Disk Management

Reading : Section 5.4 of Text Tananebaum or Ch 14 From Text Silberschatz

Disk is an I/O devices that is common to every computer.

Disk structure
Disk Scheduling
RAID
Disk Formatting & Error Handling
RAM Disks

Disk Structure

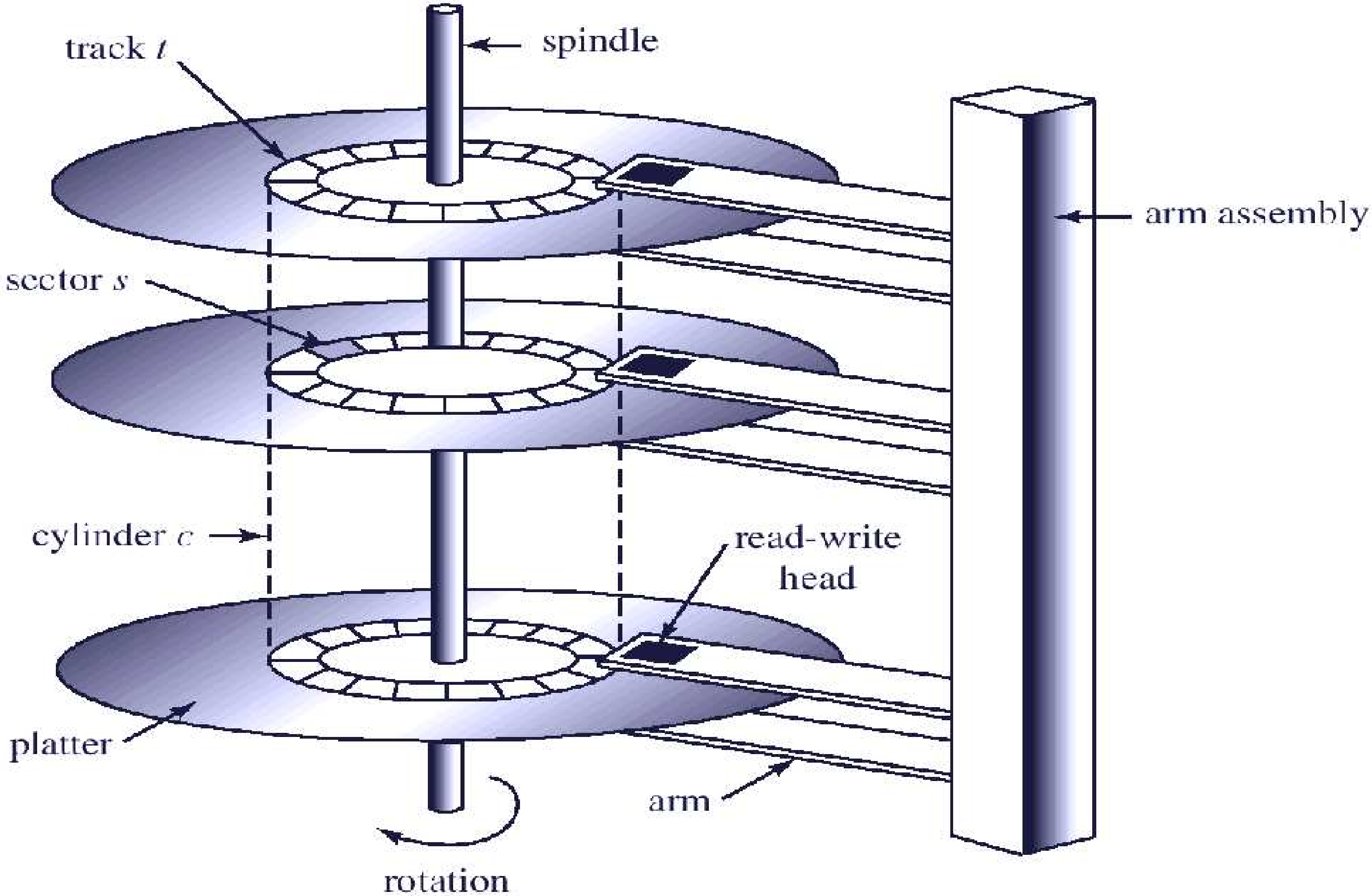
Disks comes in many sizes and speeds, and information may be stored optically or magnetically; however, all disks share a number of important features.

For Example: floppy disks, hard disks, CD-ROMs and DVDs.

Disk surface is divided into number of logical block called *sectors* and *tracks*.

The term *cylinder* refers to all the tracks at particular head position in hard disk.

Hard-Disk Structure



Disk Operations

Latency Time: The time taken to rotate from its current position to a position adjacent to the read-write head.

Seek: The processes of moving the arm assembly to new cylinder.

To access a particular record, first the arm assembly must be moved to the appropriate cylinder, and then rotate the disk until it is immediately under the read-write head.

The time taken to access the whole record is called *transmission time*.

