QUESTION BANK

Name of the Department : Electronics and CommunicationEngineering

Subject Code & Name : EC8552- Computer Architecture And Organization

Year & Semester : III& V

UNIT I BASIC STRUCTURE OF A COMPUTER SYSTEM

PART-A

1. What are the eight great ideas invented by computer architects?(May 2015)

- Design for Moore'sLaw
- Use abstraction to simplify design
- Make the common casefast
- Performance viaParallelism
- Performance viaPipelining
- Performance viaPrediction
- Hierarchy ofMemory
- Dependability viaRedundancy

2. Define power wall.

- Old conventionalwisdom
- Power isfree
- Transistors areexpensive
- New conventional wisdom: —Powerwall
- Powerexpensive
- Transistors —free (Can put more on chip than can afford to turnon)

3. What is uniprocessor?

A **uniprocessor system** is defined as a computer system that has a single central processing unitthat is used to execute computer tasks and the term uniprocessor is used to distinguish the class of computers where all processing tasks share a single CPU.

4. What is multicore processor?

A **multi-core processor** is a single computing component with two or more independent actual central processing units (called "cores"), which are the units that read and execute instructions.

The instructions are ordinary CPU instructions such as add, move data, and branch, but he multiple cores can run multiple instructions at the same time, increasing overall speed for programs amenable to parallel computing.

5. State the need for indirect addressing mode. Give an example. (Apr 2017)

This mode uses a register to hold the actual address that identifies either the source or the destination to be used in the data move. What we have here is a pointer. R0 and R1 (when they contain a valid internal memory address) can be used for internal memory.

6. Define multiprocessing.

Multiprocessing is the use of two or more central processing units (CPUs) within a single computer system. The term also refers to the ability of a system to support more than one processor and/or the ability to allocate tasks between them.

7. Differentiate supercomputer and mainframe computer.

A computer with high computational speed, very large memory and parallel structured hardware is known as a **supercomputer. EX: CDC 6600**.

Mainframe computer is the large computer system containing thousands of IC‘s. It is a room-sized machine placed in special computer centers and not directly accessible to average users. It serves as a central computing facility for an organization such as university, factory or bank.

8. Write the basic functional units of computer? (or) List the major components of a computer system? (Apr 2017)

The basic functional units of a computer are Input unit, Output unit, Memory unit, ALU , Control unit and Data path unit.

9. What is instruction register? (Nov 2016)

The **instruction register (IR)** holds the instruction that is currently being executed. Its output is available to the control circuits which generate the timing signals that control the various processing elements involved in executing the instruction.

10. What is program counter?

The **program counter (PC)** keeps track of the execution of a program. It contains the

memory address of the next instruction to be fetched and executed.

11. What is processor time?

The sum of the periods during which the processor is active is called the processor time. It doesn't count I/O or time spent running other programs. It can be broken up into system time, and user time.CPU time = Ncycles * tclock = Ncycles / fclock.

12. What is superscalar execution?

The multiple functional units are used to create parallel paths through which different instructions can be executed in parallel.

So it is possible to start the execution ofseveral

instructions in every clock cycle. This mode of operation is called **superscalar execution**.

13. What is RISC and CISC?

Reduced Instruction Set Computers (RISC) is a microprocessor that is designed to perform a **smaller number of types of computer instructions** so that it can operate at a higher speed (perform more millions of instructions per second, or MIPS).

Complex Instruction Set Computers (CISC) is a processor design where **single instructions** can execute several low-level operations (such as a load from memory, an arithmetic operation, and a memory store) or are capable of multi-step operations or addressing modes within singleinstructions.

14. List out the methods used to improve system performance?

The methods used to improve system performance are

- Processorclock
- Basic PerformanceEquation
- Pipelining
- Clock rate
- Instruction set
- Compiler

15. Define addressing modes and its various types.

The different ways in which the location of a operand is specified in an instruction is referred toas addressing modes. The various types are Immediate Addressing, Register Addressing, Base or Displacement Addressing, PC-Relative Addressing, Pseudo direct Addressing.

16. Define Base or Displacement mode addressing.

In base or displacement mode addressing, the operand is in a memory location whose address isthe sum of a register and a constant in the instruction. Eg. lw $t0,32($s3).

17. Define register mode addressing.

In register mode addressing, the name of the register is used to specify the operand. Eg. Add $s3, $s5,$s6.

Advantage:

Only a small address field is needed in the instruction and no memory is referenced.

Disadvantage:

Address space is very limited.

