William Stallings
Computer Organization and Architecture
9th Edition

+ Chapter 6

External Memory

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

Magnetic Read and Write Mechanisms

Write head is made of magnetizable material and 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 magnetizes a small area of the recording medium

Reversing the direction of the current reverses the direction of the magnetization

Inductive Write/Magnetoresistive Read Head

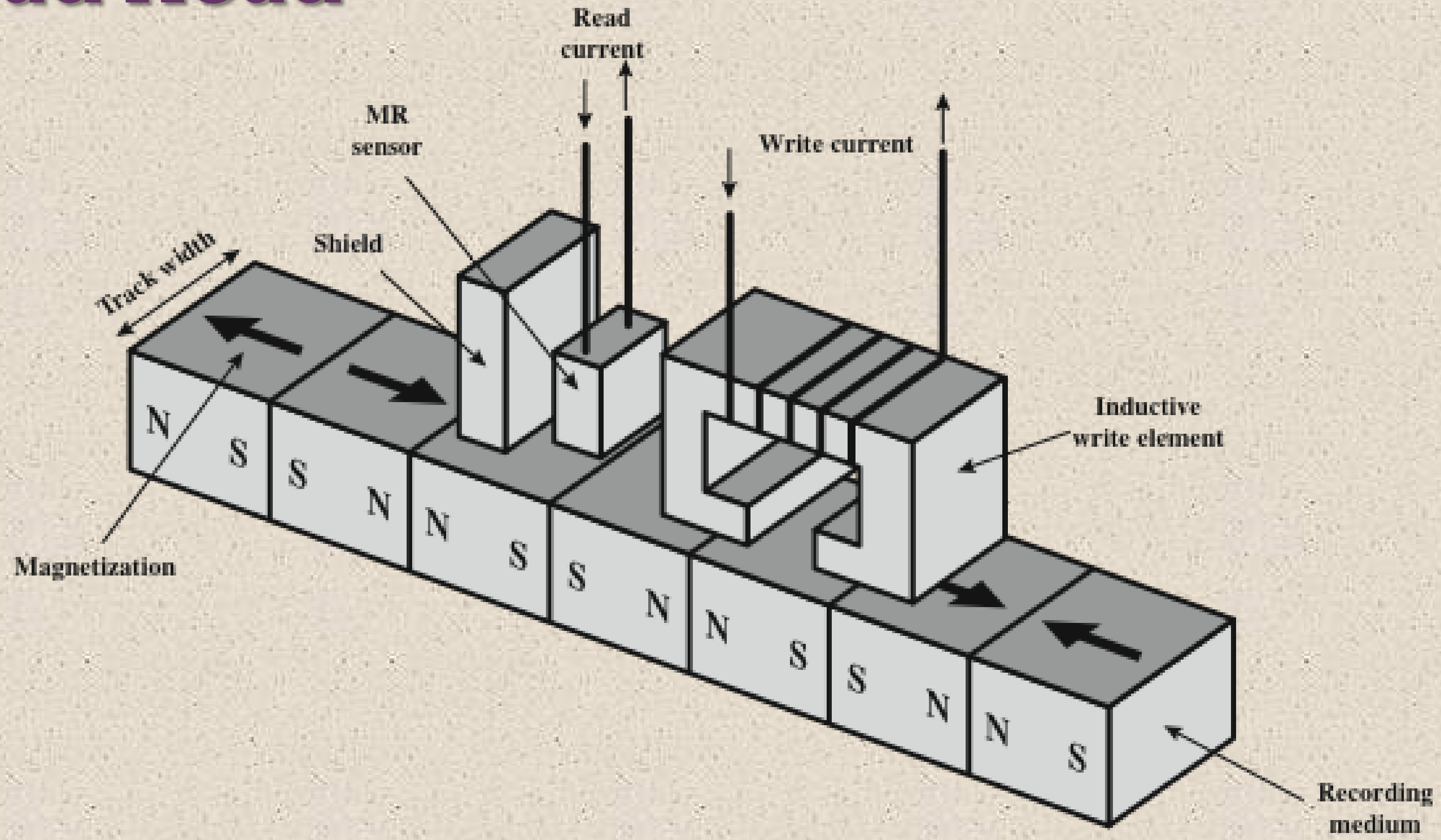


