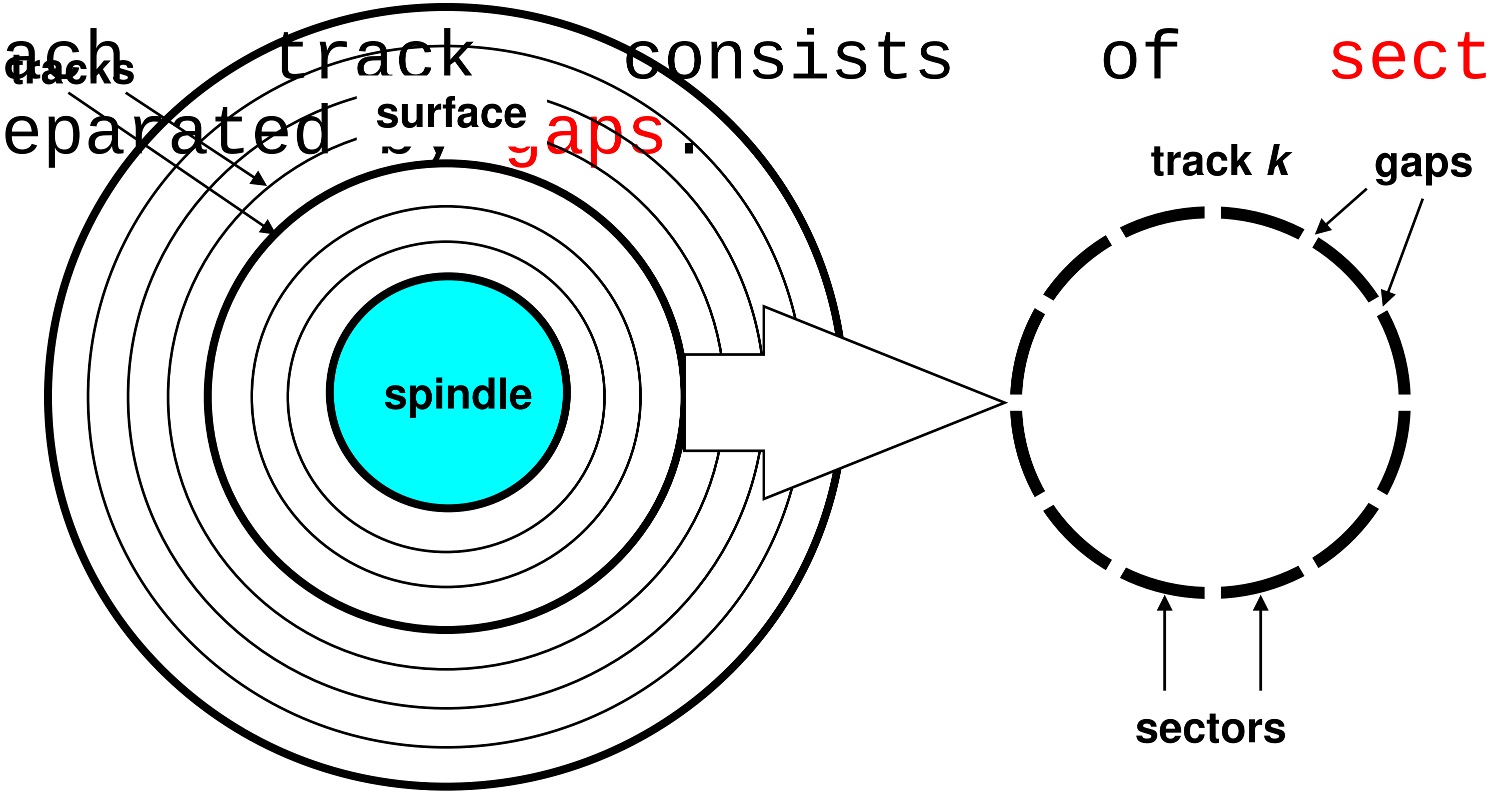
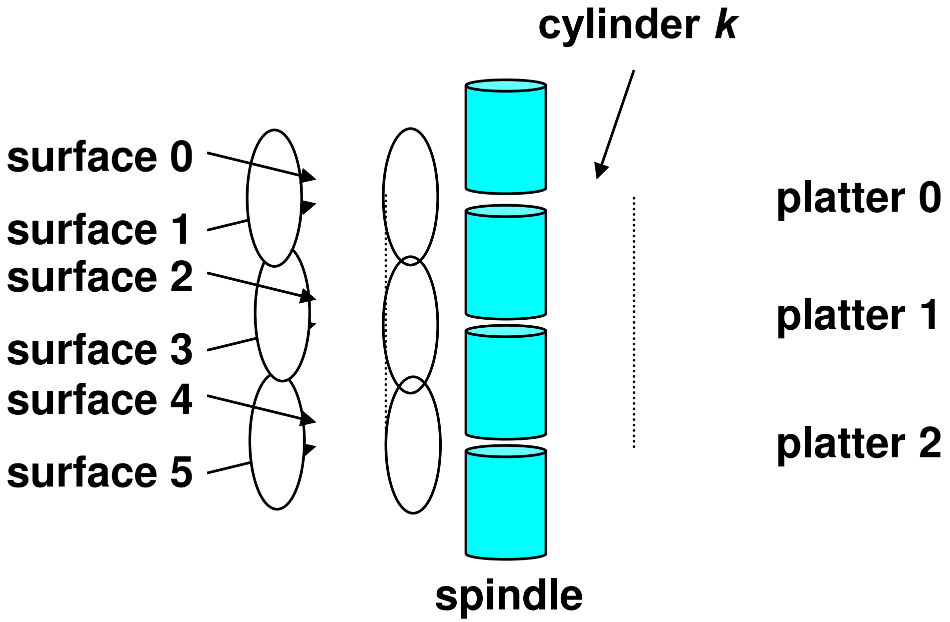
DISK GEOMETRY

- Disks consist of platters, each with two surfaces.
- Each surface consists of concentric rings called tracks.
- Each track consists of sectors separated by gaps.



DISK GEOMETRY (MULTIPLE-PLATTER VIEW)

- Aligned tracks form a cylinder.



DISK CAPACITY

- Capacity: maximum number of bits that can be stored.
  - Vendors express capacity in units of gigabytes (GB), where 1 GB = 10^9.
- Capacity is determined by these technology factors:
  - Recording density (bits/in): number of bits that can be squeezed into a 1 inch segment of a track.
  - Track density (tracks/in): number of tracks that can be squeezed into a 1 inch radial segment.
  - Areal density (bits/in2): product of recording and track density.

Cont…

- Modern disks partition tracks into disjoint subsets called recording zones
  - Each track in a zone has the same number of sectors, determined by the circumference of innermost track.
  - Each zone has a different number of sectors/track

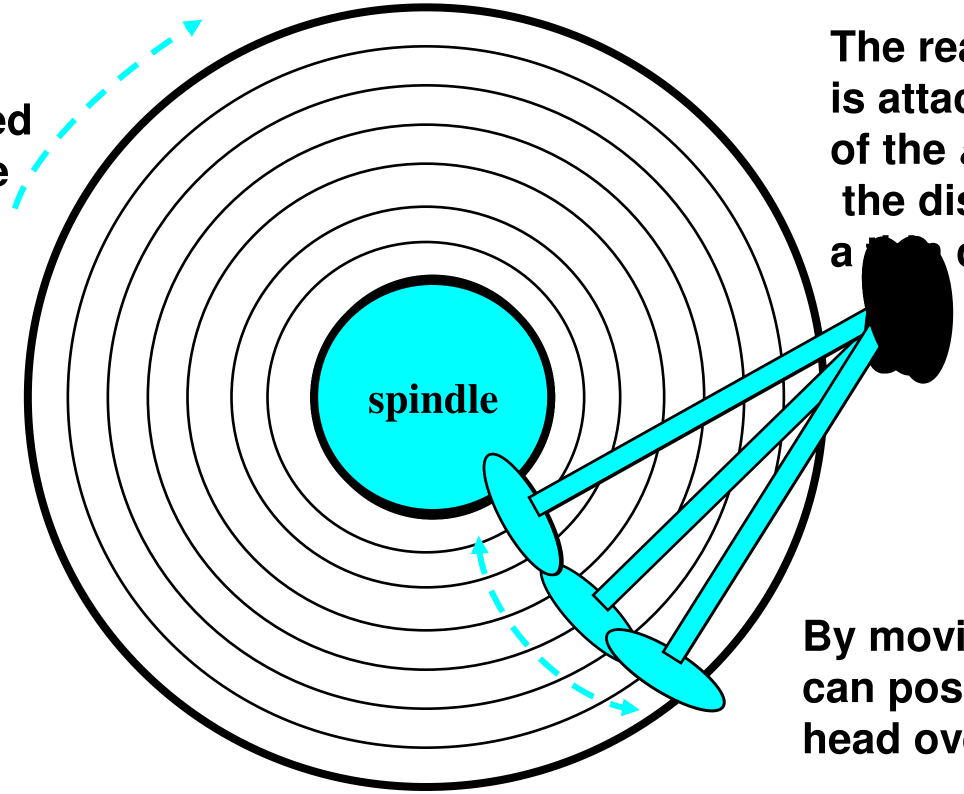
COMPUTING DISK CAPACITY

- Capacity = (# bytes/sector) x (avg. # sectors/track) x (# tracks/surface) x (# surfaces/platter) x (# platters/disk)
- Example:
  - 512 bytes/sector
  - 300 sectors/track (on average)
  - 20,000 tracks/surface
  - 2 surfaces/platter
  - 5 platters/disk
- Capacity = 512 x 300 x 20000 x 2 x 5

