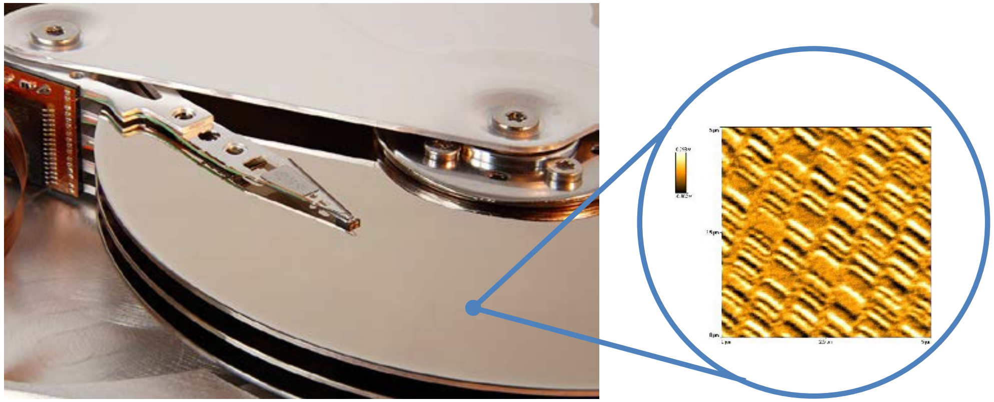
CE1006/CZ1006
Computer Organisation and Architecture

Computer Memory
Magnetic Hard Disk Drive

Oh Hong Lye
Lecturer
SCSE, Nanyang Technological University.

Magnetic Hard Disk

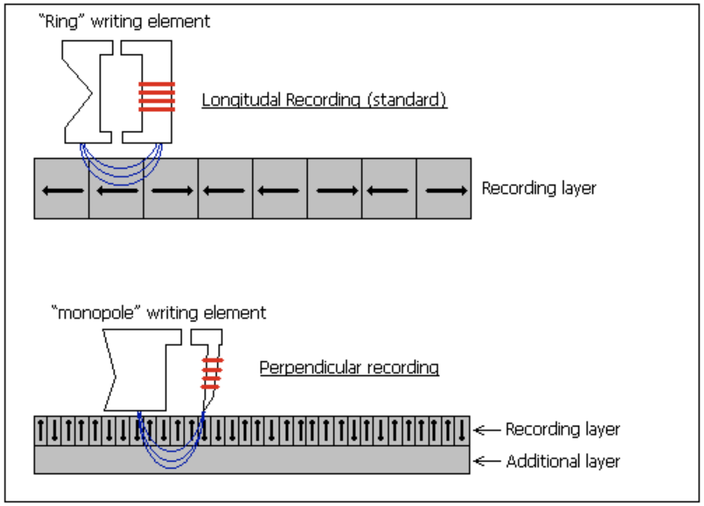


[Source] http://en.wikipedia.org/wiki/Disk_read-and-write_head

TECHNOLOGY METRIC	2010	2014 (40% /YR)
TAPE		
-- Areal Density	1.2 Gbit/in²	4.8 Gbit/in²
-- Bit Length	8000 nm	2000 nm
-- Bit Width	100 nm	100 nm
-- Minimum Feature	4000 nm	1000 nm
HDD		
-- Areal Density	635 Gbit/in²	2500 Gbit/in²
-- Bit Length	74 nm	19 nm
-- Bit Width	13.5 nm	13.5 nm
-- Minimum Feature	37 nm	10 nm
NAND Flash		
-- Areal Density	330 Gbit/in²	1300 Gbit/in²
-- Bit Length	45 nm	20 nm
-- Bit width	45 nm	20 nm
-- Minimum Feature	25 nm	12 nm

