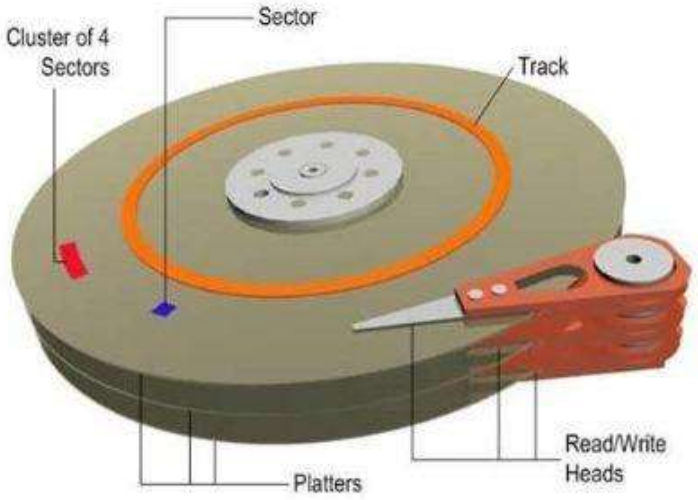
Mass-Storage Systems

S.Lakshmi Priya, AP/CSE

HDD



Sectors, Clusters and Volume.

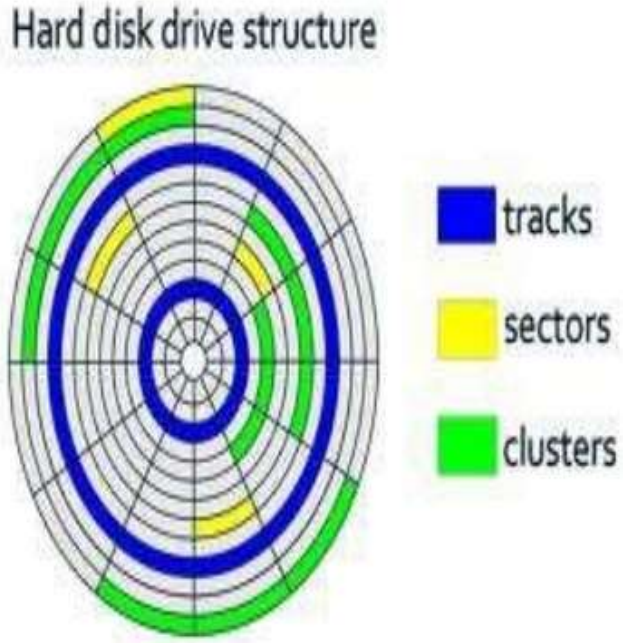
Sector: The smallest physical storage unit on the disk.

•In windows, size of each sector is 512(bytes).

•**Cluster:** Group of one or more contiguous sectors is called a cluster.

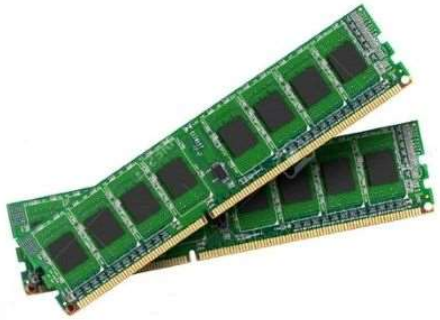
•**Volume:** logical partitioning of a disk, consisting of one or more clusters.

•Volume consist of files, system information and unallocated space, that can be allocated to files

