

1. Discuss principles of locality with example.

Answer: In computer science, locality of reference, also known as the principle of locality, is the tendency of a processor to access the same set of memory locations repetitively over a short period of time. The principle of locality can be further characterized by the following types of locality:

Temporal locality - Memory that is accessed once tends to be accessed again soon thereafter. An example would be an iterator variable (loop variable) within a loop. It is accessed during each iteration.

Spatial locality - Items whose addresses are near that of an accessed item's address tend to be referenced soon. An example would be elements within an array structure. In many cases, accessing the first element of an array directly precedes the access of the next element.

2. Discuss memory hierarchy with examples.

Ans: In computer architecture, the memory hierarchy separates computer storage into a hierarchy based on response time.

The Computer memory hierarchy looks like a pyramid structure which is used to describe the differences among memory types. It separates the computer storage based on hierarchy.

Level 0: CPU registers.

Level 1: Cache memory.

Level 2: Main memory or primary memory.

Level 3: Magnetic disks or magnetic types or tertiary memory.

Level-0 Registers:

The registers are present inside

the CPU, As they

are present

inside the CPU,

CPU registers

Level-0

Cache

Level-1

Main Memory

Level-2

Secondary Memory

Level-3

Tertiary Memory

Level-4

they have least access time. Registers are most expensive and smallest in size generally in Kilobytes. They are implemented by using Flip-Flops.

**Level-1-Cache:** Cache memory is used to store the segments of a program that are frequently accessed by the processor. It is expensive and smaller in size generally in Megabytes and is implemented by using Static RAM.

**Level-2- Primary or Main Memory:** It directly communicates with the CPU and with auxiliary memory devices through an I/O processor.

Main memory is less expensive than cache memory, and larger in size generally in Gigabytes. This memory is implemented by using dynamic RAM.

**Level-3- Secondary Storage:** Secondary storage devices like Magnetic Disk are present at level 3. They are used as backup storage. They are cheaper than main memory and larger in size generally in a few TB.

**Level-4- Tertiary Storage:** Tertiary storage devices like magnetic tape are present at level 4.

They are used to store removable files and are the cheapest and largest in size (1-20TB).

5. Discuss memory writes handling process.

Ans: A memory unit stores binary information in groups of bits called words. Data input lines provide the information to be stored into the memory, Data input lines carry the information out from the memory.

The control lines Read and Write specifies the direction of transfer of data.

Memory Write Operation: Memory write operation transfers the address of the desired word to the address lines, transfers the data bits to be stored in memory to the data input lines. Then it activates the write control line. Description of the write operation is given below:

MDR

3D

MAR

2003

Memory

4B

5C

3K

2002 2003

2002 2003

2002 2003

Before Execution

MDR

3D

MAR

2003

Memory

4B

5C

3K

2002 2003

2002 2003

2002 2003

After Execution

In the above diagram, the MAR contains 2003 and MDR contains 3D. After the execution of write instruction 3D will be written at 2003 memory location.

#### 7. Discuss RAID Levels.

Ans: RAID (Redundant array of independent disks) is a way of storing the same data in different places on multiple hard disks or solid state drives (SSDS) to protect data in the case of a drive failure. There are different RAID levels, however, not all have the goal of providing redundancy.

1. RAID 0: This configuration has striping but no redundancy of data. It offers the best performance, but it doesn't provide fault tolerance.

2. RAID 1: Also known as disk mirroring, this configuration consists of at least two drives that duplicate the storage of data. There is no striping. Read performance is improved, since either disk can be read at the same time. Write performance is the same as for single

disk storage.

3. RAID2: This configuration uses striping across disks, with some disks storing error checking and correcting (ECC) information. RAID2 also uses a dedicated Hamming code parity, a linear form of ECC. RAID2 has no advantage over RAID3, and is no longer used.

4. RAID3: This technique uses striping and dedicates one drive to storing parity information. The embedded ECC information is used to detect errors. Data recovery is accomplished by calculating the exclusive information recorded on other drives. Because an I/O operation addresses all the drives at the same time, RAID3 cannot overlap I/O. For this reason, RAID3 is best for single-user systems with long record applications.

5. RAID 4: This level uses large stripes, which means a user can record read records from any single drive. Overlapped I/O can then be used for read operations. Because all write operations are required to update the parity drive, no I/O overlapping is possible.