Disk Scheduling

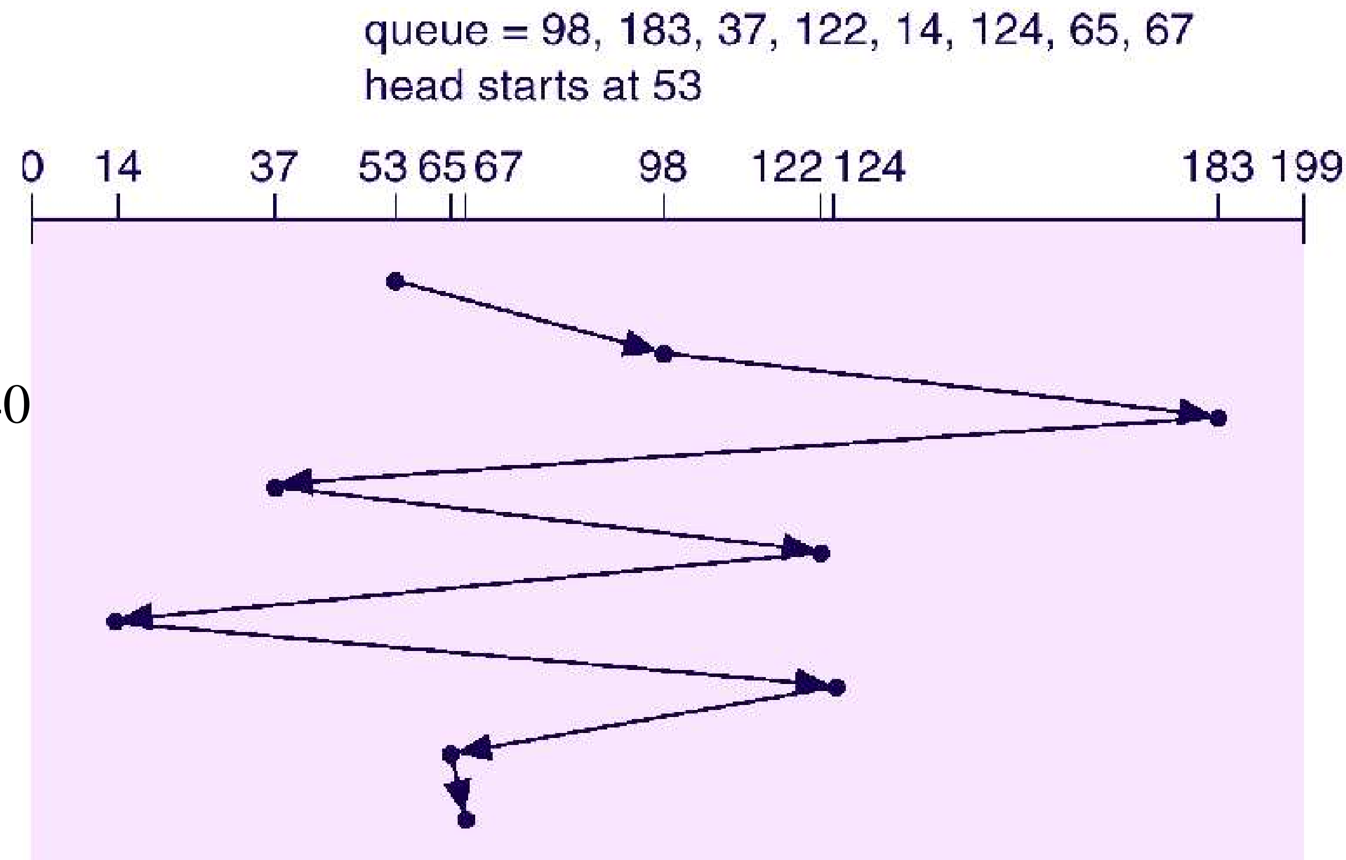
OS is responsible to use the hardware efficiently – for the disk drive this means fast seek, latency and transmission time.

For most disks, the *seek time* dominates the other two times, so reducing the *mean seek time* can improve system performance substantially.

First-Come First-Served (FCFS)

The first request to arrive is the first one serviced.

Example:



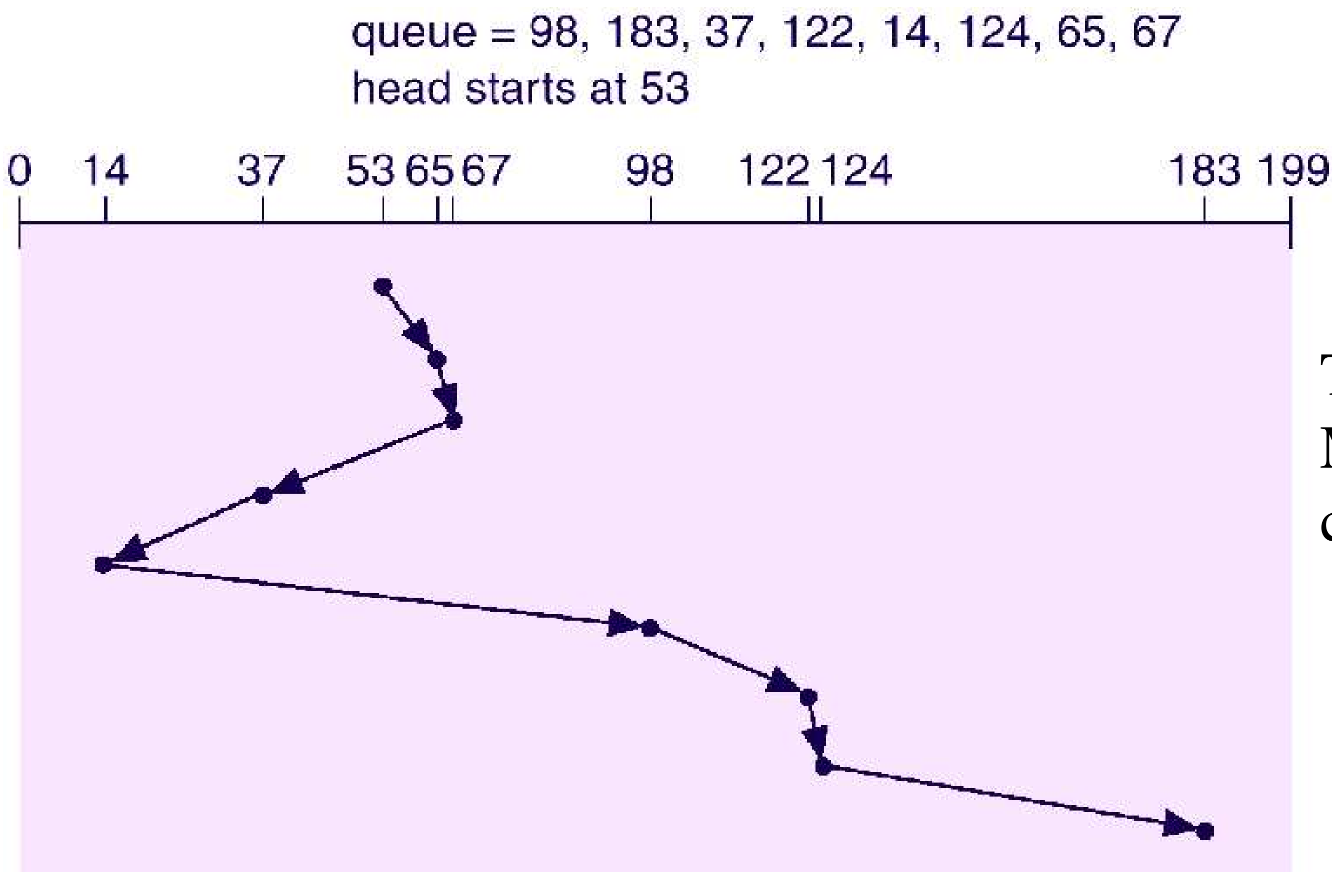
Total Head Movement = 640 cylinders

Advantages: Simple and Fair.

Problems: Does not provide fastest service.

Shortest-Seek-Time-First (SSTF)

Selects the request with the minimum seek time from the current head position.



Total Head Movement = 236 cylinders

Shortest-Seek-Time-First (SSTF)

Advantages:

Gives a substantial improvement in performance.

Problems:

SSTF scheduling is a form of SJF scheduling; may cause starvation of some requests.

Not optimal.

Used in batch system where throughput is the major consideration but unacceptable in interactive system.

SCAN

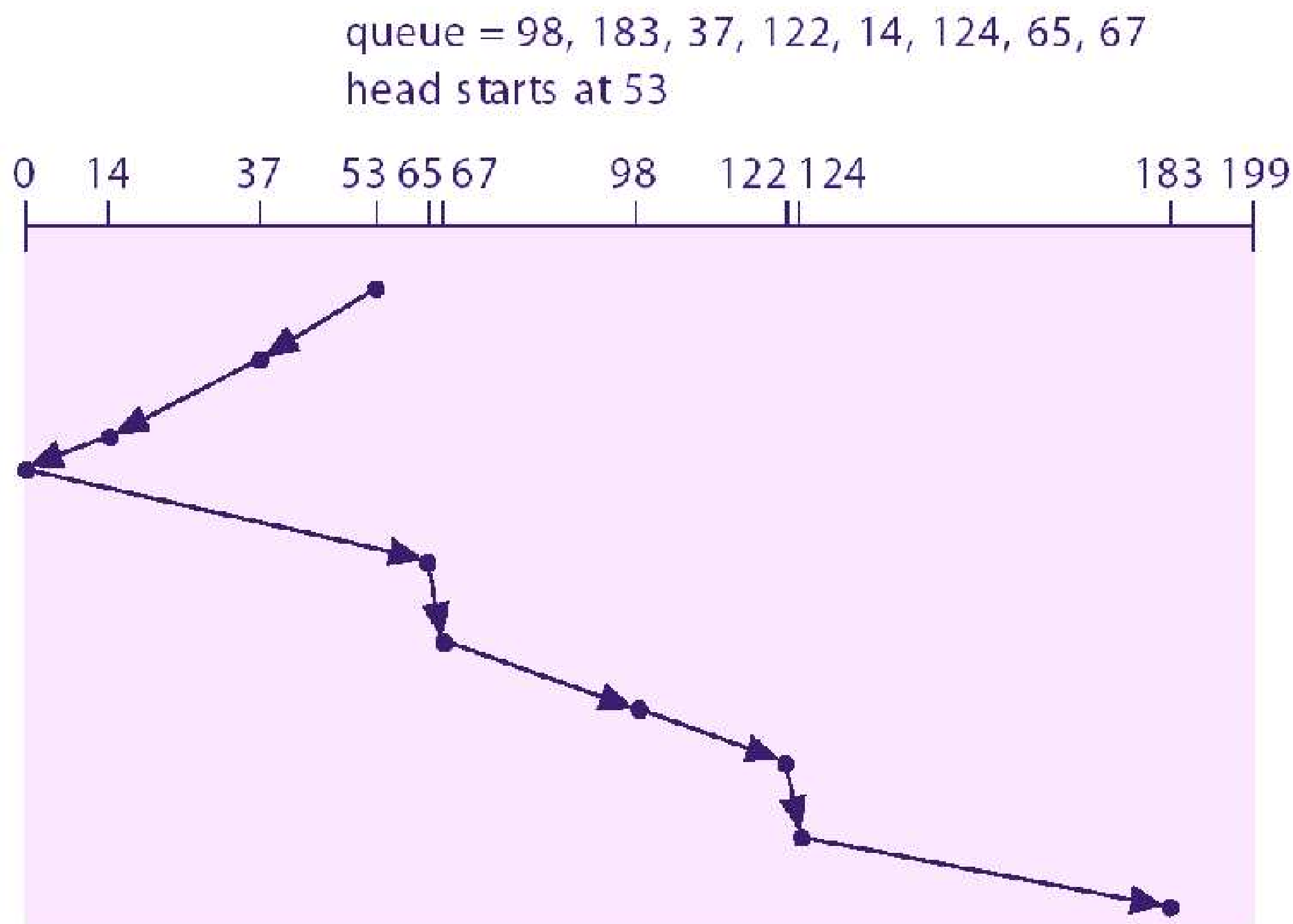
The disk arm starts at one end of the disk, and moves toward the other end, servicing requests until it gets to the other end of the disk, where the head movement is reversed and servicing continues.

