ID : 14832
Name : Syed tauseef Shah
Submitted to : Sir Amin
Subject : Computer Architecture
Semester 4th
Assignment no 4th

(I) What is the general relationship among access time, memory cost, & capcity:

As Access time becomes faster, the cost per bit increases. As memory size increases the cost per bit is smaller Also with great capcity the access time become slower.

(II) Discuss different memory access methods in detail.

Another ~~discussion~~ distinction among memory types is the method of acessing units of data. These include the following:

* Sequential Access:-

Memory is organized into units of data called records Acess must be made in a specific linear sequence. Stored adaressing information is used to separate records & assist in the retrival process. A shared read-write mechanism is used & this must be moved from its current location to the desired location passing & rejecting each intermediate record. Thus the time to access an asbitary records is highly variables.

Direct Access

As with the sequential access, direct access involves a

Shared read-write mechanism. However individuals blocks or record have a unique address based on physical location. Access is accomplished by direct access to reach a general vicinity plus sequential searching counting or waiting to reach the final location. Again access time is variable.

Random access:-

Each addressable location in memory has unique physically wired in addressing mechanism The time to access a given location is independent of the sequence of prior access & is constant. Thus any location can be selected at random & directly addressed & accessed Main memory & some cache systems are random access.

Associative:- This is a random access type of memory that enable to make a comparison of desired bit locations within word for a specified match & to do this for all word simultaneously. Thus a word is retrived based on a portion of it contents rather its address

As with ordinary random-access memory (R) (B) location, each location has it own memory addressing mechanism, & retrival time is constant independent of location or prior access patterns Cache memory may employ associative access.

III) Discuss the importance of memory hierarchy.

Memory hierarchy particularly important for understanding optimization & performances costs that happen at the hardware level storing data on disk versus main memory can impact running time. The structure of page table virtual memory & lookup cache also play a significant role

IV) How does the principal of locality relate to the use of multiple memory levels?

Slower & less expensive memory is used in higher stages with the most expensive being the registers in the processor as well as cache. Main memory is slower & less expensive & is outside of the processor

(V) How main memory is interrupted in direct associative & set associative mapping?

Direct mapping:-

* its simplest technique.
* Maps each block of main memory into only one possibe cache line.

Associative mapping:-

* permits each main memory block to be loaded into any line of the cache.
* The cache control logic interperts a memory address simply as a tag & a word field
* To determine wheather a block is in the cache the cache control logic must simultaneously examine every line's tag for a match.

* Set - associative mapping:-
* A compromise that exhibits the strength of both the direct & associative approches while reducing their disadvantges.

Q NO 2

write note each of the following.

(I) Memory unit of transfer:-

* Unit of transfer:-

its the maximum number of bits that can be read or written into the memory at a time. In case of main memory. its mostly equaly to word

Size. In case of external memory unit of tranfer is not limited to the word size. its often larger & is reffered to as blocks.

II) Memory performance parameters:.

The two most important characteristics of memory are capcity & performances. Three performances parameter are used.

* Access time (latency):.
For random-access memory, this is the time it takes to perform a read or write operation. that is the time from the instant that data have been stored or made available for use. For non-random access memory. access time is the time it takes to position the read-write mechanism at the desired location.

* Memory cycle time:-
This concept is primarily applied to random access-memory & consists of the access time plus any additional time required before a second access can commence. This additional time may be required for transients to die out on signal line or to regenerates data if they are read destructively ie

Note that memory cycle time is concerned with the system bus. not the processor.

* Transfer rate:-
This is rate at which data can be transferred into or out of a memory unit. For random access memory. it's equal to 1/(cycle time).

III) Disk Cache:-
* A portion of main memory can be used as a buffer to hold data temporarily that is to be read out to disk.
* A few large transfer of data can be used instead of many small transfer of data
* Data can be retrived rapidly from the software cache rather than slowly from the disk.

IV) principal of locality:-
The principal of locality state that data in vicinity of a referenced word are likely to be referenced in the near future.

v) logic cache & physical cache:-
A logic cache also known as a virtual cache store data using virtual addresses. The processor

access the cache directly, without going through the MMU. A physical cache stores data using main memory physical addresses. One obvious advantages of the logical cache is that cache access speed is faster than for a physical cache because the cache can respond before the MMU performs an address translation. The disadvantges has to do with the fact that most virtual memory supply each application with the same virtual memory address space. That is each application sees a virtual memory that starts at address 0. Thus the same virtual address in two different application refers to different physical address.

(VI) Replacement algorithms:-

Once the cache has been filled, when a new block is brought into the cache, one of the existing blocks must be replaced. For direct mapping, there is only one possible line for any particular block & no choice is possible. For the associative

& Set associative technique, a replacment algorithms is needed. To acheive high speed such an algorithms must be implemented in hardware

VII) Possible approches to cache memory:
Possible approches to cache memory coherency include the following.

* Bus watching with write through:-
Each controller monitors the address lines to detect write operations to memory by others bus masters. if another master write a location in shared memory that also resides in the cache memory, the cache controller invalidates that cache memory entry. This strategy depends on the use of a write-through policy by all controllers.

Hardware transparency:-
Additional hardware is used to ensure that all updates to main memory via cache are reflected in all cache. Thus if one processor modifie a word in its cache, this update is written to main memory. In addition any matching a word it in cache are similarly update.

