AAiT

Computer Architecture & Organization

Chapter 13

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

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

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

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

Inductive Write/Magnetoresistive Read Head

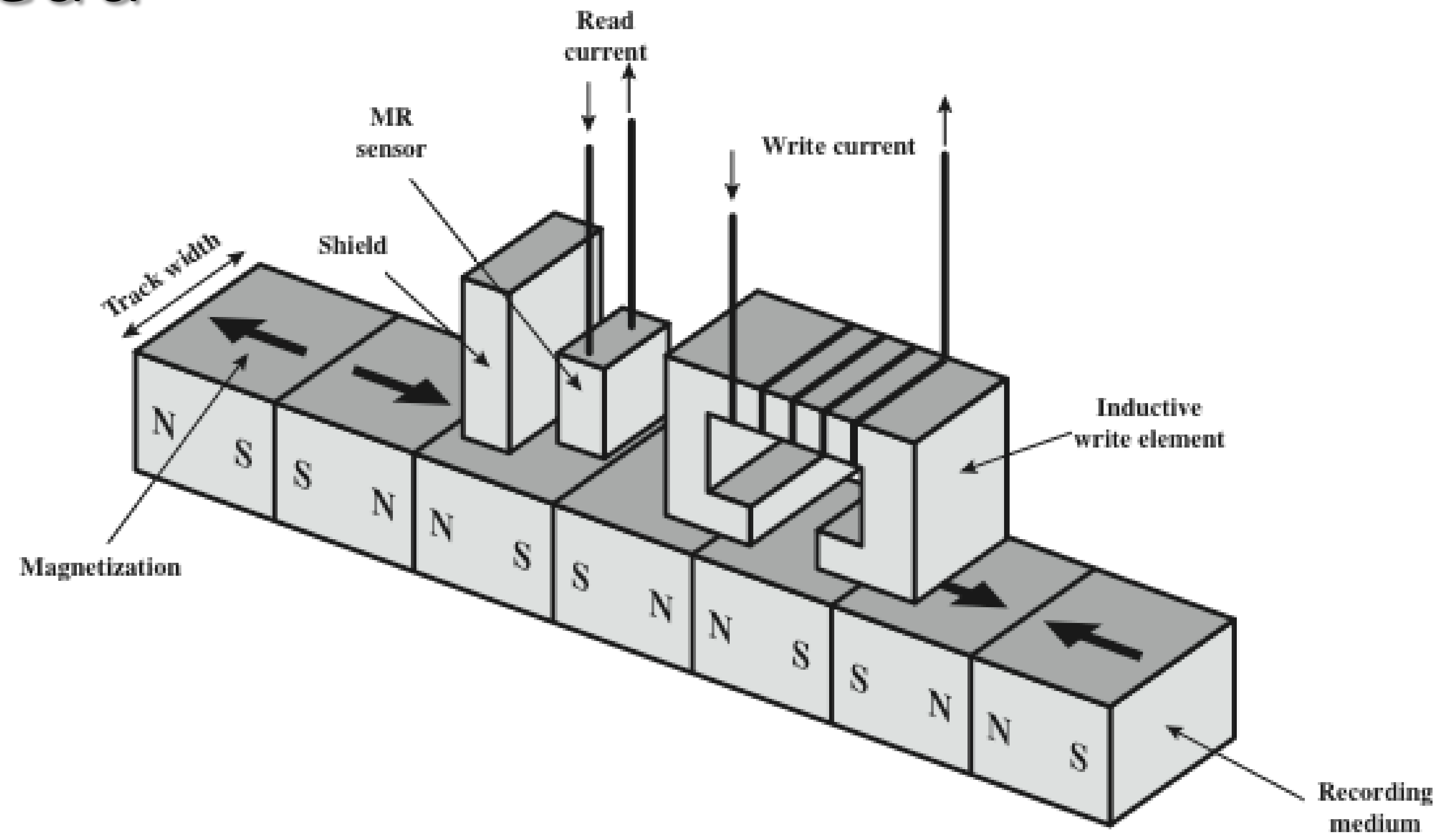


Figure 6.1 Inductive Write/Magnetoresistive Read Head

Disk Data Layout

- Adjacent tracks are separated by **gaps**
- Each track is the same width as the head.
- There are thousands of tracks per surface.
- Data are transferred to and from the disk in **sectors**
- There are typically hundreds of sectors per track,
- In most contemporary systems, fixed-length sectors are used, with 512 bytes being the nearly universal sector size
- adjacent sectors are separated by intratrack (intersector) gaps.

