Operating System Fundamentals
Prof. Santanu Chattopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 57
File System and Secondary Storage

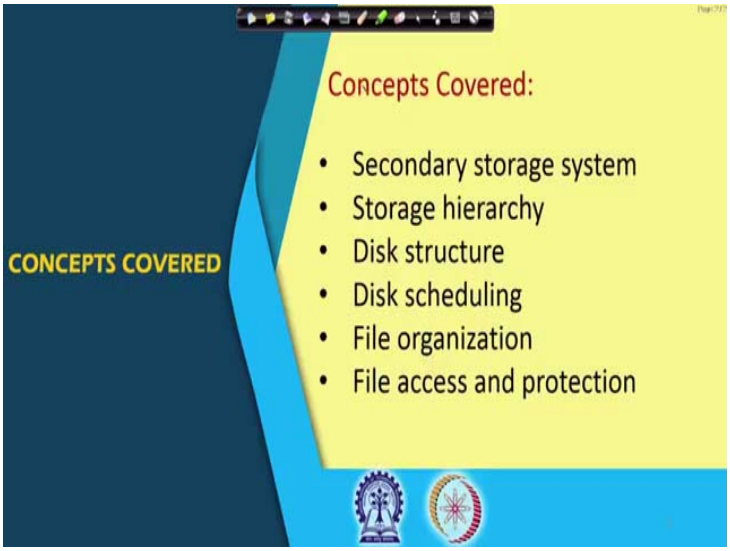
In the next part of our course, we will be discussing on File System and Mass Storage. As you know that main memory capacity is limited, so many of the programs while they are executing, so we have to use virtual memory and the disk becomes a part of it. Apart from that we need to store lot of other information like this program code, their compiled versions data bases and all, so many things and they need to be stored in the computer system for the application to run properly.

So, those information how are you going to store? And as we know that secondary storage being a magnetic or optical device, so they are much slower compared to a semiconductor memory. So, there is a huge speed gap at least and by an order of magnitude when we go from main memory to the secondary memory, but capacity wise secondary memory has got very large capacity compared to main memory. So, capacity wise it is improving, but the access time is pretty high.

So, we need to organise this secondary memory in such a fashion that this time gap that we have between main memory and secondary storage, so that reduces to some extent and also the files that that we want to search in the secondary storage, so that becomes that search becomes efficient. So, we can very easily locate a particular file.

Otherwise consider that we have got a huge secondary storage and files are all scattered like in our day to day life also we keep information organised in a proper fashion, so that when we need a particular class of information, we go to a particular almirah or maybe particular cupboard to find out the corresponding documents. So, have they been scattered over all the places you can imagine even if the information is there in the page containing the information is theree, so it is very difficult to locate that piece of information. So, this is so what is going to happen in this part of the course, so we will be looking into this file system and this mass storage organisation.

(Refer Slide Time: 02:33)



The concepts that we will be covering are like this. First we will look into the secondary storage system, then storage hierarchy, then the structure of disk. So, as far as the secondary storage is concerned some mostly we will be talking about disk storage because that is now more or less the standard one that this computer systems have. Of course, there are other forms of secondary storage like tape and say USBs and also there, but we will be primarily talking about the disk structure. Now just like a process while executing a process can generate a number of memory references.

Similarly, a process can ask for a number of accesses to the secondary storage. Now if you do not order those accesses, then there the access pattern can be random maybe we are looking for information located at location one. After that we are looking for location 50, then again coming back to 30, then going to 200. So, like that if it is random like this then secondary storage access will become slow as the read write head as we will see after a few slides. So, that needs to move to the appropriate portion of the disk.