6. RAID 5: This level is based on parity block-level striping. The parity information is striped across each drive, enabling the array to function, even if one drive were to fail. The array's architecture enables read and write operations to span multiple drives. This results in performance better than that of a single drive, but not as high as a RAID 0 array. RAID 5 requires at least three disks, but it is often recommended to use at least five disks for performance reasons.

RAID 5 arrays are generally considered to be a poor choice for use on write-intensive systems because of the performance impact associated with writing parity data. When a disk fails, it can take a long time

to rebuild a RAID 5 array.

7. RAID 6: This technique is similar to RAID 5, but it includes a second parity scheme distributed across the drives in the array.

The use of additional parity enables the array to continue functioning, even if two disks fail simultaneously. However, this extra protection comes at a cost.

RAID 6 arrays often have slower write performance than RAID 5 arrays.

8. Write the advantages of RAID system.

Ans: Advantages of RAID include the following:

1. Improved cost-effectiveness because lower-priced disks are used in large numbers.

2. Using multiple hard drives enables RAID to improve the performance of a single hard drive.

3. Increased computer speed and reliability after a crash, depending on the configuration.

4. Reads and writes can be performed faster than with a single drive with RAID 0. This is because a file system is split up and distributed across drives that work together on the same file.

5. There is increased availability and resiliency with RAID 5. With mirroring, two drives can contain the same data, ensuring one will continue to work if the other fails.

10. Write the benefits of proper memory management system.

Ans: The benefits of proper memory management system is given below:

1. There is no fragmentation in internal memory.

2. If the segments are small and it should not be joined into one page, it may be saved into its memory.

3. It is less overhead.

4. It is responsible for managing the computer's primary memory.

5. It keeps track of each as well as every memory location.

6. The average segment is greater than the average page size.

7. In terms of segment table only one as possible entry per actual segment as opposed to per page in virtual memory.

### Assignment - 2

1. What do you mean by exception handling.

Discuss its necessity.

Ans: An exception is the unwanted event that interrupts the normal flow of program.

When an exception occurs program execution gets terminated. In such cases we get a system generated error message. The good thing about exceptions is that they can be handled in Java.

By handling the executions exceptions we can provide a meaningful message

to the user about the issue rather than a system generated message, which may not be understandable to a user.

Necessity of Exception Handling is given below:

Exception handling ensures that the flow of the program doesn't break when an exception occurs.

For example, if a program has bunch of statements and an exception occurs mid way after executing certain statements then the statements after the exception will not execute and the program will terminate abruptly.

By handling we make sure that the statements execute and the flow of program doesn't break.

Assignment - Q1 (Question - 3)

Q1 Differentiate between cache hit and cache miss?

A cache is used to increase the throughput and improve performance of an application.

A cache hit is a state in which data requested for processing by a application is found in the cache memory. It is a faster means of delivering data to the processor, as the cache

already contains the requested data.

A cache hit serves data more quickly, as the data can be retrieved by reading the cache memory. The cache hit also

can be in disk caches where the requested data is stored and accessed at first query.

A cache miss is a request for data from the cache that cannot be filled because the data is not present in the cache memory.

Cache miss slows down the overall process because after a cache miss, the CPU will look for a higher level cache, such as L1, L2, L3, RAM for that data.

Q) Discuss about Virtual Memory.

Ans: The main memory can act as a "cache" for the secondary storage, usually implemented with magnetic disks. This technique is called virtual memory. Historically - there

were 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. Considered a collection of program  
-s running all at once on a computer. To allow  
multiple programs to share the same memory,  
we must be able to protect the programs from  
each other, ensuring that a program can only  
read and write the portions of the many  
programs, just as a cache contains only the  
active portion of one program. Thus, the principle  
of locality enable virtual memory as well as  
caches and virtual memory allows us to  
efficiently share the processor as well as the primary memory  
main memory. The second motivation for virtual memory  
is to allow a single user program to exceed the size of

4) Discuss the memory-mapped system with figure-

- Direct Mapping
- Fully Associative Mapping
- Set Associative Mapping.

