Disk Management and File Systems

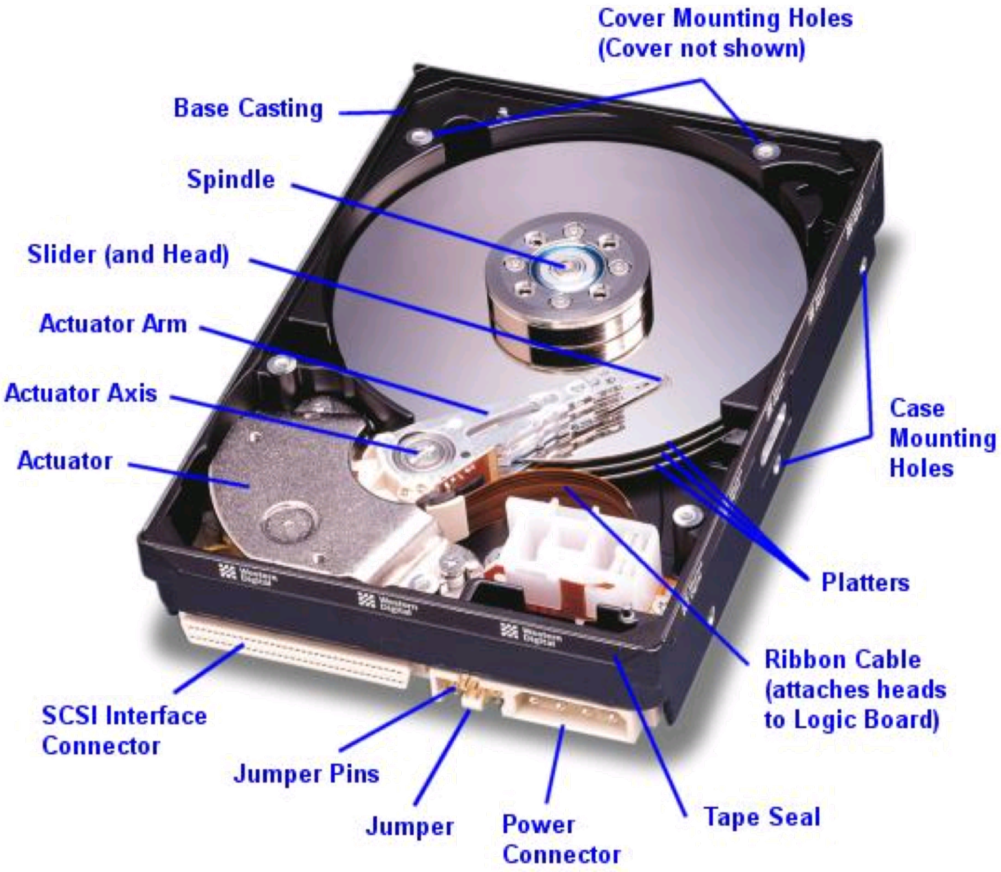
http://inst.eecs.berkeley.edu/~cs162

Goals for Today

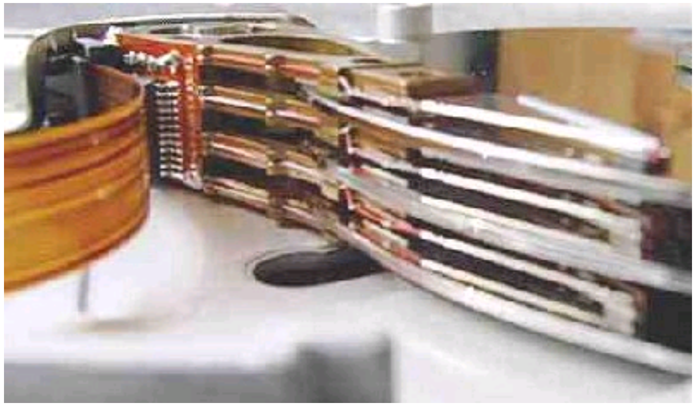
- Disk Performance
  - Hardware performance parameters

Note: Some slides and/or pictures in the following are adapted from slides ©2005 Silberschatz, Galvin, and Gagne. Many slides generated from my lecture notes by Kubiatowicz.