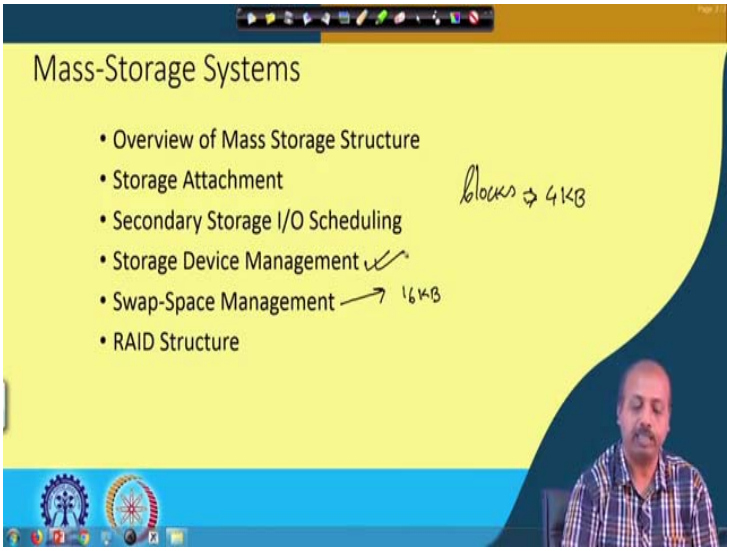
So, that way it takes time. So, scheduling becomes an important concept important issue like how do you order the accesses, how in which order you service the access requests. Then file organisation how the file are going to be organised in the secondary storage and then this file access and protection. So, access basically how do you access it like whether it is by means of some directory or it is by means of some index files like that.

So, there are several organisations. So, the access pattern will be determined by that and also who is allowed to access a particular file, who is allowed to protect, who is allowed

to modify it. So, like that the protection has to be there now um. So, if you have got a multi user system, so in the disk the files of multiple users will reside simultaneously. So, at that time it is desirable that only the genuine users or genuine owner should have the appropriate permission while others if it is intended that others will not be able to modify that should be enforced by the operating system.

So, as an owner of the file, I can tell who else can access this particular file and it is operating systems responsibility to ensure that that protection is enabled that way.

(Refer Slide Time: 05:12)



So, Mass storage Ssystems: so, we will be looking into overview of Mass Storage structure, then storage attachments, then secondary storage IO scheduling storage, device management swap space management and RAID structure. So, we will see this RAID slightly later.

So, this is a redundant array of some disk structure. So, you see that there are two specific type of management. If you see one is Storage Device Management; another is Swap Space Management.

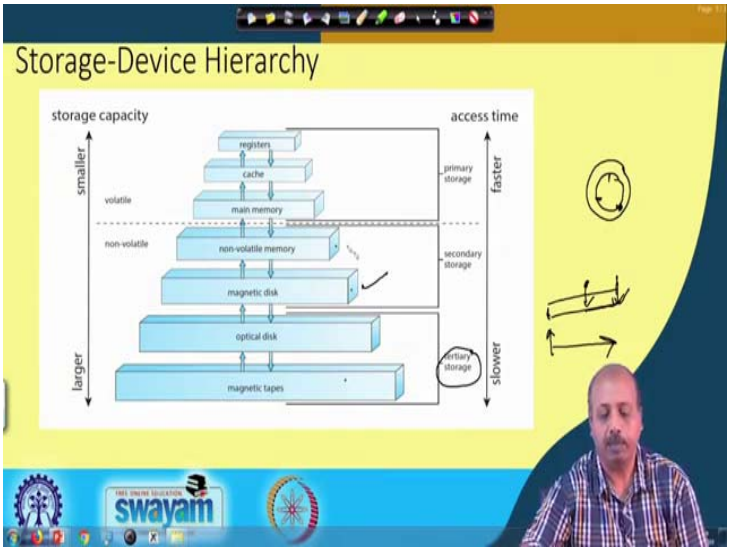
As we discussed previously while talking on virtual memory, so we have seen that is swap space is used by for the virtual memory concept and when it is used then the access pattern is different compared to the storage device. So, the main differences that we have like we know that the secondary storage it is accessed in terms of blocks ok. So, one

block may be equal to say 4 k, or so, 4 kilo byte of data. So, in one access you can get 4 kilo byte of data. So, but for the swap space management what happens is that the block size may be larger.