18. Define immediate mode addressing.

In immediate mode addressing, the operand is given explicitly in the instruction. Eg. Add $s0,

$s1,20.

Advantage

No memory reference other than the instruction fetch is required to obtain the operand.

Disadvantage

The size of the number is restricted to the size of the address field

19. Define Relative mode addressing.

The **relative addressing mode** is similar to the indexed addressing mode with the exception that the PC holds the base address. This allows the storage of memory operands at a fixed offset

from the current instruction and is useful for _short‘ jumps. Example: jump 4

20. State Amdahl’s Law.

Amdahl’s Law tells us the improvement expected from specific enhancements. The performance improvement or speedup due to improvement is calculated asfollows

Speedup= Execution time before improvement/ Execution time after improvement

21. Distinguish pipelining from parallelism.

Pipelining is a method of increasing system performance and throughput. It takes advantage of the inherent parallelism in instructions. Instructions are divided into 5 stages: IF, ID, EX, EME, WB.Parallelism means using more hardware for the executing the desired task. In

Parallel computing more than one processor are running in parallel. It increases performance but the area also increases.

22. What is Instruction set architecture?

The **ISA** serves as the boundary between the software and hardware. It is the structure of a computer that a machine language programmer (or a compiler) must understand to write a correct (timing independent) program for that machine. It also specifies a processor‘s functionality • what operations it supports • what storage mechanisms it has & how they are accessed • how the programmer/compiler communicates programs to processor.

23. Write the equation for the dynamic power required per transistor.

The power required per transistor is just the product of energy of a transition and the frequency of transistions:

Power α 1/2 * Capacitive load * volatage^2 * Frequency Switched

24. Classify the instructions based on the operations they perform and give one example to each category.

Instructions are divided into three types: R-type, I-Type and J- Type instruction

The following are the three formats used for the core instruction set:

Type	-31- format (bits) -0-					
R	opcode (6)	rs (5)	rt (5)	rd (5)	shamt (5)	funct (6)
I	opcode (6)	rs (5)	rt (5)	immediate (16)		
J	opcode (6)	address (26)				

25. How CPU execution time for a program is calculated?

CPU execution time for a program is given by the formula

CPU Execution time=Instruction Count * Clock cycles per instruction * Clock cycle time.

PART-B

1.Explain in detail about the eight ideas of computer architecture.

2. Describe the MIPS Instruction set in detail with suitable examples.

3. What is an addressing mode? What is the need for addressing in a computer system? Explain the various addressing modes with suitable examples.

4. Discuss about the various techniques to represent instructions in a computer system

5. What are the various logical operations and explain the instructions supporting the logical operations?

6. What are the various control operations and explain the instructions supporting the control operations?

7. Explain the different types of operands used in MIPS Assemble Language with examples.

8. 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 everyinstruction.

Move(R5)+, R0
Add (R5)+, R0
Move R0, (R5)
Move 16(R5), R3
Add #40, R5
Move 16(R5), R3

9..Consider the computer with three instruction classes and CPI measurements as given below and Instruction counts for each instruction class for the same program from two different compilers are given. Assume that the computer's clock rate is 4GHZ. Which Code sequence will execute faster according to executiontime?

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

(Nov 2014)

10. Consider three different processor P1, P2 and P3 executing the same instruction set. P1 has a 3 GHz clock rate and a CPI of 1.5 . P2 has a 2.5 GHz clock rate and a CPI of 1.0. P3 has a 4.0 GHz clock rate and has a CPI of 2.2.

A) Which processor has the highest performance expressed in instructions per second?

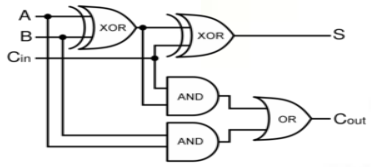
B) If the processors each execute a program in 10 seconds. Find the number of cycles and the number of instructions in each processor.

UNIT II ARITHMETIC FOR COMPUTERS

PART-A

1. Define Full Adder (FA) with logic diagram.

A **full adder** adds binary numbers and accounts for values carried in as well as out. A one-bit full adder adds three one-bit numbers, often written as *A*, *B*, and *C*in; *A* and *B* are the operands, and *C*in is a bit carried in (from a past addition).The full-adder is usually a component in a cascade of adders, which add 8, 16, 32, etc.



2. State the rule for floating point addition.

1)Choose the number with the smaller exponent and shift its mantissa right a number of steps equal to the difference in exponents.

2)Set the exponent of the result equal to the larger exponent.

3)Perform the addition on the mantissa and determine the sign of the result. Normalize the resulting value if necessary.

