| | | |
|---|---|---|
| Subject Name | : | Computer Architecture II |
| Subject Code | : | XCS 242 |
| Class | : | M.Sc Computer Technology |
| Semester | : | Fourth |

Prepared by

**V. Alan Gowri Phivin**

Lecturer

Dept. of SE & CT

NICE

# COMPUTER ARCHITECTURE – II

1. Define CISC.

A computer with a large number of instruction sets is called a Complex Instruction set Computer or CISC for short, e.g. VAX of Digital Equipment Corporation (DEC), IBM System. 370.

2. Define RISC.

A computer with limited number of instruction sets is known as Reduced Instruction Set Computers or RISC for short, e.g. Alpha Processor of DEC and 68000 series of Motorola.

3. Define Response Time.

Response time is the time spent to complete an event or an operation. It is also referred to as the execution time or latency.

4. Define Throughput.

Throughput is the amount of work done at a given time. That is, the amount of processing that can be accomplished during a given interval of time.

5. What is meant by performance of a system?

The performance of the processor is measured by elapsed time which is the time spent from the start of execution to the completion of a program.

6. What is meant by clock? Define the terms Clock interval and Clock rate.

Clock The heart of the processor circuit is the timing signal that it generates called clock.

Clock Interval The clock defines regular time intervals called clock interval.

Clock Rate The length of one clock cycle is an important parameter in the processor performance. The inverse of clock length is called clack rate which is expressed in cycles per second.

7. Give the basic performance equation.

The basic performance equation is given by

$$T = (N*S) / R$$

Where,

T- Performance parameter of an application program

N- Number of machine language instruction required to complete the execution of a program.

S- Average number of basic steps required to complete the execution of a program.

R- Clock rate of the processor in cycles per second.

8. How the clock rate can be increased.

There are two ways which the clock rates can be increased

1. Improved integrated circuits improve the logic circuit to be faster, thus reducing the time taken to complete a basic step.
2. Reducing the processing amount in one basic step helps to reduce the clock period t.

9. State Amdahl's Law

Amdahl's law states that performance improvement to be gained by using a faster mode of execution is limited by the fraction of time the faster mode can be used. Using this law, the performance gain that can be obtained by improving some portion of the computer can be calculated using the following formula.

$$\text{Speedup} = \frac{\text{Performance of entire task using the enhancement when possible}}{\text{Performance for the entire task without using the enhancement}}$$

10. What is SPEC?

A milestone in performance evaluation was the formation of the System Performance Evaluation Cooperative (SPEC) group in 1998. SPEC consists of representatives from various computer related organizations such as Apollo, Hewlett-Packard, DEC, MIPS and SUN.

11. What is Speedup?

Speedup is a measure of how fast a task will run using the machine with enhancement as opposed to the original machine without enhancement.

$$\text{Speedup} = \frac{\text{Execution time for the entire task without using enhancement}}{\text{Execution time for the entire task using the enhancement when possible}}$$

12. Define Cycles per Instruction (CPI)

A program consists of a number of CPU instructions represented by Instruction Count (IC). If the number of clock cycles and the instruction count are known, then CPI can be calculated as,

CPI = CPU clock cycles for a program / Instruction Count,

13. What are the various factors affecting the CPU time?

1. Clock cycles or Clock rate (C)
2. Clock cycles Per Instruction (CPI)
3. Instruction Count (IC)

14. What is Instruction Set and Instruction Set Architecture.

The operation of a CPU is determined by the instruction it executed called machine instruction. The collection of such instructions is called Instruction Set of a particular CPU. The complete instruction set is commonly referred to as Instruction Set Architecture.

15. What are different types of data used in an instruction?

1. Addresses
2. Numbers
3. Characters (Alpha numeric)
4. Logical data (True or False situation)

16. Explain Latency and Throughput.

Latency: Each instruction takes certain amount of time to complete. This is called as latency. It is the time difference between when as instruction is issued and when it is completed.

Throughput: The number of instructions completed in a given time is called throughput.

17. What are the major characteristics of a pipeline?

1. Pipelining cannot be implemented on a single task, as it works by splitting multiple tasks into a number of subtasks and operating on them simultaneously.
2. The speedup or efficiency is achieved by using a pipeline depends on the number of pipe stages and the number of available tasks that can be subdivided.
3. If the split-able task has uneven length of execution times, then the speedup of the pipeline is reduced.
4. Through the pipelining architecture does not reduce the time of execution of a single task, it reduces the overall time taken for the entire job to get completed.

