Computer Organization

COMP2120

Qi Zhao

March 13, 2024

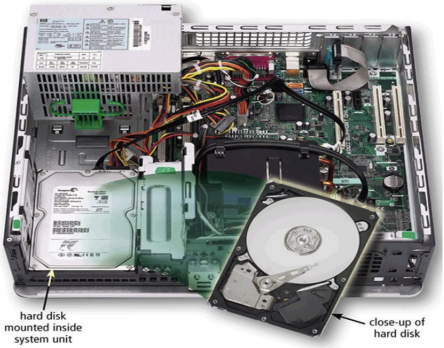
External Storage

Magnetic Disks

- consists of magnetically coated platters.
- Each surface divided into concentric set of rings, called **tracks**.
- Each track divided into sectors, usually of 512 bytes.

A computer (unopened)

A computer (opened) - Magnetic hard disk shown



A working magnetic hard disk drive

Magnetic Disks

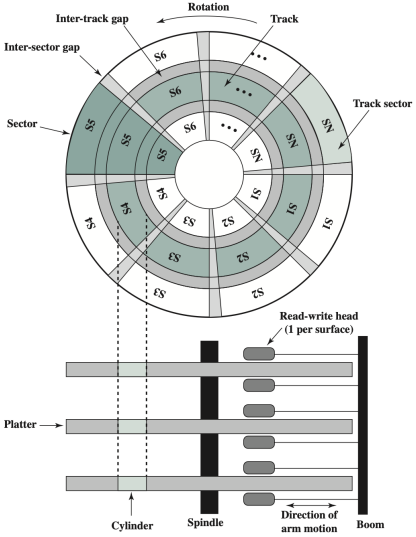
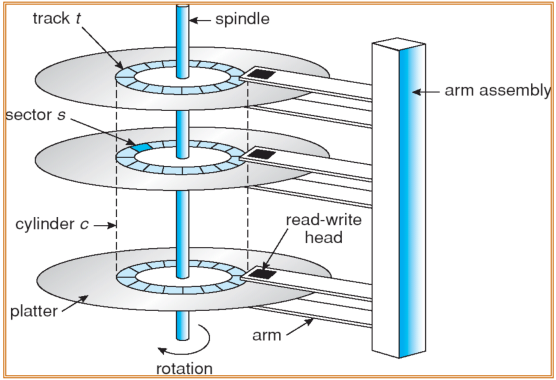


Figure 6.2 Disk Data Layout

Magnetic Disks

- All the read-write head moves together.
- Tracks of different platter under the head at the same time is called a **cylinder**.



Magnetic Disks

- **Gap**: separates sectors
- **Sync**: indicates the beginning of an information field and provides timing alignment.
- **Address mark**: Contains data to identify the sector's number and location.
- **Data**: 512 bytes , or 4096 bytes
- **ECC**: Error correction code

Gap	Sync	Ad	Data	ECC
15 bytes			Data: 512 bytes	50 bytes

Gap	Sync	Ad	Data	ECC
15 bytes			Data: 4096 bytes	100 bytes

Magnetic Disks

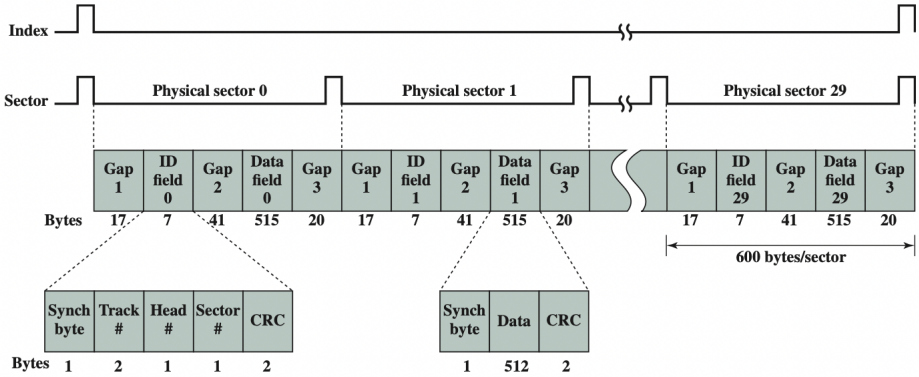
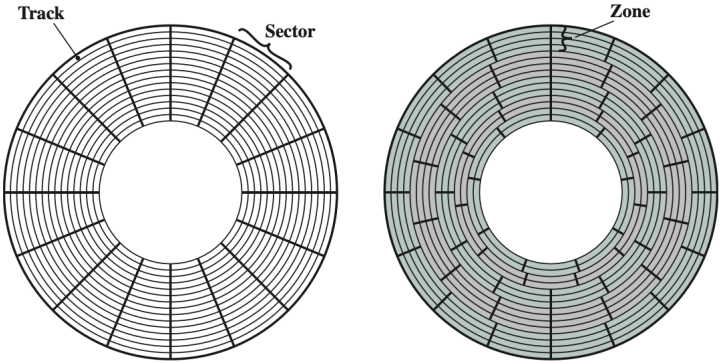