So, block size in that case may be 16 kilo byte. So, in one access you get 16 kilo byte of data because swap space access is very frequent when there is a page fault. So, this so that this page faults are served quickly; this block access the block size is increased. So, we have got per access we get more amount of data. So, number of page faults will reduce. At the same time as per as this searching is concerned so, searching is not that efficient in swap space management, but searching has to be very efficient in this storage device management.

So, the device management part searching has to be efficient because somebody may be may be looking for a particular file where as for swap space that is not the case. So, we know which memory frame for that where the information is kept. So, that is searching for only the memory frames, not arbitrarily in the disk space, ok. So, that way this management is different from the storage device to swap space. So, we will come to that.

(Refer Slide Time: 07:31)



Now, coming to this discussion so, the storage hhierarchy. So, you have got this if we go from this bottom side towards the top now, you will see that this at the lowest level we have got magnetic tapes. So, now a days of course this magnetic tapes, we can only find

in big computer centres and or things like that where they keep backup over months in the in some magnetic tapes, but it is it is becoming more and more obsolete.

But anyway the idea is that this magnetic tape; this is the volume is very high, but the access pattern is sequential. So, you can so since this is a tape, so you can start reading from one end of the tape and go towards the other end. So, if you are starting reading from here, so you can go only in this direction. You cannot randomly access say these blocks. So, first of all you have to access this, and then you have to come to this one and then access this. So, it is very difficult to come back to some other location. So, that way magnetic tape size is huge, but access time is the slowest.

On the other hand, next higher level of hierarchy you have got optical disks. So, there access time is better than magnetic tape, but the storage capacity is also less than the magnetic tape. On top of that we have got this magnetic disk. So, here we can we have got some amount of randomness. So, you can you can so disk as you can understand the disk will be structurally something like this onto which we have got tracks and on that tracks we have got these sectors.

So, you can ask for data from a particular sector since the disk is continually rotating. So, you can come to the previous block also by one rotation of the disk which is not possible for the magnetic tape type of device where the tape has to rewind and this rewinding means changing the direction of the tape movement. So, that takes lot of time. So, normally we do not do that in case of magnetic tape whereas, disk this is since it is a rotating device. So, you can come to the particular portion of information particular sector of information again and again during the rotation. So, that way it is more random compared to this magnetic tape.

On top of that we have got this non-volatile memories basically the flash devices and all. Again flash devices they are much faster compared to disk, but the capacity wise again it is less now on top of this flash. So, up to this much is the secondary storage and is optical disk and magnetic tape, they are called Tertiary Storage because most of the time after taking the backup from the system we move the tapes, we move we take out the tape or disk optical disk and keep it on some storage area.

So, this is not a no more part of the computer system. That is why they are called Tertiary Storage, then in the secondary storage. So, they remain part of the computer

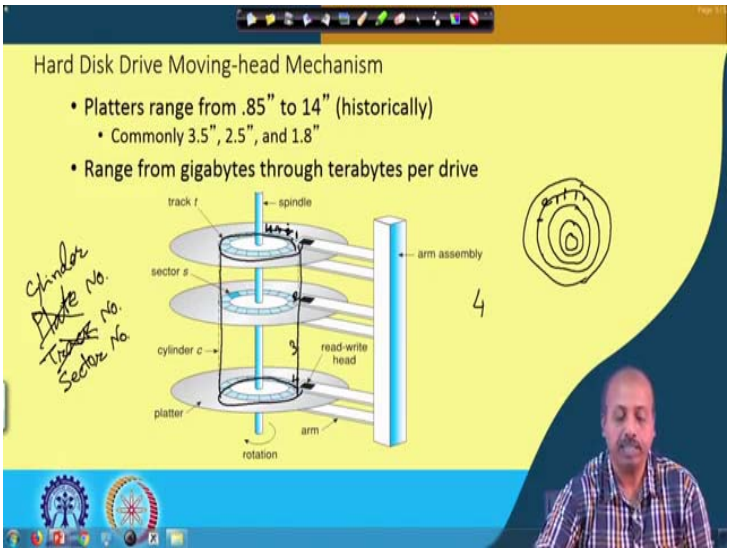
system, the disk that we have in our computers. So, they remain part of the computers even if the access the copy has been made. So, that way it remains with the system. So, that is why they are called secondary and on top of that we have got primary. So, primary means they interact directly with the CPU. So, CPU can ask for information from them directly. So, this memory cache and CPU registers, so these are the three other sources of memory which are basically primary storage.

