LAU
الجامعة اللبنانية الأميركية
Lebanese American University

External Memory

Chapter 7

Based on:
William Stallings
Computer Organization and Architecture, 11th Global Edition

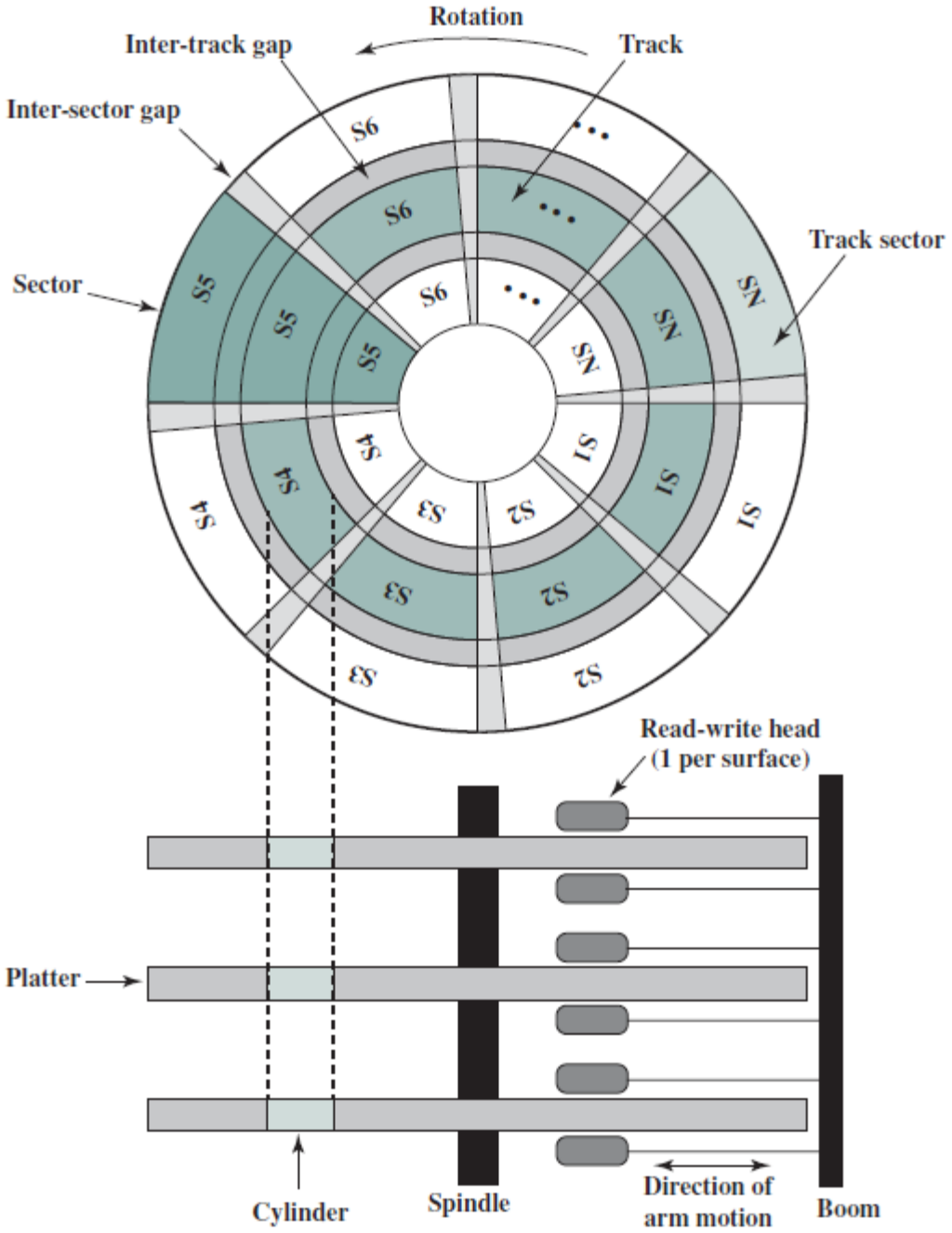
Storage

- Files (programs, data, settings)
- Virtual memory
- Performance
  - Throughput (improving, not as quickly as processor speed)
  - Latency (improving but very slowly)
- Reliability
- Very diverse types of storage
  - Magnetic disks
  - Optical disks
  - Magnetic tapes

Magnetic Disk

- Examples: hard disk (hard drive), floppy disks
- Circular platter constructed of nonmagnetic material, called the substrate, coated with a magnetizable material
- Substrate used to be aluminium
- Now glass
  - Improvement in the uniformity of the magnetic film surface to increase disk reliability
  - Significant reduction in overall surface defects to help reduce read-write errors
  - Better stiffness to reduce disk dynamics
  - Greater ability to withstand shock and damage

