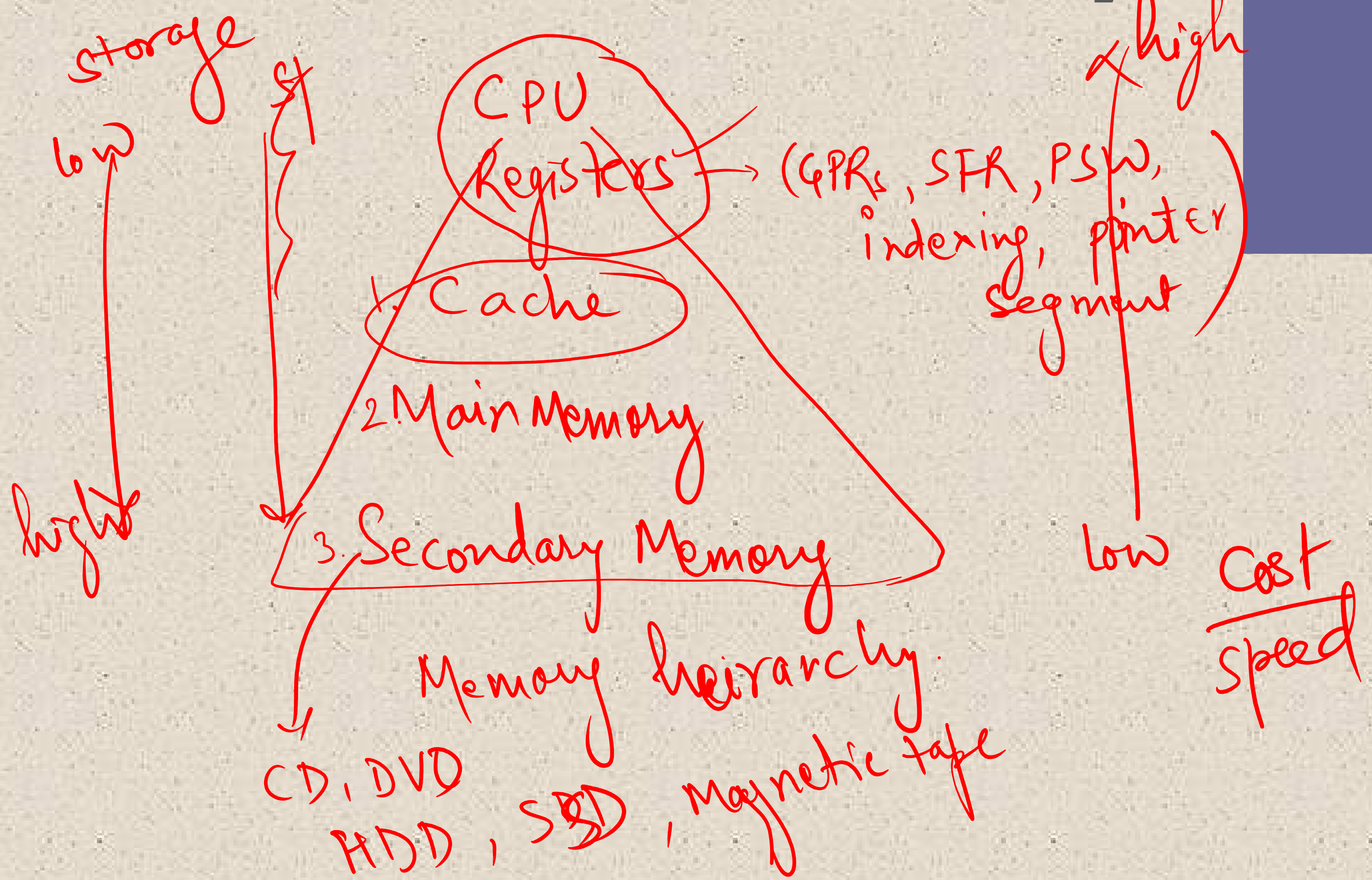
External Memory

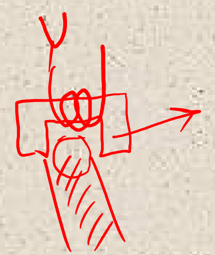


Santosh Kumar Verma
Mentor
GeeksforGeeks

Magnetic Disk

- A disk is a circular *platter* constructed of nonmagnetic material, called the *substrate*, coated with a magnetizable material
  - Traditionally the substrate has been an aluminium or aluminium alloy material
  - Recently glass substrates have been introduced
- Benefits of the glass substrate:
  - Improvement in the uniformity of the magnetic film surface to increase disk reliability
  - A significant reduction in overall surface defects to help reduce read-write errors
  - Ability to support lower fly heights
  - Better stiffness to reduce disk dynamics
  - Greater ability to withstand shock and damage

Magnetic Read and Write Mechanisms



Data are recorded on and later retrieved from the disk via a conducting coil named the *head*

- In many systems there are two heads, a read head and a write head
- During a read or write operation the head is stationary while the platter rotates beneath it

The write mechanism exploits the fact that electricity flowing through a coil produces a magnetic field

Electric pulses are sent to the write head and the resulting magnetic patterns are recorded on the surface below, with different patterns for positive and negative currents

The write head itself is made of easily magnetizable material and is in the shape of a rectangular doughnut with a gap along one side and a few turns of conducting wire along the opposite side

An electric current in the wire induces a magnetic field across the gap, which in turn magnetizes a small area of the recording medium

Reversing the direction of the current reverses the direction of the magnetization on the recording medium