Because CPU is directly talking to them on the other hand the secondary storage or tertiary storage. So, CPU is not in direct interaction of with them. CPU if it wants to access if it finds that due to due to page fault or due to some access request by a user program, it needs to access the magnetic disk. So, it will tell the disk scheduler to do that, ok. So, that way the and disk scheduler normally takes help of some other bus masters like this direct memory access and that type of policies to come to that particular block and get the information. So, CPU is not directly involved there.

On the other hand this optic, this tertiary storage, so this is accessed this is the the access to this is often determined by the system administrator maybe once in a week or so. So, they will take the backup of the entire system and put it on to that magnetic tape or the tertiary storage.

So, that way it becomes totally controlled by the administrator and the corresponding time schedule. So, this is the storage hierarchy. So, as you move higher up in the hierarchy, access time will reduce, but the storage capacity will also reduce and the thing that is not shown here is the cost to cost per bit also increases as you are moving up in this hierarchy and if you are moving down in the hierarchy, then the devices will become slower, but the capacity wise it is large and per bit cost. So, that will come down. So, all these types of storage are required in a computer system for different purposes.

(Refer Slide Time: 12:46)



So, next we will be concentrating on this Hard Disk Drive Moving Head Mechanism because this is one that we will be mostly talking about. This is because this is direct in direct interaction with the processor and this operating system really has to do something in optimising their access. So, rest of the thing that the tertiary storage access. So, operating system does not have much role to play as I have said that it is determined by the system administrator when to take the backup now in case of hard disk.

So, we have got platters range from 0.85 inch to 14 inch. So, this historically, so it was very large 14 inch diameter and then we have got the so commonly used as 3.5 inch, 2.5 inch and 1.8 inch. So, these are the diameter. So, range is from gigabytes through terabytes per drive. So, this is so we have got a number of platters arranged like this. So, every plate every plate is divided into number of tracks, ok.

So, if you have got a plate like this if you have got a plate like this, sorry this circle is really bad. So, if you have got a plate like this then it is you can think about as if there are some concentric circles here and each of them is a track. So, information is stored in the track and a track is further divided into sectors. So, the track is divided into sector. So, this maybe one sector from here to here is off. So, this is once this is sector 1, this is sector 2. So, like that we have got sectors.

So, this is the this is a sector. So, with the, so if this is the track this is the plate. So, it is divided into number of tracks and this track is further divided into sectors. Now we have got this read write head. So, for each plate there is a there is at least one read write head

sometime. So, information if it is if the information is stored on both side of the disk, then we will have two such read write heads per plate now this. So, there is a spindle such that this entire plate assembly. So, it can rotate and as it rotates this disk read write head will come to each of the sectors on a track and this read write head, so they can move backward or forward, so that it can come to a particular track.

So, to address a particular sector what you need to specify is a first of all the plate number the plate number, then you need to specify the track number and then the sector number. So, these are the three things that we need to mention when we are talking about a particular sector, and data transfer is always in terms of one sector and sector size is often made equal to the block size or the frame size in case of demand paging environment or pager environment. So, the memory frames side so they are made equal. So, that one block transfer that will be equal to one sector transfer and that is equal to one page or one frame in the main memory.

So, if you mention the plate number, so you know the which plane which plate we are talking about, then the track number. Accordingly the corresponding read write head may move backward or forward and come to the particular track and after that there is a sector number. So, this sector number will take it to the the disk will rotate and the sector, the the sector will be aligned with the read write head and from that on wards. The read write operation will start now if you if you look into the operation like.