Figure 6.1 Inductive Write/Magnetoresistive Read Head

Disk Data Layout

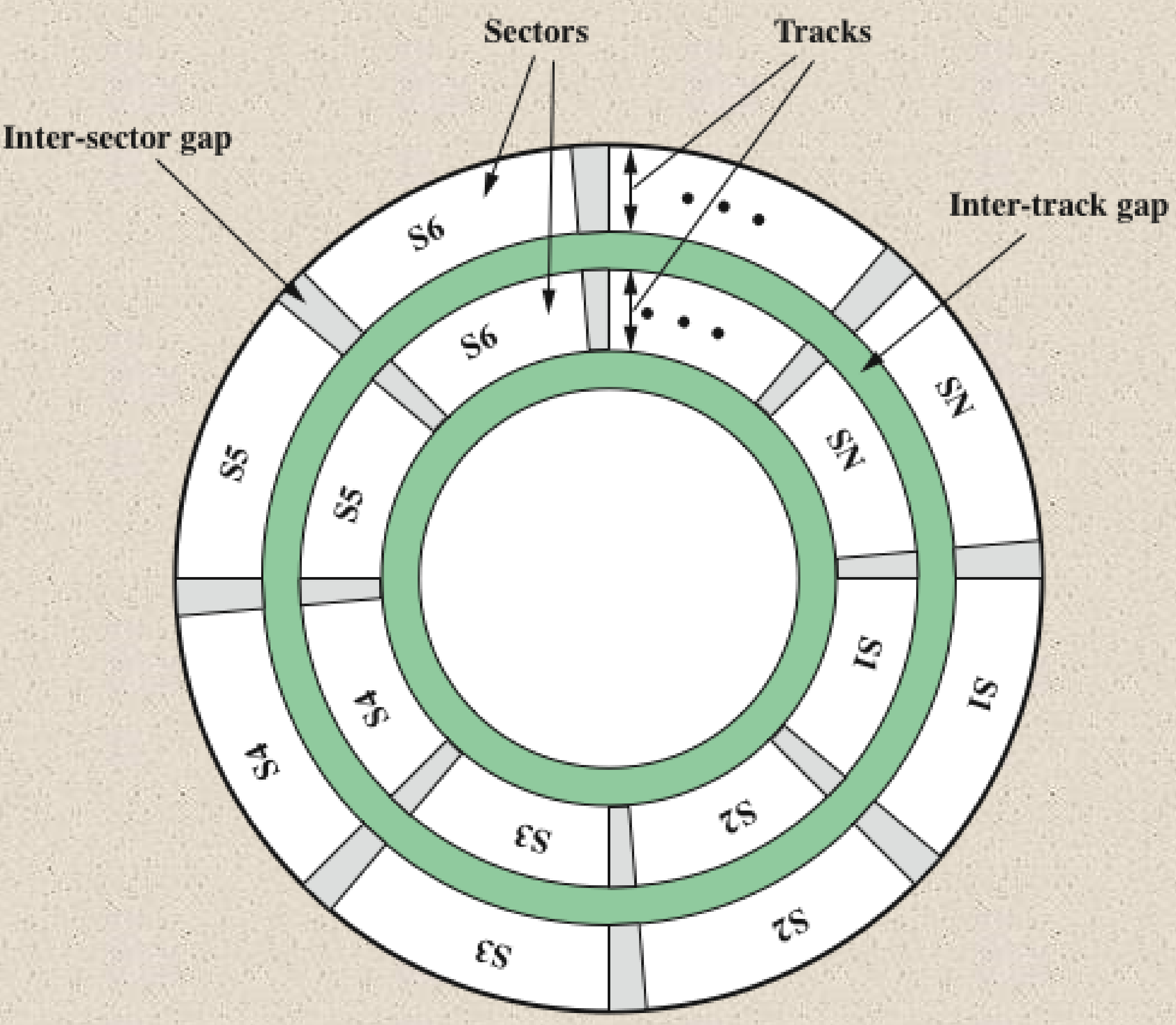


Figure 6.2 Disk Data Layout

Disk Layout Methods Diagram

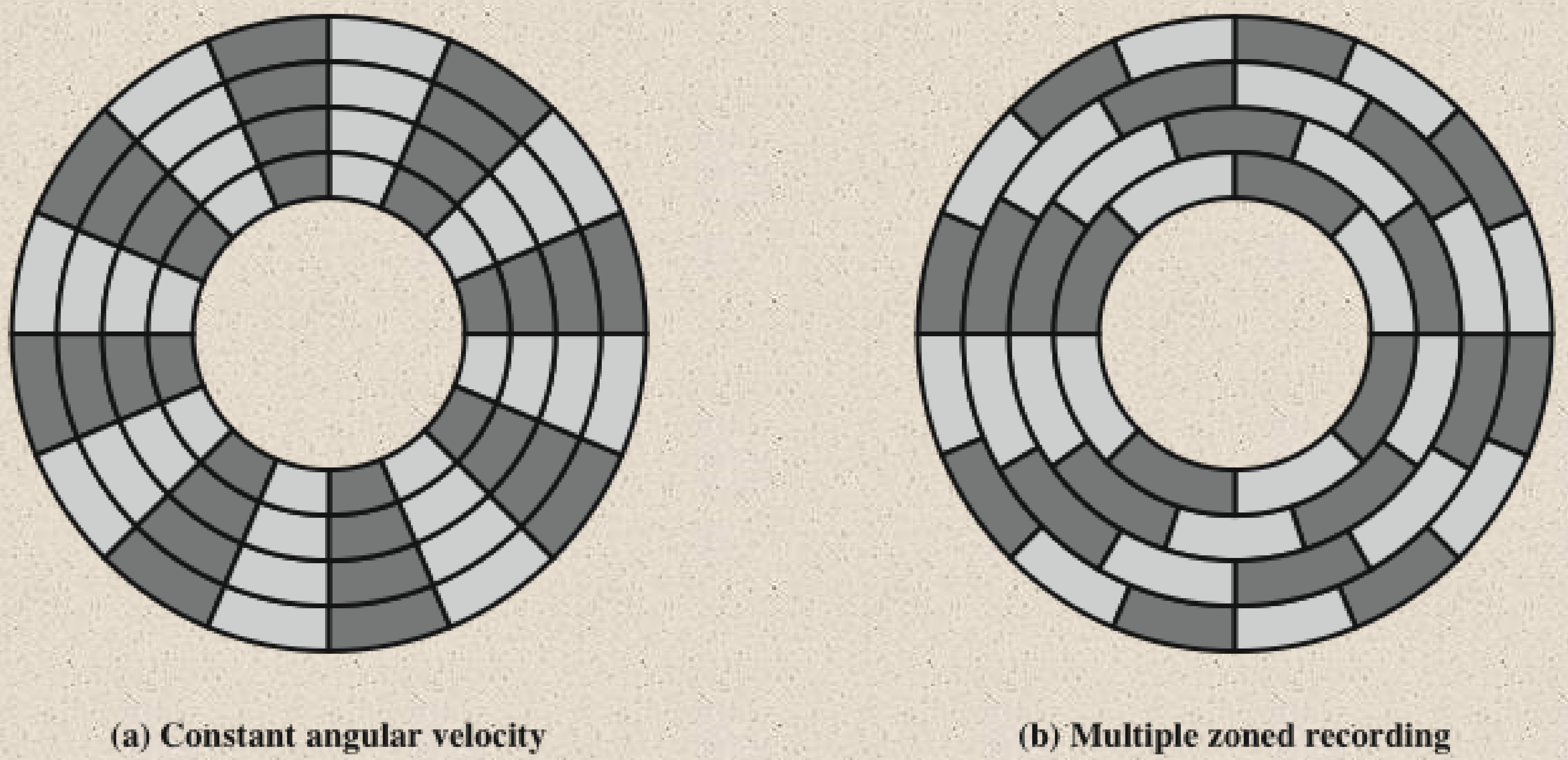


Figure 6.3 Comparison of Disk Layout Methods

Winchester Disk Format
Seagate ST506

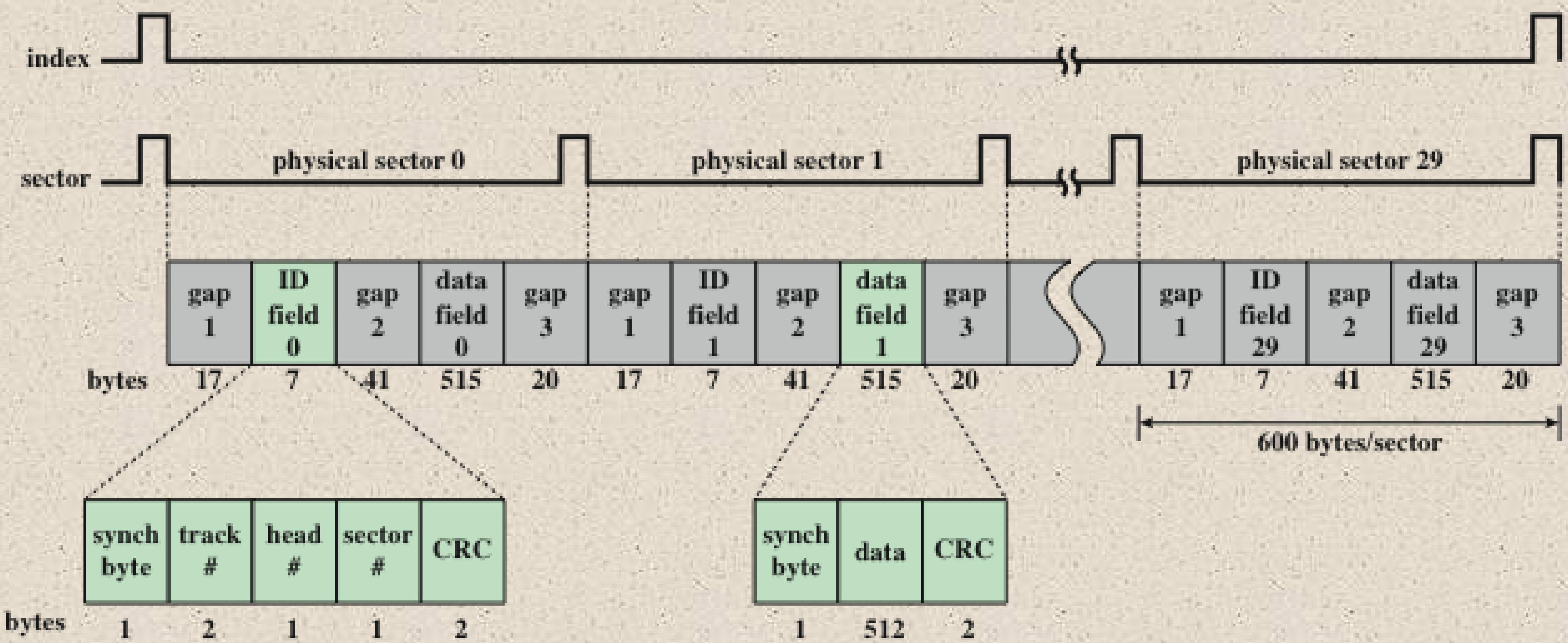


Figure 6.4 Winchester Disk Format (Seagate ST506)

+

Table 6.1 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
	Aerodynamic gap (Winchester)
Sides	
Single sided	
Double sided	

