Understanding Operating Systems Sixth Edition

Chapter 7
Device Management

Learning Objectives

After completing this chapter, you should be able to describe:

- Features of dedicated, shared, and virtual devices
- Differences between sequential and direct access media
- Concepts of blocking and buffering and how they improve I/O performance
- Roles of seek time, search time, and transfer time in calculating access time

Learning Objectives (cont'd.)

- Differences in access times in several types of devices
- Critical components of the input/output subsystem, and how they interact
- Strengths and weaknesses of common seek strategies, including FCFS, SSTF, SCAN/LOOK, C-SCAN/C-LOOK, and how they compare
- Different levels of RAID and what sets each apart from the others

Device Management

- Monitoring the status of each device
  - Storage, monitor, USB tools
- Enforcing preset policies to determine which process will get the device and how long
- Allocating device appropriately
- Deallocating each device
  - Task level
  - Job level

Types of Devices

- Three categories: dedicated, shared, and virtual
- **Dedicated Devices**
  - Device assigned to one job at a time
    - For entire time job is active (or until released)
    - Example: tape drives, printers, and plotters
  - Disadvantage
    - Inefficient if device is not used 100%
    - Allocated for duration of job's execution

Types of Devices (cont'd.)

- **Shared Devices**
  - Device assigned to several processes
    - Example: direct access storage device (DASD)
      - Processes share DASD simultaneously
      - Requests interleaved
  - Device manager supervision
    - Controls interleaving
      - Predetermined policies determine conflict resolution

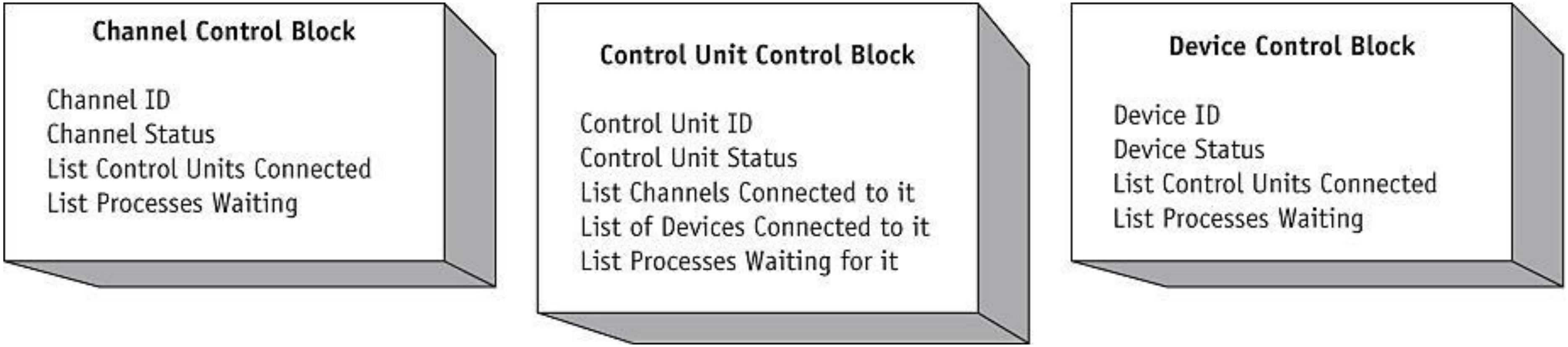
Types of Devices (cont'd.)

- **Virtual Devices**
  - Dedicated and shared device combination
  - Dedicated devices transformed into shared devices
    - Example: printer
      - Converted by spooling program
  - Spooling
    - Speeds up slow dedicated I/O devices
    - Example: **universal serial bus (USB) contro**

      - Interface between operating system, device drivers, applications, and devices attached via USB host (up to 127 devices)
      - Assigns bandwidth to each device: priority-based
        - » High, medium, or low priority

Management of I/O Requests

- I/O traffic controller
  - Watches status of devices, control units, channels
  - Three main tasks
    - Determine if path available
    - If more than one path available, determine which one to select
    - If paths all busy, determine when one is available
  - Maintains database containing each unit's status and connections grouped into:
    - Channel Control Blocks, Control Unit Control Blocks, Device Control Blocks



(figure 7.1)
Each control block contains the information it needs to manage the channels, control units, and devices in the I/O subsystem.

Management of I/O Requests (cont'd.)

- I/O scheduler
  - Same job as process scheduler (Chapter 4)
  - Allocates devices, control units, and channels
  - If requests greater than available paths
    - Decides which request to satisfy first: based on different criteria
  - In many systems
    - I/O requests not preempted
  - For some systems
    - Allow preemption with I/O request subdivided
    - Allow preferential treatment for high-priority requests

Management of I/O Requests (cont'd.)

- I/O device handler
  - Performs actual data transfer
    - Processes device (I/O) interrupts
    - Handles error conditions
    - Provides detailed scheduling algorithms
  - Device dependent
  - Each I/O device type has its own device handler algorithm
    - We will explore several algorithms for hard disk drives

I/O Devices in the Cloud

- Local operating system's role in accessing remote I/O devices
  - Essentially the same role performed accessing local devices
- Cloud provides access to many more devices

Storage media

- First secondary storage mediums were punch cards, paper tapes, then followed by magnetic tape and others
- Two groups
  - **Sequential access media**
    - Records stored sequentially
  - **Direct access storage devices (DASD)**
    - Records stored sequentially
    - Records stored using direct access files
- Vast differences
  - Speed and sharability

Sequential Access Storage Media

- **Magnetic tape**
  - Early computer systems: routine secondary storage
  - Today's use: routine archiving and data backup
  - Records stored serially
    - Record length determined by application program
    - Record identified by position on tape
    - Record access
      - Tape mount
      - Fast-forwarded to record
      - Tape rotates passing under read/write he
        access requested for read or write
    - Time-consuming process

nit Figure 90. IBM 7330 Magnetic Tape Unit

Sequential Access Storage Media (cont'd.)

(figure 7.1)

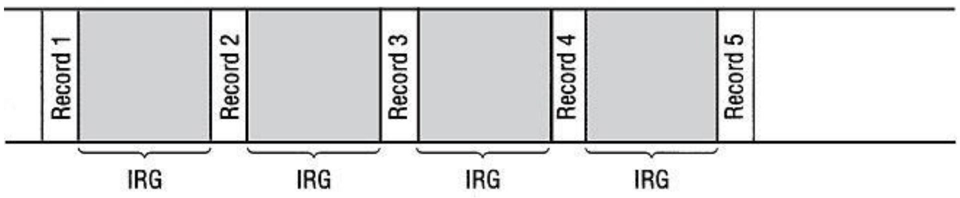
Nine-track magnetic tape with three characters recorded using odd parity. A 1/2-inch wide reel of tape, typically used to back up a mainframe computer, can store thousands of characters, or bytes, per inch.

- **Tape density**: characters recorded per inch
  - Depends upon storage method (individual or blocked)
- Tape reading/writing mechanics
  - Tape moves under read/write head when needed

Sequential Access Storage Media (cont'd.)

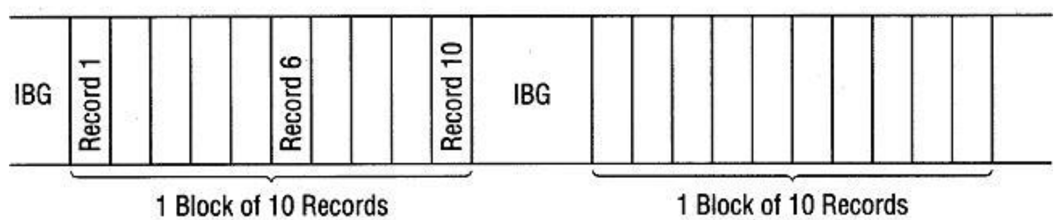
- **Interrecord gap (IRG)**
  - ½ inch gap inserted between each record
  - Same size regardless of records it separates
- **Blocking:** group records into blocks
- **Transfer rate:** (tape density) x (transport speed)
  - Byte per inch (bpi) x inch per second (ips)
- **Interblock gap (IBG)**
  - ½ inch gap inserted between each block
  - More efficient than individual records and IRG

Sequential Access Storage Media (cont'd.)



(figure 7.3)
IRGs in magnetic tape. Each record requires only 1/10 inch of tape. When 10 records are stored individually on magnetic tape, they are separated by IRGs, which adds up to 4.5 inches of tape. This totals 5.5 inches of tape.
© Cengage Learning 2014

(figure 7.4)
Two blocks of records stored on magnetic tape, each preceded by an IBG of 1/2 inch. Each block holds 10 records, each of which is still 1/10 inch. The block, however, is 1 inch, for a total of 1.5 inches.
© Cengage Learning 2014



Sequential Access Storage Media (cont'd.)

- Blocking advantages
  - Fewer I/O operations needed
  - Less wasted tape
- Blocking disadvantages
  - Overhead and software routines needed for blocking, deblocking, and record keeping
  - Buffer space wasted
    - When only one logical record needed

Sequential Access Storage Media (cont'd.)

Benchmarks	Access Time
Maximum access	2.5 minutes
Average access	1.25 minutes
Sequential access	3 milliseconds

(table 7.1)
Access times for 2400-foot magnetic tape with a tape transport speed of 200 ips.