  = 30,720,000,000 = 30.72 GB

DISK OPERATION (SINGLE-PLATTER VIEW)

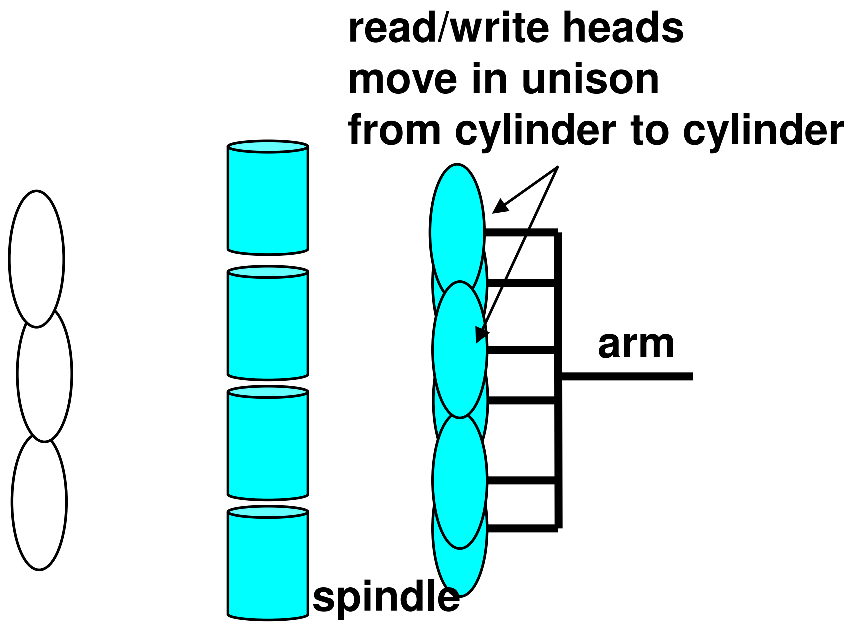
The disk
surface
spins at a fixed
rotational rate



The read/write *head*
is attached to the end
of the *arm* and flies over
the disk surface on
a [illegible] cushion of air.

By moving radially, the arm
can position the read/write
head over any track.

DISK OPERATION (MULTI-PLATTER VIEW)



DISK ACCESS TIME

- Average time to access some target sector approximated by :
  - $T_{access} = T_{avg\ seek} + T_{avg\ rotation} + T_{avg\ transfer}$
- Seek time ($T_{avg\ seek}$)
  - Time to position heads over cylinder containing target sector.
  - Typical $T_{avg\ seek}$ = 9 ms
- Rotational latency ($T_{avg\ rotation}$)
  - Time waiting for first bit of target sector to pass under r/w head.
  - $T_{avg\ rotation}$ = 1/2 x 1/RPMs x 60 sec/1 min

DISK ACCESS TIME

- Transfer time ($T_{avg\ transfer}$)
  - Time to read the bits in the target sector.
  - $T_{avg\ transfer}$ = 1/RPM x 1/(avg # sectors/track) x 60 secs/1 min.

DISK ACCESS TIME EXAMPLE

- Given:
  - Rotational rate = 7,200 RPM
  - Average seek time = 9 ms.
  - Avg # sectors/track = 400.
- Derived:
  - $T_{avg\ rotation}$ = 1/2 x (60 secs/7200 RPM) x 1000 ms/sec = 4 ms.
  - $T_{avg\ transfer}$ = 60/7200 RPM x 1/400 secs/track x 1000 ms/sec = 0.02 ms
  - T_{access} = 9 ms + 4 ms + 0.02 ms

DISK ACCESS TIME EXAMPLE

- Important points:
  - Access time dominated by seek time and rotational latency.
  - First bit in a sector is the most expensive, the rest are free.
  - SRAM access time is about 4 ns/double word, DRAM about 60 ns
    - Disk is about 40,000 times slower than SRAM,
    - 2,500 times slower then DRAM.

Types of External Memory

- Magnetic Disk
  - RAID
  - Removable
- Optical
  - CD-ROM
  - CD-Recordable (CD-R)
  - CD-R/W
  - DVD
- Magnetic Tape

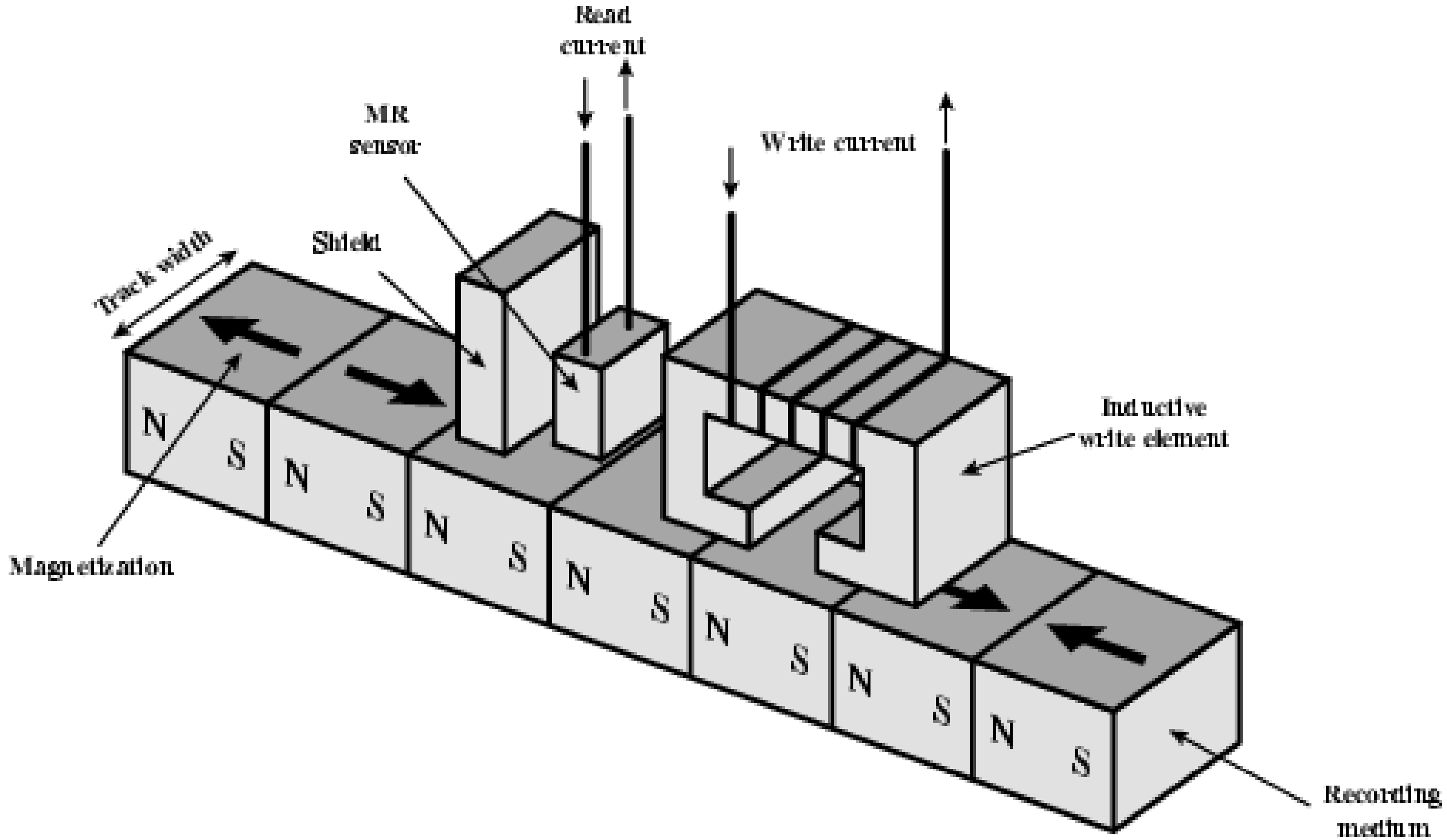
Magnetic Disk

- Disk substrate coated with magnetizable material (iron oxide…rust)
- Substrate used to be aluminium
- Now glass
  - Improved surface uniformity
    - Increases reliability
  - Reduction in surface defects
    - Reduced read/write errors
  - Lower flight heights (See later)
  - Better stiffness
  - Better shock/damage resistance