3. State the representation of double precision floating point number.

Double precision representation contains 11 bits, excess -1023 exponent E 'which has the range 1≤ E '≤ 2046 for normal values. This means that the actual exponent E is in range -1022 ≤ E ≤1023. The 53 bit mantissa provides a precision equivalent to about 16 decimal digits.

4. What is guard bit? What are the ways to truncate the guard bits?

Although the mantissa of initial operands is limited to 24 bits, it is important to retain extra bits, called as **guard bits.**

There are several ways to truncate the guard bits: **Chopping, VonNeumann rounding,Rounding.**

5. Define Booth Algorithm.

Booth's multiplication algorithm is a multiplication algorithm that multiplies two signed binary numbers in two's complement notation. Booth's algorithm can be implemented by

repeatedly adding (with ordinary unsigned binary addition) one of two predetermined values A and S to a product P, then performing a rightward arithmetic shift on P.

6. Why floating-point number is more difficult to represent and process than integer?

In floating point numbers we have to represent any number in three fields sign, exponent and mantissa. The IEEE 754 standard gives the format for these fields and according to format the numbers are to be represented. In case of any process the mantissa and exponent are considered separately.

7. When can you say that a number is normalized?

When the decimal point is placed to the right of the first (nonzero) significant digitthe number is said to be normalized.

8. What is arithmetic overflow?

In a computer, the condition that occurs when a calculation produces a result that is greater in magnitude than which a given register or storage location can store or represent. In a computer, the amount by which a calculated value is greater in magnitude than that which a given register or storage location can store or represent.

9. What is overflow and underflow case in single precision?

Underflow-The normalized representation requires an exponent less than -126

Overflow-The normalized representation requires an exponent greater than -126

10. What is Carry Save addition?

Using carry save addition, the delay can be reduced further still. The idea is to take 3 numbers that we want to add together, x+y+z, and convert it into 2 numbers c+s such that x+y+z=c+s, and do this in O (1) time. The reason why addition cannot be performed in O (1) time is because the carry information must be propagated. In carry save addition, we refrain from directly passing on the carry information until the very last step

11 .Subtract (11010)2-(10000)2 using 1's complement and 2's complement method.)

Answer:

110110 Minuend

101010 2's complement of subtrahend

Carry over 110 0 0 0 0 Result of addition

After dropping the carry-over, we get the result of subtraction to be 100000

12. Write Restoring and Non-Restoring division algorithm?

Restoring Division Algorithm:

- Shift A and Q left one binaryposition.
- Subtract M from A, and place the answer back inA.
- If the sign of A is 1, set q0 to 0 and add M back to A (that is, restore A);
- otherwise, set q0 to 1.

Non- Restoring Division Algorithm

- Do the following n times: If the sign of A is 0, shift A and Q left one bit position and subtract M from A;
- otherwise, shift A and Q left and add M to A. Now, if the sign of A is 0, set q0 to 1; otherwise, set q0 to0.
- If the Sign of A is 1, add M toA

13. Write the rules for add/sub operation on floating point numbers?

- Choose the number with the smaller exponent and shift its mantissa right a number of s steps equal to the difference inexponents.
- Set the exponent of the result equal to the largerexponent
- Perform addition / subtraction on the mantissa and determine the sign of theresult
- Normalize the resulting value, ifnecessary

14. Write the rules for multiply operation on floating point numbers?

- Add the exponents and subtract127.
- Multiply the mantissa and determine the sign of theresult.
- Normalize the resulting value, ifnecessary.
- Write the rules for divide operation on floating pointnumbers
- Subtract the exponents and subtract127.
- Divide the mantissa and determine the sign of theresult.
- Normalize the resulting value, ifnecessary.

15.Define Truncation.

To retain maximum accuracy, all extra bits during operation (called *guard bits*) are kept (e.g., multiplication). If we assume n=3 bits are used in final representation of a number, n=3extra guard bits are kept during operation. By the end of the operation, the resulting 2n=6 bits need to be truncated to n=3 bits by one of the three methods.

16. Explain how Boolean subtraction is performed?

Negate the subtrahend (i.e. in a-b, the subtrahend is b) then perform addition (2's complement)

17. Define Chopping.

There are several ways to truncate. The simplest way is to remove the guard bits and make no changes in the retained bits. This is called Chopping. Chopping discards, the least significant bits and retains the 24 most significant digits. This is easy to implement, and biased, since all values are rounded to-wards a lower mantissa value. The maximum rounding error is $0 \leq e < +1$ LSB.