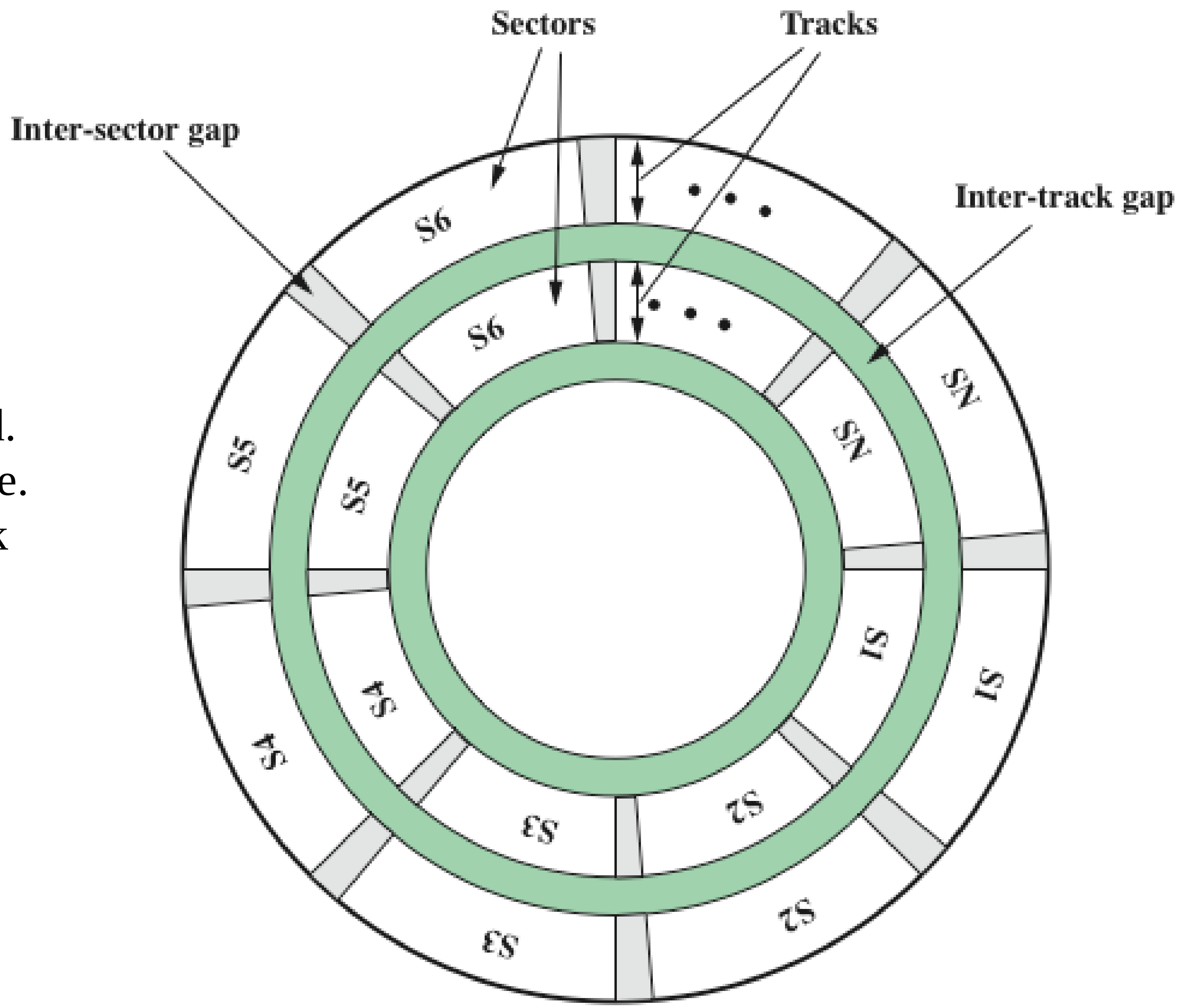
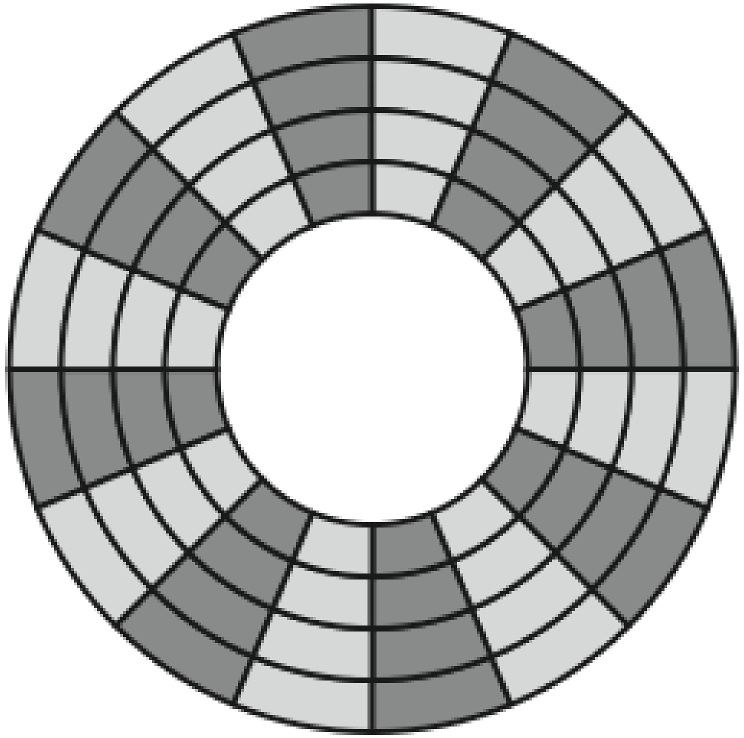


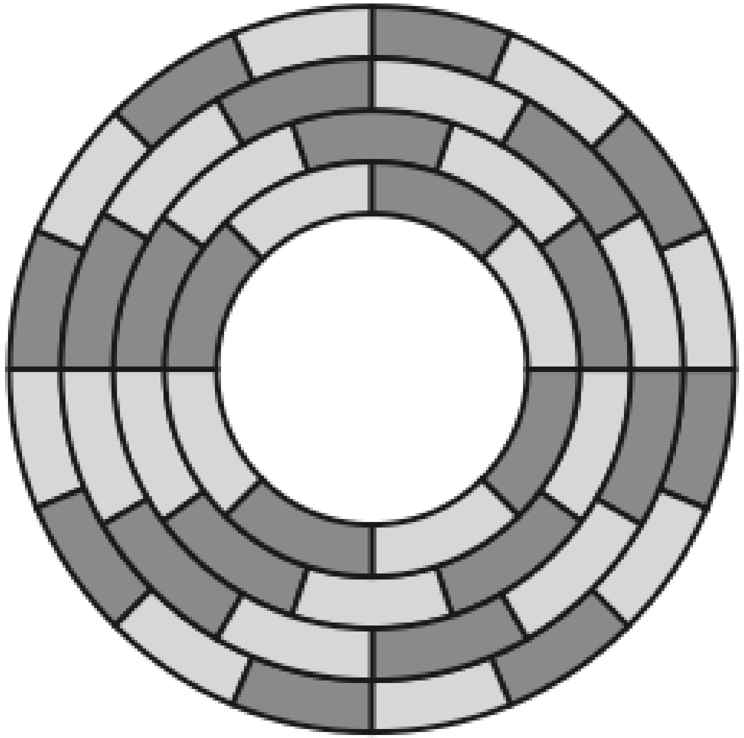
Figure 6.2 Disk Data Layout

Disk Layout Methods Diagram

- The disadvantage of CAV is that the amount of data that can be stored on the long outer tracks is the same as what can be stored on the short inner tracks.
- modern hard disk systems use a technique known as **multiple zone recording,** in which the surface is divided into a number of concentric zones.