Read and Write Mechanisms

- Recording and retrieval via conductive coil called a head
- May be single read/write head or separate ones
- During read/write, head is stationary, platter rotates
- Write
  - Current through coil produces magnetic field
  - Pulses sent to head
  - Magnetic pattern recorded on surface below
- Read (traditional)
  - Magnetic field moving relative to coil produces current
  - Coil is the same for read and write
- Read (contemporary)
  - Separate read head, close to write head
  - Partially shielded magneto resistive (MR) sensor
  - Electrical resistance depends on direction of magnetic field
  - High frequency operation
    - Higher storage density and speed

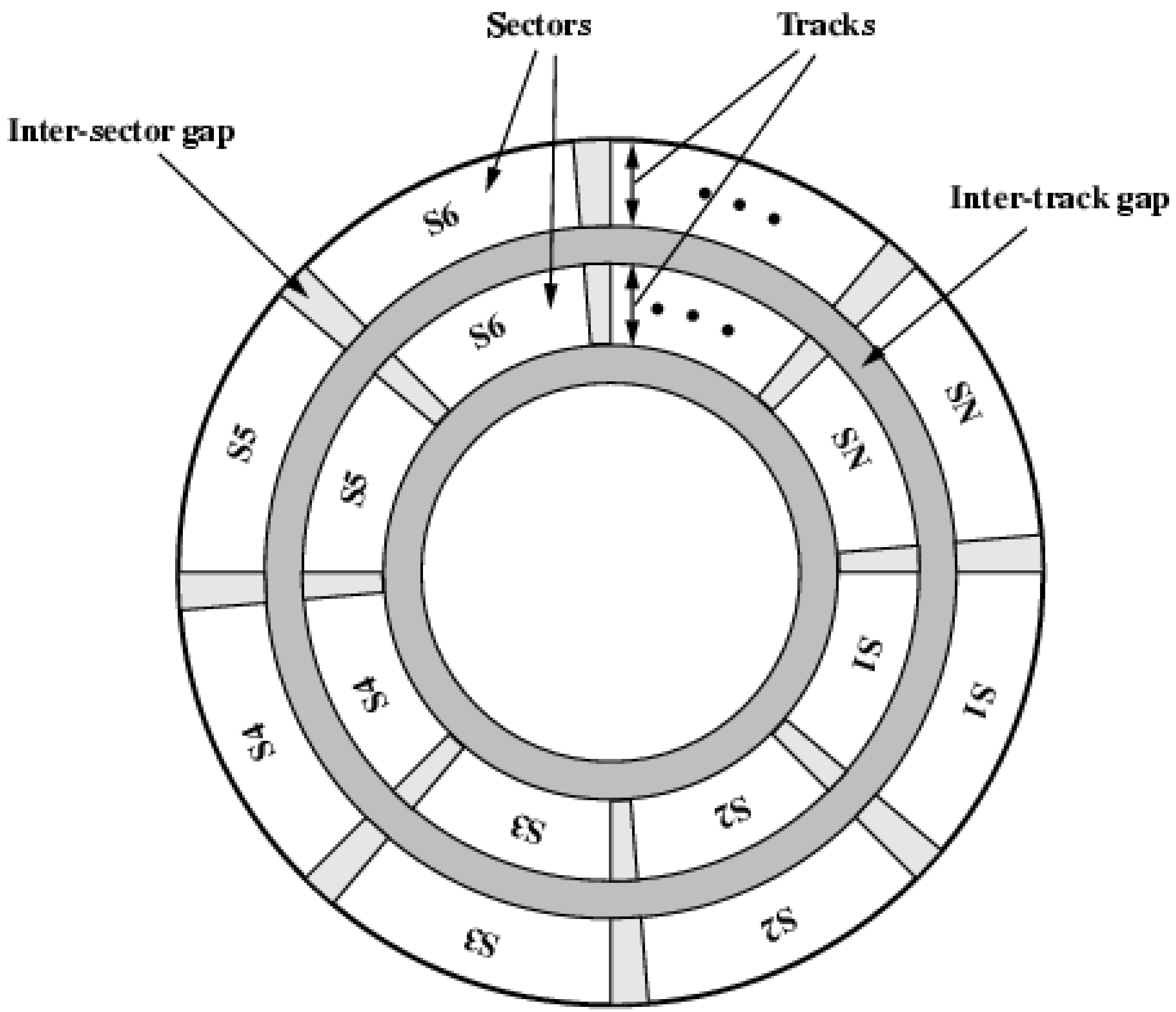
Inductive Write MR Read



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

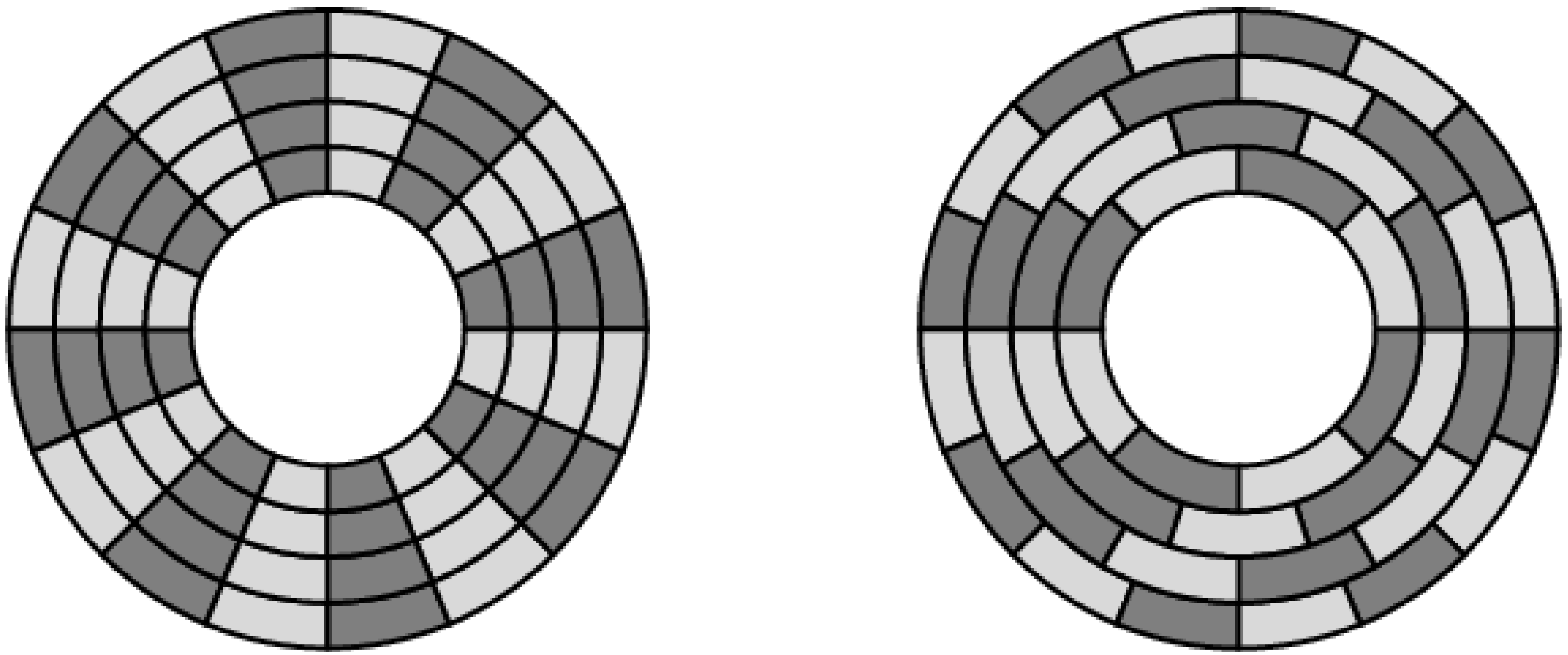
Disk Data Layout



Disk Velocity

- Bit near centre of rotating disk passes fixed point slower than bit on outside of disk
- Increase spacing between bits in different tracks
- Rotate disk at constant angular velocity (CAV)
  - Gives pie shaped sectors and concentric tracks
  - Individual tracks and sectors addressable
  - Move head to given track and wait for given sector
  - Waste of space on outer tracks
    - Lower data density
- Can use zones to increase capacity
  - Each zone has fixed bits per track
  - More complex circuitry