18. What is pipe stage?

Each step in a pipeline is called as a pipe stage.

19. What is Instruction Pipeline?

The type of pipeline which works by partitioning the instruction execution is called Instruction Pipeline.

20. 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 faulting instruction, do nor change the state of the machine.

21. What is imprecise exception?

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

22. What is Pipeline register delay?

Adding registers between pipeline stages means adding logic between stages and setup and hold times for proper operations. This delay is known as pipeline register delay.

23. Define Pipeline hazards.

The pipeline architecture works smoothly as long as it is able to take up a new task in every machine cycle. In practice there are situations when the next instruction cannot be executed in the following machine cycle. These events are called as pipeline hazards.

24. What are the types of pipeline hazards?

1. Structural Hazard
2. Data Hazard
3. Control Hazard

25. Define Structural hazard.

The structural hazard happens when there are conflicts over the same hardware resource by two different pipeline cycles.

26. Define Data hazard.

Data hazard occurs when an instruction is waiting for data that is not yet available because the previous instruction has not completed or stored.

27. Define Control Hazard.

Control hazard occurs normally due to branch instructions since the branch condition (for compare and branch) and the branch address are not available in time to fetch the next instruction on the next clock cycle.

28. List the techniques used for overcoming hazards.

1. Data forwarding
2. Adding sufficient hardware
3. Stalling instructions
4. Document to find instruction in wrong order

29. What are the techniques used to prevent Control hazards.

1. Scheduling instructions in delay slots
2. Loop unrolling
3. Conditional execution
4. Speculation (by both compiler and CPU)

30. What is instruction level parallelism?

Pipelining increases performance by overlapping execution of independent instructions. The potential to overlap instructions is called Instruction-Level Parallelism (ILP) since the instructions are evaluated in parallel.

31. Define Loop Level Parallelism.

Parallelism among iterations in a loop is called as Loop Level Parallelism (LLP).

32. What is dependence?

Dependences are the property of programs, not of CPUs or pipelines. Dependence between two instructions will always exist unless the program is changed.

33. What are the types of dependences?

1. Data dependence
2. Name dependence
3. Control dependence

34. How to over come the Data dependences.

There are two ways to over come the Data dependences

1. Maintaining the dependence but avoiding the hazard.
2. Eliminating the dependence by transforming the code.

35. What is Name dependence.

Name dependence is said to occur when two instructions use the same register or memory location and there is no flow of data between instructions that use the same name.

36. What are the types of Name dependences?

1. Anti dependence
2. Output dependence

37. What is meant by anti dependence?

Assume instruction ‘i’ precedes instruction ‘j’. Anti dependence occurs if ‘j’ writes to a register or memory location while ‘I’ reads and is executed first. This corresponds to WAR hazard.

38. What is meant by Control dependence?

Assume instruction ‘I’ precedes instruction ‘j’. Control dependence occurs when ‘I’ and ‘j’ write to the same register or memory location, resulting in a WAW hazard and instruction order is to be maintained.

39. Define Dynamic scheduling.

The CPU rearranges the instruction to reduce stalls while preserving dependences and this technique is called as dynamic scheduling. It uses a hardware based mechanism to rearrange instruction execution order to reduce stalls at run-time and enables handling cases where dependences are unknown at compile time.

40. What is Score boarding?

Score boarding is a hard ware mechanism that maintains an execution rate of one instruction pr cycle by executing an instruction as soon as its operands are made available and no hazard condition prevent it.

41. Write short notes on Tomasulo’s Approach.

This technique allows execution of instruction to proceed in the presence of hazards. This can only be applied to floating point operations. Tomasulo’s scheme allow renaming of registers dynamically. It buffers operands of instructions waiting to issue, fetch them as soon as they are available.