Sometimes called the *elevator algorithm*.

Advantages: Decreases variances in seek and improve response time.

Problem: Starvation is possible if there are repeated request in current track.

SCAN



Total Head Movement = 236 cylinders

C-SCAN

When a uniform distribution of request for cylinders, only the few request are in extreme cylinders, since these cylinders have recently been serviced.

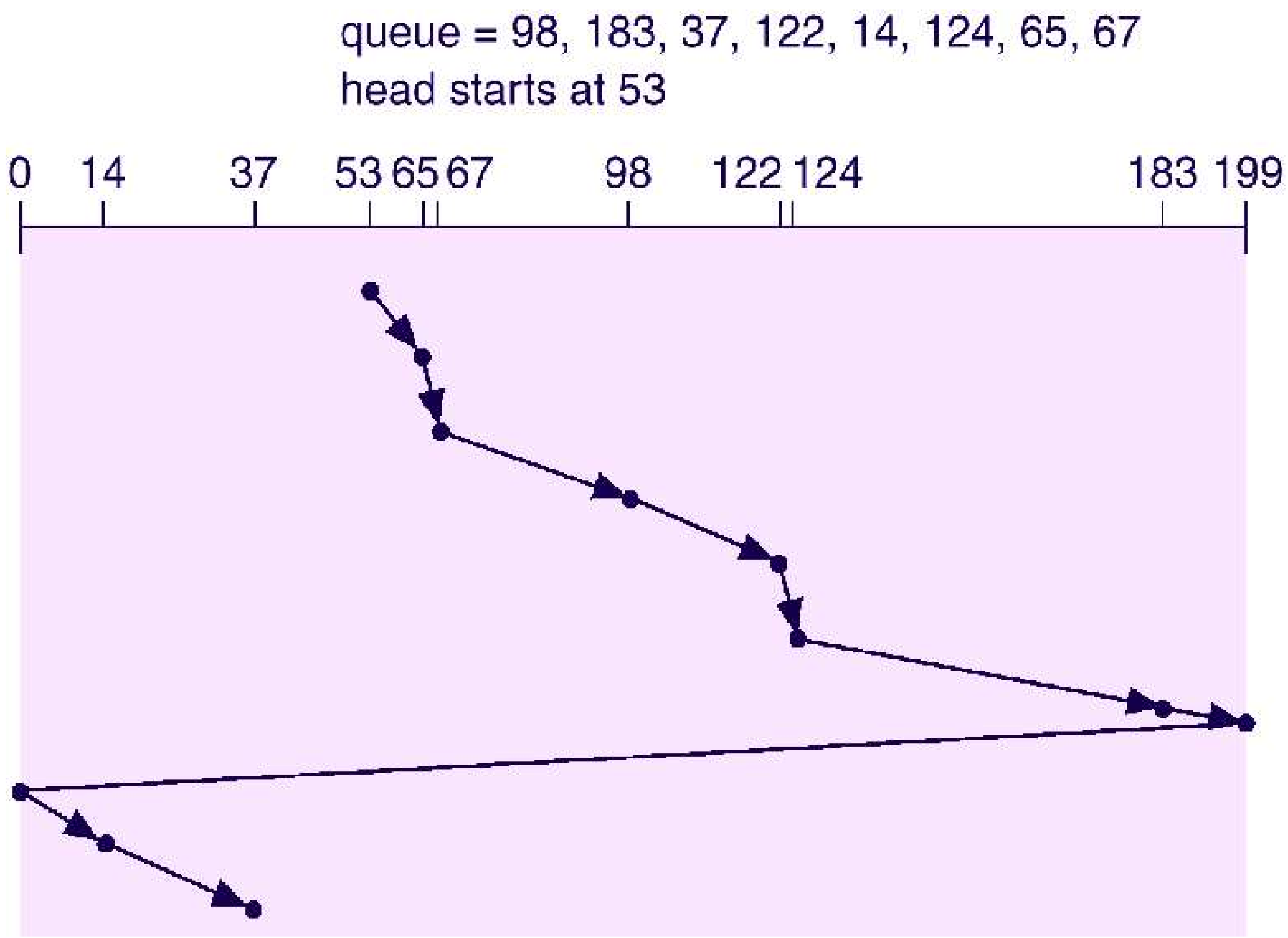
Why not go to the next extreme?

Circular SCAN is a variant of SCAN designed to provide a more uniform wait time.

The head moves from one end of the disk to the other. Servicing requests as it goes. When it reaches the other end, however, it immediately returns to the beginning of the disk, without servicing any requests on the return trip.

- Treats the cylinders as a circular list that wraps around from the last cylinder to the first one.