1. Cache Direct Mapping: In direct mapping each block from main memory has only one possible place in each cache organization in this technique. For example: every block  $i$  of the main memory can be mapped to block  $j$  of the cache using formula:

$$j = i \bmod n$$

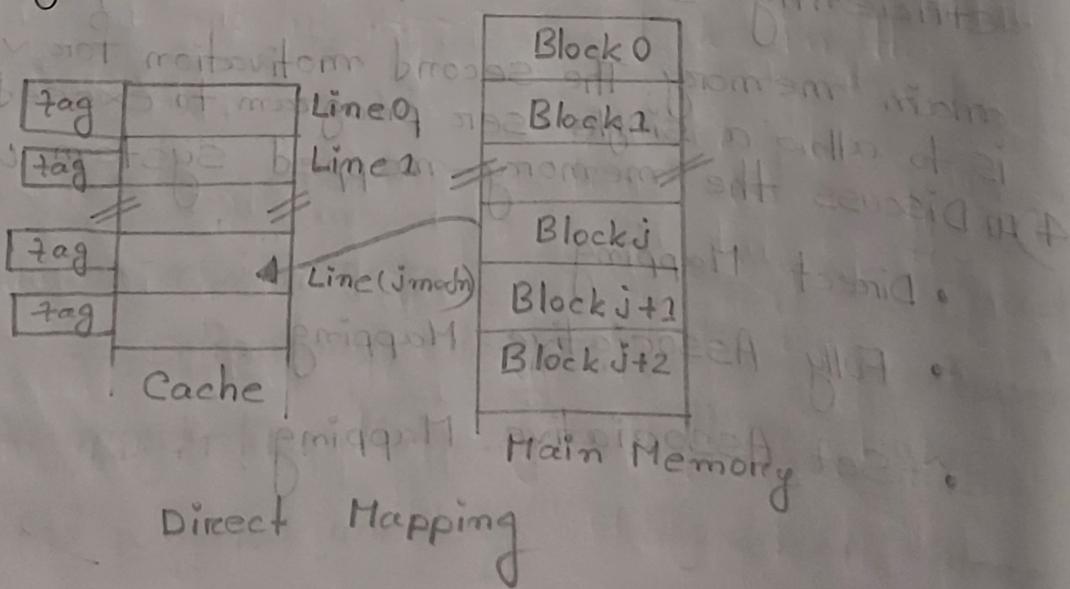
Where  $i$  = main memory block number

$j$  = cache block number.

$n$  = number of blocks in the cache.

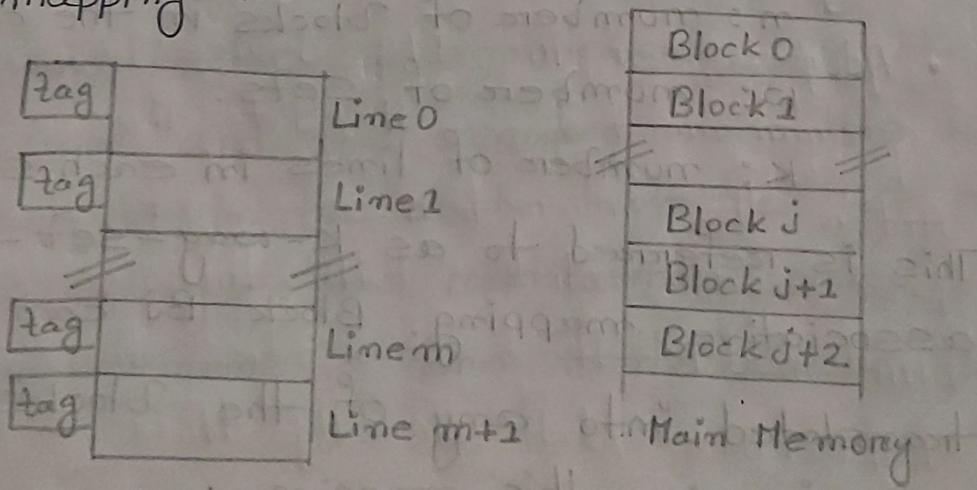
Consider cache memory is divided into ' $n$ ' number of lines. Then block ' $j$ ' of main memory can map to line number ( $j \bmod n$ )

only of the cache.