Table 6.1 Physical Characteristics of Disk Systems

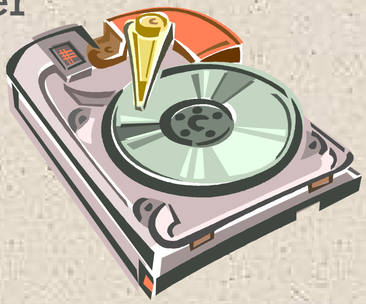
+

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

Characteristics

- 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



+

Multiple Platters

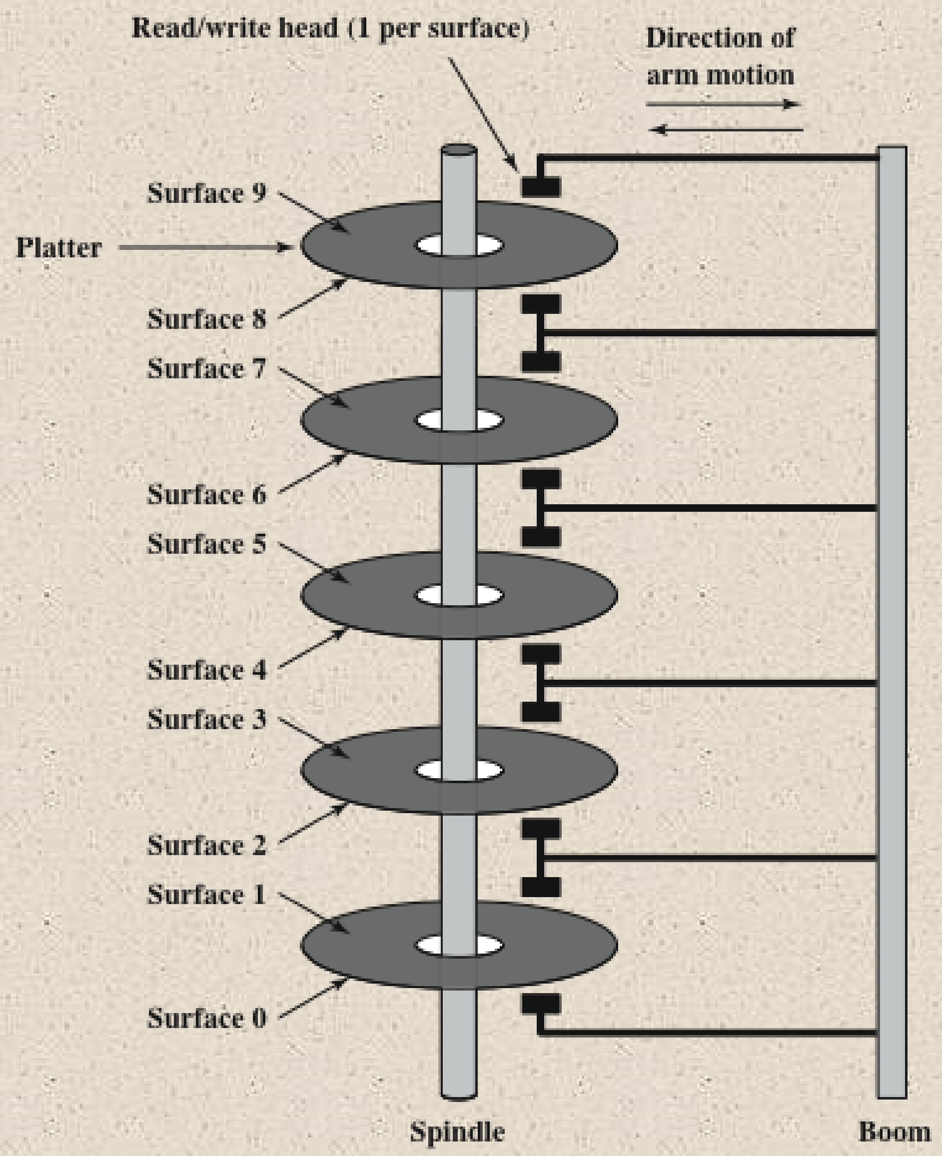


Figure 6.5 Components of a Disk Drive

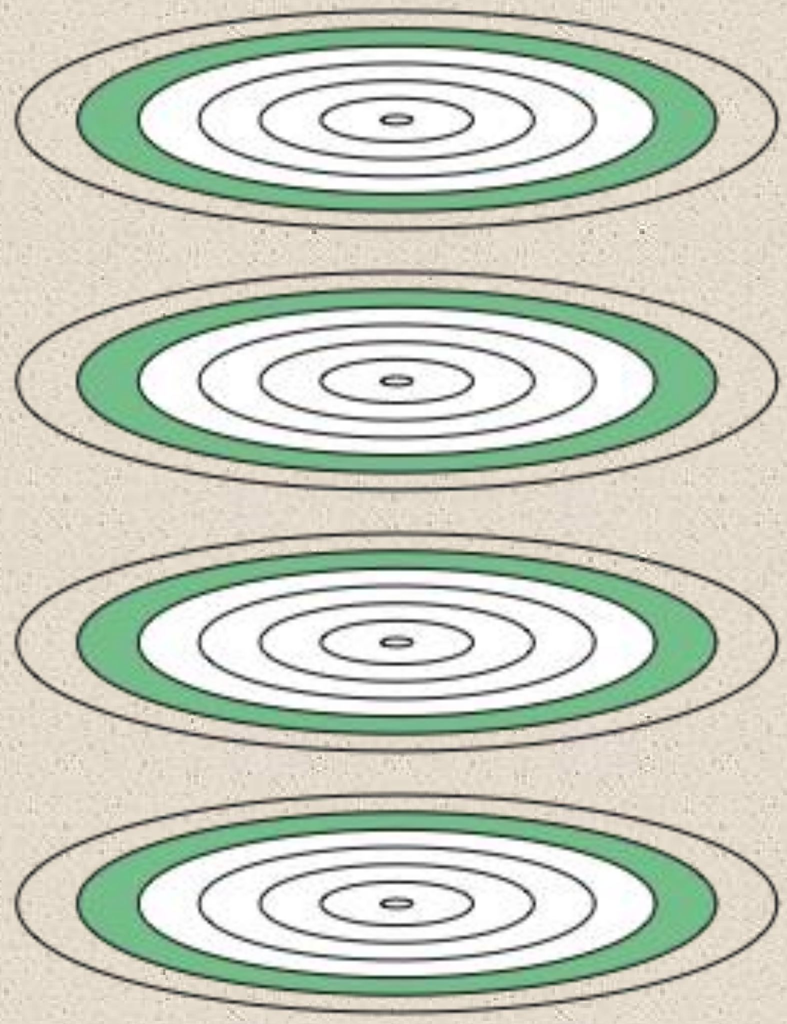


Figure 6.6 Tracks and Cylinders

Tracks

Cylinders

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

Typical Hard Disk Parameters

Characteristics	Constellation ES.2	Seagate Barracuda XT	Cheetah NS	Momentus
Application	Enterprise	Desktop	Network attached storage, application servers	Laptop
Capacity	3 TB	3 TB	400 GB	640 GB
Average seek time	8.5 ms read 9.5 ms write	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	155 MB/s	149 MB/s	97 MB/s	300 MB/s
Bytes per sector	512	512	512	4096
Tracks per cylinder (number of platter surfaces)	8	10	8	4
Cache	64 MB	64 MB	16 MB	8 MB

Table 6.2 Typical Hard Disk Drive Parameters

+

Timing of Disk I/O Transfer

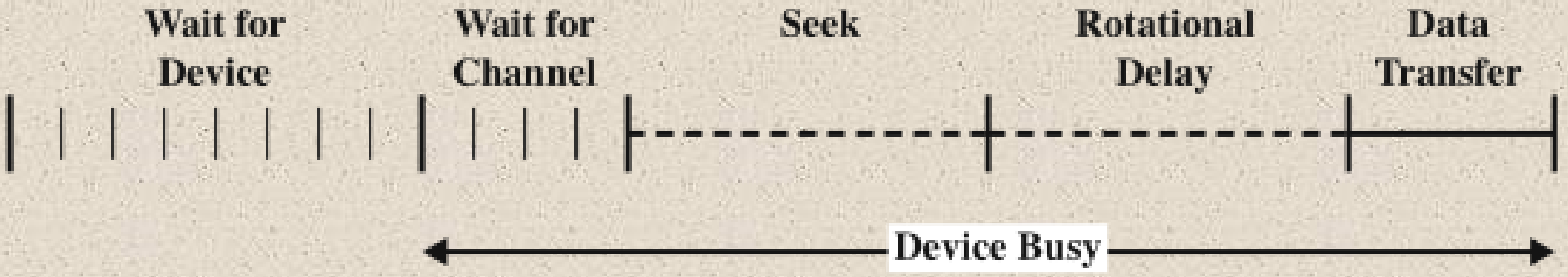


Figure 6.7 Timing of a Disk I/O Transfer

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

Disk Performance Parameters

- 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

Disk Performance Parameters

- 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

RAID

Redundant Array of Independent Disks

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

RAID

Redundant Array of Independent Disks

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

Highlights

- Increasing performance gap between CPU and IO
- Data availability a priority
- RAID meets the IO challenge:
  - Performance via parallelism
  - Data Availability via redundancy
  - Flexibility via multiple RAID levels, each offer unique performance/availability/cost tradeoffs

Table 6.3 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

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

Data Mapping for a RAID Level 0 Array

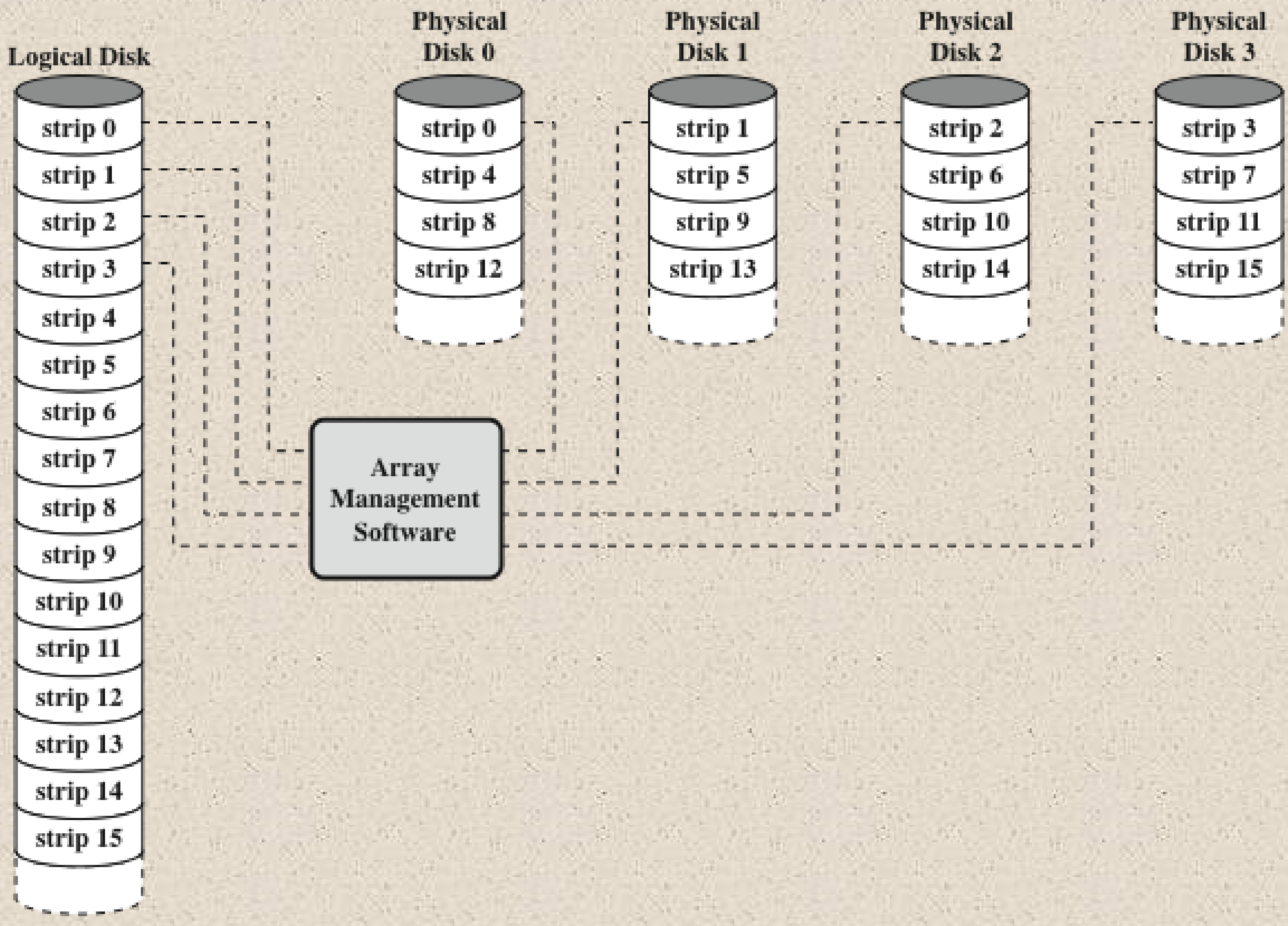


Figure 6.9 Data Mapping for a RAID Level 0 Array

+

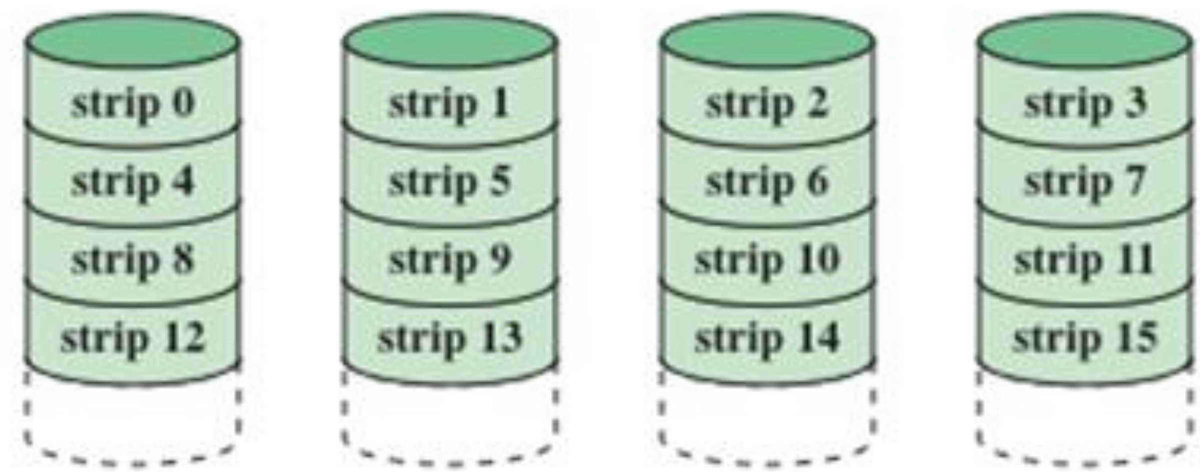
RAID Level 0

- Addresses the issues of request patterns of the host system and layout of the data
- Impact of redundancy does not interfere with analysis