[Source] Robert Fontana et al, IBM Areal Density Comparison Paper, 2010

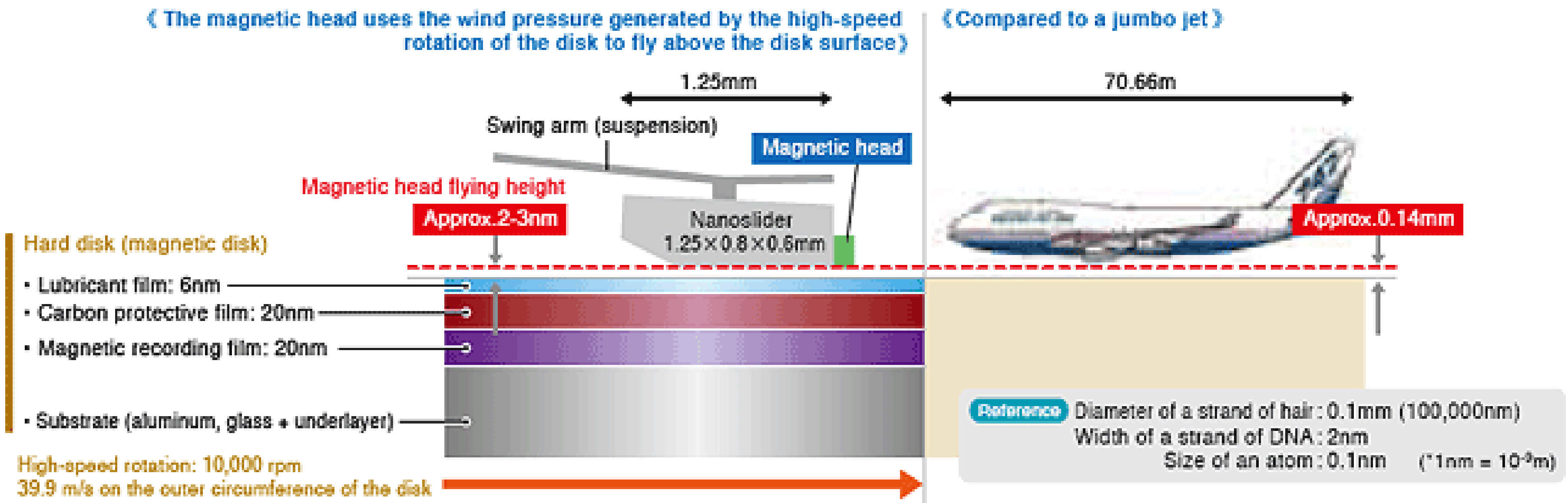
Longitudinal vs Perpendicular Recording



- Stores data by magnetising a thin film of ferromagnetic media on the circular disk known as platter.
- Video on Introduction to Hard Disk
  - https://www.youtube.com/watch?v=kdmLvl1n82U

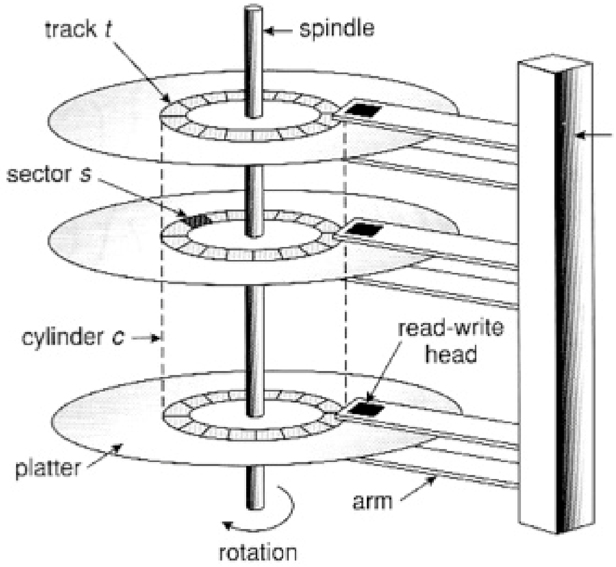
HDD Technology

Flying the Magnetic Head over the HDD media is like making a jumbo jet cruise a few mm above the ground!

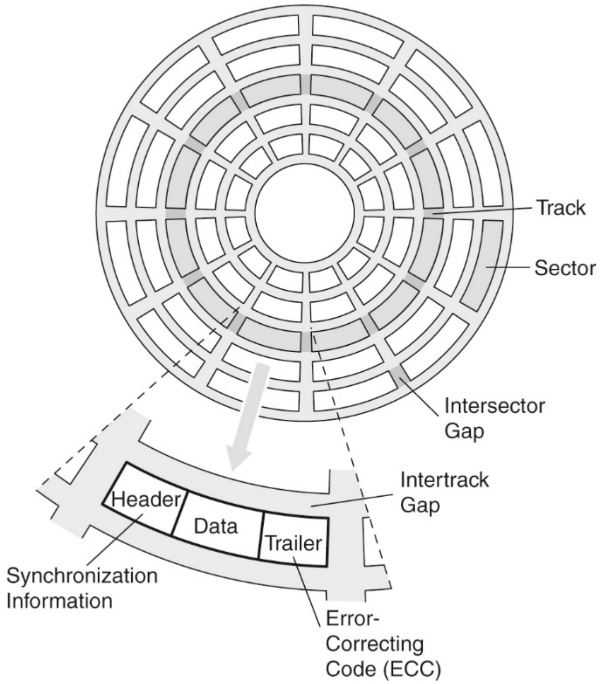


Source: http://product.tdk.com/en/environment/ecolove/eco20000.html
http://www.thestoragealchemist.com/datageddon/hdd-and-magnetic-heads/

Magnetic HDD Data Organisation



- Computers often use magnetic hard disks for large secondary storage devices.
  - One or more platters on a common spindle.
  - Platters are covered with thin magnetic film.
  - Platters rotate on spindle at constant rate.



