CS / EE 320
Computer Organization and Assembly Language
Spring 2024
Lecture 26

Shahid Masud

Topics: External Storage, Magnetic Disk, Optical Disk, RAID Array Storage

Topics

- Optical Storage
  - CDR
  - DVDR
- Magnetic Disk Storage
- RAID Arrays

QUIZ IN LAST LECTURE

MON 29 APRIL

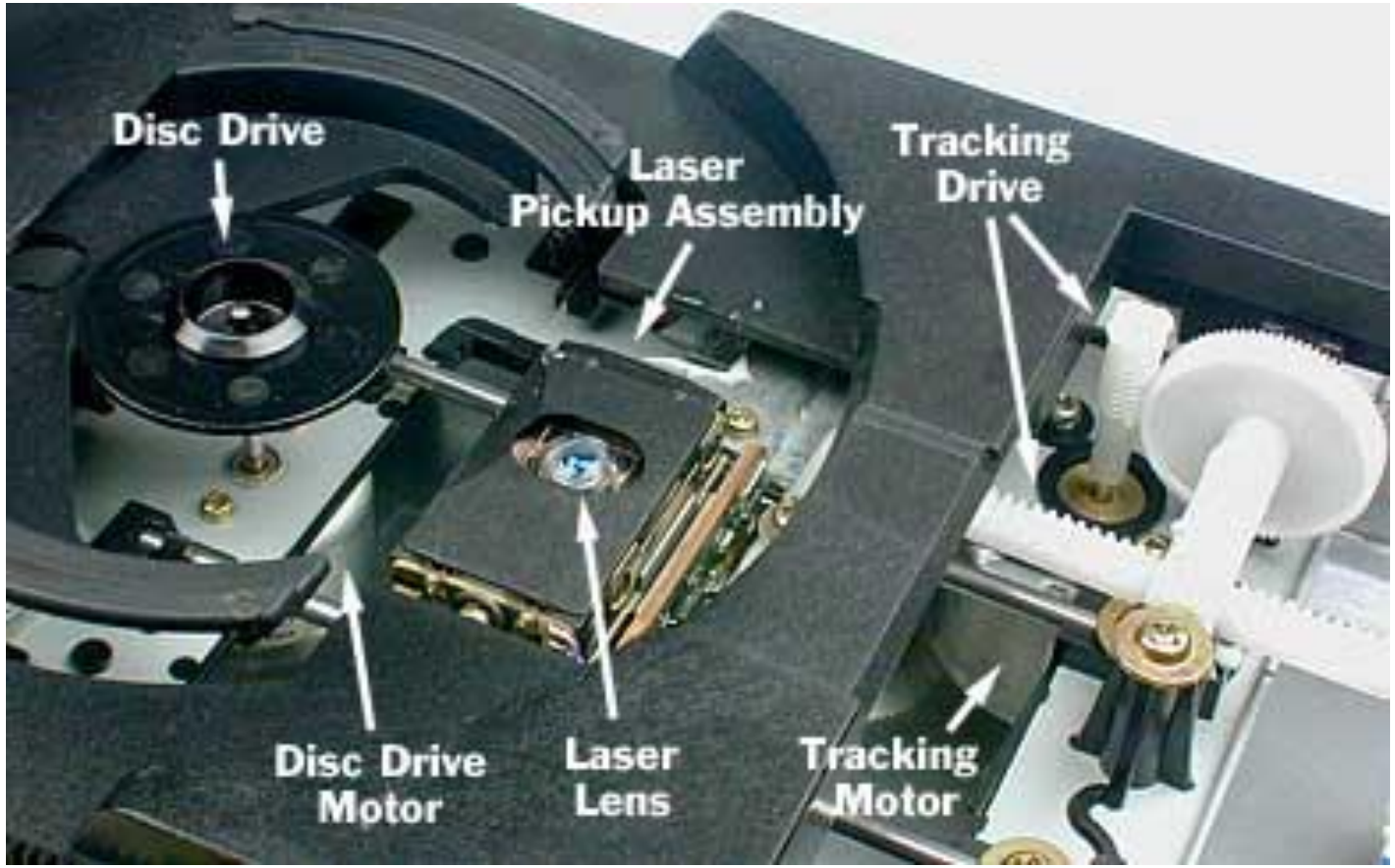
Optical Drives CD / DVD

Optical Storage CD-ROM

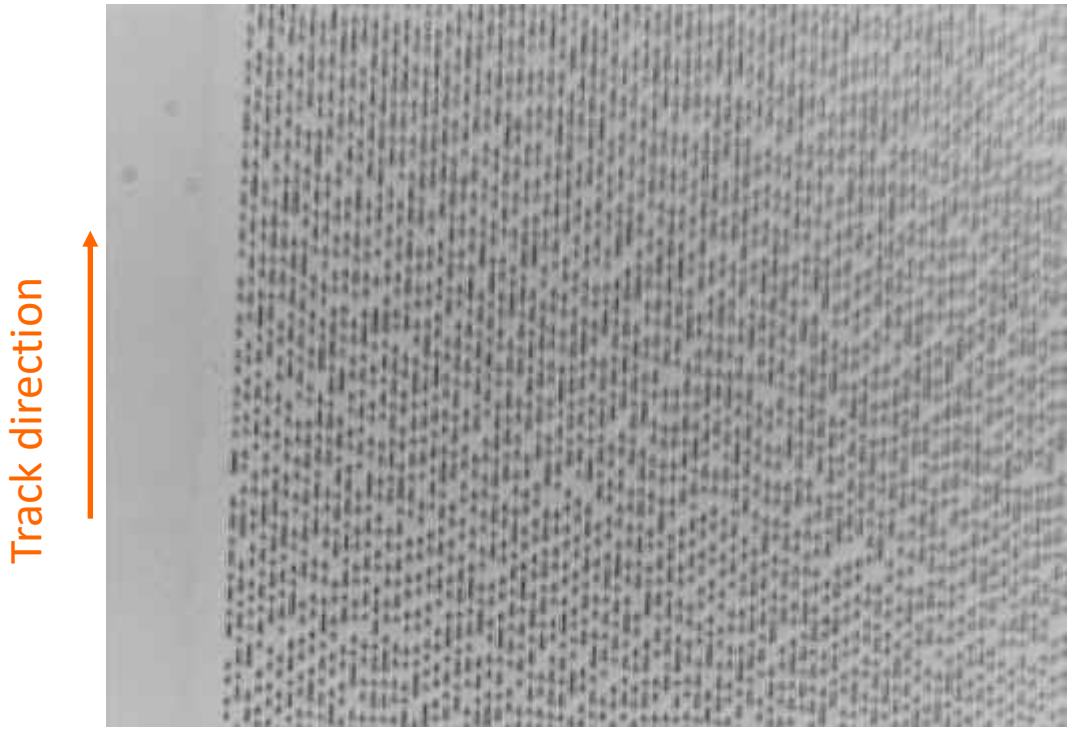
- Originally for audio
- 650Mbytes giving over 70 minutes audio
- Polycarbonate coated with highly reflective coat, usually aluminum
- Data stored as pits
- Read by reflecting laser
- Constant packing density
- Constant linear velocity

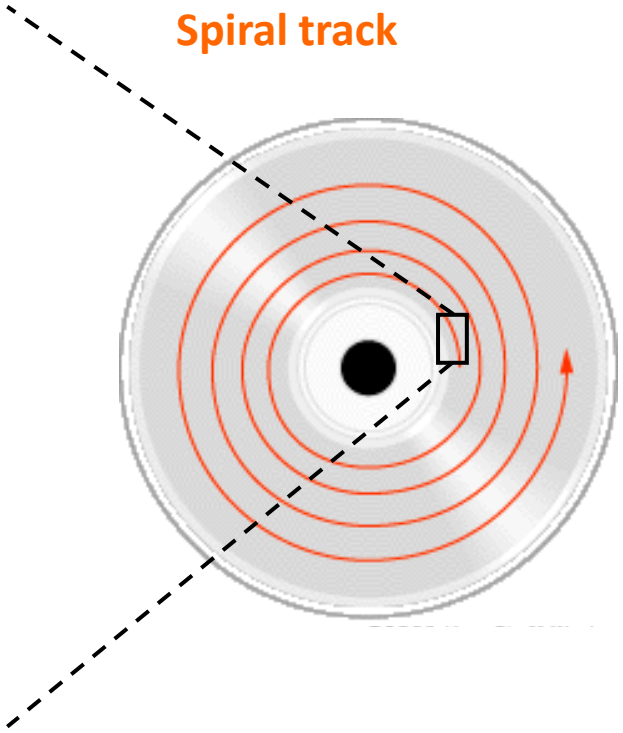
CD Components

CD Components



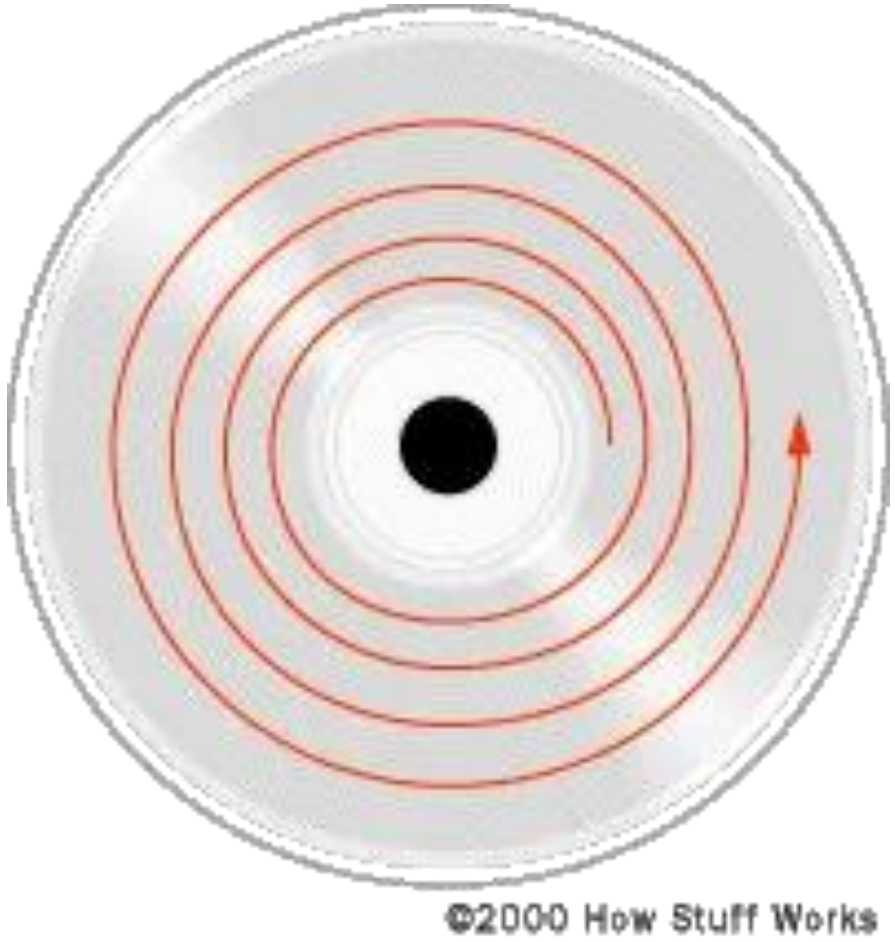
CD Surface





Low-magnification (×32) image of a CD showing an edge of the data zone.

Sprial Track of CD



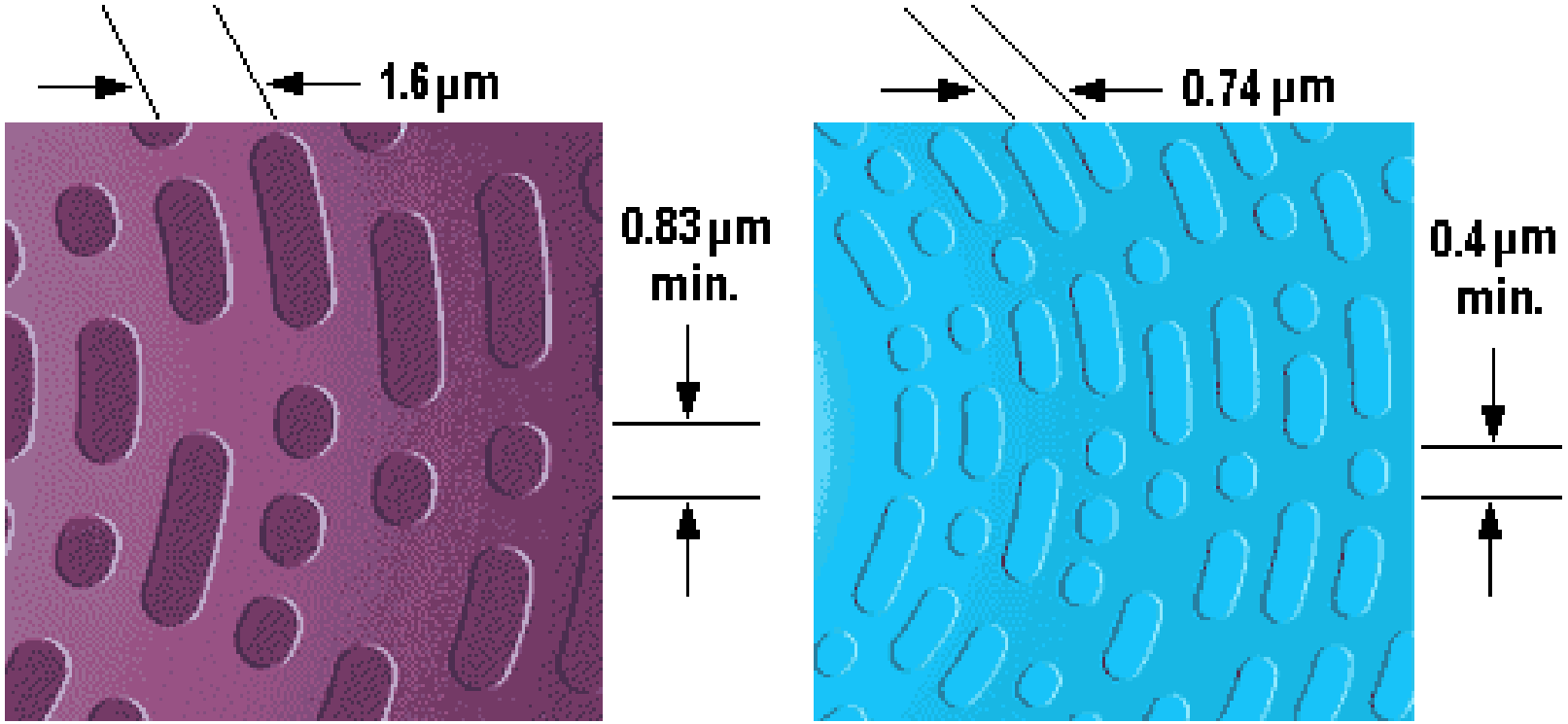
Track Density and Data Density

The CD is 12 cm in diameter, 1.2 mm thick, has a center hole 1.5 cm in diameter, and spins at a *constant linear velocity* (CLV) or *constant angular velocity* (CAV).

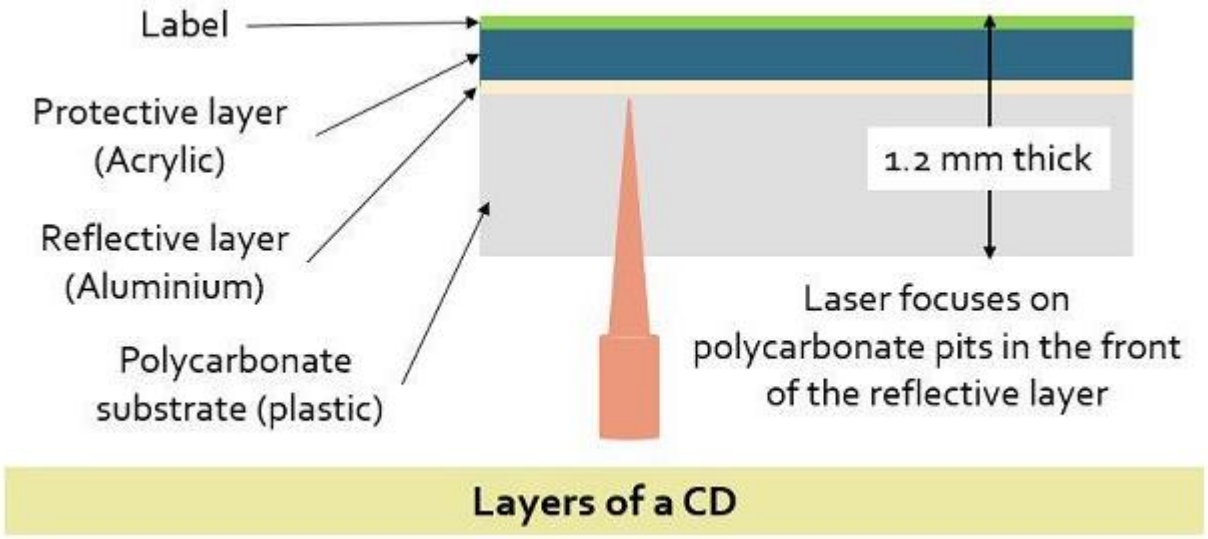
There is only one track on the optical disk and all data are stored in a spiral of about 2 billion small pits on the surface. There are about 30,000 windings on a CD - all part of the same track. This translates into about 16,000 tracks per inch and an areal density of 1 Mb/mm^2.

The total length of the track on a CD is almost 3 miles.