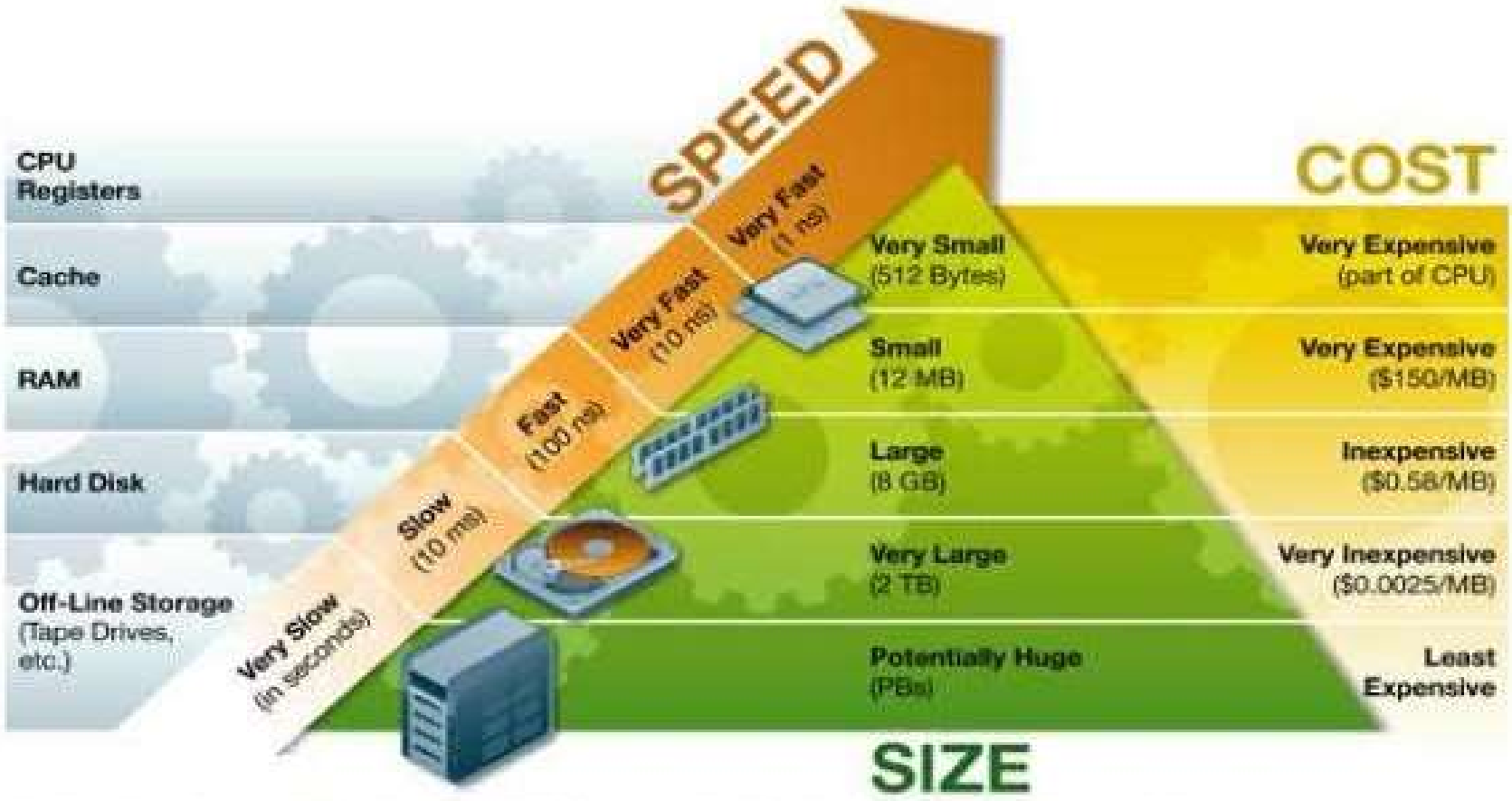


RAM vs HDD

- Components
- R/W Time
- Memory access
- Size
- Cost
- Volatility



Extended Memory Hierarchy



Source: http://www.ts.avnet.com/uk/products_and_solutions/storage/hierarchy.html

HDD vs SSD

Parameter	HDD	SSD
Full Form	HDD stands for Hard Disk Drive.	SSD stands for Solid State Drive.
Components	contains moving mechanical parts, like the arm.	No mechanical parts, only electronical parts like ICs.
R/W Time	longer R/W time.	shorter R/W time.
Size & Weight	larger in size & heavier in weight	more compact in size. & lighter in weight.
Reliability	less reliable due to possibility of mechanical failure, like head crash and susceptibility to strong magnets.	more reliable.
Cost	cheaper	Expensive

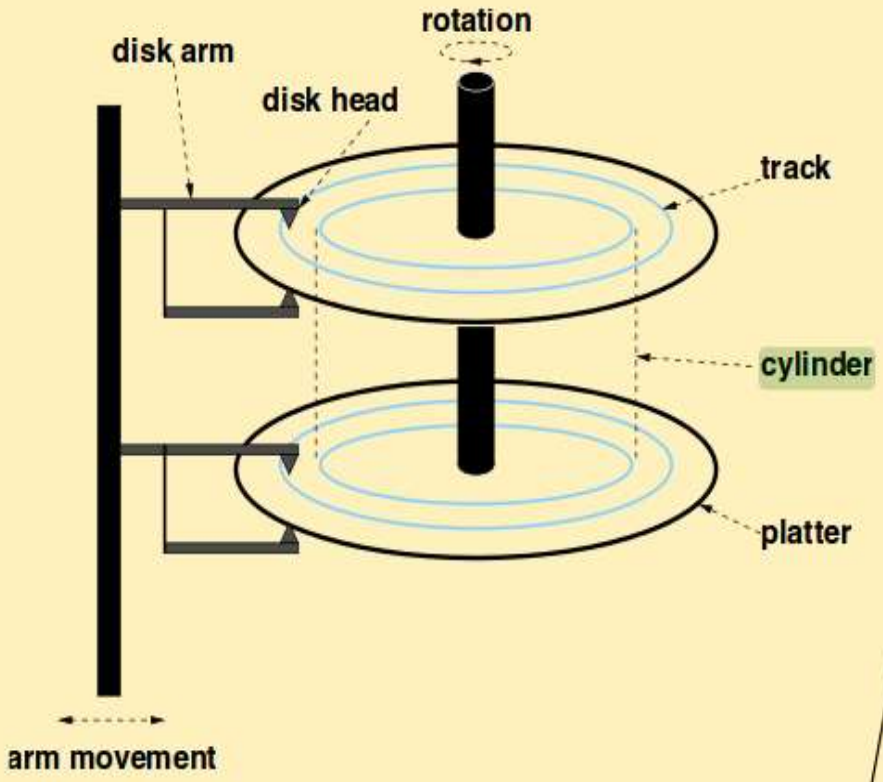
Disk Structure

- Disk drives are addressed as large 1-dimensional arrays of *logical blocks*, where the logical block is the smallest unit of transfer.

- The 1-dimensional array of logical blocks is mapped into the sectors of the disk sequentially.
  - Sector 0 is the first sector of the first track on the outermost cylinder.
  - Mapping proceeds in order through that track, then the rest of the tracks in that cylinder, and then through the rest of the cylinders from outermost to innermost.

PERMANENT MEMORY: DISK

In an electronic world, disks are a mechanical anachronism!