(a) Constant angular velocity



(b) Multiple zoned recording

Figure 6.3 Comparison of Disk Layout Methods

Winchester Disk Format Seagate ST506

- Disk is formatted with some extra data used only by the disk drive and not accessible to the user
- In this case, each track contains 30 fixed-length sectors of 600 bytes each
- ID field: unique identifier or address used to locate a particular sector
- SYNCH: delimits the beginning of the field
- Head number identifies a head, because this disk has multiple surfaces

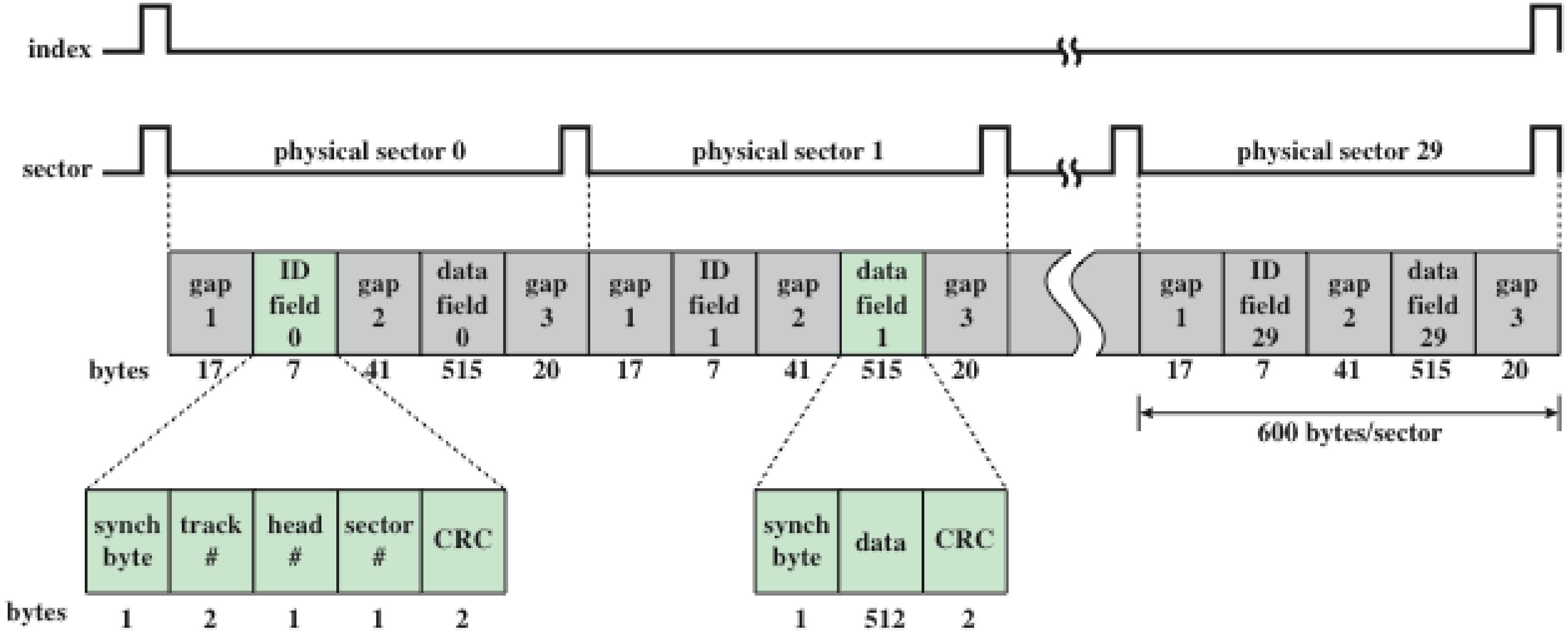


Figure 6.4 Winchester Disk Format (Seagate ST506)

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(Floppy)	Fixed gap
	Aerodynamic gap (Winchester)
Sides	
Single sided	
Double sided	

Table 6.1 Physical Characteristics of Disk Systems

Multiple Platters

- The set of all the tracks in the same relative position on all platters is referred to as a **cylinder.**