18. Define Von Neumann Rounding.

If at least one of the guard bits is 1, the least significant bit of the retained bits is set to 1 otherwise nothing is changed in retained bits and simply guard bits are dropped.

19. What do mean by Subword Parallelism?

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.

20. How overflow occur in subtraction?

When overflow occurs on integer addition and subtraction, contemporary machines invariably discard the high-order bit of the result and store the low-order bits that the adder naturally produces. Signed integer overflow of addition occurs if and only if the operands have the same

sign and the sum has assign opposite to that of the operands.

21. Define generate and propagate function.

The generate function is given by $G_i=x_iy_i$ and the propagate function is givenas $P_i=x_i+y_i$.

22. What is excess-127 format?

Instead of the signed exponent E, the value actually stored in the exponent field is and Unsigned integer $E^{¢}=E+127$.This format is calledexcess-127.

23. What is floating point numbers?

In some cases, the binary point is variable and is automatically adjusted as computation proceeds. In such case, the binary point is said to float and the numbers are called floating point Numbers.

24. Write the IEEE 754 floating point format?

The IEEE 754 standard floating point representation is almost always an approximation of

the real number. The format is:$(-1)^s(1+\text{Fraction})x2^{(\text{Exponent-Bias})}$

25. What are the overflow/underflow conditions for addition and subtraction?

- When result cannot be represented in the allocated number of bits. Overflow occurs if the Result > Max value.
- Underflow occurs if the Result < Min value. Overflow can occur when two positive numbers are added and result is out of range.
- After addition, the result will become negative. Underflow can occur when two negative numbers are added and result is out of range.

After addition, the result will become positive. While adding a positive number with a negative number. No overflow or underflow canoccur.

Unsigned number representation using n-bits

- Overflow when result > 2^n-1.
- Underflow when result <0.

Signed number representation using n-bits

- Overflow when result > $2^{n-1}-1$.
- Underflow when result <-2^{n-1}.

PART-B

1. Explain the Booth's multiplication algorithm with suitable example (or) Explain the Booth's algorithm for multiplication of signed two's complement numbers. (May 2016,April-2018) (Nov 2016)(Nov/Dec-2019)

2. Explain the various methods of performing multiplication of n-bit numbers with suitable examples. (Apr 2017).

3. Discuss in detail about division algorithm in detail with diagram and examples. (Nov 2015, Nov 2016, Apr 2017,April-2018)

4. Explain how floating point addition is carried out in a computer system. Give an example for a binary floating point addition. (May 2015, Apr 2017)

5. Describe Subword parallelism in detail. (Apr 2017)

6. Explain in detail about the multiplication algorithm with suitable example and diagram. (Nov 2015, May 2015)

7. Draw and explain the block diagram of floating point adder – subtractor unit with an

example.

8. Multiply the following pair of signed nos. using Booth's bit-pair recoding of the multiplier. A=+13 (Multiplicand) and B= -6 (Multiplier). (Nov 2014)

9. Briefly explain Carry Look-ahead adder. (Nov 2014)

10. Divide $(12)_{10}$ by $(3)_{10}$ using the Restoring and Non- restoring division algorithm with step by step intermediate results and explain. (Nov 2014)

UNIT III PROCESSOR AND CONTROL UNIT

PART-A

1. What is pipelining?

The technique of overlapping the execution of successive instruction for substantial improvement in performance is called pipelining.

2. What is precise exception?

A precise exception is one in which all instructions prior to the faulting instruction are complete and instruction following the faulting instruction, including the faulty instruction; do not change thestate of the machine.

3. Define processor cycle in pipelining.

The time required between moving an instruction one step down the pipeline is a processor cycle.

4. What is meant by pipeline bubble? (Nov 2016)

To resolve the hazard the pipeline is stall for 1 clock cycle. A stall is commonly called a pipelinebubble, since it floats through the pipeline taking space but carrying no useful work.

5. What is pipeline register delay?

Adding registers between pipeline stages me adding logic between stages and setup and hold timesfor proper operations. This delay is known as pipeline register delay.

6. What are the major characteristics of a pipeline?

The major characteristics of a pipeline are:

1. Pipelining cannot be implemented on a single task, as it works by splitting multiple tasks into a number of subtasks and operating on themsimultaneously.

2. Thespeeduporefficiencyachievedbysuingapipelinedependsonthenumberofpipestages and the number of available tasks that can be subdivided.