Track Density CD vs DVD



Surface Layers



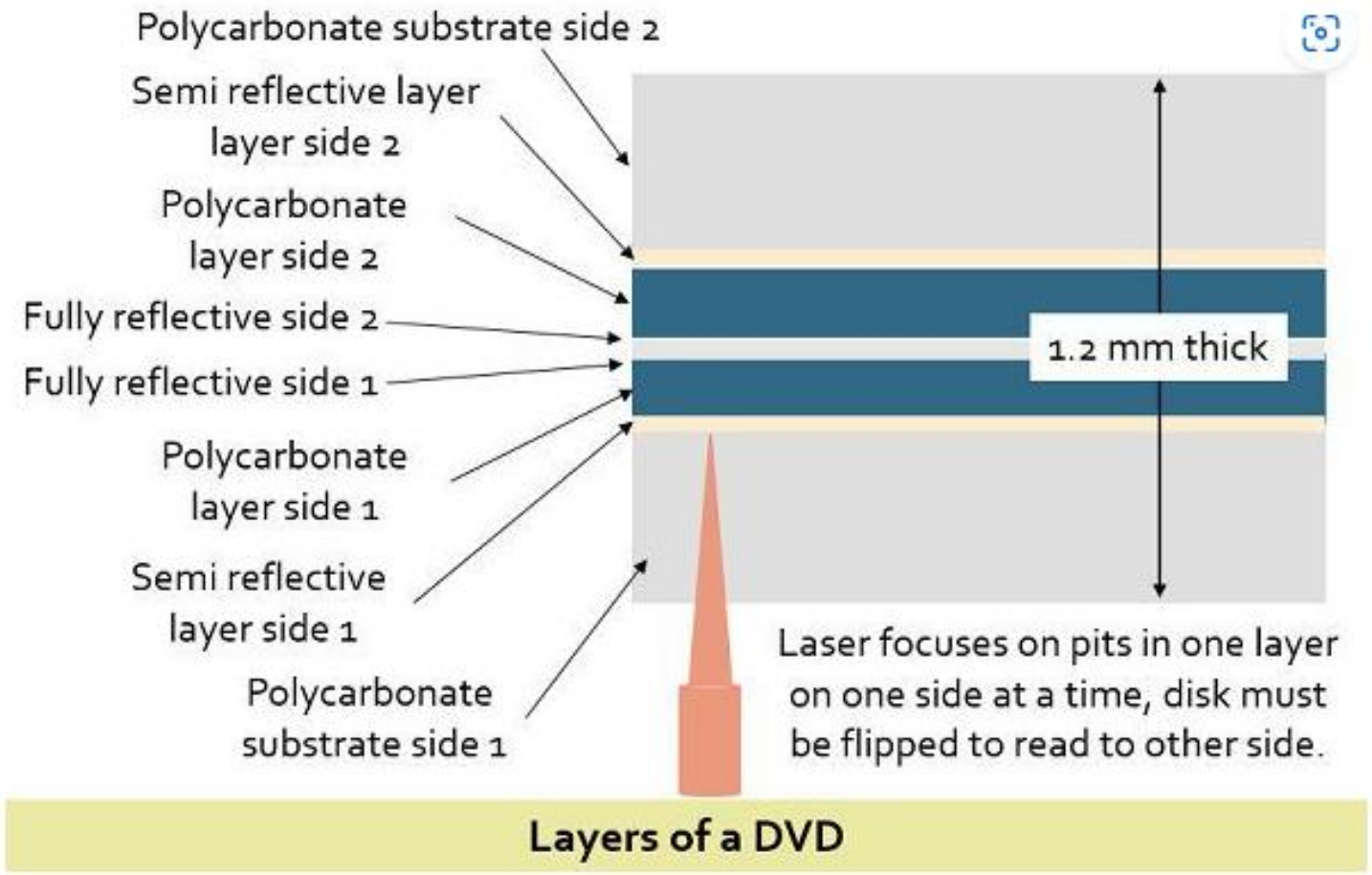
Layers of a CD

A CD can store up to 74 minutes of music, so the total amount of digital data that must be stored on a CD is:

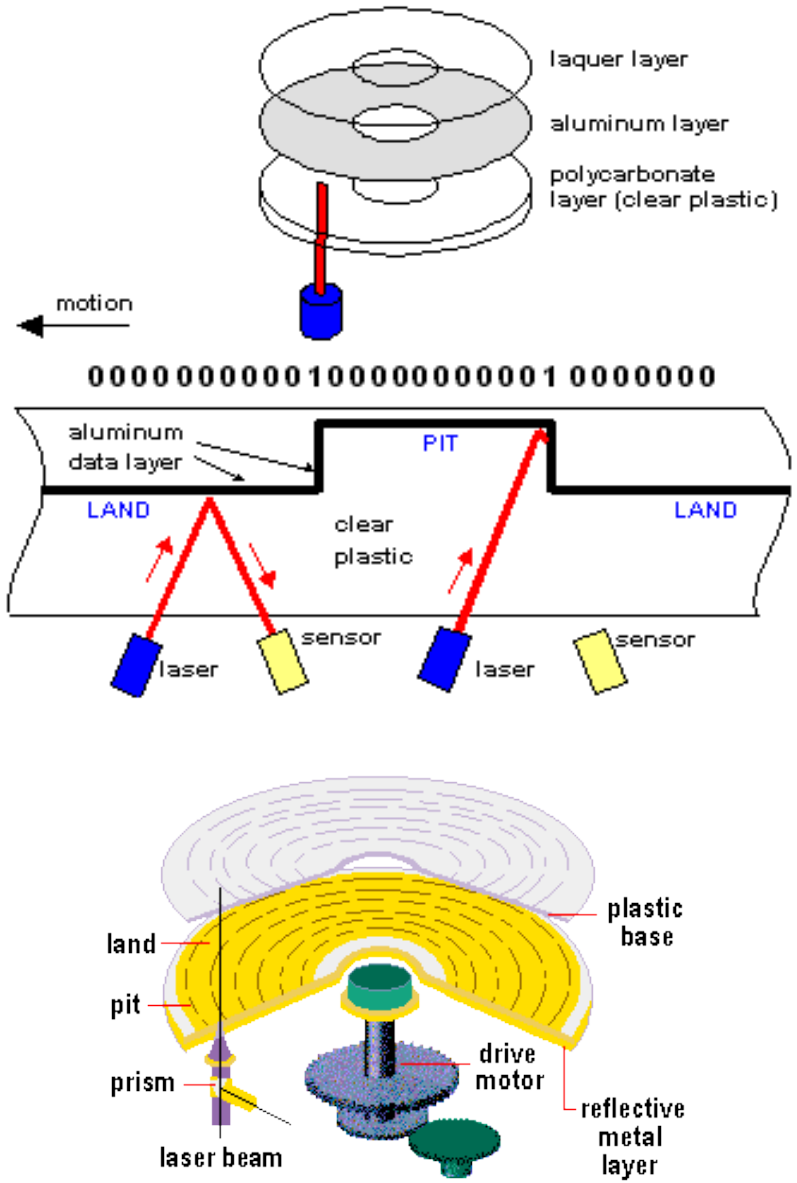
2 channels × 44,100 samples/channel/second × 2 bytes/sample × 74 minutes × 60 seconds/minute = 783,216,000 bytes

To fit more than 783 megabytes onto a disk only 12 cm in diameter requires that the individual bits be very small.

Layers of a DVD



CD Mechanism



Digital data are carved into the disc as pits (low spots) and lands (high spots). As the laser shines into the moving pits and lands, a sensor detects a change in reflection when it encounters a transition from pit to land or land to pit.
Each transition is a 1. The lack of transitions are 0s. There is only one laser in a drive. Two are used here to illustrate the difference in reflection.

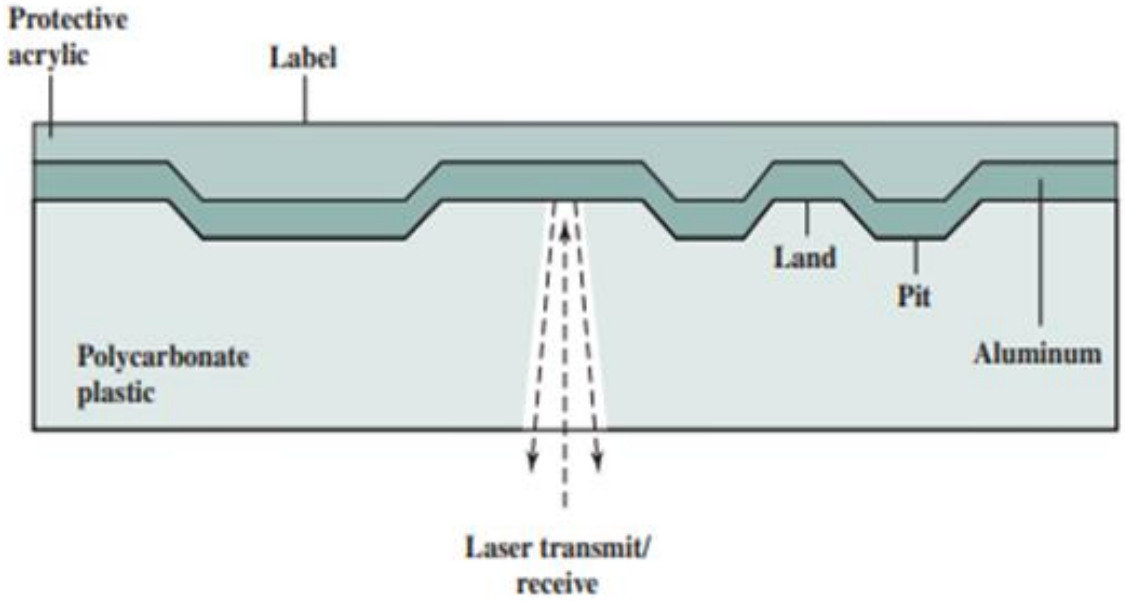
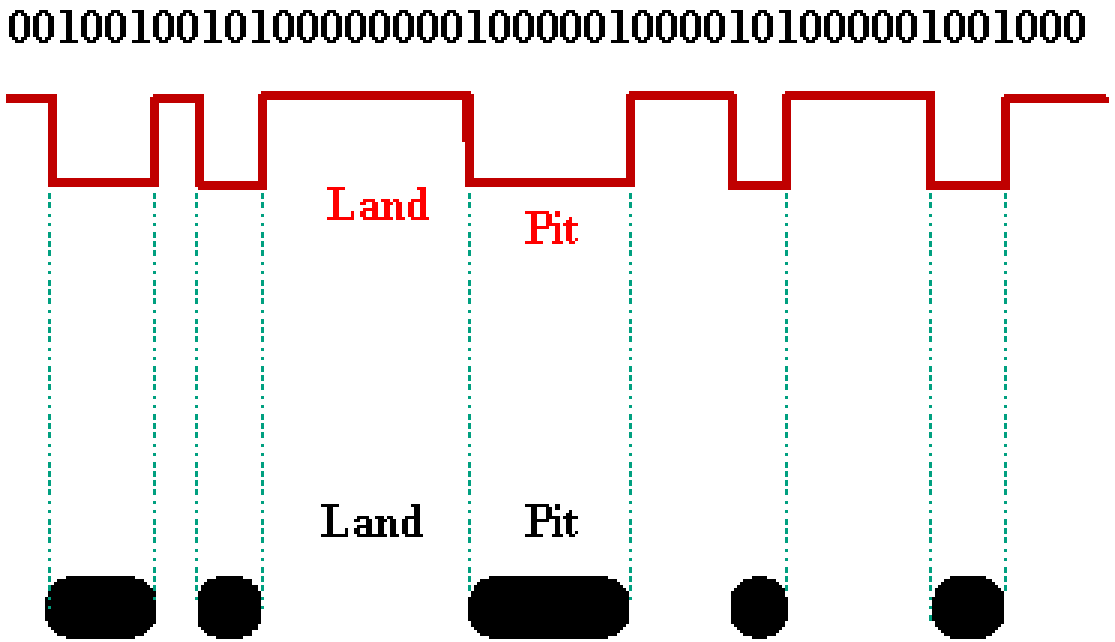


Figure 6.9 CD Operation

Data Format


PIT
LAND
PIT
LAND
PIT
1 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0


00100100101000000001000001000010100000 1001000
Land
Pit
Land
Pit

Storage Write and Read Principle

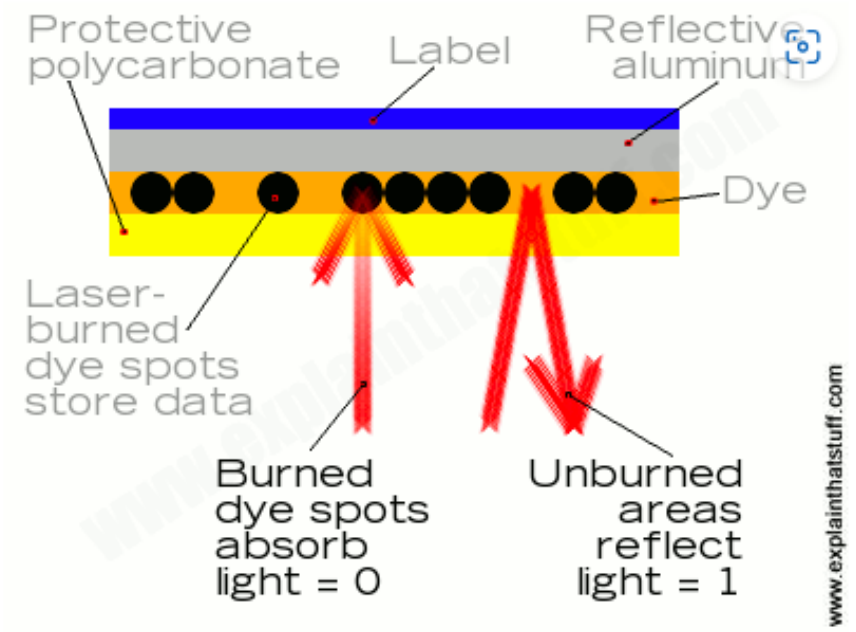
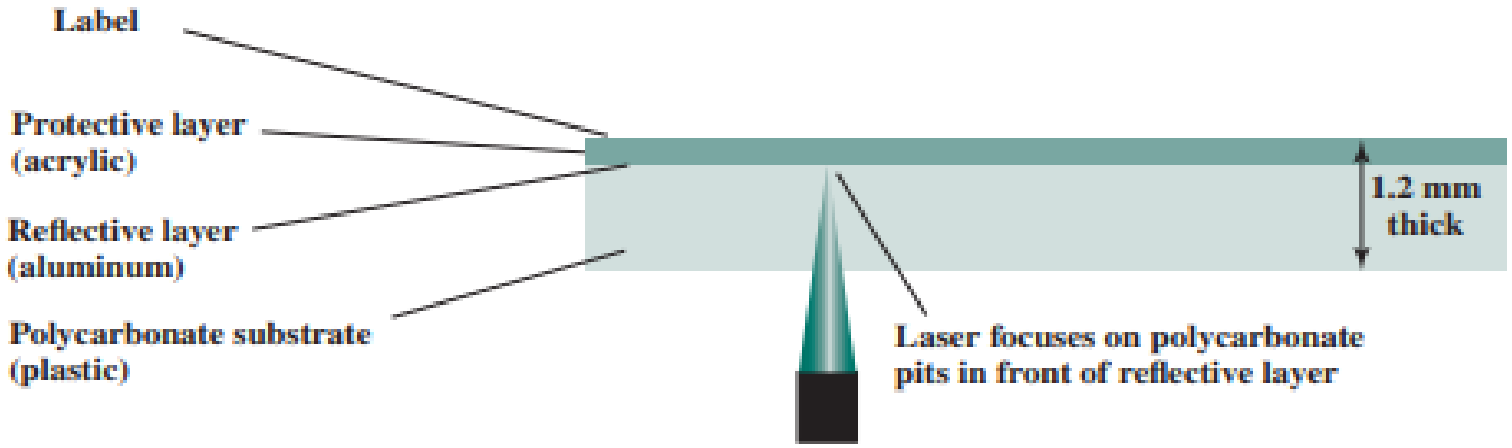
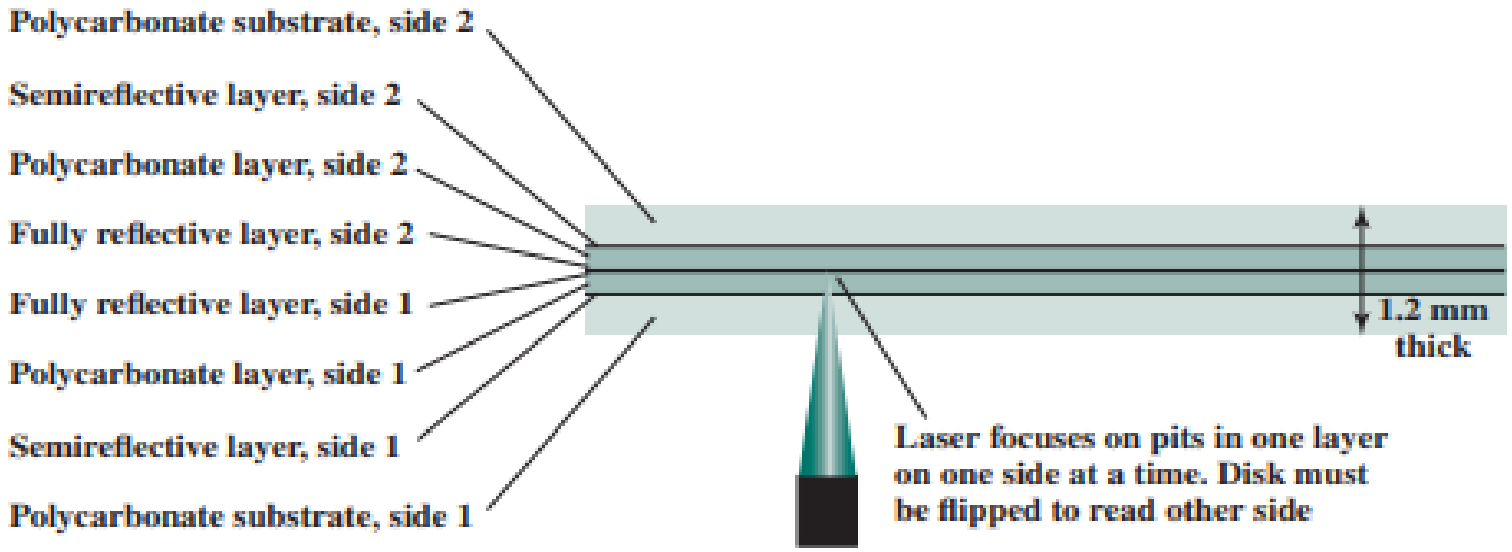


Illustration: With a CD-R, binary information is stored as "burned" areas (0) and unburned areas (1) in the dye layer sandwiched between the protective polycarbonate and the reflective aluminum.