Disk Layout Methods Diagram



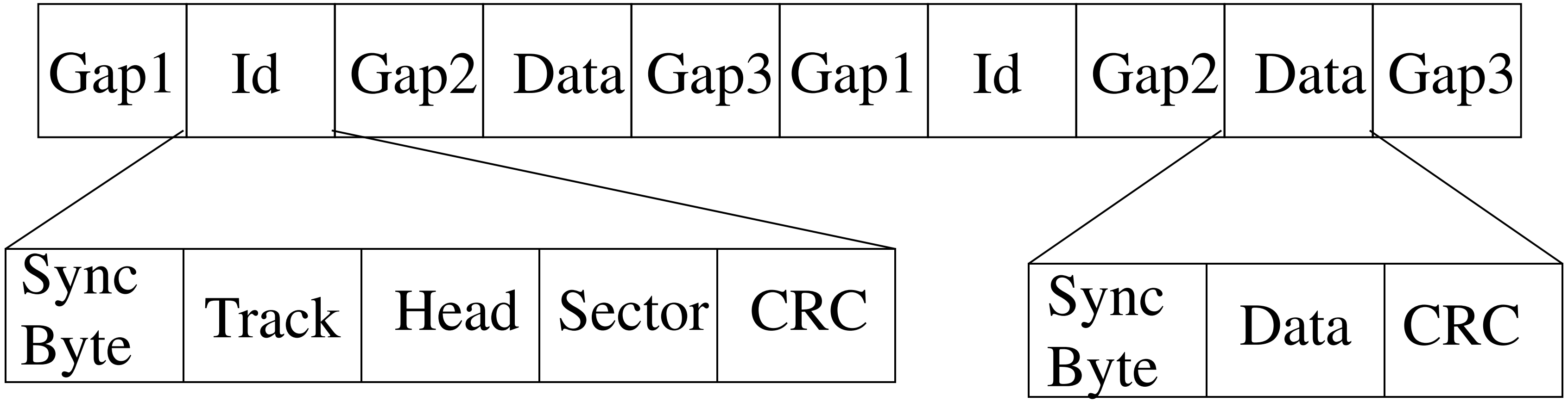
(a) Constant angular velocity

(b) Multiple zoned recording

Finding Sectors

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

ST506 format (old!)



- Foreground reading
  - Find others

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)

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

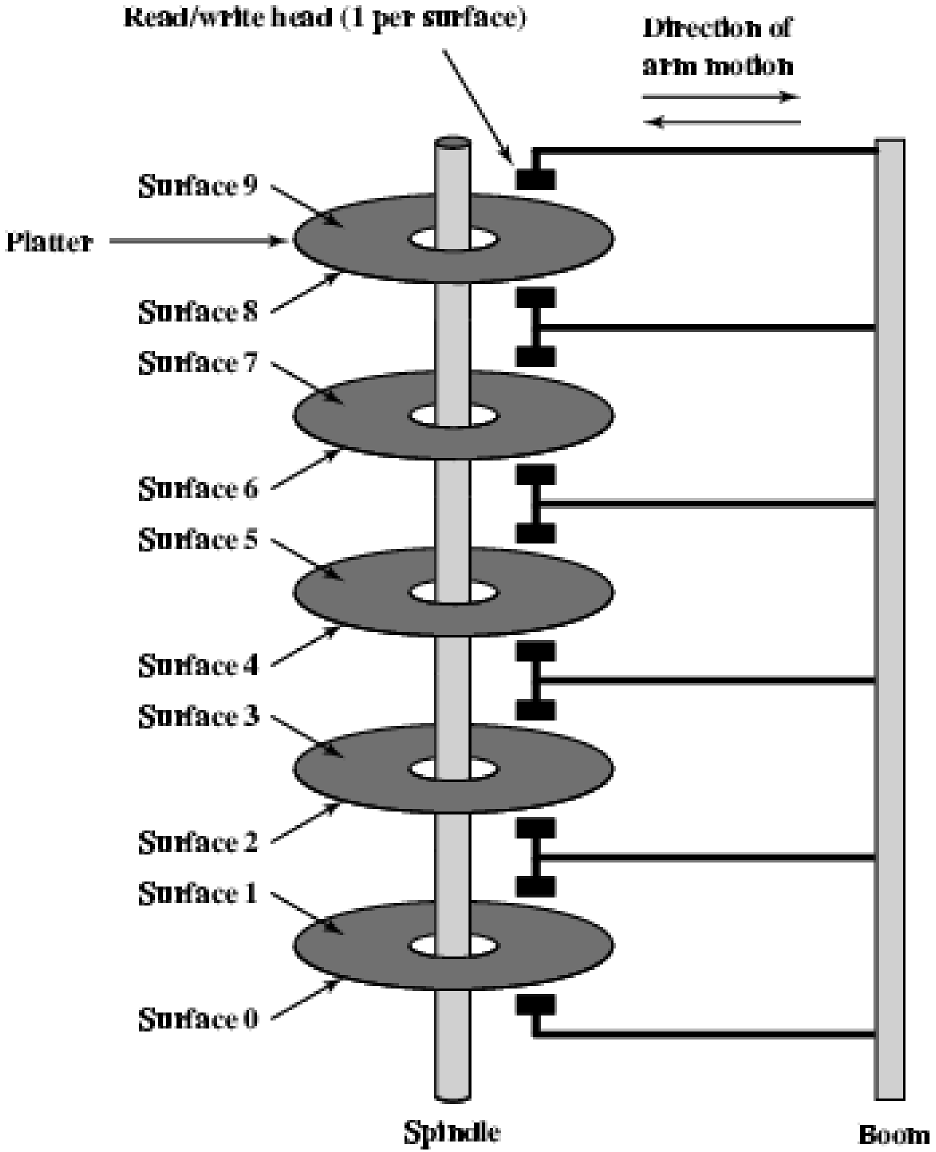
Removable or Not

- Removable disk
  - Can be removed from drive and replaced with another disk
  - Provides unlimited storage capacity
  - Easy data transfer between systems
- Nonremovable disk
  - Permanently mounted in the drive