7. What is data path? (Nov 2016, April/May-2018)

As instruction execution progress data are transferred from one instruction to another, often passing through the ALU to perform some arithmetic or logical operations. The registers, ALU, and theinterconnecting bus are collectively referred as the data path.

8. What is meant by data hazard in pipelining?

Any condition in which either the source or the destination operands of an instruction are not available at the time expected in the pipeline is called data hazard.

9.What is Instruction or control hazard?

The pipeline may be stalled because of a delay in the availability of an instruction. For example, thismay be a result of a miss in the cache, requiring the instruction to be fetched from the main memory. Such hazards are often called control hazards or instruction hazard.

10.What is the ideal CPI of a pipelined processor?

The ideal CPI of pipelined processor is almost always 1. Hence we can compute the pipelined CPI:

CPI pipelined = Ideal CPI + Pipleline stall clock cycles per instruction

= 1 + pipeline stall clock cycle per instruction.

11. What is side effect?

When a location other than one explicitly named in an instruction as a destination operand is affected, the instruction is said to have a side effect.

12. What do you mean by branch penalty?

The time lost as a result of a branch instruction is often referred to as branch penalty.

13. What is branch folding?

When the instruction fetch unit executes the branch instruction concurrently with the execution ofthe other instruction, then this technique is called branch folding.

14. What do you mean by delayed branching?

Delayed branching is used to minimize the penalty incurred as a result of conditional branch

instruction. The location following the branch instruction is called delay slot. The instructions in the delay slots are always fetched and they are arranged such that they are fully executed whether or not branch is taken.

15. What are the two types of branch prediction techniques available?

The two types of branch prediction techniques are

1) static branch prediction

2) dynamic branch prediction.

16. What is a hazard? What are its types?

Any condition that causes the pipeline to stall is called hazard. They are also called as stalls or bubbles. The various pipeline hazards are: Data hazard, StructuralHazard, ControlHazard

17. Name the control signals required to perform arithmetic operations

ALU Control	Input Function
111	and
111	or
010	add
110	sub
111	slt

18. What is branch Target Address?

The address specified in a branch, which becomes the new program counter, if the branch is taken. InMIPSthebranchtargetaddressisgivenbythesumoftheoffsetfieldoftheinstructionandthe address of the instruction following the branch.

19. What is an interrupt?

An exception is the one that comes from outside of the processor. There are two types of interrupt.They are imprecise interrupt and precise interrupt.

20. Define Pipeline speedup.

The ideal speedup from a pipeline is equal to the number of stages in the pipeline.

Speedup= Time per instruction on unpipelined machine/Number of pipe stages

21. What is meant by vectored interrupt?

An interrupt for which the address to which control is transferred is determined by the cause of the exception.

22. Define exception. Give one example of MIPS exception

The term exception is used to refer to any event that causes an interruption otherwise an unexpected change in the control flow. When an exception or interrupt occurs, the hardware begins executing code that performs an action in response to the exception. This action may involve killing a process, outputting a error message, communicating with an external device. There are following Exception

a. Execution of an undefined instruction and

b. Arithmetic overflow.

23. What is R-type instructions?

R instructions are used when all the data values used by the instruction are located in registers.

All R-type instructions have the following format:

OP rd, rs,rt. Where "OP" is the mnemonic for the particular instruction. *Rs*, and *rt* are the source registers, and *rd* is the destination register.

24. What is a branch prediction buffer?

The simplest thing to do with a branch is to predict whether or not it is taken. This helps in where the branch delay is longer than the time it takes to compute the possible target PC_s .

25. What is meant by branch prediction?

Branch Instructions may introduce branch penalty. To avoid it, prediction is done by two ways.

Static Branch prediction

The static branch prediction assumes that the branch will not take place and to continue to fetch instructions in sequential address order.

Dynamic Branch prediction

The idea is that the processor hardware assesses the likelihood of a given branch being taken by keeping track of branch decisions every time that instruction is executed. The execution history used in predicting the outcome of a given branch instruction is the result of the most recent execution of that instruction.

PART-B

1.Explain in detail about the basic MIPS implementation with suitable diagram.

2. What are the major components required to execute MIPS instruction while building a datapath.

3. Explain how the instruction pipeline works? What are the various situations where an instruction pipeline can stall? Illustrate with an example.

4. Explain the basic MIPS implementation with necessary multiplexers and control lines.

5. What is pipelining? Discuss about pipelined datapath and control.

6. What is data hazard? Explain the ways and means of handing it in pipelined datapath.

7. Why is branch prediction algorithm needed? Differentiate between the static and dynamic techniques (or) Describe the techniques for handling control hazards in pipelined datapath.