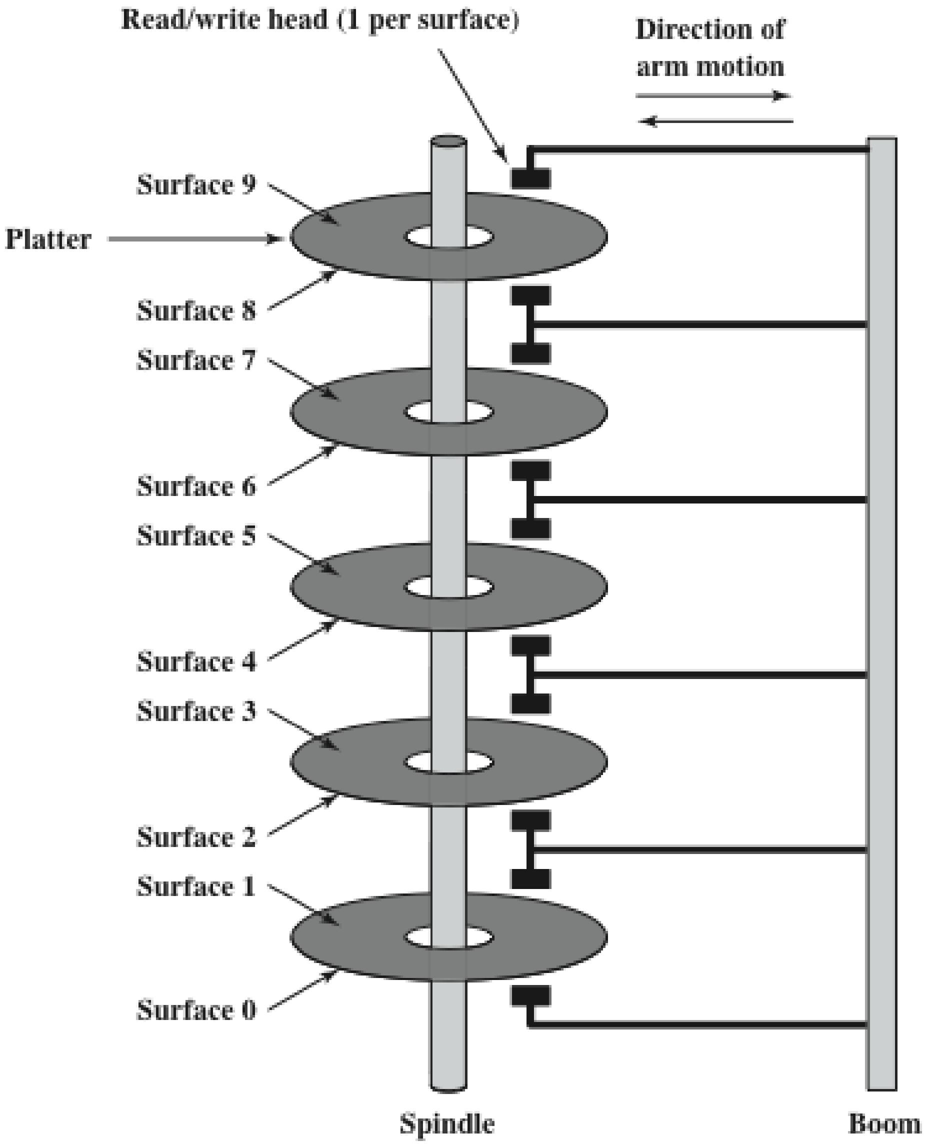


Figure 6.5 Components of a Disk Drive

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

Disk Performance Parameters

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

RAID

Redundant Array of

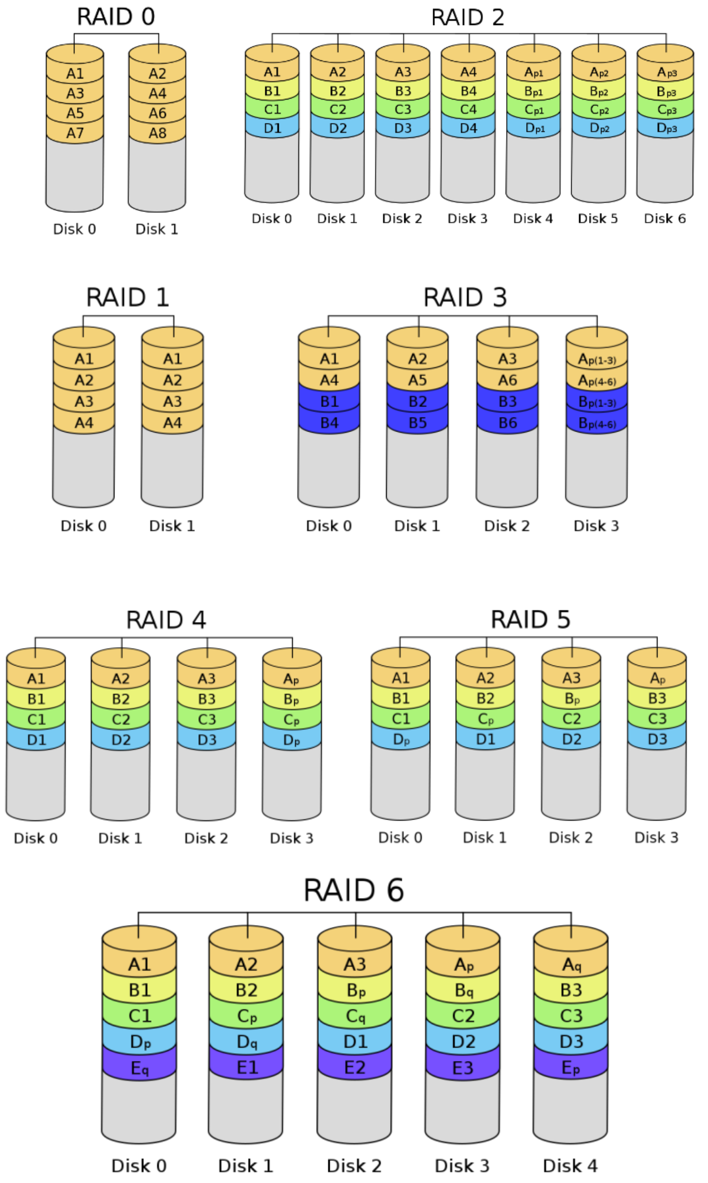
Independent Disks

- Consists of 7 levels
- Levels do not imply a hierarchical relationship but designate different design architectures that share three common characteristics:
  1) Set of physical disk drives viewed by the operating system as a single logical drive
  2) Data are distributed across the physical drives of an array in a scheme known as striping
  3) Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure

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 indicates the use of array management **software** to map between logical and physical disk space.
- This **software** may execute either in the disk subsystem or in a host computer.

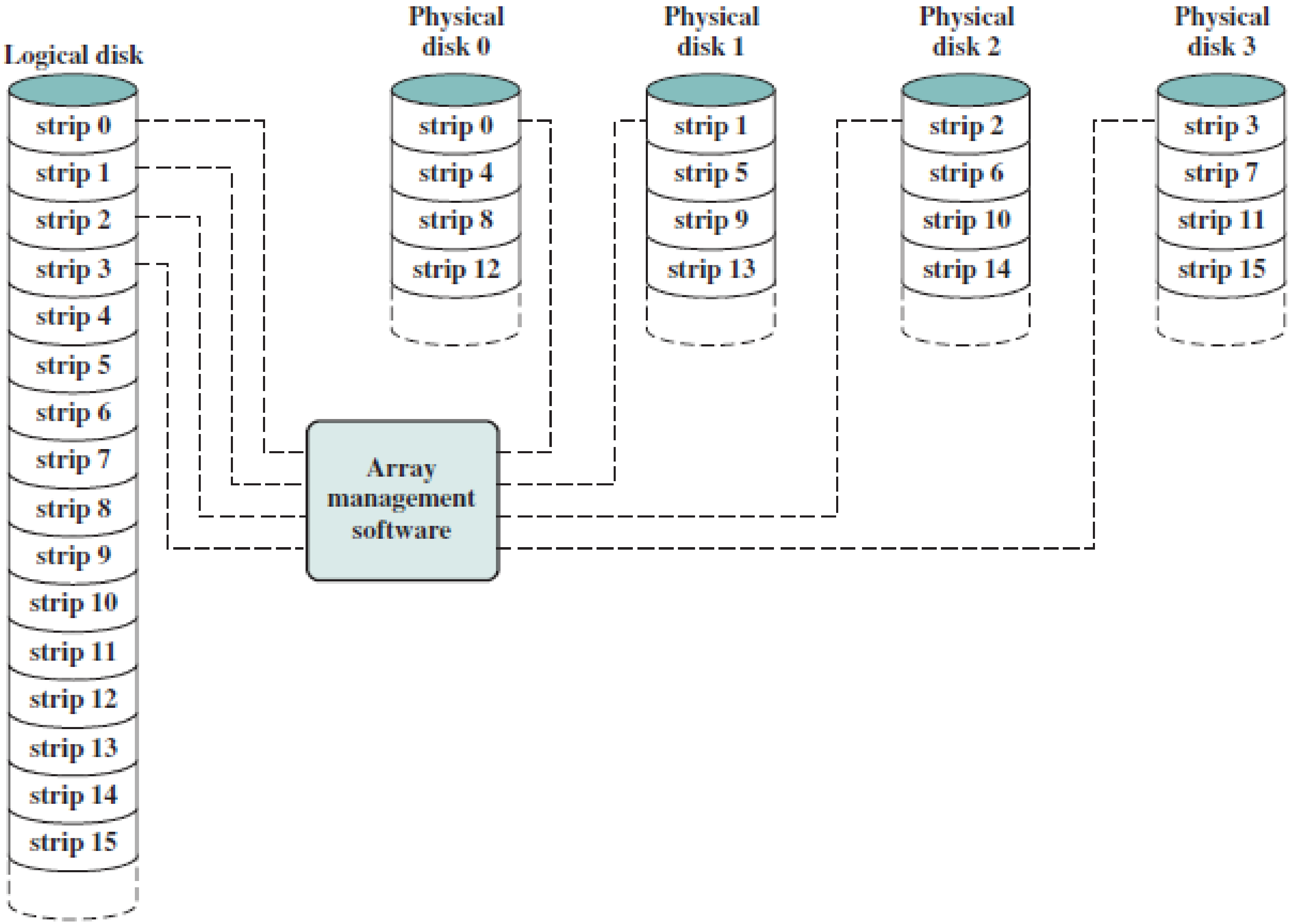


Figure 6.9 Data Mapping for a RAID Level 0 Array

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

- In a flash memory cell, a second thin oxide layer called a floating gate is added to the transistor in (a).
- Initially, the floating gate does not interfere with the operation of the transistor (b). In this state, the cell is deemed to represent binary 1
- Applying a large voltage across the oxide layer causes electrons to tunnel through it and become trapped on the floating gate, where they remain even if the power is disconnected (c). In this state, the cell is deemed to represent binary 0.