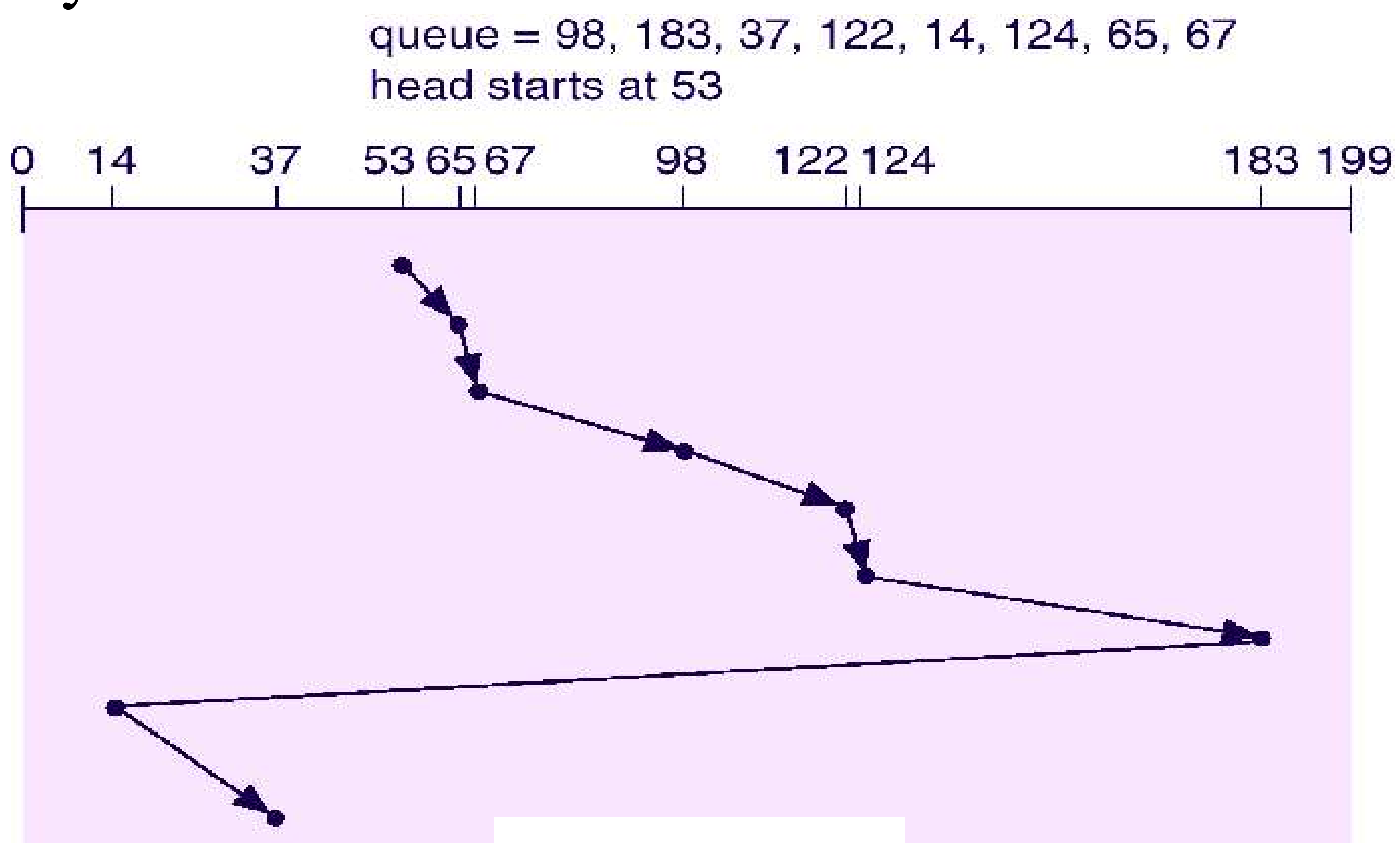
C-SCAN


queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53
0
14
37
53
65
67
98
122
124
183
199

C-LOOK

Version of C-SCAN

Arm only goes as far as the last request in each direction, then reverses direction immediately, without first going all the way to the end of the disk.



RAID

Redundant Array of Inexpensive (Independent) Disks.

Issues: Disk performance, Amount of storage required & Reliability

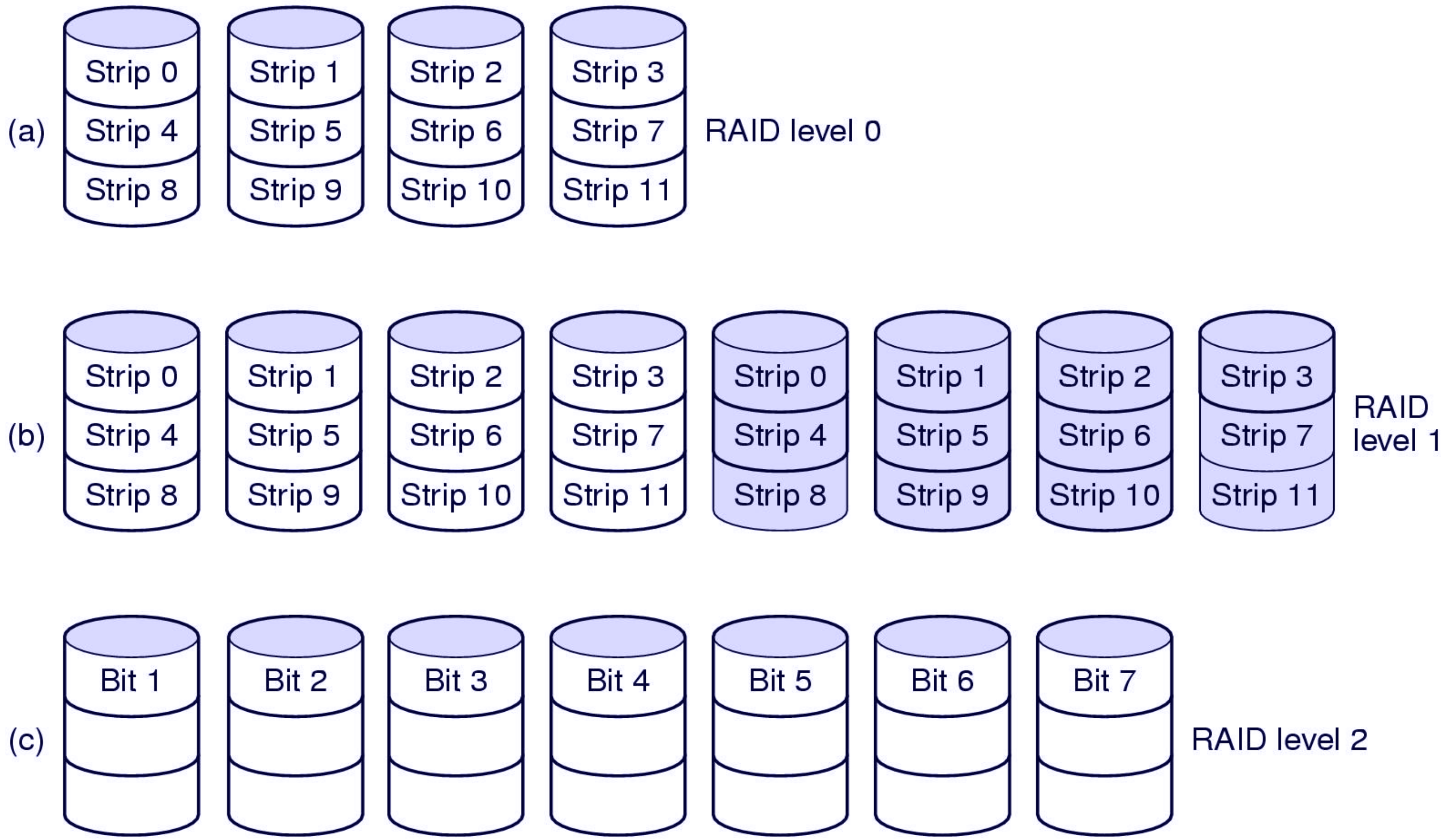
A technique of organizing multiple disks to address above issues is RAID.