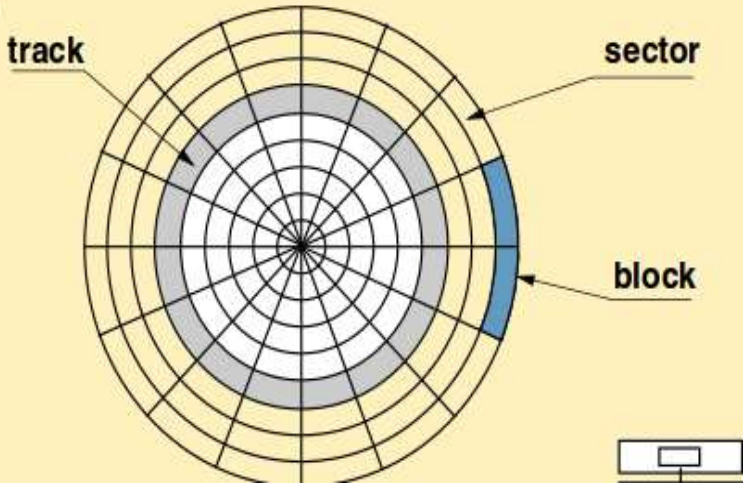


Access Time =

Seek Time (0-10 ms) +

Rotational Time (0-3 ms) +

Transfer time (.02 ms per 8K)



Disk Scheduling

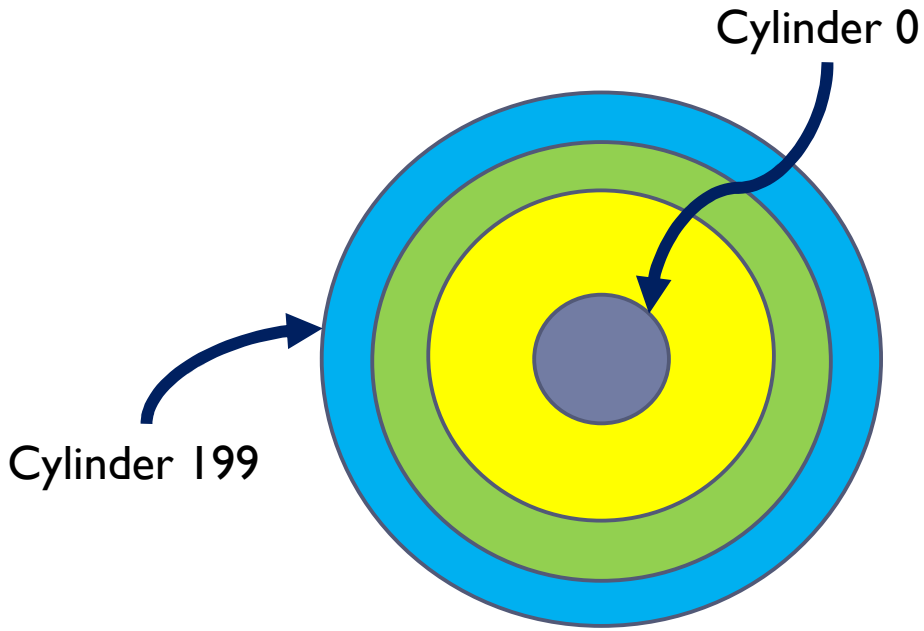
- The operating system is responsible for using hardware efficiently — for the disk drives, this means having a fast access time and disk bandwidth.
- Access time has two major components
  - *Seek time* is the time for the disk to move the heads to the cylinder containing the desired sector.
  - *Rotational latency* is the additional time waiting for the disk to rotate the desired sector to the disk head.
- Minimize seek time
- Seek time ≈ seek distance
- Disk bandwidth is the total number of bytes transferred, divided by the total time between the first request for service and the completion of the last transfer.

Disk Scheduling (Cont.)

- Several algorithms exist to schedule the servicing of disk I/O requests. We illustrate them with a request queue for a disk containing 200 cylinders (numbered from 0-199)

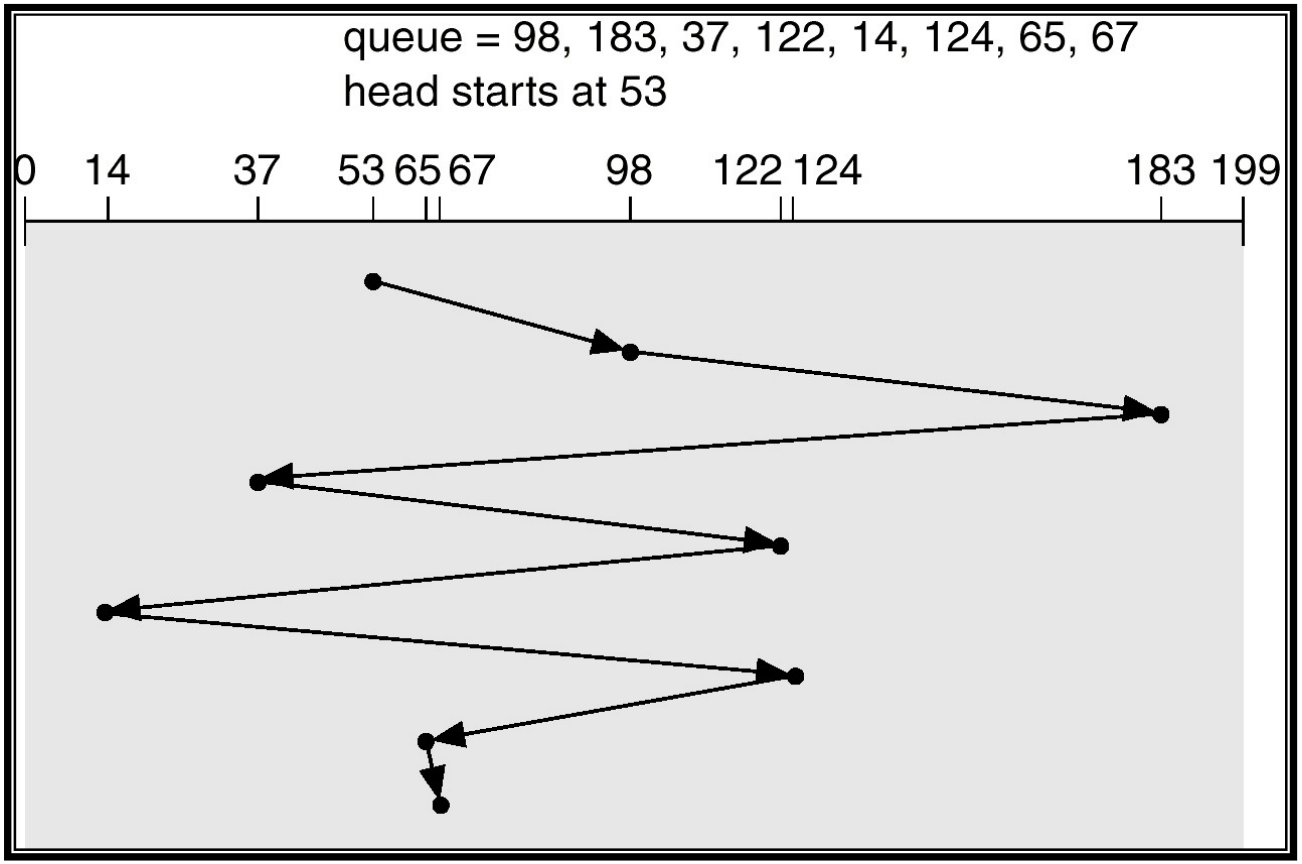
Request Queue (Track/Cylinder numbers) : 98, 183, 37, 122, 14, 124, 65, 67