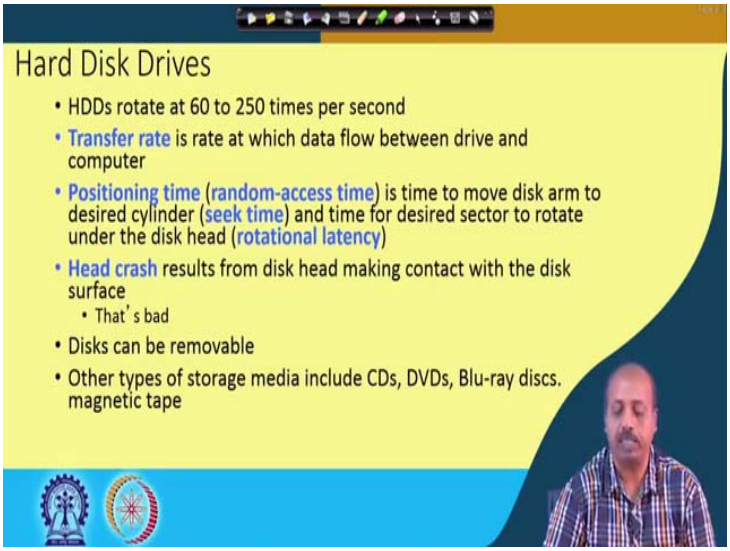
Now, you see that suppose I have to I have I have to get information about say four sectors. So, one possibility is that I put all these four sectors on the same track like this is say 1, this is 2, this is 3 and this is 4. So, what happens is that this read write head it comes, it gets aligned with sector 1 and when sector 1 reading is over, then it goes to sector 2, then it goes to sector 3 like that. Another possibility that we have is we can say that number 1 is there, 2 3 4 instead of being here.

So, 2 is on the second plate 3 is on the third plate, 4 is on the fourth plate like that, so that the read write head alignment. So, that is done in one shot and then it can start reading from all these blocks simultaneously. So, that is the thing. So, if you look into a particular track of each sector for of each of each plate like say this particular track track T of each plate, then that forms a cylinder. So, this makes a cylinder ok. So, this makes the structure of a cylinder. So, instead of talking about plate number we will be talking in

terms of cylinder number knowing well that this cylinder is distributed over number of tracks and the sector numbering will be done like this. So, for a particular cylinder the sectors will be numbered as one particular cylinder, the sectors will be named as 1 2 3 etcetera. So, this track number is also not necessary. So, we can tell the cylinder number and accordingly we know on which cylinder we are talking about and the corresponding track number. So, this cylinder number and track number sorry sector number it can identify the appropriate sector that we are talking about.

So, this way this hard disk drive structure enables us to have this information stored per sector and access is in terms of sectors. So, even if you need to access a single byte within a sector, we have to read the entire sector for that. So, that is the structure of this secondary storage.

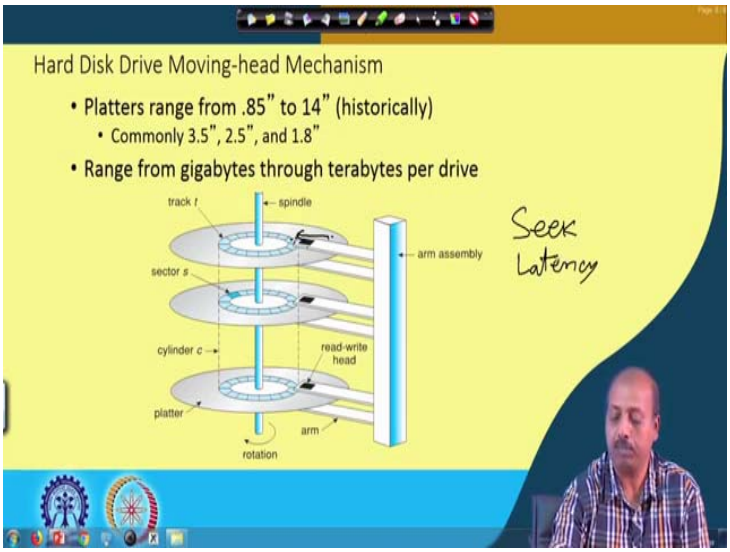
(Refer Slide Time: 19:16)



So, this hard disk drives they rotate at 60 to 250 times per second and transfer rate is the rate at which data flows between drive and computer. So, that is the transfer rate the rate at which the data can move between the computer and the drive.