CD and DVD Read Details



(a) CD-ROM–Capacity 682 MB

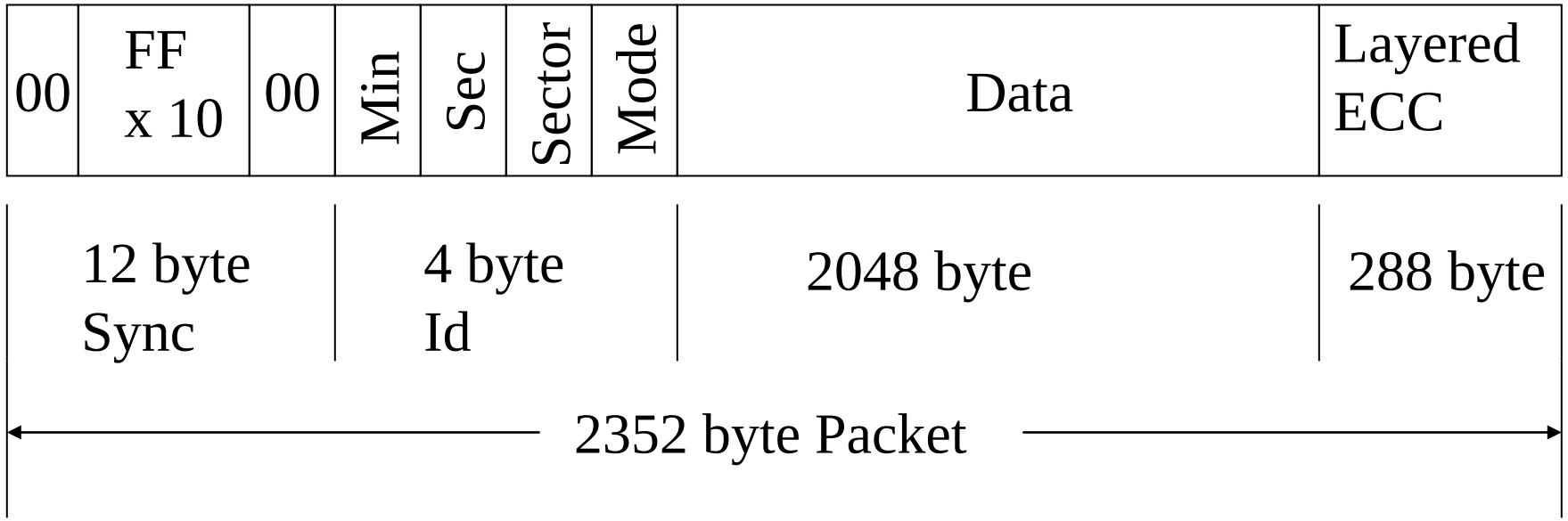


(b) DVD-ROM, double-sided, dual-layer–Capacity 17 GB

CD-ROM Drive Speeds

- Audio is single speed
  - Constant linear velocity
  - 1.2 ms^{-1}
  - Track (spiral) is 5.27km long
  - Gives 4391 seconds = 73.2 minutes
  - Date 176.4 K bytes/s total capacity 774.57 M Bytes
- Other speeds are quoted as multiples
- e.g. 24x ~= 4 M Bytes/s (data transfer rate)
- The quoted figure is the maximum the drive can achieve

CD-ROM Format Frame



- Mode 0=blank data field
- Mode 1=2048 byte data+error correction
- Mode 2=2336 byte data

Random Access on CD-ROM

- Difficult
- Move head to rough position
- Set correct speed
- Read address
- Adjust to required location

DVD

- Digital Video Disk
  - Used to indicate a player for movies
    - Only plays video disks
- Digital Versatile Disk
  - Used to indicate a computer drive
    - Will read computer disks and play video disks

DVD - technology

- Multi-layer
- Very high capacity (4.7G per layer)
- dual-layer hold 8.5 Gbytes ~> 4hr movie
- Full length movie on single disk
  - Using MPEG compression
- Finally standardized
- Movies carry regional coding
- Players only play correct region films

DVD types and Standards

- 74 min. capacity= (Beethoven's 9th)!
- 1x by definition
- 1x = 1.23 Mbit/s
- linear V = 1.2 m/s
- >12x = CAV
- 780 nm laser

- 700MB
- 1x-72x speeds

- 4.7 GB

- 1-4x Original
- 4-10x High speed
- 12-24x Ultra speed
- 32x Ultra Speed+

- 4.7 GB
- 18-20x
- 11 Mbit/s
- 650 nm laser

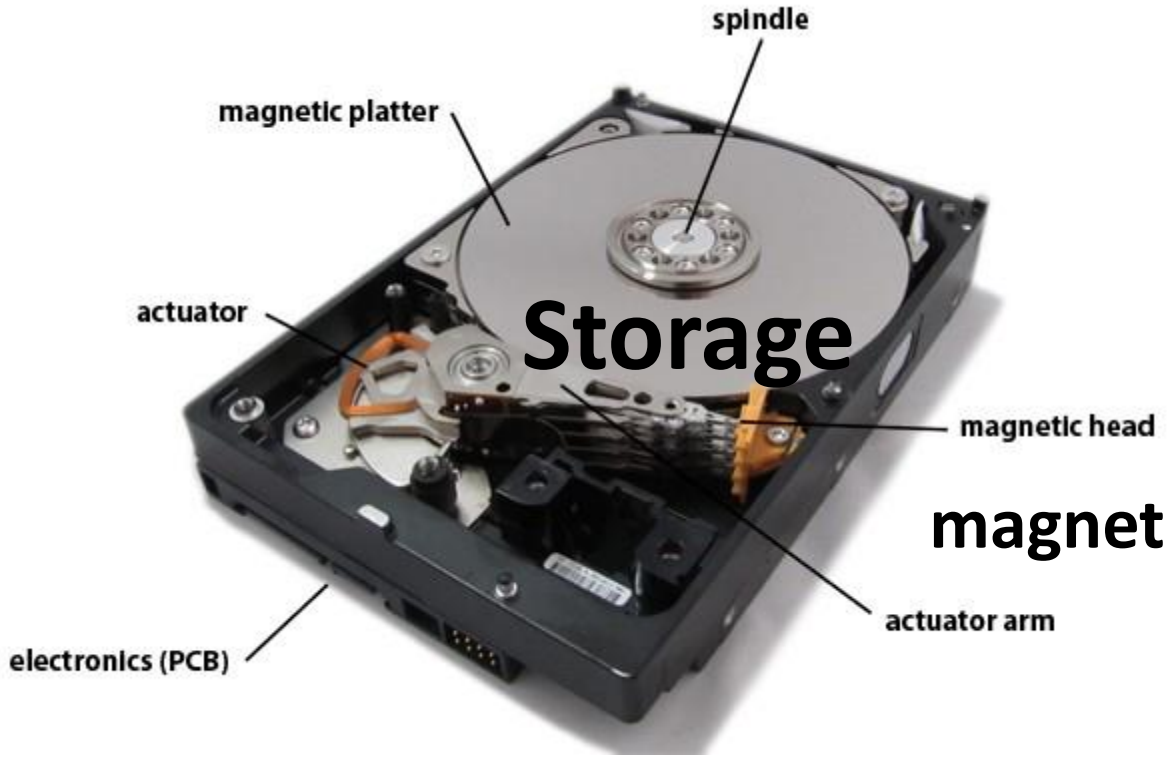
- 25 GB (1-layer)
- 50 GB (2-layers)
- 1x = 36 Mbits/s
- 1x–14x speeds
- 405 nm laser

Hard Disk Drive

External Memory – Secondary Storage Types

- Magnetic Disk
- RAID
- Optical Memory
- Magnetic Tape

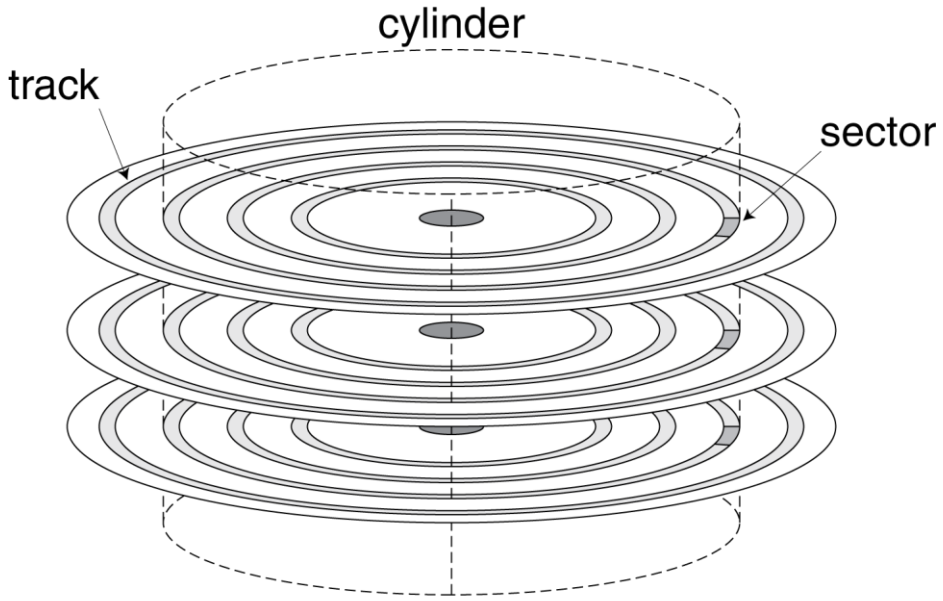
Hard Disk Drive Construction



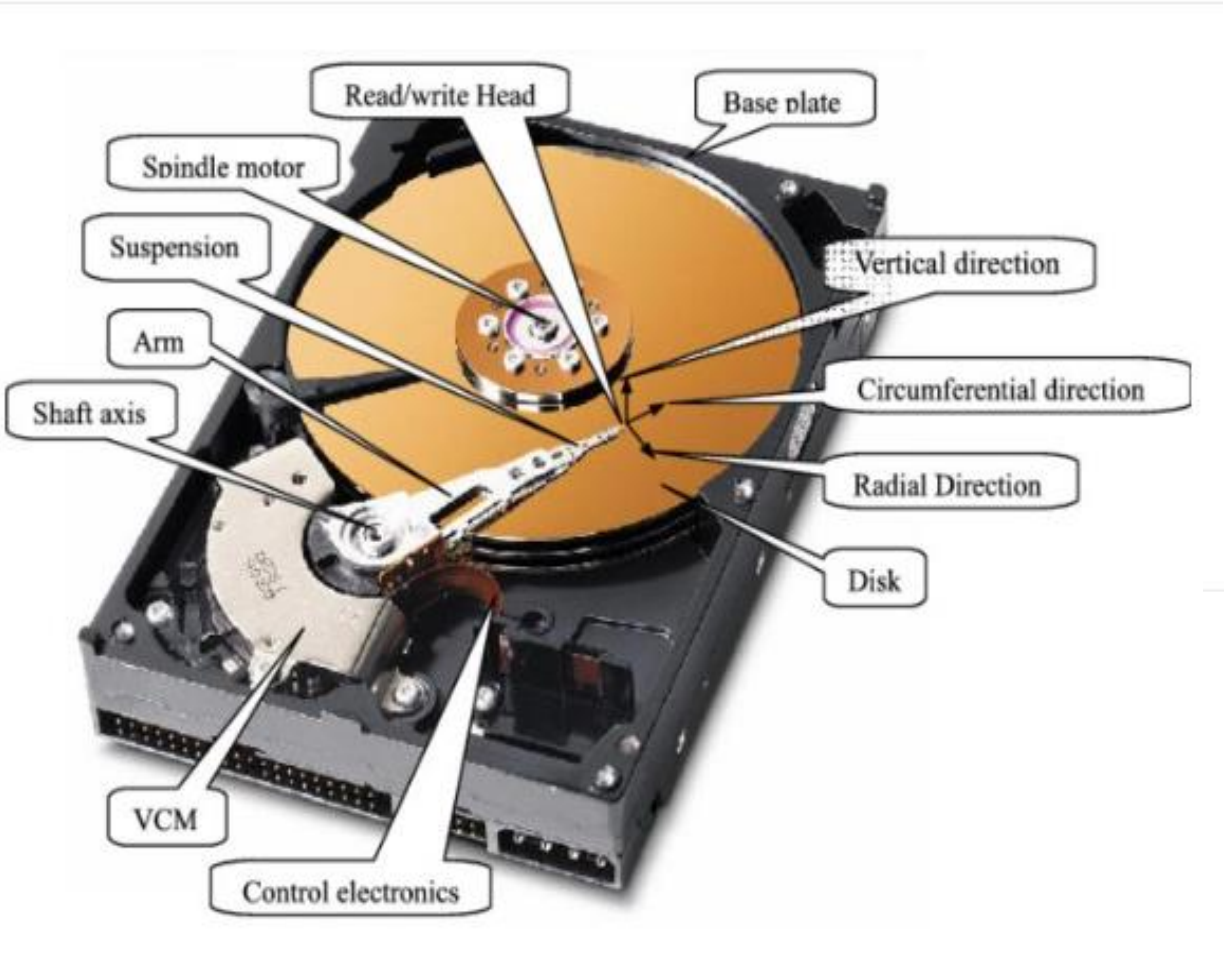
magnetic disks dominate

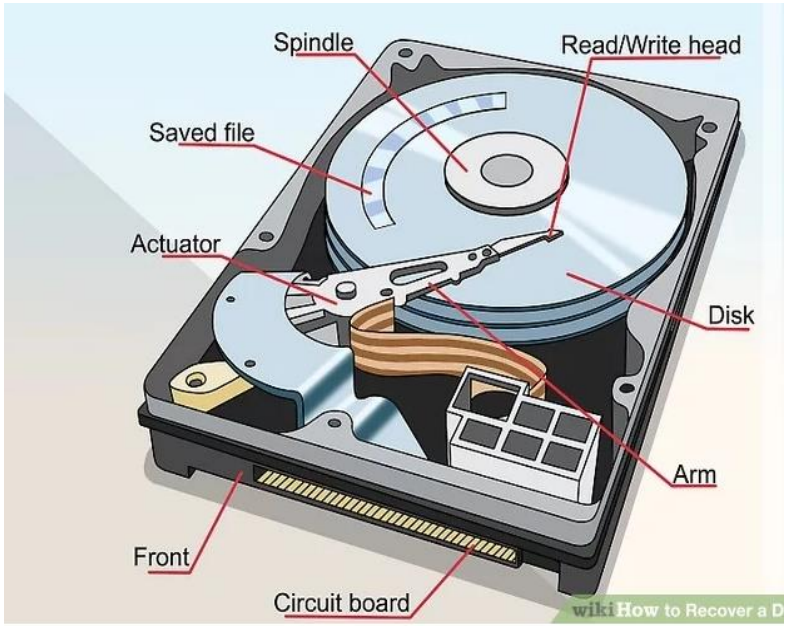
Disk Storage

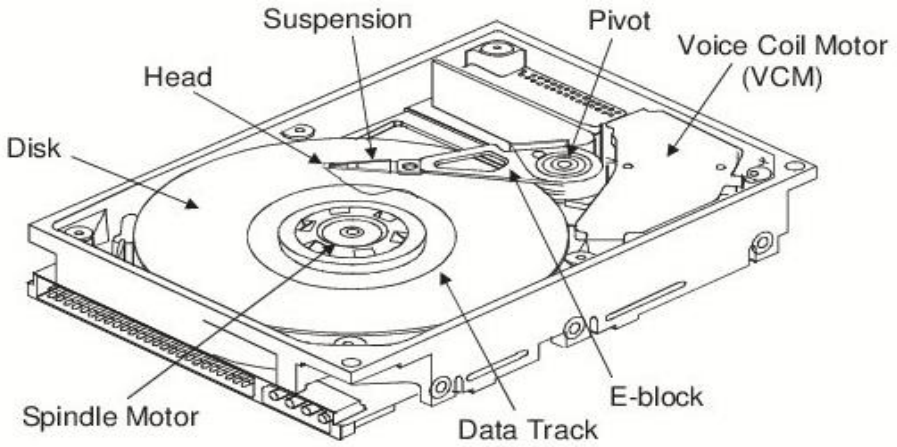
- Nonvolatile, rotating magnetic storage