Current Head pointer 53



FCFS

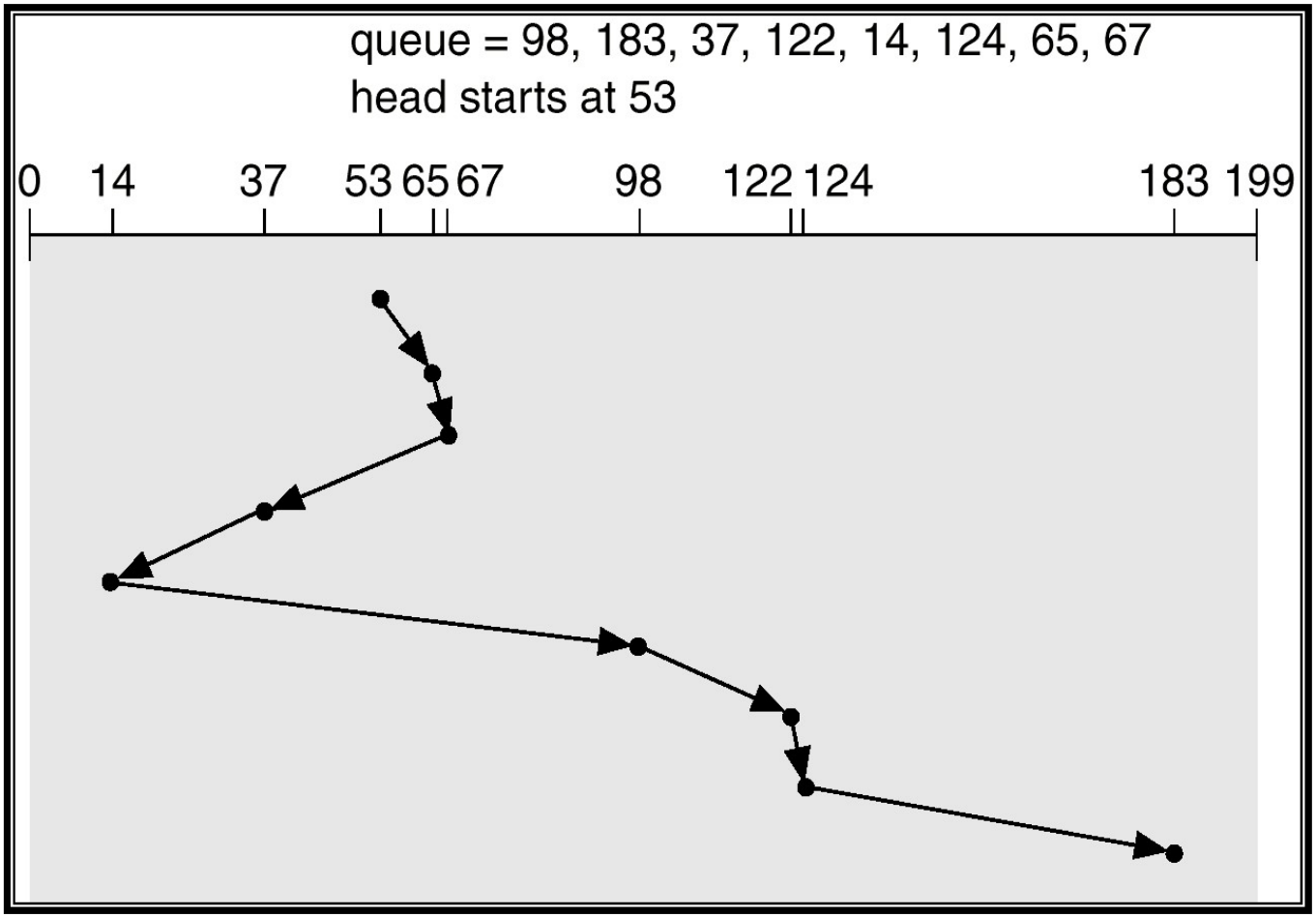
Illustration shows total head movement of 640 cylinders.



SSTF

- Selects the request with the minimum seek time from the current head position.
- SSTF scheduling is a form of SJF scheduling; may cause starvation of some requests.
- Illustration shows total head movement of 236 cylinders.