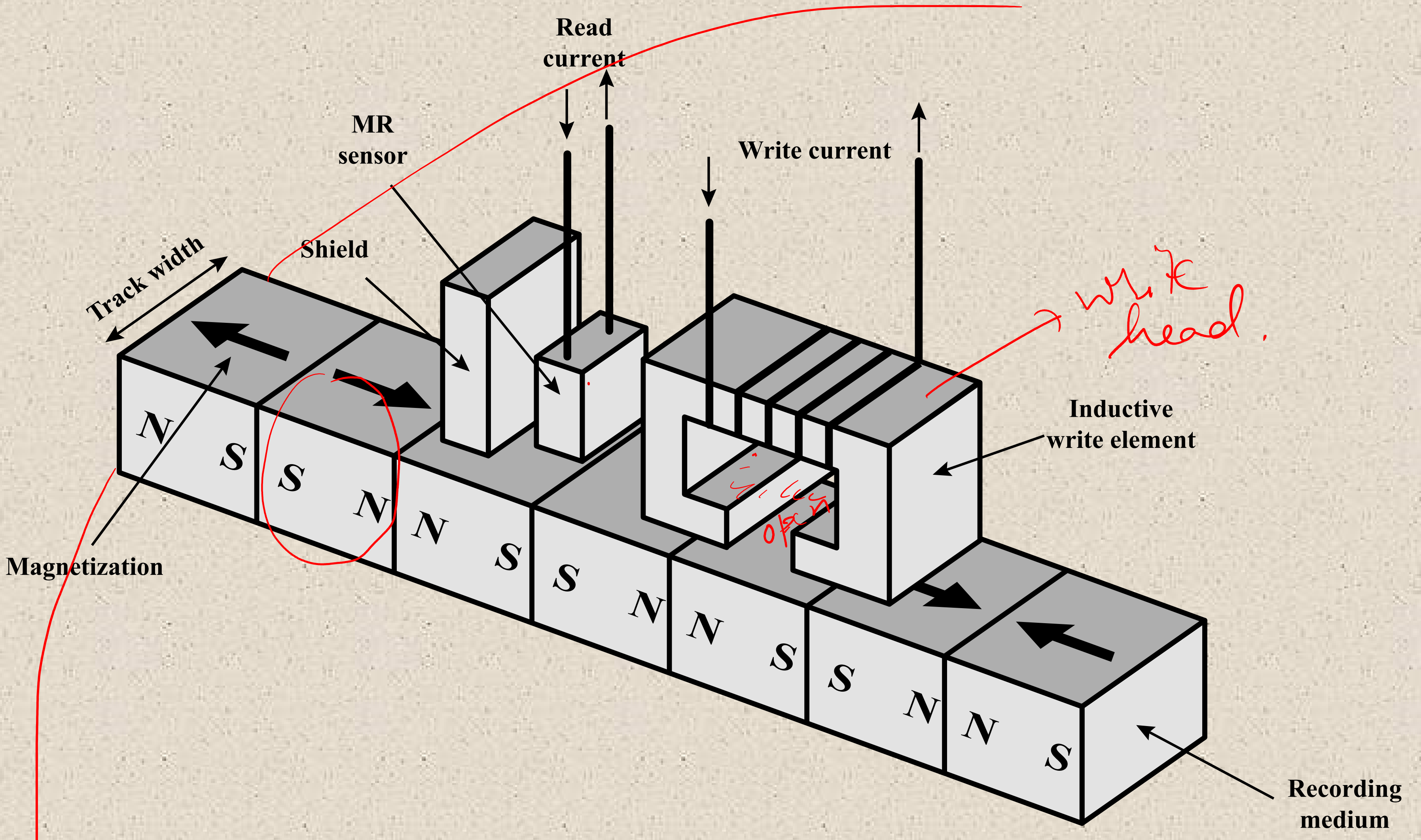


Figure 6.1 Inductive Write/Magnetoresistive Read Head