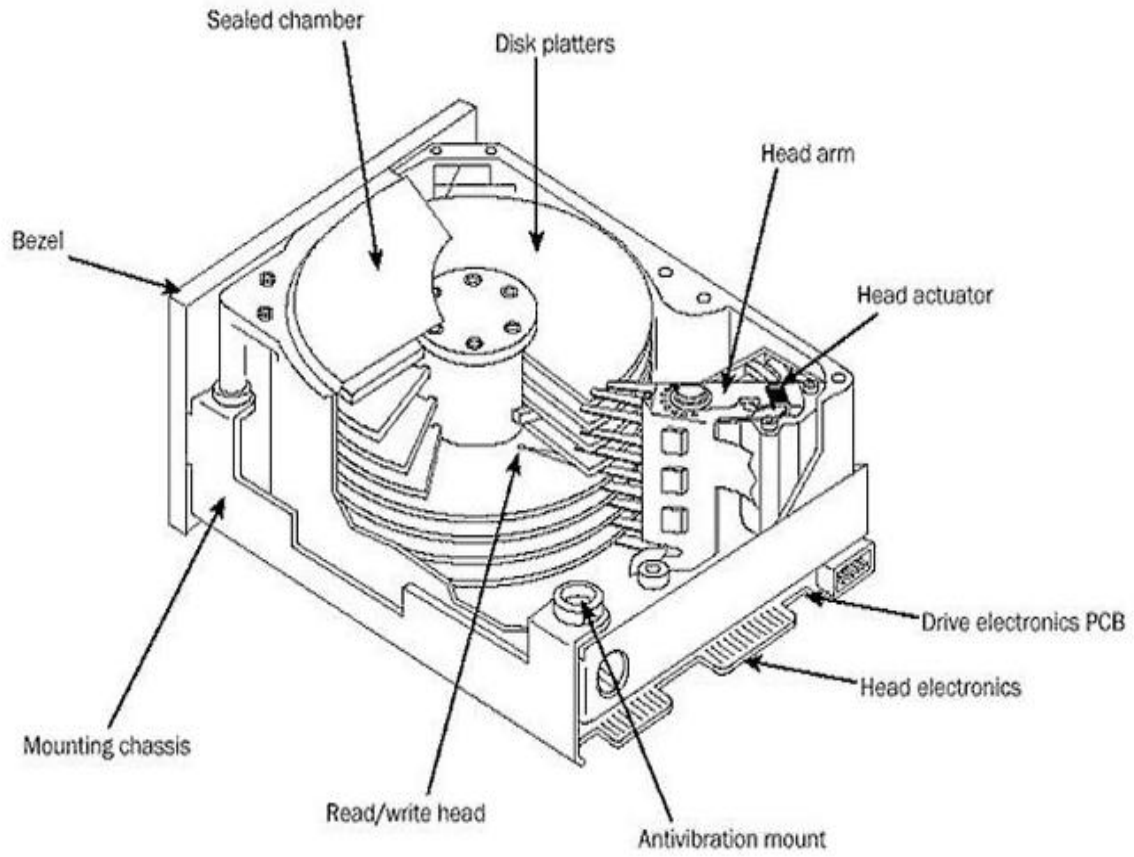
Hard disk components

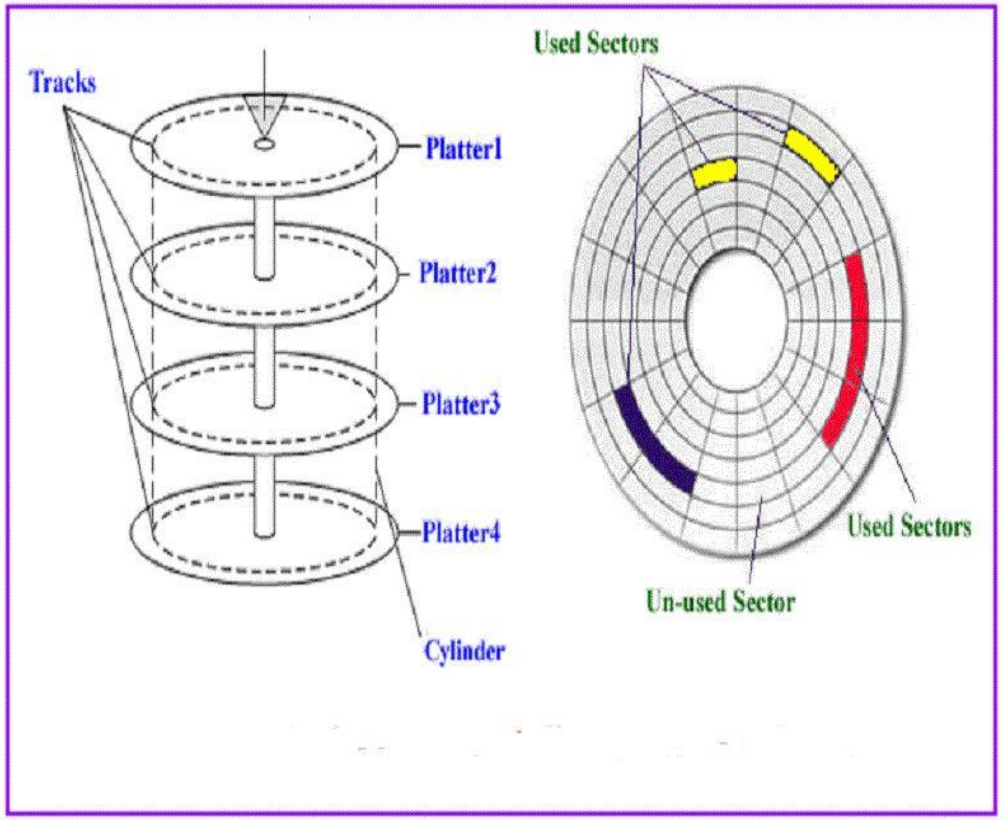






Internal Architecture of Hard Disk





Magnetic details

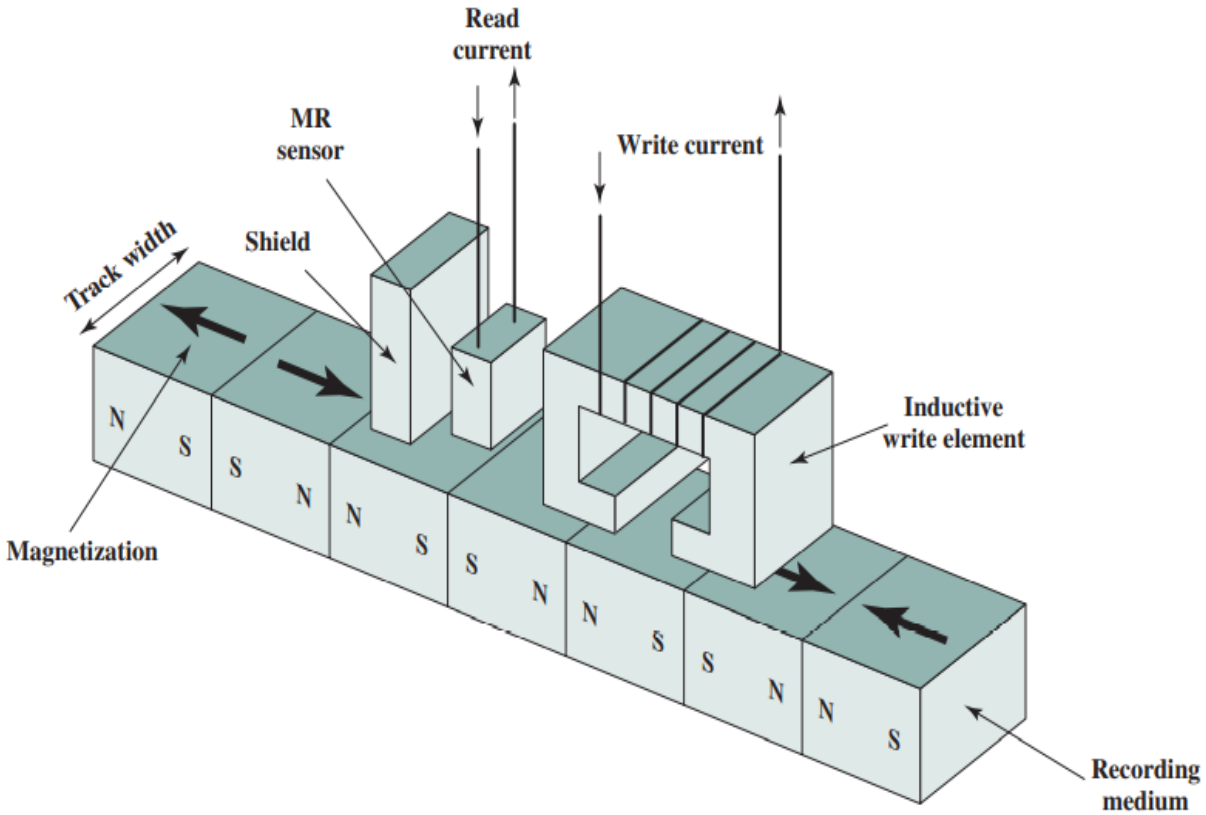


Figure 6.1 Inductive Write/Magnetoresistive Read Head

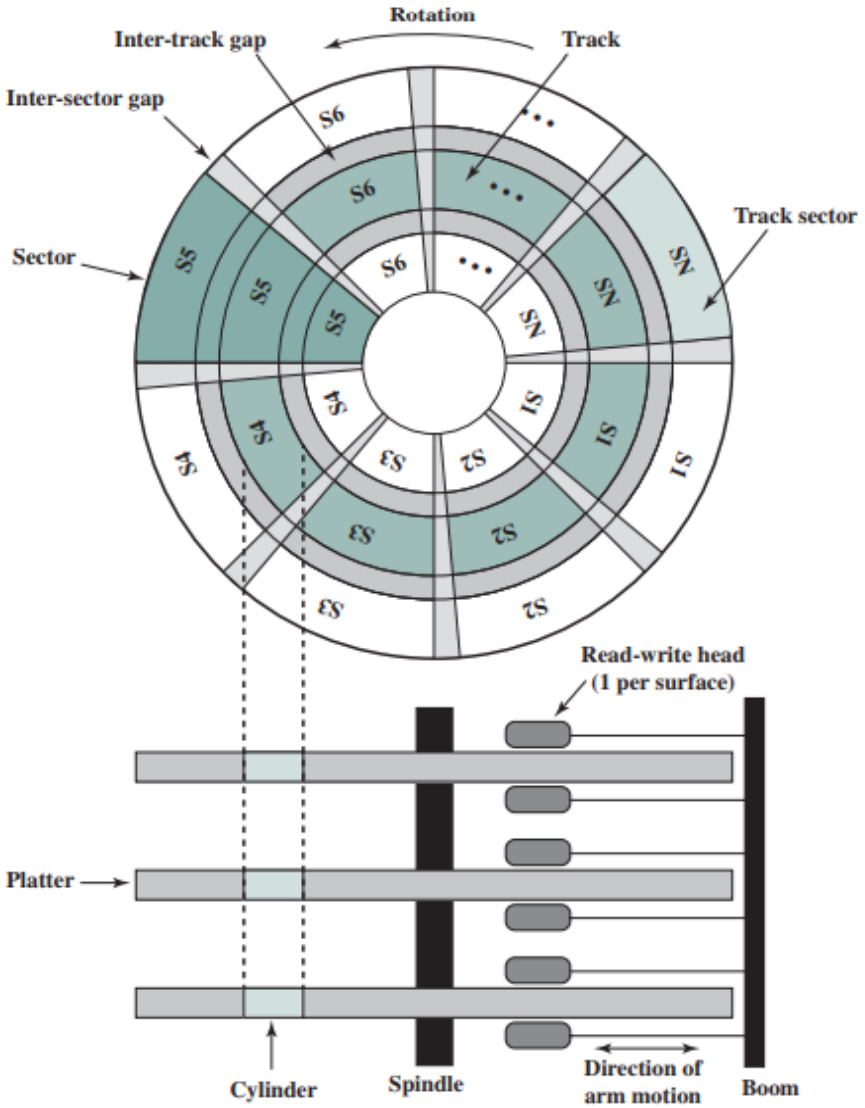
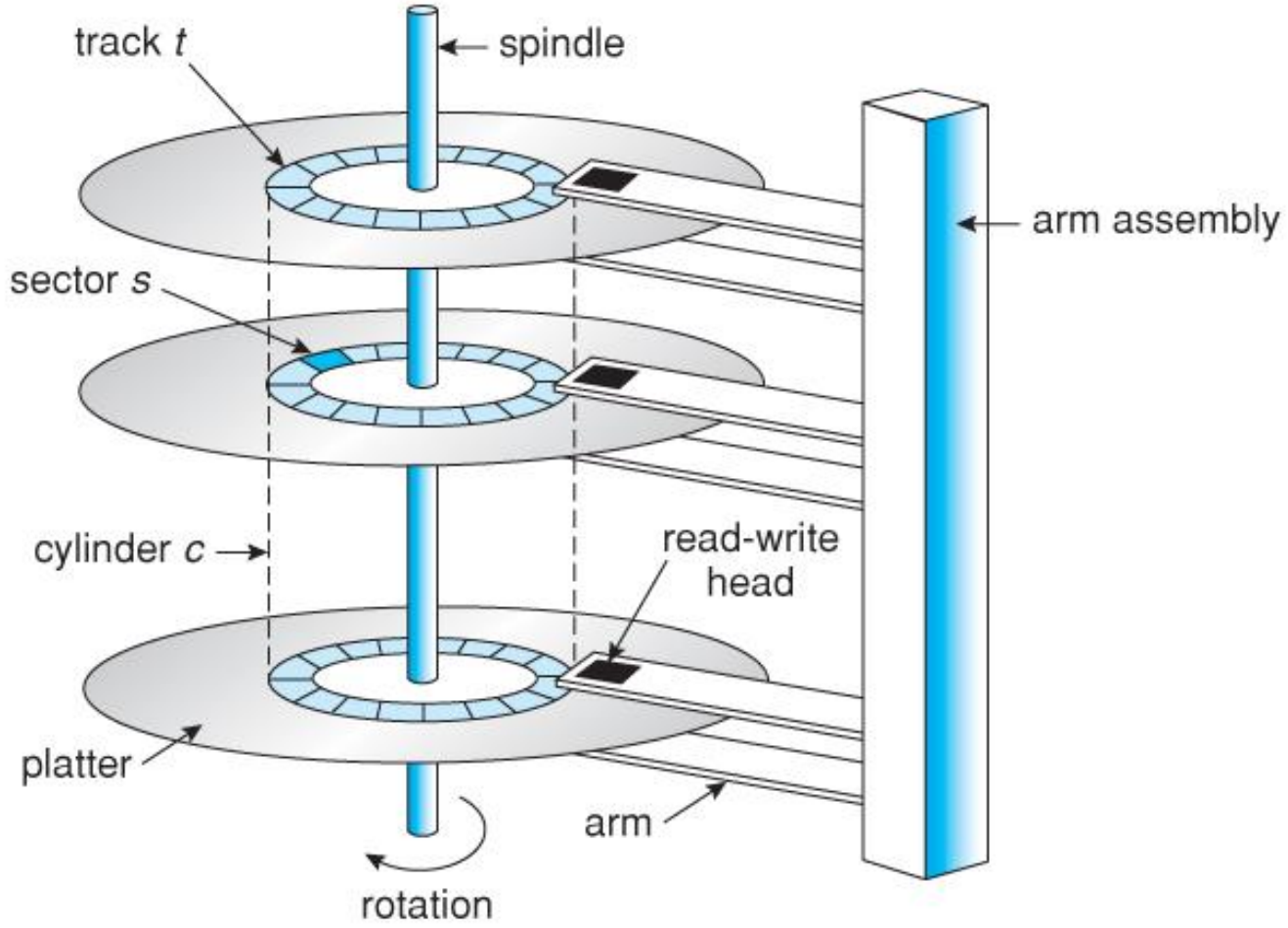


Figure 6.2 Disk Data Layout

Disk Layers



http://www.cs.uic.edu/~jbell/CourseNotes/OperatingSystems/images/Chapter10/10_01_DiskMechanism.jpg

Disk Principle and Capacity

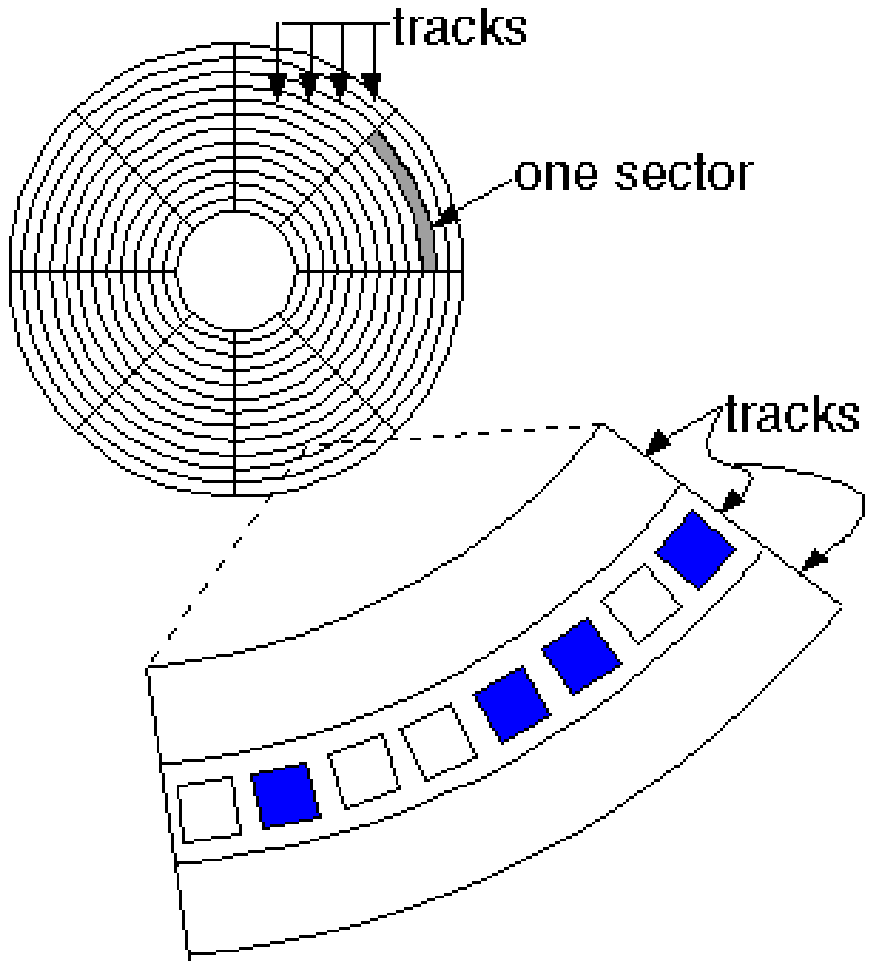
- **Areal Density**

  in 2011, the highest density

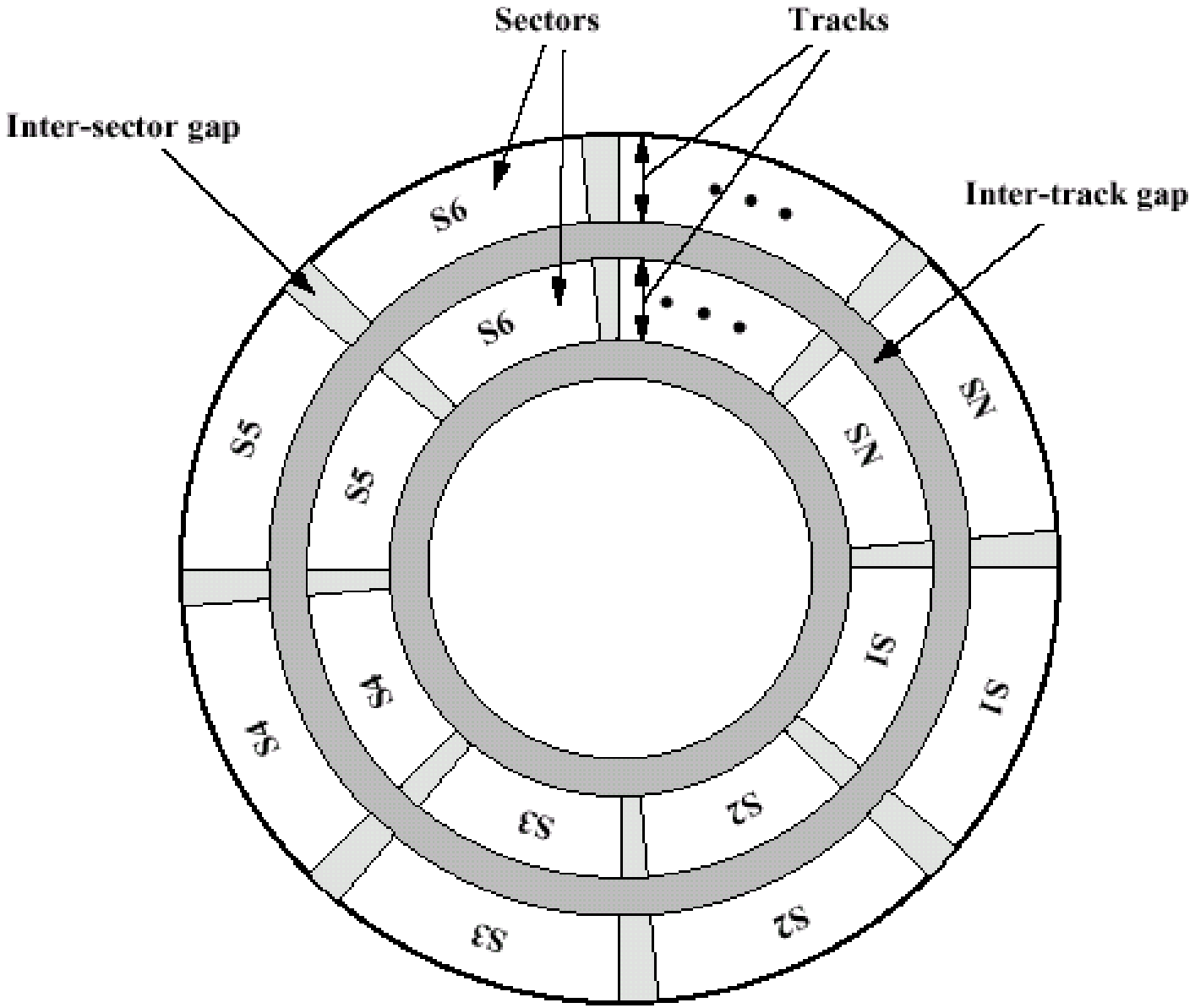
  400 billion bits per square inch
- **Costs per gigabyte**

  improved by almost since 2011

  a factor of 1,000,000



Disk Data Layout



Magnetic Disk

- Metal or plastic disk coated with magnetizable material (iron oxide…rust)
- Range of packaging
  - Floppy
  - Winchester hard disk
  - Removable hard disk

Floppy Disk

- 8”, 5.25”, 3.5”
- Small capacity
  - Up to 1.44Mbyte (2.88M never popular)
- Slow (disk rotate at 300 and 600 rpm, average delay 100/2 and 200/2 ms.)
- Universal
- Cheap

Winchester Hard Disk (1)

- Developed by IBM in Winchester (USA)
- Sealed unit
- One or more platters (disks)
- Heads fly on boundary layer of air as disk spins
- Very small head to disk gap
- Getting more robust

Winchester Hard Disk (2)

- Universal
- Cheap
- Fastest external storage (typically rotate 3600 rpm, newer faster, average rotational delay 8.3 ms.)
- Getting larger all the time
  - Multiple Gigabyte now usual

Fixed / Movable Head Disk

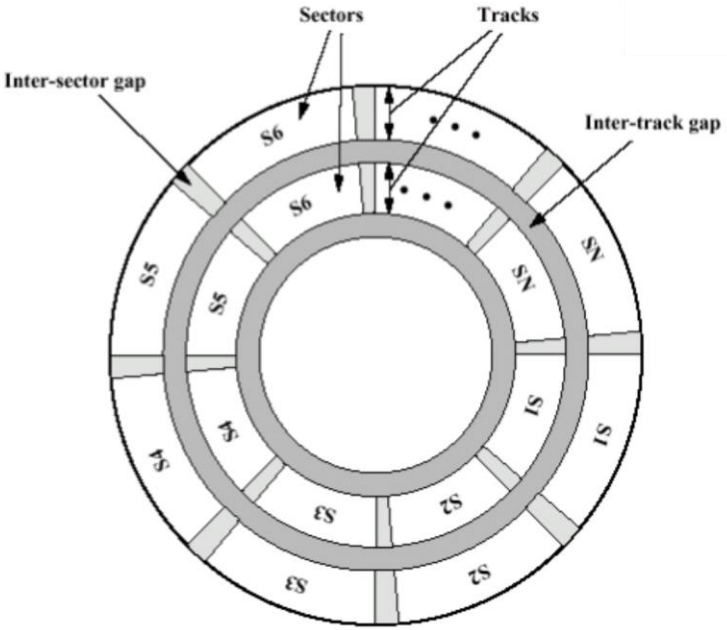
- Fixed head
  - One read write head per track
  - Heads mounted on fixed ridged arm
- Movable head
  - One read write head per side
  - Mounted on a movable arm