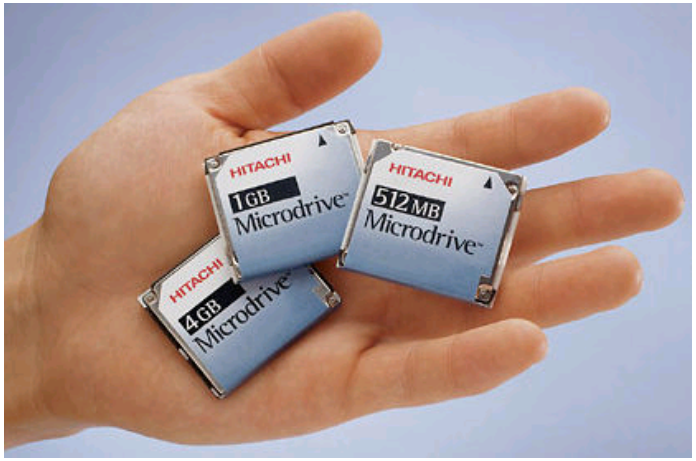
Hard Disk Drives



Western Digital Drive
http://www.storagereview.com/guide/

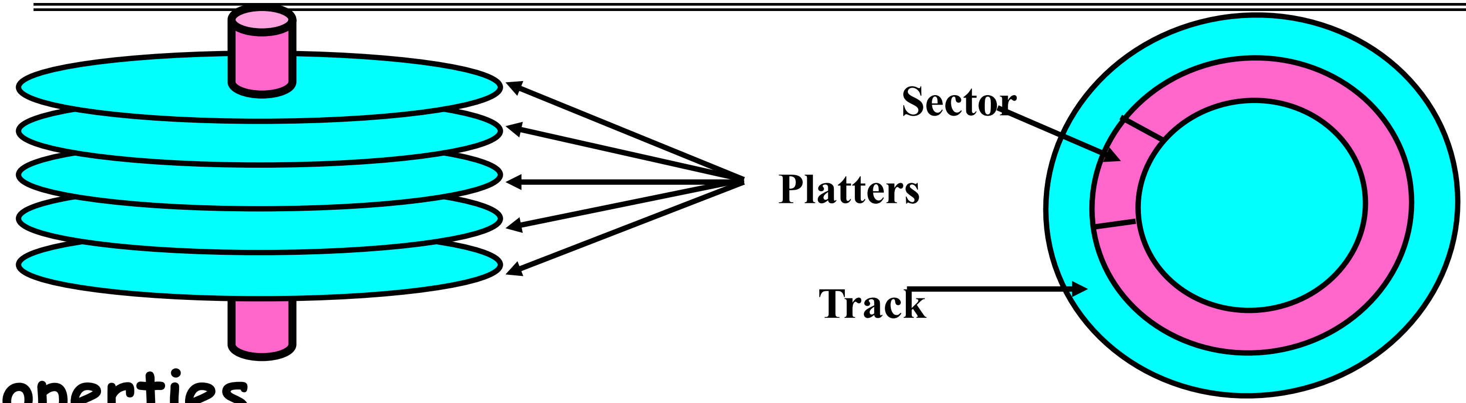


Read/Write Head
Side View



IBM/Hitachi Microdrive

Properties of a Hard Magnetic Disk



- Properties
  - Independently addressable element: sector
    - » OS always transfers groups of sectors together—"blocks"
  - A disk can access directly any given block of information it contains (random access). Can access any file either sequentially or randomly.
  - A disk can be rewritten in place: it is possible to read/modify/write a block from the disk
- Typical numbers (depending on the disk size):
  - 500 to more than 20,000 tracks per surface
  - 32 to 800 sectors per track
    - » A sector is the smallest unit that can be read or written
- Zoned bit recording
  - Constant bit density: more sectors on outer tracks
  - Speed varies with track location

Here is a primitive picture showing you how a disk drive can have multiple platters.

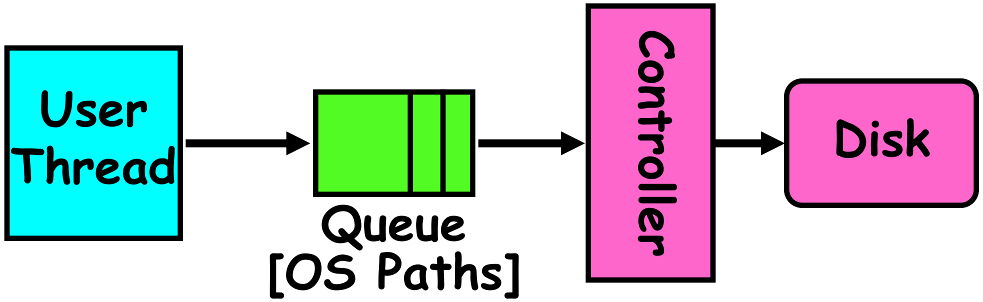
Each surface on the platter are divided into tracks and each track is further divided into sectors. A sector is the smallest unit that can be read or written. By simple geometry you know the outer track have more area and you would thing the outer tack will have more sectors.

This, however, is not the case in traditional disk design where all tracks have the same number of sectors. Well, you will say, this is dumb but dumb is the reason they do it .