Hopping the memory address to cache: The block field of the address is used to access the cache's block. Then the tag bits in the address are compared with the tag of block. For a match, a cache hit occurs as the required word is found in the cache. Otherwise, a cache miss occurs and the required word has to be brought in the cache from the main memory. The word is now stored in the cache together with the new tag. (old tag is replaced)

2. Fully Associative Mapping: In fully associative mapping, a block of main memory can map to any line of the cache that is freely available at that moment. This makes fully associative mapping more flexible than direct mapping.



Cache  
Fully Associative Mapping

In Fully Associative Mapping -

- All the lines of cache are freely available.  
Thus, any block of main memory can map to any line of the cache.

- Had all the cache lines been occupied, then one of the existing blocks will have to be replaced.

3. Set Associative Mapping: It is the combination of advantages of both direct &

associative mapping. Hence, the cache consists of a number of blocks. The relationships

$$\text{are: } n = \omega * K$$

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

where

$i$ : cache set number.

$j$ : main memory block number

$n$ : number of blocks in the cache

$\omega$ : number of sets.

$k$ : number of lines in each set.

This is referred to as  $K$ -way-set-

associative mapping. Block  $B_j$  can be

translated into any of the blocks in set  $j$  using this mapping.

Example: Consider the following example of 2-way set associative mapping -

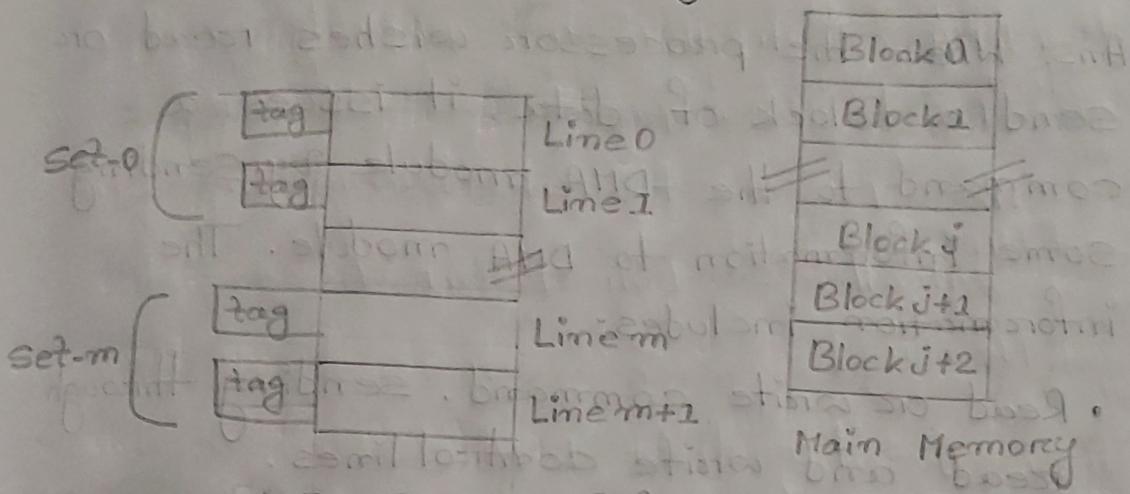


Fig-2-way Set Associative Mapping

Mapping the memory address to cache: Using the set field in the memory address, we access the particular set of the cache. Then, the tag is bits in the address are compared to with the tag of all k blocks within that set. For a match a cache hit occurs as the required word is found in the cache. Otherwise, a cache miss occurs and the required word has to be brought in the cache from the Main Memory. According to the replacement policy used, a replacement is done if the cache is full.

## Assignment - 2

### 3. Discuss about DMA operations. (X)

Ans: When the processor wishes read or send a block of data, it issues a command to the DMA module by sending some information to DMA module. The information includes:

- Read or write command, sending through read and write control lines.
- Number of words to be read or written, communicated on the data lines and stored in the data count register.
- Starting location in memory to read from or write to, communicated on data lines and stored in the address register.
- Address of the I/O device involved, communicated on the data lines.

After the information are sent, the processor continues with other work. The DMA module then transfers the entire block of data directly to or from memory without going through the processor.

When the transfer is complete, the DMA module sends an interrupt signal to the processor to inform that it has finished using the system bus.

4. Write down the advantages and disadvantages of using DMA.

Answer: Advantages of DMA:

- Computer system performance is improved by direct transfer of data between memory and I/O devices, bypassing the CPU.
- CPU is free to perform operations that do not use system buses.
- Fast memory transfer of data.
- CPU and DMA run concurrently under cache mode.
- DMA can trigger an interrupt, which frees the CPU from polling the channel.
- Improves response time for devices with large amounts of data transfers (Disks, Drums)
- Reduces CPU overhead.