8. Explain in detail how exceptions are handled in MIPS architecture.

9. Explain Data path and its control in detail.

10. What is Hazard? Explain its types with suitable examples.

UNIT IV PARALLELISIM

PART-A

1.Differentiate between Strong scaling and weak scaling.

Strong scaling:

Speed-up is achieved on a multi-processor without increasing the size of the problem.

Weak scaling:

Speed-up is achieved on a multi-processor while increasing the size of the problem proportionally to the increase in the number of processors.

2.Define Single Instruction, Single Data stream (SISD.

A sequential computer which exploits no parallelism in either the instruction or data streams. Single control unit (CU) fetches single Instruction Stream (IS) from memory. The CU then generates appropriate control signals to direct single processing element (PE) to operate on single Data Stream (DS) i.e. one operation at a time. Examples of SISD architecture are the traditional uniprocessor machines like a PC.

3.Define Single Instruction, Multiple Data streams (SIMD).

A computer which exploits multiple data streams against a single instruction stream to perform operations which may be naturally parallelized. For example, an array processor.

4.Differentiate between Strong scaling and weak scaling.

Strong scaling:

Speed-up is achieved on a multi-processor without increasing the size of the problem.

Weak scaling:

Speed-up is achieved on a multi-processor while increasing the size of the problem proportionally to the increase in the number of processors.

5. Define Single Instruction, Single Data stream (SISD).

A sequential computer which exploits no parallelism in either the instruction or data streams. Single control unit (CU) fetches single Instruction Stream (IS) from memory. The CU then generates appropriate control signals to direct single processing element (PE) to operate on single Data Stream (DS) i.e. one operation at a time. Examples of SISD architecture are the traditional uniprocessor machines like a PC.

6. Define Single Instruction, Multiple Data streams (SIMD).

A computer which exploits multiple data streams against a single instruction stream to perform operations which may be naturally parallelized. For example, an array processor.

7. Define multithreading.

Multiple threads to share the functional units of 1 processor via overlapping processor must duplicate independent state of each thread e.g., a separate copy of register file, a separate PC, and for running independent programs, a separate page table memory shared through the virtual memory mechanisms, which already support multiple processes.

8. Distinguish implicit multithreading and explicit multithreading.

Explicit multithreading concurrently execute instructions from different explicit threads. Interleave instructions from different threads on parallel execution on parallel pipelines. Implicit multithreading is concurrent execution of multiple threads extracted from single sequential program.

9. Compare UMA and NUMA multiprocessors.

The main difference between the NUMA and UMA memory architecture is the location of the Memory.

The UMA architecture nodes have first and second cache memory levels joint with the processor, next levels of the memory hierarchy are "in the other side" of the interconnection network.

10. Define GPUs.

A programmable logic chip (processor) specialized for display functions. The GPU renders images, animations and video for the computer's screen. GPUs are located on plug-in cards, in a chipset on the motherboard or in the same chip as the CPU.

11. Define Cluster.

In a computer system, a cluster is a group of servers and other resources that act like a single system and enable high availability and, in some cases, load balancing and parallel processing.

12. Give examples of topology of the interconnection network:(Nov-Dec-2019)

- Single bus
- Ring
- Mesh
- N-cube
- Crossbar Network

i. The topology of the interconnection network can be different for data and address buses.

13. What is meant by memory-mapped I/O? (Nov/Dec 2018)

Memory-mapped I/O uses the same address space to address both memory and I/O devices. The memory and registers of the I/O devices are mapped to address values. So when an address is accessed by the CPU, it may refer to a portion of physical RAM, or it can instead refer to memory of the I/O device.

14. Define a super scalar processor.

Superscalar is an advanced pipelining technique that enables the processor to execute more than one instruction per clock cycle by selecting them during execution. Dynamic multiple-issue processors are also known as superscalar processors, or simply superscalars.

15. Define multithreading. (Nov 2014,Nov 2016)

Multiple threads to share the functional units of 1 processor via overlapping processor must duplicate independent state of each thread e.g., a separate copy of register file, a separate PC, and for running independent programs, a separate page table memory shared through thevirtual memory mechanisms, which already support multiple processes.

16. What is the need for Speculation? (Nov 2014)

It is one of the most important methods for finding and exploiting more ILP. It is an approach that allows the compiler or the process to guess about the properties of an instruction, so as to enableexecution to begin for other instructions that may depend on the speculated instruction.

17. Distinguish implicit multithreading and explicit multithreading. (Apr 2017)

Explicit multithreading concurrently execute instructions from different explicit threads. Interleave instructions from different threads on parallel execution on parallel pipelines.