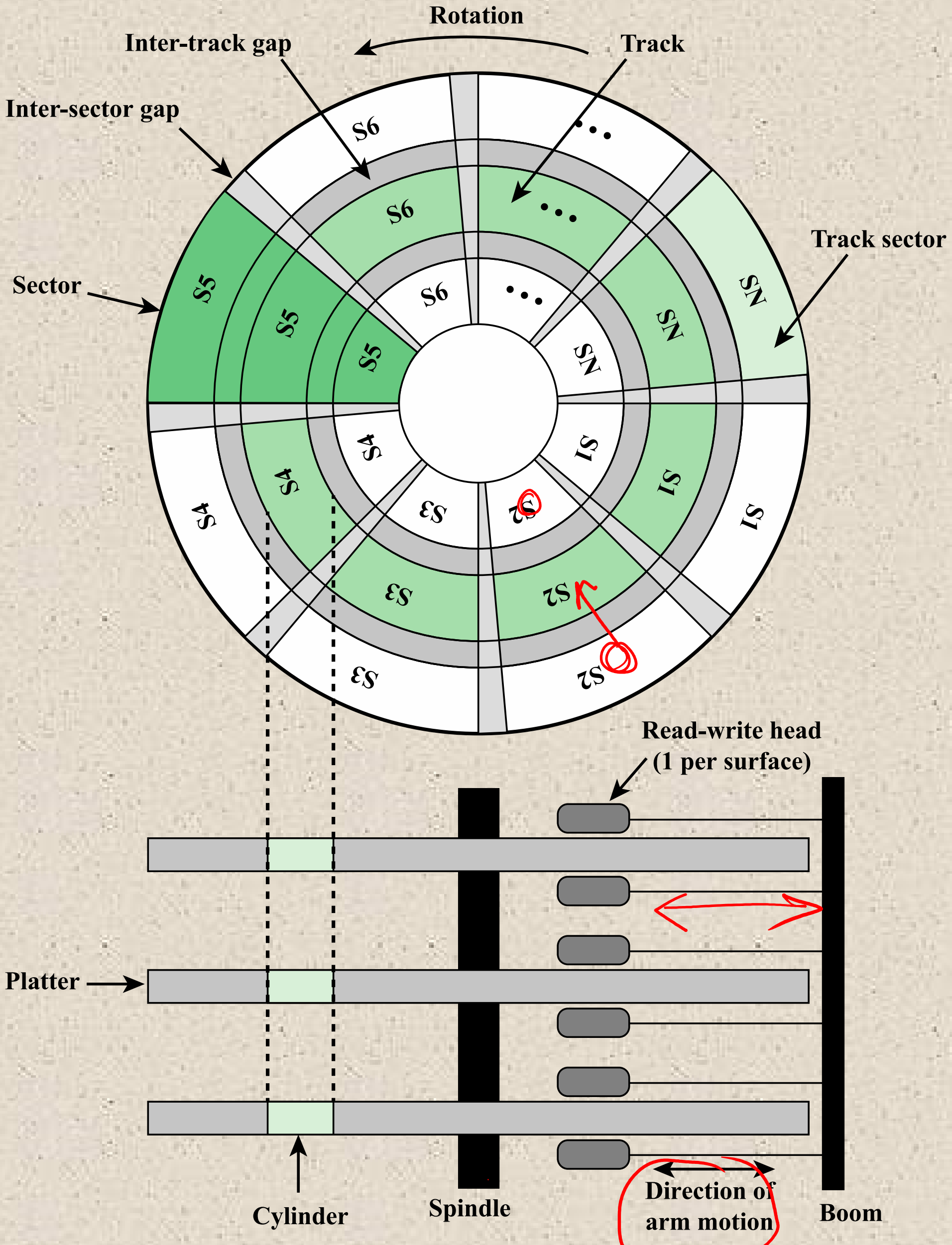
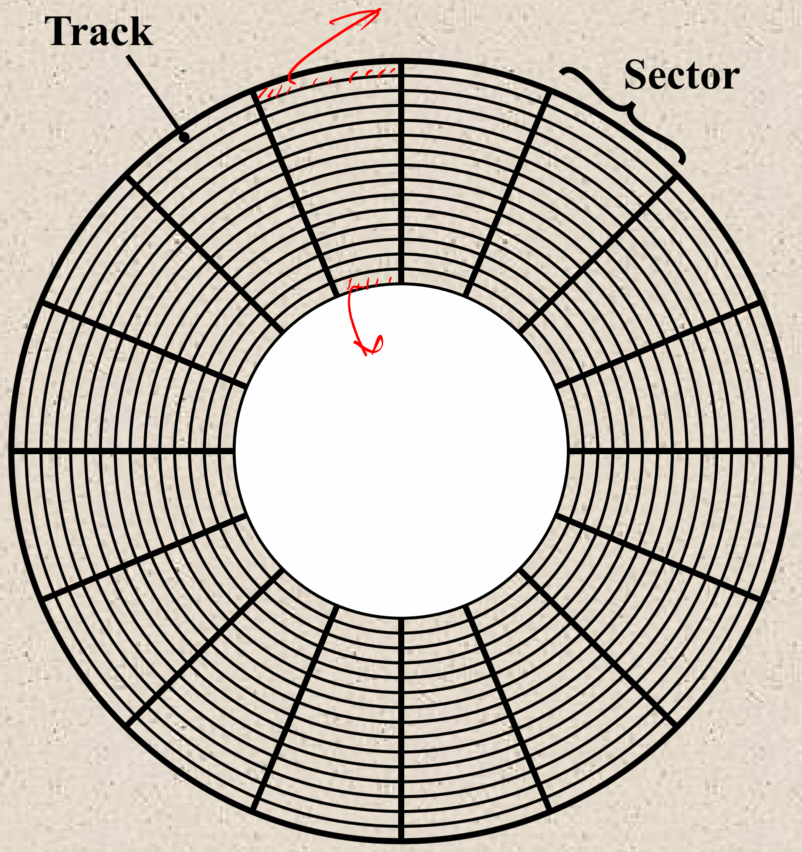