Disk Data Layout



Magnetic Read and Write Mechanisms

- Recording & 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

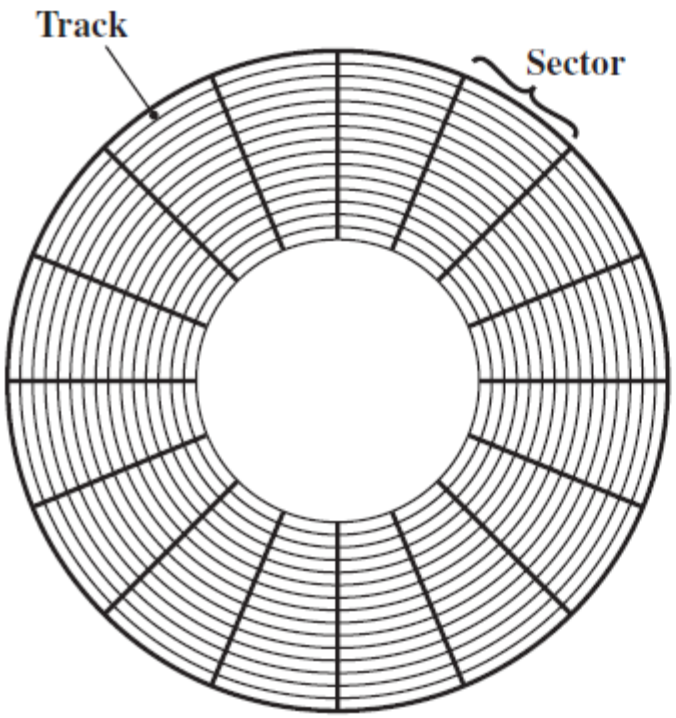
Data Organization and Formatting

- Concentric rings or tracks
  - Gaps between tracks
    - Prevent, or at least minimize, errors due to misalignment of the head or interference of magnetic fields
  - Reduce gap to increase capacity
  - Same number of bits per track (variable packing density)
- Tracks divided into sectors
- Minimum block size is one sector

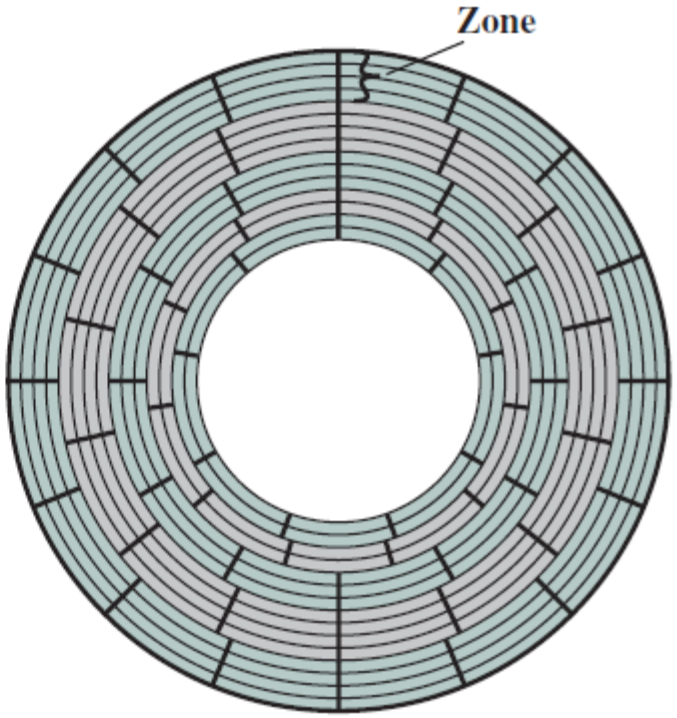
Disk Velocity

- Bit near center of rotating disk passes fixed point (such as a read-write head) slower than bit on outside of disk
- Find a way to compensate for the variation in speed so that the head can read all data bits at the same rate
- 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
  - Surface is divided into a number of concentric zones
  - Each zone contains a number of contiguous tracks
  - Within a zone, the number of bits per track is constant
  - More complex circuitry