RAID 0 for High Data Transfer Capacity

- For applications to experience a high transfer rate two requirements must be met:

1. A high transfer capacity must exist along the entire path between host memory and the individual disk drives
2. The application must make I/O requests that drive the disk array efficiently



(a) RAID 0 (non-redundant)

+

RAID Level 0

RAID 0 for High I/O Request Rate

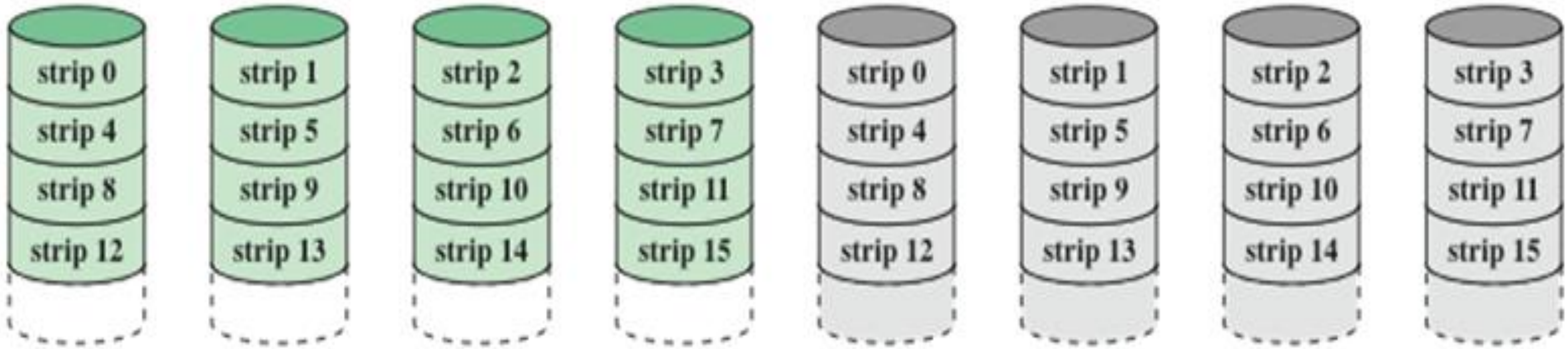
- For an individual I/O request for a small amount of data, the I/O time is dominated by the seek time and rotational latency
- A disk array can provide high I/O execution rates by balancing the I/O load across multiple disks
- If the strip size is relatively large, multiple I/O requests can be handled in parallel, reducing the queuing time.

+

RAID Level 1

Characteristics

- Differs from RAID levels 2 through 6 in the way in which redundancy is achieved
- It is achieved by duplicating the data.
- Data striping is used but each logical strip is mapped to two separate physical disks so that every disk in the array has a mirror disk that contains the same data
- It can also be implemented without data striping, although this is less common



(b) RAID 1 (mirrored)

+

RAID Level 1

Positive Aspects

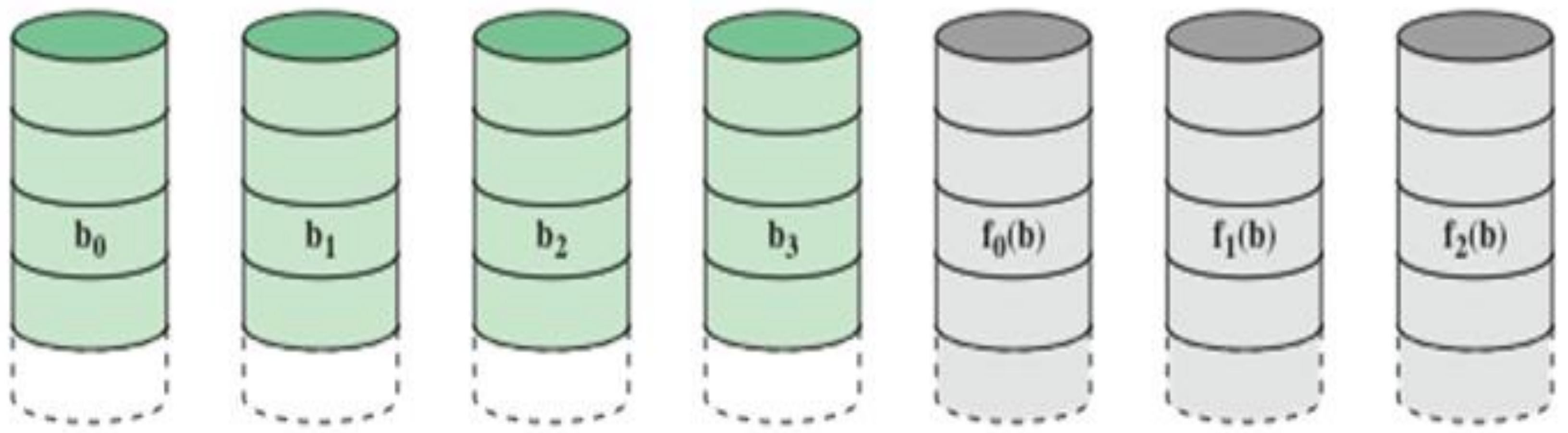
- A read request can be serviced by either of the two disks that contains the requested data
- There is no "write penalty"
- Recovery from a failure is simple, when a drive fails the data can be accessed from the second drive
- Provides real-time copy of all data
- Can achieve high I/O request rates if the bulk of the requests are reads
- Principal disadvantage is the cost

+

RAID Level 2

Characteristics

- Makes use of a parallel access technique
- In a parallel access array all member disks participate in the execution of every I/O request
- Spindles of the individual drives are synchronized so that each disk head is in the same position on each disk at any given time
- Data striping is used
  - Strips are very small, often as small as a single byte or word



(c) RAID 2 (redundancy through Hamming code)

+

RAID Level 2

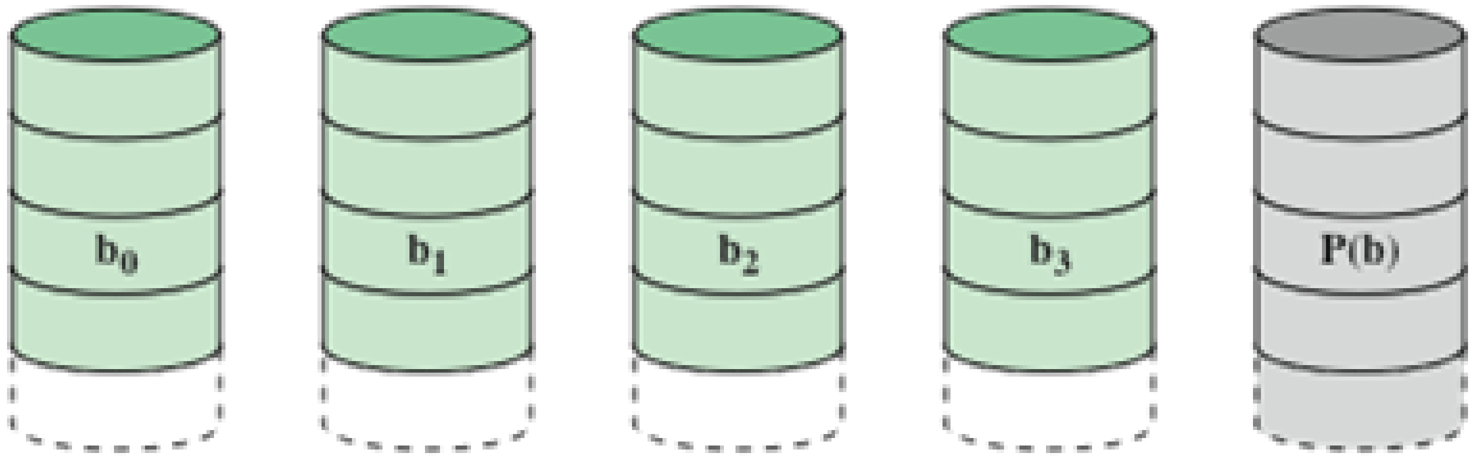
Performance

- An error-correcting code is calculated across corresponding bits on each data disk, and the bits of the code are stored in the corresponding bit positions on multiple parity disks
- Typically a Hamming code is used, which is able to correct single-bit errors and detect double-bit errors
- The number of redundant disks is proportional to the log of the number of data disks
- Would only be an effective choice in an environment in which many disk errors occur

RAID Level 3

Redundancy

- Requires only a single redundant disk, no matter how large the disk array
- Employs parallel access, with data distributed in small strips
- Instead of an error correcting code, a simple parity bit is computed for the set of individual bits in the same position on all of the data disks
- Can achieve very high data transfer rates



(d) RAID 3 (bit-interleaved parity)