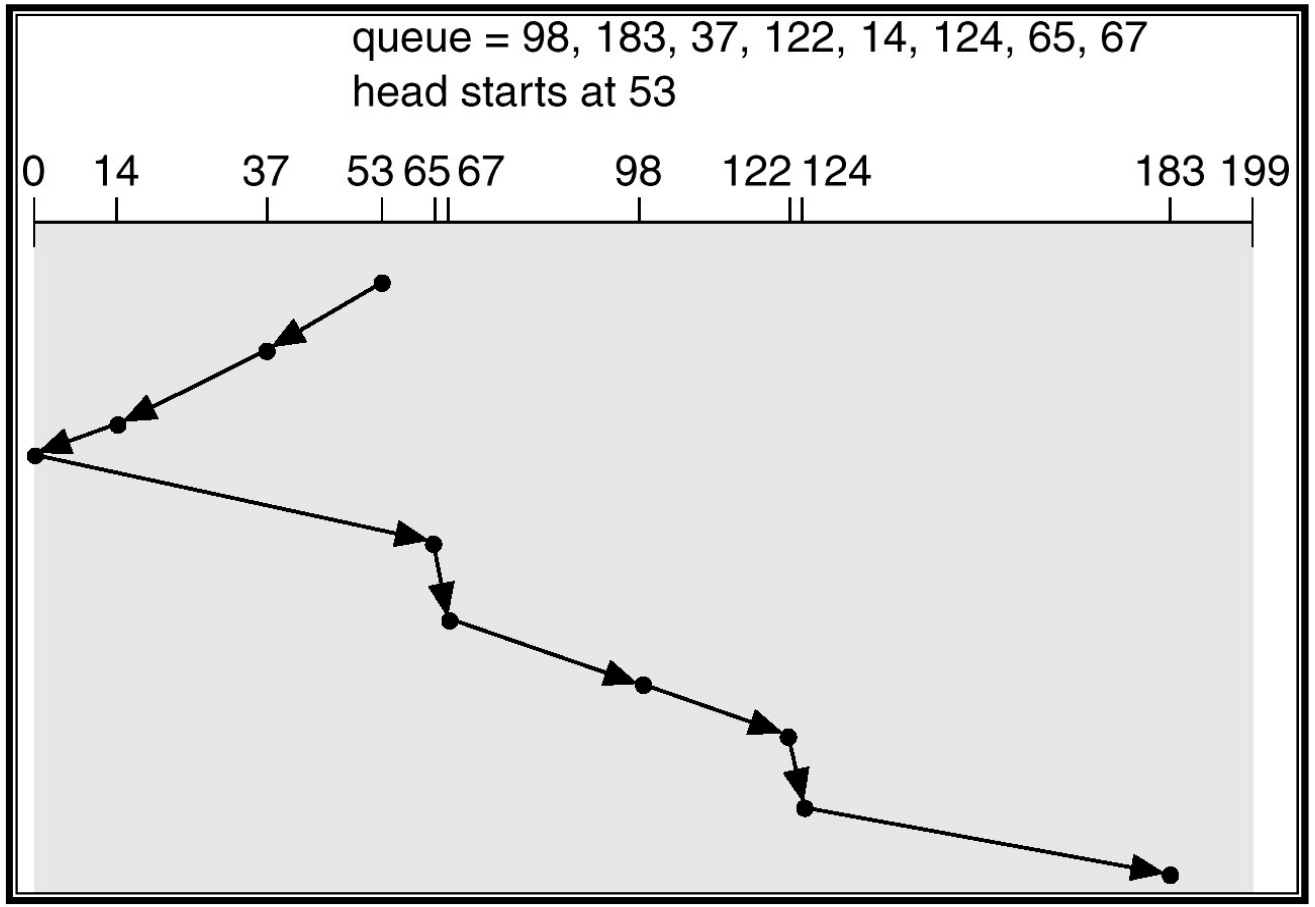
SSTF (Cont.)


queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53
0
14
37
53
65
67
98
122
124
183
199

SCAN

- The disk arm starts at one end of the disk, and moves toward the other end, servicing requests until it gets to the other end of the disk, where the head movement is reversed and servicing continues.
- Sometimes called the *elevator algorithm*.
- Illustration shows total head movement of 236 cylinders.

SCAN (Cont.)


queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53
0
14
37
53
65
67
98
122
124
183
199

SCAN (Contd.,)

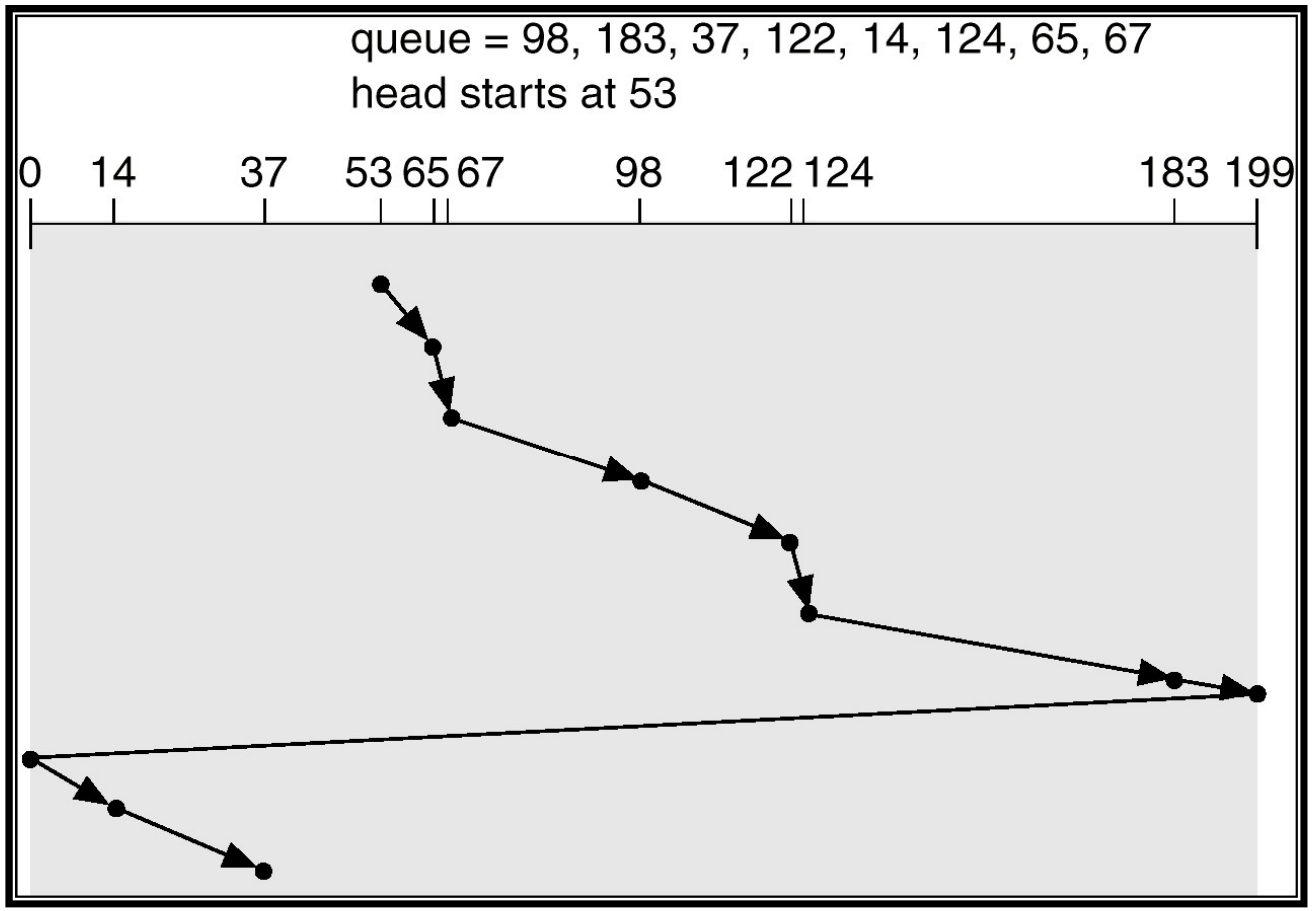
- New request just ahead of the head gets serviced immediately
- New request just behind the current head position has to wait for the arm to reverse direction
- In the eg., when cylinder 0 is reached and direction is reversed, few request lay just ahead of the head as most of them were serviced recently
- The density of requests is more near the other end and the waiting time is also more