Figure 6.4 Winchester Disk Format (Seagate ST506)

Magnetic Disks

- Constant angular velocity (CAV)
- Multiple zone recording (MZR)



(a) Constant angular velocity (b) Multiple zone recording

Figure 6.3 Comparison of Disk Layout Methods

Disk Access

Disk access time is composed of 3 parts:

- **Seek** — Move the R/W head from one cylinder to another.
  Depends on source track and destination track.
  Typical seek time: initial startup time $5 - 15$ms, $0.2 - 1$ms for consecutive track
- **Rotational delay** — time for the required sector to rotate under the R/W head.
  Average latency: half a revolution (180-degree rotation), 4.2ms for 7200 rpm hard disk
  $7200/60 = 120$ revolutions per second, $1000/120 = 8.33$ms, $8.33/2 \approx 4.2$ms
- **Data Transfer time** — typical transfer rate is 100MB/sec, i.e. require $5\mu s$ to transfer a 512 byte sector. Hence transfer time is $\ll$ seek+latency.

$$t_T = \frac{b}{N} \times \frac{1}{r}$$

where b number of bytes to be transferred, N number of bytes on a track, r rotation speed, revolutions per second.

Typical HDD Parameters

Table 6.2 Typical Hard Disk Drive Parameters

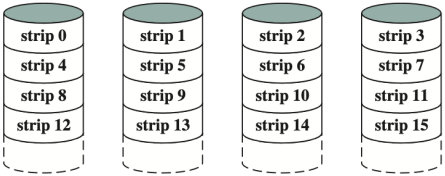
Characteristics	Seagate Enterprise	Seagate Barracuda XT	Seagate Cheetah NS	Seagate Laptop HDD
Application	Enterprise	Desktop	Network-attached storage, application servers	Laptop
Capacity	6 TB	3 TB	600 GB	2 TB
Average seek time	4.16 ms	N/A	3.9 ms read 4.2 ms write	13 ms
Spindle speed	7200 rpm	7200 rpm	10,075 rpm	5400 rpm
Average latency	4.16 ms	4.16 ms	2.98	5.6 ms
Maximum sustained transfer rate	216 MB/sec	149 MB/sec	97 MB/sec	300 MB/sec
Bytes per sector	512/4096	512	512	4096
Tracks per cylinder (number of platter surfaces)	8	10	8	4
Cache	128 MB	64 MB	16 MB	8 MB

RAID

- **RAID**: Redundant Array of Independent (Inexpensive) Disks.
- a set of physical disk drives appeared to be a single hard disk to the operating system.
- Files are **striped** across multiple disks.
- Strips may be physical blocks, sectors, or some other unit.
- Data are distributed over various drives to allow parallel operation.
- Redundant disk capacity is used to store parity information, which guarantees 7 days 24 hour operation in mission critical applications.
- 7 different RAID organization, RAID 0-6.

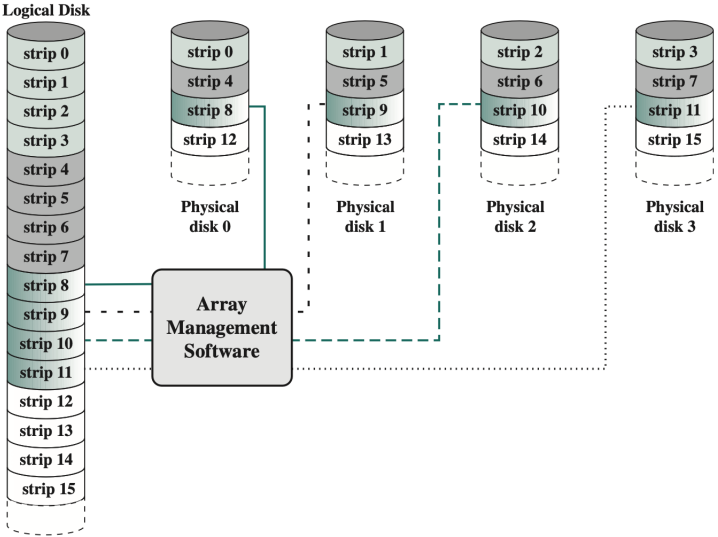
Level 0 RAID

- write consecutive sectors over the drives in a round robin fashion
- efficient when accessing a block of memory — parallelism.
- Nonredundant.



