MEMORY MANAGEMENT

TWO MARK QUESTIONS AND ANSWERS

Q.1 Name the two types of storage devices.

Ans.: The two types of storage devices are :

1. Primary memory 2. Secondary memory

Q.2 Define memory latency.

Ans.: The term memory latency is used to refer to the amount of time it takes to transfer a word of data to or from the memory. The term latency is used to denote the time it takes to transfer the first word of data. This time is usually substantially longer than the time needed to transfer each subsequent word of a block.

Q.3 Define memory bandwidth.

Ans.: Memory bandwidth is a product of the rate at which the data are transferred (and accessed) and the width of the data bus.

Q.4 What will be the width of address and data buses for a 512 K x 8 memory chip? AU: Dec.-07

Ans.: The widths of address and data buses are 19 and 8, respectively.

8.2 Memory Hierarchy

Q.5 What do you mean by memory hierarchy ? AU: May-16

Ans.: A memory hierarchy is a structure of memory that uses multiple levels of memories; as the distance from the processor increases, the size of the memories and the access time both increase.

Q.6 What is the need to implement memory as a hierarchy ? AU: May-15

Ans.: Ideally, computer memory should be fast, large and inexpensive. Unfortunately, it is impossible to meet all the three of these requirements using one type of memory. Hence it is necessary to implement memory as a hierarchy.

8.3 Memory Technologies

Q.7 Name the two types of RAM.

Ans.: The two types of RAM are: 1. Static RAM 2. Dynamic RAM

Q.8 Distinguish between static RAM and dynamic RAM. (Refer section 8.3.3) AU: June-08

Sr. No.	Static RAM	Dynamic RAM
1.	Static RAM contains less memory cells per unit area.	Dynamic RAM contains more memory cells as compared to static RAM per unit area.
2.	It has less access time hence faster memories.	Its access time is greater than static RAMs.
3.	Static RAM consists of number of flip-flops. Each flip-flop stores one bit.	Dynamic RAM stores the data as a charge on the capacitor. It consists of MOSFET and the capacitor for each cell.
4.	Refreshing circuitry is not required.	Refreshing circuitry is required to maintain the charge on the capacitors after every few milliseconds. Extra hardware is required to control refreshing. This makes system design complicated.
5.	Cost is more.	Cost is less.

Q.9 Distinguish between asynchronous DRAM and synchronous RAM. (Refer section 8.3.6.2)AU:

Dec08			
Sr. No.	DRAMs	SDRAMs	
1.	Timing for DRAM circuits is asynchronous in nature.	Timing for SDRAM circuits is synchronous in nature.	
2.	Simple control logic.	Complex control logic.	
3.	Uses memory controllers optimized for specific system as a control logic.	Use of clock signal for synchronization enables the design of programmable synchronous state machine as the control logic.	
4.	More time access.	Less access time.	
5.	Does not support burst mode transfer.	Supports burst mode transfer.	
6.	Requires external refresh circuit.	It has built-in refresh circuit.	
7.	Not applicable.	Length of data burst and its type are programmable.	
8.	No multiple internal independent banks.	Multiple internal, independent banks support pipelining.	
9.	Less costly.	Costly.	

Table 8.3.2 Comparison between DRAMs and SDRAMs

Q.10 What are static memories ?

Ans.: Memories that consists of circuits capable of retaining their state as long as power is applied are known as static memories.

Q.11 What are static and dynamic memories ?

Ans.: Static memory are memories which require periodic no refreshing. Dynamic memories are memories, which require periodic refreshing.

Q.12 Give the features of a ROM cell.AU: May-08

Ans.: The main features of a ROM cell are:

•It can hold one bit data.

•It can hold data even if power is turned off.

• We can not write data in ROM cell; it is read only.

Q.13 Name the different types of ROMs.

Ans.: There are four types of ROMs are:

1. Masked ROM 2. PROM 3. EPROM and 4. EEPROM or E2PROM.

Q.14 What is the function of memory controller ?

Ans.: The memory controller is connected between processor and the DRAM memory. It does the task of generating multiplexed address as well as task of generating control signals such as \overrightarrow{RAS} , \overrightarrow{CAS} , $\overrightarrow{R/W}$, \overrightarrow{CS} and so on for DRAM.

Q.15 What is refreshing overhead?