C-SCAN

- Provides a more uniform wait time than SCAN.
- The head moves from one end of the disk to the other, servicing requests as it goes. When it reaches the other end, however, it immediately returns to the beginning of the disk, without servicing any requests on the return trip.
- Treats the cylinders as a circular list that wraps around from the last cylinder to the first one.

C-SCAN (Cont.)

Total head movement = 208

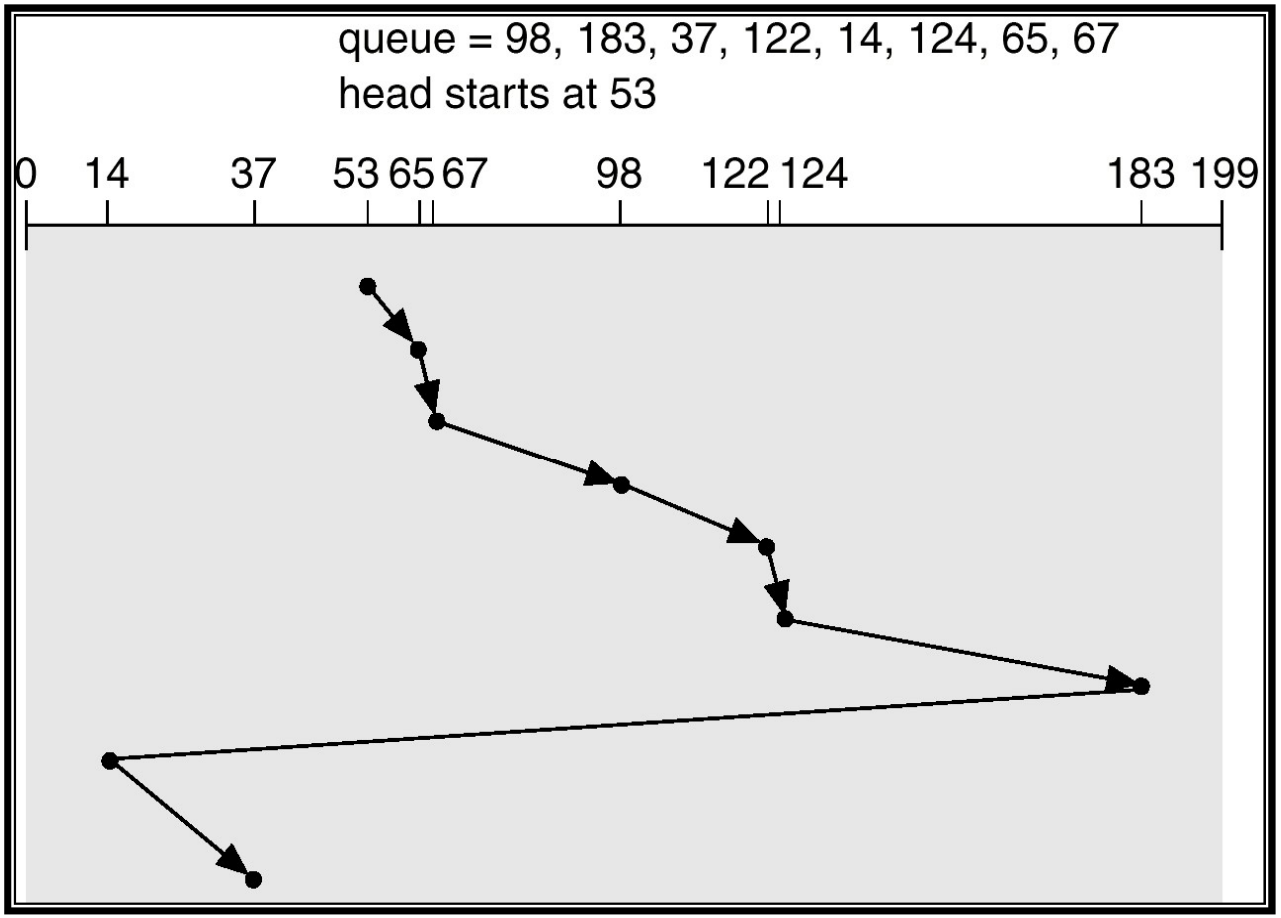


C-LOOK

- Version of C-SCAN
- Arm only goes as far as the last request in each direction, then reverses direction immediately, without first going all the way to the end of the disk.