Figure 6.2 Disk Data Layout



(a) ~~Constant~~ angular velocity

(b) Multiple zone recording

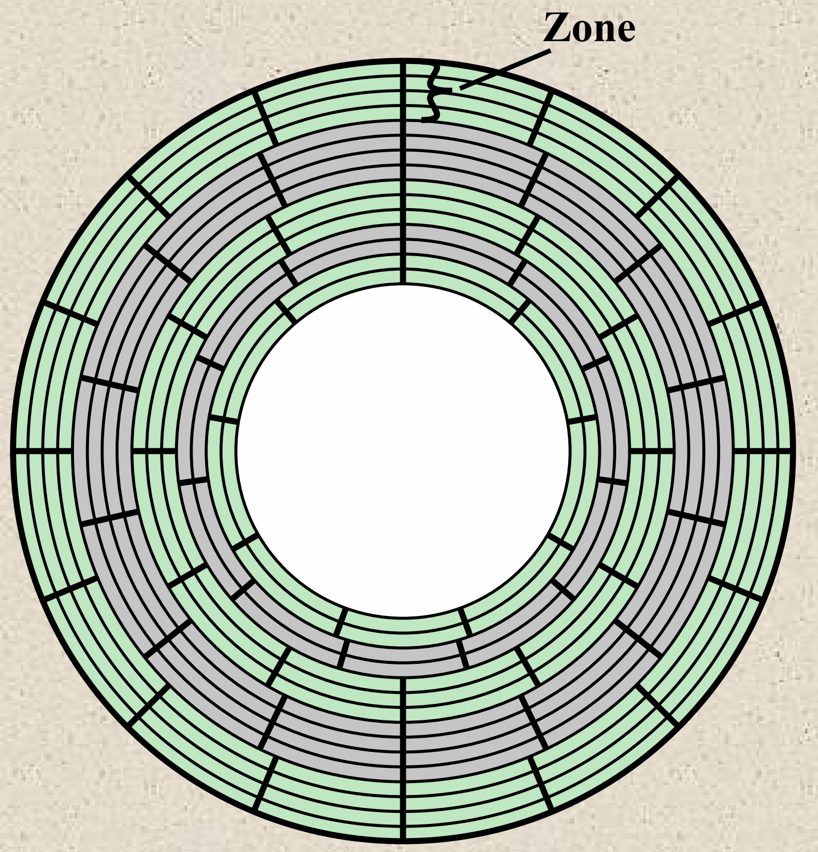


Figure 6.3 Comparison of Disk Layout Methods

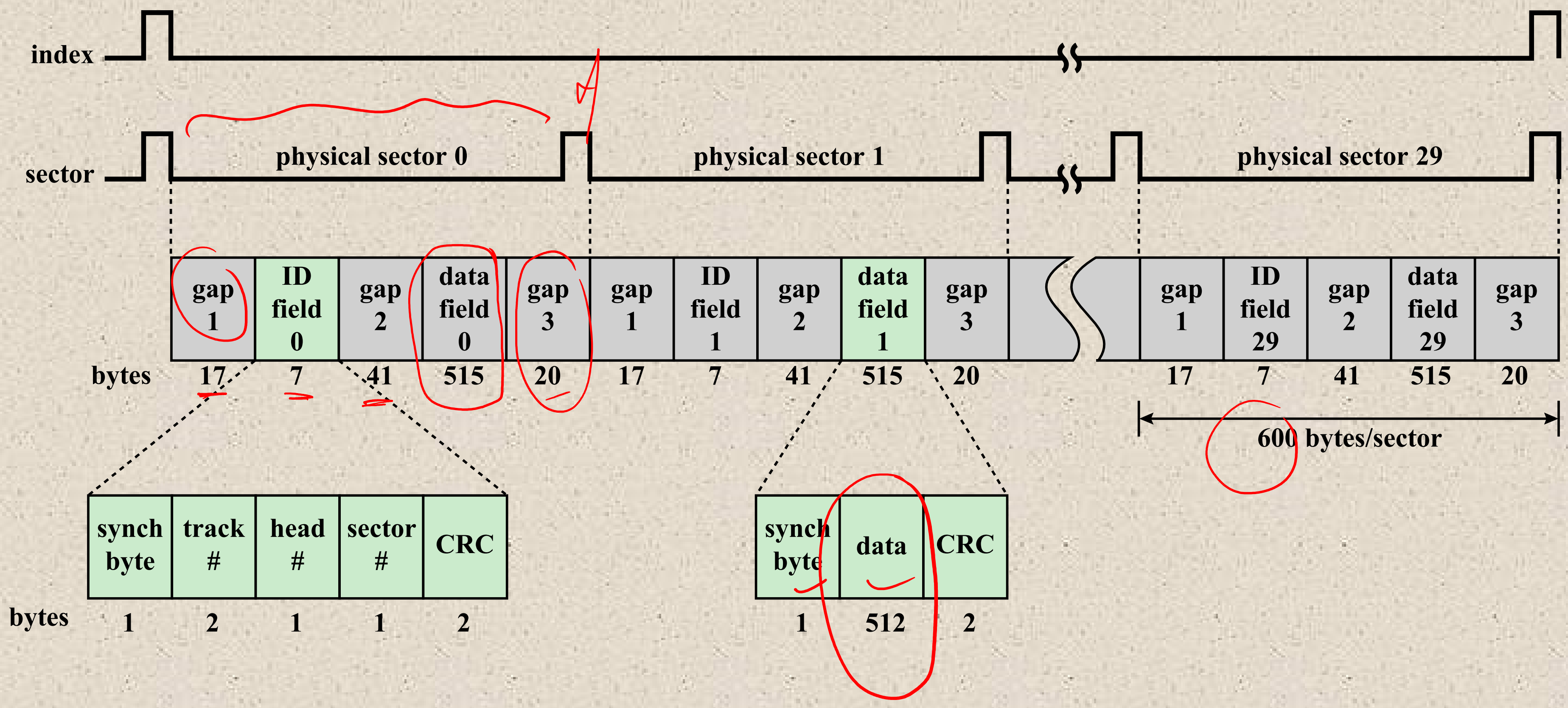


Figure 6.4 Winchester Disk Format (Seagate ST506)

Head Motion	**Platters**
Fixed head (one per track)	Single platter
Movable head (one per surface)	Multiple platter
Disk Portability	**Head Mechanism**
Nonremovable disk	Contact (floppy)
Removable disk	Fixed gap
	Aerodynamic gap (Winchester)
Sides	
Single sided	
Double sided	

Table 6.1
Physical Characteristics of Disk Systems

Characteristics

- Fixed-head disk
  - One read-write head per track
  - Heads are mounted on a fixed ridged arm that extends across all tracks
- Movable-head disk
  - One read-write head
  - Head is mounted on an arm
  - The arm can be extended or retracted
- Non-removable disk
  - Permanently mounted in the disk drive
  - The hard disk in a personal computer is a non-removable disk
- Removable disk
  - Can be removed and replaced with another disk
  - Advantages:
    - Unlimited amounts of data are available with a limited number of disk systems
    - A disk may be moved from one computer system to another
  - Floppy disks and ZIP cartridge disks are examples of removable disks
- Double sided disk
  - Magnetizable coating is applied to both sides of the platter

Disk Classification

The head mechanism provides a classification of disks into three types

- The head must generate or sense an electromagnetic field of sufficient magnitude to write and read properly
- The narrower the head, the closer it must be to the platter surface to function
  - A narrower head means narrower tracks and therefore greater data density
- The closer the head is to the disk the greater the risk of error from impurities or imperfections

Winchester Heads

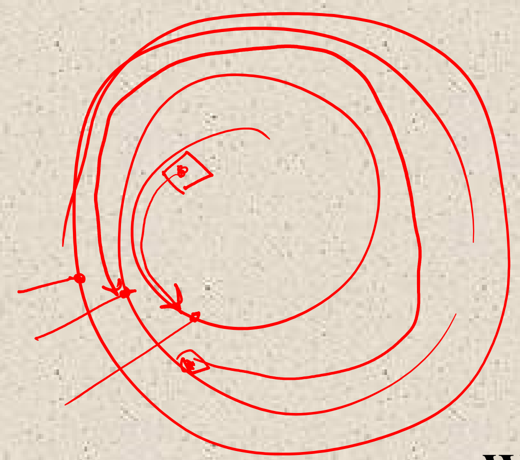
- Used in sealed drive assemblies that are almost free of contaminants
- Designed to operate closer to the disk's surface than conventional rigid disk heads, thus allowing greater data density
- Is actually an aerodynamic foil that rests lightly on the platter's surface when the disk is motionless
  - The air pressure generated by a spinning disk is enough to make the foil rise above the surface

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/s	149 MB/s	97 MB/s	300 MB/s
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