Comparison of Disk Layout Methods



(a) Constant angular velocity



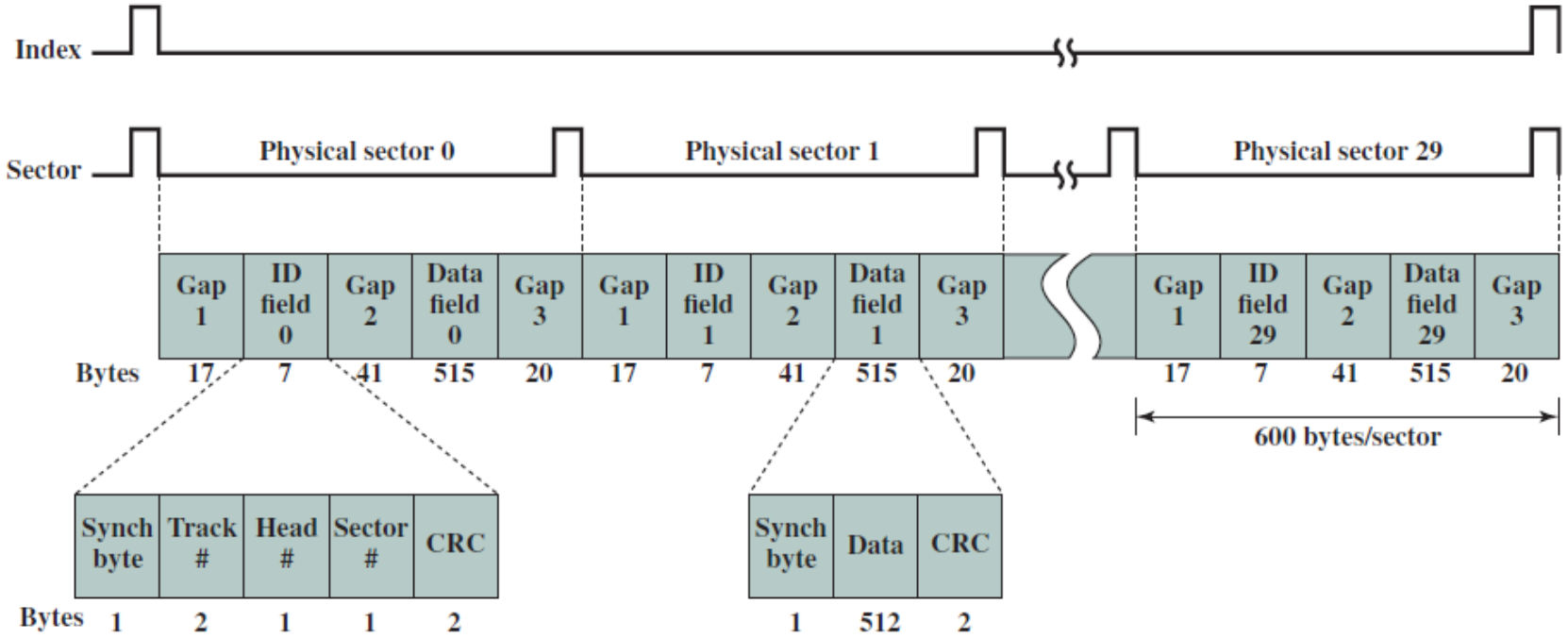
(b) Multiple zone recording

Zones farther from the center contain more bits (more sectors) than zones closer to the center

- 15 tracks organized into 5 zones
- innermost 2 zones have 2 tracks each, each track having 9 sectors
- next zone has 3 tracks, each with 12 sectors
- outermost 2 zones have 4 tracks each, each track having 16 sectors

Finding Sectors

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



Winchester Disk Format
(Seagate ST506)

Physical Characteristics of Disk Systems

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
Sides	Aerodynamic gap (Winchester)
Single sided	
Double sided	

Characteristics

- Fixed-head disk
  - One read-write head per track
  - Heads mounted on fixed ridged arm that extends across all tracks
- Moveable-head disk
  - One read-write head per side
  - Mounted on a movable arm (can be extended or retracted)
- Non-removable disk
  - Permanently mounted in the disk drive
  - E.g. Hard disk in a personal computer
- Removable disk
  - Can be removed and replaced with another disk
  - Provides unlimited storage capacity
  - Easy data transfer between systems
  - E.g. floppy disks
- Double sided disk
  - Magnetizable coating is applied to both sides of the platter