Implicit multithreading is concurrent execution of multiple threads extracted from single sequential program.

18. State the need for Instruction Level Parallelism (May 2016)

A new way to improve uniprocessor performance. The proposals such as VLIW, superscalar, and even relatively old ideas such as vector processing try to improve computer performance by exploiting instruction-level parallelism.

They take advantage of this parallelism by issuing more than one instruction per cycle explicitly (as in VLIW or superscalar machines) or implicitly (as in vectormachines).

19. What is meant by hardware multithreading? (Nov/Dec-2019)

Hardware multithreading allows multiple threads to share the functional units of a single processor in an overlapping fashion to try to utilize the hardware resources efficiently. To permit this sharing, the processor must duplicate the independent state of each thread.

20. Define warehouse-scale computer (WSC)?

- A warehouse-scale computer (WSC) is a cluster comprised of tens of thousands of servers.
- The cost may be on the order of $150M for the building, electrical and cooling infrastructure, the servers, and the networking equipment that houses 50,000 to 100,000 servers.

PART-B

1. Explain in detail Flynn's classification of parallel hardware.
2. Explain SISD and SIMD with suitable example. (May 2015)
3. Explain MISD and MIMD with suitable example. (May 2015)
4. Discuss the centralized and distributed shared memory multiprocessors with suitable diagrams (or) Discuss Shared memory multiprocessor with a neat diagram.

5. Discuss briefly about the motivation of Multi-core computing.

6. Suppose you want to perform matrix subtraction for a pair of two dimensional arrays with dimensions 20 by 20 on 40 processors. Assume that the matrix subtraction is parallelizable. To achieve a speed-up of 20.5 with 40 processors, assume that the load is perfectly balanced among all the processors. Then each of the 40 processors had 2.5% of work to do.

i) Show the impact on speed-up if one processor has twice the load than all the rest of the processors.

ii) Show the impact on speed-up if one processor has five times the load than all the rest of the processors.

7. What is hardware multithreading? Compare and contrast Fine grained multithreading and Coarsed grained multithreading.

8. Explain the Dynamic & Static multiple issue processor and their scheduling with block diagram

9. Explain Instruction Level Parallel Processing. State the challenges of parallel processing.

10. Explain the term: (i)Multicore Processor (ii) Hardware Multithreading

UNIT V MEMORY & I/O SYSTEMS

PART-A

1.What is principle of locality?

The principle of locality states that programs access a relatively small portion of their address space at any instant of time. Two different types of locality have been observed:

Temporal locality: states that recently accessed items are likely to be accessed in the near future.

S partial locality: says that items whose addresses are near one another tend to be referenced close together in time.

2.Define temporal locality.

- The principle stating that a data location is referenced then it will tend to be referenced again soon.
- Temporal locality is found in instruction loops, data stacks and variable accesses.

3.Define spatial locality.

The locality principle states that if a data location is referenced, data locations with nearby addresses will tend to be referenced soon.

4.What is the need to implement Memory as Hierarchy?

- It is a structure that uses multiple levels of memory with different speeds andsizes.
- The memory unit is a essential component in a digital computer since it is needed for storing program and data.
- They are used for storing system programs, large data files, and other backupinformation.
- Only programs and data currently needed by the processor reside in mainmemory.

5. Define Hit and Miss.

The performance of cache memory is frequently 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, then it is in main memory and it counts as a miss.

6. What is cache memory?

It is a fast memory that is inserted between the larger slower main memory and the processor. It holds the currently active segments of a program and their data.

7. Define memory interleaving

In computing, interleaved memory is a design made to compensate for the relatively slow speed of dynamic random-access memory (DRAM) or core memory, by spreading memory addresses evenly across memory banks.

8. Define write through.

It is a scheme in which writes always update both the cache and the next lower level of the memory hierarchy, ensuring the data is always consistent between the two.

9.Define write buffer.

It is a queue t\hat holds data while the data is waiting to be written to memory.

10.What is write-back?

It is a scheme that handles writes by updating values only to the block in the cache, then writingthe modified block to the lower level of the hierarchy when the block is replaced.

11. Define virtual memory.

The data is to be stored in physical memory locations that have addresses different from those specified by the program. The memory control circuitry translates the address specified by the program into an address that can be used to access the physical memory.

12. Summarize the sequence of events involved in handling an interrupt requestfrom a single device

The sequence of events involved in handling an interrupt request from a single device is interrupt service routine, interrupt enable instruction and interrupt disable instruction.