C-LOOK (Cont.)

Total head movement = 322



Selecting a Disk-Scheduling Algorithm

- SSTF is common and has a natural appeal
- SCAN and C-SCAN perform better for systems that place a heavy load on the disk.
- Performance depends on the number and types of requests.
- Requests for disk service can be influenced by the file-allocation method.
- The disk-scheduling algorithm should be written as a separate module of the operating system, allowing it to be replaced with a different algorithm if necessary.
- Either SSTF or LOOK is a reasonable choice for the default algorithm.

Exercise – Disk Scheduling

- For the following disk accesses, compute the number of head movements for the following list of seeks to disk cylinder: 26 37 100 14 88 33 99 12
- Assume head is initially positioned over 26.
- Total cylinders : 101 (0-100)
- SCAN AND LOOK – (service in increasing order)

Disk Management

Disk Initialization, Disk formatting and booting, bad block recovery

Disk Formatting

- *Initially a disk is blank-just a platter coated with magnetic material*
- *Low-level formatting*, or *physical formatting* — Dividing a disk into sectors that the disk controller can read and write.
  - Fills the disk with special data structure for each sector
  - Usually a header, data-area and a trailer
  - Header has sector number and trailer has ECC
  - When data is read, ECC is recalculated to check for any inconsistencies. One or two byte changes are usually corrected by controller and is reported as soft error
  - Low level formatting is done by manufacturer itself
  - Sector sizes maybe 256, 512, 1024 bytes.
  - Larger the sector, fewer header and trailers and hence more space for data.

Disk formatting – Contd.,

- To use a disk to hold files, the operating system still needs to record its own data structures on the disk.
- Done in 2 steps
  1. *Partition* the disk into one or more groups of cylinders.
  2. *Logical formatting* or "making a file system" –free and allocated memory maps
- Filesystem groups blocks to form clusters

Boot Block

- Boot block initializes system.
  - The bootstrap program is stored in ROM
  - Issue: Cannot change the bootstrap even if we want to
  - *Solution: Bootstrap loader* program alone is kept in ROM and bootstrap program is kept in "Boot Block" on disk.
- A disk that has a boot partition is called a boot disk or system disk.
- Eg: Windows - one partition identified as the boot partition—contains the operating system and device drivers.
- The first sector on the hard disk - master boot record, or MBR has two things
  - Boot code
  - Partition table

Boot Block – Contd.,

- Partition table has a list of partitions in the hard disk and a flag indicating which partition the system is to be booted from

- First sector from boot partition - called the **boot sector** is read and the remainder of the boot process continues, which includes loading the various subsystems and system services.

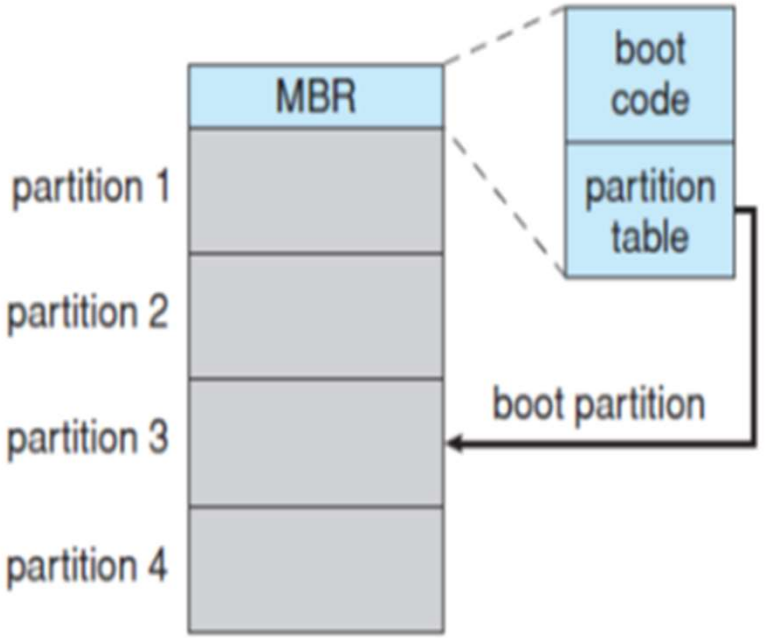
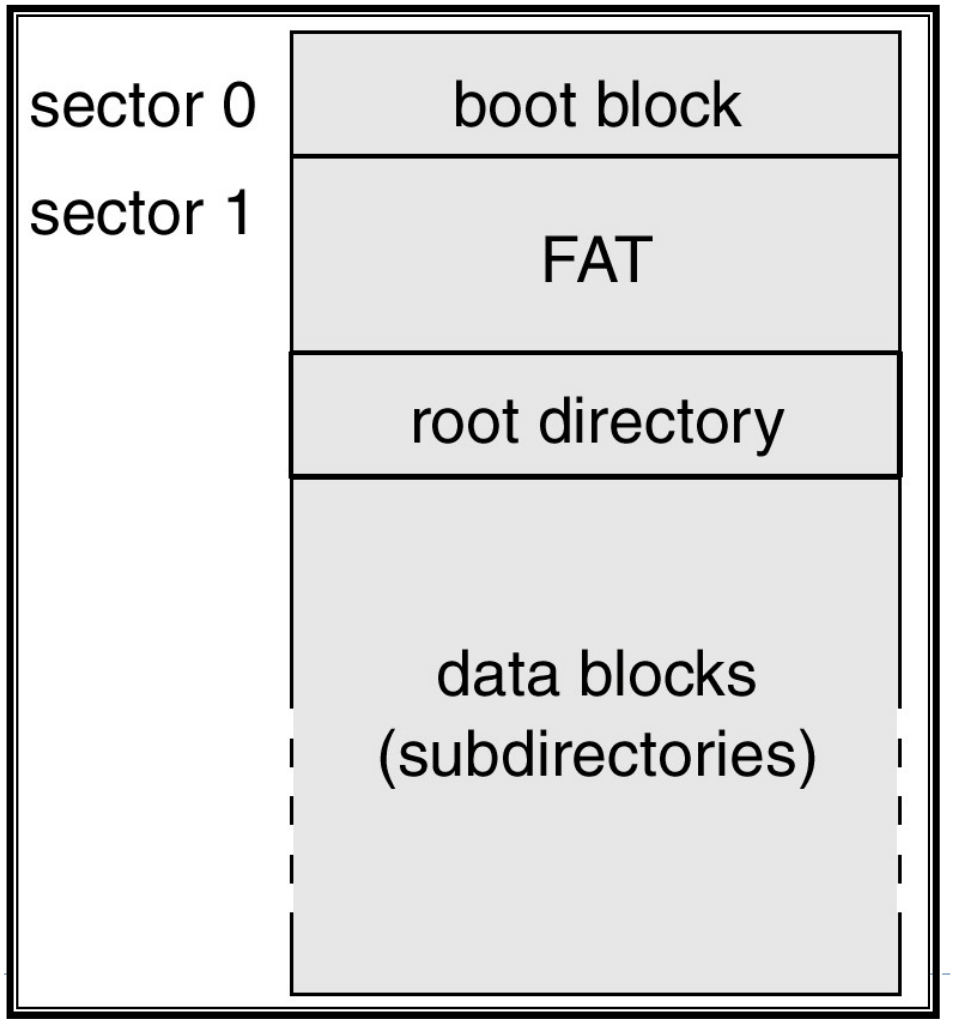