Disk Performance Parameters

- When the disk drive is operating the disk is rotating at constant speed
- To read or write the head must be positioned at the desired track and at the beginning of the desired sector on the track
  - Track selection involves moving the head in a movable-head system or electronically selecting one head on a fixed-head system
  - Once the track is selected, the disk controller waits until the appropriate sector rotates to line up with the head
- Seek time
  - On a movable–head system, the time it takes to position the head at the track
- Rotational delay *(rotational latency)*
  - The time it takes for the beginning of the sector to reach the head
- Access time
  - The sum of the seek time and the rotational delay
  - The time it takes to get into position to read or write
- Transfer time
  - Once the head is in position, the read or write operation is then performed as the sector moves under the head
  - This is the data transfer portion of the operation



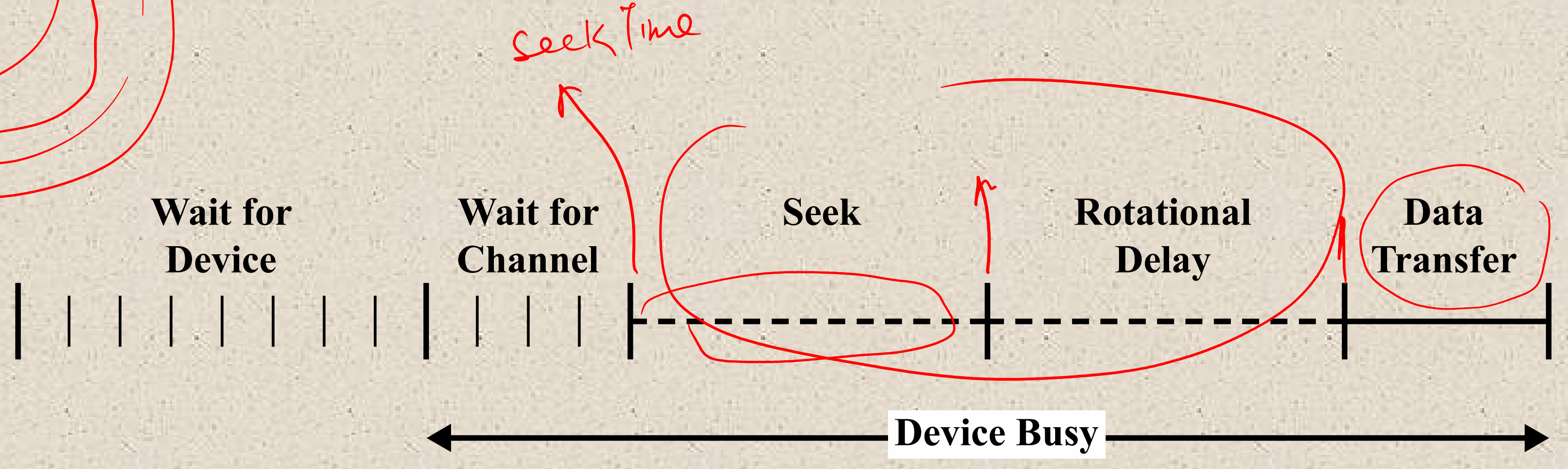


Figure 6.5 Timing of a Disk I/O Transfer

Avg Time to Read/ write = Avg Seek Time + Avg Rotation delay + Effective Transfer Time

RAID

Redundant Array of Independent Disks

- Consists of 7 levels
- Levels do not imply a hierarchical relationship but designate different design architectures that share three common characteristics:

1) Set of physical disk drives viewed by the operating system as a single logical drive
2) Data are distributed across the physical drives of an array in a scheme known as striping
3) Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure

RAID Levels

Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code $m = \log N$	8 $N + m$ 3	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N + 2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