Positioning time or random access time is time to move disk arm to the desired cylinder. So, this is also called seek time and time to time for desired sector to rotate under the disk head. So, that is called Rotationa Latency. So, you see that if you are so if suppose my hard my disk head is somewhere here. So, suppose my disk head is somewhere here.

(Refer Slide Time: 20:09)



And then I need to access this particular cylinder, then this disk head has to move by this much amount and that is done by a mechanical motor and an electrical motor.

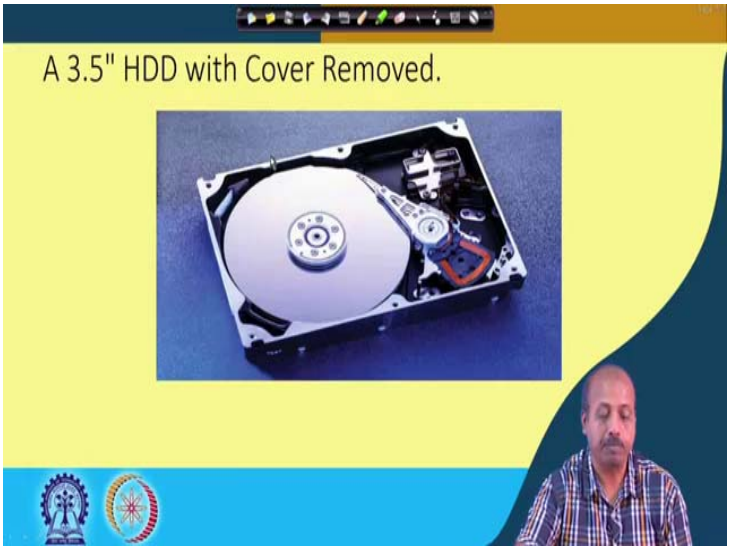
So, that electrical pulses are given and as then this head moves and this head moves one by one track by track. So, this takes some time for the disk head to come to a particular track or come to a particular cylinder. So, that way, so that that time is called the seek time. So, time required by the read write head to reach a particular cylinder. So, that is called the seek time. After that once it has reached the particular cylinder, now it has to go to a particular track particular sector. So, the going to the particular sector, that means this disk has to be rotated and then that rotation will take the head to the particular sector.

So, that time is called the rotational latency ok. So, you have got two times here seek time and latency time. So, that is so what is explained the next slide. So, when we are talking about this seek and rotational latency, so seek time is the time to move the disk arm to the desired cylinder and the rotational latency is the time for the desired sector to rotate under the disk head. So, disk head does not rotate. So, disk head just advances by cylinders or comes back by cylinders and then this disk the disk moves disk rotates and this it gets aligned to a particular read write head, the sector gets aligned. So, head crash results from disk head making contact with the disk surface. So, it is so this disk does not disk head does not make a contact with the disk surface because the; if you do that then this rotational speed it will be limited. So, to take care of that, so there is a small gap

between the two and then by some magnetic induction. So, it gets the corresponding bits stored.

So, this is, so the head crash should not occur. Disks can be removable. So, you should be able to remove the disk from the system like a if you have got say a optical disk. So, as you know in computer systems, so we can remove the disk, but the head remains there, the disk drive remains there. From the drive I should be able to remove the disk, other types of storage media like CDs DVD Blu-ray discs magnetic tapes. So, these are also there, but their principle of operation is slightly different. So, we do not have much time to go into that. So, if you are interested, we can look into the corresponding literature for that purpose.