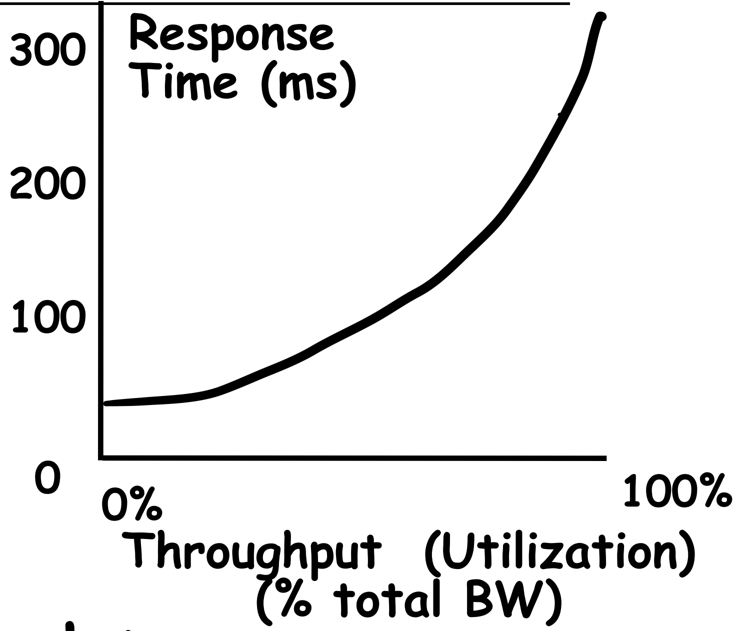
By keeping the number of sectors the same, the disk controller hardware and software can be dumb and does not have to know which track has how many sectors.

With more intelligent disk controller hardware and software, it is getting more popular to record more sectors on the outer tracks. This is referred to as constant bit density.

Disk I/O Performance

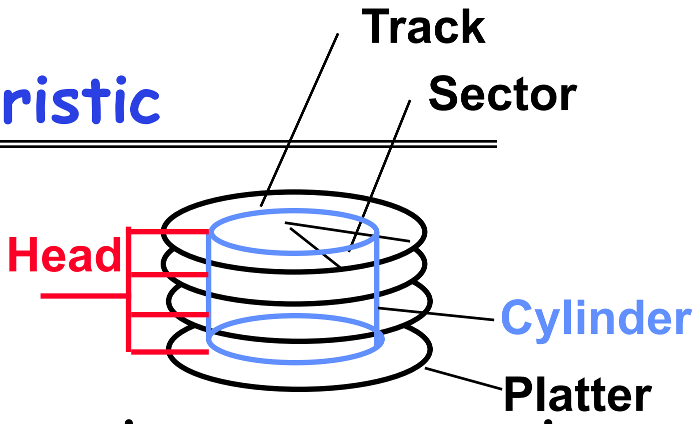


Response Time = Queue+Disk Service Time

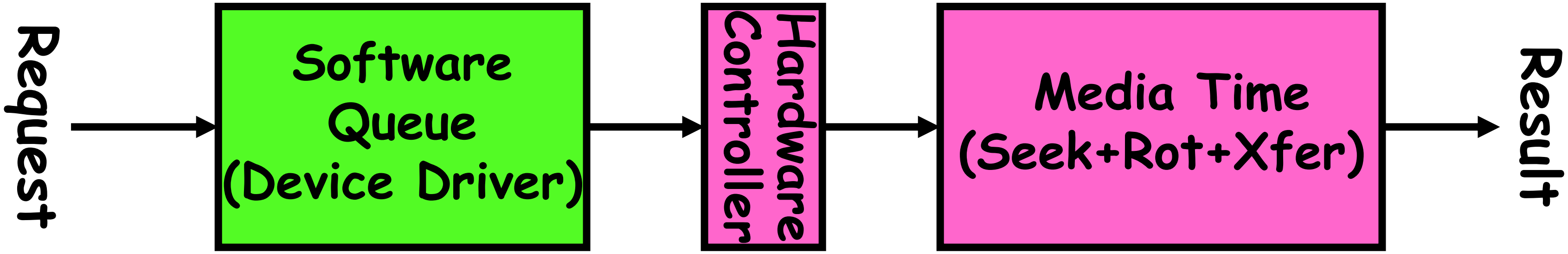


- Performance of disk drive/file system
  - Metrics: Response Time, Throughput
  - Contributing factors to latency:
    - » Software paths (can be loosely modeled by a queue)
    - » Hardware controller
    - » Physical disk media
- Queuing behavior:
  - Can lead to big increases of latency as utilization approaches 100%

Magnetic Disk Characteristic



- Cylinder: all the tracks under the head at a given point on all surface
- Read/write data is a three-stage process:
  - Seek time: position the head/arm over the proper track (into proper cylinder)
  - Rotational latency: wait for the desired sector to rotate under the read/write head
  - Transfer time: transfer a block of bits (sector) under the read-write head
- Disk Latency = Queueing Time + Controller time + Seek Time + Rotation Time + Xfer Time



Request

Software Queue (Device Driver)

Hardware Controller

Media Time (Seek+Rot+Xfer)

Result

- Highest Bandwidth:
  - Transfer large group of blocks sequentially from one track

To read write information into a sector, a movable arm containing a read/write head is located over each surface.

