Name Shahab Zaib

ID 14590

deptt Computer Science.

Assignment No 04

Date 14/6/2020.

Q:NO:01

I

When access time become faster cost increases As memory size increase the cost per bit is smaller with greater capacity the access time become slower

II

Memory access methods are.

Sequencial access:

Memory is organized into small units of data called records. Access must be made in a linear sequence stored addressing information is used to separater records

and assist in the relevant process. A shared read mechanism is used and must be moved from its current to desired location, passing and rejecting each intermediate record.

Direct Access:

Direct access involves shared read write mechanism. However individual blocks or record have a unique address based on physical address. Can be direct access to reach a general vicinity plus Sequential. Time is variable.

Random Access:

Each location in memory has a unique physical wired-in address mechanism.

The time to access a location given is independant of the sequence prior and is constant. Thus any location can be access random and directly. Main memory and some cache systems are random access.

Associative:

The is a random access type of memory that enables one to make a comparison of desired bit locations within a word for a specified match and to do this for all words simultaneously. Thus a word is retrived based on a portion of its contents rather than its address. As with ordinary random access memory.

each location has its own addressing mechanism and retrieval time is independent and constant of location prior access pattern. cache memory may employ associative access.

III

Memory hierarchy is used for understanding optimization and performance costs that happen at hardware level. Runing time can be impact by storing data on disk vs main memory the structure of page tables virtual memory, and lookup caches also play a significant role.

IV:

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

V

1. Direct Mapping:
- It is the simplest technique
- Maps each block of main memory into only one possible cache line

Associative Mapping:
- Permits each main memory block to be loaded into any line of the cache
- The cache control logic

Interprets a memory address simply as a tag and a word field.

- To determine whether a block is in the cache the cache control logic must simultaneously examine every lines tag for a match

Set-Associative Mapping:

- A compromise that exhibits the strengths of both the direct and associative approaches while reducing their disadvantages.

Q:No:02:

I:

Unit of transfer:

It is the maximum number of bits that can be read or write into the memory at a

time. In case of main memory, it is mostly equal to word size. In case of external memory, unit of transfer is not limited to the word size. It is often larger and refer to as blocks.

II:

Memory performance parameters:

Two important characteristics of memory are capacity and performance.

Access time (Latency):

It is the time for instant that an address is presented to the memory to the instant that data have been stored or made available for use.

For random access memory.

this is the time taken to perform read write operations.

For non random it is the time takes to read write mechanism into desired location.

Memory Cycle Time:

Memory cycle time is applied to random access memory and consist of access time plus any additional time required before a second access can commence. Additional time required may be of regenerating data if they are read destructive or transient to die out on signal lines. This is concerned with signal of system and bus not processor.

Transfer rate:

The rate at which data can be transfer into or out of memory unit.

III:

Disk cache:

- A portion of main memory can be used as a buffer to hold data temporarily that is to be read out ~~of~~ to disk
- Transfer of few large data can be used instead of many small transfer data
- Data can be retrive rapidly from the software cache rather than slowly from the disk

V:

Logical cache and physical cache.

Also known as

virtual cache. It stores 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. The advantage of logical cache is that cache access speed is faster than that of physical because the cache can respond before the MMU performs an address translation.

Disadvantage is that most virtual memory system supply each application with same virtual memory address space, each application sees a virtual memory that start at address 0. Thus same virtual address in two different application

refer to two different physical addresses.

(VI):

Replacement Algorithms:

When the cache is filled, a new block is brought into cache, one of existing blocks must be replaced. For direct mapping there is only one possible line for new block, and no choice is possible. For associate and set-associate a replacement algorithm is needed. For high speed, such algorithm must be implemented in hardware.

VII:

Possible approaches to Cache Coherency:

There are some of the following.

III

Bus watching with write through:

The address lines are monitored by cache controller to detect write operations ~~by~~ to memory by other bus masters. If another writes to a location in shared memory that also resides in the cache memory, the cache controller invalidates that cache entry. This strategy depend on the use of write through policy by all cache controller.

Hardware Transparency:

Additional hardware is used to ensure that all updates to main memory via cache are reflected in all caches.

Thus if one processor modifies a word in its cache. This update is written to main memory. In addition any matching word in other caches are similarly updated.

Non Cacheble Memory:

Non Cacheble design mean that only one portion of main memory is shared by more than one processor. In such systems all accesses to main memory are cache misses, because the shared memory is never copied into the cache. The non cacheble memory can be identified using chip select logic or high address bit

Q:NO: 03

(I)

Memory access methods are:

Sequencial access:

Memory is organized into small units of data called records. Access must be made in a linear sequence stored addressing information is used to separates records

and assist in the retrieval process. A shared read mechanism is used and must be moved from its present to desired location, passing and rejecting each intermediate record.

Direct Access:

Direct access involves shared read write mechanism. However individual blocks or record have a unique address based on physical address. Can be direct access to reach a general vicinity plus sequential. Time is variable.

Random Access:

Each location in memory has a unique physical wised in address mechanism.

The time to access a location given is independant of the sequence prior and is constant. Thus any location can be access randomly directly. Main memory and some cache systems are random access.

IV

1 Direct Mapping:

- It is the siplest technique
- Maps each block of main memory into only one possible cache line

Associative Mapping:

- Permits each main memory block to be loaded into any line of the cache
- The cache control logic

Interprets a memory address simply as a tag and a word field.

- To determine whether a block is in the cache the cache control logic must simultaneously examine every lines tag for a match

Set-Associative Mapping:

- A compromise that exhibits the strengths of both the direct and associative approaches while reducing their disadvantages.

III:

Split cache:

- One dedicated to Instructions.
- One dedicated to data
- Both exist at the same level, typically as two L1 cache
- Trend is toward split cache at the L1 and unified caches for higher levels.
- Advantages:

=> Eliminates cache contention between instruction fetch/decode unit and execution unit

=> Important in pipelining.

Unified Cache:

- Trend is toward unified cache for higher level
- Advantages of unified cache

I) Balance load of instruction and data fetches automatically.

II) Only one cache need to be designed and implemented.

IV:

Write through:

- Simplest technique.
- All write operations are made to main memory as well as to the cache
- The main disadvantage of this technique is

that it generates substantial memory traffic and may create a bottleneck

Write back:

- Minimizes memory writes.
- updates are made only in the cache.
- Portion of main memory are invalid and hence accesses by I/O modules can be allowed only through the cache

This makes for complex circuitry and a potential bottleneck