- Data is stored on the surface of the platter in concentric rings called tracks.
  - Gaps between tracks to minimise interferences from adjacent tracks.
- Tracks are divided into sectors.
  - Minimum data block size is a sector.
- Stored data consists header, data and trailer.

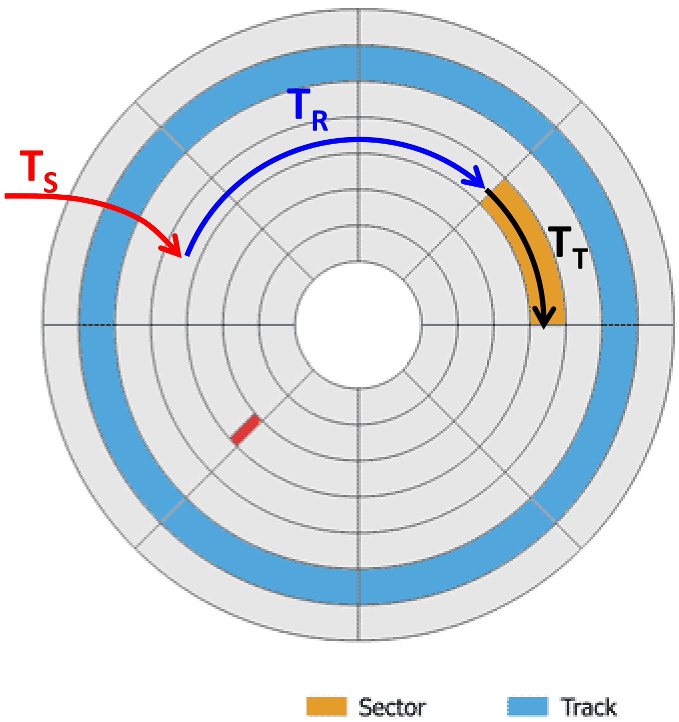
CE1006/CZ1006
Computer Organisation and Architecture

Computer Memory
HDD Transfer Rate

Oh Hong Lye
Lecturer
SCSE, Nanyang Technological University.

HDD Transfer Rate

- Seek time (T_S)
  - Time taken for the head to move to the correct track.
- Rotational Delay (T_R)
  - Time taken for the disk to rotate until the read/write head reaches the starting position of the target sector.
- Access Time (T_A)
  - Time from request to the time the head is in position ($T_S + T_R$)
- Transfer Time (T_T)
  - Time required to transfer the required data after the head is positioned.



HDD Transfer Rate

- T_R is dependent on the rotational speed, Revolutions Per Minute (RPM), of the disk. For calculations, RPM is usually converted to Revolutions Per Second (RPS). i.e. RPS = RPM/60
- For a random section, average rotational delay $T_{R,AV}$ may be calculated as

$$T_{R,AV} = \frac{0.5}{RPS} \quad \text{seconds}$$

- T_T is dependent on the rotational speed of the disk, the Track Density D_T (number of sectors per track), Sector Density D_S (number of bytes per sector) and the number of bytes N for the transfer.

$$T_T = \frac{N}{RPS * D_T * D_S}$$

HDD Transfer Rate Example

- A magnetic hard disk rotates at 15000 RPM, with the following properties:
  - Average Seek Time, T_S = 4ms
  - Track density, D_T = 500 sectors per track
  - Sector density, D_S = 512 bytes per sector
- Calculate the total time T_{TOTAL} it takes to read a 3 KB file stored in consecutive sectors on the same track?

[Solution]

RPS = 15000/60 = 250 per second

Access time $T_A = T_S + T_R$ = 4 ms + (0.5/250) = 6 ms

Transfer time T_T = 3072/(250*500*512) = 48 µs

Total time, $T_{TOTAL} = T_A + T_T$ = 6.048 ms

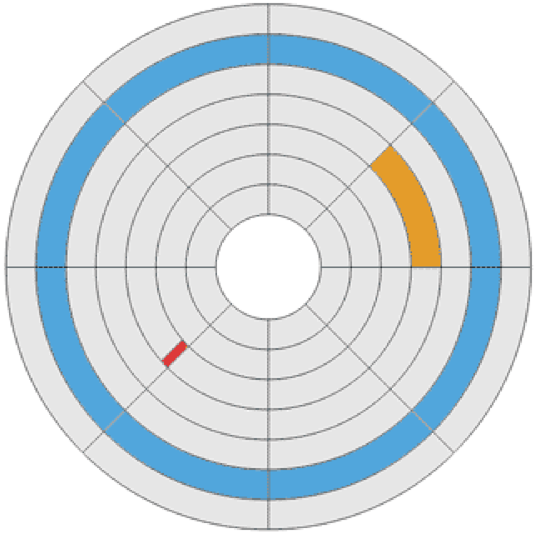
Notice $T_A \gg T_T$. This example shows that accessing random files on the magnetic hard disk can be the bottleneck, since it has significant access time.