13. How does a processor handle an interrupt?

Assume that an interrupt request arises during execution of instruction i. steps to handle interrupt by the processor is asfollow:

1. Processor completes execution of instructioni
2. Processor saves the PC value, program status on tostack.
3. It loads the PC with starting address ofISR
4. After ISR is executed, the processor resumes the main program execution by reloading PC with (i+1)th instructionaddress.

14. What is SCSI?

Small Computer System Interface, a interface standard. SCSI interfaces provide for faster data transmission rates (up to 80 megabytes per second) than standard serial and parallel ports. In addition, you can attach many devices to a single SCSI port, so that SCSI is really an I/O,bus rather than simply an interface.

15. Define USB.

Universal Serial Bus, an external bus standard that supports data transfer rates of 12 Mbps. A single USB port can be used to connect up to 127 peripheral devices, such as mice, modems, and keyboards. USB also supports Plug-and-Play installation and hot plugging.

16. Distinguish between isolated and memory mapped I/O.

The **isolated I/O** method isolates memory and I/O addresses so that memory address values are not affected by interface address assignment since each has its own address space.

In **memory mapped I/O**, there are no specific input or output instructions. The CPU can manipulate I/O data residing in interface registers with the same instructions that areused to

manipulate memory words.

17. What is meant by vectored interrupt?

Vectored Interrupts are type of I/O interrupts in which the device that generates the interrupt request (also called IRQ in some text books) identifies itself directly to the processor.

18. Compare Static RAM and Dynamic RAM

Static RAM is more expensive, requires four times the amount of space for a given amount of data than dynamic RAM, but, unlike dynamic RAM, does not need to be power-refreshed and is therefore faster to access

Dynamic RAM uses a kind of capacitor that needs frequent power refreshing to retain its charge. Because reading a DRAM discharges its contents, a power refresh is required after each read.

19. What is DMA?

DMA (Direct Memory Access) provides I/O transfer of data directly to and from the memory unit and the peripheral. The following DMA transfer combinations are possible:

- Memory tomemory
- Memory toperipheral
- Peripheral tomemory
- Peripheral toperipheral

20. Differentiate Programmed I/O and Interrupt I/O.

- In programmed I/O all data transfers between the computer system and external devices are completely controlled by the computer program. Part of the program will check to see if any external devices require attention and act accordingly.
- Interrupt I/O is a way of controlling input/output activity in which a peripheral or terminal that needs to make or receive a datatransfer sends a signal that causes a program interrupt to be set.

21. What is the purpose of Dirty/Modified bit in Cache memory?

During Write back the information is written only to the block in the cache. The modified cache block is written to main memory only when it is replaced. To reduce the frequency of writing back blocks on replacement, a dirty bit is commonly used. This status bit indicates whether the block is dirty(modifiedwhileinthecache)orclean(notmodified).Ifitiscleantheblockisnotwrittenonamiss.

22. Point out how DMA can improve I/O speed?

Direct memory access (DMA) is a feature of computer systems that allows certain hardware subsystems to access main system memory (RAM) independently of the central processing unit(CPU).

23. Differentiate physical address from logical address.

- Physical address is an address in main memory.
- Logical address (or) virtual address is the CPU generated addresses that corresponds to a location in virtual space and is translated by address mapping to a physical address when memory is accessed.

24. What are the various memory technologies?

- SRAM
- DRAM
- Magnetic Disks

25. Define Hit ratio.

The **hit rate**, or hit ratio, is the fraction of memory accesses found in the upper level; it is often used as a measure of the performance of the memory hierarchy.

PART-B

1.Explain in detail about the basic structure of a memory level hierarchy with suitable diagram.

2. Elaborate on the various memory technologies and its relevance.

3. Explain in detail about the different ways of measuring and improving the performance of cache memory.

4. Discuss the steps involved in the address translation of Virtual Memory with necessary block diagram.

5. Explain in detail about internal Organization of Memory Chip.

6. Describe dependency and the various types of dependencies in detail.

7. Explain about Parallel Bus architecture in detail

8. i)Draw different memory address layouts and brief about the technique used to increase the average rate of fetching words from the mainmemory.

ii)Explainindetailaboutanytwostandardinputandoutputinterfacesrequiredtoconnectthe I/O device to the Bus.

9.What is virtual memory? Explain in detail about how virtual memory is implemented with neat diagram?

10.Define Cache Memory? Explain various mapping techniques associated with cache memory and Explain in detail about the Bus Arbitration techniques in DMA.