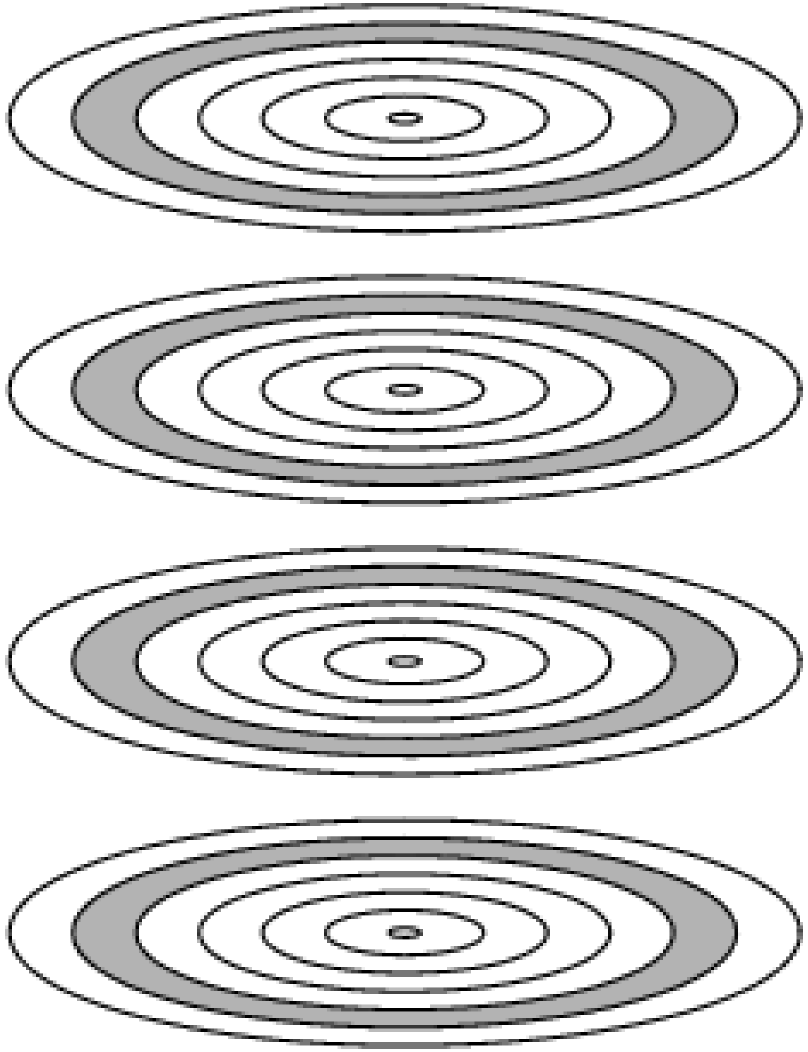
Multiple Platter

- 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)

Multiple Platters



Cylinders



Floppy Disk

- 8”, 5.25”, 3.5”
- Small capacity
  - Up to 1.44Mbyte (2.88M never popular)
- Slow
- Universal
- Cheap
- Obsolete?

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
- Getting larger all the time
  - Multiple Gigabyte now usual

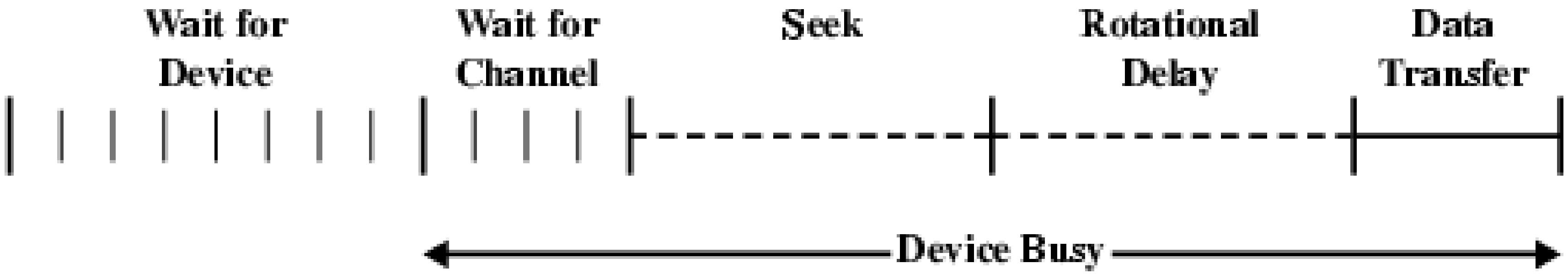
Removable Hard Disk

- ZIP
  - Cheap
  - Very common
  - Only 100M
- JAZ
  - Not cheap
  - 1G
- L-120 (a: drive)
  - Also reads 3.5″ floppy
  - Becoming more popular?
- All obsoleted by CD-R and CD-R/W?

Speed

- Seek time
  - Moving head to correct track
- (Rotational) latency
  - Waiting for data to rotate under head
- Access time = Seek + Latency
- Transfer rate
- See timing comparison on p.173
  - Sequential organization = .064 seconds to read 1.28 MB
  - Random organization = 4.008 seconds to 1.28 MB
  - Why periodic defragmentation is important!

Timing of Disk I/O Transfer



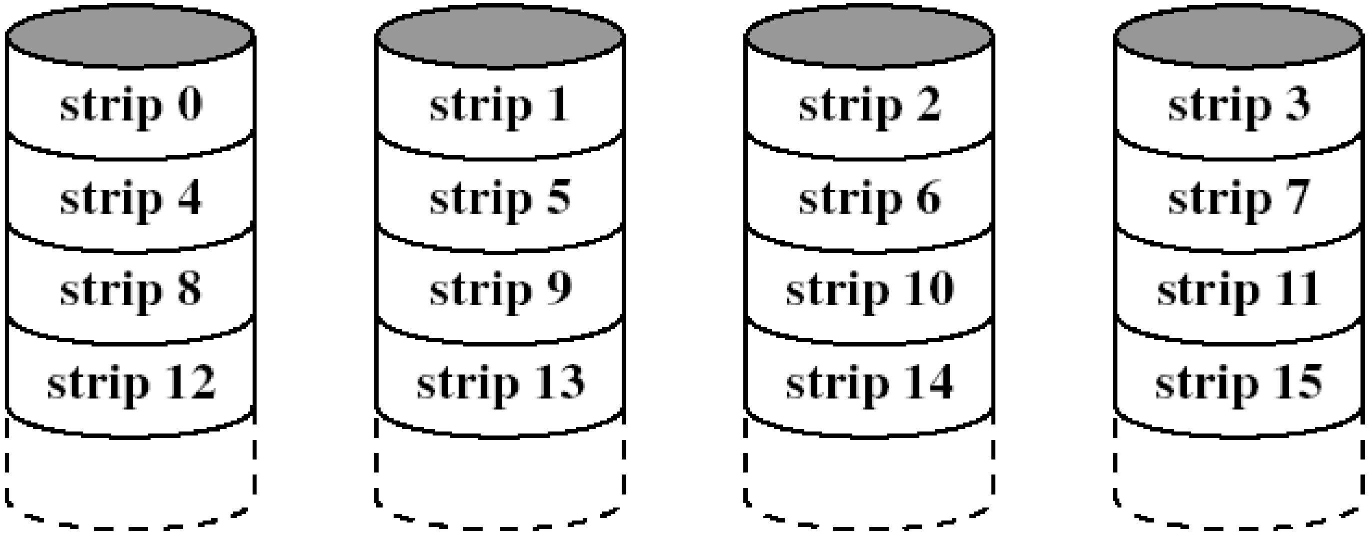
RAID

- Redundant Array of Independent Disks
- Redundant Array of Inexpensive Disks
- 6 levels in common use
- Not a hierarchy—but different design architectures with 3 common characteristics:
  - Set of physical disks viewed as single logical drive by O/S
  - Data distributed across physical drives of an array
  - Can use redundant capacity to store parity information, which guarantees data recoverability in case of a disk failure (except RAID 0)