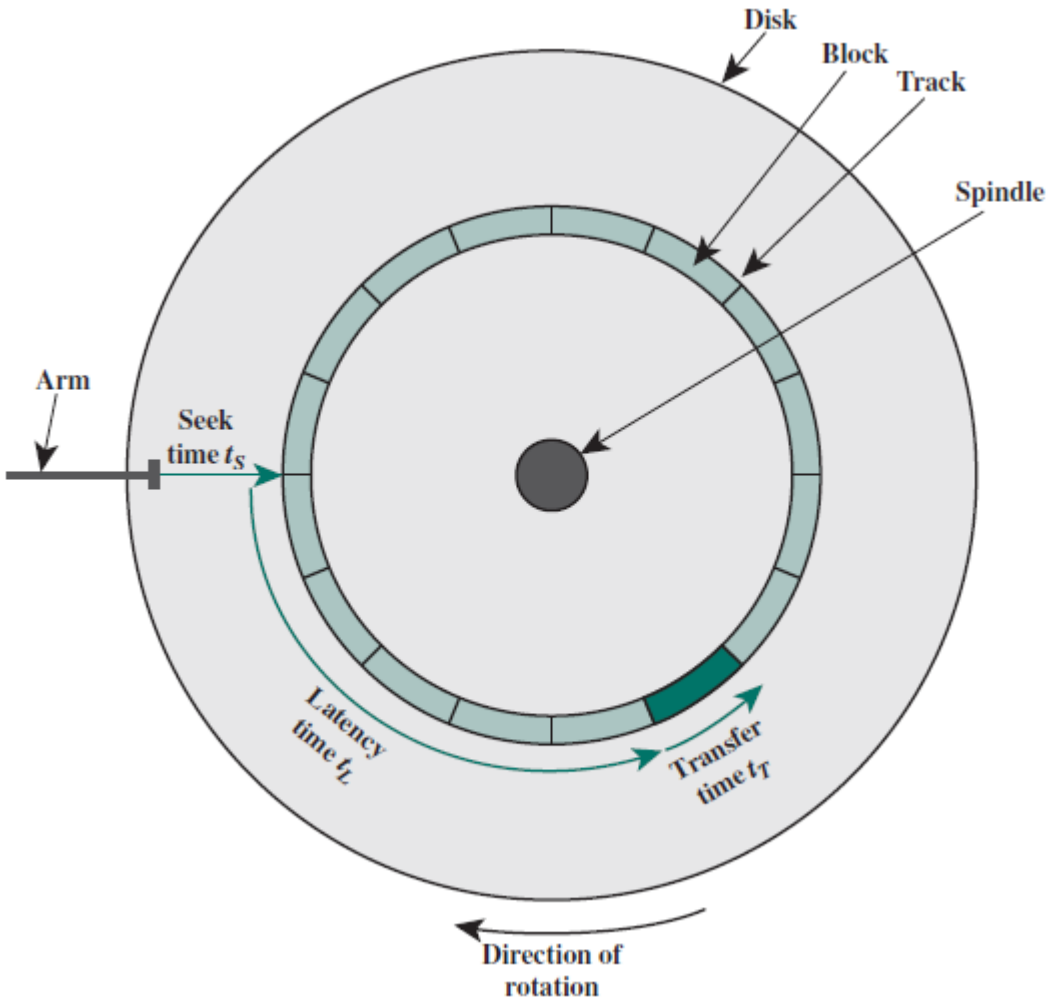
Typical Hard Disk Drive Parameters

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

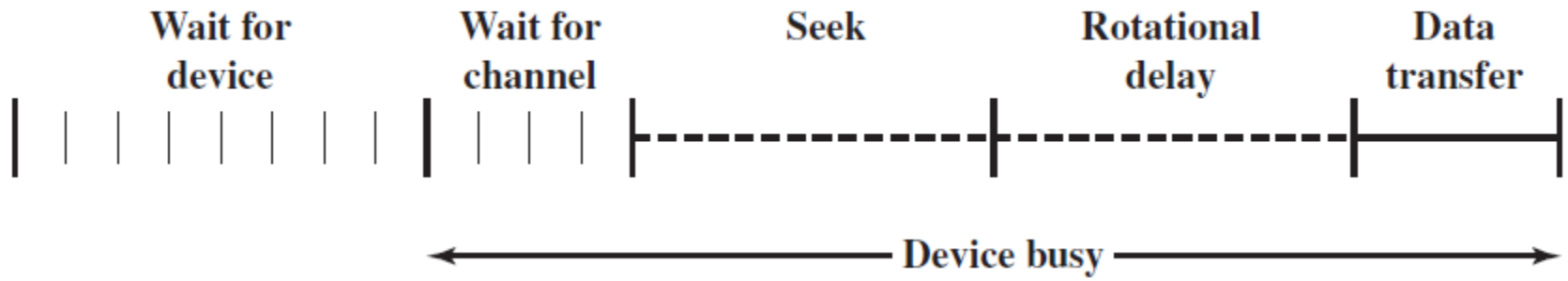
Disk Performance Parameters

- To read or write, the head must be positioned at the desired track and at the beginning of the desired sector on the track
- **Seek time**
  - Moving head to correct track
- **Rotational delay** (rotational latency)
  - Waiting for data to rotate under head
- **Access time** = Seek + Latency
  - Getting into position to read or write
- **Transfer time**
  - Data transfer portion of the operation
  - Read until end of sector is seen by the head
  - Depends on how fast the disk is spinning and how many sectors per track

Timing of a Disk I/O Transfer



Timing of a Disk I/O Transfer



- Access to disk are happening one at a time
- Can’t access another track until we are done with this one and then move to another track

Example

- 1000 cylinders, 10 sectors/track
- Head assembly at cylinder 0 initially
- Head moves at 10 μs/cylinder
- Disk rotates 100 times/second

- What is the average time to read a randomly chosen byte from this disk?

RAID

- Redundant Array of Independent Disks
- 7 levels
- Not a hierarchy
- Set of physical disks viewed as single logical drive by OS
- Data distributed across physical drives
- Can use redundant capacity to store parity information

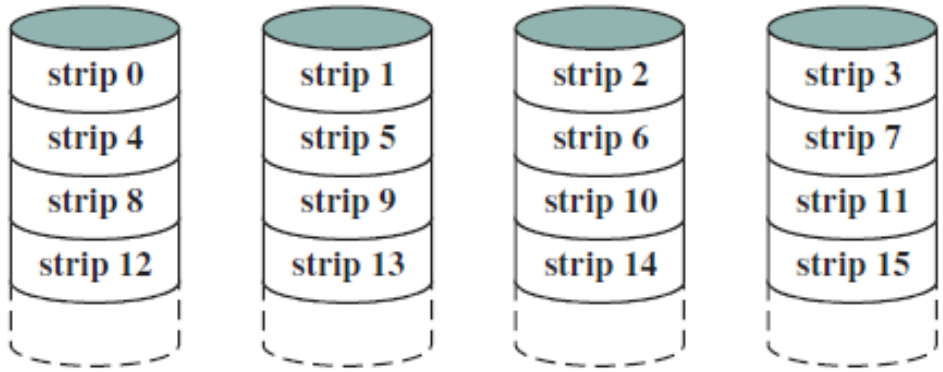
RAID Levels

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	$N + m$	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