42. What are the differences between Score boarding and Tomasulo’s Approach.

| S.No | Score boarding | Tomasulo’s Approach |
|---|---|---|
| 1 | Score boarding must wait for WAR and WAW hazards to clear. | Register renaming is used to eliminate WAR and WAW hazards. |
| 2 | Hazard detection and execution control is centralized. | Hazard detection and execution control is distributed to each functional unit. |
| 3 | Result is forwarded to the register file. | Forwards results directly to the functional units. |

43. What is BTB?

A BTB (Branch Target Buffer) is an associative memory where the addresses of branch instructions are stored along with their target addresses.

44. What are the drawbacks of pipelined computers?

Modern pipelined computers achieve high performance by breaking up the execution of an instruction into separate operations and concurrently executing those operations in parallel. This can lower the cycle time and allow one instruction to complete in every clock cycle. They issue only one instruction in one clock cycle and this places an upper limit on performance of these computers to keep the CPI closer to 1.

45. Differentiate Superscalar from VLIW

| S.No | Superscalar | VLIW |
|---|---|---|
| 1 | Smaller code size | Simplified hardware for decoding, issuing instructions |
| 2 | Binary compatibility across generations of hardware | No interlock hardware |
| 3 | No hardware changes required | More registers, but simplified hardware for register ports |

46. List the limitations of VLIW.

Technical problems Increase in code size from open slots like wasted bits for unused functional units, increases the memory bandwidth requirements unnecessarily.

Logistical Problems Binary compatibility is a problem.

47. Define Vector Processors.

There are certain problems whose complex computations are beyond the capabilities of the normal conventional computer. These problems are characterized by the fact that they require a vast number of computations. Many science and engineering applications that require such complex computation need a different kind of approach and the same can be formulated in terms of vectors and matrices leading to vector processing. Computers with vector processing capabilities are called as Vector Processors.

48. Give the instruction format for Vector Processor.

| Operation Code | Base Address Source 1 | Base Address Source 2 | Base Address Destination | Vector length |
|---|---|---|---|---|

49. Define Vectorization.

The process to replace a block of sequential code by vector instruction is called Vectorization.

50. State some static compiler techniques.

1. Static Pipeline scheduling
2. Loop unrolling
3. Static branch prediction
4. Static multiple instruction issue-VLIW
5. Conditional or predicted instructions
6. Loop level parallelism analysis
7. Software pipelining

51. What is internetworking?

The connection of two or more interconnection network is called internetworking, which relies an communication standards to convert in formation from one kind to network to another. (eg) Internet.

52. What are multiprocessor systems?

A multiprocessor system contains two or more processors of approximately comparable capabilities connected together using a common bus. All processors share access to common sets of memory modules, I/O channels and peripheral devices, The entire system is controlled by a single integrated operating system that provides interaction between processors and their programs at various level.

53. Give the Flynn's classification of digital computers.

1. Single Instruction Single Data stream (SISD)
2. Single Instruction Multiple Data stream (SIMD)
3. Multiple Instruction Single Data stream (MISD)
4. Multiple Instruction Multiple Data stream (MIMD)

54. Give the classification of MIMD machines.

1. Tightly coupled systems (Centralized shared memory)
2. Loosely coupled systems (Distributed shared memory)

55. Write short notes on Centralized shared memory architecture.

In this architecture, there are few dozens of processors that share a bus and a single main memory. Large caches allow bus and memory organization to satisfy the memory demands of a small number of processors.

56. Write short notes on Distributed shared memory architecture.

This architecture supports large processor counts by distributing the memory and allowing multiple memories to work in parallel. Because of the increase in processor band width requirements, the number of processors for which the scheme is preferred is slowly decreasing.

57. Give the advantages of Distributed shared memory architecture.

1. It is cost effective way to use the entire memory band width when most accesses are to the local memory in the node.
2. It reduces latency for accesses to local memory due to less contention for resources.

58. Classify distributed memory machines based on the communication models.

1. Shared memory with Non Uniform Memory Access (NUMA)
2. Message passing Multi computers

59. Define : SAN.

Originally " System area Network" but more recently "Storage area network", it connects computers and /or storage devices in a machine room. FC –AL and Infiniband are SANS.

60. Define Communication latency.

Communication latency is given by the following factors.

Communication latency = Sender overhead + Time of flight + Transport latency + Recovery overhead.

61. List any 3 advantages of Shared memory.

1. Compatibility with well-understood mechanism in use in centralized multiprocessors, all of which use shared memory communication.
2. Ease of programming, particularly for systems in which communication patterns are complex or change dynamically during execution.
3. Low overhead of communication and better use of band width when communicating small items.

