the execution of thread instructions.

17. Write the advantages and disadvantages of coarse grained multithreading.

Advantage

It is more useful for reducing the penalty of high cost stalls. Disadvantage

It is limited in its ability to overcome throughput losses especially from shorter stalls.

18. What is simultaneous multithreading?

Simultaneous multithreading is a variation on hardware multithreading that uses the resources of a multiple issue, dynamically scheduled micro architecture.

19. What is a multiple issue?

A multiple issue is a scheme used to place multiple instructions in one clock cycle. It has two types

- Static multiple issue
- Dynamic multiple issue

20. What is static multiple issue?

Static multiple issue is an approach of implementing a multiple issue processor where many decisions are made by the complier before execution.

21. What is dynamic multiple issue?

Dynamic multiple issue is an approach of implementing a multiple issue processor where many decisions are made during execution by the processor.

PART: B

1. Explain with diagrammatic illustration Flynn's classification.**(Nov/Dec 2017)(May/June 2016)(Nov/Dec 2015)(April/May 2015)(April/May 2017)**
2. Describe Simultaneous Multithreading (SMT) with an example. **(Nov/Dec 2017)**
3. Explain in detail about the memory technologies. **(Nov/Dec 2017)(April/May 2015)(Nov/Dec 2014)**
4. Discuss shared memory multiprocessor with a neat diagram. **(Nov/Dec 2016)**
5. Write short notes on (a) Hardware Multi-threading (b) Multicore processors. **(May/June 2016)(Nov/Dec 2015)(April/May 2015)(Nov/Dec 2014)(April/May 2017)**
6. Explain instruction level parallel processing. State the challenges of parallel processing. **(Nov/Dec 2014)**
7. Discuss the challenges in parallel processing with necessary examples. **(April/May 2017)**

UNIT V
MEMORY AND I/O SYSTEMS

PART A

1. What is meant by address mapping? (Nov/Dec 2016)

The correspondence between the main memory blocks and those in the cache is specified by address mapping. There are three commonly used methods to translate main memory addresses to cache memory addresses. They are:

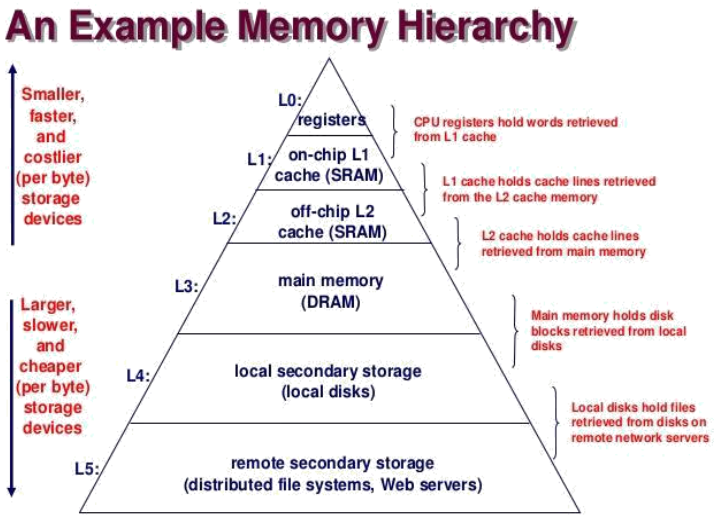
- Direct mapping
- Associative mapping
- Set-associative mapping

2. What is cache memory? (Nov/Dec 2016)

A Cache memory is a small and very fast temporary storage memory. It is designed to speed up the transfer of data and instructions. It is faster than RAM and the data/instructions that are most recently or most frequently used by CPU are stored in cache.

3. **Define memory hierarchy.** **(May/June 2016)(April/May 2015)**

In computer architecture, the memory hierarchy separates computer storage into a hierarchy based on response time. Since response time, complexity, and capacity are related, the levels may also be distinguished by their performance and controlling technologies.



4. **State the advantages of virtual memory.** **(May/June 2016)**

- Virtual memory allows processes whose aggregate memory requirement is greater than the amount of physical memory, as infrequently used pages can reside on the disk
- Virtual memory allows speed gain when only a particular segment of the program is required for the execution of the program
- It is very helpful in implementing multiprogramming environment.

5. **What are the various memory technologies?** **(Nov/Dec 2015)**

- SRAM (Static Random Access Memory)
- DRAM (Dynamic Random Access Memory)
- ROM (Read Only Memory)
- Flash Memory
- Magnetic Disk

6. **Point out how DMA can improve I/O speed.** **(April/May 2015)**

DMA allow the peripherals to directly communicate with each other using the memory buses, removing the intervention of the CPU. During DMA the CPU is idle and it has no control over the memory buses. So, the DMA controller takes control over the buses to manage the transfer directly between the I/O devices and the memory unit for improving the speed.