→ Data transferred directly to memory.

→ DMA controller has priority over CPU to user processor - Memory Bus.

The CPU must wait during the DMA Controller's Infrequent Access to Memory.

#### Disadvantages of DMA:

→ In case of Burst Mode data transfer the CPU is rendered inactive for relatively long periods of time.

→ DMA is useful only for Data commands All non-data commands have to be executed by CPU.

→ Data has to be stored in continuous locations in memory.

→ CPU's intervention is required for initializing DMA logic for every continuous data block transfer. In other words Data changing is not possible.

→ DMA controller is much slower than the main CPU.

2. Discuss Three primary data transfer mechanism of DMA.

Ans: Three primary data transfer mechanism of DMA are given below:

1. Polling
2. Interrupt Driven
3. DMA

1. Polling: Polling is a form of foreground data acquisition in which the processor is dedicated to acquiring the incoming data. often by waiting in a loop.

2. Interrupt Driven: CPU is kept busy unnecessarily. This situation can very well be avoided by using an interrupt driven method for data transfer. By using interrupt facility and special commands to issue an interrupt request. Signal whenever data is available from any device. The CPU can proceed for any other program execution. The interface keeps monitoring the

device. Whenever it is determined that the device is ready for data transfer initiates an interrupt request signal to the computer.

3. DMA: The data transfer between a fast storage media such as magnetic disk and memory unit is limited by the speed of the CPU. Peripherals directly communicate with each other using the memory bus removing the intervention of the CPU. This type of data transfer technique is known as DMA or Directly Memory Access.

7. Write about DMA channels and their uses.

Ams: Direct Memory Access (DMA) channels are system pathways used by many devices to transfer information directly to and from memory. DMA channels are not nearly as "famous" as IRQs as system resources go.

This is mostly true for a good reason: There are fewer of them and they are used by many fewer problems. with system setup. However, conflicts on DMA channels can cause very strange system problems and can be very difficult to diagnose. DMA's are used most commonly today by floppy disk drives, tape drives and sound cards.

11. Differentiate between hardware control and Microprogrammed control system.

Ans:

| Hardware Control Unit  | Microprogrammed Control Unit  |
|--|---|
| 1. The hardware control unit induces the control signals required for the process. | 1. The microprogrammed control unit induces the control signals through micro instructions. |
| 2. Hardware control unit is faster than a microprogrammed control unit.            | 2. Microprogrammed control unit is slower than hardware control unit.                       |

## Hardwired Control Unit      Microprogrammed Control Unit

- |  |   |
|--|---|
| 3. It is hard to modify.   | 3. It is easy to modify.                                      |
| 4. It is more expensive as compared to the microprogrammed control unit.                                       | 4. It is affordable as compared to the hardware control unit. |
| 5. It faces difficulty in managing the complex instructions because the design of the circuit is also complex. | 5. It can easily manage complex instruction.                  |
| 6. It can use limited instructions.  | 6. It can generate control signals for many instructions.     |

6. Discuss about virtual memory.

Ans: The main memory can act as a "cache" for the secondary storage, usually implemented with magnetic disks. This technique is called Virtual Memory. Virtual Memory is a common technique used in a computer's operating system. Virtual memory uses both hardware and software to enable a computer to compensate for physical memory shortages, temporarily transferring data from random access memory (RAM) to disk storage. Mapping chunks of memory to disk files enables a computer to treat secondary memory as though it were main memory.

Today most personal computers (PCs) come with at least 8 GB of RAM. But, sometimes this is not enough to run several programs at one time. This is where virtual memory comes in. Virtual Memory frees up RAM by swapping

data that has not been used recently over to a storage device such as a hard drive or solid-state-drive (SSD).

7. Write about DMA channels and their uses.

Ans: Direct Memory Access (DMA) channels

are system pathways used by many devices to transfer information directly

to and from memory. DMA channels are not nearly as "famous" as

IRQs as system resources go. This is mostly for a good reason. There are

fewer of them and they are used by many fewer devices and hence

they usually cause fewer problems.

with system setup. However, Conflicts

on DMA channels can cause very

strange system problems and can be

very difficult to diagnose. DMAs are

used most commonly today by

floppy disk drives, tape drives and

## Sound cards.

The table below provides information about the 8DMA channel numbers in a typical PC.