62. List any 3 advantages of message passing.

1. Explicit communication gives simple understanding.
2. Synchronization is associated with sending messages, thus reducing errors.
3. Sender-initiated communication can also be used easily to generate advantages in performance.

63. List the advantages of replication.

1. Reduced latency and memory band width requirements.
2. Reduced contention for data items that are read by multiple processors simultaneoualy.

64. List any two Cache coherence protocols.

1. Directory based protocol
2. Snooping protocol

65. What are the two ways to maintain the coherence requirements?

1. Write invalidate
2. Write Broadcast

66. What is barrier synchronization?

Barrier synchronization allows multiple processes on multiple CPUs to wait until a certain number of processes have reached a barrier. When sufficient processes arrive at barrier, all waiting processes are released and allowed to continue.

67. What is meant by Cache Coherence protocols?

The protocols a maintain coherence of cache (i.e. consistent view of memory) for multiple processors are called Cache coherence protocols.

68. What is directory based Cache coherence protocol?

The sharing status of a block of physical memory is kept in one location called the directory. Each directory is responsible for tracking caches that share memory addresses. The directory may communicate with the processor and memory over a common bus. It may also have a separate port of memory for communication.

69. What is snooping?

Every cache that has a copy of data from a block of physical memory also has a copy of the sharing status of the block and no centralized state is maintained. ie. sharing status is distributed and kept with the block in each cache.

70. Define Fully associative cache.

The cache is said to be fully associative if the block can go anywhere in the cache.

71. Define Set Associative cache.

In a set associative cache, a block can go in any one of a set of places in the cache. A set is a group of blocks in the cache.

72. How a block can be identified in the memory?

Data in the cache is identified using the components of an address as they relate to the cache. Table

| Block Address | | Offset |
|---|---|---|
| Tag | Index | |
| Stored in cache and used in comparison with CPU address | Select set | Selects data within the block. |

73. What is meant by block replacement?

When a cache miss occurs, the controller must select a block to be replaced with the desired data. Only one block frame is checked for a hit and only that block can be replaced. This is called block replacement.

74. State any three strategies for block replacement.

1. Least Recently used
2. First in first out
3. Random

75. State the strategies used for Cache writes

1. Write through
2. Write back

76. List the advantages of Write through cache.

1. Read misses do not result in writes
2. Memory hierarchy is consistent
3. It is simple to implement

77. What is write back cache?

In write back scheme, only one block in cache is modified. The main memory is modified only when the block must be replaced in the cache. This requires the use of dirty bit to keep track of blocks that have been modified.

78. List of advantages to write back cache.

1. Writes occur at the speed of cache.
2. The main memory bandwidth is smaller when multiple writes occur to the same block.

79. What steps are to be adopted when there is a write miss?

If a miss occurs on write, there are two options

1. Write allocate on a miss, the block is loaded into the cache before anything else occurs.
2. Write around (on write allocate) The block is written only to main memory. It is not stored in the cache.

In general, write back caches use write allocate while write-through caches use write around.

80. Give the formula to calculate average memory access time?

Average memory access time = Hit time + Miss rate * Miss penlty

Where,

Hit time the time taken to hit a valid block in the cache.

Miss rates the fraction of references that is not satisfied in the upper level (cache). They require to access to the lower level main memory and slower level secondary storage to satisfied.

Miss penalty the length of time taken to access the lower level main memory. A low miss rate does not help much in improving the performance if the miss penalty is very high.

81. Differentiate unified cache from split cache.

Unified cache - All memory requests go through a single cache. This requires less hardware, but it has lower bandwidth and more opportunities for collision.

Split instruction and data cache - separate cache are used for instruction and data.

82. List some techniques for improving cache performance

1. Reducing the miss rate
2. Reducing the miss penalty
3. Reducing the miss rate or miss penalty using parallelism
4. Reducing the time to hit in a cache.

83. Define conflict

When the cache has sufficient space for data, but the block cannot be kept because the set is full, a conflict miss will occur.