RAID Level 3

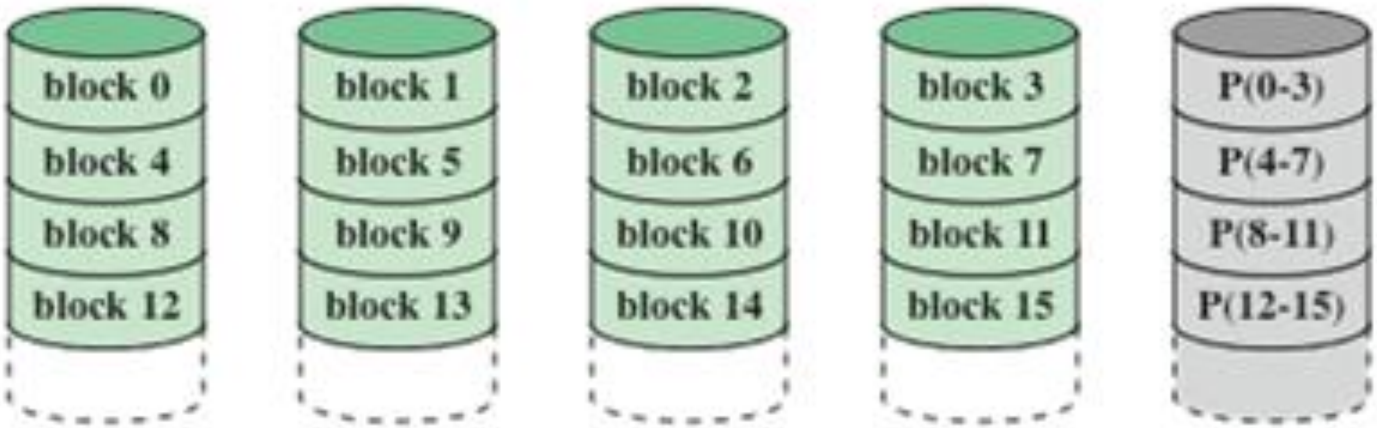
Performance

- In the event of a drive failure, the parity drive is accessed and data is reconstructed from the remaining devices
- Once the failed drive is replaced, the missing data can be restored on the new drive and operation resumed
- In the event of a disk failure, all of the data are still available in what is referred to as *reduced mode*
- Return to full operation requires that the failed disk be replaced and the entire contents of the failed disk be regenerated on the new disk
- In a transaction-oriented environment performance suffers

RAID Level 4

Characteristics

- Makes use of an independent access technique
  - In an independent access array, each member disk operates independently so that separate I/O requests can be satisfied in parallel
- Data striping is used
  - Strips are relatively large
- To calculate the new parity the array management software must read the old user strip and the old parity strip



(e) RAID 4 (block-level parity)

RAID Level 4

Performance

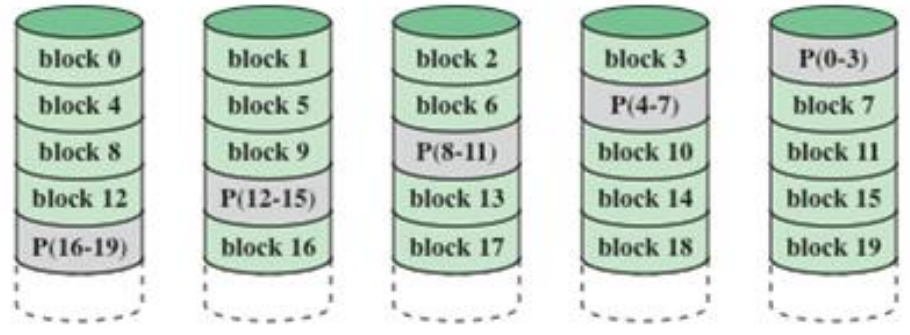
- Involves a write penalty when an I/O write request of small size is performed
- Each time a write occurs the array management software must update the user data and the corresponding parity bits
- Thus each strip write involves two reads and two writes

+

RAID Level 5

Characteristics

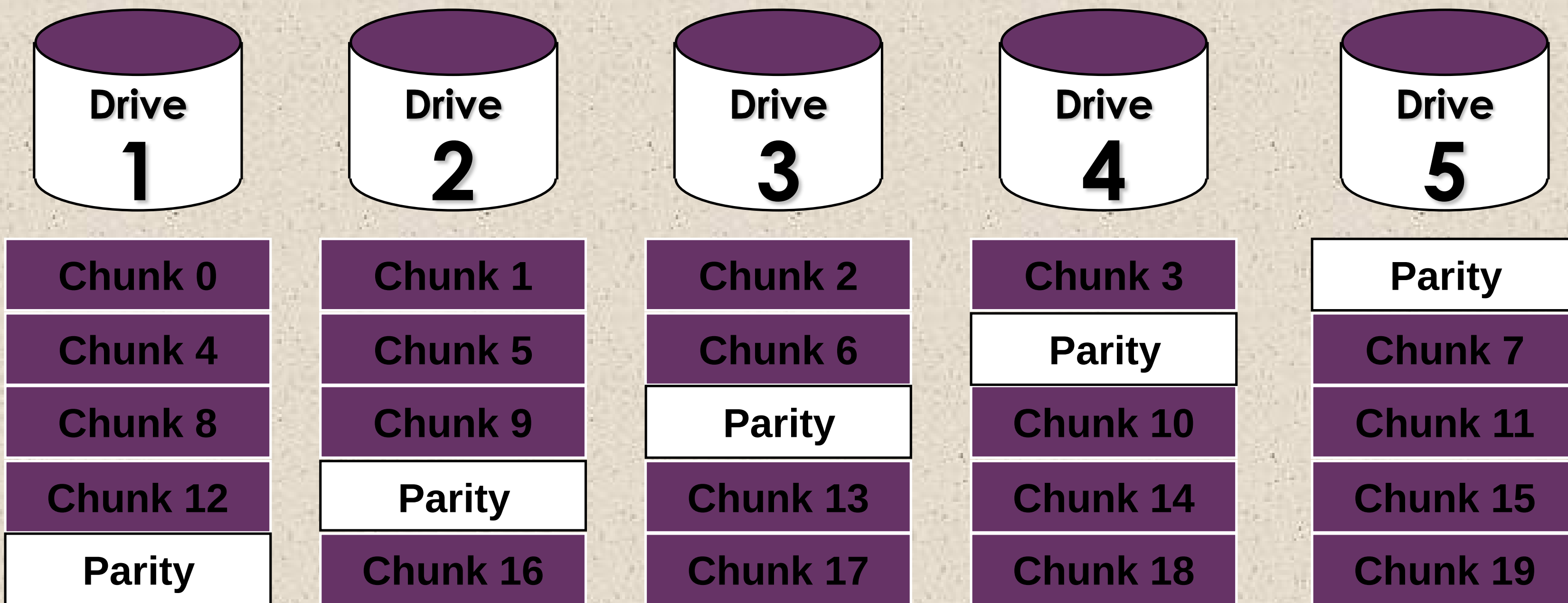
- Organized in a similar fashion to RAID 4
- Difference is distribution of the parity strips across all disks. A typical allocation is a round-robin scheme
- The distribution of parity strips across all drives avoids the potential I/O bottleneck found in RAID 4



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

RAID 5 Striping

- *Chunk size is tuned such that typical IO aligns on single disk.*
- *Parity rotates amongst disks to avoid write bottleneck*



RAID 5 - Write Operation

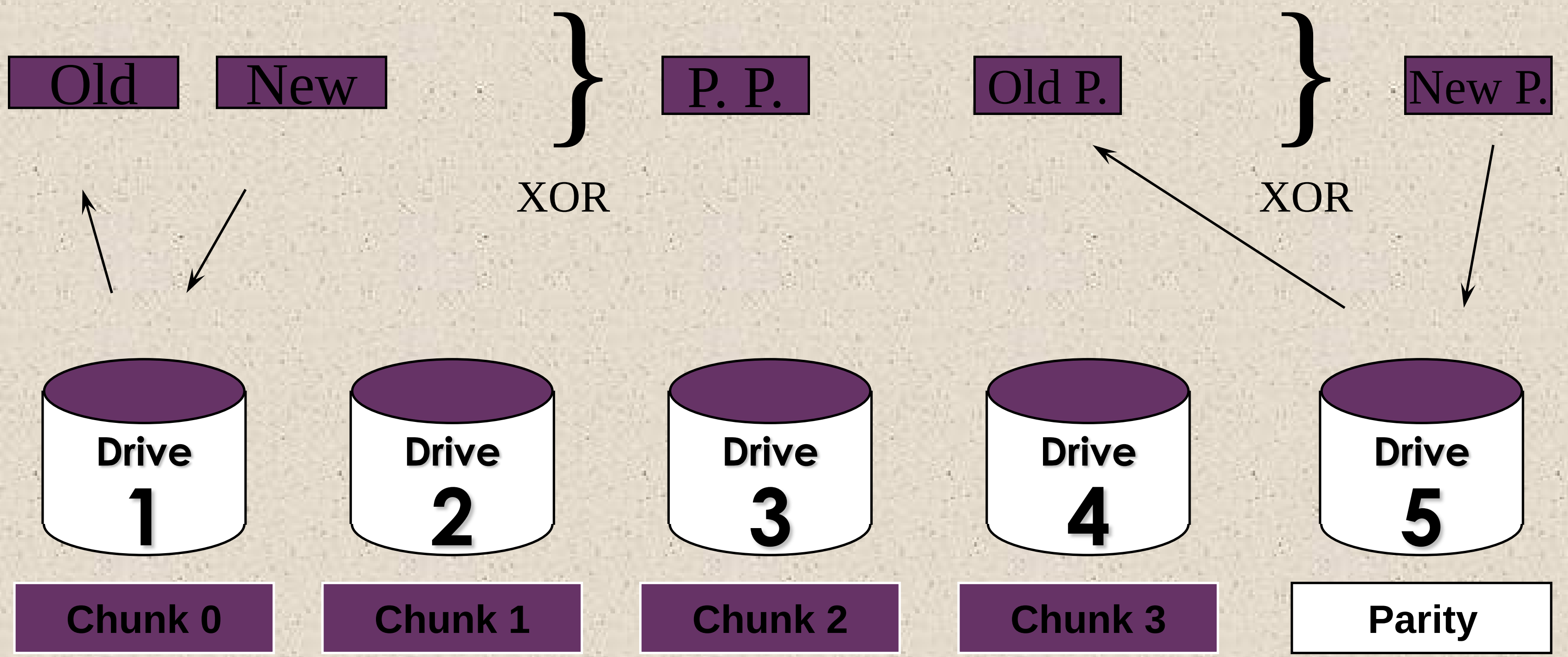
1. Read old data.