(Refer Slide Time: 22:49)



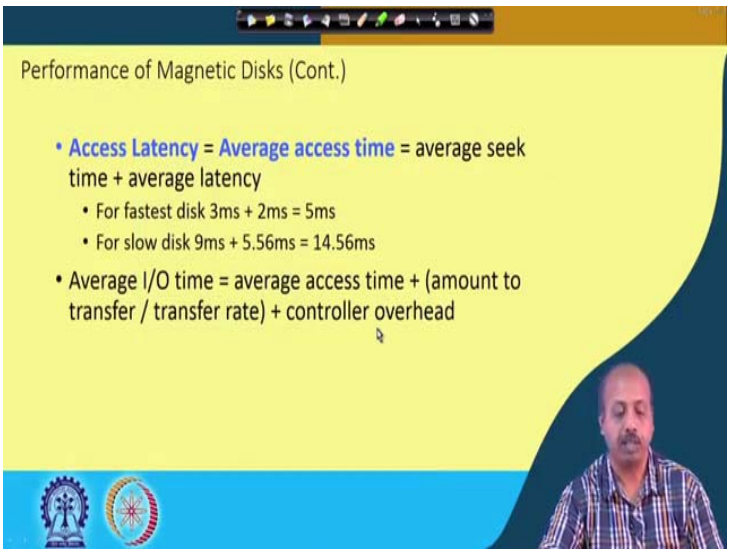
So, this is how a disk drive looks like. So, this is your disk ok. So, and this is the read write assembly, the head assembly. So, as you are giving pulses to this read write head, so it can go move from one cylinder to the next cylinder. So, it can move like this and this head this disk can move this and as a result the corresponding sector will come below the read write head from where it will be able to start doing the read operation.

(Refer Slide Time: 23:19)

So, this is the first commercial disk drive. So, that was in 1956 IBM computer, RAMDAC computer is included the IBM model 350 disks storage 5 million characters. Each character is 7 bit.

So, 50 into 24 inch platters. So, that is a huge amount in fact and access time is less the less or equal one second. So, that was the scenario so, the big system. So, this is the disk drive that we have. So, it is a big system, but now the size has diminished. So, it has become a very very small and we can have it onto the computers attached very easily.

(Refer Slide Time: 24:02)

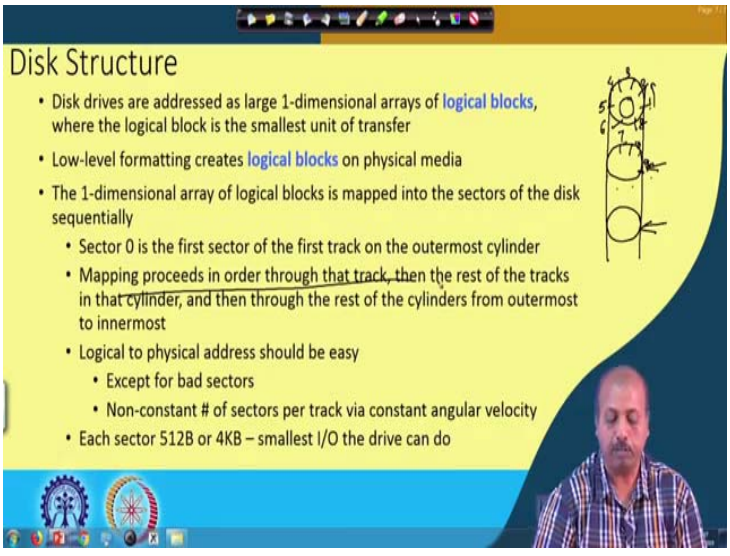


Now, if you are trying to judge the performance of magnetic disk, then this access latency is that is also sometimes called average access time is the average seek time plus average latency time for fastest disk. So, average seek time is about 3 millisecond and latency time is 2 millisecond.