84. How to reduce the conflict misses and compulsory misses?

1. Large cache blocks
2. Higher associativity
3. Victim caches
4. Pseudo associative caches
5. Complier optimization

85. What is victim cache?

A victim cache is a small (usually, but not necessarily) fully associative cache that holds a few of the most recently replaced blocks or victims from the main cache.

86. How compiler optimization can be achieved?

1. Mergine arrays
2. Loop interchange
3. Loop fusion
4. Blocking

87. Give some methods to reduce cache miss penalty.

1. Giving priority to read misses.
2. Using sub-blocks to reduce fetch time
3. Early restart and critical word first
4. Having second level caches.

88. What is virtual memory?

Virtual memory is another level of memory hierarchy and it allows main memory to cache pages (blocks) to be normally stored on disk. As with caches, the operations performed by virtual memory are transparent to users.

89. Write the differences between Cache and Virtual Memory.

| **Caches** | **Virtual Memory** |
|---|---|
| Replacement is primarily controlled by the hardware. | Replacement is primarily comtro9lled by the OS. |
| The cache size is independent of the address size. | The number of bits in the address determines the size of VM. |

90. List some characteristics of RISC processors.

1. Relatively few instructions
2. Relatively few addressing modes
3. Memory access limited to load and store instructions
4. All operations done within the registers of the CPU

91. What is payload?

The middle part of the message that contains user information.

92. What are the two advantages of Load/Store Architecture?

1. Simple fixed length encodings and so simplifies decoding.
2. Similar number of clock cycles per instructions which simplifies control.

93. What are the different types of addressing modes used in the instruction set design?

1. Immediate Addressing
2. Direct Addressing
3. Register Addressing
4. Register Indirect Addressing
5. Displacement Addressing
6. Relative Addressing
7. Base Register Addressing
8. Indexing

94. What is time of flight?

The time for the first bit of the message to arrive at the receiver.

95. What is transmission time?

The time for the message to pass through the network. (not including time of flight).

96. Write the formula for finding time latency –in interconnection network.

$$\text{Total latency} = \text{Sender Overhead} + \text{Time of flight} + \frac{\text{Message size}}{\text{Band width}} + \text{Receiver overhead.}$$

97. Write the three network media.

Coaxial Cable, Twisted Pair, Fiber optics.

98. Write some examples for interconnection networks.

Ether net, Infiniband, ATM.

99. Define Bandwidth.

Maximum rate the network can propagate information once the message enters it.

100. What is stack processor?

All processor shave been using general purpose registers to hold data operands and instructions can access them in any desired order. Some years ago, the Hewlett Packard Company designed and manufactured a computer called HP3000m whose main architectural feature is an instruction set that is keyed to processing operands held in a stack data structure. Access to operands is restricted to only those operands reading from the top of the stack and results are always returned to the top of the stack such computers are called as Stack processor.

## Question Answer Sixteen Marks

1. Explain the fundamental of Computer Design.

A computer may consist of the following hardware.

1. Memory Unit
2. General and Special purpose register.
3. Control circuits consisting of flip-flops, decoders etc.
4. Common data and address buses.

A Computer consists of five functionally interdependent units.

1. Input Unit
2. Output Unit
3. Memory Unit
4. Arithmetic and Logic Unit (ALU)
5. Control Unit (CU).

2. What is CPU Performance? Derive the CPU Performance equation.

CPU performance is the measure the speed of CPU.

CPU time = CPU Clock cycles for a program x Clock cycle time.

Or

CPU time = CPU Clock cycles for a program / Clock rate.

The units of CPU time is seconds / program

CPU Clock cycles = $\sum_{i=1}^{n} (CPIi \times Ci)$.

3. Explain different types of Operands.

The assembly and high – level language programs use the following patterns.

1. Address.
2. Numbers.
3. Characters.
4. Logical data.

4. Explain the type of OPCODES.

Many machines follow their own variety of opcodes and may be classified as follows.

1. Arithmetic
2. Logical
3. Data transfer
4. conversion
5. System control
6. Input / Output control
7. Control transfer.

5. Explain different types of Instruction Set Architecture.

1. Accumulator Architecture.
2. Stack Architecture.
3. Register – Register Architecture.
4. Register – Memory Architecture.

6. Explain Instruction formats.

Three - Address Instruction.

Two - Address Instruction.

One - Address Instruction.

Zero - Address Instruction.