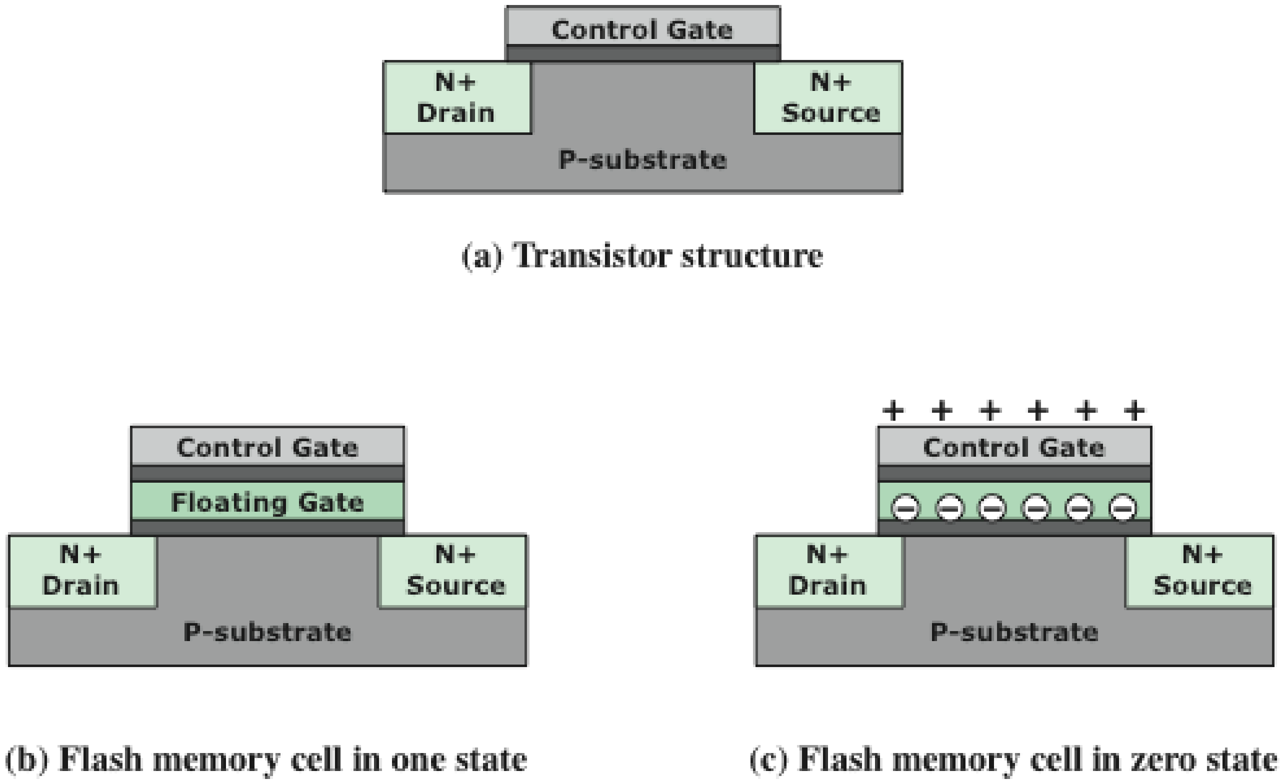


Figure 6.10
Flash Memory Operation

Solid State Drive (SSD)

AAiT

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

Table6.5 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

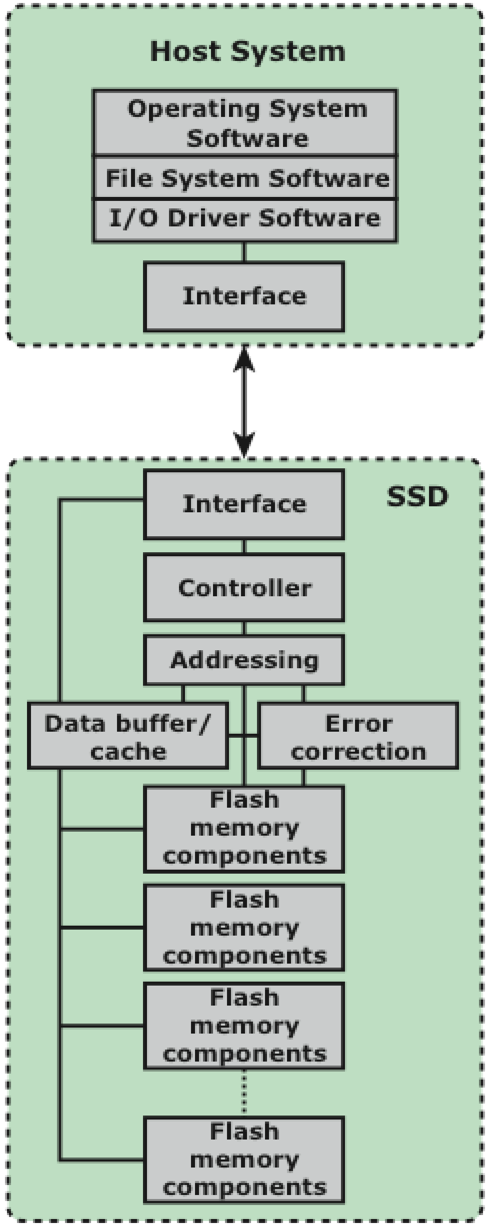


Figure 6.11 Solid State Drive Architecture

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

Optical Disk Products

- In 1983, one of the most successful consumer products of all time was introduced: the compact disk (CD) digital audio system

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

- Information is retrieved from a CD or CD-ROM by a low-powered laser housed in an optical-disk player, or drive unit
- If the laser beam falls on a pit, the light scatters and a low intensity is reflected back to the source.
- The areas between pits are called lands. A land is a smooth surface, which reflects back at higher intensity.
- The change between pits and lands is detected by a photo sensor and converted into a digital signal.
- The sensor tests the surface at regular intervals.
- The beginning or end of a pit represents a 1; when no change in elevation occurs between intervals, a 0 is recorded.