HDD - Important concepts

- Circular plate of metal or plastic coated with magnetizable material
- Data is read and written through a magnetic coil called Head
- Electrical pulses are sent to Head that magnetizes the surface material as 1 or 0
- Data is organized as circular tracks. Width of track is same as that of Head. 500 to 2000 tracks per surface.
- Adjacent tracks are separated by Gaps. This is to prevent errors due to Head mis-alignment.
- Same number of bits stored per track – thus bit density is higher in inner tracks and reduces outward

HDD – concepts continued

- Blocks – Data is transferred as blocks. One track is divided into some number of blocks. One block size region is called a SECTOR. There is some gap in between Sectors. Usually there are 10 to 100 Sectors per track.
- Control Data – is stored on disk to identify start of new Blocks, etc.
- Each Track contains 30 fixed length Sectors of 600 Bytes each
- Each SECTOR holds 512 bytes data plus some control information



Multiple Platter

- One head per side
- Heads are joined and aligned
- Aligned tracks on each platter form cylinders
- Data is striped by cylinder
  - reduces head movement
  - Increases speed (transfer rate)

Data Organization and Formatting

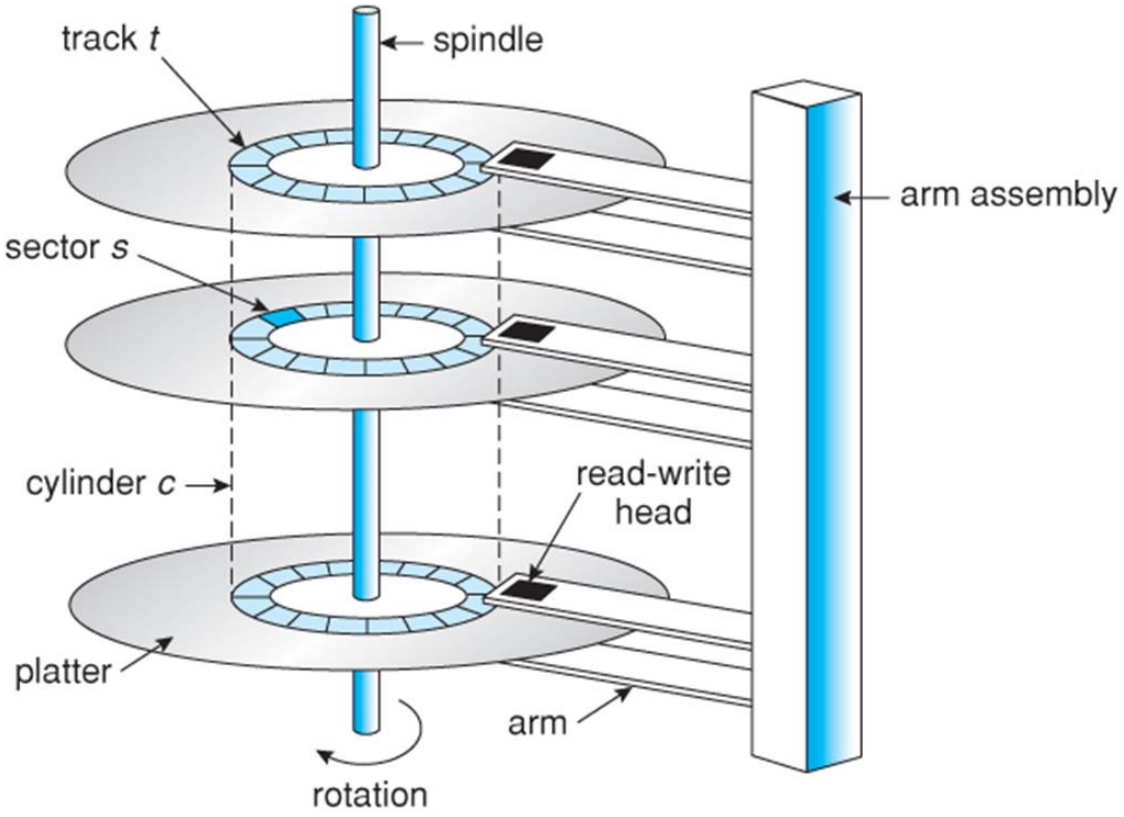
- Concentric rings or tracks
  - Gaps between tracks
  - Reduce gap to increase capacity
  - Same number of bits per track (variable packing density)
  - Constant angular velocity
- Tracks divided into sectors
- Minimum block size is one sector
- May have more than one sector per block

Finding Sectors

- Must be able to identify start of track and sector
- Format disk
  - Additional information not available to user
  - Marks tracks and sectors

Characteristics

- Fixed (rare) or movable head
- Removable or fixed
- Single or double (usually) sided
- Single or multiple platter
- Head mechanism
  - Contact (Floppy)
  - Fixed gap
  - Flying (Winchester)



Disk's Competitor

- **Flash Memory**

  non-volatile semiconductor memory;

  same bandwidth as disks;

  100 to 1000 times faster;

  15 to 25 times higher cost/gigabyte;
- **Wear out**

  limited to 1 million writes
- **Popular in cell phones,**

  but not in desktop and server

Examples of Hard Disk Performance

Magnetic Timing

SEEK TIME Seek time is the time required to move the disk arm to the required track. It turns out that this is a difficult quantity to pin down. The seek time consists of two key components: the initial startup time, and the time taken to traverse the tracks that have to be crossed once the access arm is up to speed. Unfortunately, the traversal time is not a linear function of the number of tracks, but includes a settling time (time after positioning the head over the target track until track identification is confirmed).

Much improvement comes from smaller and lighter disk components. Some years ago, a typical disk was 14 inches (36 cm) in diameter, whereas the most common size today is 3.5 inches (8.9 cm), reducing the distance that the arm has to travel. A typical average seek time on contemporary hard disks is under 10 ms.

ROTATIONAL DELAY Disks, other than floppy disks, rotate at speeds ranging from 3600 rpm (for handheld devices such as digital cameras) up to, as of this writing, 20,000 rpm; at this latter speed, there is one revolution per 3 ms. Thus, on the average, the rotational delay will be 1.5 ms.

TRANSFER TIME The transfer time to or from the disk depends on the rotation speed of the disk in the following fashion:

$$T = \frac{b}{rN}$$

where

T = transfer time
b = number of bytes to be transferred
N = number of bytes on a track
r = rotation speed, in revolutions per second

Thus the total average read or write time T_{total} can be expressed as

$$T_{total} = T_s + \frac{1}{2r} + \frac{b}{rN} \qquad \textbf{(6.1)}$$

where T_s is the average seek time. Note that on a zoned drive, the number of bytes per track is variable, complicating the calculation.[1]

Magnetic Disk Timing

- Seek time
  - Moving head to correct track
- (Rotational) latency
  - Waiting for data to rotate under head
- Access time = Seek + Latency
- Transfer rate T = (number of bytes to be transferred)/(rotation speed)/(number of bytes on a track) = b/(rN)
- total access time $T_a = T_s + 1/(2r) + b/(rN)$

HDD Timing Calculation

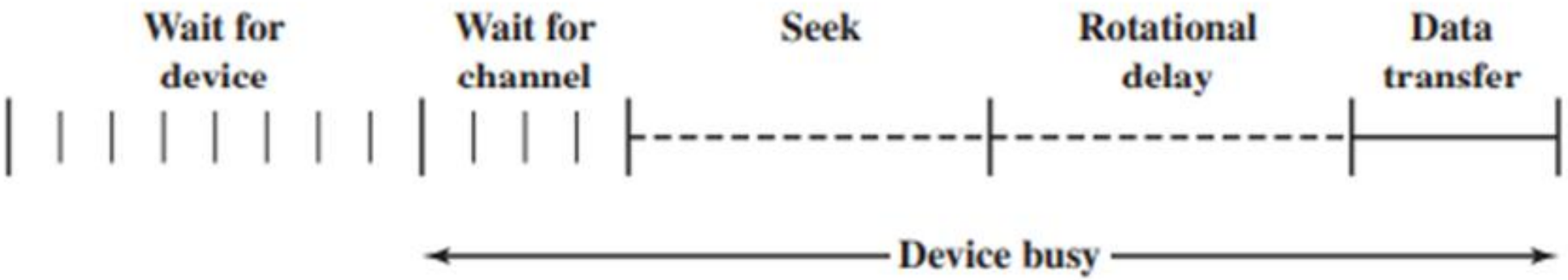
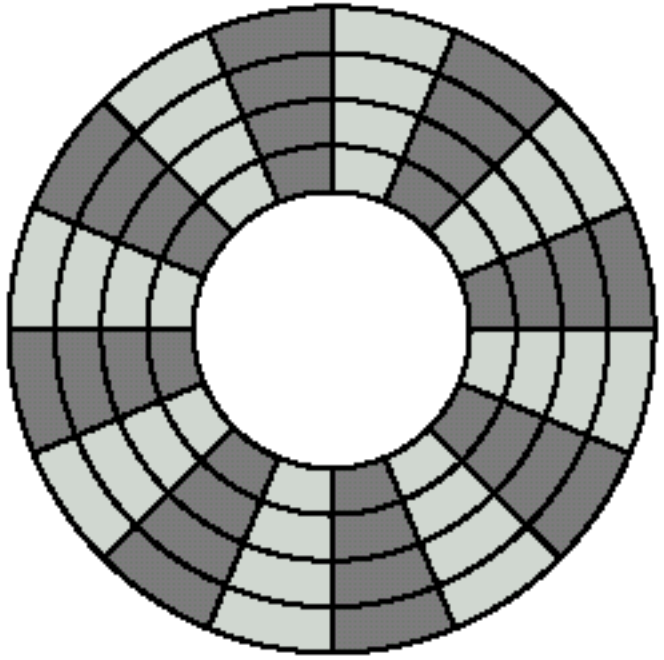
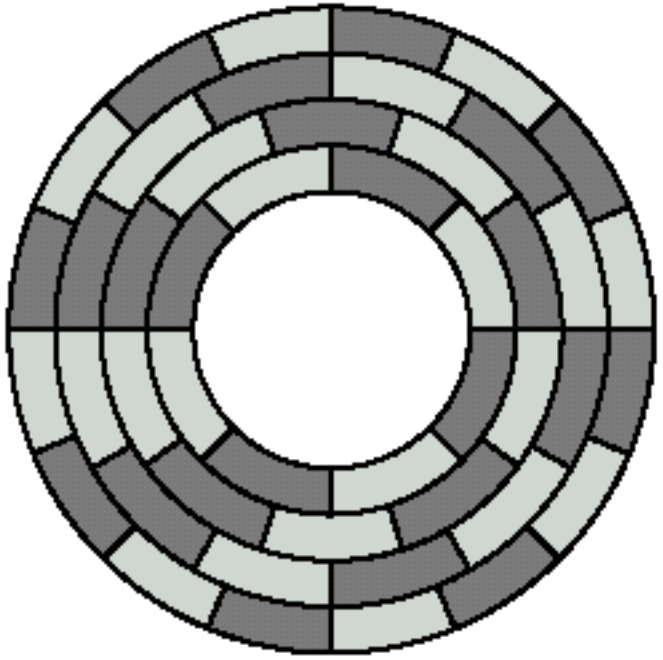


Figure 6.5 Timing of a Disk I/O Transfer

Constant Angular Velocity vs. Constant Linear Velocity



(a) Constant angular velocity



(b) Constant linear velocity

Disk performance formulas

Once the head has reached the correct track, we must wait for the desired sector to rotate under the read/write head. This time is called the **rotational latency** or **rotational delay**. The average latency to the desired information is halfway around the disk. Disks rotate at 5400 RPM to 15,000 RPM. The average rotational latency at 5400 RPM is

$$\text{Average rotational latency} = \frac{0.5 \text{ rotation}}{5400 \text{ RPM}} = \frac{0.5 \text{ rotation}}{5400 \text{ RPM}/\left(60 \frac{\text{seconds}}{\text{minute}}\right)}$$

$$= 0.0056 \text{ seconds} = 5.6 \text{ ms}$$

Disk Access Example

- Given
  - 512B sector, 15,000rpm, 4ms average seek time, 100MB/s transfer rate, 0.2ms controller overhead, idle disk
- Average read time
  - 4ms seek time
    + ½ / (15,000/60) = 2ms rotational latency
    + 512 / 100MB/s = 0.005ms transfer time
    + 0.2ms controller delay
    = 6.2ms
- If actual average seek time is 1ms
  - Average read time = 3.2ms

Sequential organization vs. random access e.g.

- e.g. a hard disk has average seek time of 20 ms, a transfer rate of 1 M byte/s, and 512 byte sectors with 32 sectors per track. Need to read a file consisting 256 sectors for a total of 128 K bytes. What is the total time for the transfer?

- Case 1: Sequential Organization (256 sectors on 8 tracks x 32 sectors/tracks)
  - Average seek time = 20.0 ms
  - Rotational delay = 8.3 ms
  - Read 32 sections (one track) = 16.7 ms
  - total time to read first track = 45 ms
  - Total time = 45 ms + 7*(8.3 + 16.7) ms = 0.22 s

Time required for random access on highly fragmented organization

- Case 2: random access rather than sequential access
  - Average seek time = 20.0 ms
  - Rotational delay = 8.3 ms
  - Read 1 sector = 16.7/32 = 0.5 ms
  - time to read one sector = 28.8 ms
  - Total time = 256 * 28.8 ms = 7.37 s

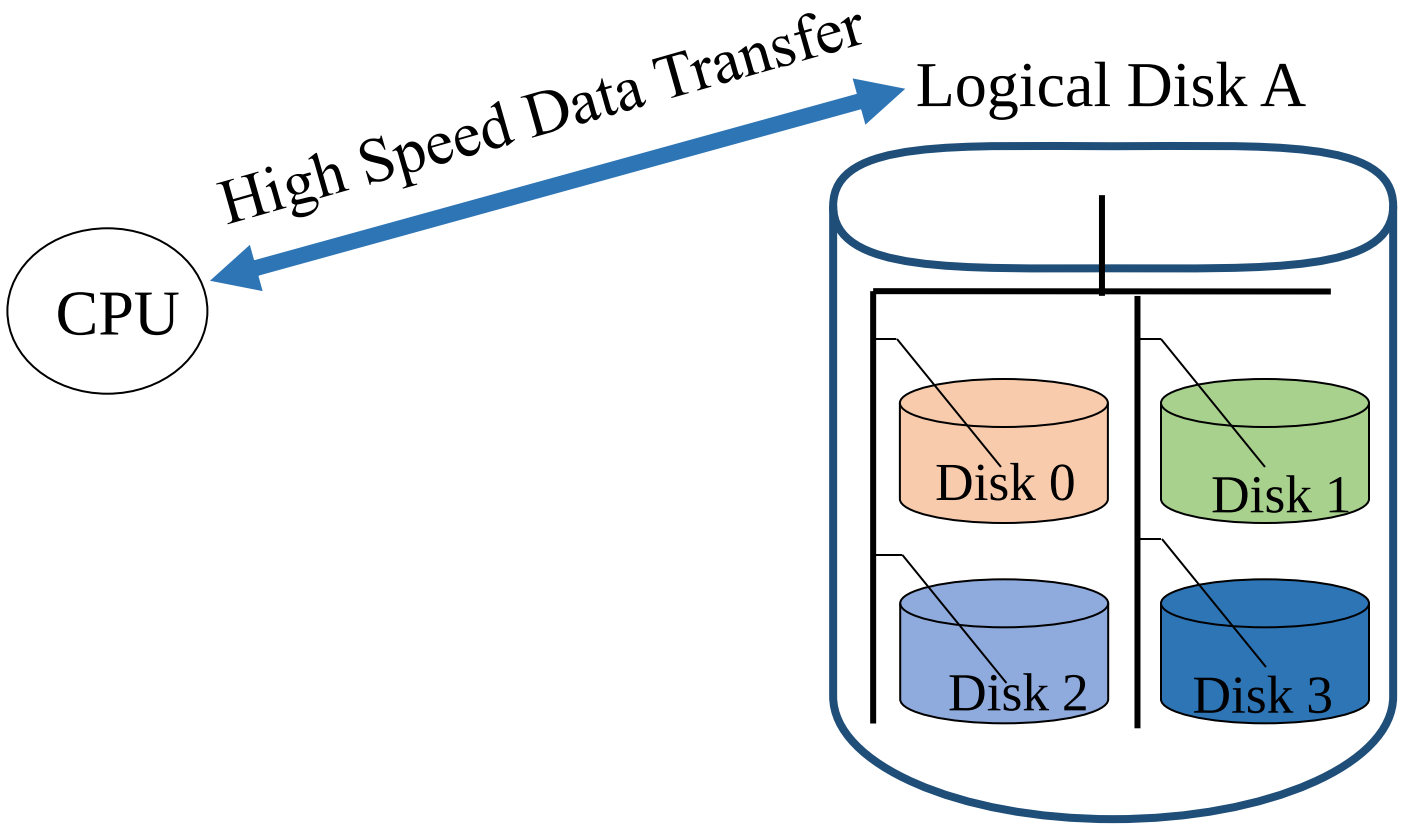
- De-fragment you hard disk!