- Advantages
  - Low cost, compact storage capabilities, good for magnetic disk backup and long-term archival
- Disadvantages
  - **Access time**
  - Poor for routine secondary storage except files with very high (90 to 100 percent) sequential activity
  - Poor for interactive applications

Direct Access Storage Devices

- Directly read or write to specific disk area
  - Random access storage devices
- Three categories
  - Magnetic disks
  - Optical discs
  - Solid state (Flash) memory
- Access time variance
  - Not as wide as magnetic tape
    - Still: Record location directly affects access time

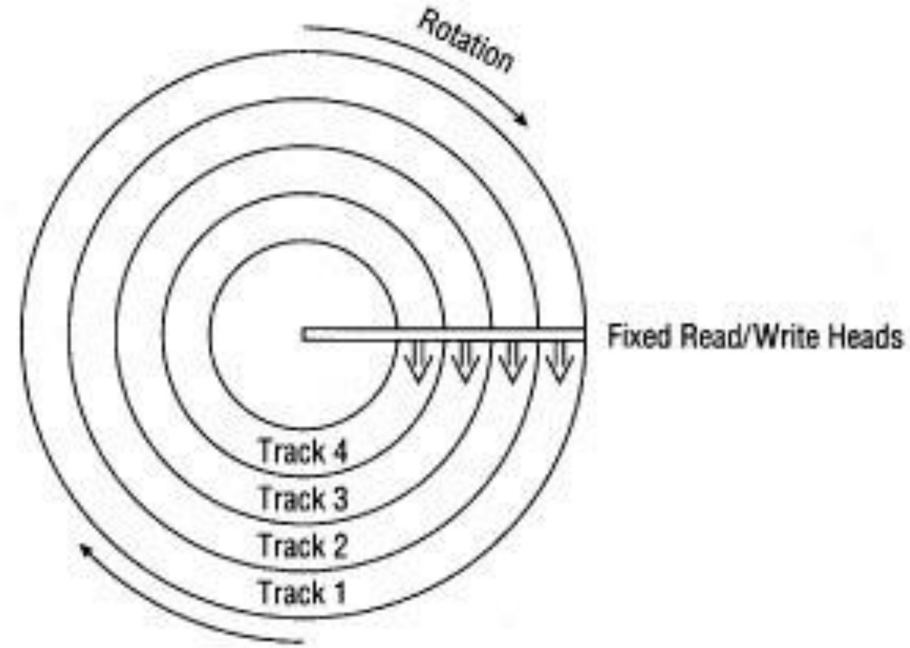
Fixed-Head Magnetic Disk Storage

- Looks like a large CD or DVD
- Used around 1980s https://www-03.ibm.com/ibm/history/exhibits/storage/storage_2305.html
  - Covered with magnetic film
  - Formatted
    - Both sides (usually) in concentric circles called **tracks**
  - Data recorded serially on each track
    - Fixed read/write head positioned over data
- Advantages
  - Fast (more so than movable head)
- Disadvantages
  - High cost and reduced storage

Fixed-Head Magnetic Disk Storage (cont'd.)

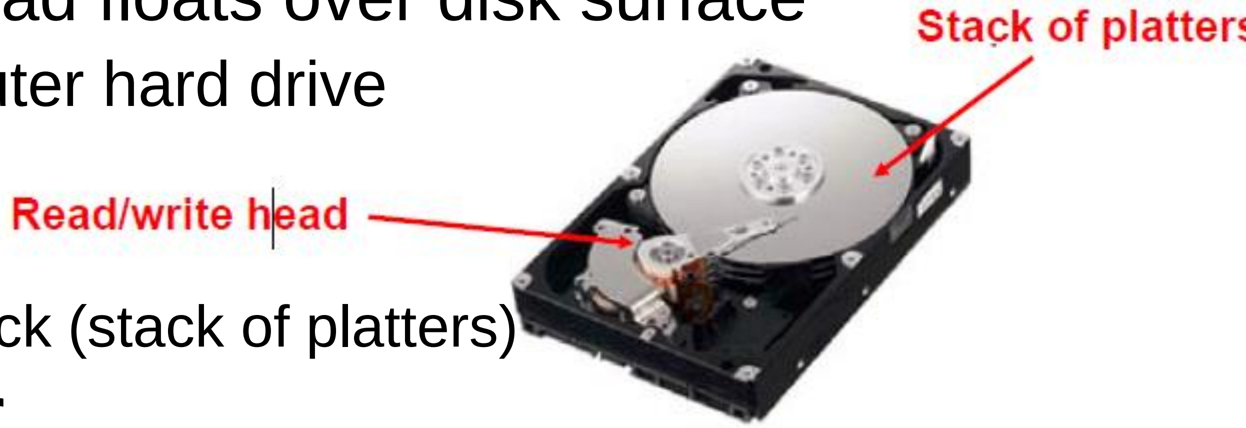
(figure 7.4)

A fixed-head disk with four read/write heads, one per track.



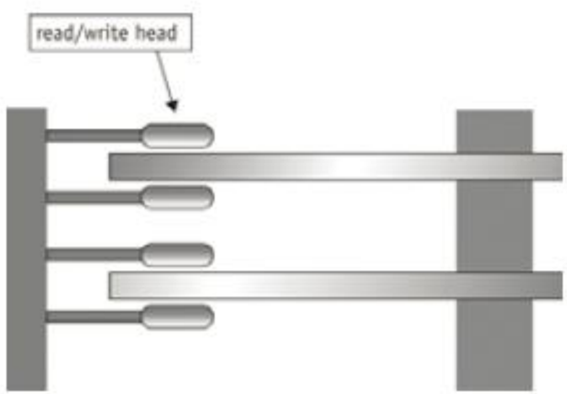
Movable-Head Magnetic Disk Storage

- One read/write head floats over disk surface
  - Example: computer hard drive
  - Disks
    - Single platter
    - Part of disk pack (stack of platters)
- **Disk pack platter**
  - Two recording surfaces
    - Exception: top and bottom platters
  - Surface formatted with concentric tracks
  - Track number varies
    - 1000+ (high-capacity disk)



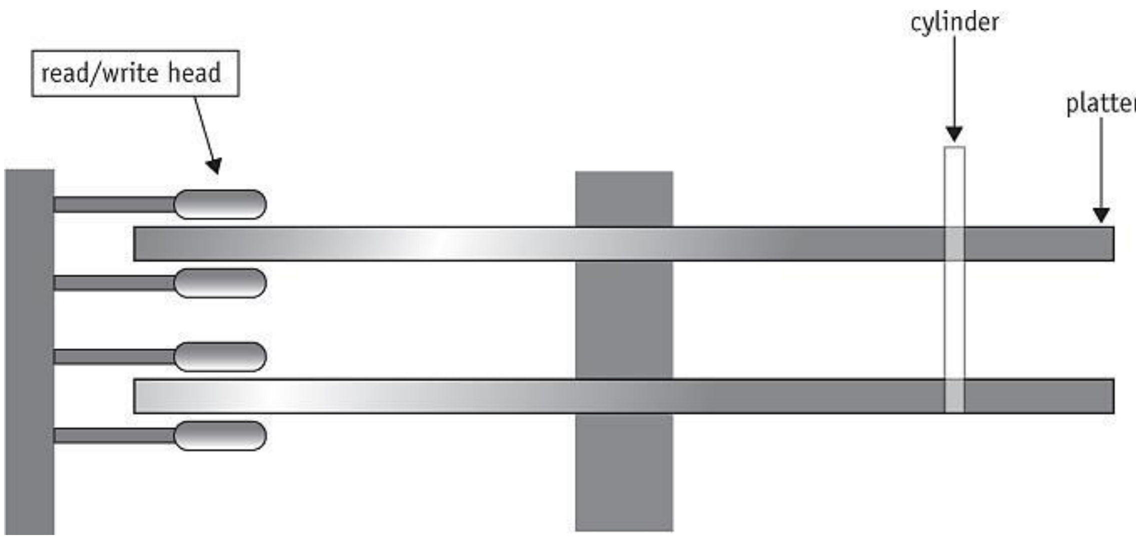
Movable-Head Magnetic Disk Storage (cont'd.)

- Disk pack platter (cont'd.)
  - Track surface number (Two surfaces at each disk)
    - Track zero: outermost concentric circle on each surface
    - Center: contains highest-numbered track



Magnetic Disk Storage

- Computer hard drives
- Single platter or stack of magnetic platters

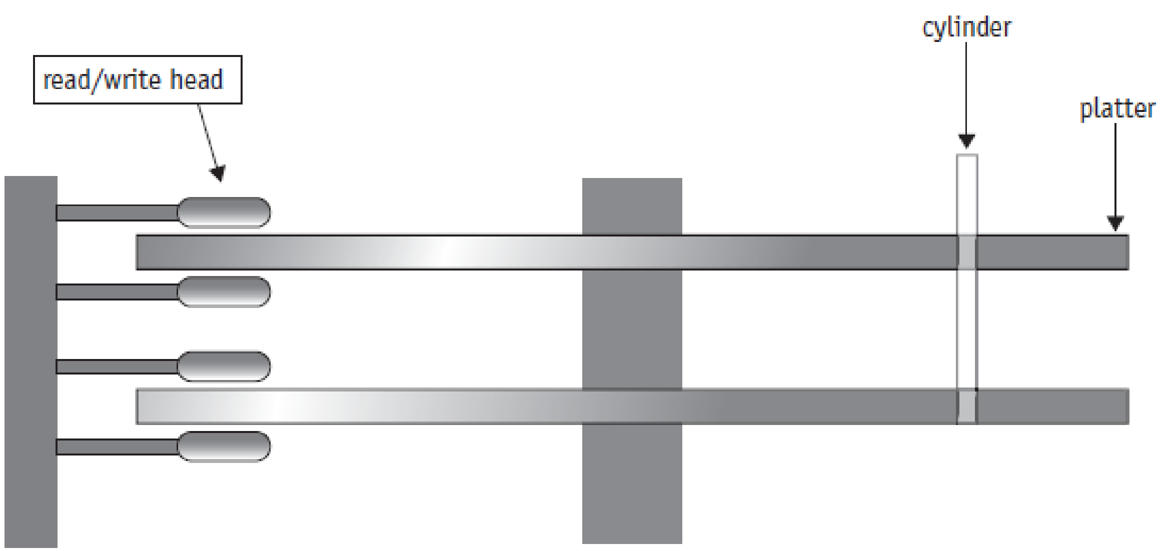


(figure 7.5)
A disk pack is a stack of magnetic platters. The read/write heads move between each pair of surfaces, and all of the heads are moved in unison by the arm.

Movable-Head Magnetic Disk Storage (cont'd.)

- Arm moves over all heads in unison
  - Slower: fill disk pack surface-by-surface
  - Faster: fill disk pack track-by-track
- Virtual cylinder: fill track zero

- Record access system requirements
  - Cylinder number, surface number, sector number



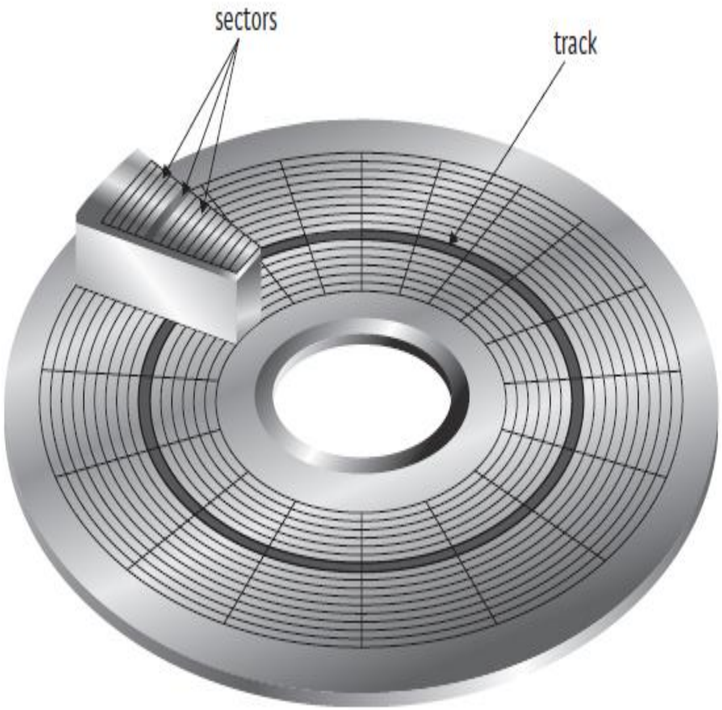
(figure 7.5)

A disk pack is a stack of magnetic platters. The read/write heads move between each pair of surfaces, and all of the heads are moved in unison by the arm.