RAID 0

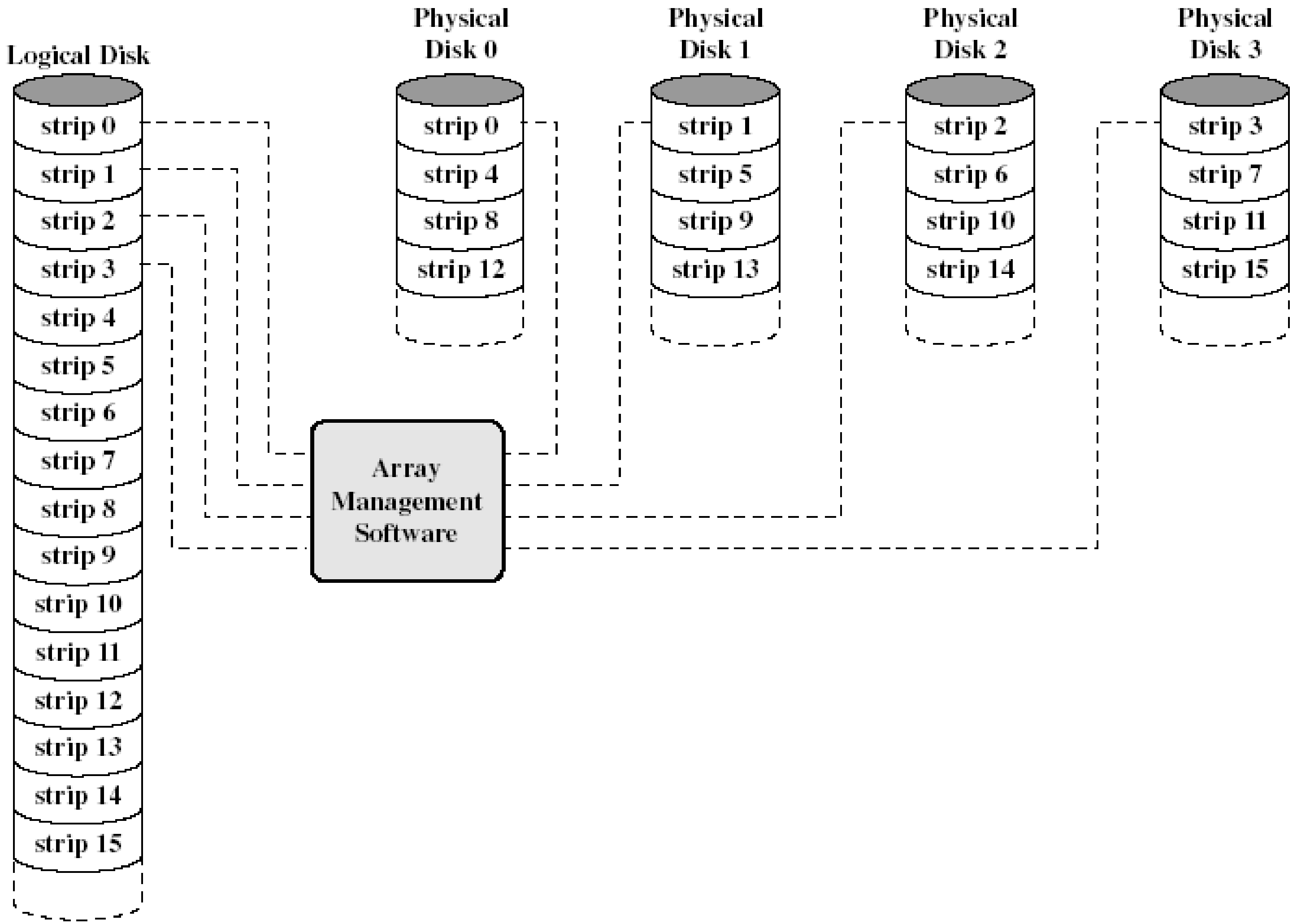
- No redundancy
- Data striped across all disks
- Round Robin striping
- Increase speed
  - Multiple data requests probably not on same disk
  - Disks seek in parallel
  - A set of data is likely to be striped across multiple disks
- **disk striping:** The procedure of combining a set of same-size disk partitions that reside on separate disks (from 2 to 32 disks) into a single volume, forming a virtual "stripe" across the disks that the operating system recognizes as a single drive. Disk striping enables multiple I/O operations in the same volume to proceed concurrently, thus offering enhanced performance. Microsoft Press« Computer and Internet Dictionary, 4th Edition,

RAID 0 Striping



(a) RAID 0 (non-redundant)

RAID 0 Logical Disk



RAID 1

- Mirrored Disks
- Data is striped across disks
- 2 copies of each stripe on separate disks
- Read from either to speed up reads
- Write to both at the same time in parallel
- Recovery is simple
  - Swap faulty disk & re-mirror
  - No down time
- Expensive—two complete sets of drives

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 commercially available

RAID 3

- Similar to RAID 2
- Only *one* redundant parity 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 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
- Not commercially available

RAID 5

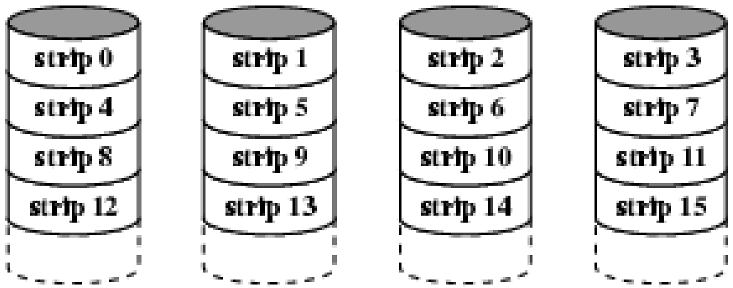
- Like RAID 4
- Parity striped across all disks
- Round robin allocation for parity stripe
- Avoids RAID 4 bottleneck at parity disk
- Commonly used in network servers

- N.B. DOES NOT MEAN 5 DISKS!!!!!

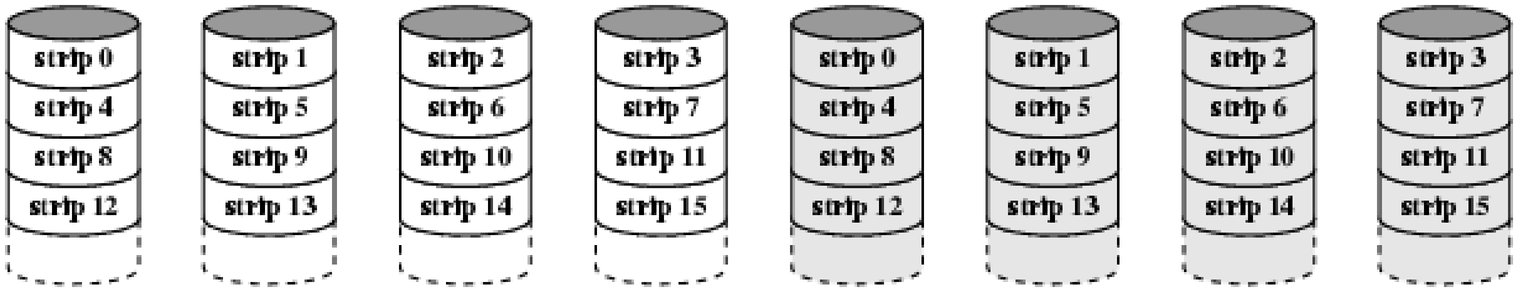
RAID 6

- Two parity calculations
- Stored in separate blocks on different disks
- User requirement of N disks needs N+2
- High data availability
  - Three disks need to fail for data loss
  - Significant write penalty

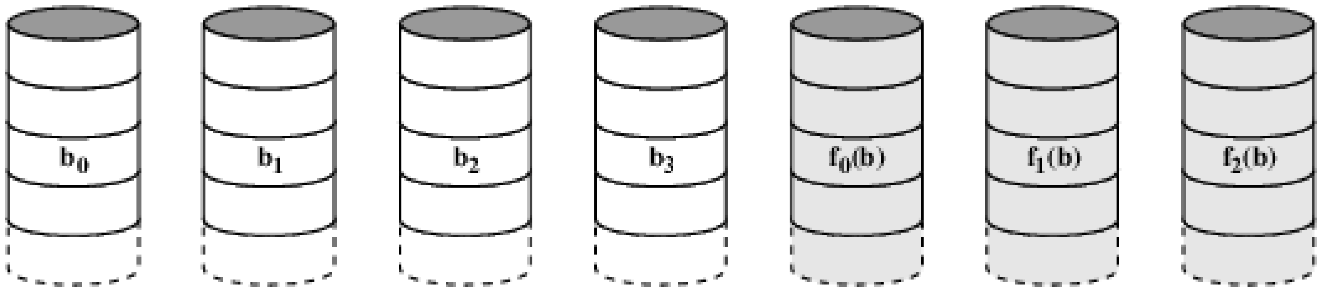
RAID 0, 1, 2



(a) RAID 0 (non-redundant)

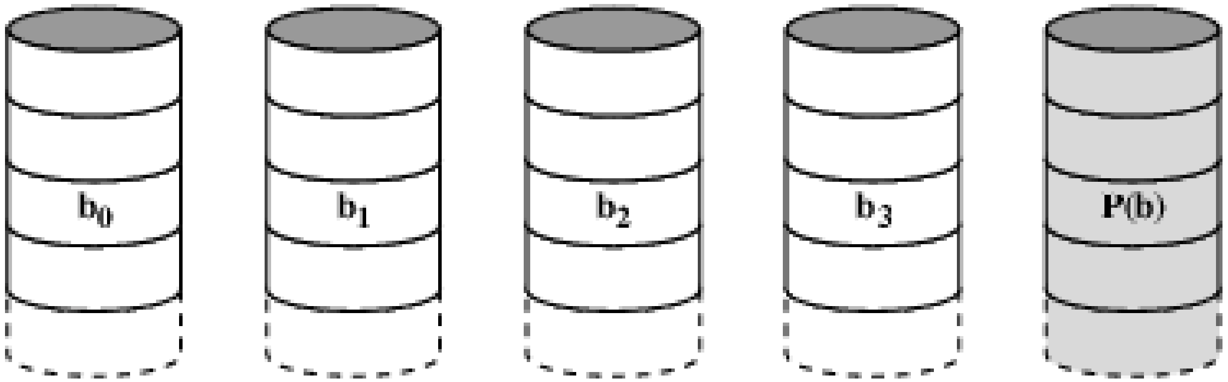


(b) RAID 1 (mirrored)