Storage Systems
RAID Array

Disk Arrays

- Disk arrays with **redundant disks** to tolerate faults
- If a single disk fails, the lost information is reconstructed from redundant information
- **Striping**: simply spreading data over multiple disks
- **RAID:** redundant array of inexpensive/independent disks

Improving Performance of External Storage



Make an Array of External Hard Disks

RAID IDEA

- RAID = Redundant Array of Independent Disks
- Use multiple hard disks in parallel to:
  - Increase the data storage space
  - Parallel data access to improve speed of data transfer
  - Built-In Backup and Redundancy for Data Recovery
  - Increased Reliability of Disk Based Storage Systems

RAID

- Redundant Array of Independent Disks
- Redundant Array of Inexpensive Disks
- 6 levels in common use
- Not a hierarchy
- Set of physical disks viewed as single logical drive by O/S
- Data distributed across physical drives
- Can use redundant capacity to store parity information

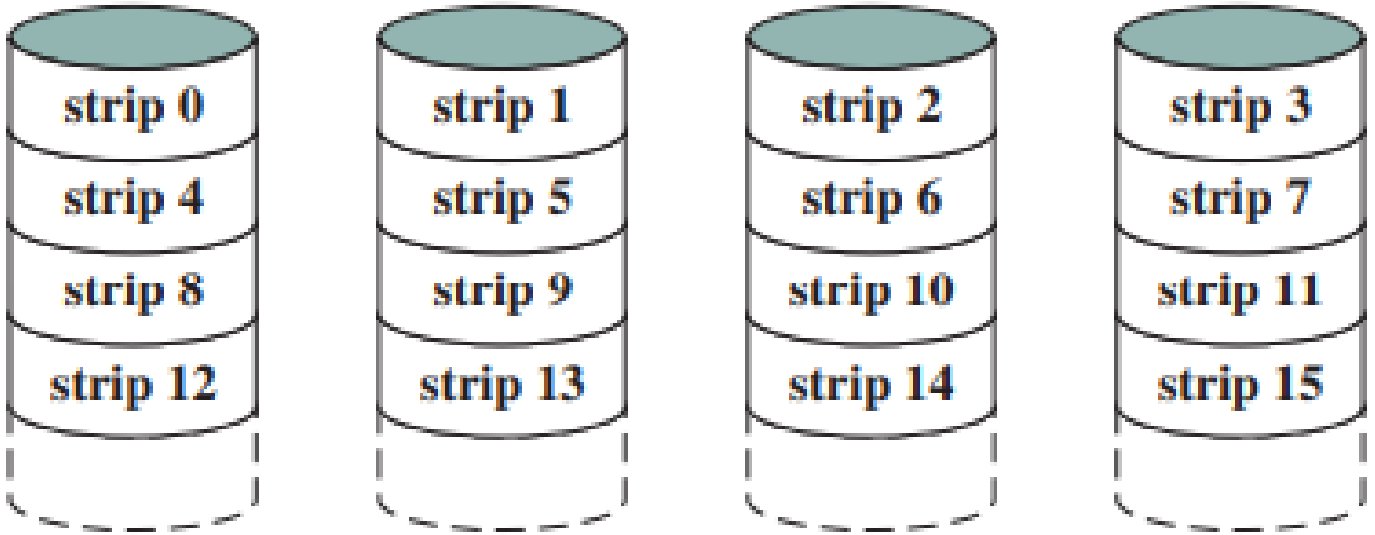
RAID 0

- No redundancy
- Data striped across all disks
- Round Robin striping
- Easier incremental increase in Capacity
- Increase speed
  - Multiple data requests probably not on same disk
  - Disks seek in parallel
  - A set of user and system data is likely to be striped across multiple disks
- Useful in high performance and low cost demands

RAID 0

- **JBOD =** just a bunch of disks
- No redundancy
- No failure tolerated
- Measuring stick for other RAID levels in terms of cost, performance, and dependability

RAID Level 0



(a) RAID 0 (Nonredundant)

Data Mapping in RAID 0

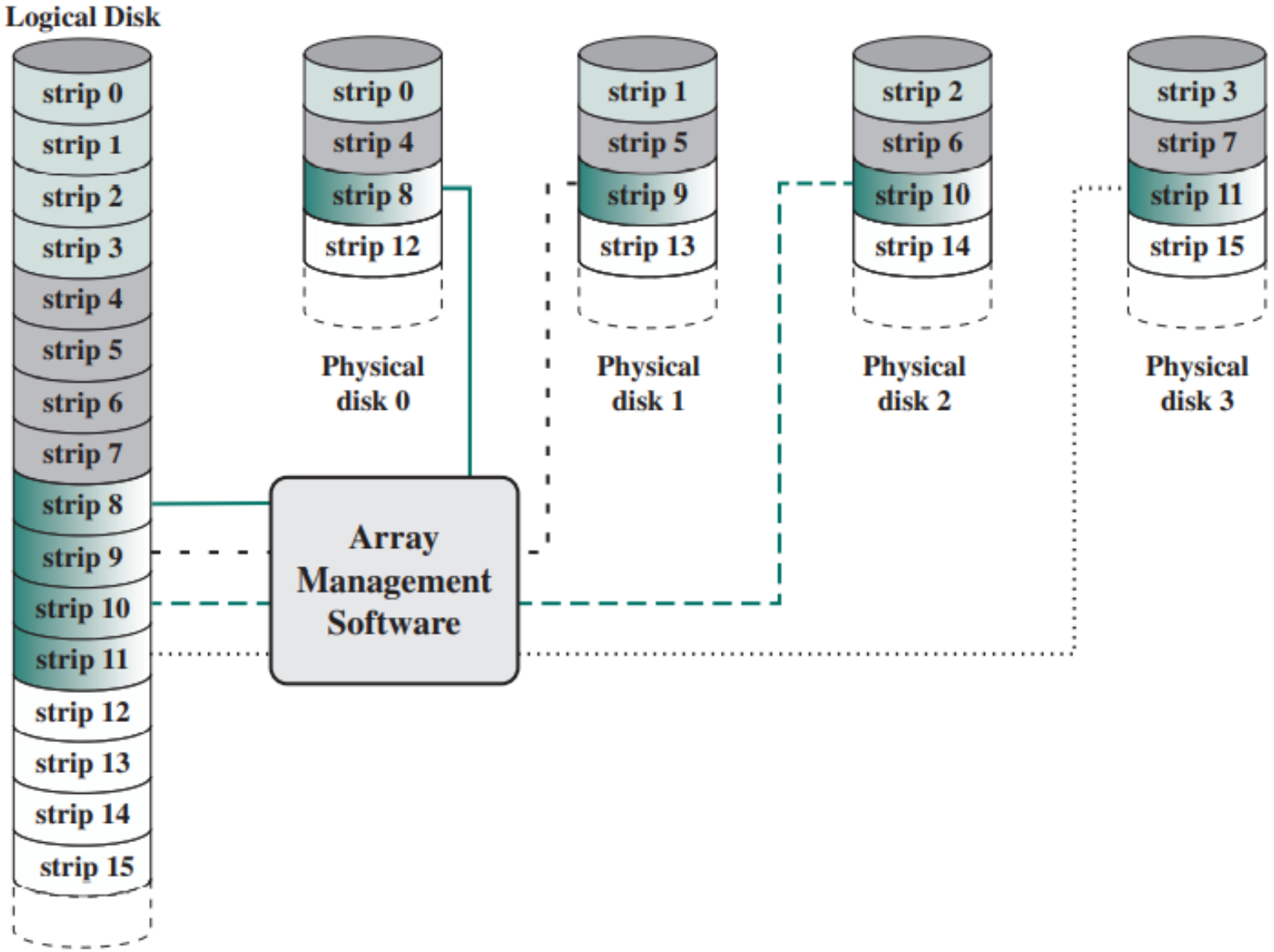


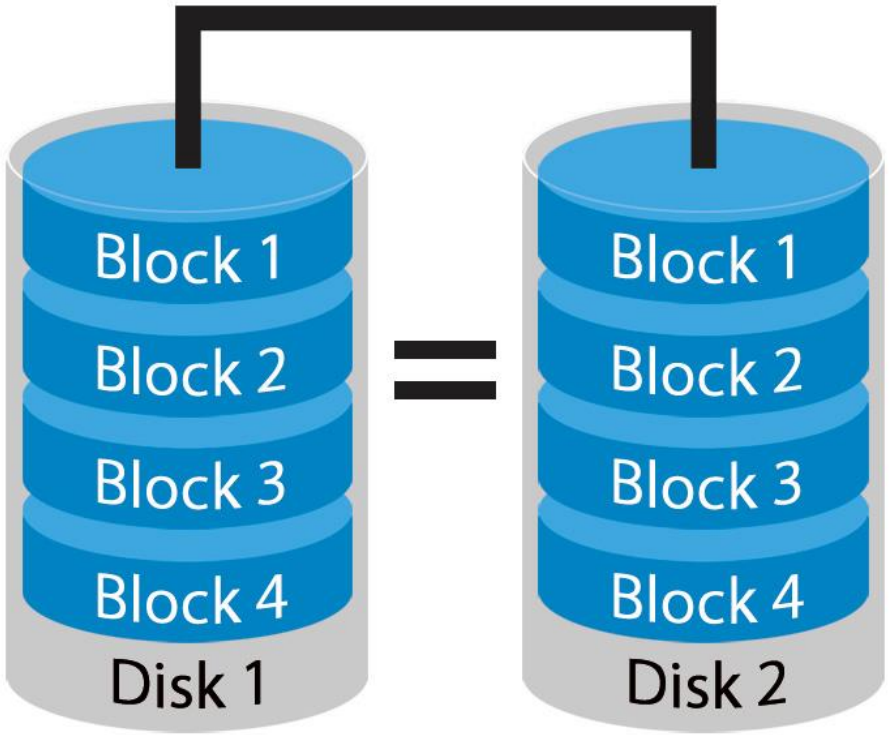
Figure 6.7 Data Mapping for a RAID Level 0 Array

RAID 1

- Mirrored Disks
- Data is striped across disks
- 2 copies of each stripe on separate disks
- Read from either
- Write to both
- Recovery is simple
  - Swap faulty disk & re-mirror
  - No down time
- Expensive

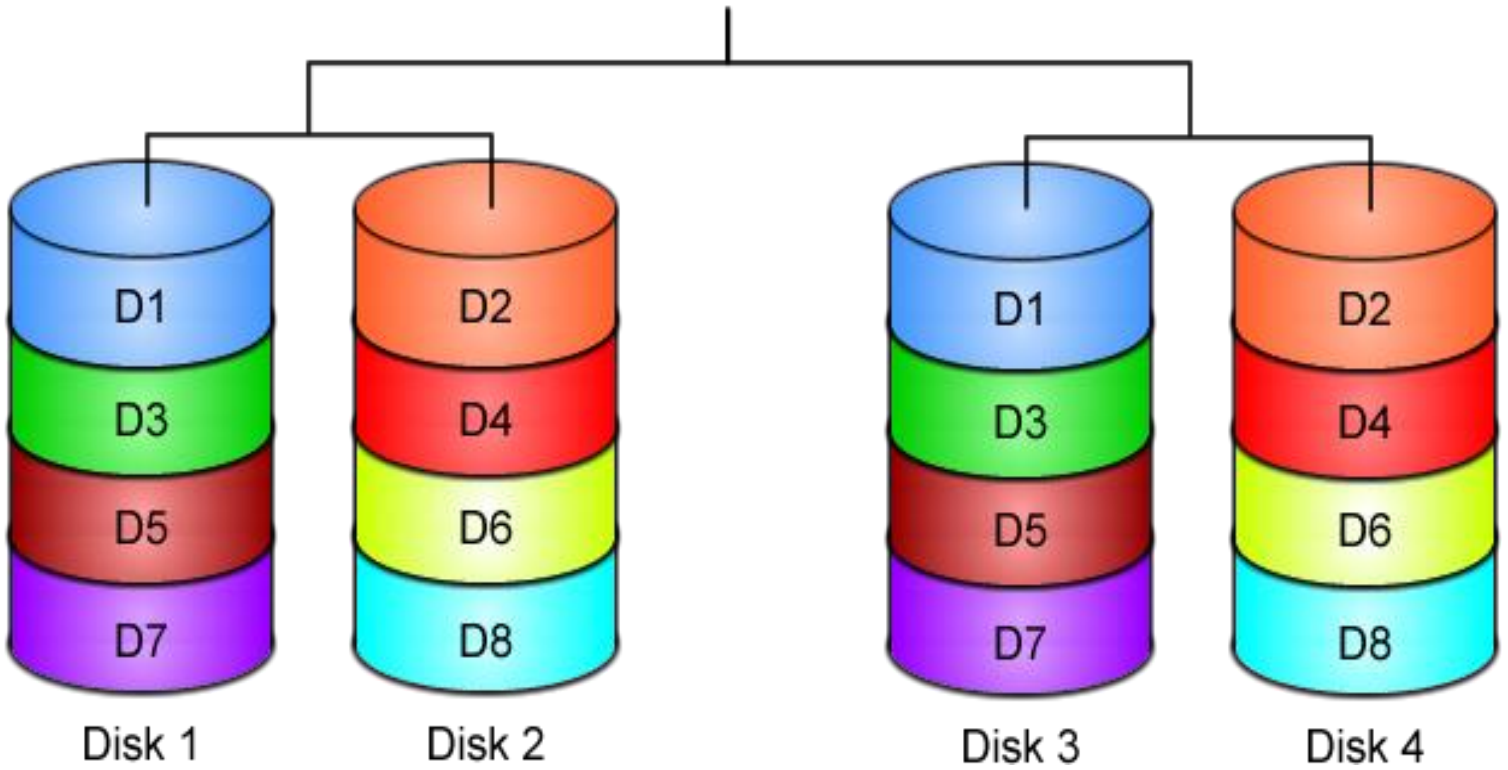
RAID 1

- **Mirroring** or **Shadowing**
- Two copies for every piece of data
- one logical write = two physical writes
- 100% capacity/space overhead



http://www.petemarovichimages.com/wp-content/uploads/2013/11/RAID1.jpg

RAID 0+1 (Stripe+Mirror)

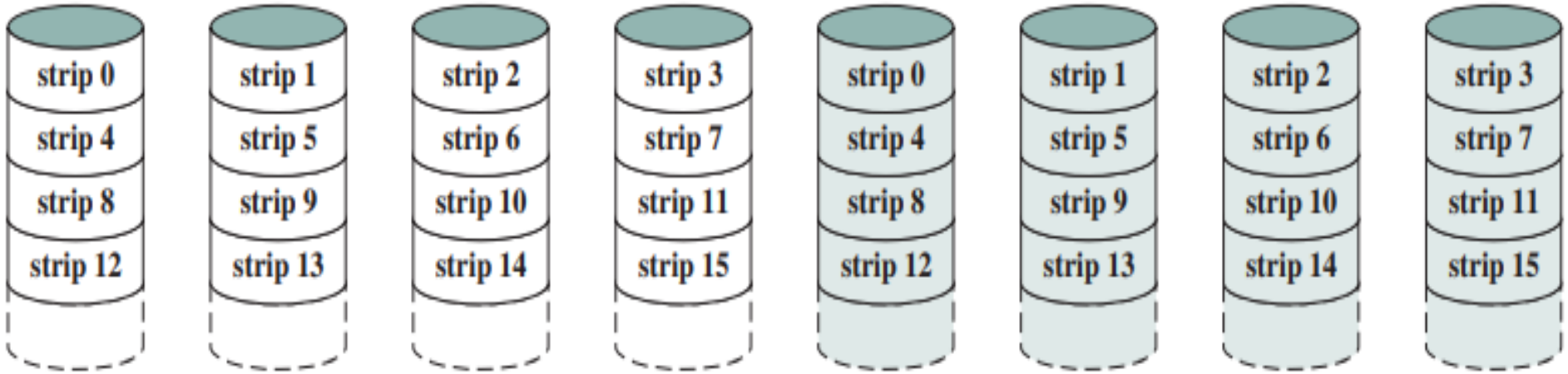


https://www.icc-usa.com/content/raid-calculator/raid-0-1.png

RAID 2

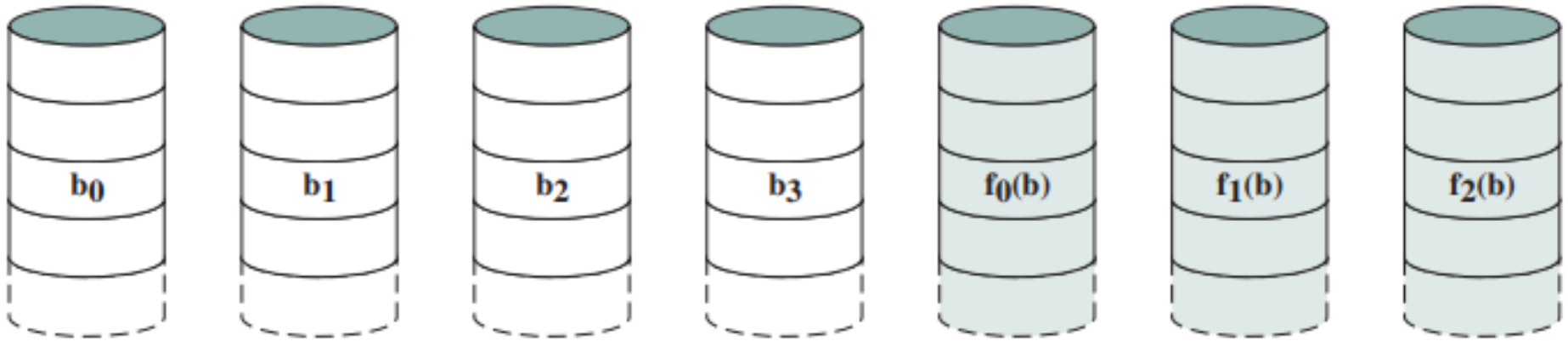
- Disks are synchronized
- Very small stripes
  - Often single byte/word
- Error correction calculated across corresponding bits on disks
- Multiple parity disks store Hamming code error correction in corresponding positions
- Lots of redundancy
  - Expensive
  - Not used

RAID Level 1



(b) RAID 1 (Mirrored)

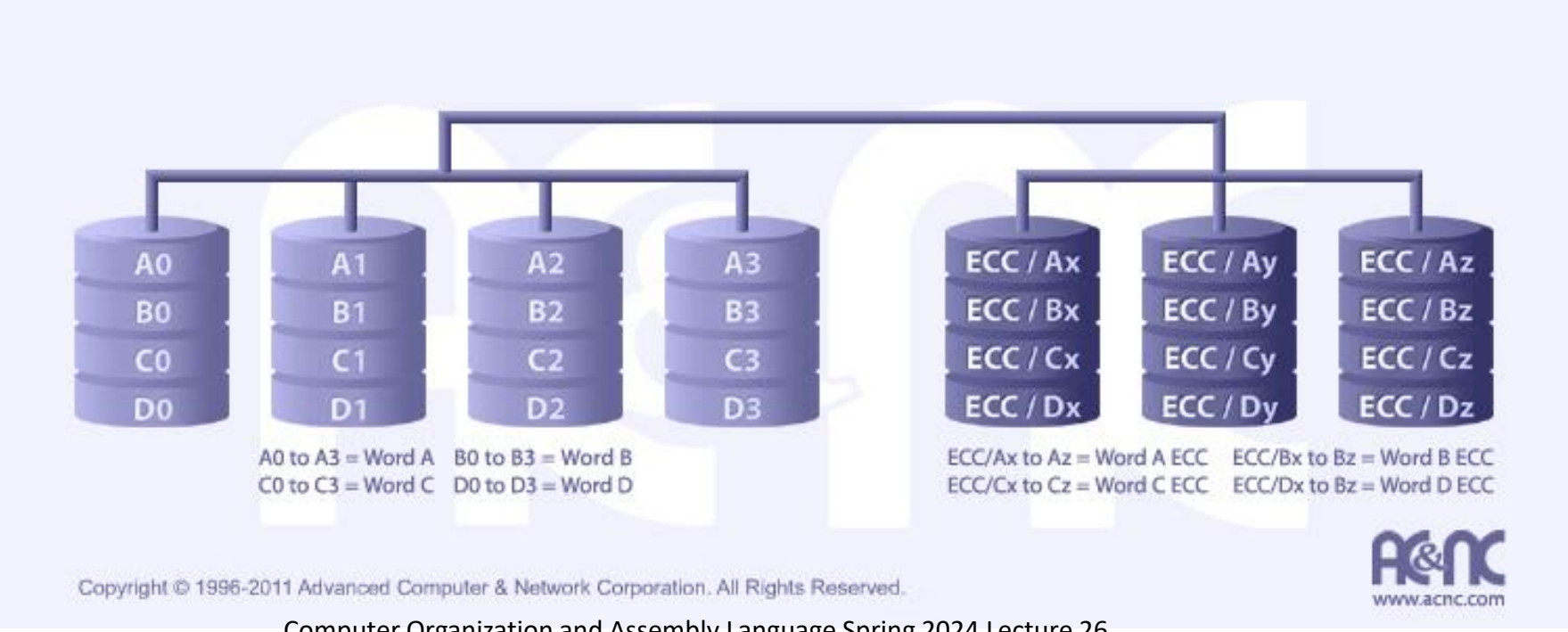
RAID Level 2



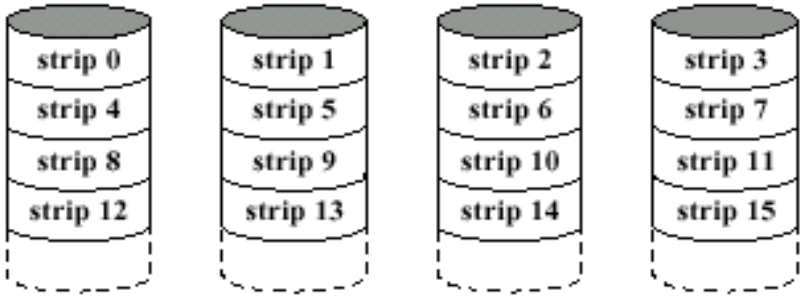
(c) RAID 2 (Redundancy through Hamming code)

