Disk Management

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

Disk Structure

Disk Scheduling

RAID

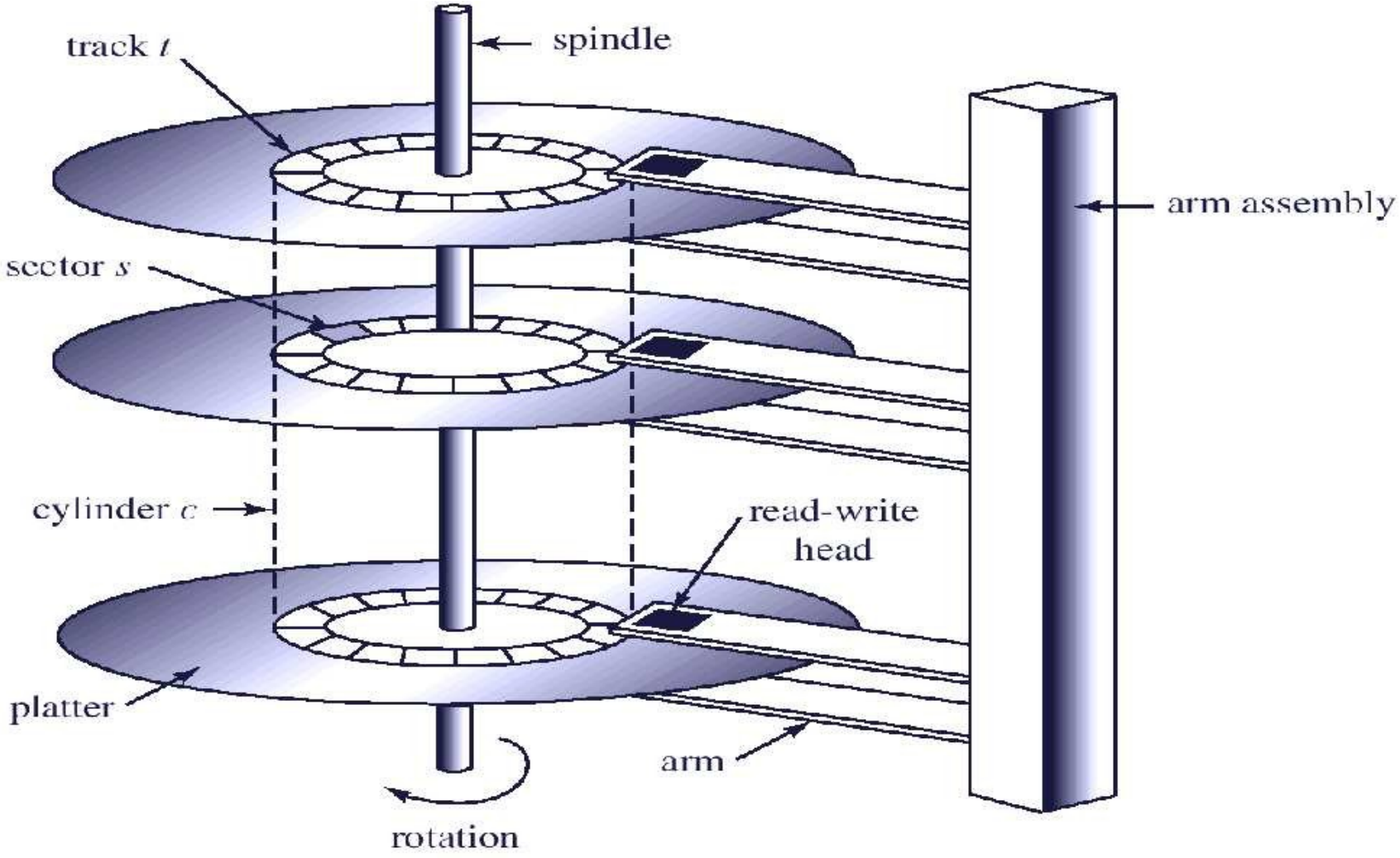
Disk Formatting & Error Handling

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.

