IQRA NATIONAL UNIVERSITY

INU

DEFINING POSSIBILITIES, CREATING OPPORTUNITIES

Name : Muhammad Musa

Department : BS(CS)

Semester : 4th

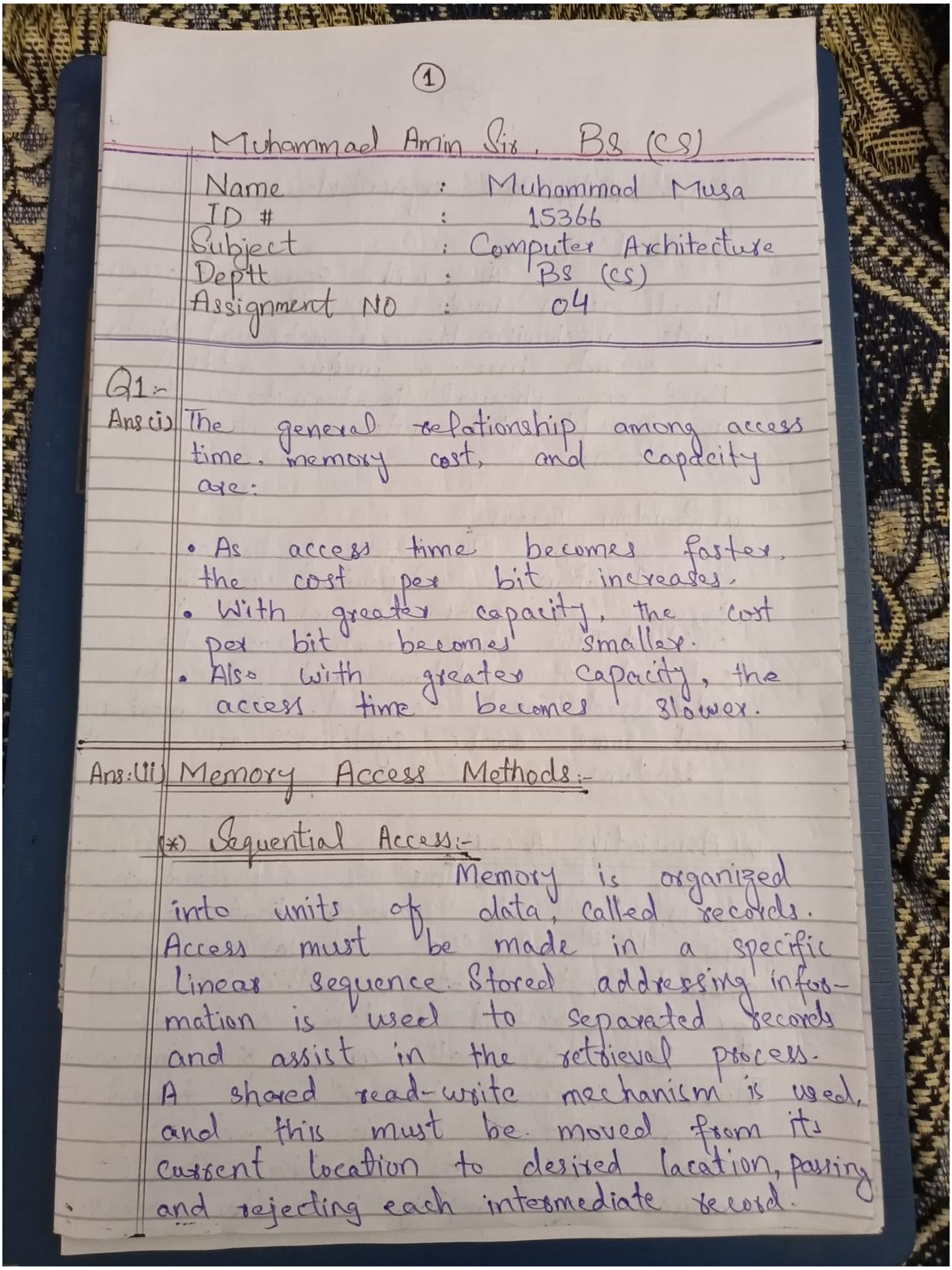
ID # : 15366

Sessional Assignment No : 4th

Subject : Computer Architecture

Submitted To :Muhammad Amin Sir

Dated : 13th June 2020



Muhammad Amin Sir, Bs (CS)