Magnetic Storage

(figure 7.7)

On a magnetic disk, the sectors are of different sizes: bigger at the rim and smaller at the center. The disk spins at a constant angular velocity (CAV) to compensate for this difference. Some optical discs can read and write on multiple layers, greatly enhancing storage capacity.



- To access a given record system needs:
  - Cylinder number
  - Surface number
  - Sector number

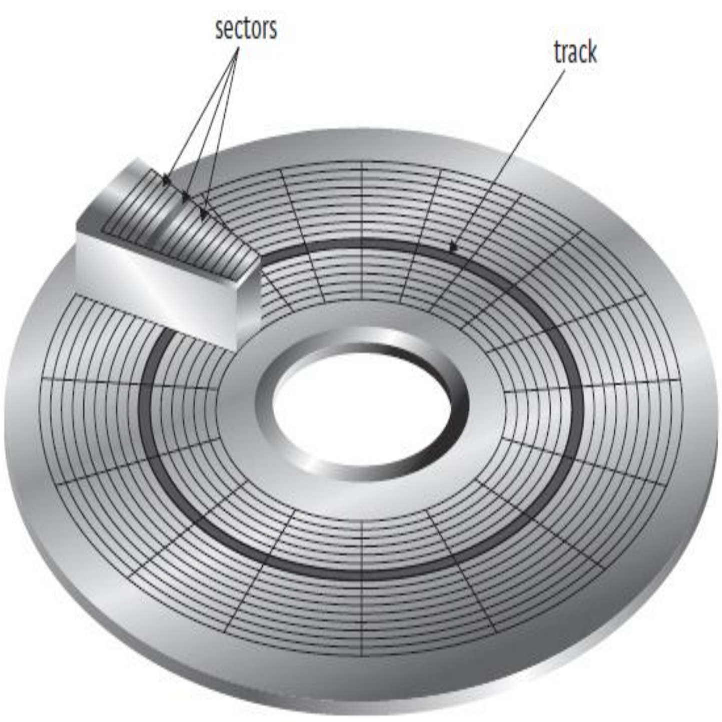
Access Times

- File access time factors
  - Seek time (slowest)
    - Time to position read/write head on track
    - Does not apply to fixed read/write head devices
  - Search time
    - Rotational delay
    - Time to rotate DASD
    - Rotate until desired record under read/write head
  - Transfer time (fastest)
    - Time to transfer data
    - Secondary storage to main memory transfer

Optical Disc Storage vs. Magnetic Storage

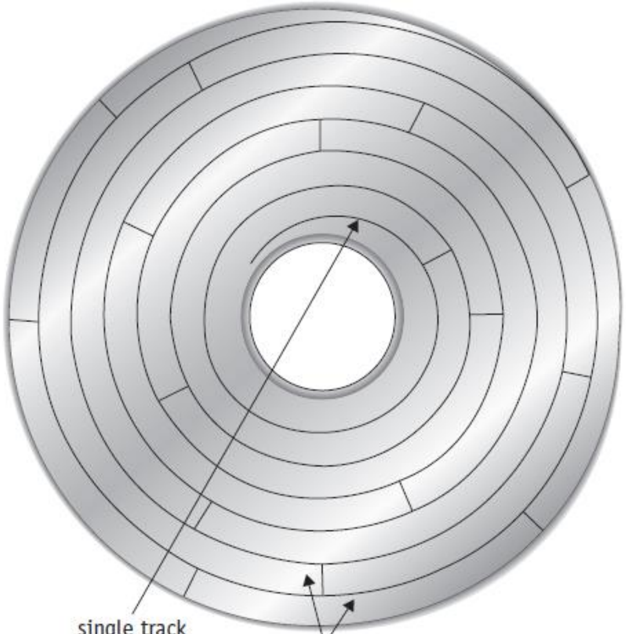
(figure 7.7)

On a magnetic disk, the sectors are of different sizes: bigger at the rim and smaller at the center. The disk spins at a constant angular velocity (CAV) to compensate for this difference. Some optical discs can read and write on multiple layers, greatly enhancing storage capacity.



- Design difference
  - Magnetic disk
    - Concentric tracks of sectors
    - Spins at constant angular velocity (CAV)
    - Wastes storage space but fast data retrieval

Optical Disc Storage (cont'd.)

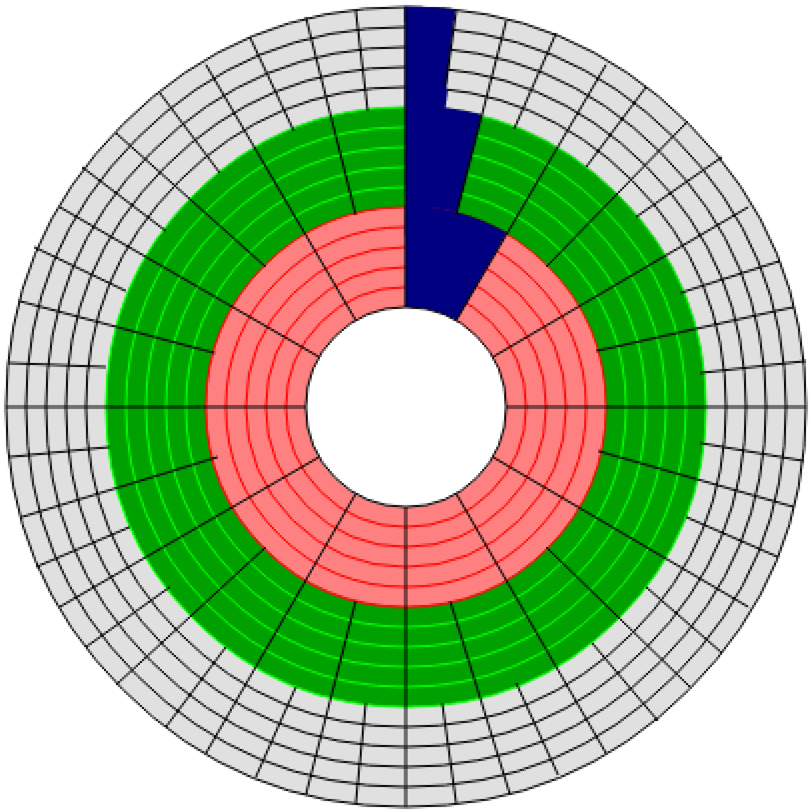


(figure 7.8)

On an optical disc, the sectors (not all sectors are shown here) are of the same size throughout the disc. The disc drive changes speed to compensate, but it spins at a constant linear velocity (CLV).

- Design features
  - Optical disc
    - Single spiralling track of same-sized sectors running from center to disc rim
    - Spins at constant linear velocity (CLV)
    - More sectors and more disc data

New type of Magnetic discs

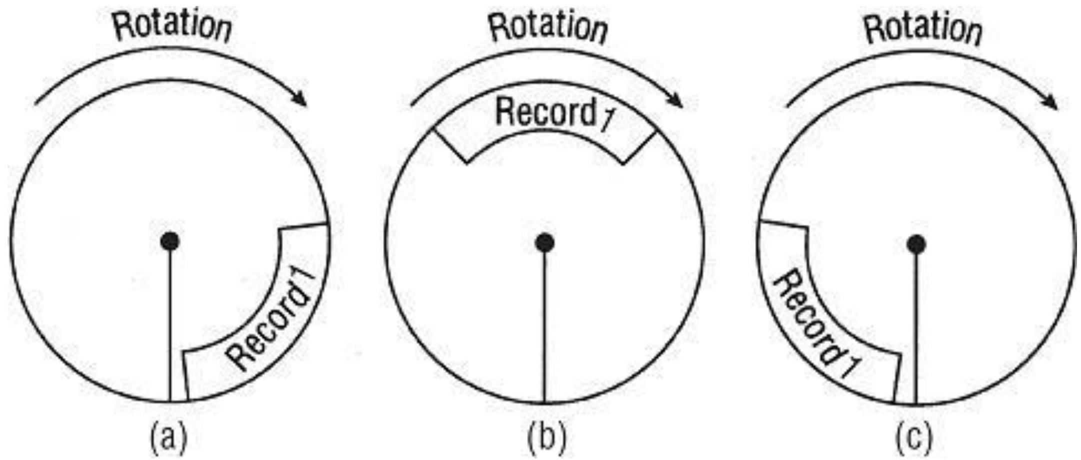


- Zone bit recording
  - All modern drives since 1990

Access Times: Fixed-Head Devices

- Record access requires two items
  - Track number and record number
- Access time = search time + transfer time
- Total access time
  - Rotational speed dependent
- DASDs rotate continuously
  - Three basic positions for requested record
    - In relation to read/write head position
- DASD has little access variance
  - Good candidates: low activity files, random access
- Blocking used to minimize access time

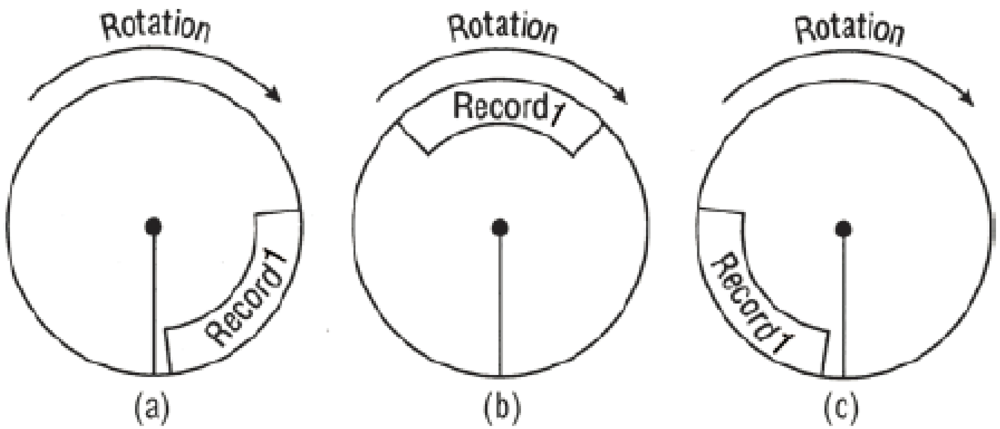
(figure 7.8)
As a disk rotates, Record 1 may be near the read/write head and ready to be scanned, as seen in (a); in the farthest position just past the head, (c); or somewhere in between, as in the average case, (b).
© Cengage Learning 2014



Benchmarks	Access Time
Maximum access time	16.8 ms + 0.00094 ms/byte
Average access time	8.4 ms + 0.00094 ms/byte
Sequential access time	Depends on the length of the record (known as the transfer rate)

(table 7.2)
Access times for a fixed-head disk drive at 16.8 ms/revolution.
© Cengage Learning 2014

Fixed-Head Devices (cont'd.)