2. Write new data

3. XOR old and new data to create “Partial Product”.

4. Read old parity data.

5. XOR old parity with partial product, writing out result as new parity.



RAID Level Review

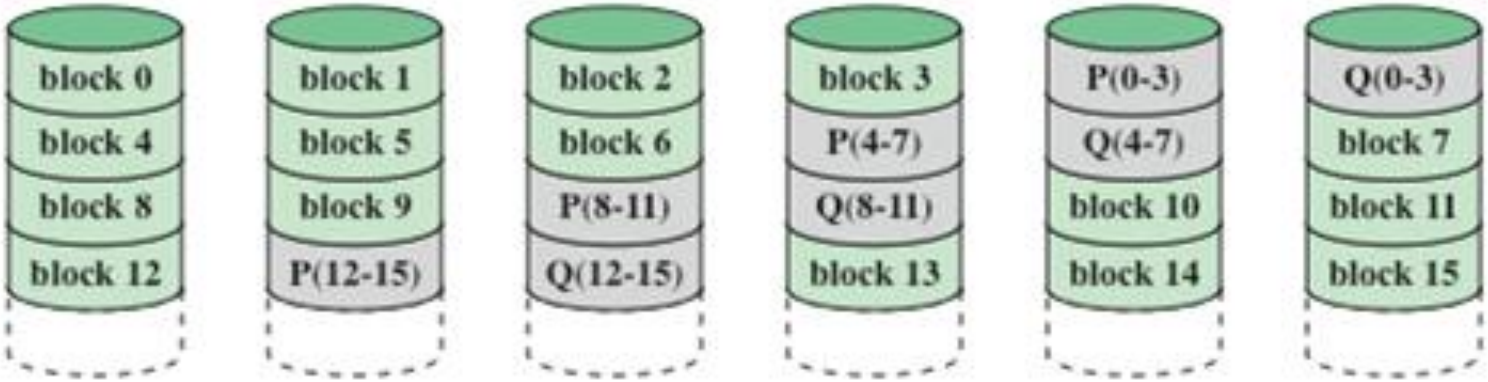
- RAID 0 - Data striping, Non-redundant.
  - *High Performance, Low Availability*
- RAID 1 - Mirroring
  - *Moderate Performance, Expensive High Availability*
- RAID 1/0 - Striping and Mirroring
  - *High Performance, Expensively High Availability*
- RAID 3 - Striping, single parity disk.
  - *High **Bandwidth** Performance, Cheap Availability*
- RAID 5 - Striping, rotating parity disk.
  - *High **Thruput** Performance, Cheap Availability*

+

RAID Level 6

Characteristics

- Two different parity calculations are carried out and stored in separate blocks on different disks
- Advantage is that it provides extremely high data availability
- Three disks would have to fail within the mean time to repair (MTTR) interval to cause data to be lost
- Incurs a substantial write penalty because each write affects two parity blocks



(g) RAID 6 (dual redundancy)

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

Table 6.4

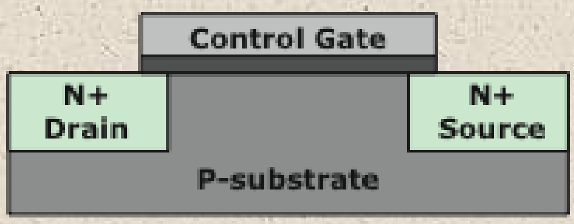
RAID Comparison (page 1 of 2)

Level	Advantages	Disadvantages	Applications
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

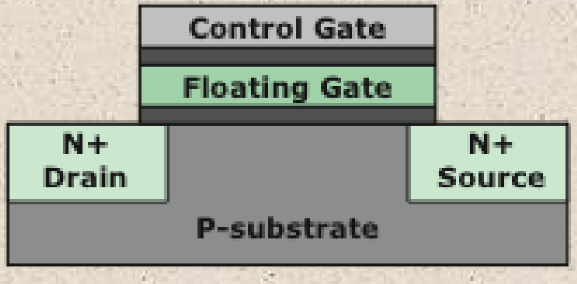
Table 6.4

RAID Comparison (page 2 of 2)

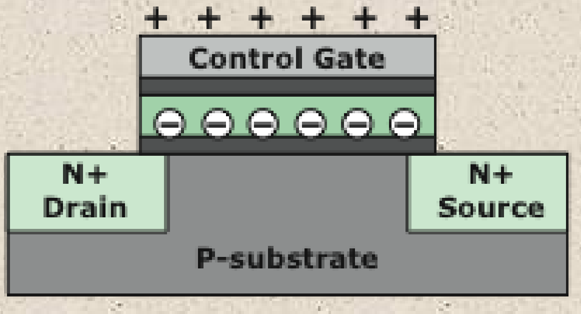
Flash Memory



(a) Transistor structure



(b) Flash memory cell in one state



(c) Flash memory cell in zero state

+

Figure 6.10
Flash Memory Operation

Solid State Drive (SSD)

A memory device made with solid state components that can be used as a replacement to a hard disk drive (HDD)

The term *solid state* refers to electronic circuitry built with semiconductors

Flash memory

A type of semiconductor memory used in many consumer electronic products including smart phones, GPS devices, MP3 players, digital cameras, and USB devices

Cost and performance has evolved to the point where it is feasible to use to replace HDDs

Two distinctive types of flash memory:

NOR

- The basic unit of access is a bit
- Provides high-speed random access
- Used to store cell phone operating system code and on Windows computers for the BIOS program that runs at start-up

NAND

- The basic unit is 16 or 32 bits
- Reads and writes in small blocks
- Used in USB flash drives, memory cards, and in SSDs
- Does not provide a random-access external address bus so the data must be read on a block-wise basis

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

Table 6.5

Comparisons

	NAND Flash Drives	Disk Drives
I/O per second (sustained)	Read: 45,000 Write: 15,000	300
Throughput (MB/s)	Read: 200+ Write: 100+	up to 80
Random access time (ms)	0.1	4–10
Storage capacity	up to 256 GB	up to 4 TB

SSD Organization

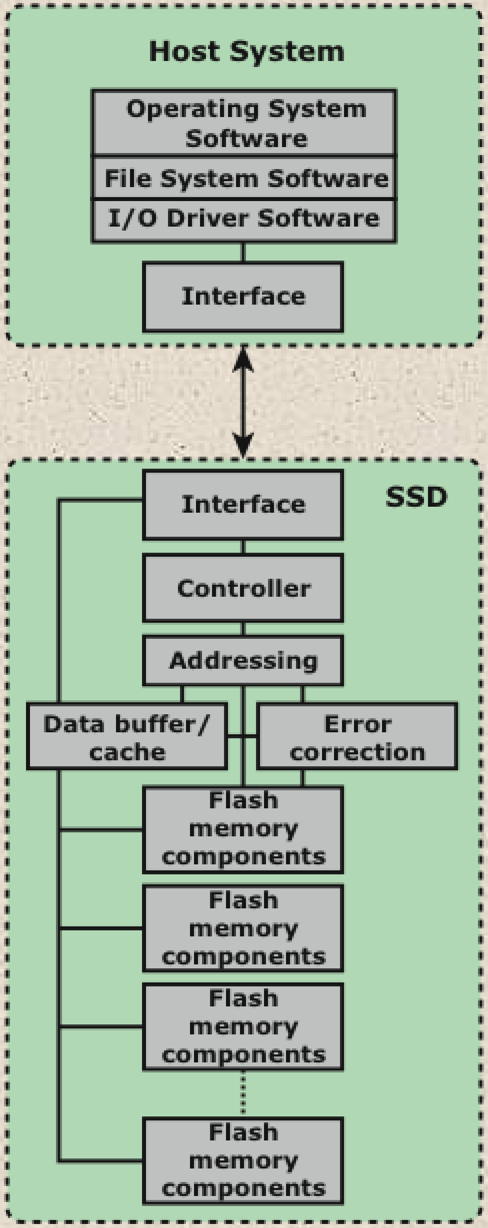


Figure 6.11 Solid State Drive Architecture

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.

Compact Disk Read-Only Memory (CD-ROM)

- Audio CD and the CD-ROM share a similar technology
  - The main difference is that CD-ROM players are more rugged and have error correction devices to ensure that data are properly transferred

- Production:
  - The disk is formed from a resin such as polycarbonate
  - Digitally recorded information is imprinted as a series of microscopic pits on the surface of the polycarbonate
    - This is done with a finely focused, high intensity laser to create a master disk
  - The master is used, in turn, to make a die to stamp out copies onto polycarbonate
  - The pitted surface is then coated with a highly reflective surface, usually aluminum or gold
  - This shiny surface is protected against dust and scratches by a top coat of clear acrylic
  - Finally a label can be silkscreened onto the acrylic