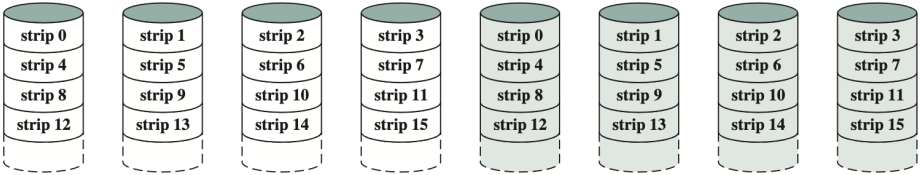
(a) RAID 0 (Nonredundant)

Level 0 RAID


Logical Disk
strip 0
strip 1
strip 2
strip 3
strip 4
strip 5
strip 6
strip 7
strip 8
strip 9
strip 10
strip 11
strip 12
strip 13
strip 14
strip 15
strip 0
strip 4
strip 8
strip 12
Physical disk 0
strip 1
strip 5
strip 9
strip 13
Physical disk 1
strip 2
strip 6
strip 10
strip 14
Physical disk 2
strip 3
strip 7
strip 11
strip 15
Physical disk 3
Array Management Software

Level 1 RAID

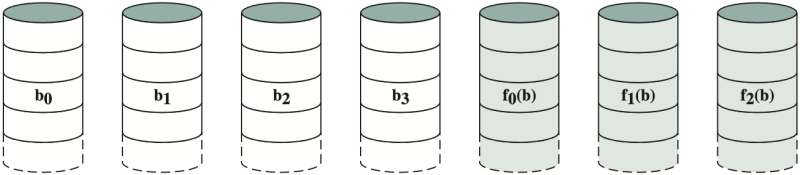
- duplicate the disk for backup: (mirrored)
- Fault tolerant.
- During reading, either copy can be used, hence reduced seek time.
- sacrifice on write: 1 Logical write = two physical writes



(b) RAID 1 (Mirrored)

Level 2 RAID

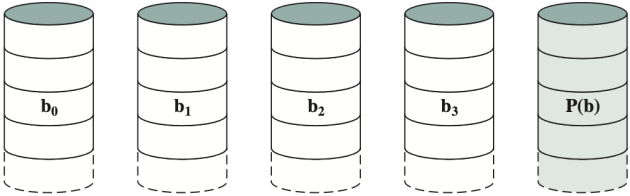
- Error Correction Code (Hamming code) embedded in extra HDDs.
- Expensive, as multiple HDDs required, hence not implemented.
- Redundant disks: log of the number of data disks



(c) RAID 2 (Redundancy through Hamming code)

Level 3 RAID

- The extra HDD contains the parity bit for the corresponding bits of all other HDDs.
- If any HDD fails, can easily recontruct the content of the failed HDD.
- How to reconstruct?



(d) RAID 3 (Bit-interleaved parity)

Level 3 RAID

- For example, an array of 4 drives $\mathrm{X0}, \mathrm{X1}, \mathrm{X2}, \mathrm{X3}, \mathrm{X3}$ is the parity disk.

$$\mathrm{X3(i)} = \mathrm{X2(i)} \oplus \mathrm{X1(i)} \oplus \mathrm{X0(i)}$$

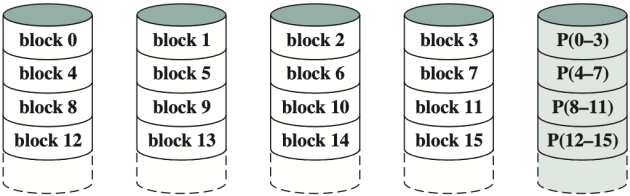
  where $\oplus$ is the exclusive-OR function. Why not OR/AND
- When $\mathrm{X1}$ failed, we add $\mathrm{X3(i)} \oplus \mathrm{X1(i)}$ to both sides of the equation,

$$\mathrm{X1(i)} = \mathrm{X3(i)} \oplus \mathrm{X2(i)} \oplus \mathrm{X0(i)}$$

- This principle is true for RAID levels 3 through 6.
- How to read and write in the event of a disk failure?

Level 4 RAID

- Parity strip used for error correction.
- They are stored in the same HDD.
- Each member disk operates independently.
- Not used. Use RAID 5 instead.



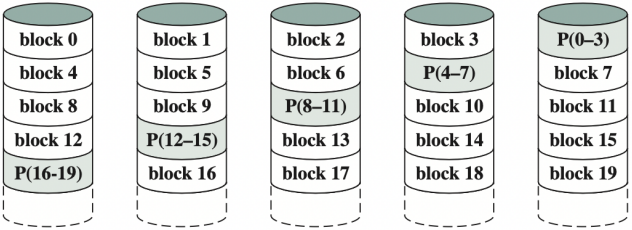
(e) RAID 4 (Block-level parity)

Level 4 RAID

- How to write:
  Option 1:
  read other data disks, create new sum and write to Parity Disk
  Option 2:
  since P has old sum, compare old data to new data, add the difference to P
  Which one is better?
- Write penalty: two reads and two writes
- Every write operation must involve the parity disk, which become a bottleneck.