Note: N = number of data disks; m proportional to log N

RAID 0

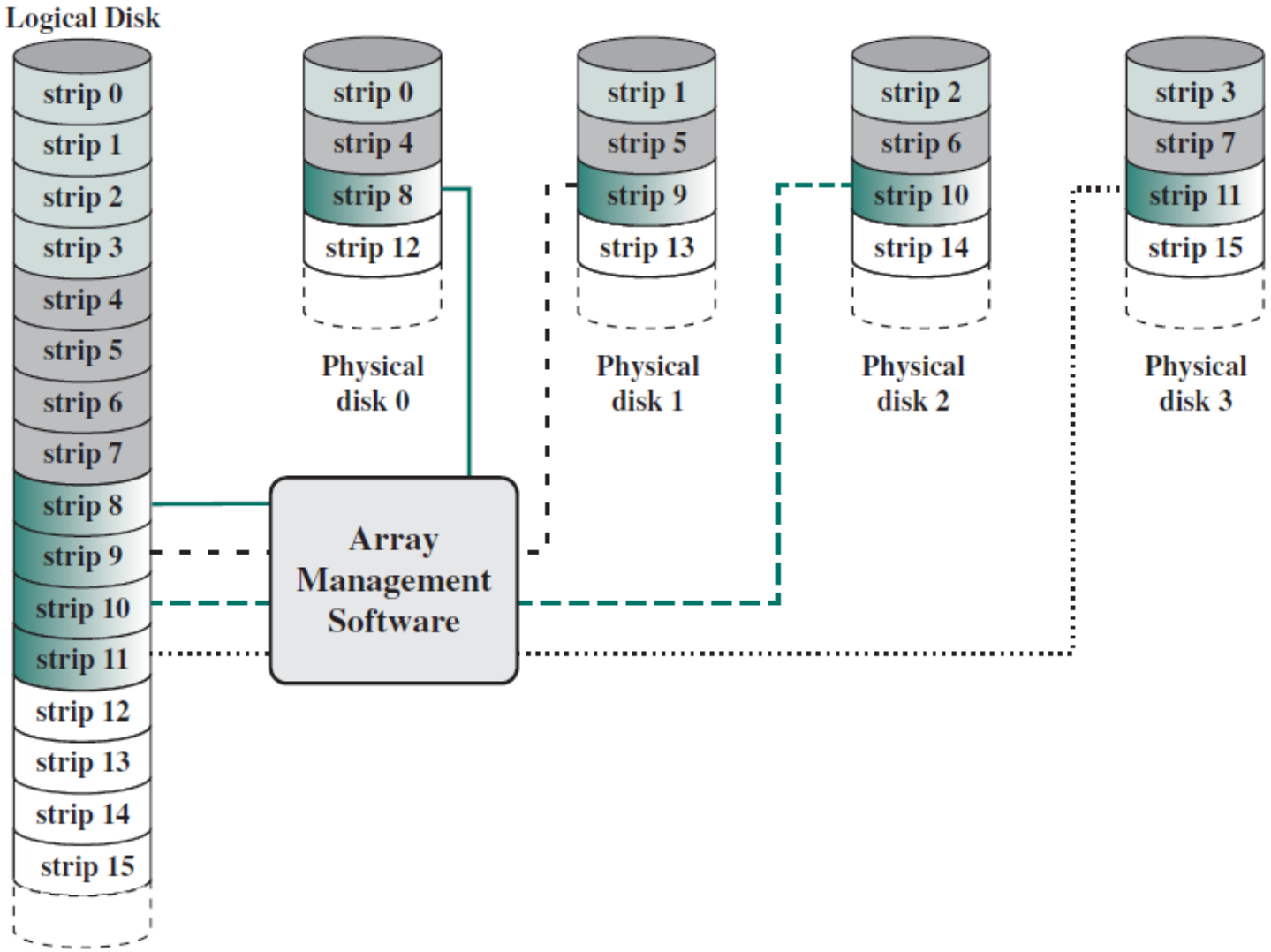
- No redundancy
- Data striped across all disks
- Round Robin striping



(a) RAID 0 (Nonredundant)

- 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

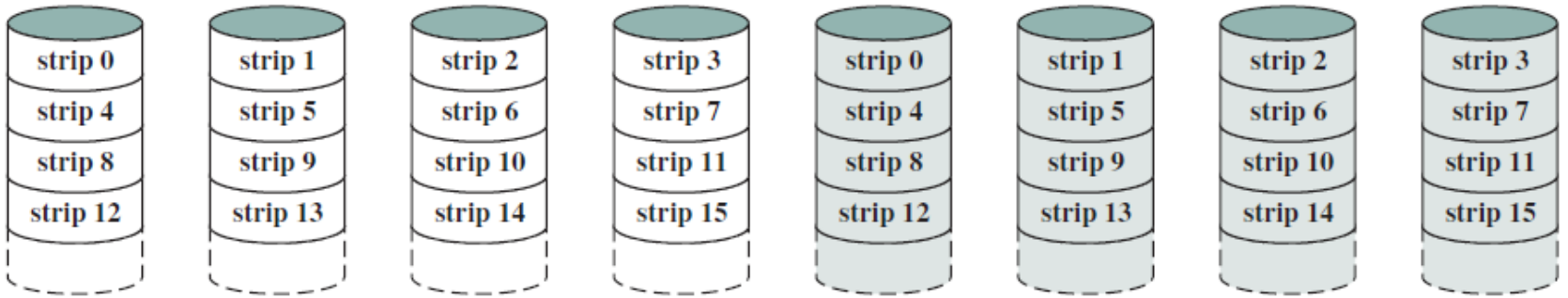
Data Mapping for RAID 0



RAID 1

- Mirrored Disks
- Data is striped across disks
- 2 copies of each stripe on separate disks
- Read from either
- Write to both