7. Explain different types of Addressing modes.

1. Immediate addressing.
2. Direct addressing.
3. Indirect addressing.
4. Index addressing.
5. Register addressing.
6. Displacement addressing.
7. Auto – increment addressing.
8. Auto – decrement addressing.

8. Explain and derive the performance of pipeline with Stalls.

$$\text{Speed up due to pipelining} = \frac{\text{Average instruction time without pipelining}}{\text{Average instruction time with pipelining}}$$

$$\text{Overall Speed up} = \frac{\text{Depth of pipeline}}{1 + \text{Pipeline stall cycles per instruction}}$$

9. Explain the different types of Hazards.

**Define Pipeline hazards.**

The pipeline architecture works smoothly as long as it is able to take up a new task in every machine cycle. In practice there are situations when the next instruction cannot be executed in the following machine cycle. These events are called as pipeline hazards.

**Types of pipeline hazards**.

1. Structural Hazard
2. Data Hazard
3. Control Hazard

**Structural hazard**.

The structural hazard happens when there are conflicts over the same hardware resource by two different pipeline cycles.

**Data hazard.**

Data hazard occurs when an instruction is waiting for data that is not yet available because the previous instruction has not completed or stored.

**Control Hazard**.

Control hazard occurs normally due to branch instructions since the branch condition (for compare and branch) and the branch address are not available in time to fetch the next instruction on the next clock cycle.

10. How to overcome the hazards. Explain with suitable example.

**The techniques used for overcoming hazards**.

5. Data forwarding
6. Adding sufficient hardware
7. Stalling instructions
8. Document to find instruction in wrong order

**The techniques used to prevent Control hazards**.

5. Scheduling instructions in delay slots
6. Loop unrolling
7. Conditional execution
8. Speculation (by both compiler and CPU)

11. Explain the type of dependences.

Dependences are the property of programs, not of CPUs or pipelines. Dependence between two instructions will always exist unless the program is changed.

**Types of dependences.**

4. Data dependence
5. Name dependence
6. Control dependence

**There are two ways to over come the Data dependences**

3. Maintaining the dependence but avoiding the hazard.
4. Eliminating the dependence by transforming the code.

**Name dependence.**

Name dependence is said to occur when two instructions use the same register or memory location and there is no flow of data between instructions that use the same name.

**Types of Name dependences.**

3. Anti dependence
4. Output dependence

**Anti dependence.**

Assume instruction ‘i’ precedes instruction ‘j’. Anti dependence occurs if ‘j’ writes to a register or memory location while ‘I’ reads and is executed first. This corresponds to WAR hazard.

**Control dependence.**

Assume instruction ‘I’ precedes instruction ‘j’. Control dependence occurs when ‘I’ and ‘j’ write to the same register or memory location, resulting in a WAW hazard and instruction order is to be maintained.

12. What is Dynamic scheduling. Explain with Tomasulo's approach.

**Dynamic scheduling.**

The CPU rearranges the instruction to reduce stalls while preserving dependences and this technique is called as dynamic scheduling. It uses a hardware based mechanism to rearrange instruction execution order to reduce stalls at run-time and enables handling cases where dependences are unknown at compile time.

**Score boarding.**

Score boarding is a hard ware mechanism that maintains an execution rate of one instruction pr cycle by executing an instruction as soon as its operands are made available and no hazard condition prevent it.

**Tomasulo's Approach.**

This technique allows execution of instruction to proceed in the presence of hazards. This can only be applied to floating point operations. Tomasulo's scheme allow renaming of registers dynamically. It buffers operands of instructions waiting to issue, fetch them as soon as they are available.

| S.No | Score boarding | Tomasulo's Approach |
|---|---|---|
| 1 | Score boarding must wait for WAR and WAW hazards to clear. | Register renaming is used to eliminate WAR and WAW hazards. |
| 2 | Hazard detection and execution control is centralized. | Hazard detection and execution control is distributed to each functional unit. |
| 3 | Result is forwarded to the register file. | Forwards results directly to the functional units. |

13. Explain about Superscalar.

1. Two – issue Superscalar.
2. In – Order Issue in superscalar

14. Differentiate Superscalar from VLIW

| S.No | Superscalar | VLIW |
|---|---|---|
| 1 | Smaller code size | Simplified hardware for decoding, issuing instructions |
| 2 | Binary compatibility across generations of hardware | No interlock hardware |
| 3 | No hardware changes required | More registers, but simplified hardware for register ports |

15. Explain Vector Processor

There are certain problems whose complex computations are beyond the capabilities of the normal conventional computer. These problems are characterized by the fact that they require a vast number of computations. Many science and engineering applications that require such complex computation need a different kind of approach and the same can be formulated in terms of vectors and matrices leading to vector processing. Computers with vector processing capabilities are called as Vector Processors.