RAID allows more than one disk to be used for a given operation, and allows continued operation and even automatic recovery in the face of disk failure.

Implemented in hardware or in OS.

RAID Levels

There are six types of organizations called RAID Levels.



RAID Levels

RAID Level 0:

RAID level 0 creates one large virtual disk from a number of smaller disks.

Storage is grouped into logical units called strips with the size of a strip being some multiple of sector size.

The virtual storage is sequence of strips interleaved among the disks in the array.

Advantages: Can create large disk; Performance benefit can be achieved.

Disadvantages: Reliability decrease.

RAID Levels

RAID Level 1:

Stores duplicate copy of each strip, with each copy on a different disk.

Advantages: Excellent reliability; if drive crashes, the copy is used. Read performance can be achived.

Disadvantages: Write Performance is no better than in single drive.

RAID Levels

RAID Level 2:

An error-correcting code is used for corresponding bits on each data disks.

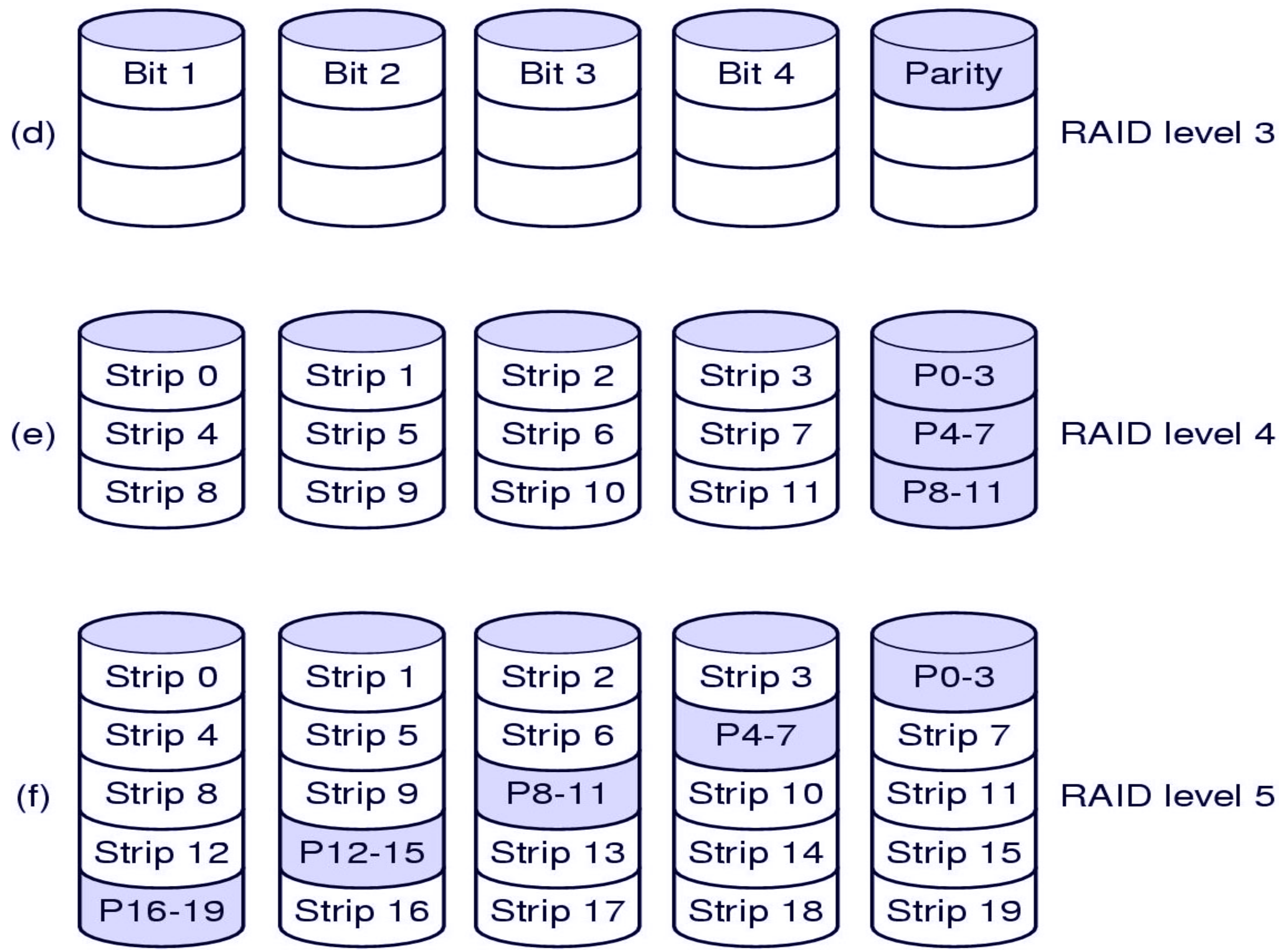
Error-correcting scheme store two or more extra bits, and can reconstruct the data if a single bit get damaged.