CD Operation

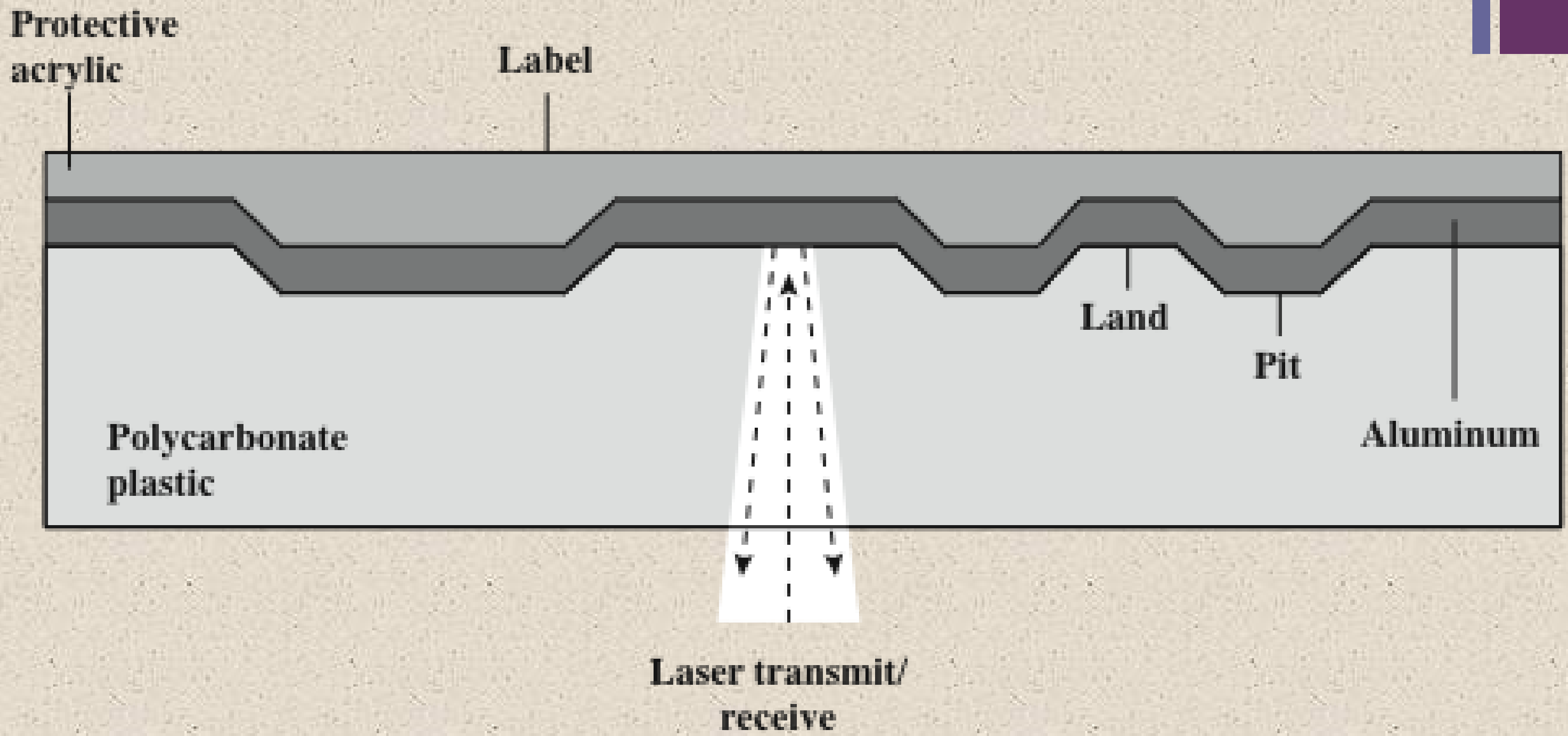


Figure 6.12 CD Operation

CD-ROM Block Format

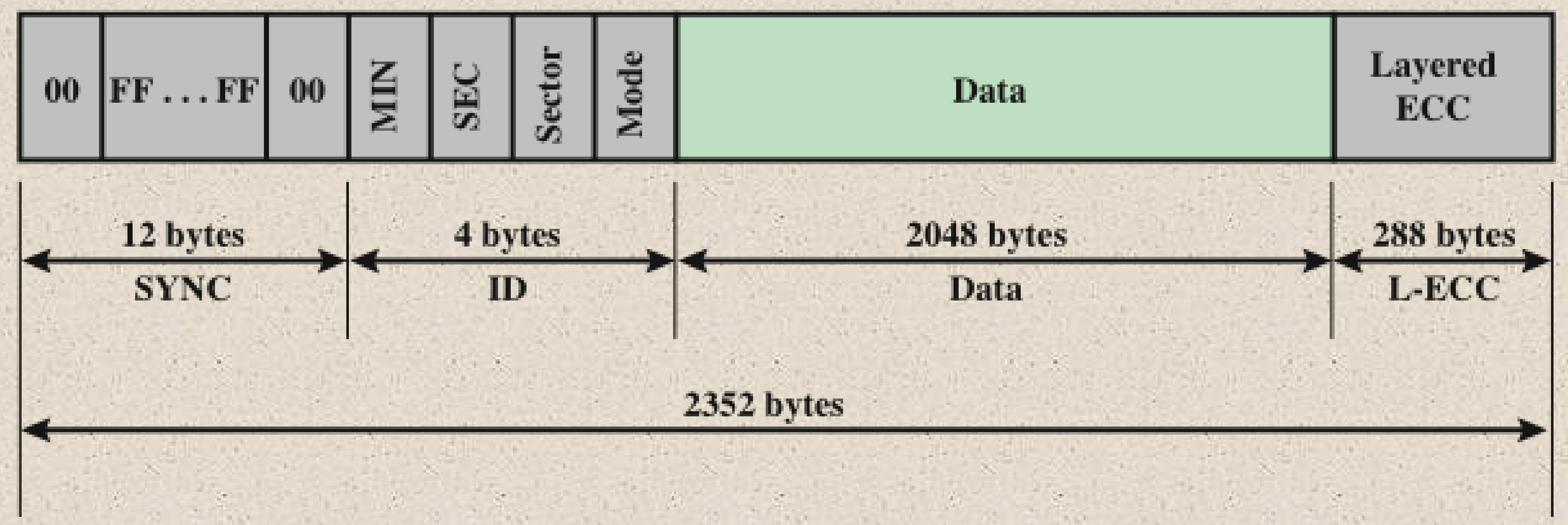
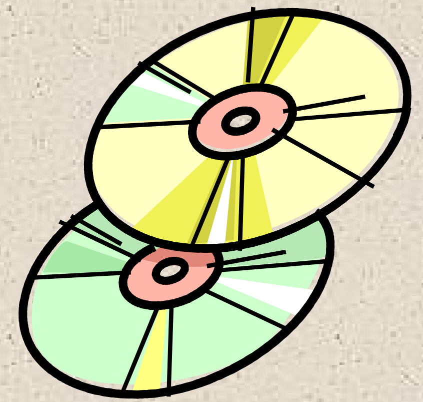


Figure 6.13 CD-ROM Block Format

CD-ROM

- CD-ROM is appropriate for the distribution of large amounts of data to a large number of users
- Because the expense of the initial writing process it is not appropriate for individualized applications
- The CD-ROM has two advantages:
  - The optical disk together with the information stored on it can be mass replicated inexpensively
  - The optical disk is removable, allowing the disk itself to be used for archival storage
- The CD-ROM disadvantages:
  - It is read-only and cannot be updated
  - It has an access time much longer than that of a magnetic disk drive



+

CD Recordable (CD-R)

- Write-once read-many
- Accommodates applications in which only one or a small number of copies of a set of data is needed
- Disk is prepared in such a way that it can be subsequently written once with a laser beam of modest-intensity
- Medium includes a dye layer which is used to change reflectivity and is activated by a high-intensity laser
- Provides a permanent record of large volumes of user data

CD Rewritable (CD-RW)

- Can be repeatedly written and overwritten
- Phase change disk uses a material that has two significantly different reflectivities in two different phase states
- Amorphous state
  - Molecules exhibit a random orientation that reflects light poorly
- Crystalline state
  - Has a smooth surface that reflects light well
- A beam of laser light can change the material from one phase to the other
- Disadvantage is that the material eventually and permanently loses its desirable properties
- Advantage is that it can be rewritten

Digital Versatile Disk (DVD)