Search time = 0 Search time = t/2 Search time = t

Benchmarks	Access Time
Maximum access	16.8 ms + 0.00094 ms/byte
Average access	8.4 ms + 0.00094 ms/byte
Sequential access	Depends on the length of the record; generally less than 1 ms (known as the transfer rate)

Fixed Head devices access time using blocking

- Assume rotational delay = 8.4 ms and transfer time for 100 bytes is 0.094 ms
- If no blocking – 10 records stored seperately
  - Access time = 8.4 + 0.094 = 8.494 ms
  - For 10 records = 10* (8.4+0.094) = 84.940
- If the 10 records are blocked
  - Access time = 8.4 + (0.094 * 10) = 9.34 ms

Movable-Head Devices

- Record access requires three items
  - Seek time + search time + transfer time
- Search time and transfer time calculation
  - Same as fixed-head DASD
- Blocking is a good way to minimize access time

Reading 10 records that stored separately

Access time = **25** + 8.4 + 0.094 = 33.494 ms for one record

Total access time = 10 (33.494) = **334.94** ms for ten records

Reading 10 records that stored in a cylinder

Total access time = 25 + 8.4 + 10 x 0.094 = **34.34** ms for ten records

Optical Disc Storage (cont'd.)

- Two important performance measures
  - **Sustained data-transfer rate**
    - Speed to read massive data amounts from disc
    - Measured in megabits per second (Mbps) or Megabytes per second (MBps)
  - **Average access time**
    - Average time to move head to specific disc location
    - Expressed in milliseconds (ms)
- Third feature
  - **Cache size** (hardware)
    - Buffer to transfer data blocks from disc

CD and DVD Technology

- **CD**
  - Data recorded as zeros and ones
    - **Pits**: indentations
    - **Lands**: flat areas
  - Reads with low-power laser
    - Light strikes land and reflects to photodetector
    - Pit is scattered and absorbed
    - Photodetector converts light intensity into digital signal

CD and DVD Technology (cont'd.)

- **CD-Recordable technology** (CD-R)
  - Records data using write-once technique
  - Data cannot be erased or modified
  - Disk
    - Contains several layers
    - Gold reflective layer and dye layer
    - Records with high-power laser
    - Permanent marks on dye layer
    - CD cannot be erased after data recorded
  - Data read on standard CD drive (low-power beam)

CD and DVD Technology (cont'd.)

- **CD-Rewritable technology** (CD-RW)
  - Data written, changed, erased
  - Uses phase change technology
    - Amorphous and crystalline phase states
  - Record data: beam heats up disc
    - State changes from crystalline to amorphous
  - Erase data: low-energy beam to heat up pits
    - Loosens alloy to return to original crystalline state
  - Drives read standard CD-ROM, CD-R, CD-RW discs
  - Drives store large quantities of data, sound, graphics, multimedia

CD and DVD Technology (cont'd.)

- **DVD technology** (Digital Versatile Disc)
- CD-ROM comparison
  - Similar in design, shape, size
  - Differs in data capacity
    - Dual-layer, single-sided DVD holds 13 CDs
    - Single-layer, single-sided DVD holds 8.6 GB (MPEG video compression)
  - Differs in laser wavelength
    - Uses red laser (smaller pits, tighter spiral)
- DVDs cannot be read by CD or CD-ROM drives
- DVD-R and DVD-RW provide rewritable flexibility

Blu-Ray Disc Technology

- Same physical size as DVD/CD
- Smaller pits
- More tightly wound tracks
- Use of blue-violet laser allows multiple layers
- 50GB-500GB
- 432 Mbps
- Formats: BD-ROM, BD-R, BD-RE

Flash Memory Storage

- Electronically erasable programmable read-only memory (EEP)
  - Nonvolatile and removable
  - Emulates random access
    - Difference: data stored securely (even if removed)
- Data stored on microchip card or “key”
  - Compact flash, smart cards, memory sticks
  - Often connected through USB port
- Write data: electric charge sent through floating gate
- Erase data: strong electrical field (flash) applied

Magnetic Disk Drive Access Times

- File access time factors
  - **Seek time** (slowest)
    - Time to position read/write head on track
    - Does not apply to fixed read/write head devices
  - **Search time**
    - Rotational delay
    - Time to rotate DASD
    - Rotate until desired record under read/write head
  - **Transfer time** (fastest)
    - Time to transfer data
    - Secondary storage to main memory transfer

Device Handler Seek Strategies

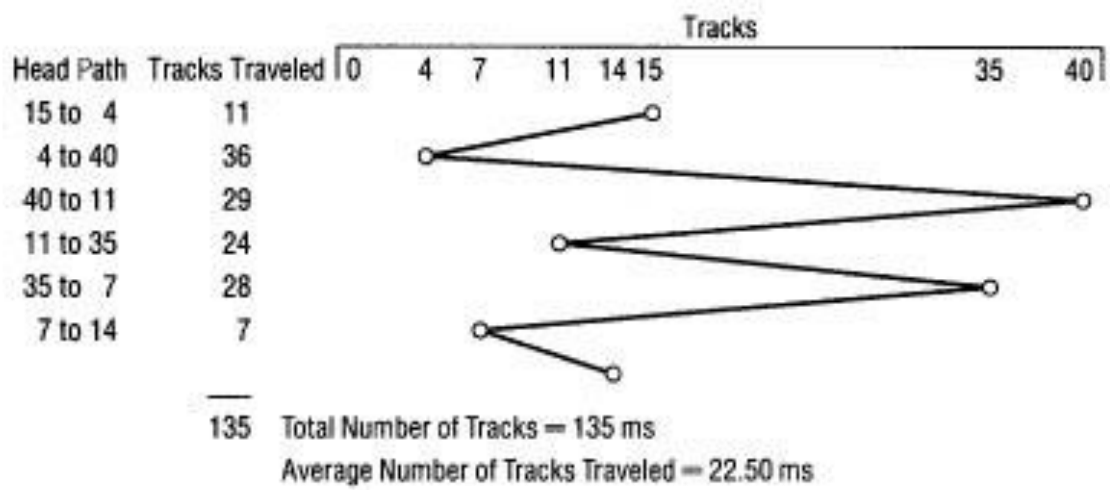
- Predetermined device handler
  - Determines device processing order
  - Goal: minimize seek time
- Types
  - First-come, first-served (FCFS), shortest seek time first (SSTF), SCAN (including LOOK, N-Step SCAN, C-SCAN, and C-LOOK)
- Scheduling algorithm goals
  - Minimize arm movement
  - Minimize mean response time
  - Minimize variance in response time

Device Handler Seek Strategies (cont'd.)

- **FCFS**
  - On average: does not meet three seek strategy goals
  - Disadvantage: extreme arm movement

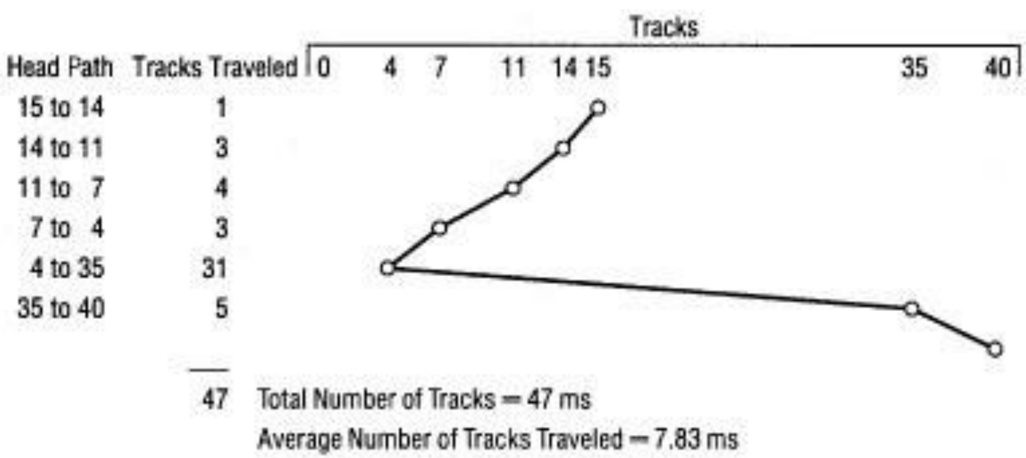
(figure 7.15)

The arm makes many time-consuming movements as it travels from track to track to satisfy all requests in FCFS order.



Device Handler Seek Strategies (cont'd.)

- **Shortest Seek Time First (SSTF)**
  - Request with track closest to one being served
  - Minimizes overall seek time
  - Postpones traveling to out of way tracks



(figure 7.16)

Using the SSTF algorithm, with all track requests on the wait queue, arm movement is reduced by almost one third while satisfying the same requests shown in Figure 7.15 (using the FCFS algorithm).

Device Handler Seek Strategies (cont'd.)