For Example, the first bit of each byte is stored in disk 1, second bit in disk 2, and until eight bit in disk 8, and error correcting bits are stored in further disks. If one of the disk fail, the remaining bits of the byte and associated error-correction bits can be read from other disks and be used to reconstruct the damage data.

Advantages: Total parallelism.

Disadvantages: Requires substantial number of drives.

RAID Levels


(d)
Bit 1
Bit 2
Bit 3
Bit 4
Parity
RAID level 3
(e)
Strip 0
Strip 1
Strip 2
Strip 3
P0-3
Strip 4
Strip 5
Strip 6
Strip 7
P4-7
Strip 8
Strip 9
Strip 10
Strip 11
P8-11
RAID level 4
(f)
Strip 0
Strip 1
Strip 2
Strip 3
P0-3
Strip 4
Strip 5
Strip 6
P4-7
Strip 7
Strip 8
Strip 9
P8-11
Strip 10
Strip 11
Strip 12
P12-15
Strip 13
Strip 14
Strip 15
P16-19
Strip 16
Strip 17
Strip 18
Strip 19
RAID level 5

RAID Levels

RAID Level 3

Simplified version of Level 2.

A single parity bit is used instead of error-correcting code, hence required just one extra disk.

If any disk in the array fails, its data can be determined from the data on the remaining disks.

It is as good as Level 2 but is less expensive in the number of extra disks.

RAID Levels

RAID Level 4

It uses block-level striping, as in Level 0, and in addition keeps a parity block on separate disk for corresponding blocks from other disks.

If one of the disks fails, the parity block can be used with the corresponding blocks from other disks to restore the blocks of the fail disks.

The transfer rate for large read as well as large write is high since reads and writes in parallel but small read and write can not be in parallel.

Problem: Parity bottleneck

RAID Levels

RAID Level 5

Similar to level 4 but parity information is distributed in all disks.

For each block one of the disk stores parity and other stores data.

For example, with an array of five disks, the parity for nth blocks is stored in disks (n mod 5) + 1; the nth block of the other four disks stores actual data for that block.

Disk Formatting

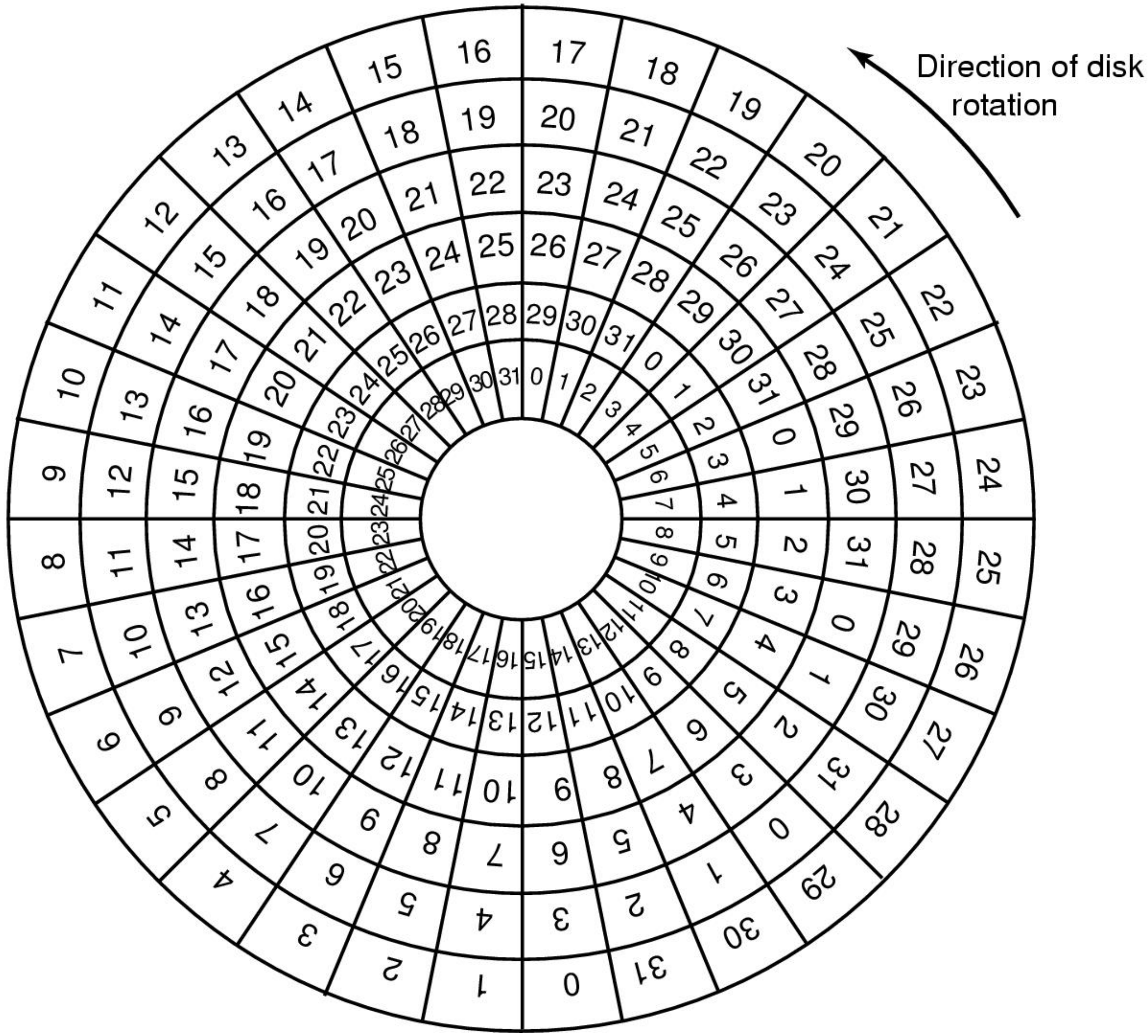
Before a disk can store data, it must be divided into sectors that the disk controller can read and write, called low-level formating.