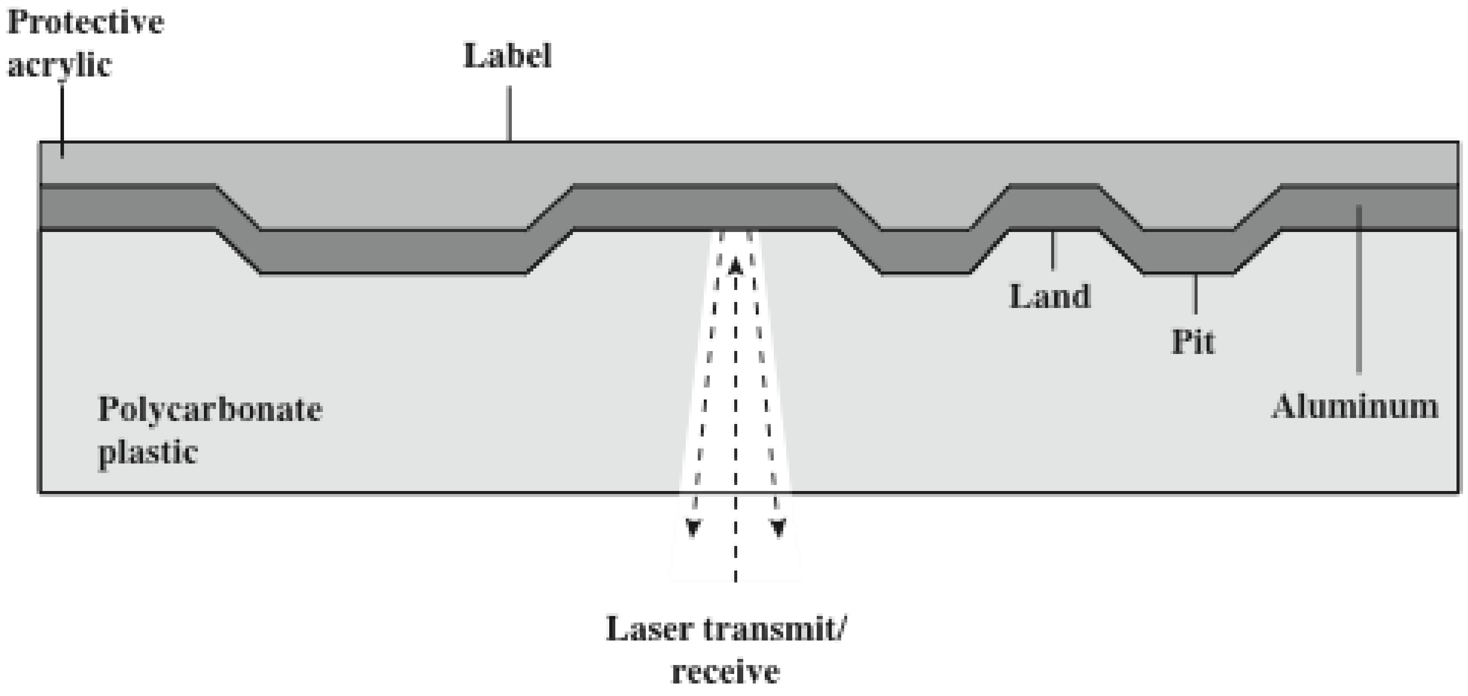


Figure 6.12 CD Operation

CD-ROM Block Format

- **Sync:** The sync field identifies the beginning of a block. It consists of a byte of all 0s, 10 bytes of all 1s, and a byte of all 0s
    - Mode 0 specifies a blank data field;
    - Mode 1 specifies the use of an error-correcting code and 2048 bytes of data;
    - Mode 2 specifies 2336 bytes of user data with no error-correcting code.

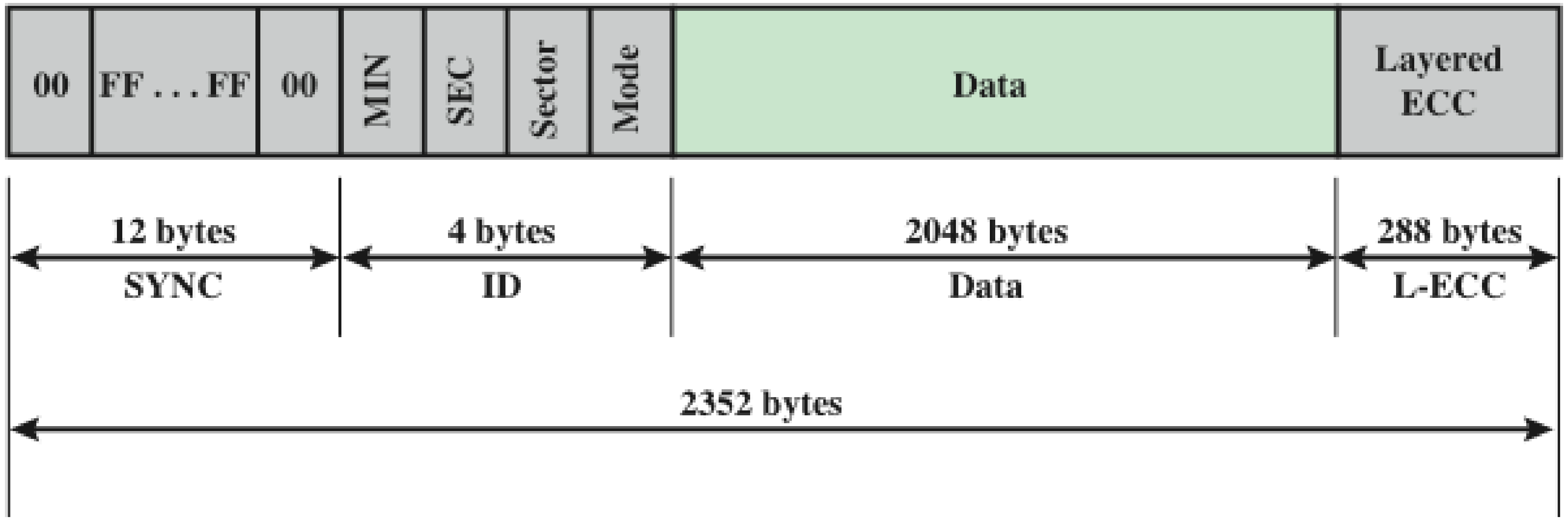


Figure 6.13 CD-ROM Block Format

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)