Ans.: Refreshing overhead = Time required for refresh / Interval time

Q.16 Compare SDRAM with DDR SDRAM. AU: May-09

Ans.: The standard SDRAM performs its operations on the rising edge of the clock signal. On the other hand, the DDR SDRAMS transfers data on both the edges of the clock signal. The latency of the DDR SDRAMs is same as that for standard SDRAM. However, since they transfer data on both the edges of the clock signal, their bandwidth is effectively doubled for long burst transfer.

Q.17 Define track and sectors on the disk.

Ans.: Each disk surface is divided into concentric circles, called tracks. There are typically tens of thousands of tracks per surface. Each track is in turn divided into sectors that contain the information; each track may have thousands of sectors. Sectors are typically 512 to 4096 bytes in size.

Q.18 Define rotational latency or rotational delay.

Ans.: Rotational latency or rotational delay is the time required for the desired sector of a disk to rotate under the read/write head; usually assumed to be half the rotation time.

Q.19 How do you construct a 8 M x 32 memory using 512 K x 8 memory chips ?AU : May-07

Ans.: By connecting four 512 x 8 chips in parallel we can expand the word size to 32 and such sixteen blocks consisting of four memory chips we can construct 8 M \times 32 memory.

Q.20 What are the characteristics of semiconductor RAM memories ?

Ans.: They are available in a wide range of speeds.

Their cycle time range from 100 ns to less than 10 ns.

They replaced the expensive magnetic core memories.

They are used for implementing memories.

Q.21 Why SRAMS are said to be volatile ?

Ans.: SRAMs are said to be volatile because their contents are lost when power is interrupted.

Q.22 What are the characteristics of SRAMS ?

Ans.: 1. SRAMS are fast

2. They are volatile

3. They are of high cost

4. Less density

Q.23 What are the characteristics of DRAMs ?

Ans.:1. Low cost 2. High density 3. Refresh circuitry is needed.

Q.24 Define refresh circuit.

Ans.: Refresh circuit is a circuit which ensures that the contents of a DRAM are maintained when each row of cells are accessed periodically.

Q.25 What are asynchronous DRAMs ?

Ans.: In asynchronous DRAMs, the timing of the memory device is controlled asynchronously. A specialized memory controller circuit provides the necessary control signals RAS and CAS that govern the timing. The processor must take into account the delay in the response of the memory. Such memories are asynchronous DRAMs.

Q.26 What are synchronous DRAMs?

Ans.: Synchronous DRAMS are those whose operation is directly synchronized with a clock signal.

Q.27 Define memory access time.

Ans.: The time required to access one word is called the memory access time. Or It is the time that elapses between the initiation of an operation and the completion of that operation.

Q.28 Define memory cycle time.

Ans.: It is the minimum time delay required between the initiation of two successive memory operations. That is the time between two successive read operations.

Q.29 Define memory cell.

Ans.: A memory cell is capable of storing one bit of information. It is usually organized in the form of an array.

Q.30 What is a word line?

Ans.: In a memory cell, all the cells of a row are connected to a common line called word line.

Q.31 What is double data rate SDRAMs ? AU: Dec.-11

Ans.: Double data rates SDRAMS are those which can transfer data on both edges of the clock and their bandwidth is essentially doubled for long burst transfers.

Q.32 Define ROM.

Ans.: It is a non-volatile and read only memory.

Q.33 What are the features of PROM?

Ans.: They are :

- 1. programmed directly by the user
- 2. Faster
- 3. Less expensive
- 4. More flexible.

Q.34 Why EPROM chips are mounted in packages that have transparent window ?

Ans.: Since the erasure requires dissipating the charges trapped in the transistors of memory cells. This can be done by exposing the chip to UV light.

Q.35 What are the disadvantages of EPROM?

Ans.: The chip must be physically removed from the circuit for reprogramming and its entire contents are erased by the ultraviolet light.

Q.36 What are the advantages and disadvantages of using EEPROM ?

Ans.: The advantages are that EEPROMS do not have to be removed for erasure. Also it is possible to erase the cell contents selectively. The only disadvantage is that different voltages are needed for erasing, writing and reading the stored data.

Q.37 Define flash memory.

Ans.: Flash memories are read/write memories. In flash memories it is possible to read the contents of a single cell, but it is only possible to write an entire block of cells. A flash cell is based on a single transistor controlled by trapped charge.