CE1006/CZ1006
Computer Organisation and Architecture

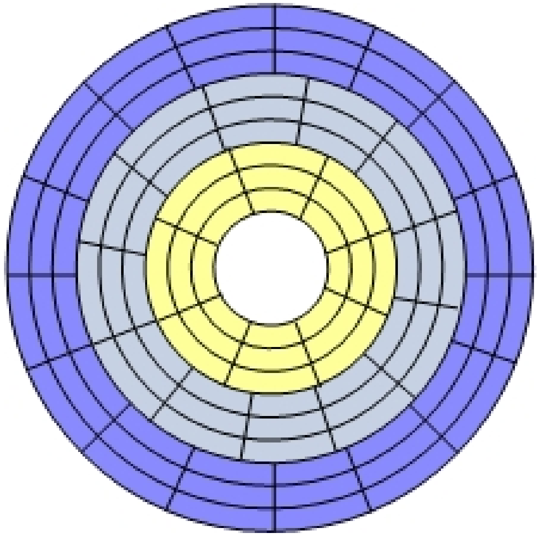
Computer Memory
HDD Layouts and SSD

Oh Hong Lye
Lecturer
SCSE, Nanyang Technological University.

Physical Layout



- Early HDD
  - Physical Layout refers to the actual layout design on the HDD.
  - Equal number of sectors per track and each sector has the same data size.
  - Same-numbered Tracks from different platters formed a cylinder.
  - The early hard disks were implemented using this topology to simplify the controller design.
- Modern HDD
  - Having equal number of sectors per track means that sectors at the outer tracks are wider.
  - Waste physical space as bit density of those sectors are not optimal.
  - Zone bit recording technique addresses space wastage.
  - Tracks are divided into zones, with differing number of sectors per track for different zones.

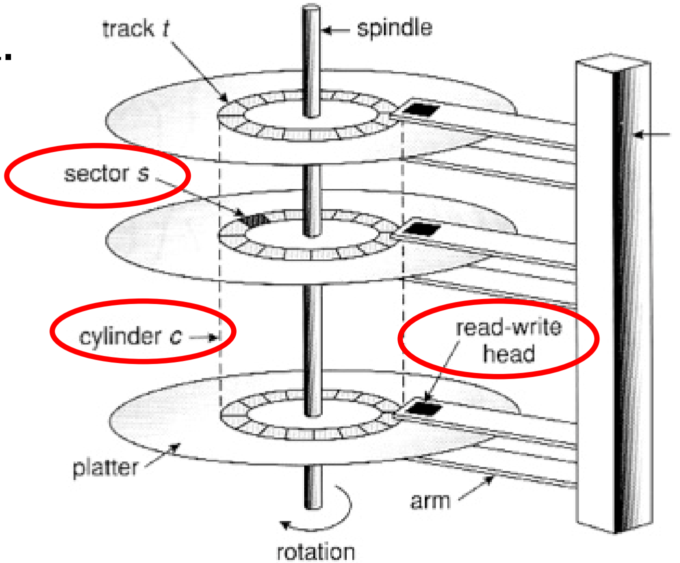


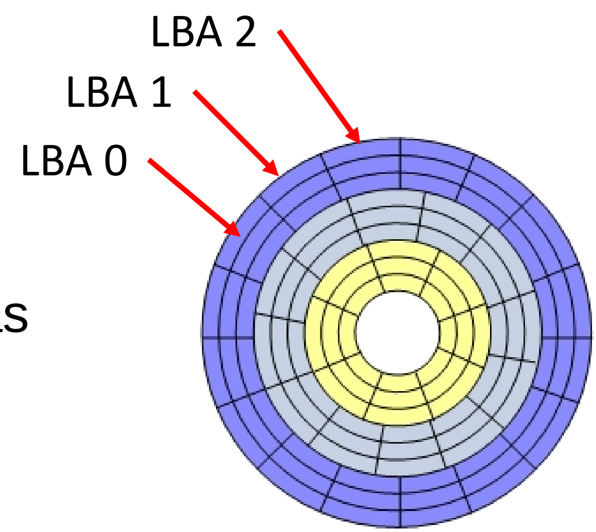
Logical Layout

- How the software see and address the HDD data.
- Address translations are needed to map the physical to logical locations. Two addressing scheme are CHS and LBA.