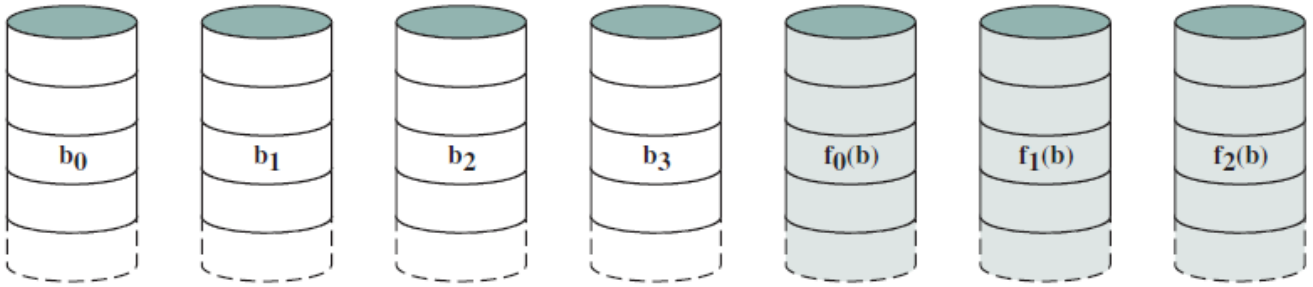
- Recovery is simple
  - Swap faulty disk & re-mirror
  - No down time
- Expensive



(b) RAID 1 (Mirrored)

RAID 2

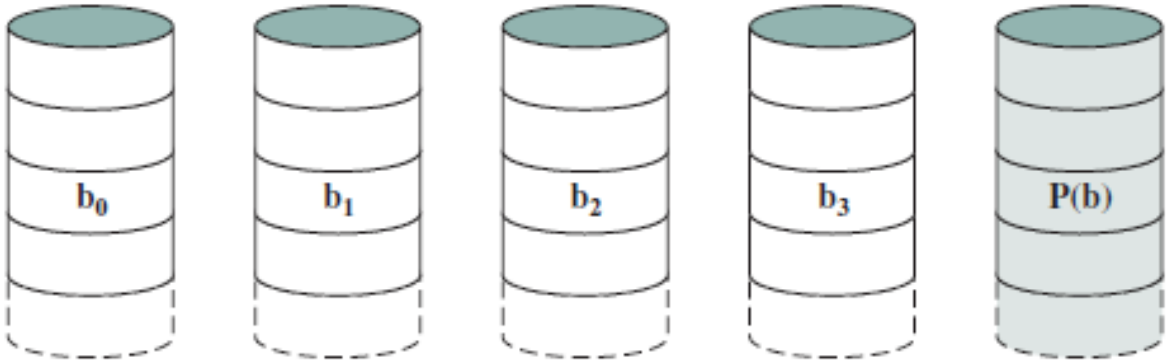
- Very small strips
  - Often single byte/word
- Error correction calculated across corresponding bits on disks
- Multiple parity disks store Hamming code error correction in corresponding positions
- Lots of redundancy
  - Expensive
  - Not used



(c) RAID 2 (Redundancy through Hamming code)

RAID 3

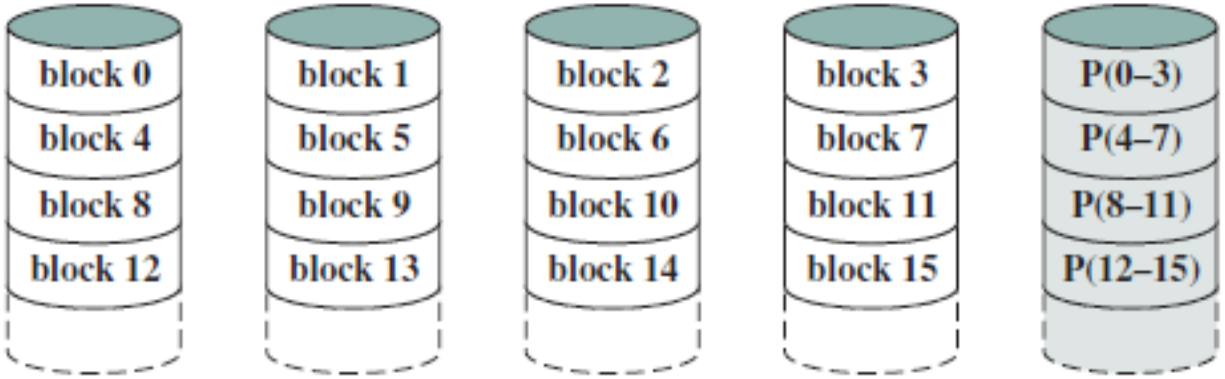
- Similar to RAID 2
- Only one redundant disk, no matter how large the array
- Simple parity bit for each set of corresponding bits
- Data on failed drive can be reconstructed from surviving data and parity info



(d) RAID 3 (Bit-interleaved parity)