**Instruction format for Vector Processor.**

| Operation Code | Base Address Source 1 | Base Address Source 2 | Base Address Destination | Vector length |
|---|---|---|---|---|

16. Explain Centralized – shared memory Architecture.

In this architecture, there are few dozens of processors that share a bus and a single main memory. Large caches allow bus and memory organization to satisfy the memory demands of a small number of processors.

**Cache coherence protocols.**

3. Directory based protocol
4. Snooping protocol

**Two ways to maintain the coherence requirements.**

3. Write invalidate
4. Write Broadcast

**Cache Coherence protocols.**

The protocols a maintain coherence of cache (i.e. consistent view of memory) for multiple processors are called Cache coherence protocols.

Directory based Cache coherence protocol.

The sharing status of a block of physical memory is kept in one location called the directory. Each directory is responsible for tracking caches that share memory addresses. The directory may communicate with the processor and memory over a common bus. It may also have a separate port of memory for communication.

**Snooping.**

Every cache that has a copy of data from a block of physical memory also has a copy of the sharing status of the block and no centralized state is maintained. ie. sharing status is distributed and kept with the block in each cache.

17. Explain Distributed – Shared memory Architecture.

This architecture supports large processor counts by distributing the memory and allowing multiple memories to work in parallel. Because of the increase in processor band width requirements, the number of processors for which the scheme is preferred is slowly decreasing.

**Advantages of Distributed shared memory architecture.**

3. It is cost effective way to use the entire memory band width when most accesses are to the local memory in the node.
4. It reduces latency for accesses to local memory due to less contention for resources.

**Classify distributed memory machines based on the communication models.**

3. Shared memory with Non Uniform Memory Access (NUMA)
4. Message passing Multi computers

**State the strategies used for Cache writes**

3. Write through
4. Write back

**Advantages of Write through cache.**

4. Read misses do not result in writes
5. Memory hierarchy is consistent
6. It is simple to implement

18. Explain Synchronization.

Barrier synchronization allows multiple processes on multiple CPUs to wait until a certain number of processes have reached a barrier. When sufficient processes arrive at barrier, all waiting processes are released and allowed to continue.

19. How to improve cache performance.

Fully associative cache.

The cache is said to be fully associative if the block can go anywhere in the cache.

**Associative cache.**

In a set associative cache, a block can go in any one of a set of places in the cache. A set is a group of blocks in the cache.

**Identified in the memory.**

Data in the cache is identified using the components of an address as they relate to the cache. Table

| Block Address | | Offset |
|---|---|---|
| Tag | Index | |
| Stored in cache and used in comparison with CPU address | Select set | Selects data within the block. |

**Block replacement.**

When a cache miss occurs, the controller must select a block to be replaced with the desired data. Only one block frame is checked for a hit and only that block can be replaced. This is called block replacement.

**State any three strategies for block replacement.**

4. Least Recently used
5. First in first out
6. Random

**State the strategies used for Cache writes.**

1.Write through

2.Write back

**Advantages of Write through cache.**

1. Read misses do not result in writes
2. Memory hierarchy is consistent
3. It is simple to implement

Write back cache.

In write back scheme, only one block in cache is modified. The main memory is modified only when the block must be replaced in the cache. This requires the use of dirty bit to keep track of blocks that have been modified.

20. Explain the typical RISC Architecture.

RISC architecture makes an attempt to reduce execution time by simplifying the instruction set of the computer. The major characteristics of a RISC processor are.

1. Relatively few instructions.
2. Relatively few addressing modes.
3. Memory access limited to load and store instructions.
4. All operations done within the registers of the CPU.
5. Fixed length instruction format making it easy to decode.
6. Single-cycle instruction execution.
7. Hardwired rather than micro-programmed control.