- Cylinder-Head-Sector (CHS)
  - Legacy scheme using the old HDD physical structure.
  - Made Obsolete in recent standards due to incompatibility with HDD Physical layout and more complex formatting.

- Logical Block Addressing (LBA)
  - Simple linear addressing starting from LBA=0 as first block, LBA=1 as second block and so on.
  - 48-bit LBA standard allows addressing up to 128PByte. 1PByte=2^50 Bytes.





HDD Physical Layout Information

- Not much Physical layout information of modern HDD is revealed in their datasheets these days as it create more confusion than information.

Description	500GB	400GB	320GB
Physical Layout			
Bytes per Sector	512	512	512
Number of Heads	4	4	3
Number of Disks	2	2	2
Logical Layout			
Number of Heads	16	16	16
Number of Sectors/ Track	63	63	63
Number of Cylinders	16,383	16,383	16,383
Number of Sectors	976,773,168	781,422,768	625,142,448
Total Logical Data Bytes	500,107,862,016	400,088,457,216	320,072,933,376

[Source] Hitachi Travelstar 5K500.B SATA OEM Specification, Revision 1.2, 17 March 2009

Solid State Drive (SSD)

- Solid-state drives (SSD) are becoming popular
  - Memory array based on NAND-FLASH or NOR-FLASH.
  - Floating gate transistor
  - Still more expensive than magnetic hard-disk.
- Limited program-erase cycles of Flash Memory (3,000 to 1,000,000).
- Various techniques used to extend the life of SSD
  - Wear levelling is a technique used to extend the life of the SSD disk by distributing data and erase/write cycles evenly over the entire disk.
  - Use External RAM as buffer to minimise the number of writes to Flash in SSD.

HDD vs SSD

- HDD
  - Pros
    - Lower Cost Per Bit,
    - Almost infinite Erasure cycles
  - Cons
    - Consist of Moving Mechanical Parts so more prone to crashing if HDD is dropped or shaken.
    - Heavier and larger physical profile.
    - Slower transfer rate
- SSD
  - Pros
    - No moving parts. More robust to movement.
    - Lighter and occupy less space
    - Higher Transfer rate
  - Cons
    - More Costly compared to HDD
    - Finite number of Erasure cycles