N = number of data disks; m proportional to $\log N$

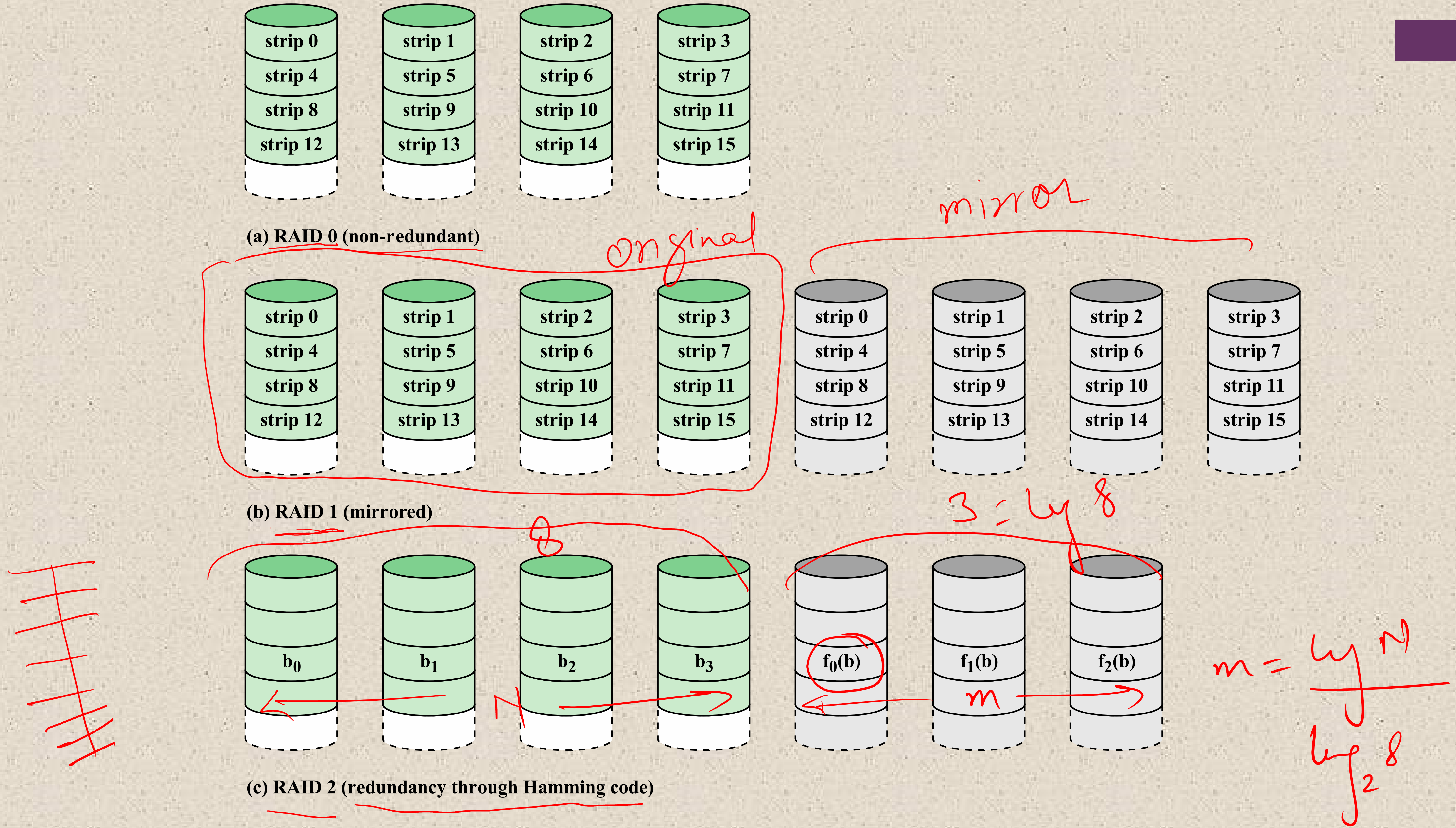


Figure 6.6 RAID Levels (page 1 of 2)

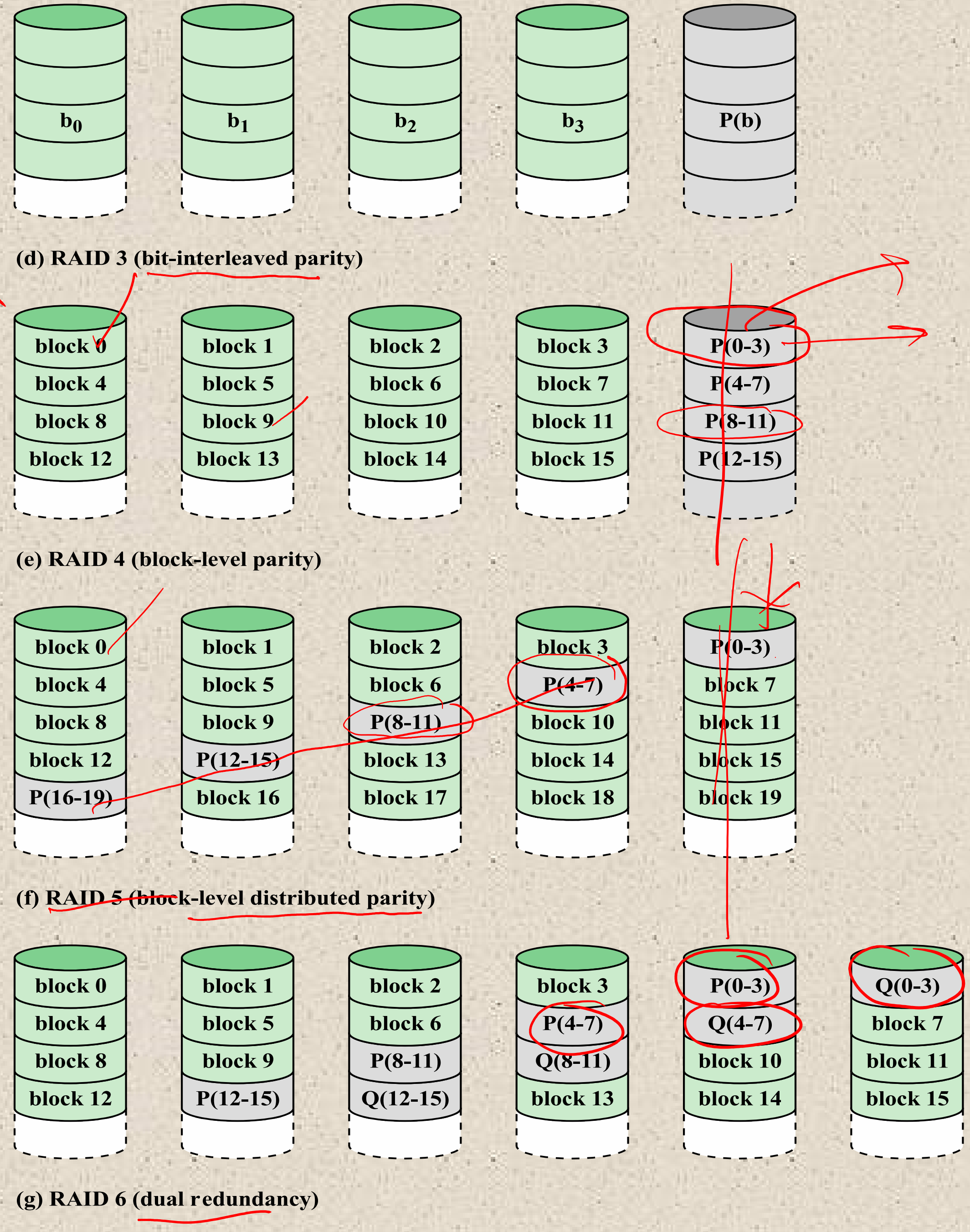


Figure 6.6 RAID Levels (page 2 of 2)