RAID 2

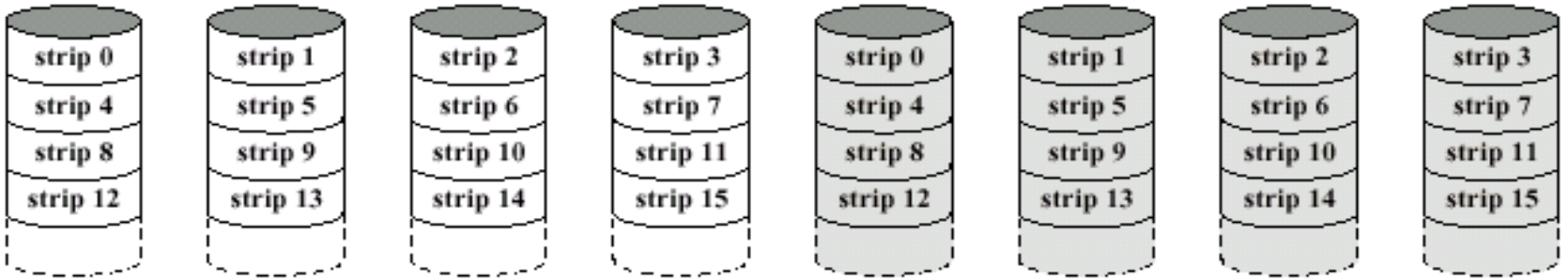
- http://www.acnc.com/raidedu/2
- Each bit of data word is written to a data disk drive
- Each data word has its (Hamming Code) ECC word recorded on the ECC disks
- On read, the ECC code verifies correct data or corrects single disks errors



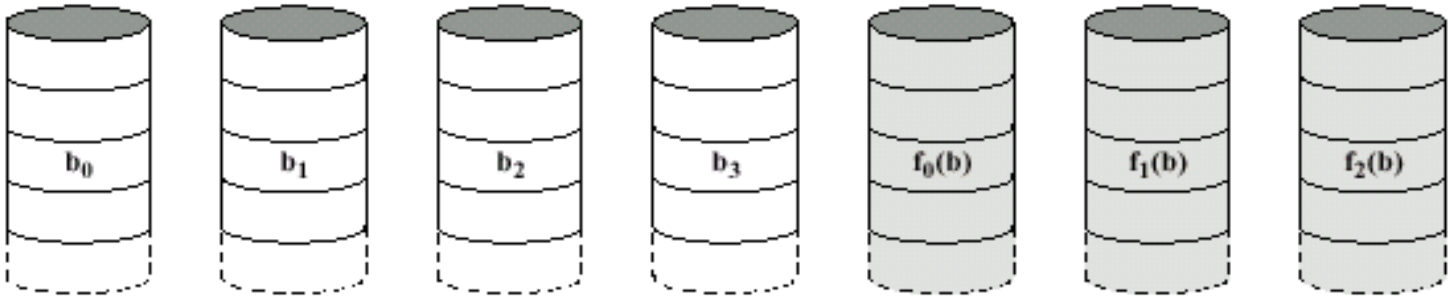
RAID Levels 0, 1, 2



(a) RAID 0 (non-redundant)



(b) RAID 1 (mirrored)

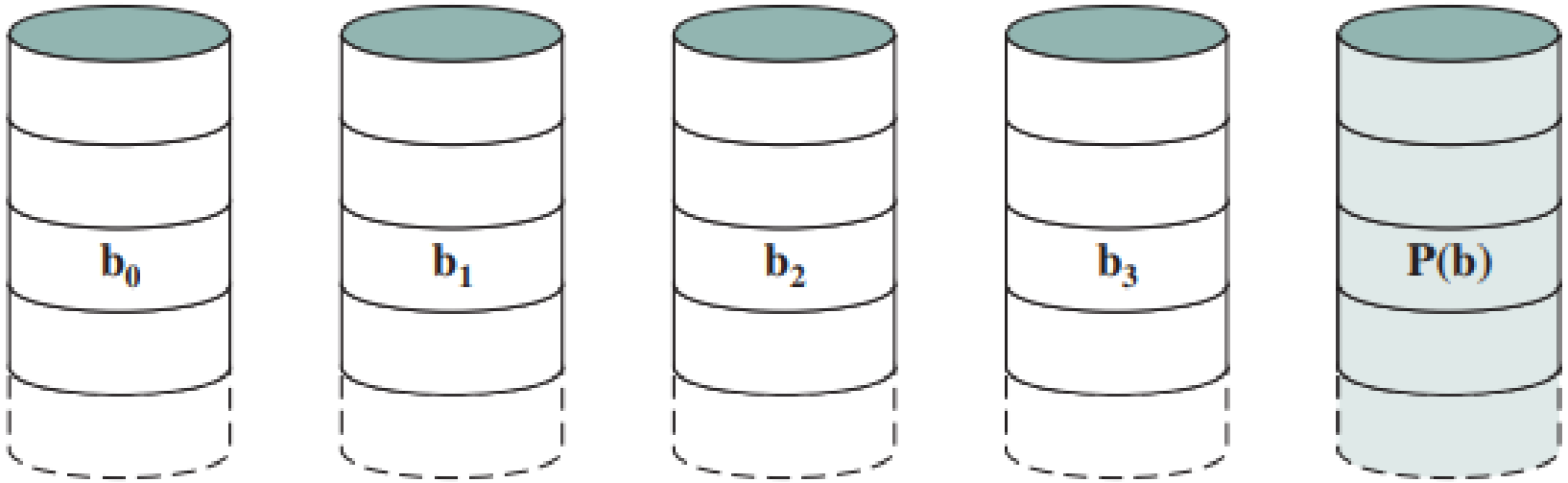


(c) RAID 2 (redundancy through Hamming code)

RAID 3

- Similar to RAID 2
- Only one redundant disk, no matter how large the array
- Simple parity bit for each set of corresponding bits
- Data on failed drive can be reconstructed from surviving data and parity info
- Very high transfer rates

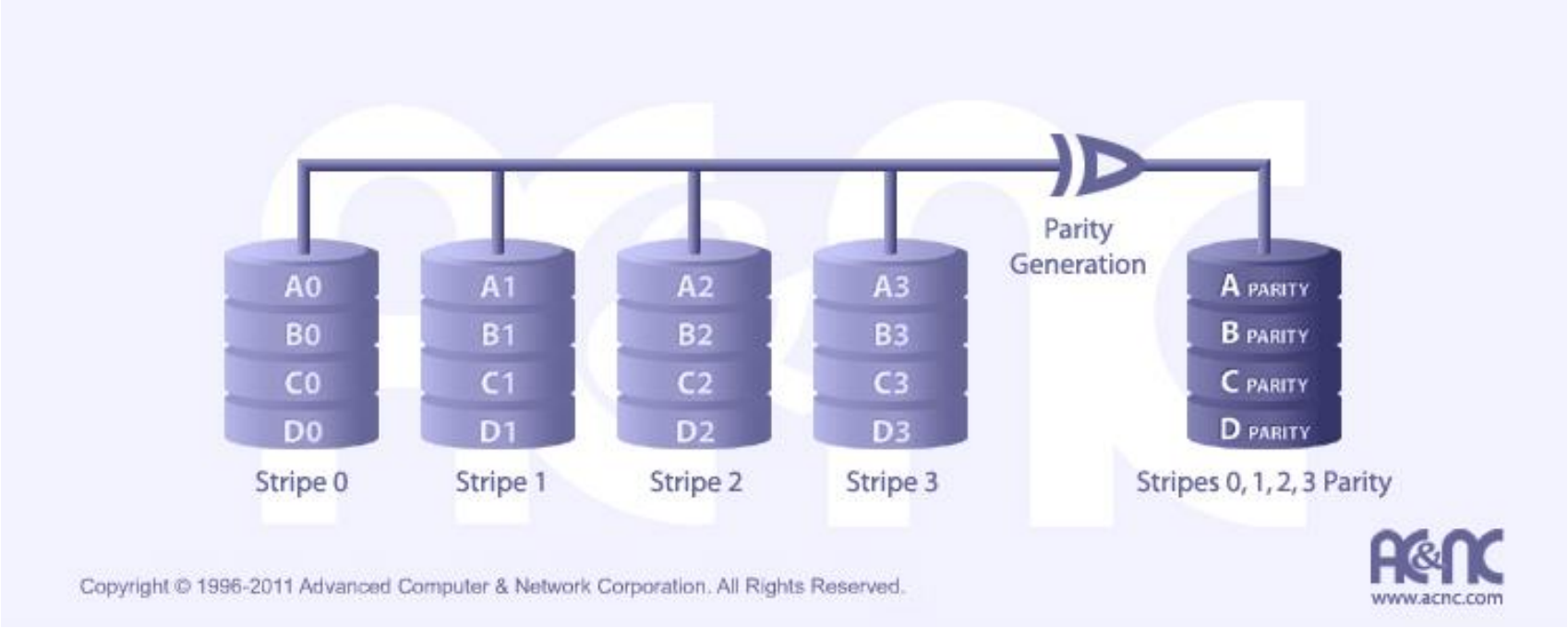
RAID Level 3



(d) RAID 3 (Bit-interleaved parity)

RAID 3

- http://www.acnc.com/raidedu/3
- Data striped over all data disks
- Parity of a stripe to parity disk
- Require at least 3 disks to implement



RAID 3

- Even Parity

  parity bit makes

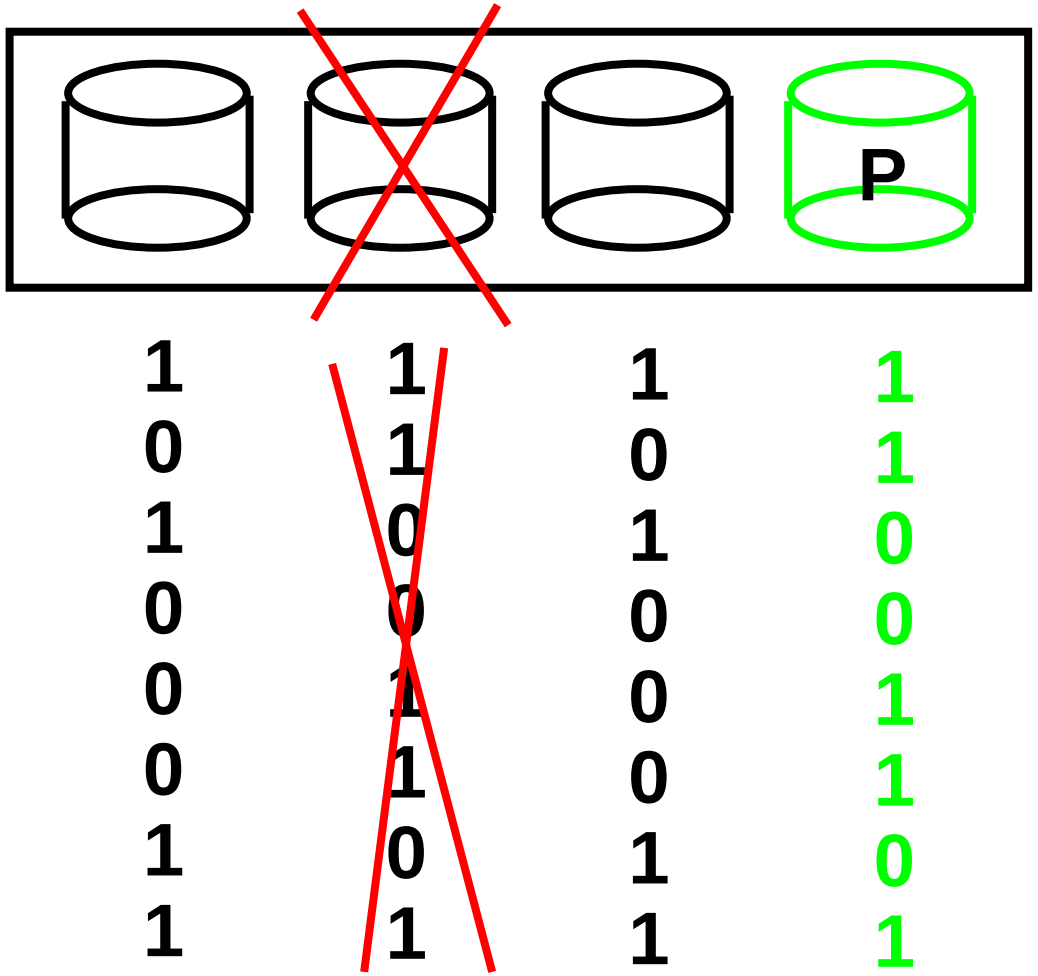
  the No. of 1 even
- p = sum(data1) mod 2
- Recovery

  if a disk fails,

  “subtract” good data

  from good blocks;

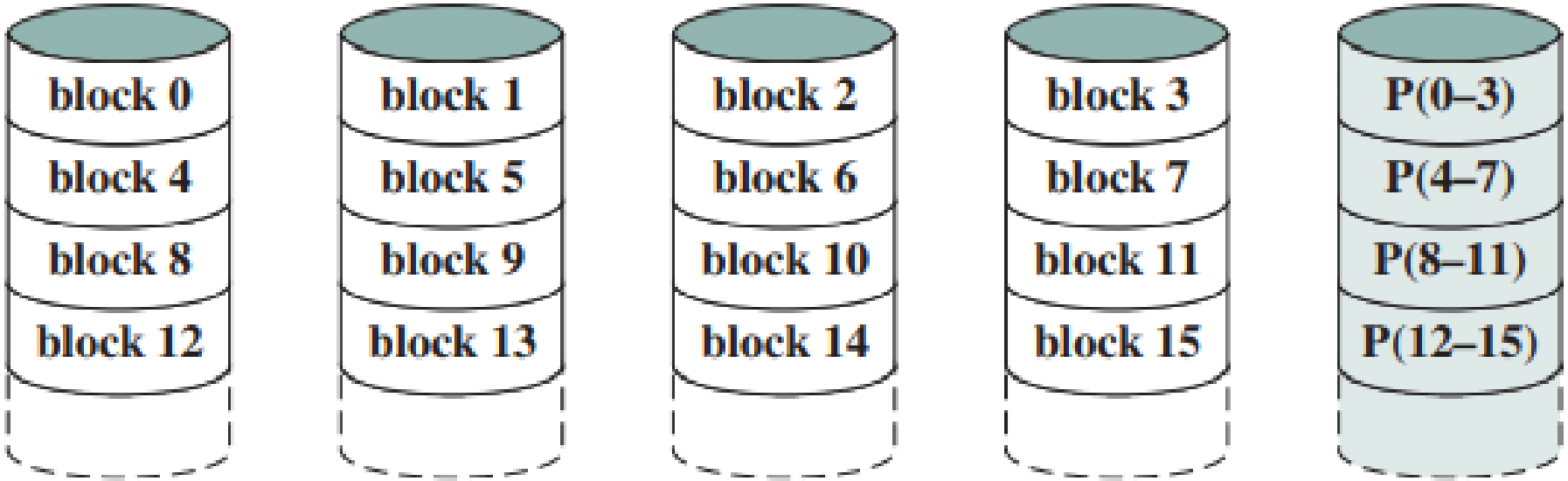
  what remains is missing data;



RAID 4

- Each disk operates independently
- Good for high I/O request rate
- Large stripes
- Bit by bit parity calculated across stripes on each disk
- Parity stored on parity disk

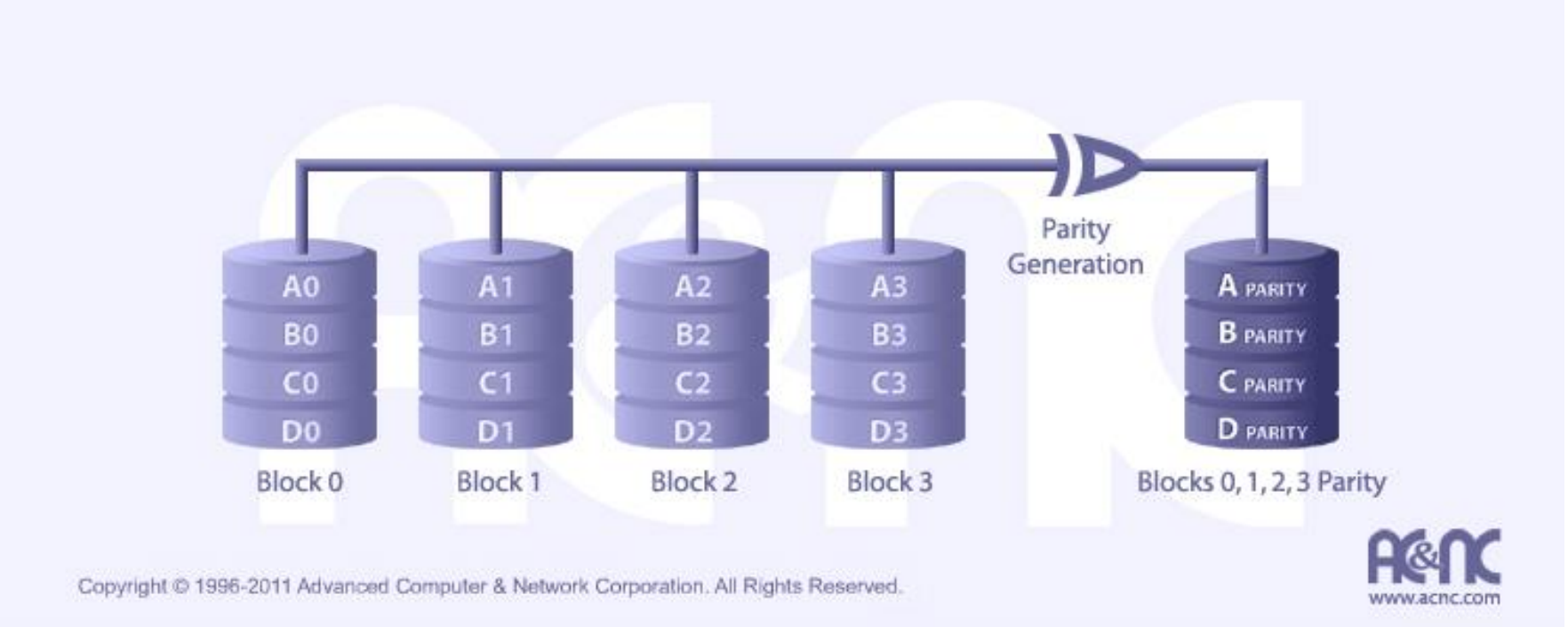
RAID Level 4



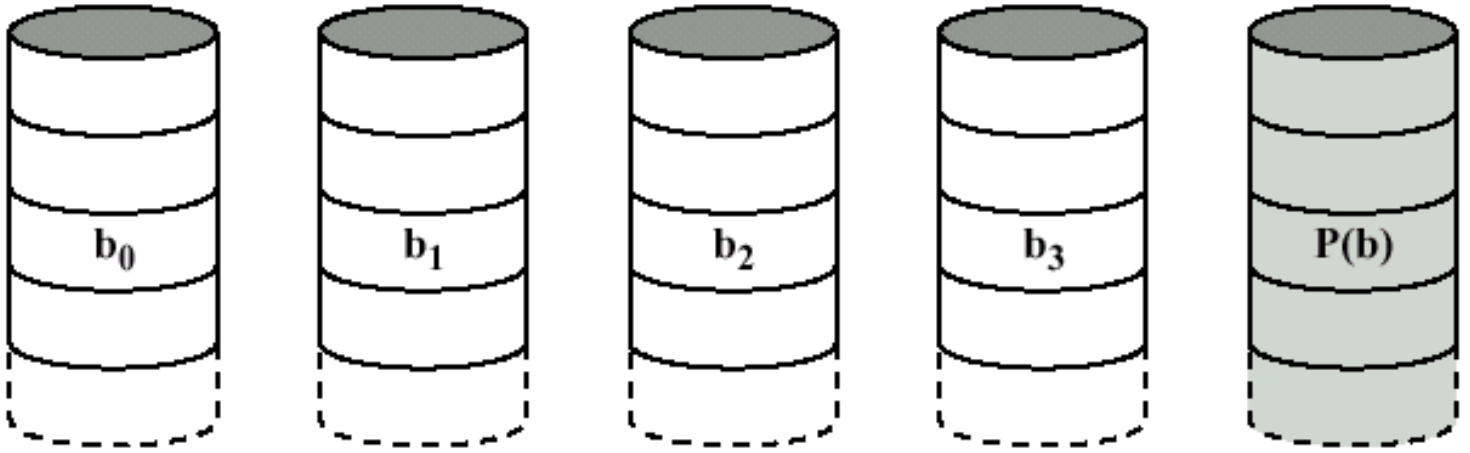
(e) RAID 4 (Block-level parity)

