www.AllAbtEngg.com

25.What is SMT?

Simultaneous Multithreading (SMT) is a variation on hardware multithreading that uses the resources of a multiple-issue, dynamically scheduled pipelined processor to exploit thread-level parallelism. It also exploits instruction level parallelism.

26. Define SMP

Shared memory multiprocessor (SMP) is one that offers the programmer a single physical address space across all processors - which is nearly always the case for multicore chips. Processors communicate through shared variables in memory, with all processors capable of accessing any memory location via loads and stores

27. Differentiate UMA from NUMA.

Uniform memory access (UMA) is a multiprocessor in which latency to any word in main memory is about the same no matter which processor requests the access. Non uniform memory access (NUMA) is 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.

PART-B

1. Explain Instruction level parallelism.
2. Explain challenges in parallel processing.
3. Explain in detail about Hardware multithreading.
4. Discuss in detail about Flynn's classification.
5. Explain SISD and SIMD with an example.
6. Explain MISD and MIMD with an example
7. Explain shared memory multiprocessor
8. Explain Multicore processors
9. . Explain the different types of multithreading

UNIT-V
Memory and 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.

2. What are the temporal and spatial localities of references?

Temporal locality: The principle stating that a data location is referenced then it will tend to be referenced again soon.

www.AllAbtEngg.com Study Materials for Anna University, Polytechnic & School

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

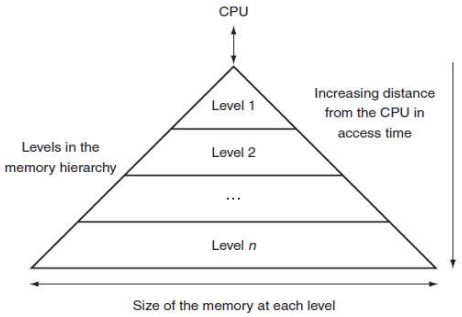
3. What are the various memory technologies?

The various memory technologies are:

1. SRAM semiconductor memory
2. DRAM semiconductor memory
3. Flash semiconductor memory
4. Magnetic disk

4. Define Memory Hierarchy and give the structure of memory hierarchy

A structure that uses multiple levels of memory with different speeds and sizes. The faster memories are more expensive per bit than the slower memories.



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. What is direct-mapped cache?

Direct-mapped cache is a cache structure in which each memory location is mapped to exactly one location in the cache. For example, almost all direct-mapped caches use this mapping to find a block, (Block address) modulo (Number of blocks in the cache)

www.AllAbtEngg.com Study Materials for Anna University, Polytechnic & School

8. What are the writing strategies in cache memory?

Write-through is a scheme in which writes always update both the cache and the next lower level of the memory hierarchy, ensuring that data is always consistent between the two. Write-back is a scheme that handles writes by updating values only to the block in the cache, then writing the modified block to the lower level of the hierarchy when the block is replaced.

9. Define write through ,write buffer and write-back.

Write through :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.

Write buffer: A queue that holds data while the data is waiting to be written to memory.

Write-back: A scheme that handles writes by updating values only to the block in the cache, then writing the modified block to the lower level of the hierarchy when the block is replaced.

10. 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

11. Distinguish between memory mapped I/O, I/O mapped I/O and Isolated I/O.

Memory mapped I/O: When I/O devices and the memory share the same address space, the arrangement is called memory mapped I/O. The machine instructions that can access memory is used to trfer data to or from an I/O device.

I/O mapped I/O: Here the I/O devices the memories have different address space. It has special I/O instructions. The advantage of a separate I/O address space is that I/O devices deals with fewer address lines.

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.

12.What are the various block placement schemes in cache memory?

Direct-mapped cache is a cache structure in which each memory location is mapped to exactly one location in the cache.

Fully associative cache is a cache structure in which a block can be placed in any location in the cache. Set-associative cache is a cache that has a fixed number of locations (at least two) where each block can be placed.

www.AllAbtEngg.com Study Materials for Anna University, Polytechnic & School

13. What is the use of DMA?

DMA (Direct Memory Access) provides I/O transfer of data directly to and from the memory unit and the peripheral.

14. 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) identifies itself directly to the processor. An interrupt for which the address to which control is transferred is determined by the cause of the exception.

15. 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. One source gives a typical access time as 25 nanoseconds in contrast to a typical access time of 60 nanoseconds for dynamic RAM. (More recent advances in dynamic RAM have improved access time.) Static RAM is used mainly for the level-1 and level-2 caches that the microprocessor looks in first before looking in dynamic RAM.

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. Apart from reading, just to maintain the charge that holds its content in place, DRAM must be refreshed about every 15 microseconds. DRAM is the least expensive kind of RAM.

16. What are the steps to be taken in an instruction cache miss?

The steps to be taken on an instruction cache miss are

1. Send the original PC value (current PC + 4) to the memory.
2. Instruct main memory to perform a read and wait for the memory to complete its access.
3. Write the cache entry, putting the data from memory in the data portion of the entry, writing the upper bits of the address (from the ALU) into the tag field, and turning the valid bit on.
4. Restart the instruction execution at the first step, which will refetch the instruction, this time finding it in the cache

17. Define AMAT

Average memory access time is the average time to access memory considering both hits and misses and the frequency of different accesses.It is equal to the following answer AMT only

$$\text{AMAT} = \text{Time for a hit} + \text{Miss rate} \times \text{Miss penalty}$$

www.AllAbtEngg.com Study Materials for Anna University, Polytechnic & School

18. What is meant by address mapping?

Address translation also called address mapping is the process by which a virtual address is mapped to an address used to access memory.

19. Define Availability

Availability is then a measure of service accomplishment with respect to the alternation between the two states of accomplishment and interruption. Availability is statistically quantified as answer availability

$$\text{AMAT} = \text{Time for a hit} + \text{Miss rate} \times \text{Miss penalty}$$

$$\text{Availability} = \frac{\text{MTTF}}{(\text{MTTF} + \text{MTTR})}$$

20. Define virtual memory.

Virtual memory is a technique that uses main memory as a "cache" for secondary storage. Two major motivations for virtual memory: to allow efficient and safe sharing of memory among multiple programs, and to remove the programming burdens of a small, limited amount of main memory.

21. Define TLB

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

22. 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.

PART- B

1. Explain in detail about memory technologies
2. Explain in detail about memory Hierarchy with neat diagram
3. Describe the basic operations of cache in detail with diagram
4. Discuss the various mapping schemes used in cache design A byte addressable computer has a small data cache capable of holding eight 32-bit words. Each cache block contains 132-bit word. When a given program is executed, the processor reads data from the following sequence of hex addresses – 200, 204, 208, 20C, 2F4, 2F0, 200, 204,218, 21C, 24C, 2F4. The pattern is repeated four times. Assuming that the cache is initially empty, show the contents of the cache at the end of each pass, and compute the hit rate for a direct mapped cache.
5. Discuss the methods used to measure and improve the performance of the cache.
6. Explain the virtual memory address translation and TLB with necessary diagram.

www.AllAbtEngg.com Study Materials for Anna University, Polytechnic & School