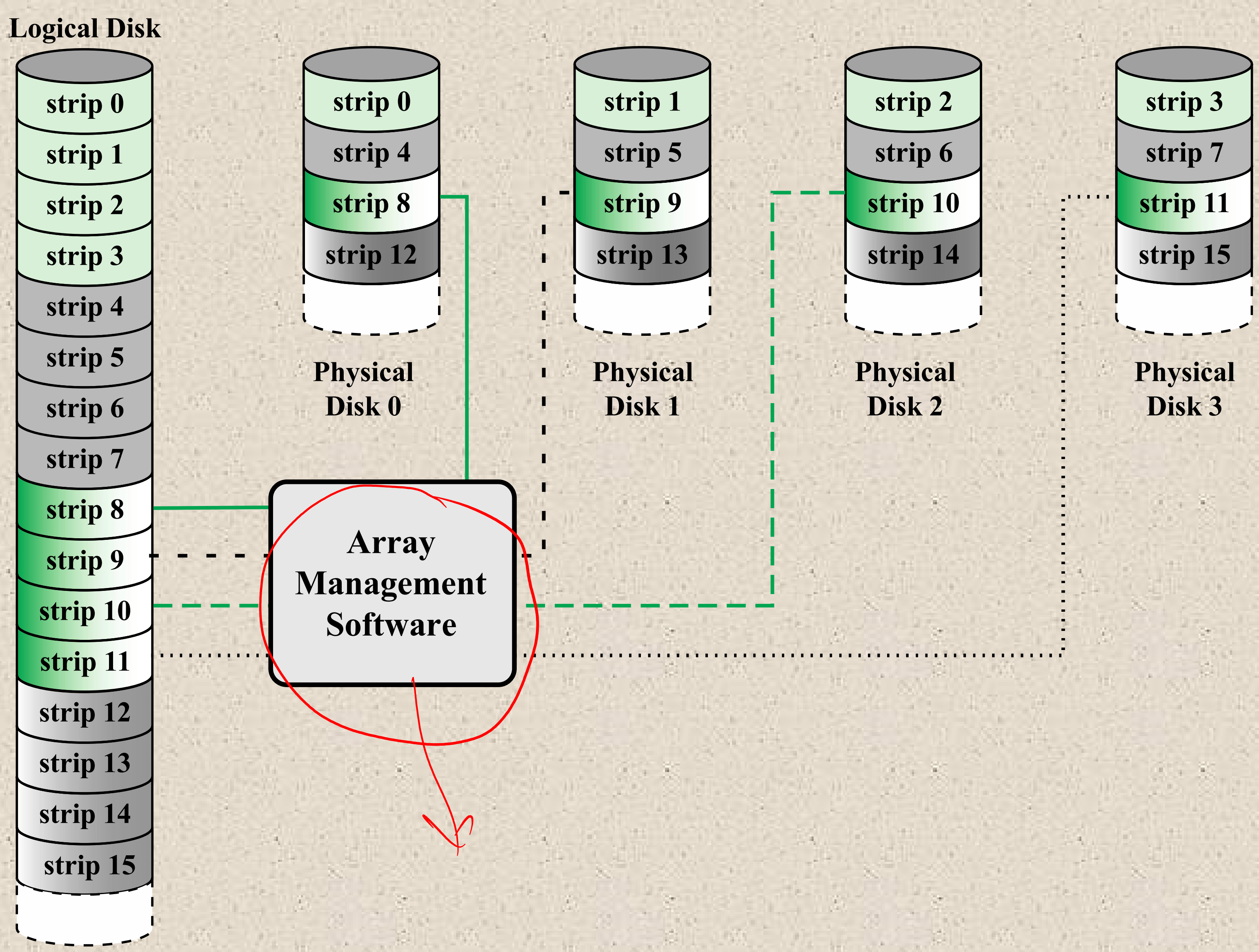


Figure 6.7 Data Mapping for a RAID Level 0 Array

RAID Comparison (page 1 of 2)

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
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

RAID Comparison (page 2 of 2)

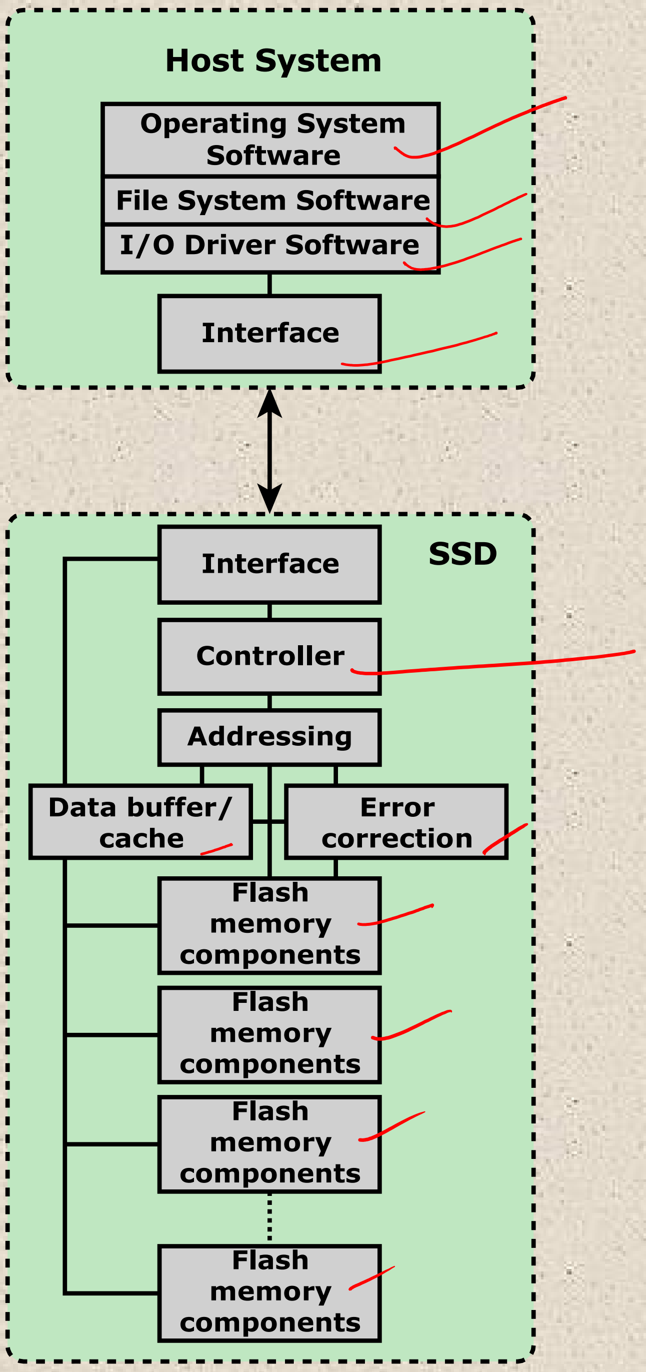
3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput
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