Non-cachable memory:-
only a portion of main memory is shared by more than one processor

& this is designated as non-cacheable. In such a system, all access to shared memory are cache misses, because the shared memory is never copied into the cache. The non cacheable memory can be identified using chip-select logic or high-address bits.

Q NO 3

Differentiate each of the following:

(I) Servential, direct & random access methods.

* Servential access:-

Memory is organized into units of data called records. Access must be made in a specific linear Serviences. Stored addressing information is used to separate records & assist in the retreival process. A shared read write mechanism is used, & this must be moved from its current location to the desire location, passing & rejecting each intermediate record. Thus, the time to access an arbitary record is highly variable.

Direct access:-

As a with sequential access, direct access include shared read-write mechanism. However individuals blocks or records have a unique address based on physical location. Access is accomplished by direct access to reach a general vicinity plus sequential searching, counting, or waiting to reach the final location Again, access time is variable.

Random access:-

Each addressable location in memory has a unique physically wired in addressing mechanism. The time to access a given location is independent of the sequences of prior access & is constant. Thus any location can be selected at random & directly address & accessed. Main memory & some cache systems are random access.

II) Direct, associative & set associative mapping

Direct mapping:-

the direct & associative approches while reducing their disadvantges.

III) Split cache & unified cache

* Split cache:-
* Has become common to split cache
* one dedicated to instructions
* one didicated to data
* BOth exist at the same level, typically as two L1 caches.
* Trend is toward split cache at the L1 & unified cache for highes level.
* Advantges of Split Cache:
* Eliminates cache contention between instruction fetch/decode unit & execution unit.
* important is pipeling
* unified cache
* Trend is toward unified cache for higher levels
* Advantges for unified cache
* Higher hit rate
* Balances load of instructions & data fetches automatically
* only one cache needs to be designed & implemented.

VI) Write through & write back

* write through
* Simplest technique
* All write operations are made to main memory as well as to cache.

The direct mapping technique is simple & inexpensive to implement. Its main disadvantges is that there is a fixed cache location for any given block. Thus if a program happen to reference word repeatedly from two different blocks that map into the same line then the blocks will be continually swapped in the cache, & the hit ratio will be low (a phenomen known as thrashing).

* Associative mapping:-

With associative mapping, there is flexibility as to which block to replace when a new block is read into the cache. Replacements algorithms, discussed later in this section are designed to maximize the hit ratio. The principal disadvantges of associative mapping is the complex circuitry required to examine the tags of all cache line in parallel.

* Set-associative mapping

Set associative mapping is a compromise that exhibits the strength of both

* The main disadvantges of this technique is that it generates substanial memory traffic & may create bottleneck.
* Write back
* Minimize memory writes
* Updates are made only in the cache.
* portion of main memory are invalid & hence access by I/O modules can be allowed only through the cache.
* This make for complex circuitry & a potential bottleneck.

Q NO 4

Solve each of the following.

Suppose that the processor has access to two levels of memory level-1 contains 1000 words & has an access time of 0.01 μs level-2 contains 100,000 words & has an access time of 0.1 μs Assume that if a word to be accessed is in level 1 then the processor access it directly if it is in level 2 then the word is first transferred to level 1 & then accessed by the processor. Suppose 95% of the memory access are found in level 1 Then find the average time to access a word.

In our example, Suppose 95% of the memory access are found in level 1

Then the average time to access a word can be expressed as

$(0.95)(0.01ms) + (0.05)(0.01ms + 0.1ms) = 0.0095 + 0.0055$

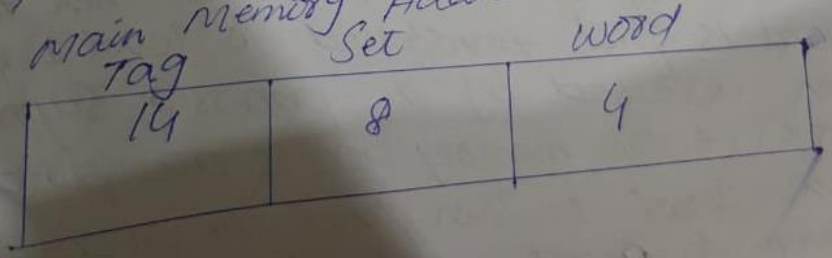
$= 0.015\ ms$

The average access time is much closer to 0.01 ms than to 0.1ms as desired

(II) A two way set associative cache has lines of 16 bytes & a total size of 8K bytes the 64Mbytes main memory is byte addresable Show the format of main memory address.

There are a total of 8K bytes / 16 bytes = 512 lines in the cache. Thus the cache consists of 256 sets of line each therefore 8 bits needed to identify the set numbers For the 64Mbytes main memory a 26-bit addressed is needed. Main memory consists of 64 Mbytes / 16 bytes = 2^{22} blocks. Therefore the set plus tag length must be 22 bits so the tag length is 14 bits & the word field length is 4 bits.

Main Memory Address =



Q NO #4

Address (H) BBBBBB

Address (binary) 101110111011101110111011

(a) Tag (8)/line (14)/word 2 BBH/2EEH/3H

(b) Tag (22)/word (2) 2EEEEEH/3H

(c) Tag (9)/Set (13)/word (2) 177H/0EEH/3H