RAID 4

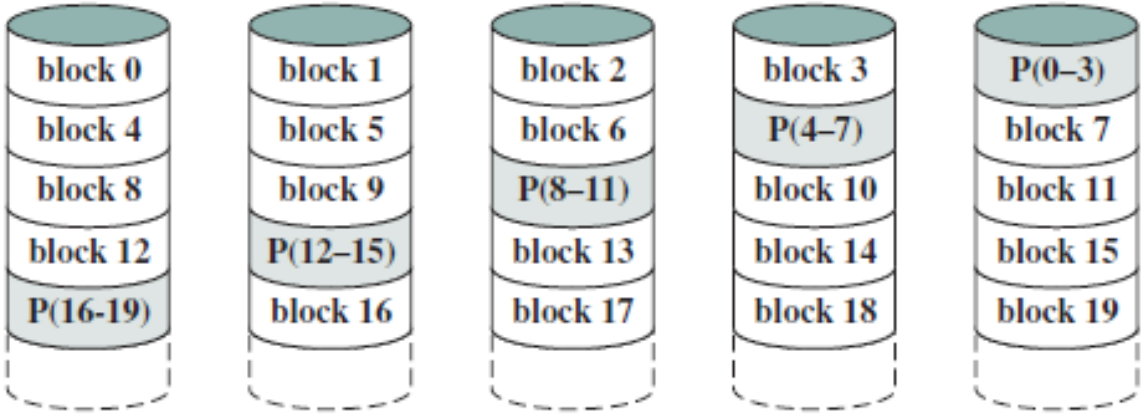
- Each disk operates independently
- Good for high I/O request rate; separate I/O requests can be satisfied in parallel
- Large stripes
- Bit by bit parity calculated across stripes on each disk
- Parity stored on parity disk



(e) RAID 4 (Block-level parity)

RAID 5

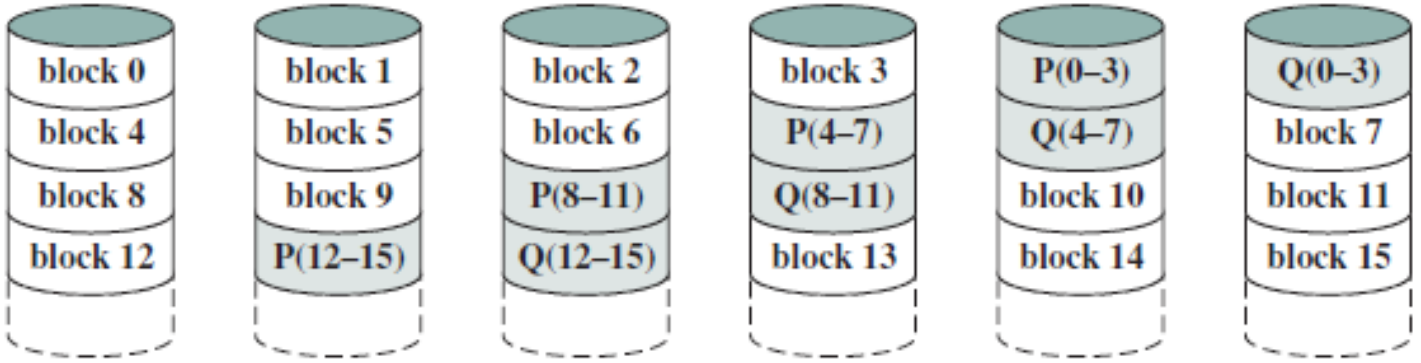
- Like RAID 4
- Parity striped across all disks
- Round robin allocation for parity stripe
- Avoids RAID 4 bottleneck at parity disk
- Commonly used in network servers



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

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



(g) RAID 6 (Dual redundancy)

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
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 (cont'd)

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

Solid State Drives

- Most significant development to complement or even replace hard disk drives (HDDs) both as internal and external memory
- Solid state: electronic circuitry built with semiconductors
- Use NAND flash memory
- As the cost of flash-based SSDs has dropped and the performance and bit density increased, SSDs have become increasingly competitive with HDDs.

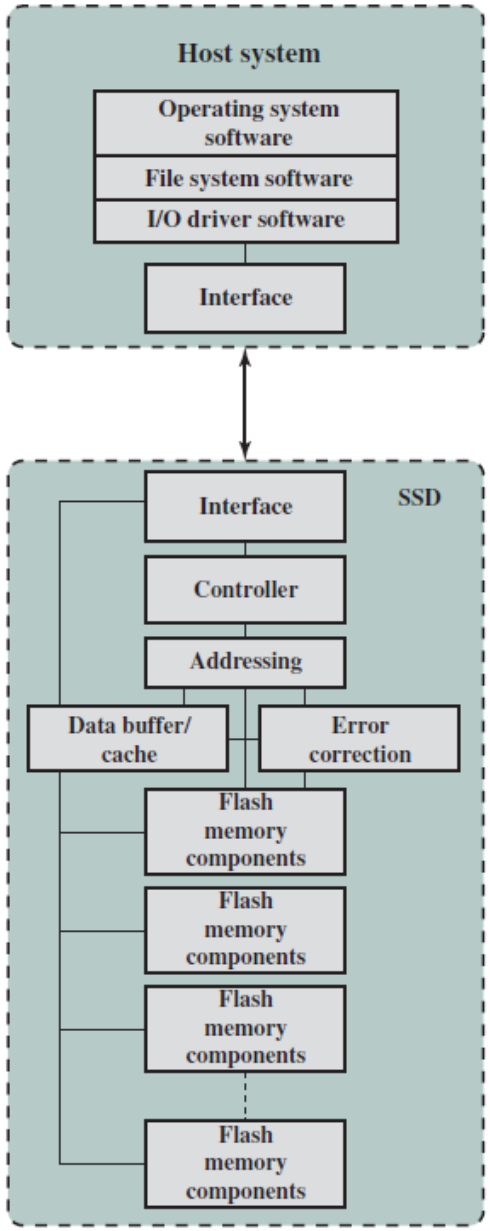
SSD Compared to HDD

- SSDs have the following advantages over HDDs:
  - High-performance input/output operations per second (IOPS)
  - Durability:
    - Less susceptible to physical shock and vibration
  - Longer lifespan:
    - SSDs are not susceptible to mechanical wear
  - Lower power consumption:
    - SSDs use considerably less power than comparable size HDDs
  - Quieter and cooler running capabilities:
    - Less space required, lower energy costs, and a greener enterprise
  - Lower access times and latency rates:
    - Over 10 times faster than the spinning disks in an HDD