Q.38 Differentiate flash devices and EEPROM devices.

ans.		
Flash devices	EEPROM devices	
It is possible to read the contents of a single cell, but it is only possible to write an entire block of cells.	It is possible to read and write the contents of a single cell.	
Greater density which leads to higher capacity.	Relatively lower density	
Lower cost per bit.	Relatively more cost.	
Consumes less power in their operation and makes it more attractive for use in portable equipments that is battery driven.	Consumes more power.	

8.4 Cache Memories

Q.39 What is cache memory? AU: Dec.-16

Ans.: In the memory system small section of SRAM is added along with main memory, referred to as cache memory.

Q.40 Define hit rate or hit ratio. AU: Dec.-15

Ans.: The percentage of accesses where the processor finds the code or data word it needs in the cache memory is called the hit rate or hit ratio.

Q.41 Define miss rate and miss penalty.

Ans.: The percentage of accesses where the processor does not find the code or data word it needs in the cache memory is called the miss rate. Extra time needed to bring the desired information into the cache is called the miss penalty.

Q.42 Define multi-level cache.

Ans.: A memory hierarchy with multiple levels of caches, rather than just a cache and main memory is called multi-level cache.

Q.43 Define global miss rate.

Ans.: The fraction of references that miss in all levels of a multilevel cache is called global miss rate.

Q.44 Define local miss rate.

Ans.: In a multilevel cache system, the fraction of references to one level of a cache that miss is called local miss rate.

Q.45 What is program locality?

Ans.: In cache memory system, prediction of memory location for the next access is essential. This is possible because computer systems usually access memory from the consecutive locations. This prediction of next memory address from the current memory address is known as program locality. Program locality enables cache controller to get a block of memory instead of getting just a single address.

Q.46 Define locality of reference. What are its types ? AU: June-08, May-14

Ans.: The program may contain a simple loop, nested loops, or a few procedures that repeatedly call each other. The point is that many instructions in localized area of the program are executed repeatedly during some time period and the remainder of the program is accessed relatively infrequently. This is referred to as locality of reference.

Q.47 Define the terms: spatial locality and temporal locality. AU: May-08

Ans.: Locality of reference manifests itself in two ways: temporal and spatial. The temporal means that a recently executed instruction is likely to be executed again very soon. The spatial means that instructions stored near by to the recently executed instruction are also likely to be executed soon.

Q.48 Explain the concept of block fetch.

Ans.: Block fetch technique is used to increase the hit rate of cache. A block fetch can retrieve the data located before the requested byte (look behind) or data located after the requested byte (look ahead), or both. When CPU needs to access any byte from the block, entire block that contains the needed byte is copied from main memory into cache.

Q.49 Name the common replacement algorithms.

Ans.: The four most common replacement algorithms are:

1. Least-Recently Used (LRU)

- 2. First-In-First-Out (FIFO)
- 3. Least-Frequently-Used (LFU)
- 4. Random
- Q.50 What is a mapping function ? AU: June-08

Ans.: Usually, the cache memory can store a reasonable number of blocks at any given time, but this number is small compared to the total number of blocks in the main memory. The correspondence between the main memory blocks and those in the cache is specified by a mapping function.

Q.51 List the mapping techniques.

Ans.: There are two main mapping techniques which decides the cache organisation:

1. Direct-mapping technique 2. Associative-mapping technique

The associative mapping technique is further classified as fully associative and set associative techniques.

Q.52 Define direct-mapped cache.

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

Q.53 Define fully associative cache.

Ans.: Fully associative cache is cache structure is which a block can be placed in any location in the cache.

Q.54 Define set-associative cache.

Ans.: Set-associative cache is a cache that has a fixed number of locations (at least two) where each block can be placed.

Q.55 Define tag field related to memory.

Ans.: A field in a table used for a memory hierarchy that contains the address information required to identify whether the associated block in the hierarchy corresponds to a requested word is called tag bit. Q.56 Define valid bit related to memory.

Ans.: A field in the tables of a memory hierarchy that indicates that the associated block in the hierarchy contains valid data is called valid bit.

Q.57 Define split cache.

Ans.: A scheme in which a level of the memory hierarchy is composed of two independent caches that operate in parallel with each other, with one handling instructions and one handling data is called split cache.

Q.58 What is the need of cache updating?

Ans.: In a cache system, two copies of same data can exist at a time, one in cache and one in main memory. If one copy is altered and other is not, two different sets of data become associated with the same address. To prevent this cache updating is needed.

Q.59 Name the different cache updating systems.