So, this is the about 5 millisecond if you looking for slow disk that you have today. So, this is 9 millisecond of seek time and 5.56 millisecond of average latency. So, that makes a 14.56 millisecond access latency. So, average I/O time is average access time plus amount to transfer or transfer amount to transfer divided by transfer rate plus the controller over head like if you want to do a some actual I/O operation the disk drive has to be informed by the operating system that I want to do this transfer. So, accordingly that controller will have some have its own delay. So, that is the controller delay.

Then these average access time. So, that is seek time plus latency time. So, those values will be coming plus for actual data transfer. So, after the read write head has been aligned with the particular sector, the data transfer will take place at a certain rate. So, that rate if I divide by that quantity that amount of data that we need to transfer. So, that will give me the transfer time. So, that is the average I/O time that we have.

(Refer Slide Time: 25:31)



Coming to the structure of disk : so, disk drives are addressed as large one dimensional array of logical blocks as I said that unlike primary storage. secondary storage is not accessed in terms of bytes or words. So, they are accessed it in terms of blocks and one

block may be 1 kilobyte 2 kilobyte 4 kilobyte like that. So, it can be the block size can be determined by the system designer. So, that is so disk drives are addressed as one large block, a one dimensional array of logical blocks and low level formatting creates logical blocks on physical media. So, when you are doing may be aware that when you when you have got a new disk drive or new secondary storage, it needs to be formatted.

So, formatting actually marks these logical blocks where the logical blocks will be located. The one dimensional array of logical blocks is mapped onto the sectors of the disk sequentially. So, as I was telling, so looking into a particular cylinder so, these blocks are marked sequentially one by one on each cylinder. So, sector zero is the first sector of the first track on the outermost cylinder. So, sector zero is the outermost one outermost track and the first sector there and mapping proceeds in order through that track and then rest of the tracks in that cylinder and then through the rest of the cylinders from the outermost to innermost. So, like that so the way the ordering is done.

So, it is like this suppose so this is one track this is. So, if you consider one particular cylinder. So, this is one cylinder and each disk it has got a track for that now this um. So, this is this track is now divided into number offs sectors, this track is divided into number of sectors like this. So, it is say sector 1 2 3 4 5 6 7 8, then it will start from here.

So, now this will be 9, this is 10, this is 11. So, it will go like this. Why? Because once the disk read write head has been placed here, so you see that it can start reading the block sequentially. So, so once this one has been read your read write head is already here. So, from there it can start reading the second block, from there it can start reading the third block. So, it for one full track of data the read write can proceed without requiring any movement of the read write head across the across tracks.

However, once this is done now if the next sector was assigned on this disk itself, then you need to move the read write head again. So, that will again take the seek time. So, instead of that so this next sector is assigned or the next logical block is assigned or the first sector of the next plate, ok. So, that way so, the 9 is here. So, the so when the first when for the first disk read write head got aligned to 1 for this, here also this read write head got aligned to 9.

And this did not move. So, this is remaining here. So, once up to 8 you have read. So, you can immediately start reading from sector 9 onwards so, the block 9 onward. So, it

can go like this. Now once this is done, you can come to the next track on the next plate and on the same cylinder on the next plate and you can start reading from there. So, this is what is told here that mapping proceeds in order through that track and then rest of the tracks on that in that cylinder and then through the rest of the cylinders from outermost to innermost. So, logical to physical address should be easy. So, there are there are some bad sectors. So, while doing the initial formatting, these bad sectors are avoided and there is a notice kept that these are bad sectors.

No data will be written there non constant number of sectors per track via constant angular velocity. So, if we so naturally the number of sectors per track will vary because for this inner tracks the track size is small and for outer track the track size is size is larger, ok. So, as a result the number of sectors will vary across the tracks. Each sector is 512 byte to 4 kilobyte. So, that can vary and the smallest I/O that the drive can do, so it is in terms of block. So, so whatever be the block size determined so at that rate only it can do the transfer. So, we will continue with this discussion in the next class on this disk drive.