Name	:	Muhammad Musa
ID #	:	15366
Subject	:	Computer Architecture
Deptt	:	Bs (CS)
Assignment NO	:	04

Q1:-

Ans (i) The general relationship among access time, memory cost, and capacity are:

- As access time becomes faster, the cost per bit increases.
- With greater capacity, the cost per bit becomes smaller.
- Also with greater capacity, the access time becomes slower.

Ans:(ii) Memory Access Methods:-

(*) Sequential Access:-

Memory is organized into units of data, called records. Access must be made in a specific linear sequence. Stored addressing information is used to separated records and assist in the retrieval process. A shared read-write mechanism is used, and this must be moved from its current location to desired lacation, passing and rejecting each intermediate record.

Tape units are sequential access.

(*) Direct Access:-

As with sequential access, direct access involves a shared read-write mechanism. However, individual blocks or records have a unique address based on physical location. Access time is variable. Dist units are direct access.

(*) Random Access:-

The time to access a given location is independent of the sequence of prior accesses and is constant. Thus, any location can be selected at random and directly addressed and accessed. Main memory and some cache systems are random access.

(*) Associative Access:-

This 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 retrieved based on a portion of its contents rather them its address. Cache memories may employ associative access.

Ans: (iii) Importance of Memory hierarchy:-

- It is possible to organize data across the hierarchy such that the percentage of accesses to each successively lower level is substantially less than that of the level above.
- The three forms of memory just described are, volatile and employ Semiconductor technology. The use of three levels exploits the fact that semiconductor memory comes in a variety of types, which differ in speed and cost.
- Secondary memory or auxiliary memory are used to store program and data files and are usually visible the programmer only in terms of files and records.

(iv) Ans: (iv) Slower also less expensive memory is utilized within higher stages, for the majority expensive continously the registers in the processor and additionally reserve. Fundamental memory may be slower and less expensive, furthermore will be outside of the processor.

Ans:(v)

Direct Mapping:-

The simplest techniques, known as direct mapping, maps each block of main memory into only one possible cache line. The mapping is expressed as:

$$i = j \text{ modulo } m$$

Associative Mapping:-

Associative mapping overcomes the disadvantages of direct mapping by permitting each main memory block to be loaded into any line of the cache.

In this case, the cache control Logic interprets a memory address simply as a Tag and a Word field.

Set-Associative Mapping:-

Set-Associative mapping is a compromise that exhibits the strengths of both the direct and associative approaches while reducing their disadvantages.

In this case, the cache consists of number sets, each of which consists of a number of lines. The relationships are