The sector typically consists of preamble, data and ECC. The preamble contains the cylinder and sector number and the ECC contains redundant information that can be used to recover from read error.

The size depends upon the manufacturer, depending on reliability.

Preamble	Data	ECC

Disk Formatting



Disk Formatting

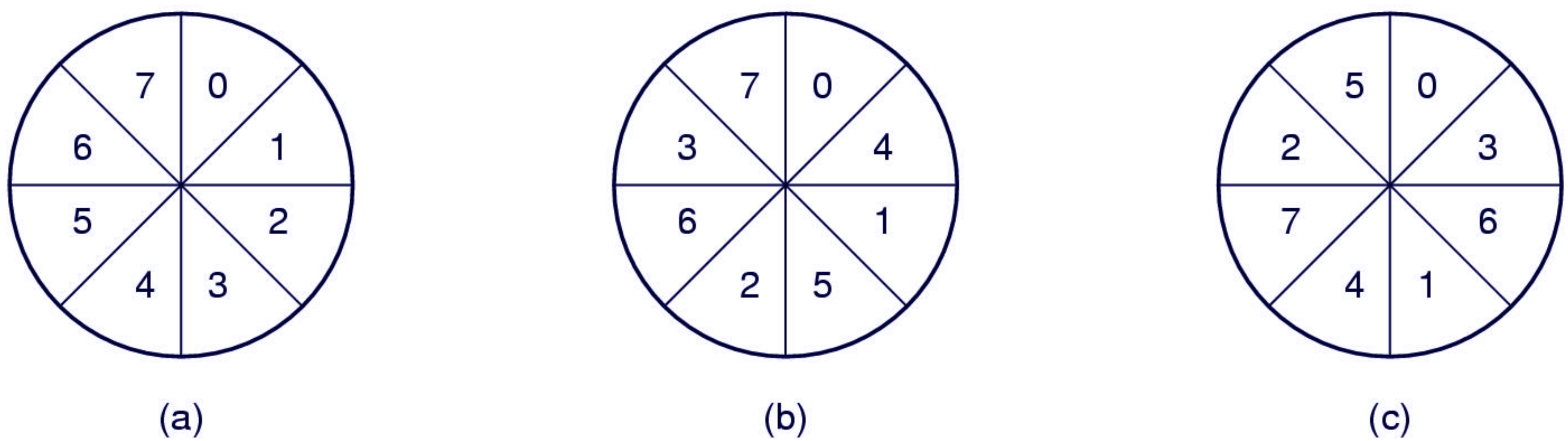
If disk I/O operations are limited to transferring a single sector at a time, it reads the first sector from the disk and doing the ECC calculation, and transfers to main memory, during this time the next sector will fly by the head.

When transferring completes the controller will have to wait almost an entire rotation for the second sector to come around again.

This problem can be eliminated by numbering the sectors in an interleaved fashion when formating the disk.

According to the copying rate, interleaving may be of single or double.

Disk Formatting



a) No interleaving b) Single interleaving c) Double interleaving

Error Handling

Bad Blocks

Most frequently, one or more sectors becomes defective or most disks even come from factory with bad blocks.

Depending on the disk and controller in use, these blocks handled in variety of ways.

1. Bad blocks are handled manually- For example run MS-DOS chkdsk command to find bad block, and format command to create new block – data resided on bad blocks usually are lost.
2. Using bad block recovery- The controller maintains the list of bad blocks on the disk and for each bad block, one of the spares is substituted.

RAM Disks

RAM Disk is virtual block device created from main memory.

Commands to read or write disks blocks are implemented by RAM disk driver.

It completely eliminates seek and rotational delays suffered in disk devices.

RAM disks are particularly useful for storing files that are frequently accessed or temporary.

RAM disks are especially used in high performance applications.

Some OS define the RAM disks at boot time, other dynamically.

Disadvantages: cost and volatility.

The volatility is solve by providing battery backups.

Home Works

HW #13:

1. 14, 15, 17, 19, & 24 from Text Book (Tanenbaum) Ch. 5.
2. Suppose that a disk drive has 5,000 cylinders, numbered 0 to 4999. The drive is currently serving a request at cylinder 143, and the previous request was at cylinder 125. The queue of the pending requests, in FIFO order, is

   86, 1470, 913, 1774, 948, 1502, 1022, 1750, 130.

   Starting from the current head position, what is the total distance (in cylinders) that the disk arm moves to satisfy all the pending requests for each of the following disk-scheduling algorithms?

   a) FCFS b) SSTF c) SCAN d) C-SCAN e) C-LOOK
3. A disk has 8 sectors per track and spins at 600 rpm. It takes the controller 10ms from the end of one I/O operation before before it can issue a subsequent one. How long does it take to read all 8 sectors using the following interleaving system?

   a) No interleaving b)Single interleaving c)Double interleaving