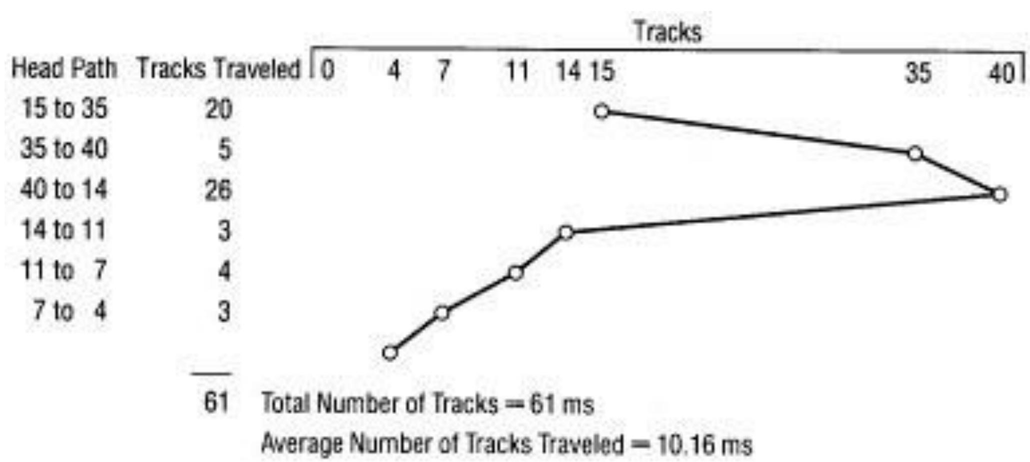
- **SCAN**
  - Directional bit
    - Indicates if arm moving toward/away from disk center
  - Algorithm moves arm methodically
    - From outer to inner track, services every request in its path
    - If reaches innermost track, reverses direction and moves toward outer tracks
    - Services every request in its path

Device Handler Seek Strategies (cont'd.)

- **LOOK**
  - Most common variation of scan
  - Arm does not go to either edge
    - Unless requests exist
  - Eliminates indefinite postponement

(figure 7.17)

The LOOK algorithm makes the arm move systematically from the first requested track at one edge of the disk to the last requested track at the other edge. In this example, all track requests are on the wait queue.



Device Handler Seek Strategies (cont'd.)

- **N-Step SCAN**
  - Holds all requests until arm starts on way back
    - New requests grouped together for next sweep
- **C-SCAN (Circular SCAN)**
  - Arm picks up requests on path during inward sweep
  - Provides more uniform wait time
- **C-LOOK**
  - Inward sweep stops at last high-numbered track request
  - No last track access unless required

Device Handler Seek Strategies (cont'd.)

- Best strategy
  - FCFS best with light loads
    - Service time unacceptably long under high loads
  - SSTF best with moderate loads
    - Localization problem under heavy loads
  - SCAN best with light to moderate loads
    - Eliminates indefinite postponement
      - Throughput and mean service times SSTF similarities
  - C-SCAN best with moderate to heavy loads
    - Very small service time variances

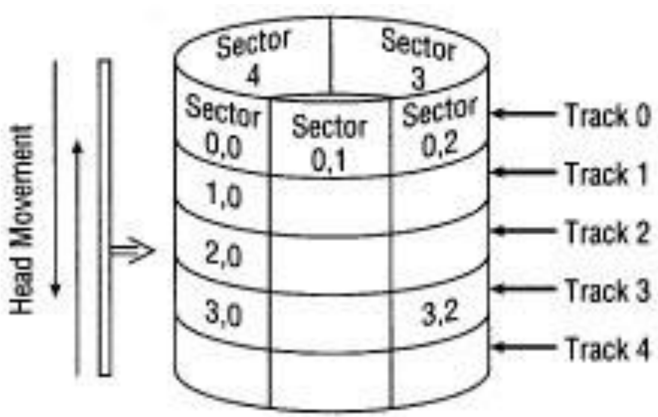
Search Strategies: Rotational Ordering

- **Rotational ordering**
  - Optimizes search times
    - Orders requests once read/write heads positioned
  - Read/write head movement time
    - Hardware dependent
- Reduces time wasted
  - Due to rotational delay
  - Request arrangement
    - First sector requested on second track is next number higher than one just served

Search Strategies: Rotational Ordering

(figure 7.18)

This movable-head cylinder takes 5 ms to move the read/write head from one track to the next. The read/write head is initially positioned at Track 0, Sector 0. It takes 5 ms to rotate the cylinder from Sector 0 to Sector 4 and 1 ms to transfer one sector from the cylinder to main memory.



-5 ms from one track to next
-5 ms to rotate from sector 0 to sector 4

Request List

Track	Sector
0	1
1	4
1	3
2	0
2	3
2	4
3	2
3	0

Search Strategies: Rotational Ordering (cont'd.)

(table 7.5)

It takes 36 ms to fill the eight requests on the movable-head cylinder shown in Figure 7.18.

Request (Track, Sector)		Seek Time		Search Time		Data Transfer		Total Time
1.	0,1	0		1		1		2
2.	1,4	5		2		1		8
3.	1,3	0		3		1		4
4.	2,0	5		1		1		7
5.	2,3	0		2		1		3
6.	2,4	0		0		1		1
7.	3,2	5		2		1		8
8.	3,0	0		2		1		3
	TOTALS	15 ms	+	13 ms	+	8 ms	=	36 ms

Search Strategies: Rotational Ordering (cont'd.)

Request (Track, Sector)		Seek Time	Search Time	Data Transfer	Total Time
1.	0,1	0	1	1	2
2.	1,3	5	1	1	7
3.	1,4	0	0	1	1
4.	2,0	5	0	1	6
5.	2,3	0	2	1	3
6.	2,4	0	0	1	1
7.	3,0	5	0	1	6
8.	3,2	0	1	1	2
	TOTALS	15 ms +	5 ms +	8 ms =	28 ms

(table 7.6)

It takes 28 ms to fill the same eight requests shown in Table 7.5 after the requests are ordered to minimize search time, reducing it from 13 ms to 5 ms.

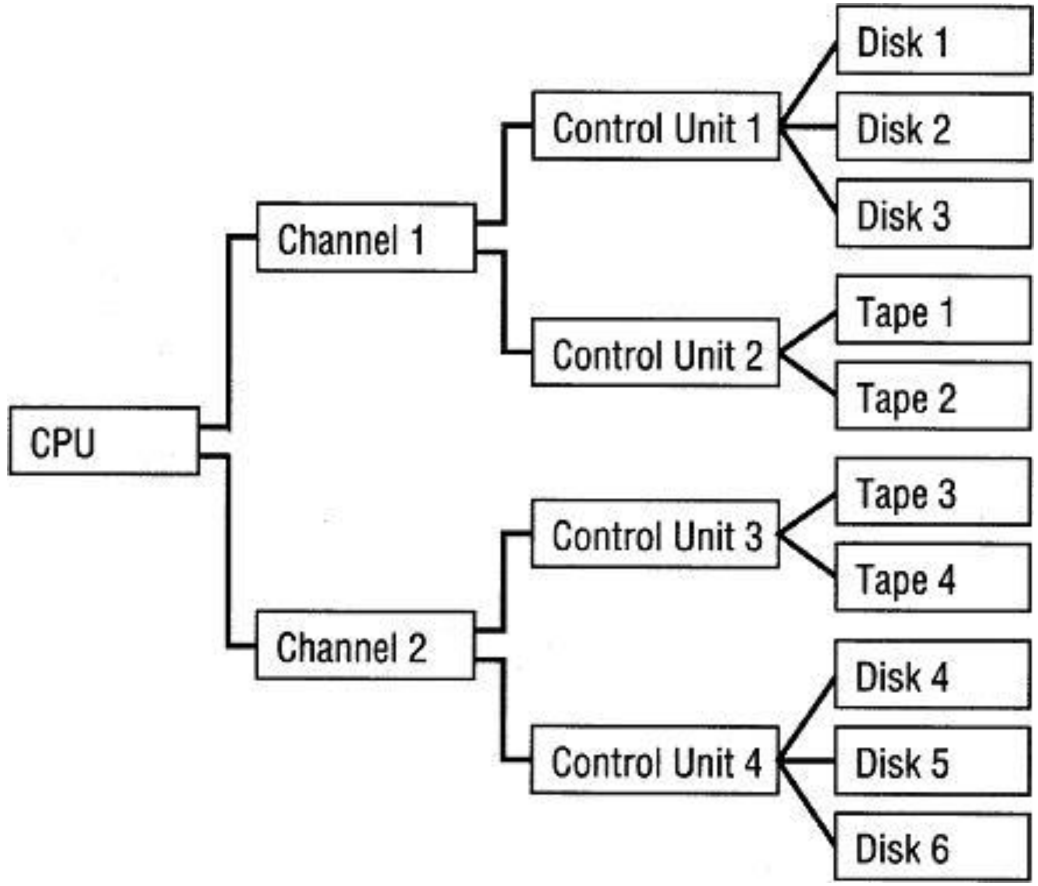
Components of the I/O Subsystem

- **I/O Channel**
  – Programmable units
    - Positioned between CPU and control unit
  – Synchronizes device speeds
    - CPU (fast) with I/O device (slow)
  – Manages concurrent processing
    - CPU and I/O device requests
  – Allows overlap
    - CPU and I/O operations

Components of the I/O Subsystem (cont'd.)

- **I/O channel programs**
  - Specifies action performed by devices
  - Controls data transmission
    - Between main memory and control units
- **I/O control unit**: receives and interprets signal
- **Disk controller** (disk drive interface)
  - Links disk drive and system bus
- Entire path must be available when I/O command initiated
- I/O subsystem configuration
  - Multiple paths increase flexibility and reliability

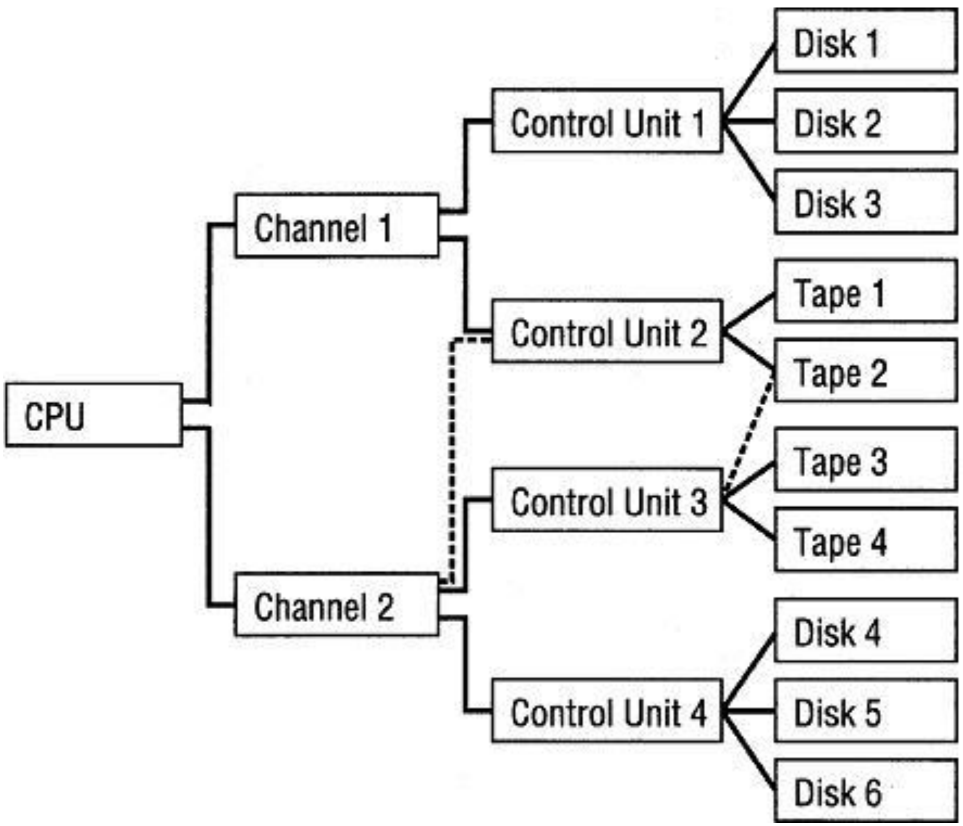
Components of the I/O Subsystem (cont'd.)



