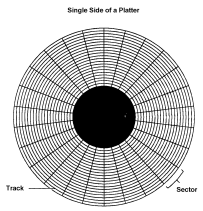
IO Requirements: A Differentiating Characteristic!

- Computer systems are often labeled with names in which the distinguishing factor is the I/O requirement
  - "workstation" vs. "server" (modem)
  - "minicomputer" vs. "mainframe" (late 70's, early 80's)
- Modern I/O requirements are usually based on disk service need or network service need
- Issues in Disk I/O
  - Capacity
  - Transaction time
  - Availability
  - Reliability

Disk Layout

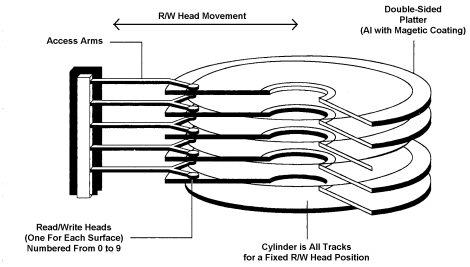
- *"Platter"* – Circular Piece of Magnetic Material – Coated Aluminum
- One Platter Has 2 *"Sides"*
- Each Side is Divided into Concentric Rings Called *"Tracks"*
- Each Track is Divided into Arcs Called *"Sectors"*
- Each Sector Can Store Some Number of *"Bytes"* of Data
- *"Media Capacity"* Determines Amount of Storage per Sector



Physical Disk Coordinates

- A Platter can be *"Double Sided"* or *"Single Sided"*
- A Single Fixed Disk can Contain Multiple Platters
- Need Three-Dimensional Coordinate
- A Stack of Tracks On Top of Each Other is a *"Cylinder"*
- Each Side of a Platter Must Have a *"Read/Write Head"*
  Contains Induction Coil – Floats Over Surface
  (Except for Floppy Where it Actually Touches)
- Position of Read/Write Head Given by *"Cylinder"* Number
- Physical Location on Disk: `<cylinder, head, sector>`

Multi-Platter Disk Layout



Disk Performance - Bandwidth

- *"Seek Time"* – (Average) Time Required to Position R/W Head

  Over a Track/Sector (5-15 ms)
- *"Rotational Latency"* – After Seek Wait for Sector to be Under R/W Head

  Typical Disks Rotate at 3600, 5400, 7200, 10800 RPM

  Average Rotational Latency = ½ Rotation (4-8ms)
- *"Transfer Time"* – Time for 1 Sector of Data to be Read/Written

  After Seek and Rotational Latency (25-50µs)
- *"Access Time"* is Sum of Above – Dominated by Seek and

  Rotational Latency

Disk Performance - Capacity

- *"Linear Bit Density"* – Number of Bits/cm that can be Stored Along a Track
- Linear Bit Density *(Typically 50,000-100,000 bits/cm)* Depends on:

  Purity of Magnetizable Media

  Air Quality

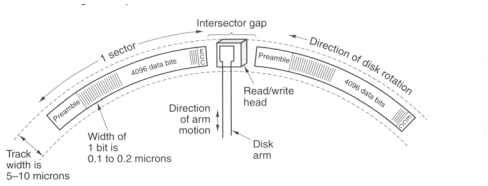
  Data Encoding
- *"Winchester Disk"* – Sealed at Factory to Preserve Air/Surface Quality

  -First "Winchester Disk" Was 30 MB Fixed and 30 MB Removable (IBM)

  -.30-30 Winchester
- *"Unformatted Capacity"* – Disregards Preamble, ECC and Intersector Gap

  -Marketing Hype

Capacity Along a Sector



- Preamble – Used to Synchronize R/W Head

  "formatting" a disk is "writing" all preambles on it

- 4096 bits – Actual Data Storage Area (512 bytes is typical)
- ECC – Error Correcting Code – Usually a Reed-Solomon Code

Disk Controllers

- Each Disk Drive Has a Controller Associated With It

  Purpose of Controller:

  1) Accept Commands from Software (READ, WRITE, FORMAT)
  2) Control Arm Motion
  3) Detect/Correct Errors
  4) Convert parallel data into Serial Bit-Stream (vice-versa)
  5) Remap bad Sectors to Spares (each track has spares)
  6) Caching Sectors

- IDE – "Integrated Drive Electronics" – Most Controller Functionality is Inside the Drive Enclosure
- EIDE – "Extended IDE"
- SCSI – "Small Computer System Interface" – Actually a Bus Standard – Disk Media is Same

Some Calculations

Assume a disk has a read/write bandwidth of 5 MB/s average (transfer rate).

Assume that seek time + rotational latency is 10 ms.

What % of the bandwidth of a 20MB/s SCSI bus is used for a 64KB transfer?

Disk IO Time = seek/rotational latency + transfer time
= 10 ms + 64KB/(5 MB/s) = 22.8 ms.

IO transfer rate = Transfer size/Transfer time
= 64KB/22.8 ms = 2.74 MB/s

% bandwidth needed = IO transfer rate/ Bus transfer rate
= 2.74 MB/s / 20 MB/s = 13.7%

More Calculations

In the previous calculation, assume that only 50% of the 20 MB/s bandwidth of the SCSI bus is actually achievable.

Assuming the previous disk transfer rates, I/O transactions – how many disks could I place on the SCSI bus before it saturates?

Achievable SCSI bandwidth = 0.5 * 20 MB/s = 10 MB/s

# of SCSI Disks = 10 MB/s / 2.74 MB/s = 3.6 = 3 disks

Assume I get new disks that have a 10 MB transfer rate. How does this change things?

Disk IO Time = seek/rotational latency + transfer time
= 10 ms + 64KB/(10 MB/s) = 16.4 ms.

IO transfer rate = Transfer size/Transfer time
= 64KB/16.4 ms = 3.9 MB/s

# of SCSI disk = 10 MB/s / 3.9 MB/s = 2.6 = 2 disks

RAID – Redundant Array of Inexpensive Disks

- Used to increase both performance and reliability
- Increases performance by providing parallel access to multiple disks
- Increases reliability via
  - replication – data mirrored on disks
  - striping with error correction – data striped across disk along with error correction data. Can support failure of any single disk drive (disks are hot swappable!).

RAID Levels

- Level 0: simply parallel disks in which data is striped with no redundancy
- Level 1: data is mirrored to 2nd disk. Requires twice as many disks as level 0. Can survive 1 disk failure.
- Level 3: data is striped at byte level across several drives, one drive holds checksum information that allows recovery from one failed drive (including checksum drive).
- Level 5: data is striped at block level across several drives, and parity is also distributed.
  - Better than level 3 for multiprocessing systems because can support multiple concurrent read accesses if data is limited to block size.