RAID 4

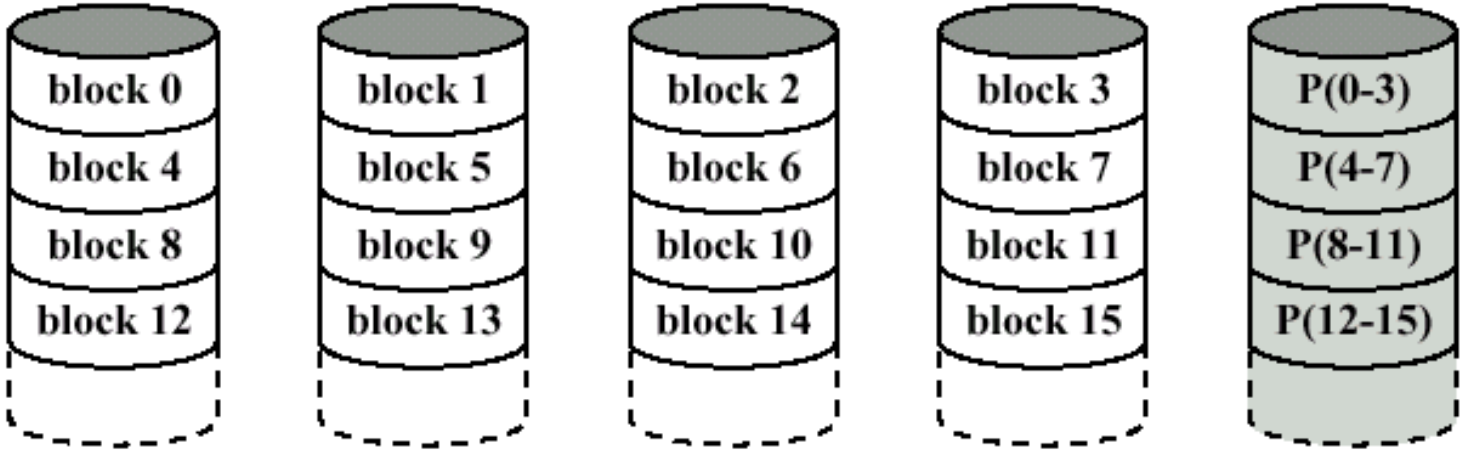
- http://www.acnc.com/raidedu/4
- Favor small accesses
- Allows each disk to perform independent reads, using sectors' own error checking



RAID Levels 3, 4



(d) RAID 3 (bit-interleaved parity)

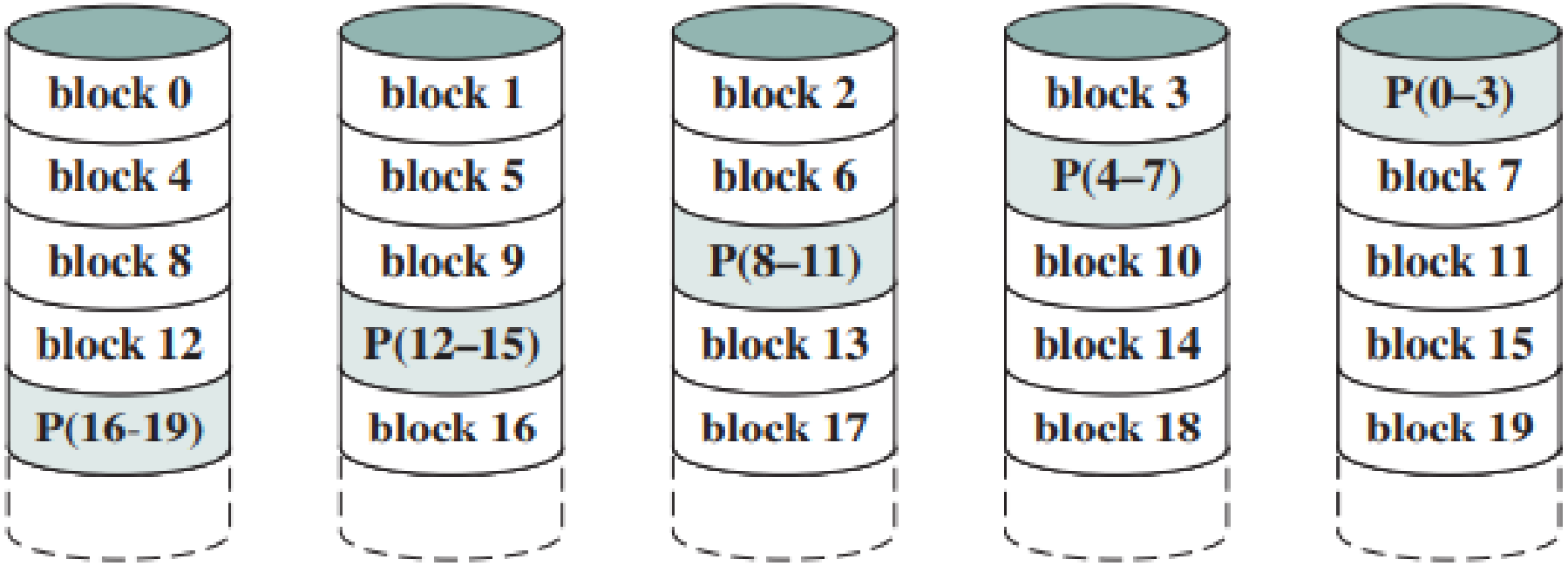


(e) RAID 4 (block-level parity)

RAID 5

- Like RAID 4
- Parity striped across all disks
- Round robin allocation for parity stripe
- Avoids RAID 4 bottleneck at parity disk
- Commonly used in network servers

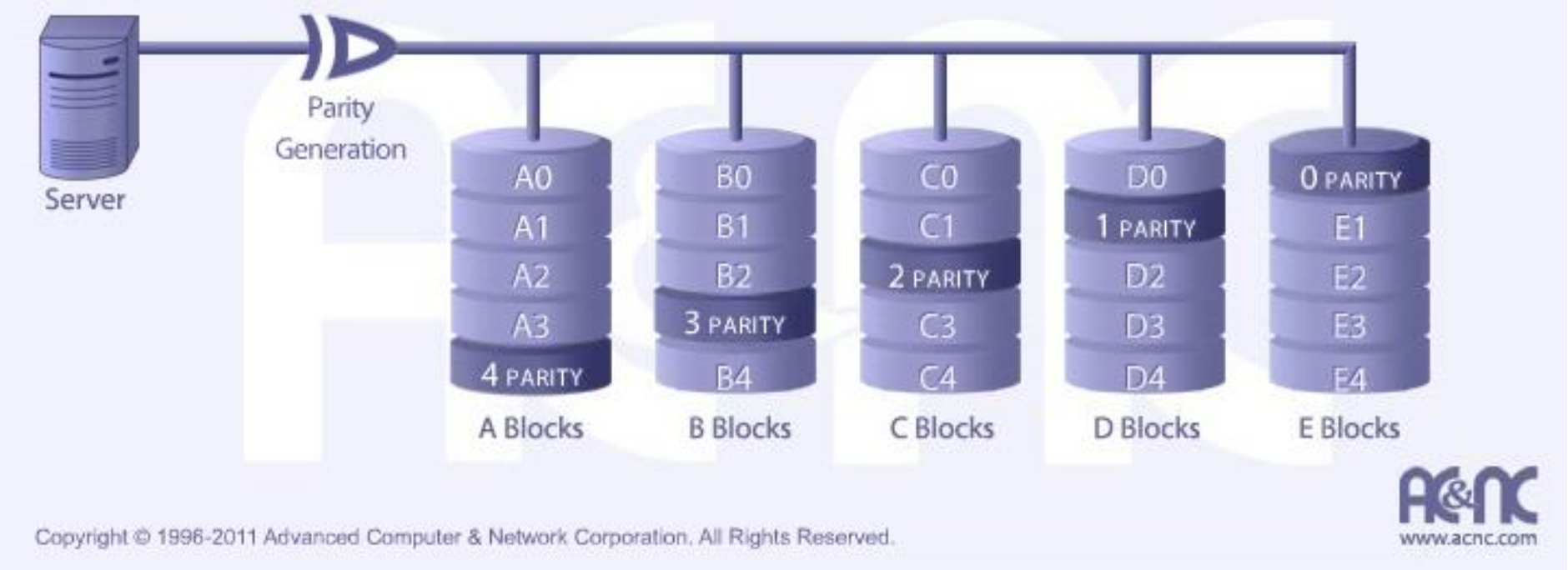
RAID Level 5



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

RAID 5

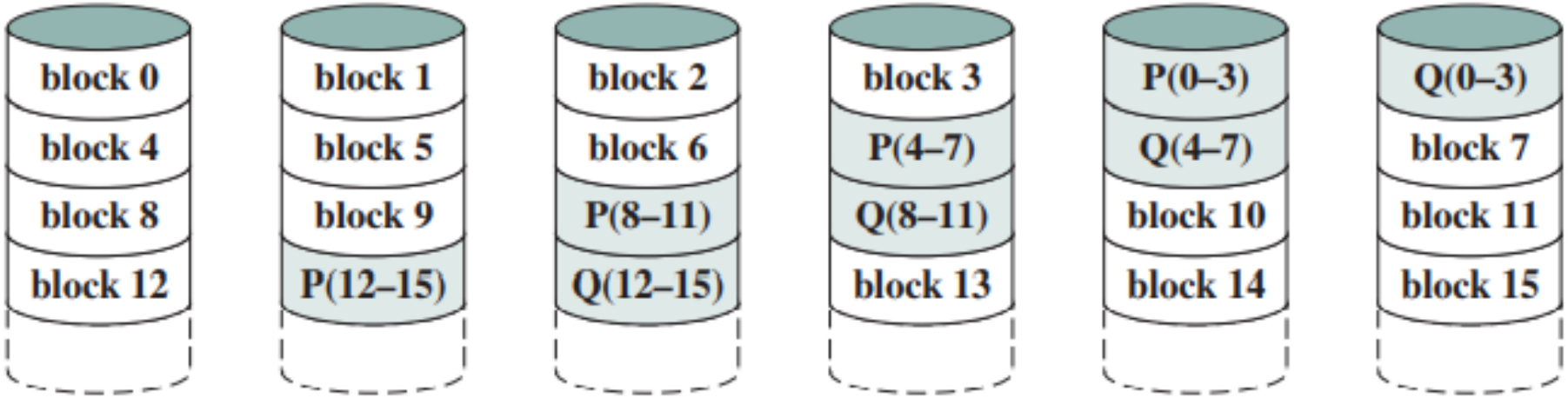
- http://www.acnc.com/raidedu/5
- Distributes the parity info across all disks in the array
- Removes the bottleneck of a single parity disk as RAID 3 and RAID 4



RAID 6

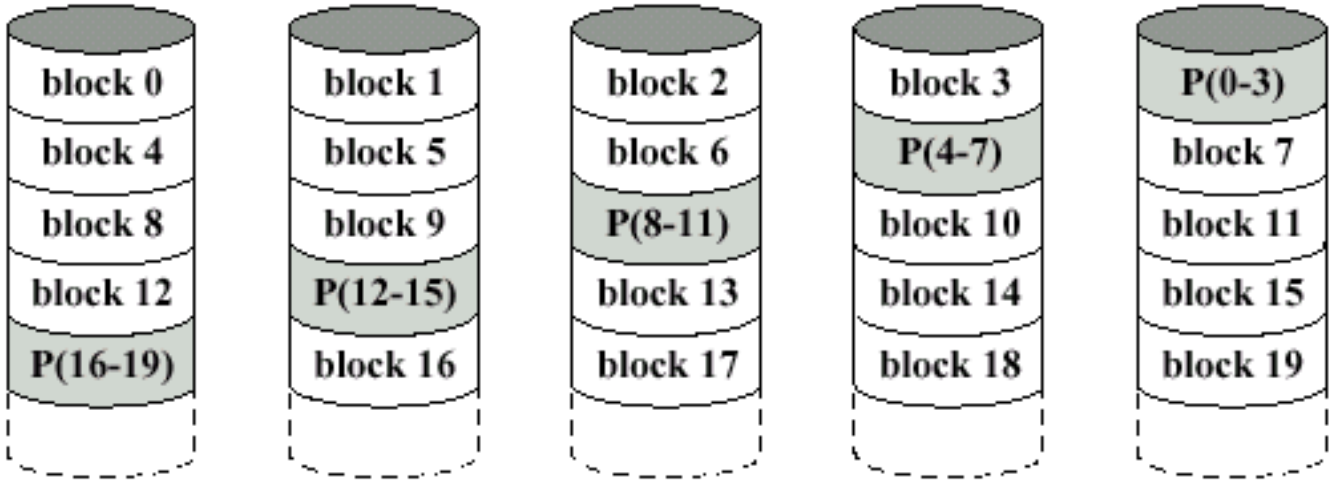
- Two different parity calculations are carried out and
- stored in separate blocks on different disks.
- Able to regenerate data even if two disks containing user data fail

RAID Level 6

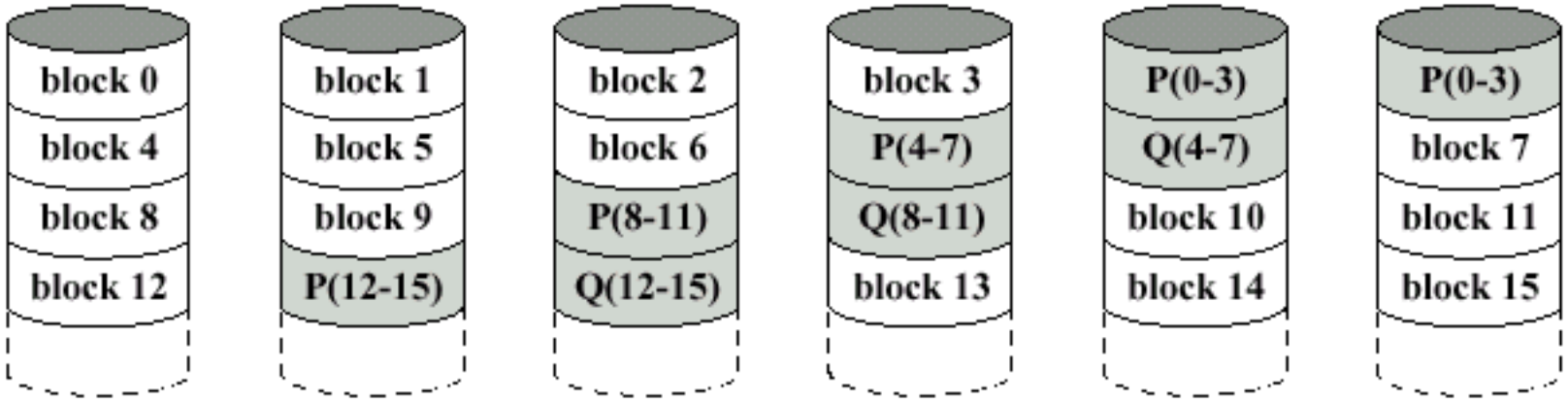


(g) RAID 6 (Dual redundancy)

RAID Levels 5, 6



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



(g) RAID 6 (dual redundancy)

Data Reconstruction RAID 3 to RAID 6

Data reconstruction is simple. Consider an array of five drives in which X0 through X3 contain data and X4 is the parity disk. The parity for the ith bit is calculated as follows:

$$X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

where $\oplus$ is exclusive-OR function.

Suppose that drive X1 has failed. If we add $X4(i) \oplus X1(i)$ to both sides of the preceding equation, we get

$$X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$$

Thus, the contents of each strip of data on X1 can be regenerated from the contents of the corresponding strips on the remaining disks in the array. This principle is true for RAID levels 3 through 6.

RAID

RAID level		Disk failures tolerated, check space overhead for 8 data disks	Pros	Cons	Company products
0	Nonredundant striped	0 failures, 0 check disks	No space overhead	No protection	Widely used
1	Mirrored	1 failure, 8 check disks	No parity calculation; fast recovery; small writes faster than higher RAIDs; fast reads	Highest check storage overhead	EMC, HP (Tandem), IBM
2	Memory-style ECC	1 failure, 4 check disks	Doesn't rely on failed disk to self-diagnose	~ Log 2 check storage overhead	Not used
3	Bit-interleaved parity	1 failure, 1 check disk	Low check overhead; high bandwidth for large reads or writes	No support for small, random reads or writes	Storage Concepts
4	Block-interleaved parity	1 failure, 1 check disk	Low check overhead; more bandwidth for small reads	Parity disk is small write bottleneck	Network Appliance
5	Block-interleaved distributed parity	1 failure, 1 check disk	Low check overhead; more bandwidth for small reads and writes	Small writes → 4 disk accesses	Widely used
6	Row-diagonal parity, EVEN-ODD	2 failures, 2 check disks	Protects against 2 disk failures	Small writes → 6 disk accesses; 2× check overhead	Network Appliance

RAID: Further Readings

- Raid Types – Classifications

  BytePile.com

  https://www.icc-usa.com/content/raid-calculator/raid-0-1.png
- RAID

  JetStor

  http://www.acnc.com/raidedu/0