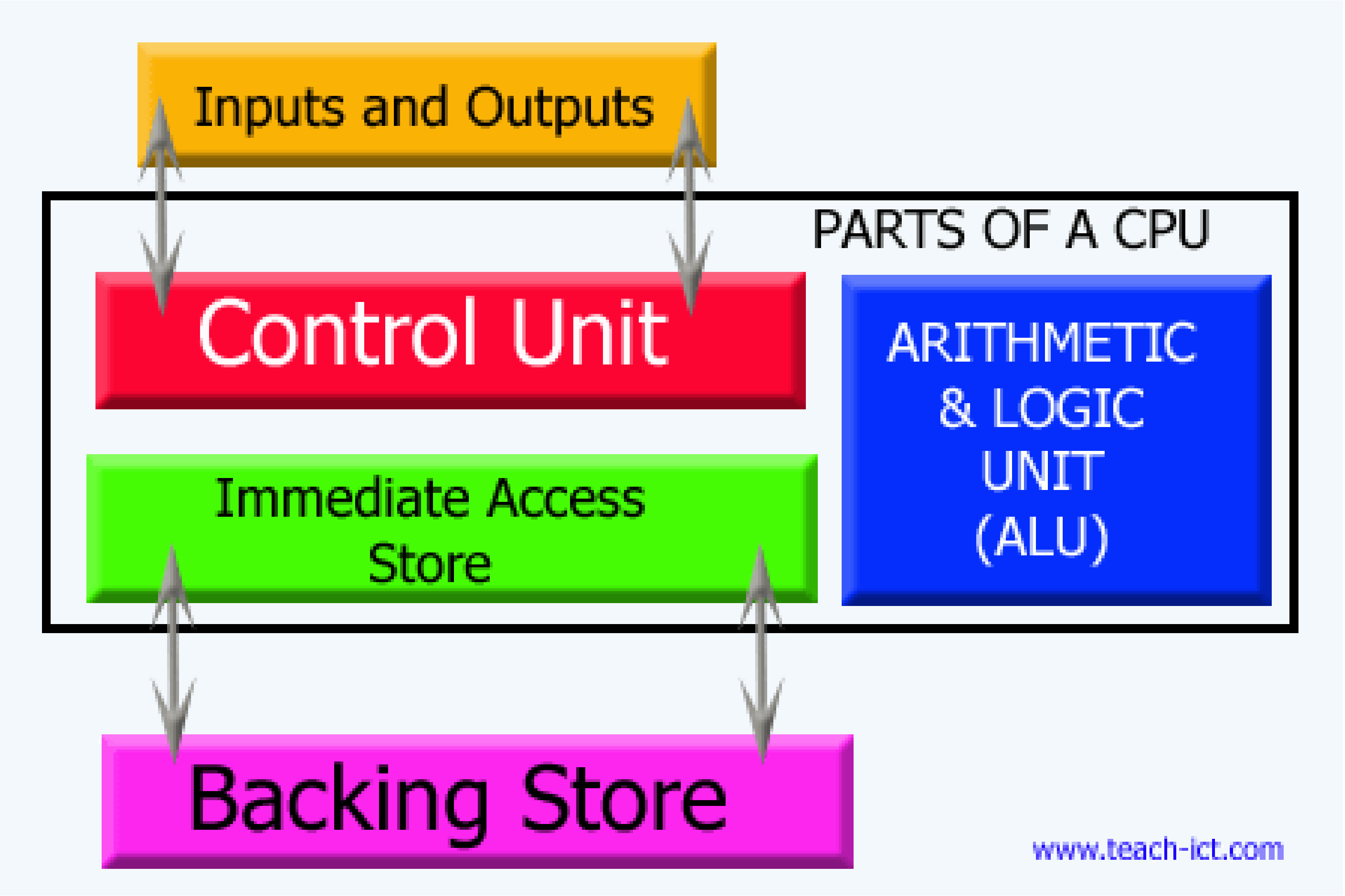
(figure 7.12)

Typical I/O subsystem configuration.

Components of the I/O Subsystem (cont'd.)



(figure 7.18)
I/O subsystem configuration with multiple paths, which increase both flexibility and reliability. With two additional paths, shown with dashed lines, if Control Unit 2 malfunctions, then Tape 2 can still be accessed via Control Unit 3.


Inputs and Outputs
PARTS OF A CPU
Control Unit
ARITHMETIC
& LOGIC
UNIT
(ALU)
Immediate Access
Store
Backing Store
www.teach-ict.com

Communication Among Devices

- Problems to resolve
  - Know which components are busy/free
    - Solved by structuring interaction between units
  - Accommodate requests during heavy I/O traffic
    - Handled by buffering records and queuing requests
  - Accommodate speed disparity between CPU and I/O devices
    - Handled by buffering records and queuing requests

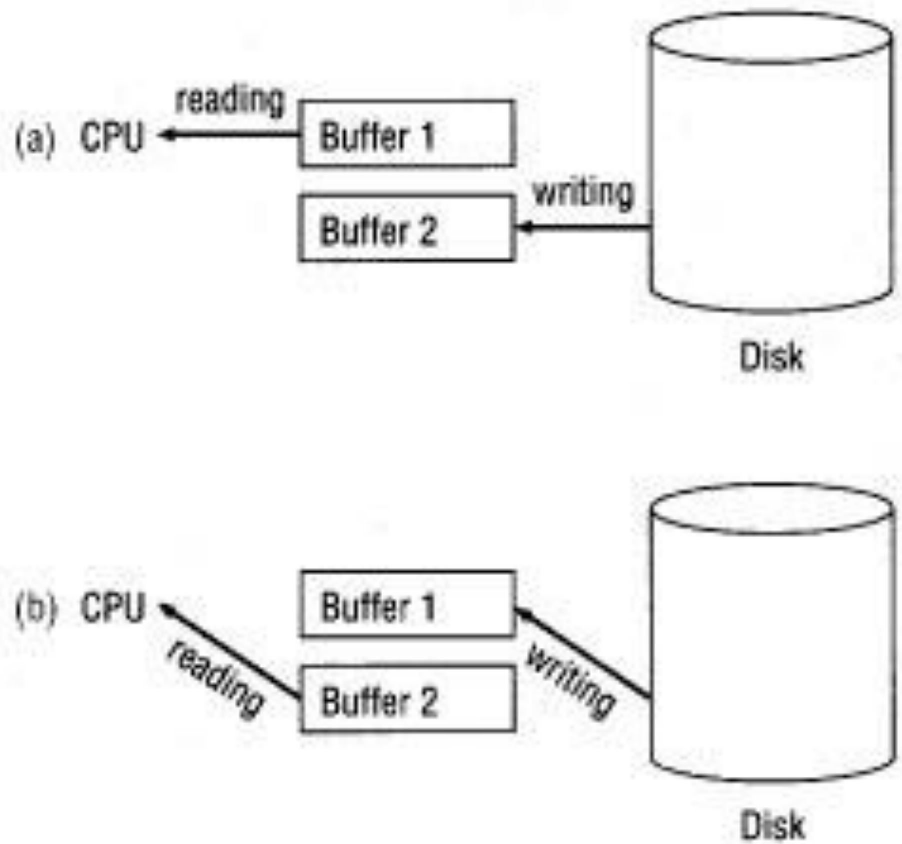
Communication Among Devices (cont'd.)

- I/O subsystem units finish independently of others
- CPU processes data while I/O performed
- Success requires device completion knowledge
  - Hardware flag tested by CPU
    - **Channel status word (CSW)** contains flag
    - Three bits in flag represent I/O system component (channel, control unit, device)
    - Changes zero to one (free to busy)
  - Flag tested using polling and interrupts
    - **Interrupts** are more efficient way to test flag

Communication Among Devices (cont'd.)

- Direct memory access (DMA)
  - Allows control unit to directly access main memory
  - Transfers data without the intervention of CPU
  - Used for high-speed devices (disk)
- Buffers
  - Temporary storage areas in main memory, channels, control units
  - Improves data movement synchronization
    - Between relatively slow I/O devices and very fast CPU
  - Double buffering: processing of record by CPU while another is read or written by channel

Double buffering



(figure 7.14)

Example of double buffering: (a) the CPU is reading from Buffer 1 as Buffer 2 is being filled; (b) once Buffer 2 is filled, it can be read quickly by the CPU while Buffer 1 is being filled again.

Management of I/O Requests

- **I/O traffic controller**
  - Watches status of devices, control units, channels
  - Three main tasks
    - Determine if path available
    - If more than one path available, determine which one to select
    - If paths all busy, determine when one is available
  - Maintain database containing unit status and connections

Management of I/O Requests (cont'd.)

(table 7.4)

Each control block contains the information it needs to manage its part of the I/O subsystem.

Channel Control Block	Control Unit Control Block	Device Control Block
• Channel identification • Status • List of control units connected to it • List of processes waiting for it	• Control unit identification • Status • List of channels connected to it • List of devices connected to it • List of processes waiting for it	• Device identification • Status • List of control units connected to it • List of processes waiting for it

Management of I/O Requests (cont'd.)