Figure 10.9 Booting from disk in Windows.

MS-DOS Disk Layout



Bad Blocks

- Bad blocks – controller not able to read/write data from that block
  - might be due to mechanical failure
- How to handle bad blocks?
  - Scan the disk, mark the bad blocks, let OS know so that it doesn't allocate these blocks in future
  - Sector sparing – replace with a spare
  - Sector slipping – find a spare and remap all
- Some sectors are placed as spare sectors during low level formatting and are invisible to OS. Controller maintains a list of bad blocks and the list of spares.

Bad Blocks – Contd.,

- Sector sparing or forwarding
  - A typical bad-sector transaction might be as follows:
    - OS tries to read logical block 87.
    - Controller calculates the ECC and finds that the sector is bad. It reports this finding to the OS
    - The next time the system is rebooted, a special command is run to tell the controller to replace the bad sector with a spare.
    - After that, whenever the system requests logical block 87, the request is translated into the replacement sector's address by the controller.

Bad Blocks – Contd.,

- Sector slipping
- Eg: Logical block 17 becomes defective and the first available spare follows sector 202.
- remaps all the sectors from 17 to 202, moving them all down one spot.
- i.e., sector 202 is copied into the spare, then sector 201 into 202, then 200 into 201, and so on, until sector 18 is copied into sector 19.
- Slipping the sectors in this way frees up the space of sector 18 so that sector 17 can be mapped to it.

Tertiary Storage Devices

- Low cost is the defining characteristic of tertiary storage.
- Generally, tertiary storage is built using *removable media*
- Common examples of removable media are floppy disks and CD-ROMs; other types are available.

Removable Disks

- Floppy disk — thin flexible disk coated with magnetic material, enclosed in a protective plastic case.
  - Most floppies hold about 1 MB; similar technology is used for removable disks that hold more than 1 GB.
  - Removable magnetic disks can be nearly as fast as hard disks, but they are at a greater risk of damage from exposure.

Removable Disks (Cont.)

- A magneto-optic disk records data on a rigid platter coated with magnetic material.
  - Laser heat is used to amplify a large, weak magnetic field to record a bit.
  - Laser light is also used to read data (Kerr effect).
  - The magneto-optic head flies much farther from the disk surface than a magnetic disk head, and the magnetic material is covered with a protective layer of plastic or glass; resistant to head crashes.
- Optical disks do not use magnetism; they employ special materials that are altered by laser light.

WORM Disks

- The data on read-write disks can be modified over and over.
- WORM ("Write Once, Read Many Times") disks can be written only once.
- Thin aluminum film sandwiched between two glass or plastic platters.
- To write a bit, the drive uses a laser light to burn a small hole through the aluminum; information can be destroyed by not altered.
- Very durable and reliable.
- *Read Only* disks, such ad CD-ROM and DVD, com from the factory with the data pre-recorded.

Tapes

- Compared to a disk, a tape is less expensive and holds more data, but random access is much slower.
- Tape is an economical medium for purposes that do not require fast random access, e.g., backup copies of disk data, holding huge volumes of data.
- Large tape installations typically use robotic tape changers that move tapes between tape drives and storage slots in a tape library.
  - stacker – library that holds a few tapes
  - silo – library that holds thousands of tapes
- A disk-resident file can be *archived* to tape for low cost storage; the computer can *stage* it back into disk storage for active use.

Tape Drives

- The basic operations for a tape drive differ from those of a disk drive.
- **locate** positions the tape to a specific logical block, not an entire track (corresponds to **seek**).
- The **read position** operation returns the logical block number where the tape head is.
- The **space** operation enables relative motion.
- Tape drives are "append-only" devices; updating a block in the middle of the tape also effectively erases everything beyond that block.
- An EOT mark is placed after a block that is written.