CE1006/CZ1006

Computer Organisation and Architecture

Computer Memory
Reliability and Redundancy

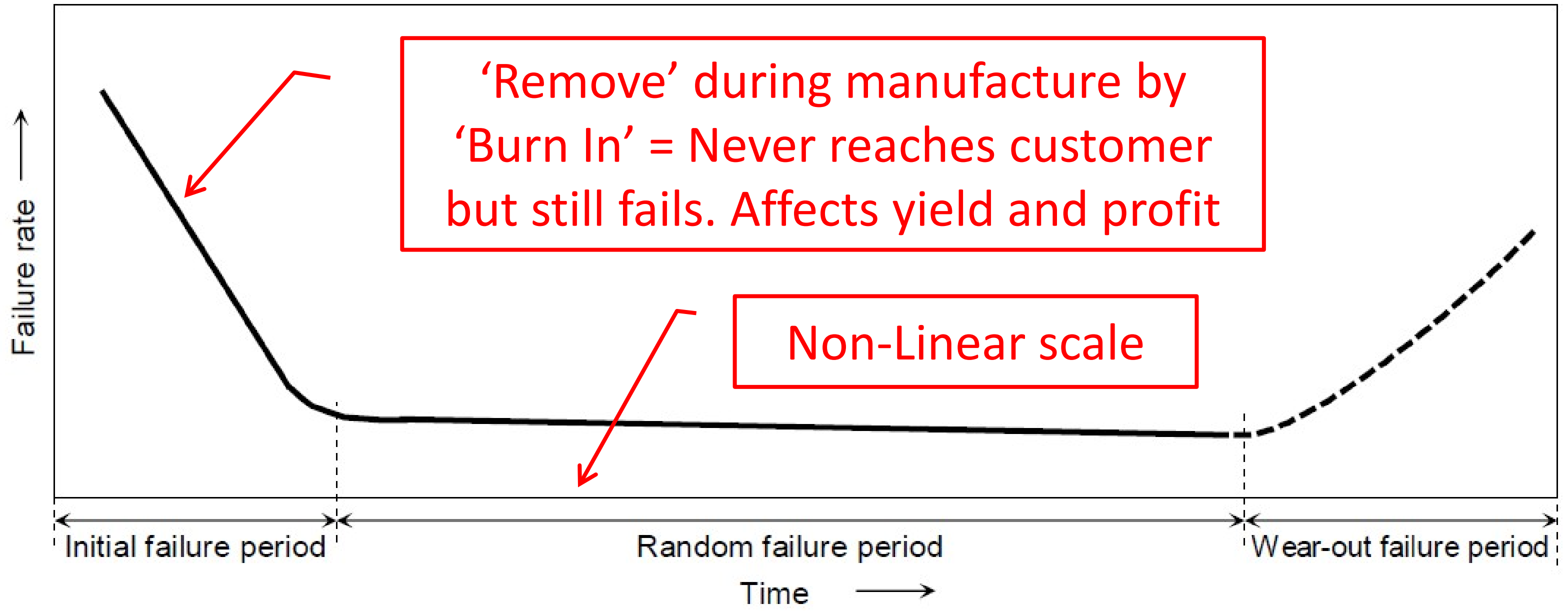
Oh Hong Lye

Lecturer

SCSE, Nanyang Technological University.

Reliability – Bath Tub Curve

- Typical reliability curve for most products.



- Initial failure, e.g. Design/Manufacturing/Testing
- Random failure e.g. Corner cases failure, user issue.
- Wear-out failure e.g. Component aging/degradation

Reliability -

- Many parameters used to quantify reliability
  - MTTF, MTBF, MTTR etc
  - MTTF - Mean (Average) Time To Failure.

- Source of much confusion
  - MTTF of 10 years does not mean the typical life is 10 years
  - The probability that a unit survives is often modelled as

$$R(t) = e^{-\frac{time}{MTTF}}$$

  - This means that the probability a particular unit survives to its MTBF is

$$R(t) = e^{-\frac{MTTF}{MTTF}} = e^{-1} = 0.368$$

  - At time=MTTF, only 37% chance the unit is still functioning.

CE1006/CZ1006
Computer Organisation and Architecture

Computer Memory
HDD RAID System

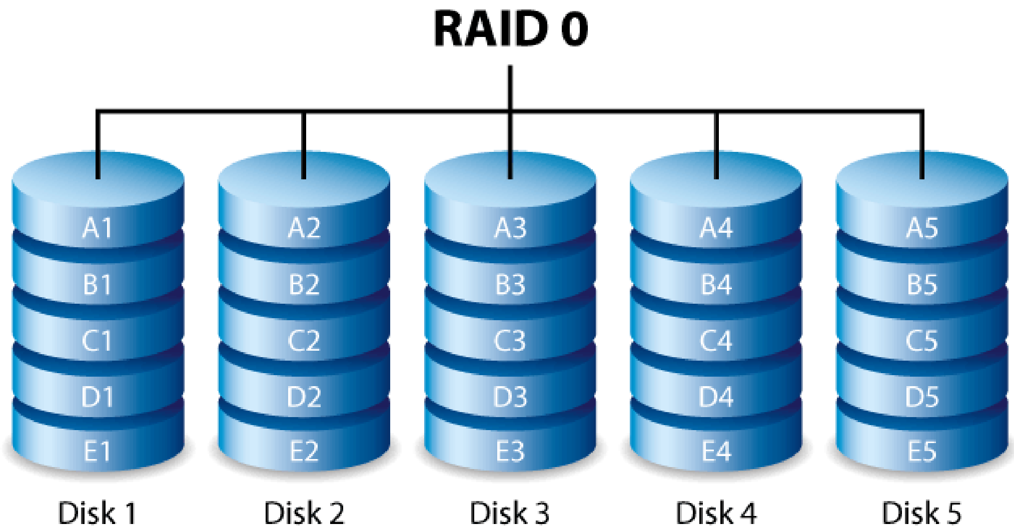
Oh Hong Lye
Lecturer
SCSE, Nanyang Technological University.

Redundancy and RAID

- A single magnetic hard disk suffers from:
  - The access times for moving the heads to the correct position significantly lowers the transfer rate of the disk.
  - Magnetic hard disk are mechanical devices that suffers "easily" from crashes.

- Redundant Array of Independent Disks (RAID)
  - Coined in a 1988 paper "A Case for Redundant Arrays of Inexpensive Disks", RAID was devised to improve both reliability and performance.
  - Use multiple physical disk as a single logical drive.
  - Striping is the segmentation of logically sequential data so that segments are written across multiple physical drives.
  - The segments can be a single bit, or blocks of a specific size.
  - RAID configurations: Level 0 to 6.