- **I/O scheduler**
  - Same job as process scheduler (Chapter 4)
  - Allocates devices, control units, channels
  - If requests greater than available paths
    - Decides which request to satisfy first: based on different criteria
  - In many systems
    - I/O requests not preempted (major difference with process scheduler)
  - For some systems
    - Allow preemption with I/O request subdivided
    - Allow preferential treatment for high-priority requests

Management of I/O Requests (cont'd.)

- **I/O device handler**
  - Performs actual data transfer
    - Processes device interrupts
    - Handles error conditions
    - Provides detailed scheduling algorithms
  - Device dependent
  - Each I/O device type has device handler algorithm

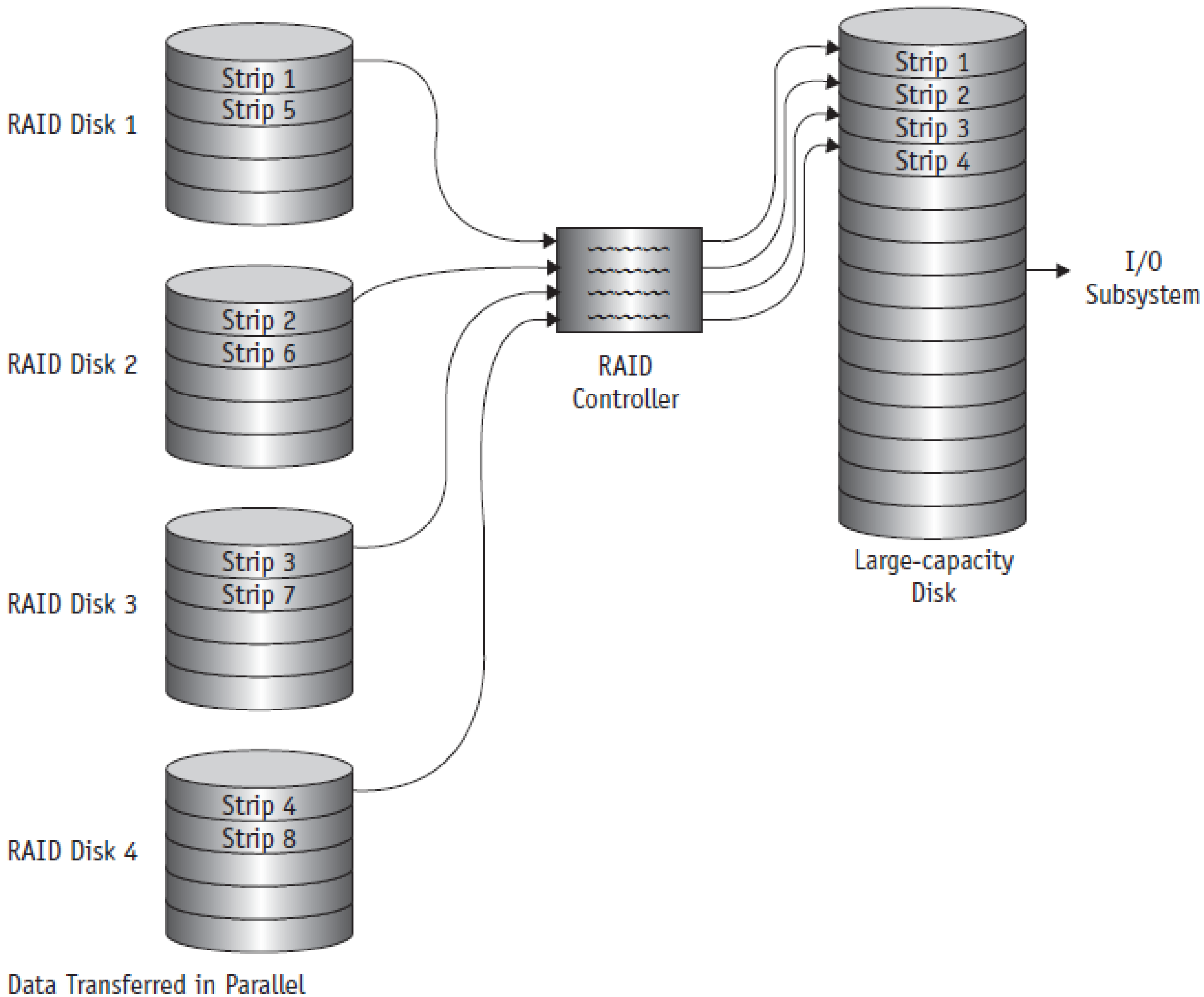
RAID

- Physical disk drive set viewed as single logical unit
  - Preferable over few large-capacity disk drives
- Improved I/O performance
- Improved data recovery
  - Disk failure event
- Introduces redundancy
  - Helps with hardware failure recovery
- Significant factors in RAID level selection
  - Cost, speed, system's applications
- Increases hardware costs

RAID (cont'd.)

(figure 7.19)

Data being transferred in parallel from a Level 0 RAID configuration to a large-capacity disk. The software in the controller ensures that the strips are stored in correct order.



RAID (cont'd.)

RAID Level	Error Correction Method	I/O Request Rate	Data Transfer Rate
0	None	Excellent	Excellent
1	Mirroring	Read: Good Write: Fair	Read: Fair Write: Fair
2	Hamming code	Poor	Excellent
3	Word parity	Poor	Excellent
4	Strip parity	Read: Excellent Write: Fair	Read: Fair Write: Poor
5	Distributed strip parity	Read: Excellent Write: Fair	Read: Fair Write: Poor
6	Distributed strip parity and independent data check	Read: Excellent Write: Poor	Read: Fair Write: Poor

(table 7.7)

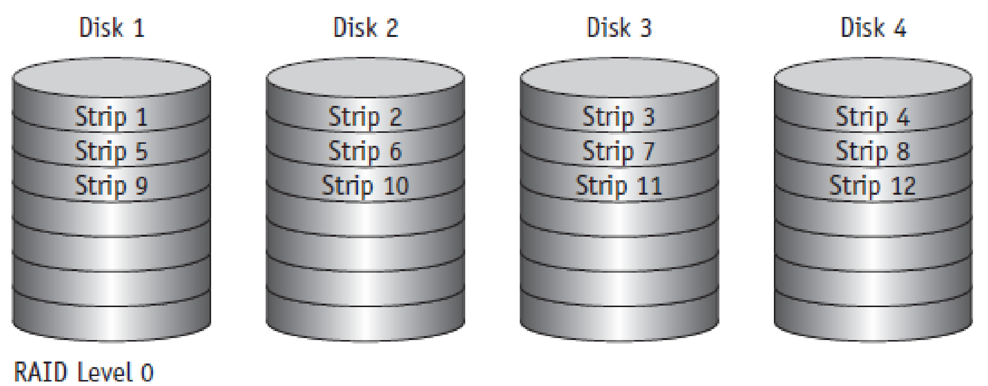
The seven standard levels of RAID provide various degrees of error correction. Cost, speed, and the system's applications are significant factors to consider when choosing a system.

Level Zero

- Uses data striping (not considered true RAID) but
  - No parity and error corrections
  - No error correction/redundancy/recovery
- Benefits
  - Devices appear as one logical unit
  - Improves system performance
  - Best for large data quantity non-critical data

(figure 7.20)

RAID Level 0 with four disks in the array. Strips 1, 2, 3, and 4 make up a stripe. Strips 5, 6, 7, and 8 make up another stripe, and so on.

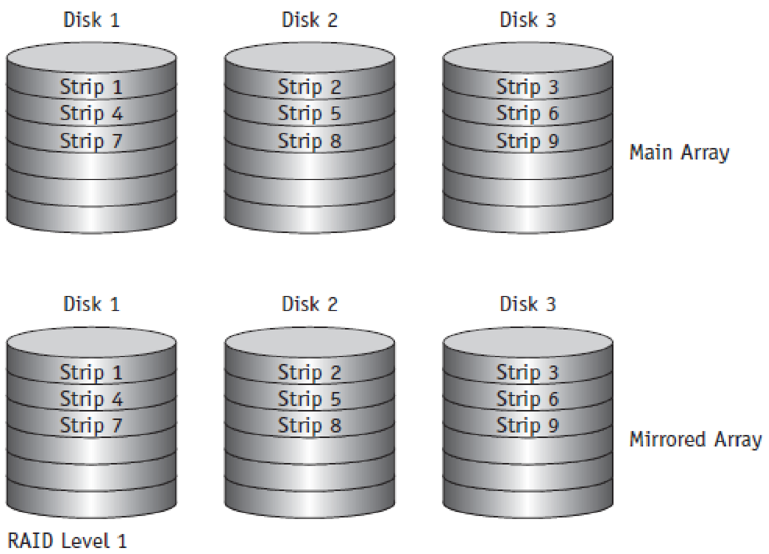


Level One

- Uses data striping (considered true RAID)
  - Mirrored configuration (backup)
    - Duplicate set of all data (expensive)
  - Provides redundancy and improved reliability

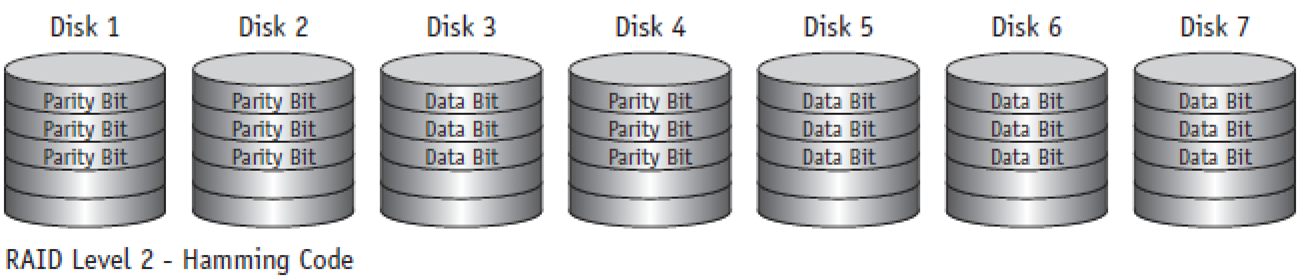
(figure 7.21)

RAID Level 1 with three disks in the main array and three corresponding disks in the backup array, the mirrored array.



Level Two

- Uses small stripes (considered true RAID)
- **Hamming code**: error detection and correction
- Expensive and complex
  - Size of strip determines number of array disks



(figure 7.22)

RAID Level 2. Seven disks are needed in the array to store a 4-bit data item, one for each bit and three for the parity bits. Each disk stores either a bit or a parity bit based on the Hamming code used for redundancy.