DMA channel usage organized by device instead of DMA number.

| DMA | Bus Line   | Typical Default Use              | Other Common Use  |
|-----|--|----------------------------------|---|
| 0   | No   | Memory Refresh                   | None  |
| 1   | font face<br>= "Arial,<br>Arial, Hei<br>vetica") | 8/16-bit Sound Card<br>(Low DMA) | SCSI host adapters,<br>ECP Parallel ports, tape<br>accelerator cards,<br>network cards, voice<br>modems.  |
| 2   | 8/16-bit   | Floppy disk controller           | Tape accelerator cards.   |
| 3   | 8/16-bit   | None                             | ECP parallel ports, SCSI<br>host adapters, tape acce<br>-lerator cards, sound<br>card (Low DMA) network<br>cards, hard disk contro<br>ller on old PC/XT |
| 4   | No   | None; Cascade<br>for DMAs 0-3    | None  |
| 5   | 16-bit<br>only                                   | None                             | Sound cards (high DMA)<br>network cards   |
| 6   | 16-bit<br>only                                   | None                             | Sound   |

| DMA | Bus Line    | Typical Default Use   | Other Common Use                         |
|-----|-------------|-----------------------|--|
| 5   | 16-bit only | Sound card (high DMA) | SCSI host adapters<br>network cards      |
| 6   | 16-bit-only | None                  | Sound Cards (high DMA)<br>network cards. |
| 7   | 16-bit-only | None                  | Sound Cards (high DMA)<br>network cards  |

### 3. Discuss about DMA Operations.

Ans: Direct Memory Access involves transfer of data between I/O devices and memory by an external circuitry system

called DMA Controller without involving the microprocessor. Microprocessor itself initiates the DMA Control process by providing starting address, size

of data block and direction of data flow. DMA contains a control unit to deals with the control

3. DMA Operations: functions during DMA operations such as, read, write and interrupt. The address register of DMA Controller is used to the Data block. The count register counts and hold number of data transfers.

5 Steps for DMA Operation:

- (a) DMA Controller first make a Bus request (BR) by sending a control signal HOLD to the Control Line.
- (b) On receiving the BR through HOLD Pin high, the microprocessor completes the current instructions execution and afterward it generates HLDA control signal and sends it to the DMA Controller. This event switches over the control from microprocessor to DMA controller.
- (c) As soon as, DMA Controller receive HLDA through BG Line, it takes the control of system bus and start transferring the

not 8 data blocks between memory and I/O devices. Without involving the microprocessor

(d) On completion of data transfer, the DMA controller sends a low signal to HOLD pin and microprocessor makes the HLDA pin Low and takes the control over system bus.