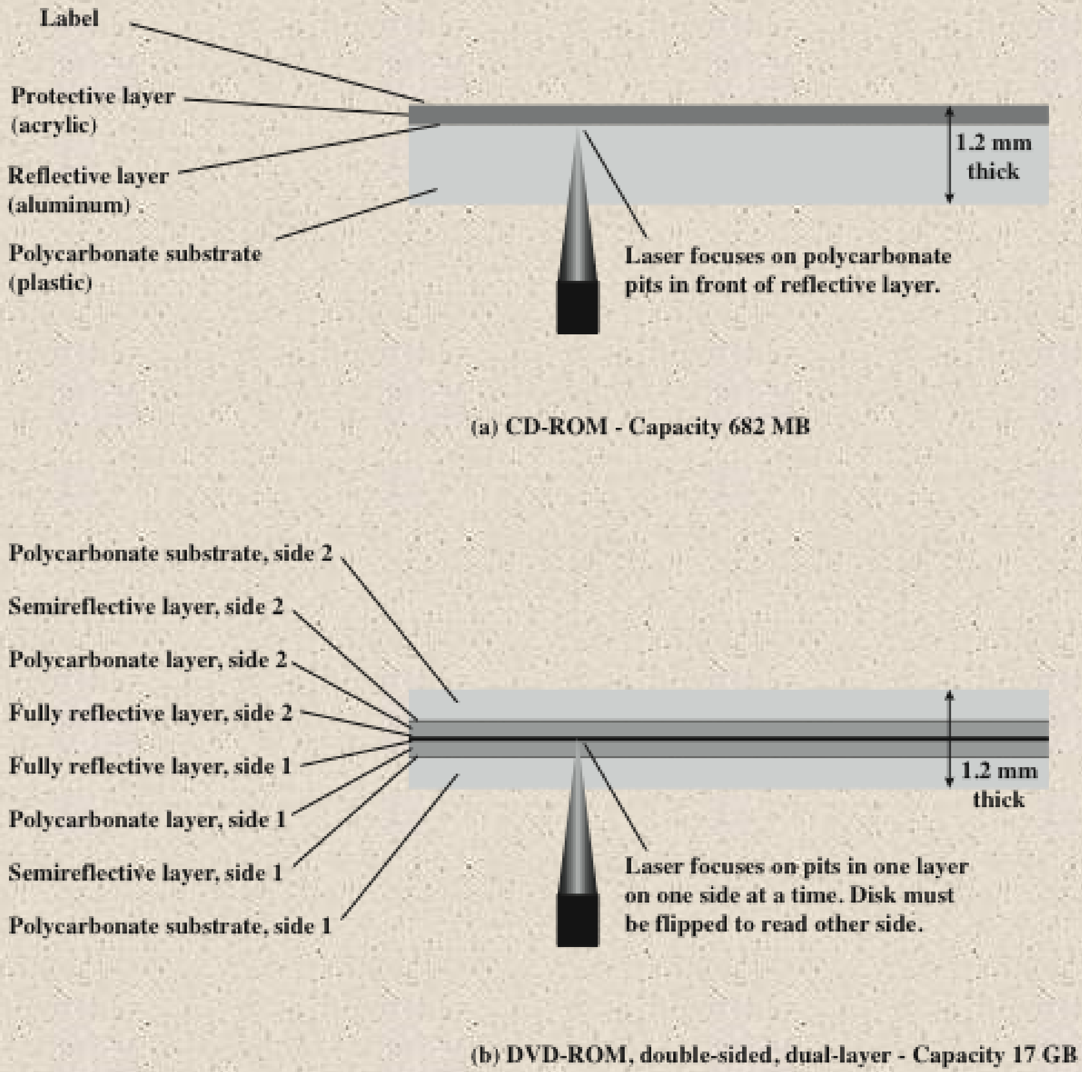


Figure 6.14 CD-ROM and DVD-ROM

High-Definition Optical Disks

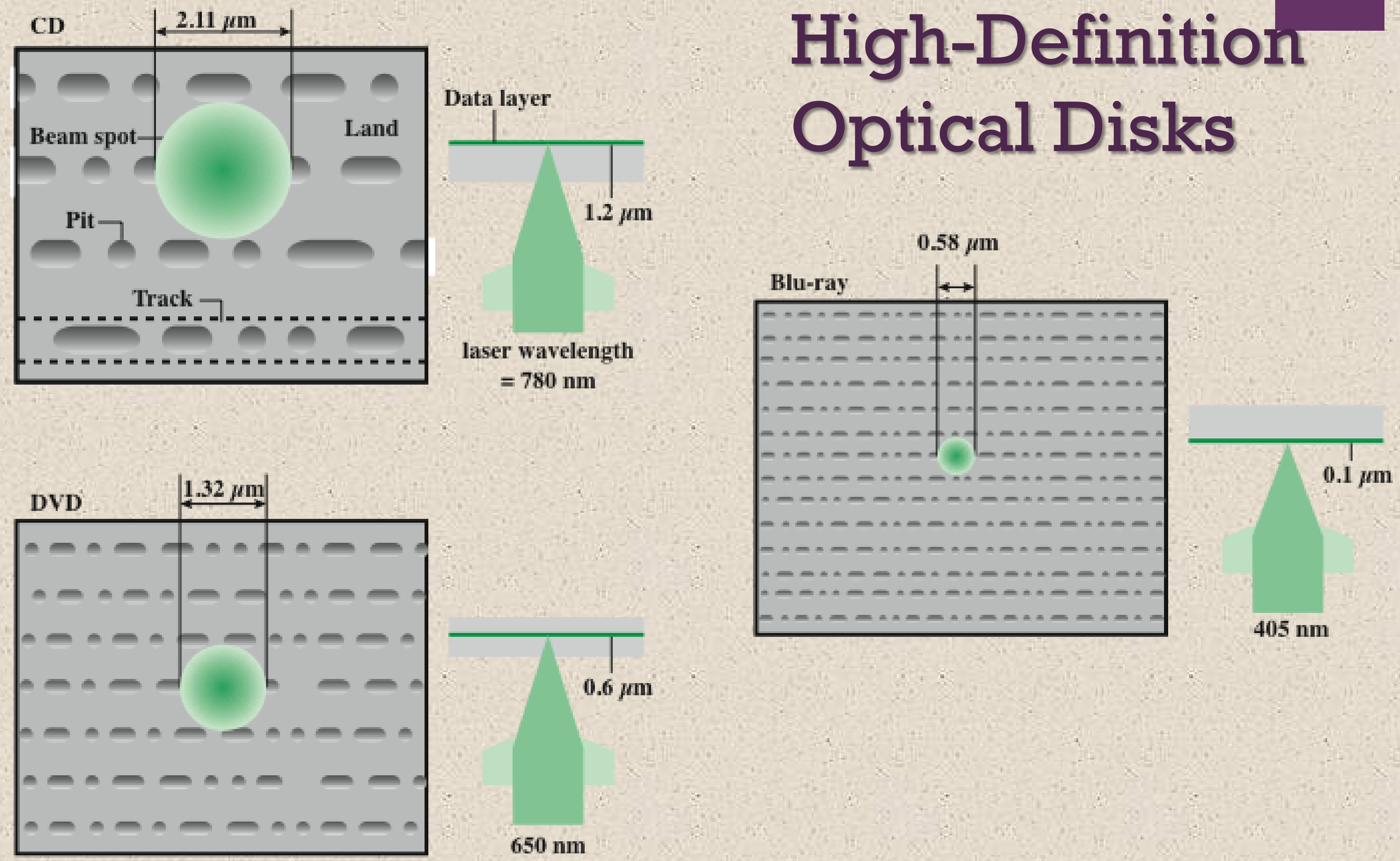


Figure 6.15 Optical Memory Characteristics

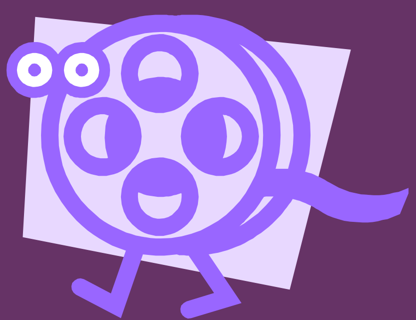
+

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*

+

Magnetic Tape Features



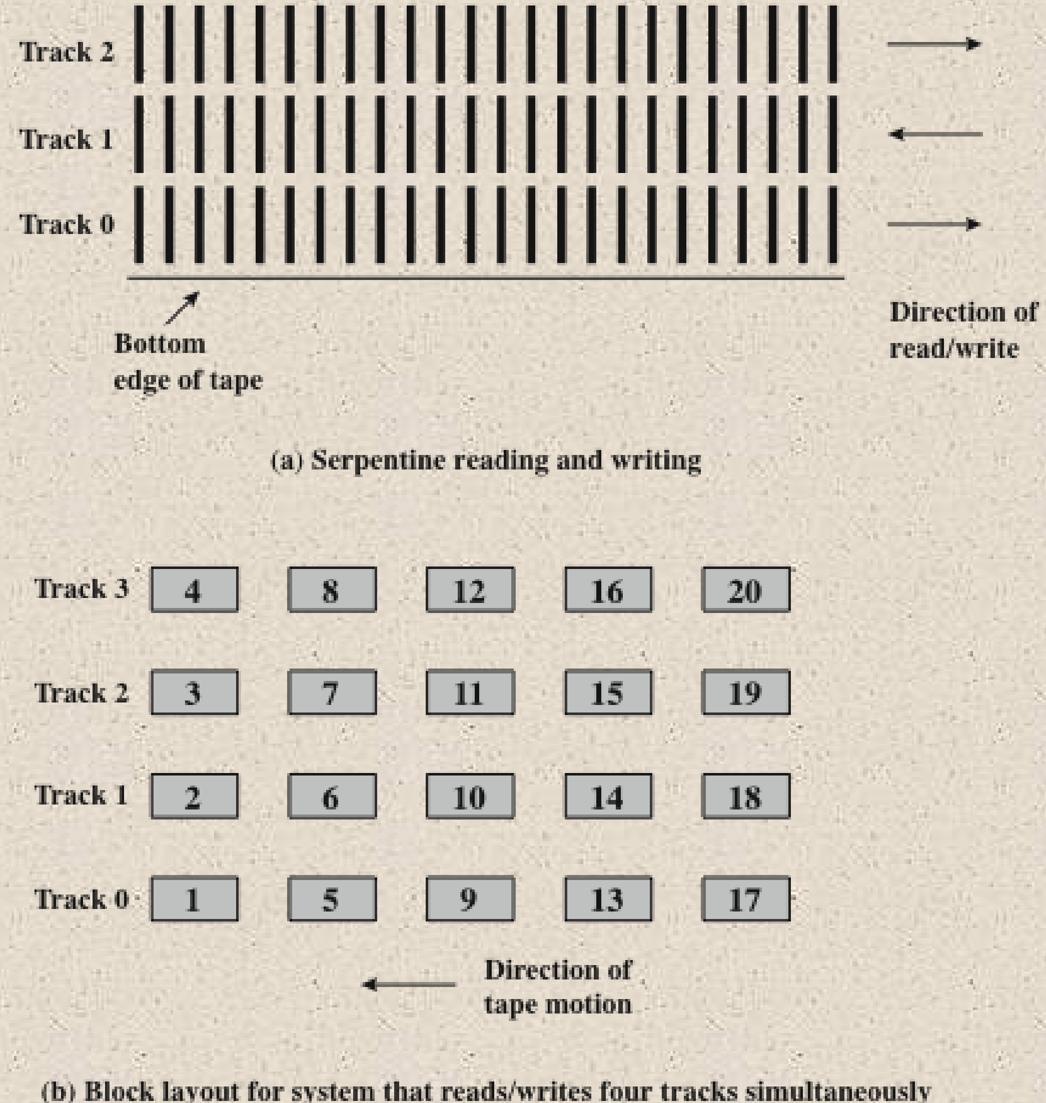


Figure 6.16 Typical Magnetic Tape Features

+ Summary

Chapter 6

External Memory

- Magnetic disk
  - Magnetic read and write mechanisms
  - Data organization and formatting
  - Physical characteristics
  - Disk performance parameters
- Solid state drives
  - Flash memory
  - SSD compared to HDD
  - SSD organization
  - Practical issues
- Magnetic tape
- RAID
  - RAID level 0
  - RAID level 1
  - RAID level 2
  - RAID level 3
  - RAID level 4
  - RAID level 5
  - RAID level 6
- Optical memory
  - Compact disk
  - Digital versatile disk
  - High-definition optical disks