- The DVD has replaced the videotape used in video cassette recorders (VCRs) and, more important for this discussion, replace the CD-ROM in personal computers and servers

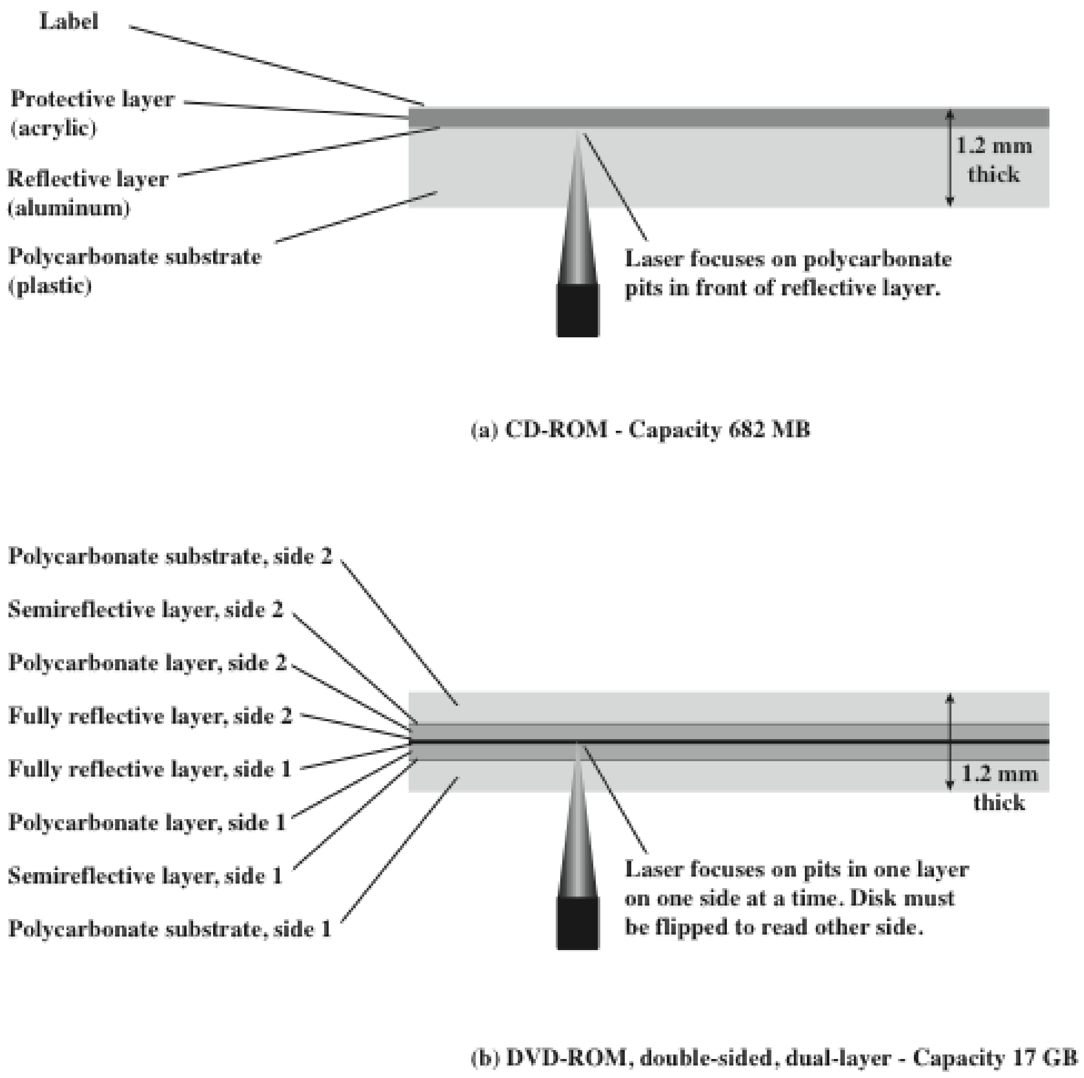


Figure 6.14 CD-ROM and DVD-ROM

High-Definition Optical Disks

- High-definition optical disks are designed to store high-definition videos and to provide significantly greater storage capacity compared to DVDs
- The HD DVD scheme can store 15 GB on a single layer on a single side
- Blu-ray can store 25 GB on a single layer