SSD Compared to HDD

- SSDs have the following advantages over HDDs:
- High-performance input/output operations per second (IOPS)
- Durability
- Longer lifespan
- Lower power consumption
- Quieter and cooler running capabilities
- Lower access times and latency rates

	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 maximum 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

Comparison of Solid State Drives and Disk Drives


Host System
Operating System Software
File System Software
I/O Driver Software
Interface
SSD
Interface
Controller
Addressing
Data buffer/ cache
Error correction
Flash memory components
Flash memory components
Flash memory components
Flash memory components

+

Practical Issues

There are two practical issues peculiar to SSDs that are not faced by HDDs:

- SDD performance has a tendency to slow down as the device is used
  - The entire block must be read from the flash memory and placed in a RAM buffer
  - Before the block can be written back to flash memory, the entire block of flash memory must be erased
  - The entire block from the buffer is now written back to the flash memory

- Flash memory becomes unusable after a certain number of writes
  - Techniques for prolonging life:
    - Front-ending the flash with a cache to delay and group write operations
    - Using wear-leveling algorithms that evenly distribute writes across block of cells
    - Bad-block management techniques
  - Most flash devices estimate their own remaining lifetimes so systems can anticipate failure and take preemptive action

Table 6. 6 Optical Disk Products

CD

Compact Disk. A nonerasable disk that stores digitized audio information. The standard system uses 12-cm disks and can record more than 60 minutes of uninterrupted playing time.

CD-ROM

Compact Disk Read-Only Memory. A nonerasable disk used for storing computer data. The standard system uses 12-cm disks and can hold more than 650 Mbytes.

CD-R

CD Recordable. Similar to a CD-ROM. The user can write to the disk only once.

CD-RW

CD Rewritable. Similar to a CD-ROM. The user can erase and rewrite to the disk multiple times.

DVD

Digital Versatile Disk. A technology for producing digitized, compressed representation of video information, as well as large volumes of other digital data. Both 8 and 12 cm diameters are used, with a double-sided capacity of up to 17 Gbytes. The basic DVD is read-only (DVD-ROM).

DVD-R

DVD Recordable. Similar to a DVD-ROM. The user can write to the disk only once. Only one-sided disks can be used.

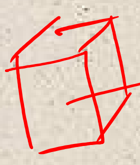
DVD-RW

DVD Rewritable. Similar to a DVD-ROM. The user can erase and rewrite to the disk multiple times. Only one-sided disks can be used.

Blu-Ray DVD

High definition video disk. Provides considerably greater data storage density than DVD, using a 405-nm (blue-violet) laser. A single layer on a single side can store 25 Gbytes.

Holographic Memory



→ 1TB

+

Magnetic Tape

- Tape systems use the same reading and recording techniques as disk systems
- Medium is flexible polyester tape coated with magnetizable material
- Coating may consist of particles of pure metal in special binders or vapor-plated metal films
- Data on the tape are structured as a number of parallel tracks running lengthwise
- Serial recording
  - Data are laid out as a sequence of bits along each track
- Data are read and written in contiguous blocks called *physical records*
- Blocks on the tape are separated by gaps referred to as *inter-record gaps*

Gate CSE 2015 set 2

Consider a typical disk that rotates at 15000 rotations per minute (RPM) and has a transfer rate of 50×10^6 $bytes/\mathrm{sec}$. If the average seek time of the disk is twice the average rotational delay and the controller's transfer time is 10 times the disk transfer time, the average time (in milliseconds) to read or write a 512-byte sector of the disk is ______________________.

avg time = Avg seek time + Avg Rotational delay + Effective Tranfer time.

ST = 2 · D Rot delay

Controller's Transfer time = 10 × disk transfer time

Rotation / minute = 15000

Time Rotation / ~~sec~~ (delay) = $\frac{60}{15000}$ = 4 ms

Avg rotational delay = $\frac{1}{2}$ Rotational delay = 2 ms

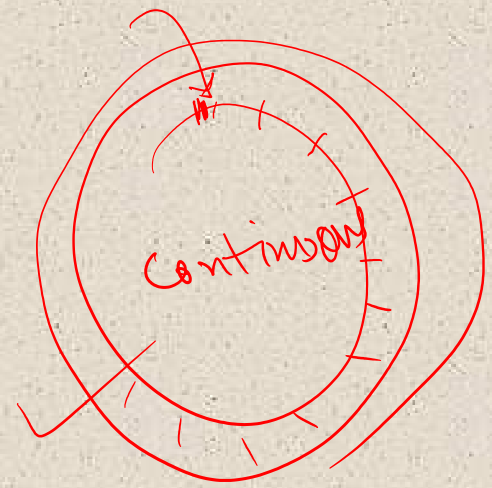
Avg seek Time = 2 × Avg Rotationdelay = 4ms

4 ms + 2 ms + $\frac{512}{50\times10^6}$ × 10 → 0.0102 ms

$\boxed{6.102 \text{ ms}}$

Gate CSE 2015 set 1

Consider a disk pack with a seek time of 4 milliseconds and rotational speed of 10000 rotations per minute (RPM). It has 600 sectors per track and each sector can store 512 bytes of data. Consider a file stored in the disk. The file contains 2000 sectors. Assume that every sector access necessitates a seek, and the average rotational latency for accessing each sector is half of the time for one complete rotation. The total time (in milliseconds) needed to read the entire file is ____________.



Avg rotational delay $= \frac{0.006}{2} = \frac{6\text{ ms}}{2} = 3\text{ ms}$

for 2000 sector $= 2000 \times 3\text{ m} = 6000\text{ ms}$

Avg seek time $= 2000 \times 4\text{ ms} = 8000\text{ ms}$

Data transfer rate in 1 rotation $= \frac{600 \times 512}{6\text{ ms}} = 51200$ Byte/ms.

Transfer rate for 512 Byte $= \frac{512}{51200} = .01\text{ ms} = .01\text{ ms}$

Transfer rate for 2000 sector $= .01 \times 2000 = 20\text{ ms}$ $\times 2000$

14020 ms