5. Draw and Discuss the internal architecture of a DMA (Working principles)

Ans: Internal Architecture of a DMA:-

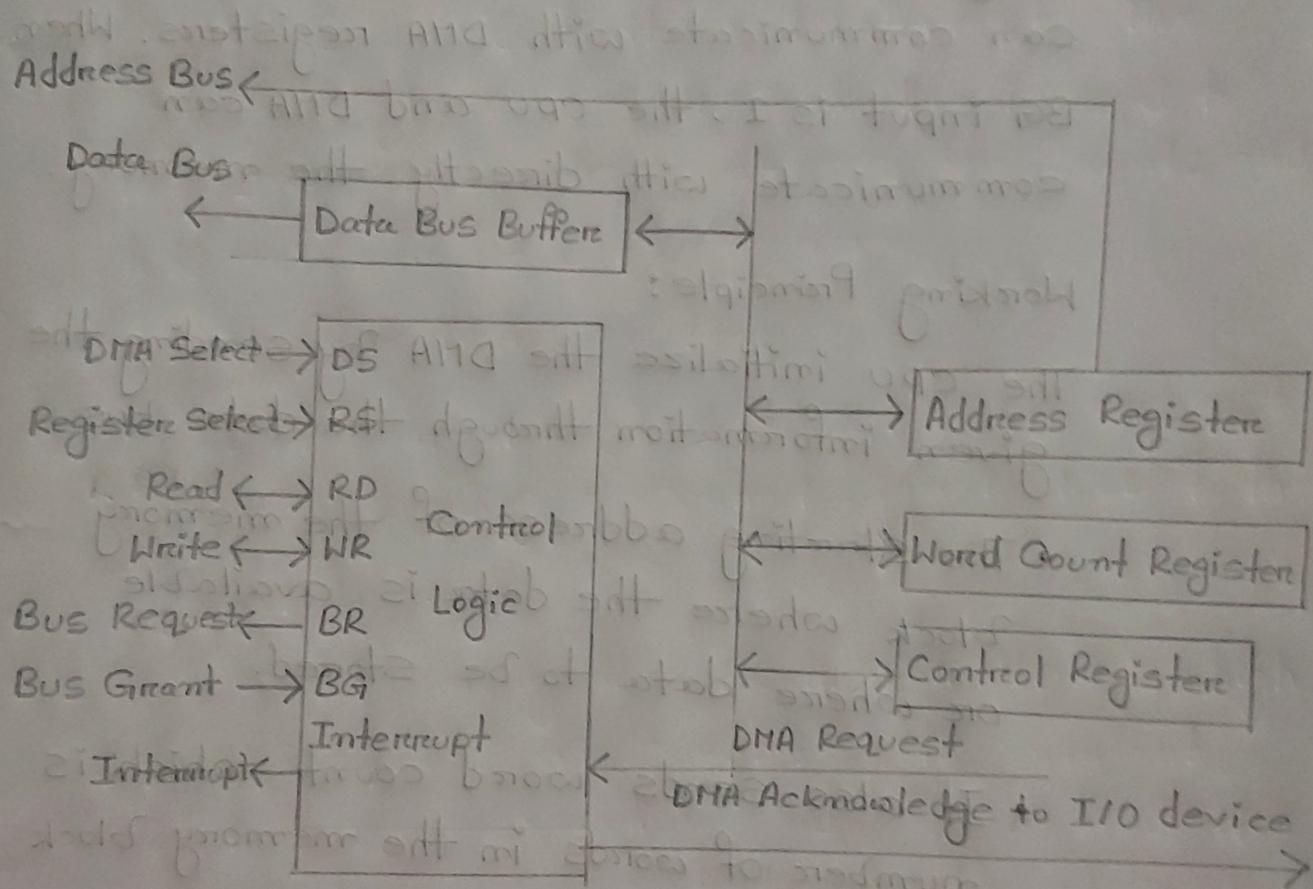


Fig 1 : Block Diagram

Fig-1 below shows the block diagram of the DMA controller. The unit communicates with the CPU through data bus and control lines.

Through the use of the address bus and allowing the DMA and RS register to select

inputs, the register within the DMA is chosen by the CPU. RD and WR are two -way inputs. When BG input is 0, the CPU can communicate with DMA registers. When BG input is 1, the CPU and DMA can communicate with directly the memory.

#### Working Principle:

The CPU initialize the DMA by sending the given information through the data bus.

- The starting address of the memory block where the data is available or where data to be stored.
- It also sends word count which is number of words in the memory block to be read or write.
- Control to define the mode of transfer such as read or write.
- A control to begin the DMA transfer.

□ Physical Address: An address in main memory. The programs sharing the memory change dynamically while the programs are running. Because of this dynamic interaction, compile each program into its own address space - a separate range of memory locations accessible only to this program. Virtual memory implements the translation of a program's address space to Physical Address.

□ Protection: A set of mechanisms for ensuring that multiple processes sharing the processor, memory or I/O devices control interfaces, intentionally or unintentionally, with one another by reading or writing each other's data. These mechanisms also isolate the operating system from a user process. This process enforces protection of a program's address space from other programs.

□ Page fault: A virtual memory block is called a page, and a virtual memory miss is called a page Page fault. An event that occurs when an accessed page is not

present in main memory.

◻ Virtual Address: With virtual memory,

the processor produces a virtual address which is translated by a combination of hardware and software to a physical address, which in turn can be used to access main memory.

◻ Address Translation: The process by which

a virtual address is mapped to an address used to access main memory. This process is called Address Translation.

Address Translation is also called

Address Mapping.

◻ Logical Address: An IP address is a

logical address that is assigned by a

software residing in the router or

server, and that logical address can

change from time to time.