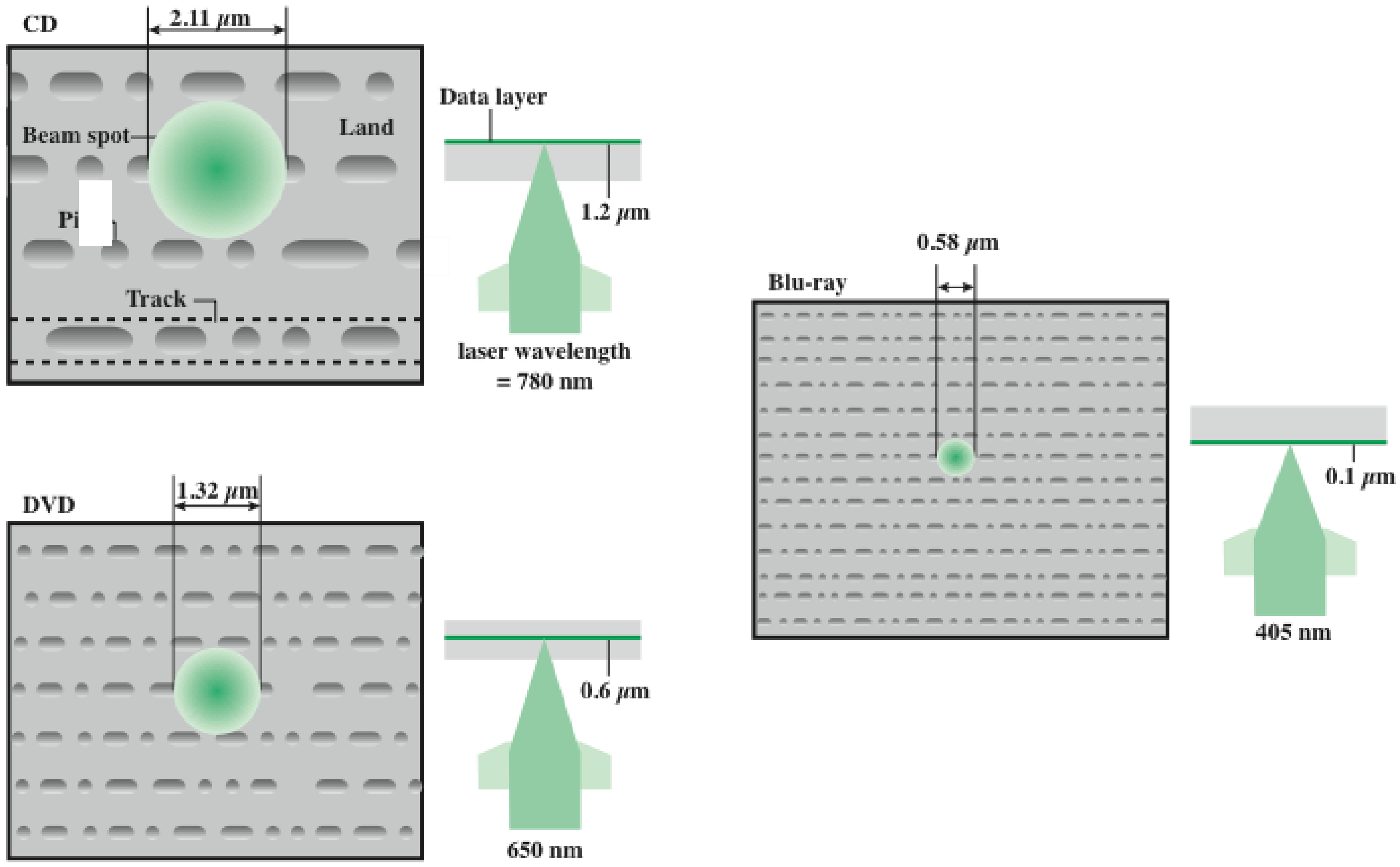


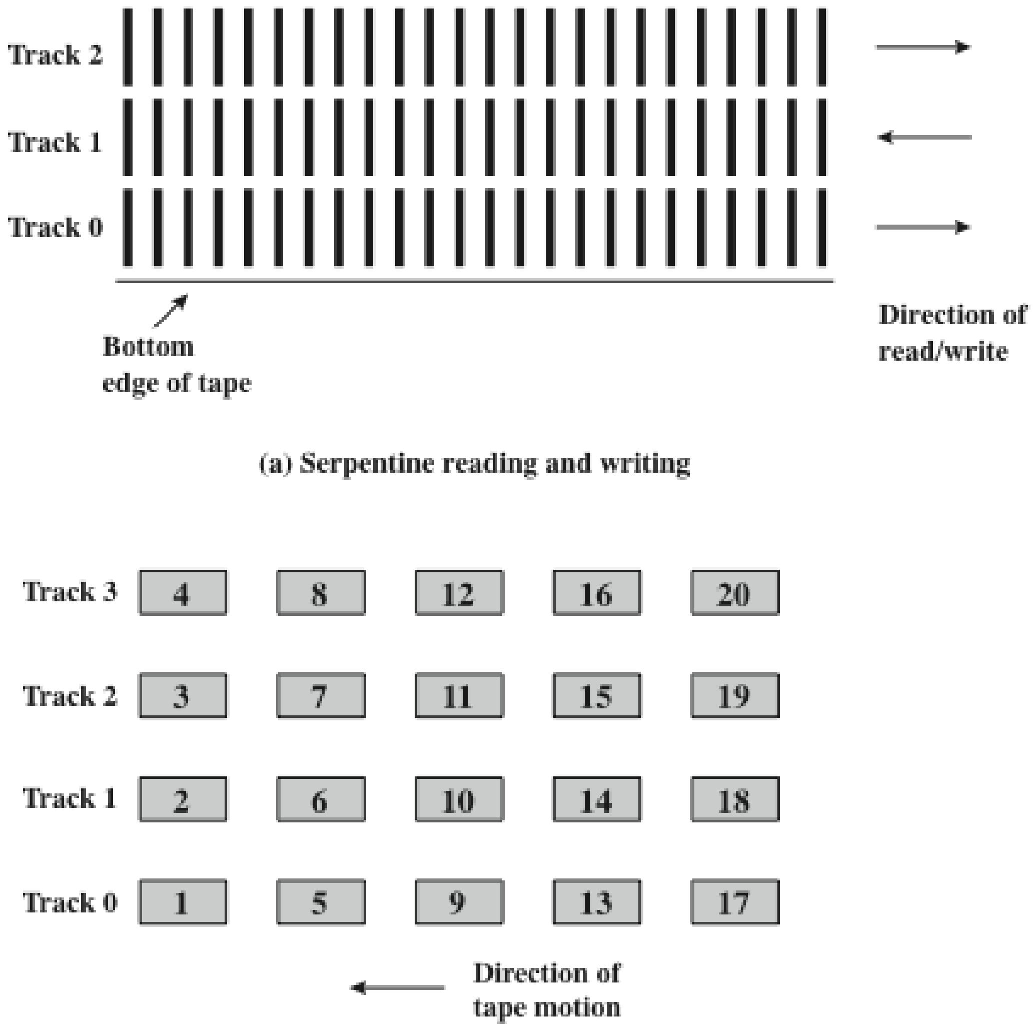
Figure 6.15 Optical Memory Characteristics

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

- A tape drive is a *sequential-access device*
- In **serpentine recording**
  - First set of bits is recorded along the whole length of the tape.
  - When the end of the tape is reached, the heads are repositioned to record a new track, and the tape is again recorded on its whole length, this time in the opposite direction
  - That process continues, back and forth, until the tape is full



(b) Block layout for system that reads/writes four tracks simultaneously

Figure 6.16 Typical Magnetic Tape Features

Table 6.7
LTO Tape Drives

- The dominant tape technology today is a cartridge system known as linear tape-open (LTO).

	LTO-1	LTO-2	LTO-3	LTO-4	LTO-5	LTO-6	LTO-7	LTO-8
Release date	2000	2003	2005	2007	2010	TBA	TBA	TBA
Compressed capacity	200 GB	400 GB	800 GB	1600 GB	3.2 TB	8 TB	16 TB	32 TB
Compressed transfer rate	40 MB/s	80 MB/s	160 MB/s	240 MB/s	280 MB/s	525 MB/s	788 MB/s	1.18 GB/s
Linear density (bits/mm)	4880	7398	9638	13250	15142			
Tape tracks	384	512	704	896	1280			
Tape length (m)	609	609	680	820	846			
Tape width (cm)	1.27	1.27	1.27	1.27	1.27			
Write elements	8	8	16	16	16			
WORM?	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Encryption Capable?	No	No	No	Yes	Yes	Yes	Yes	Yes
Partitioning?	No	No	No	No	Yes	Yes	Yes	Yes