RAID level 0

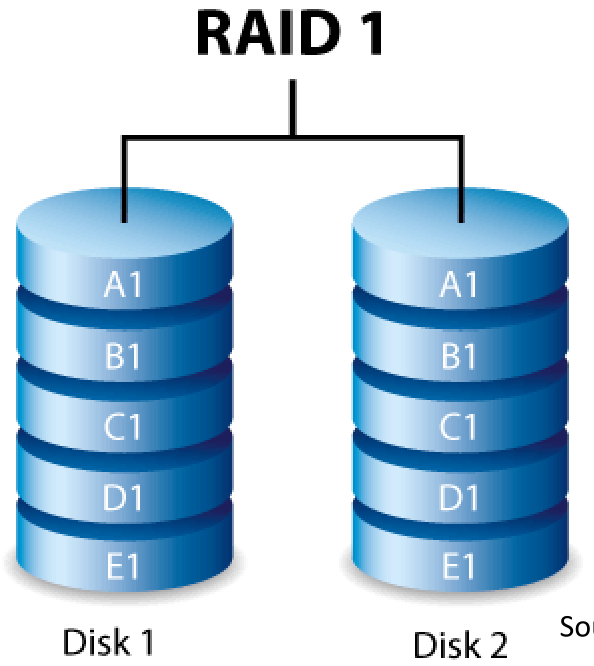


Source: Seagate

- No redundancy
  - Best performance.
  - No fault tolerance. Lack of reliability is a key concern, making this configuration suitable for noncritical data such as video editing.
- Needs a minimum of 2 disks.
- Data striped as blocks across all disks.

RAID Level 1

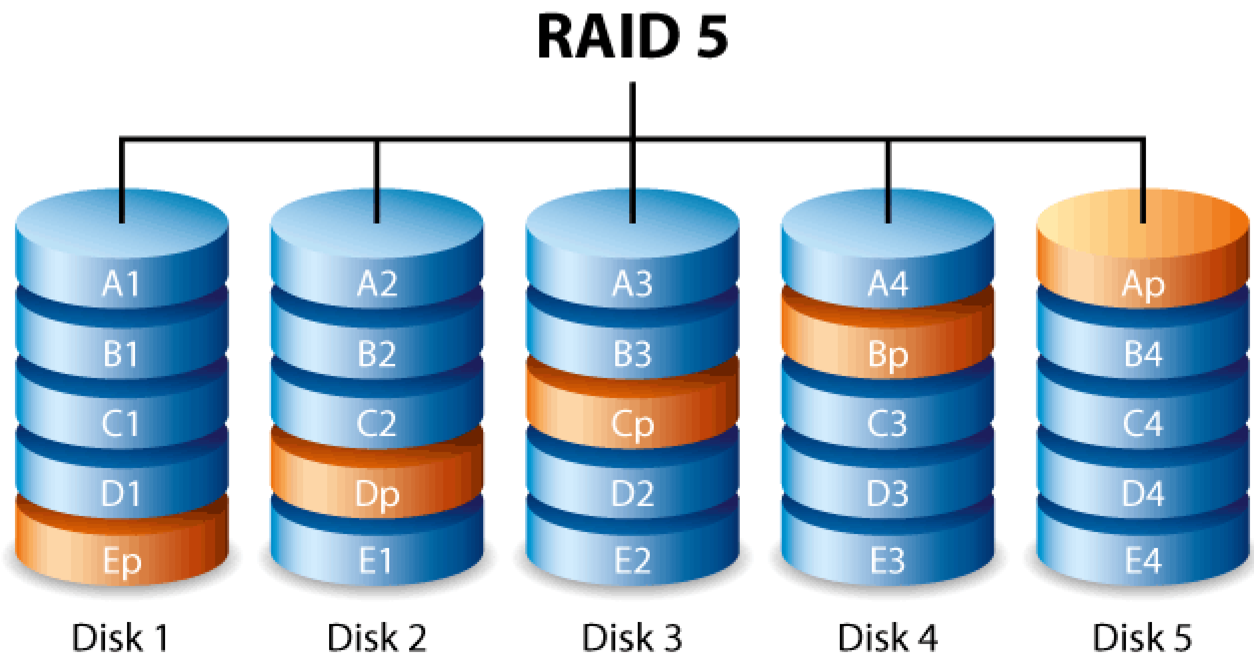
- Mirrored disks
  - 100% redundancy.
  - Data is written identically to each disk.
  - Write throughput is slower because every drive must be updated.
- Needs a minimum of 2 disks.



Source: Seagate

Advantage	Disadvantage
Faster read operations ▪ Read from either disk	Expensive
Disk recovery is simple ▪ Swap faulty disk & re-mirror ▪ No down time	Write throughput is slower because every drive must be updated.