Ans.: The different cache updating systems are :

•Write through system

•Buffered write through system and

•Write-back system

Q.60 What is a hit?

Ans.: A successful access to data in cache memory is called hit.

Q.61 Define cache line or cache block.

Ans.: Cache block is used to refer to a set of contiguous address locations of some size that can be either present or not present in a cache. Cache block is also referred to as cache line.

Q.62 What are the two ways in which the system using cache can proceed for a write operation?

Ans.: Write through protocol technique.

Write-back or copy back protocol technique.

Q.63 What is write through protocol?

Ans.: In write through updating system, the cache controller copies data to the main memory immediately after it is written to the cache. Due to this main memory always contains a valid data, and any block in the cache can be overwritten immediately without data loss.

Q.64 What is write-back or copy back protocol ?

Ans.: In a write-back System, the alter bit in the tag field is used to keep information of the new data. Cache controller checks this bit before overwriting any block in the cache. If it is set, the controller copies the block to main memory before loading new data into the cache.

Q.65 When does a read miss occur?

Ans.: When the addressed word in a read operation is not in the cache, a read miss occur.

Q.66 What is write miss ?

Ans.: During the write operation if the addressed word is not in cache then said to be write miss.

Q.67 What is load-through or early restart ?

Ans.: When a read miss occurs for a system with cache the required word may be sent to the processor as soon as it is read from the main memory instead of loading into the cache. This approach is called load through or early restart and it reduces the processor's waiting period.

Q.68 What is replacement algorithm ?

Ans.: When the cache is full and a memory word that is not in the cache is referenced, the cache control hardware must decide which block should be removed to create space for the new block that contains the reference word. The collection of rules for making this decision constitutes the replacement algorithm.

Q.69 What do you mean by Least Recently Used (LRU)?

Ans.: A replacement scheme in which the block replaced is the one that has been unused for the longest time is called least recently used replacement scheme.

Q.70 What are the various memory technologies ?AU: Dec.-15

Ans.: Computer memory is classified as primary memory and secondary memory. Memory technologies in use for primary memory are static RAM, Dynamic RAM, ROM and its types (PROM, EPROM, EEPROM) etc. Memory technologies in use for secondary memory are optical disk, flash-based removal memory card, hard-disk drives, magnetic tapes etc.

8.5 Virtual Memory

Q.71 What is virtual memory ?AU: Dec.-07, 08, 17, June-07, 09, 10, May-15

Ans.: In modern computers, the operating system moves program and data automatically between the main memory and secondary storage. Techniques that automatically swaps program and data blocks between main memory and secondary storage device are called virtual memory. The addresses that processor issues to access either instruction or data are called virtual or logical address.

Q.72 What is the function of memory management unit?

Ans.: The memory management unit controls this virtual memory system. It translates virtual address into physical address.

Q.73 What is meant by address mapping?AU: May-16

Ans.: The virtually addressed memory with pages mapped to main memory. This process is called address mapping or address translation.

Q.74 What is page fault?

Ans.: Page fault is an event that occurs when an accessed page is not present in main memory.

Q.75 What is TLB (Translation Look-aside Buffer) ?AU: Dec.-11

Ans.: To support demand paging and virtual memory processor has to access page table which is kept in the main memory. To avoid the access time and degradation of performance, a small portion of the page table is accommodated in the memory management unit. This portion is called Translation Lookaside Buffer (TLB).

Q.76 What is the function of a TLB (Translation Look-aside Buffer) ? AU: June-06

Ans.: TLB is used to hold the page table entries that corresponds to the most recently accessed pages. When processor finds the page table entries in the TLB it does not have to access page table and saves substantial access time

Q.77 What is virtual address ?

Ans.: Virtual address is an address that corresponds to allocation in virtual space and is translated by address mapping to a physical address when memory is accessed.

Q.78 What is virtual page number ?

Ans.: Each virtual address generated by the processor whether it is for an instruction fetch is interpreted as a virtual page.

Q.79 What is page frame?

Ans.: An area in the main memory that can hold one page is called page frame.

Q.80 What is MMU?

Ans.: MMU is the Memory Management Unit. It is a special memory control circuit used for implementing the mapping of the virtual address space onto the physical memory.

Q.81 What are pages ?

Ans.: All programs and data are composed of fixed length units called pages. Each page consists of blocks of words that occupies contiguous locations in main memory.