$$m = v \times k$$

$$i = j \text{ modulo } v$$

Q2:-

Ans: (i) **Memory Unit of Transfer:-**

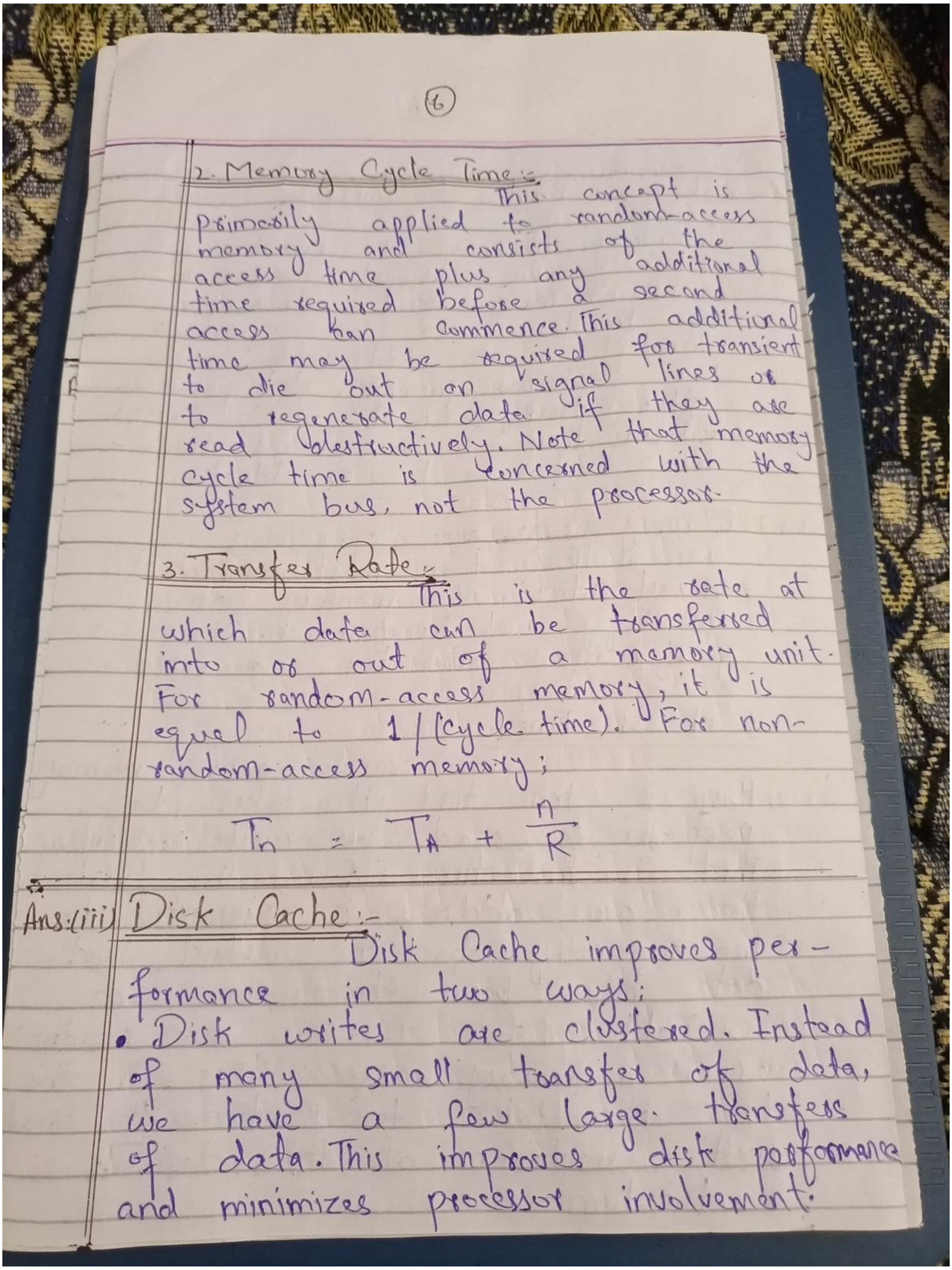
For main memory, this is the number of bits read out of or written into memory at a time. The unit of transfer need not equal a word or an addressable unit. For external memory, data are often transferred in much larger units than a word, and these are referred to as blocks.

Ans: (ii) **Memory Performance Parameters:-**

The two most important characteristics of memory are capacity and performance. Three performance parameters are used:

1. **Access Time:-**

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 are address is presented to the memory to 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.



2. Memory Cycle Time:-

This concept is primarily applied to random-access memory and consists of the access time plus any additional time required before a second access kan commence. This additional time may be required for transient to die out on signal lines or to regenerate data if they are read destructively. Note that memory cycle time is concerned with the system bus, not the processor.

3. Transfer Rate:-

This is the rate at which data can be transferred into or out of a memory unit. For random-access memory, it is equal to 1/(cycle time). For non-random-access memory;

$$T_n = T_A + \frac{n}{R}$$

Ans:(iii) Disk Cache:-

Disk Cache improves performance in two ways:

- Disk writes are clustered. Instead of many small transfer of data, we have a few large transfers of data. This improves disk performance and minimizes processor involvement.