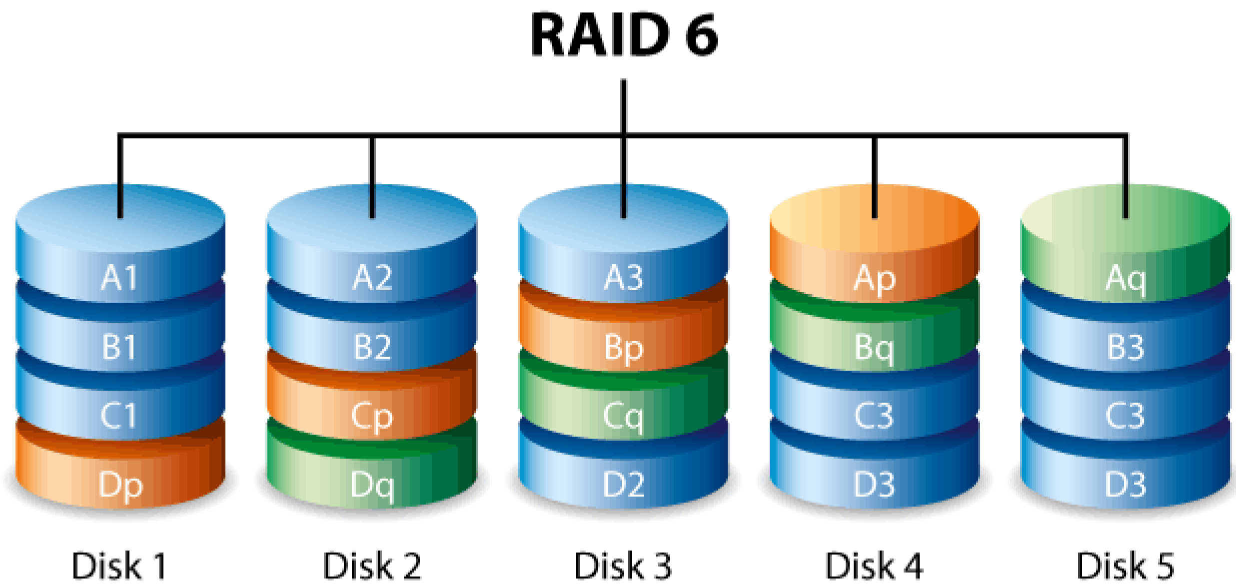
RAID Level 5



Source: Seagate

- Block level stripping with single distributed parity
- Data on single failed drive can be reconstructed from surviving data and parity info.
- Commonly used in network servers.
- RAID 3 & 4 are similar in concepts but stores parity information on dedicated disks.

RAID Level 6

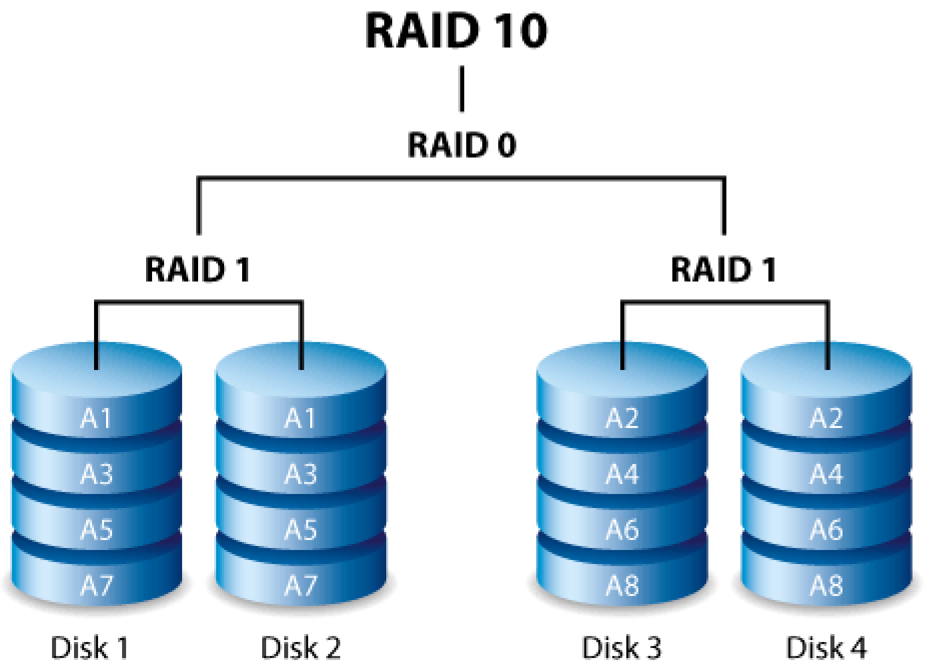


Source: Seagate

- May recover from multiple disk failures.
- Similar to RAID-5 but uses block-level striping with double distributed parity.

Advantage	Disadvantage
High data availability	Significant write penalty

RAID Level 10



Source: Seagate

- RAID-1 + RAID-0
- Combines the protection of RAID 1 with the performance of RAID 0.