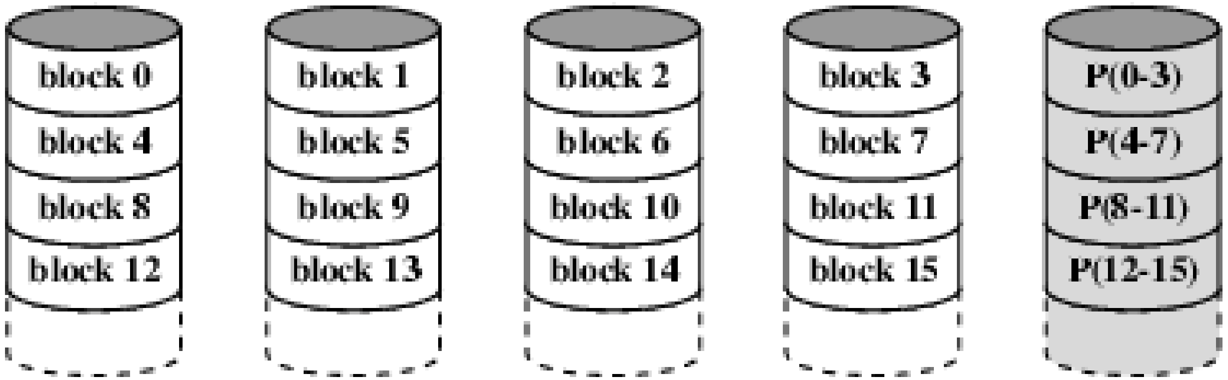


(c) RAID 2 (redundancy through Hamming code)

RAID 3 & 4

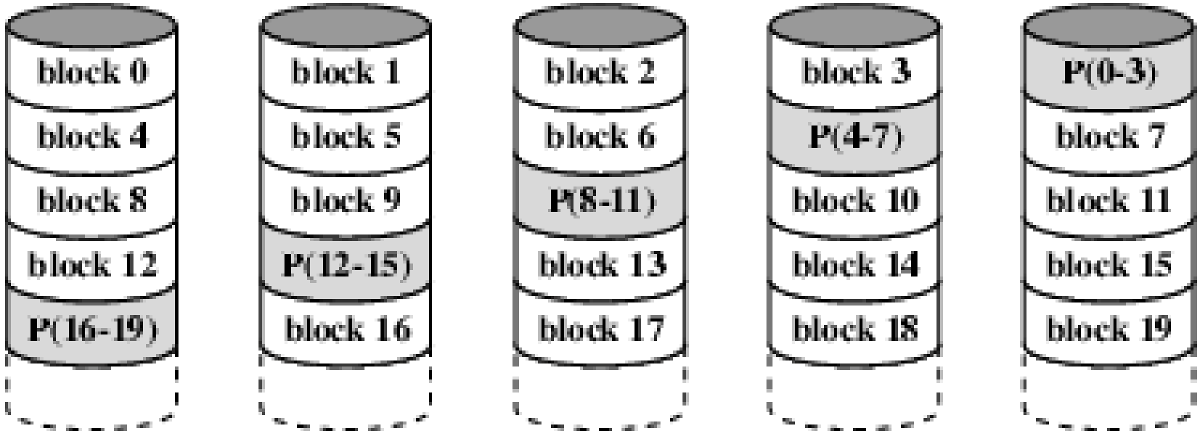


(d) RAID 3 (bit-interleaved parity)

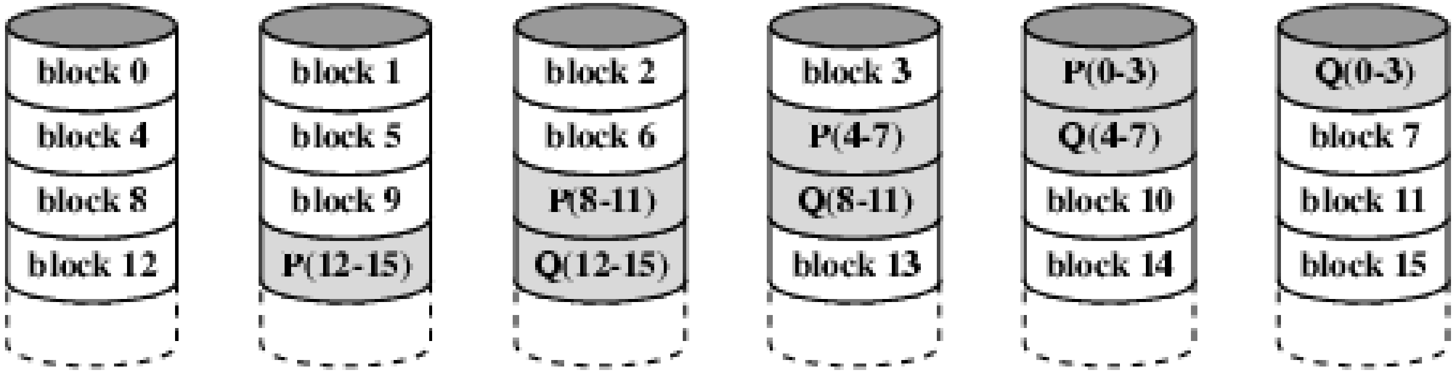


(e) RAID 4 (block-level parity)

RAID 5 & 6



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



(g) RAID 6 (dual redundancy)

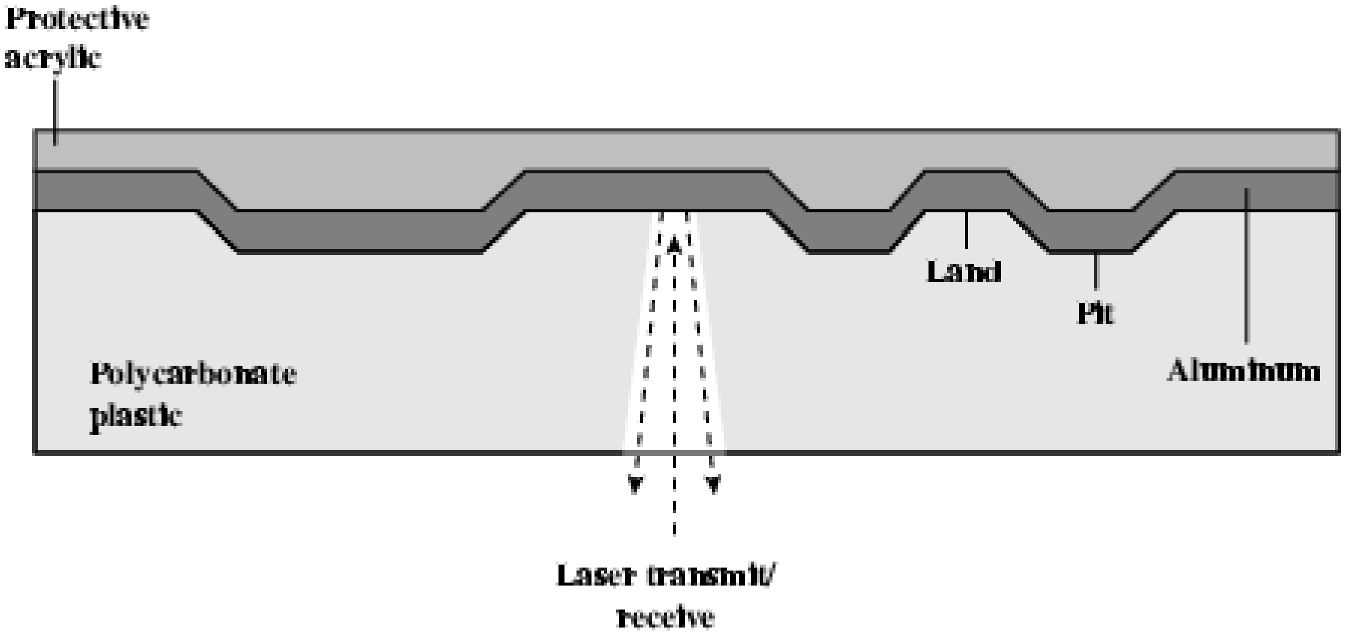
RAID Levels

Category	Level	Description	I/O Request Rate (Read/Write)	Data Transfer Rate (Read/Write)	Typical Application
Striping	0	Non-redundant	Large strips: Excellent	Small strips: Excellent	Applications requiring high performance for non-critical data
Mirroring	1	Mirrored	Good/Fair	Fair/Fair	System drives; critical files
Parallel access	2	Redundant via Hamming code	Poor	Excellent	
	3	Bit-interleaved parity	Poor	Excellent	Large I/O request size applications, such as imaging, CAD
Independent access	4	Block-interleaved parity	Excellent/Fair	Fair/Poor	
	5	Block-interleaved distributed parity	Excellent/Fair	Fair/Poor	High request rate, read-intensive, data lookup
	6	Block-interleaved dual distributed parity	Excellent/Poor	Fair/Poor	Applications requiring extremely high availablity