Q.82 What is dirty or modified bit? AU: Dec.-14

Ans.: The cache location is updated with an associated flag bit called dirty bit.

Q.83 An address space is specified by 24-bits and the corresponding memory space by 16-bits:

How many words are there in the a) Virtual memory b) Main memory

Ans.: a) Words in the virtual memory $= 2^{24} = 16$ M words

b) Words in the main memory $= 2^{16} = 64$ K words

Q.84 What do you mean by virtually addressed cache ?AU: May-12

Ans.: The processor can index the cache with an address that is completely or partially virtual. This is called a virtually addressed cache. It uses tags that are virtual addresses and hence such a cache is virtually indexed and virtually tagged.

Q.85 What do you mean by physically addressed cache?

Ans.: A cache that is addressed by a physical address is called physically addressed cache.

Q.86 What is cache aliasing ?

Ans.: In case of virtually address cache. The pages are shared between programs and which may access them with different virtual addresses. Cache aliasing occurs when two addresses are assigned for a same page.

Q.87 Define supervisor/kernel/executive state.AU : May-14

Ans.: The state which indicates that a running process is an operating system process is called supervisor/kernel/ executive state. Special instructions are available only in this state.

Q.88 Define system call.

Ans.: It is a special instruction that transfers control from user mode to a dedicated location in supervisor code space, invoking the exception mechanism in the process.

Q.89 State the advantages of virtual memory ?AU: May-16

Ans.: 1. It allows to load and execute a process that requires a larger amount of memory than what is available by loading the process in parts and then executing them.

2. It eliminates external memory fragmentation.

8.6 Accessing I/O

Q.90 Define interface.

Ans.: The word interface refers to the boundary between two circuits or devices.

Q.91 What is the necessity of an interface? OR What are the functions of a typical I/O interface? AU: Dec.-06, 07, May-07, 09

Ans.: An interface is necessary to coordinate the transfer of data between the CPU and external devices. The functions performed by an I/O interface are:

1. Handle data transfer between much slower peripherals and CPU or memory.

2. Handle data transfer between CPU or memory and peripherals having different data formats and word lengths.

3. Match signal levels of different I/O protocols with computer signal levels.

4. Provides necessary driving capabilities - sinking and sourcing currents.

Q.92 What is an I/O channel ?

Ans.: An I/O channel is actually a special purpose processor, also called peripheral processor. The main processor initiates a transfer by passing the required information in the input output channel. The channel then takes over and controls the actual transfer of data.

Q.93 Name the two interfacing techniques.

Ans.: I/O devices can be interfaced to a computer system I/O in two ways:

• Memory mapped I/O

• I/O mapped I/O

Q.94 What is memory mapped I/O ?AU May-14, Dec.-18

Ans.: The technique in which the total memory address space is partitioned and part of this space is devoted to I/O addressing is called memory mapped I/O technique.

Q.95 What is I/O mapped I/O ?

Ans.: The technique in which separate I/O address space, apart from total memory space is used to access I/O is called I/O mapped I/O technique.

Q.96 What are the components of an I/O interface ? AU Dec.-11

Ans.: The components of an I/O interface are :

- 1. Data register
- 2. Status/control register
- 3. Address decoder and
- 4. External devices interface logic

Q.97 Specify the different I/O transfer mechanisms available.

Ans.: Different I/O transfer mechanisms available are:

- 1. Polling I/O transfer
- 2. Interrupt driven I/O transfer
- 3. DMA transfer
- 4. Serial I/O transfer

Q.98 Distinguish between isolated and memory-mapped I/O.

(Refer section 8.6.3)AU: May-13

Sr. No.	Memory mapped I/O	I/O mapped I/O
1.	Memory and I/O share the entire address range of processor.	Processor provides separate address range for memory and I/O devices.
2.	Usually, processor provides more address lines for accessing memory. Therefore more decoding is required control signals.	Usually, processor provides less address lines for accessing 1/O. Therefore, less decoding is required.
3	Memory control signals are used to control read and write I/O operations.	I/O control signals are used to control read and write I/O operations.

Sr. No.	Memory bus	I/O bus
1.	Memory address bus shares entire address range.	I/O bus shares only I/O address range.
2.	Memory address bus width is greater than I/O address bus width	I/O address bus width is smaller than memory address bus width.
3.	Memory bus includes data bus, address bus and control signals to access memory.	I/O bus includes data bus, address bus and control signals to access I/O.