Hamming Code

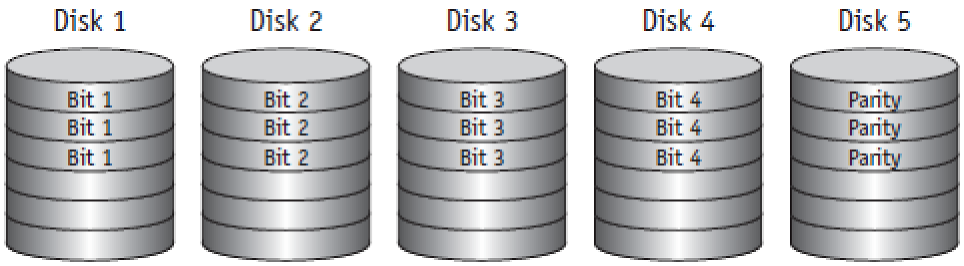
Bit position		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Encoded data bits		p1	p2	d1	p4	d2	d3	d4	p8	d5	d6	d7	d8	d9	d10	d11	p16	d12	d13	d14	d15	
Parity bit coverage	p1	X		X		X		X		X		X		X		X		X		X		
	p2		X	X			X	X			X	X			X	X			X	X		...
	p4				X	X	X	X					X	X	X	X					X	
	p8								X	X	X	X	X	X	X	X						
	p16																X	X	X	X	X	

Level Three

- Modification of Level 2
  - Requires one disk for redundancy
    - One parity bit for each strip

(figure 7.23)

RAID Level 3. A 4-bit data item is stored in the first four disks of the array. The fifth disk is used to store the parity for the stored data item.



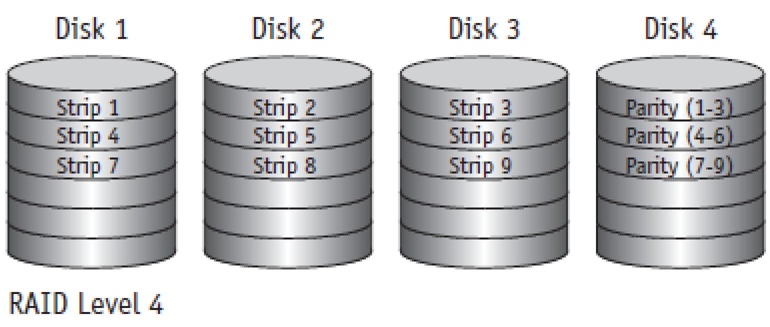
RAID Level 3

Level Four

- Same strip scheme as Levels 0 and 1
  - Computes parity for each strip
  - Stores parities in corresponding strip
    - Has designated parity disk

(figure 7.24)

RAID Level 4. The array contains four disks: the first three are used to store data strips, and the fourth is used to store the parity of those strips.

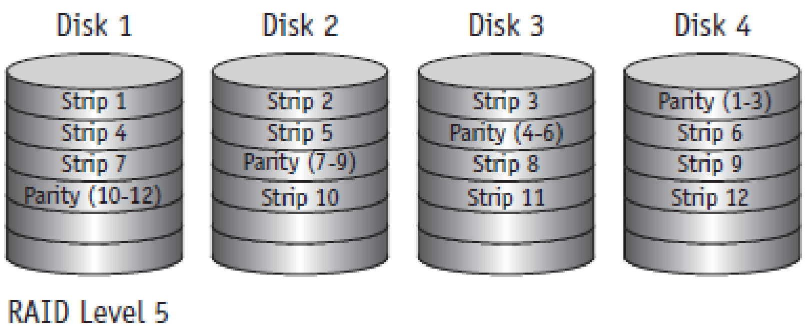


Level Five

- Modification of Level 4
- Distributes parity strips across disks
  - Avoids Level 4 bottleneck
- Disadvantage
  - Complicated to regenerate data from failed device

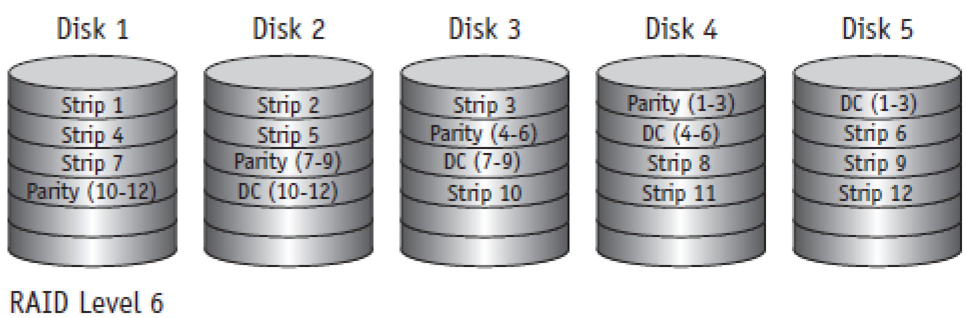
(figure 7.25)

RAID Level 5 with four disks. Notice how the parity strips are distributed among the disks.



Level Six

- Provides extra degree of error protection/correction
  - Two different parity calculations (double parity)
    - Same as level four/five and independent algorithm
  - Parities stored on separate disk across array
    - Stored in corresponding data strip
- Advantage: data restoration even if two disks fail



(figure 7.26)

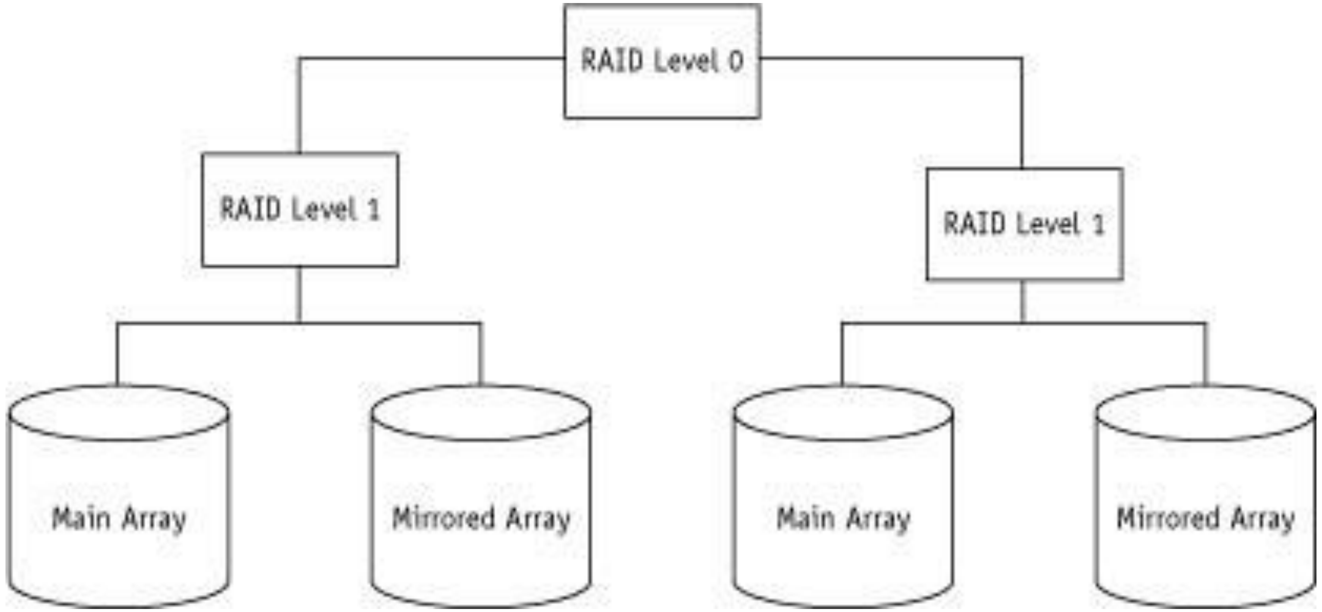
RAID Level 6. Notice how parity strips and data check (DC) strips are distributed across the disks.

Nested RAID Levels

- Combines multiple RAID levels (complex)

(figure 7.27)

A RAID Level 10 system.



Nested RAID Levels (cont'd.)

RAID Level	Combinations
01 (or 0+1)	A Level 1 system consisting of multiple Level 0 systems
10 (or 1+0)	A Level 0 system consisting of multiple Level 1 systems
03 (or 0+3)	A Level 3 system consisting of multiple Level 0 systems
30 (or 3+0)	A Level 0 system consisting of multiple Level 3 systems
50 (or 5+0)	A Level 0 system consisting of multiple Level 5 systems
60 (or 6+0)	A Level 0 system consisting of multiple Level 6 systems

(table 7.8)

Some common nested RAID configurations. Important: RAID 01 and 03 are not to be confused with RAID Levels 1 and 3, respectively.

Summary

- Device Manager
  - Manages every system device effectively as possible
- Devices
  - Vary in speed and sharability degrees
  - Direct access and sequential access
- Magnetic media: one or many read/write heads
  - Heads in a fixed position (optimum speed)
  - Move across surface (optimum storage space)
- Optical media: disk speed adjusted
  - Data recorded/retrieved correctly

Summary (cont'd.)

- Flash memory: device manager tracks USB devices
  - Assures data sent/received correctly
- I/O subsystem success dependence
  - Communication linking channels, control units, devices
- SCAN: eliminates indefinite postponement problem
  - Best for light to moderate loads
- C-SCAN: very small service time variance
  - Best for moderate to heavy loads
- RAID: redundancy helps hardware failure recover
  - Consider cost, speed, applications

Summary (cont'd.)

Strategy	Advantages	Disadvantages
FCFS	• Easy to implement • Sufficient for light loads	• Doesn't provide best average service • Doesn't maximize throughput
SSTF	• Throughput better than FCFS • Tends to minimize arm movement • Tends to minimize response time	• May cause starvation of some requests • Localizes under heavy loads
SCAN/LOOK	• Eliminates starvation • Throughput similar to SSTF • Works well with light to moderate loads	• Needs directional bit • More complex algorithm to implement • Increased overhead
N-Step SCAN	• Easier to implement than SCAN	• The most recent requests wait longer than with SCAN
C-SCAN/C-LOOK	• Works well with moderate to heavy loads • No directional bit • Small variance in service time • C-LOOK doesn't travel to unused tracks	• May not be fair to recent requests for high-numbered tracks • More complex algorithm than N-Step SCAN, causing more overhead

(table 7.9)

Comparison of DASD seek strategies discussed in this chapter.