- Some data destined for write-out may be referenced by a program before the next dump to disk. In that case, the data are retrieved rapidly from the software cache rather than slowly from the disk.

Ans:(iv) Principle of Locality:- The principle of locality states that data in the vicinity of a referenced word are likely to be referenced in the near future.

OR

An implication of locality is that we can predict with reasonable accuracy what instructions and data a program will use in the near future based on its accesses in the recent past.

Ans:(v) Logical Cache and Physical Cache:-

A Logical cache, also known as a virtual cache, stores data using virtual addresses. The processors accesses the cache directly, without going through the MMU.

A physical cache stores data using main memory physical addresses.

Advantage of 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 disadvantage has to do with the fact that most virtual memory systems supply each application with the same virtual memory address space.

Ans:(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 possible line for any particular block, and no choice is possible. For the associative and set-associative techniques, a replacement algorithm is needed. To achieve high speed, such an algorithm must be implemented in hardware.

Ans.(vii) **Possible approaches to Cache Coherency:-**

Possible approaches to cache coherency include the following:

- **Bus watching with write through:-**

Each cache controller monitors the address lines to detect write operations to memory by other bus masters.

- **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.

- **Noncacheable Memory:-** Only a portion of main memory is shared by more than one processor, and this is designated as noncacheable. In such a system, all accesses to shared memory are cache misses, because the shared memory is never copied into the cache.

Q3:-

Ans: (i) **Sequential Access:-** Memory is organized into units of data, called records. Access must be made in a specific linear sequence. Stored addressing information is used to separate records and assist in the retrieval process.

Direct Access:- As with sequential access, direct access involves a shared read-write mechanism. However, individual 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.

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 sequence of prior accesses and is constant.

Ans:(ii) 1) Direct Mapping:- The direct mapping technique is simple and inexpensive to implement. Its main disadvantage is that there is a fixed cache location for any given block. Thus, if a program happens to reference words repeatedly from two different blocks that map into the same line.

Associative Mapping:- With associative mapping, there is flexibility as to which block to replace when a new block is read into the cache. Replacement algorithms, discussed in this Section, are designed to maximize the hit ratio. The disadvantage is that the complex circuitry required to examine the tags of all cache lines in parallel.

Set-Associative Mapping:-

For set-associative mapping, the cache control logic interprets a memory address as three fields: Tag, Set and Word. The d set bits specify one of $v = 2^d$ sets. With fully ~~a~~ set-associative mapping, the tag in a memory address is much smaller and is only compared to the k tags within a single set.

(iii) Ans: Split Cache and Unified Cache:-

→ Has become common to split Cache:

- One dedicated to instructions.
- One dedicated to data.
- Both exist at the same level, typically as two L1 caches.

→ Advantages of Unified Cache:

- Higher hit rate because it balances the load between instruction and data fetches automatically.

- Only one cache needs to be designed and implemented.

→ Advantages of Split cache:

- Eliminates cache contention between instruction fetch/decode unit and execution unit.
- This is important in any design that relies on the pipelining of instructions.

Ans: (iv) Write Through and Write Back.

→ 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.
- Portions of main memory are invalids and hence accesses by I/O modules can be allowed only through the cache.

Q4:-

Ans:(i) In example, 95% of the memory accesses are found in level 1. Then the average time to access a word can be expressed as;

$$(0.95)(0.01\mu s) + (0.05)(0.01\mu s + 0.1\mu s)$$

$$= 0.0095 + 0.0055$$

$$= 0.015\ \mu s$$

The average time is much closer to 0.01μs than to 0.1μs, as desired.

Ans:(ii) There are a total of 8kbytes/16 byte = 512 lines in the cache. Thus the cache consists of 256 sets of 2 lines each. Therefore 8 bits are needed to identify the set number. For 64-Mbyte main memory, a 26-bit address is needed. Main memory consists of 64-Mbyte/16 bytes = 2^{22} blocks. Therefore the set plus tag lengths must be 22 bits, so the tag length is 14bits and word field length is 4bits.

Main memory Address=

TAG	SET	WORD
14	8	4