The term cylinder is used to refer to all the tracks under the read/write head at a given point on all surfaces.

To access data, the operating system must direct the disk through a 3-stage process.

(a) The first step is to position the arm over the proper track. This is the seek operation and

the time to complete this operation is called the seek time.

(b) Once the head has reached the correct track, we must wait for the desired sector to

rotate under the read/write head. This is referred to as the rotational latency.

(c) Finally, once the desired sector is under the read/write head, the data transfer can begin.

The average seek time as reported by the manufacturer is in the range of 12 ms to 20ms and is calculated as the sum of the time for all possible seeks divided by the number of possible seeks.

This number is usually on the pessimistic side because due to locality of disk reference, the actual average seek time may only be 25 to 33% of the number published.

Typical Numbers of a Magnetic Disk

- Average seek time as reported by the industry:
  - Typically in the range of 8 ms to 12 ms
  - Due to locality of disk reference may only be 25% to 33% of the advertised number
- Rotational Latency:
  - *Most* disks rotate at 3,600 to 7200 RPM (Up to 15,000RPM or more)
  - Approximately 16 ms to 8 ms per revolution, respectively
  - An average latency to the desired information is halfway around the disk: 8 ms at 3600 RPM, 4 ms at 7200 RPM
- Transfer Time is a function of:
  - Transfer size (usually a sector): 512B – 1KB per sector
  - Rotation speed: 3600 RPM to 15000 RPM
  - Recording density: bits per inch on a track
  - Diameter: ranges from 1 in to 5.25 in
  - Typical values: 2 to 50 MB per second
- Controller time depends on controller hardware
- Cost drops by factor of two per year (since 1991)

As far as rotational latency is concerned, most disks rotate at 3,600 RPM or approximately 16 ms per revolution.
Since on average, the information you desired is half way around the disk, the average rotational latency will be 8ms.
The transfer time is a function of transfer size, rotation speed, and recording density.
The typical transfer speed is 2 to 4 MB per second.
Notice that the transfer time is much faster than the rotational latency and seek time.
This is similar to the DRAM situation where the DRAM access time is much shorter than the DRAM cycle time.
***** Do anybody remember what we did to take advantage of the short access time versus cycle time? Well, we interleave!

New International Disk Drive, Equipment, and Materials Association standard is 4KB sectors instead of 512 byte sectors

Disk Performance

- Assumptions:
  - Ignoring queuing and controller times for now
  - Avg seek time of 5ms, avg rotational delay of 4ms
  - Transfer rate of 4MByte/s, sector size of 1 KByte
- Random place on disk:
  - Seek (5ms) + Rot. Delay (4ms) + Transfer (0.25ms)
  - Roughly 10ms to fetch/put data: 100 KByte/sec
- Random place in same cylinder:
  - Rot. Delay (4ms) + Transfer (0.25ms)
  - Roughly 5ms to fetch/put data: 200 KByte/sec
- Next sector on same track:
  - Transfer (0.25ms): 4 MByte/sec
- Key to using disk effectively (esp. for filesystems) is to minimize seek and rotational delays

Disk Tradeoffs

- How do manufacturers choose disk sector sizes?
  - Need 100-1000 bits between each sector to allow system to measure how fast disk is spinning and to tolerate small (thermal) changes in track length
- What if sector was 1 byte?
  - Space efficiency – only 1% of disk has useful space
  - Time efficiency – each seek takes 10 ms, transfer rate of 50 – 100 Bytes/sec
- What if sector was 1 KByte?
  - Space efficiency – only 90% of disk has useful space
  - Time efficiency – transfer rate of 100 KByte/sec
- What if sector was 1 MByte?
  - Space efficiency – almost all of disk has useful space
  - Time efficiency – transfer rate of 4 MByte/sec

Summary

- I/O Controllers: Hardware that controls actual device
  - Processor Accesses through I/O instructions or load/store to special physical memory
- Notification mechanisms
  - Interrupts
  - Polling: Report results through status register that processor looks at periodically
- Disk Performance:
  - Queuing time + Controller + Seek + Rotational + Transfer
  - Rotational latency: on average ½ rotation
  - Transfer time: spec of disk depends on rotation speed and bit storage density