Level 5 RAID

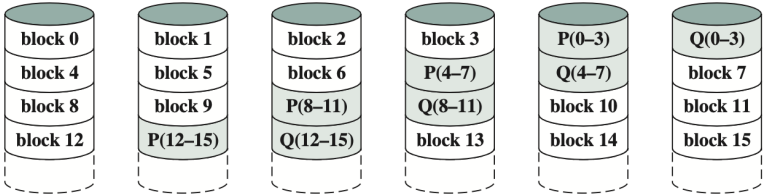
- Parity strip used for error correction.
- Difference from RAID 4: Parity strips are distributed across different HDDs.
- Difference from RAID 3: 3 requires to access all disk (bit level) while 5 require to read only one of them (block level).
- can endure one HDD failure.
- commonly used in NAS(Network Attached Storage).



(f) RAID 5 (Block-level distributed parity)

Level 6 RAID

- Similar to Level 5 RAID, but use two Parity Strips, calculated using different methods
- The two Parity Strips are distributed across different HDDs. Need two extra HDDs
- Can endure two HDD failure.



(g) RAID 6 (Dual redundancy)

RAID Comparison

Table 6.4 RAID Comparison

Level	Advantages	Disadvantages	Applications
0	I/O performance is greatly improved by spreading the I/O load across many channels and drives No parity calculation overhead is involved Very simple design Easy to implement	The failure of just one drive will result in all data in an array being lost	Video production and editing Image Editing Pre-press applications Any application requiring high bandwidth
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability

RAID Comparison

2	Extremely high data transfer rates possible The higher the data transfer rate required, the better the ratio of data disks to ECC disks Relatively simple controller design compared to RAID levels 3, 4, & 5	Very high ratio of ECC disks to data disks with smaller word sizes—inefficient Entry level cost very high—requires very high transfer rate requirement to justify	No commercial implementations exist/not commercially viable
3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput

RAID Comparison

4	Very high Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency	Quite complex controller design Worst write transaction rate and Write aggregate transfer rate Difficult and inefficient data rebuild in the event of disk failure	No commercial implementations exist/not commercially viable
5	Highest Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency Good aggregate transfer rate	Most complex controller design Difficult to rebuild in the event of a disk failure (as compared to RAID level 1)	File and application servers Database servers Web, e-mail, and news servers Intranet servers Most versatile RAID level
6	Provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications

Solid State Drives

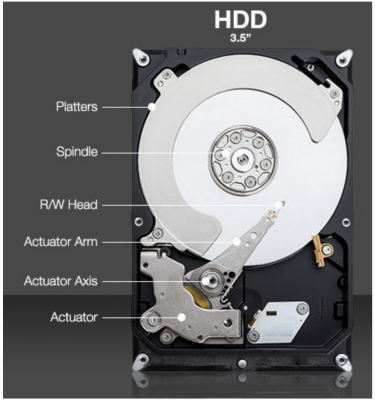
- Semiconductor storage.
- based on Flash Memory (Non-volatile).
- has a limited number of write cycles.

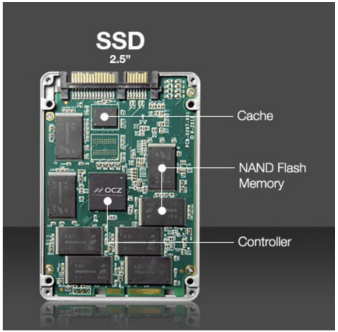
Table 6.5 Comparison of Solid State Drives and Disk Drives

	NAND Flash Drives	Seagate Laptop Internal HDD
File copy/write speed	200–550 Mbps	50–120 Mbps
Power draw/battery life	Less power draw, averages 2–3 watts, resulting in 30+ minute battery boost	More power draw, averages 6–7 watts and therefore uses more battery
Storage capacity	Typically not larger than 512 GB for notebook size drives; 1 TB max for desktops	Typically around 500 GB and 2 TB max for notebook size drives; 4 TB max for desktops
Cost	Approx. $0.50 per GB for a 1-TB drive	Approx. $0.15 per GB for a 4-TB drive

SSD and HDD

HDD: Mechanical Hard drives





Advantages over HDD

- High-performance I/O Operations.
- Durability — less susceptible to physical shock and vibration.
- Longer Lifespan – no mechanical wear (but electronic wear for write operations)
- Lower power consumption — no moving parts, only semiconductor. Quieter and Cooler Operation
- Lower access times — 10 times faster than HDD.