7. **Define memory interleaving.**(April/May 2017)

Memory interleaving is the technique used to increase the throughput. The memory system is split into independent banks, which can answer read or write requests independents in parallel. There are two-address format for memory interleaving the address space. They are:

- Low order interleaving
- High order interleaving

8. **Summarize the sequence of events involved in handling an interrupt request from a single device.(April/May 2017)**
   - The device raises an interrupt request.
   - The processor interrupts the program currently being executed.
   - Interrupts are disabled by changing the control bits in the Processor Status Register.
   - The device is informed that its request has been recognized, and in response, it deactivates the interrupt-request signal.
   - The action requested by the interrupt is performed by the interrupt-service routine.
   - Interrupts are enabled and execution of the interrupted program is resumed.

9. **Differentiate Programmed I/O and Interrupt I/O.(Nov/Dec 2014)**

Programmed I/O	Interrupt I/O
The CPU manually checks if there are any I/O requests available periodically. If there are no I/O requests it keeps executing its normal workflow. If there is any I/O request, it handles the IO request.	The CPU doesn't need to manually check for I/O requests. When there is an I/O request available, the CPU is immediately notified using interrupts, and the request is immediately handled using interrupt service routines.

10. **What is the purpose of dirty/modified bit in cache memory? (Nov/Dec 2014)**

A dirty bit or modified bit is a bit that is associated with a block of computer memory and indicates whether or not the corresponding block of memory has been modified. Dirty bits are used by the CPU cache and in the page replacement algorithms of an operating system.

11. **What is virtual memory? (Nov/Dec 2017)**

Virtual memory is a memory management technique that is implemented using both hardware and software and it uses main memory as a cache for secondary storage. Techniques that automatically move program and data blocks into the physical main memory when they are required for execution are called virtual memory techniques.

12. **How many total bits are required for a direct mapped cache with 16KB of data and 4-word blocks, assuming a 32bit address? (Nov/Dec 2017)**

No. of cache lines= data memory size of cache /data size of 1 cache line

$= 16KB/4B = 16x1024/4 = 4096 = 2^{12}$ lines

No. of bits needed to represent cache line = $\log_2 (2^{12}) = 12$

bits No. of bits needed to represent a word in a line = $\log_2 4 =$

2 bits No. of bits needed for tag = 32-12-2 = 18 bits

Size of tag memory = No. of tag bits * No. of lines = $18 * 2^{12}$ bits = **72K bits** Size of data memory = 16 KB = 16 x 8 = **128K bits**

Total memory needed for cache = 128 K bits + 72K bits = 200Kbits

13. **Define hit ratio. (Nov/Dec 2015)**

The hit ratio is the fraction of accesses which are a hit. The miss ratio is the fraction of accesses which are a miss. It holds that miss rate = 1 − hit rate. The (hit/miss) latency (also known as access time) is the time it takes to fetch the data in case of a hit/miss.

14. **Define hit rate and miss rate.**

The fraction of memory accesses found in a level of the memory hierarchy is called hit rate. The fraction of memory accesses not found in a level of the memory hierarchy is called miss rate.

15. What is TLB?

Translation Look-aside Buffer (TLB) is a cache that keeps track of recently used address mapping which tries to avoid an access to the page table.

16. What are the difference between DRAM and SRAM memory technology?

S.No	DRAM	SRAM
1	Single transistor is used to access the stored charge.	It uses six to eight transistors per bit.
2	It is much denser and cheaper per bit than SRAM.	It is high cost per bit.
3	It need periodic refresh.	It neednot to be refreshed.
4	The value cannot be kept indefinitely.	The value can be kept indefinitely.

17. What are the methods used to improving cache performance?

There are two different techniques available for improving cache performance:

- Reducing the miss penalty by adding an additional level to the hierarchy.
- Reducing the miss rate by reducing the probability that two different memory blocks will content for the same cache location.

18. Define interrupts.

Interrupt is a process that causes a CPU to temporarily transfer control from its current program to another program. It improves the computer's IO performance.

19. What are the two I/O interfacing techniques?

The two I/O interfacing techniques are

- Memory mapped I/O
- I/O mapped I/O

20. Differentiate memory mapped I/O and I/O mapped I/O.

Memory mapped I/O	I/O mapped I/O
In this technique, the total memory address space is partitioned and part of this space is allocated to I/O addressing.	In this technique, different space is allocated for I/O address apart from total memory space.

PART: B

1. What is cache memory? How to improve the cache performance? Discuss. **(Nov/Dec 2017)(May/June 2016)(Nov/Dec 2014)(April/May 2017)**
2. Discuss DMA controller with block diagram. **(Nov/Dec 2016)(May/June 2016)(Nov/Dec 2015)(Nov/Dec 2014)(April/May 2017)**
3. Discuss the steps involved in the address translation of virtual memory with necessary block diagram. **(Nov/Dec 2016)(Nov/Dec 2015)(April/May 2015)(April/May 2017)**
4. Design and explain parallel priority interrupt hardware for a system with eight interrupt sources.**(Nov/Dec 2016)**
5. Explain in detail about any two standard input and output interfaces required to connect the I/O device to the bus. **(Nov/Dec 2014)**