Optical Storage CD-ROM

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

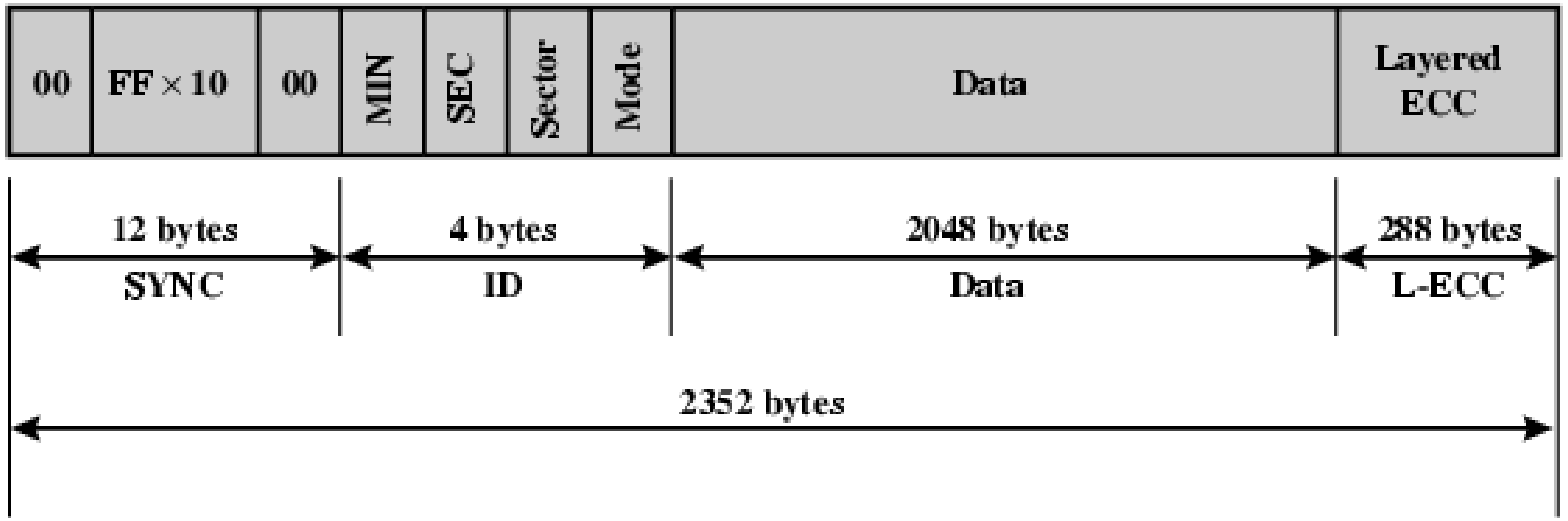
CD Operation



CD-ROM Drive Speeds

- Audio is single speed
  - Constant linier velocity
  - 1.2 ms^{-1}
  - Track (spiral) is 5.27km long
  - Gives 4391 seconds = 73.2 minutes
- Other speeds are quoted as multiples
- e.g. 24x
- Quoted figure is maximum drive can achieve

CD-ROM Format



- 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
- (Yawn!)

CD-ROM for & against

- Large capacity (?)
- Easy to mass produce
- Removable
- Robust

- Expensive for small runs
- Slow
- Read only

Other Optical Storage

- CD-Recordable (CD-R)
  - WORM
  - Now affordable
  - Compatible with CD-ROM drives
- CD-RW
  - Erasable
  - Getting cheaper
  - Mostly CD-ROM drive compatible
  - Phase change
    - Material has two different reflectivities in different phase states

DVD - what’s in a name?

- 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
- Dogs Veritable Dinner
- Officially - nothing!!!

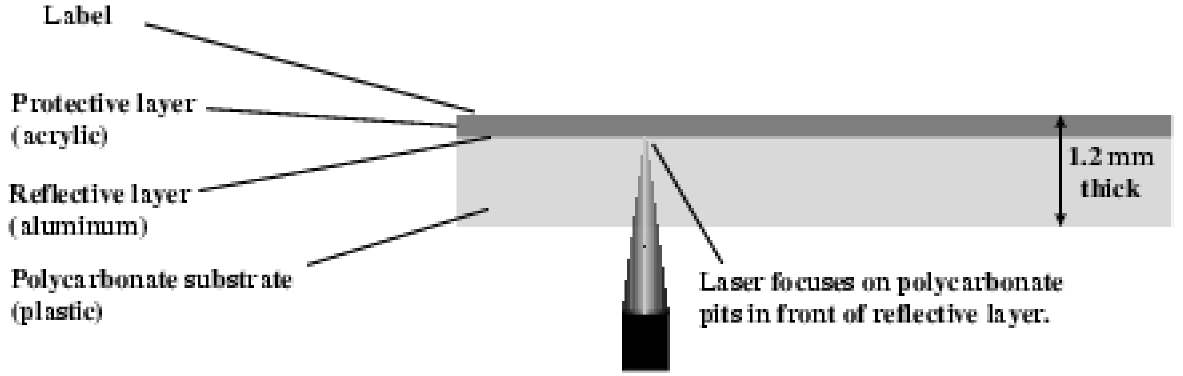
DVD - technology

- Multi-layer
- Very high capacity (4.7G per layer)
- Full length movie on single disk
  - Using MPEG compression
- Finally standardized (honest!)
- Movies carry regional coding
- Players only play correct region films
- Can be “fixed”

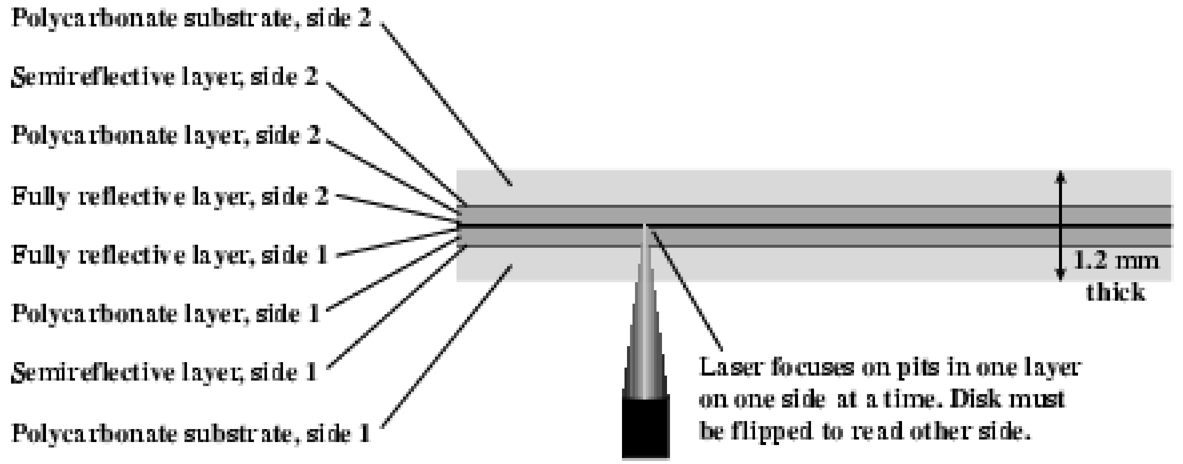
DVD – Writable

- Loads of trouble with standards
- First generation DVD drives may not read first generation DVD-W disks
- First generation DVD drives may not read CD-RW disks
- Wait for it to settle down before buying!

CD and DVD



(a) CD-ROM - Capacity 682 MB



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

Magnetic Tape

- Serial access
- Slow
- Very cheap
- Backup and archive

Digital Audio Tape (DAT)

- Uses rotating head (like video)
- High capacity on small tape
  - 4Gbyte uncompressed
  - 8Gbyte compressed
- Backup of PC/network servers