Solid State Drive Architecture

SSD contains the following components:

- Interface to the host system
- Controller
- Addressing
- Data buffer/cache
- Error correction
- Flash memory components



Practical Issues

- SSD performance has a tendency to slow down as the device is used

- To write a page onto flash memory
  - 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

- Most flash devices estimate their own remaining lifetimes so systems can anticipate failure and take preemptive action

Hybrid Magnetic-Flash

- Magnetic disk
  - Low $/GB
  - Huge capacity
  - Power hungry
  - Slow (mechanical movement)
  - Sensitive to impacts while spinning (head likely to scratch surface of disk)

- Flash
  - Fast
  - Power efficient
  - No moving parts

- Have both
  - Use flash as cache for disk
  - Most of data is on disk
  - Data we frequently access is on the flash

Flash vs Disk vs both - Example

- Play game for 2 hours (reads to 2 GB, writes another 10MB)
- Watch movie for 2 hours (read 1 GB sequentially)
- Repeat 4 times!
- Disk: read 100MB/s sequential, read/write 1 MB/s random
- Flash: 1GB/s

1) total access time with disk

2) total access time with flash

3) total access time with disk and 4GB flash

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.

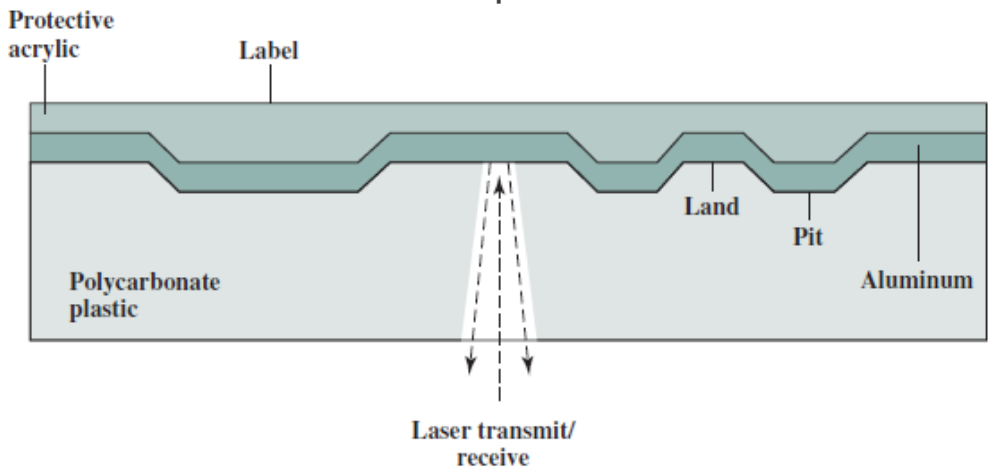
DVD-RW
DVD Rewritable. Similar to a DVD-ROM. The user can erase and rewrite to the disk multiple times. Only one-sided disks can be used.

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.

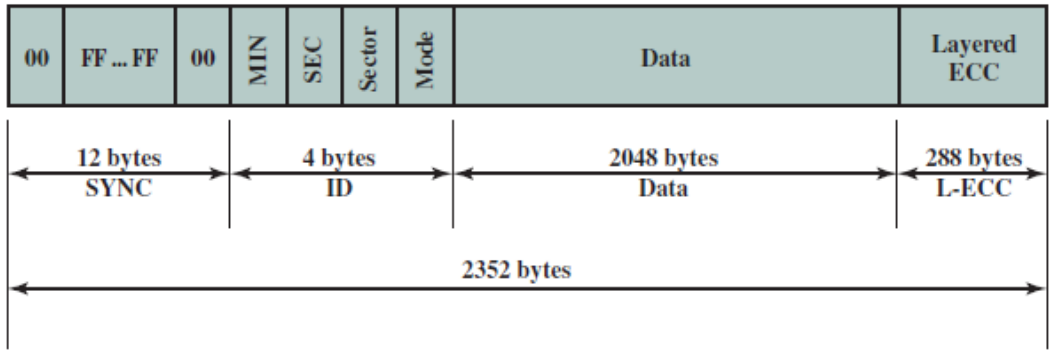
Compact Disk Read Only Memory (CD-ROM)

- Originally for audio
- Polycarbonate coated with highly reflective coat, usually aluminium
- Data read by reflecting laser
- To achieve greater capacity, the disk contains a single spiral track, beginning near the center and spiraling out to the outer edge of the disk
- Data capacity about 680 MB

CD operation



CD-ROM block format



Magnetic Tape

- Backup and archive (secondary storage)
- Large capacity, replaceable
- Slow
- Sequential access
- Dying out
  - Low production volume
    - Cost not dropping as rapidly as disks
  - Cheaper to use disks
    - USB drives

Image from IEEE Spectrum

Disk, Tape, Both, Neither

- Read 1 GB file from start to end

  ☐ Disk ☐ Tape

- Read just first and last byte of 1 GB file

  ☐ Disk ☐ Tape

- Make a cat happy

  ☐ Disk ☐ Tape