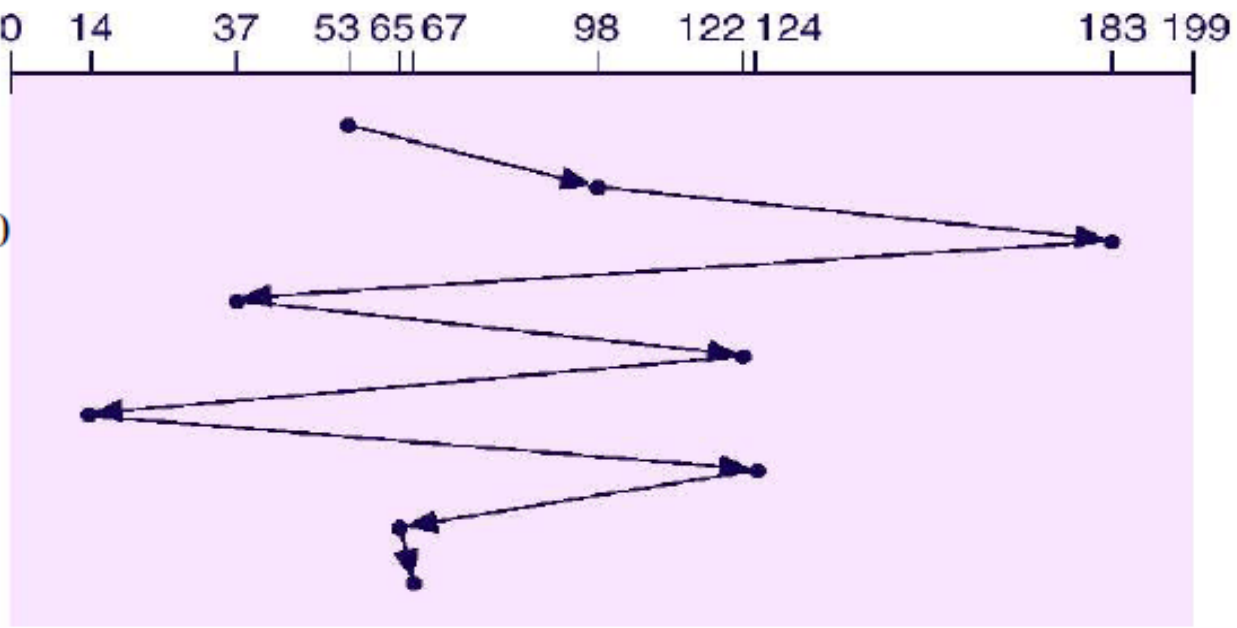
- For most disks, the seek time dominates the other two times, so reducing the mean seek time can improve system performance substantially.

First-Come First-Served (FCFS)

The first request to arrive is the first one serviced.

Example:

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



Total Head Movement = 640 cylinders

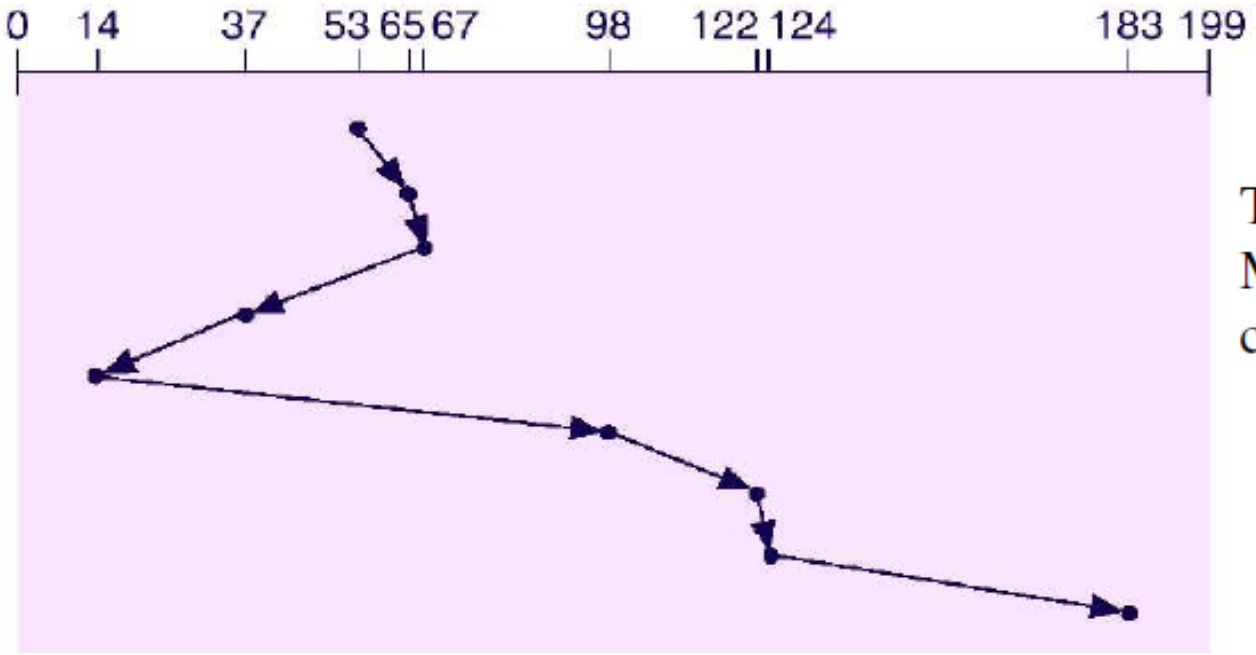
Advantages: Simple and Fair.

Problems: Does not provide fastest service.

Shortest-Seek-Time-First (SSTF)

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

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



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

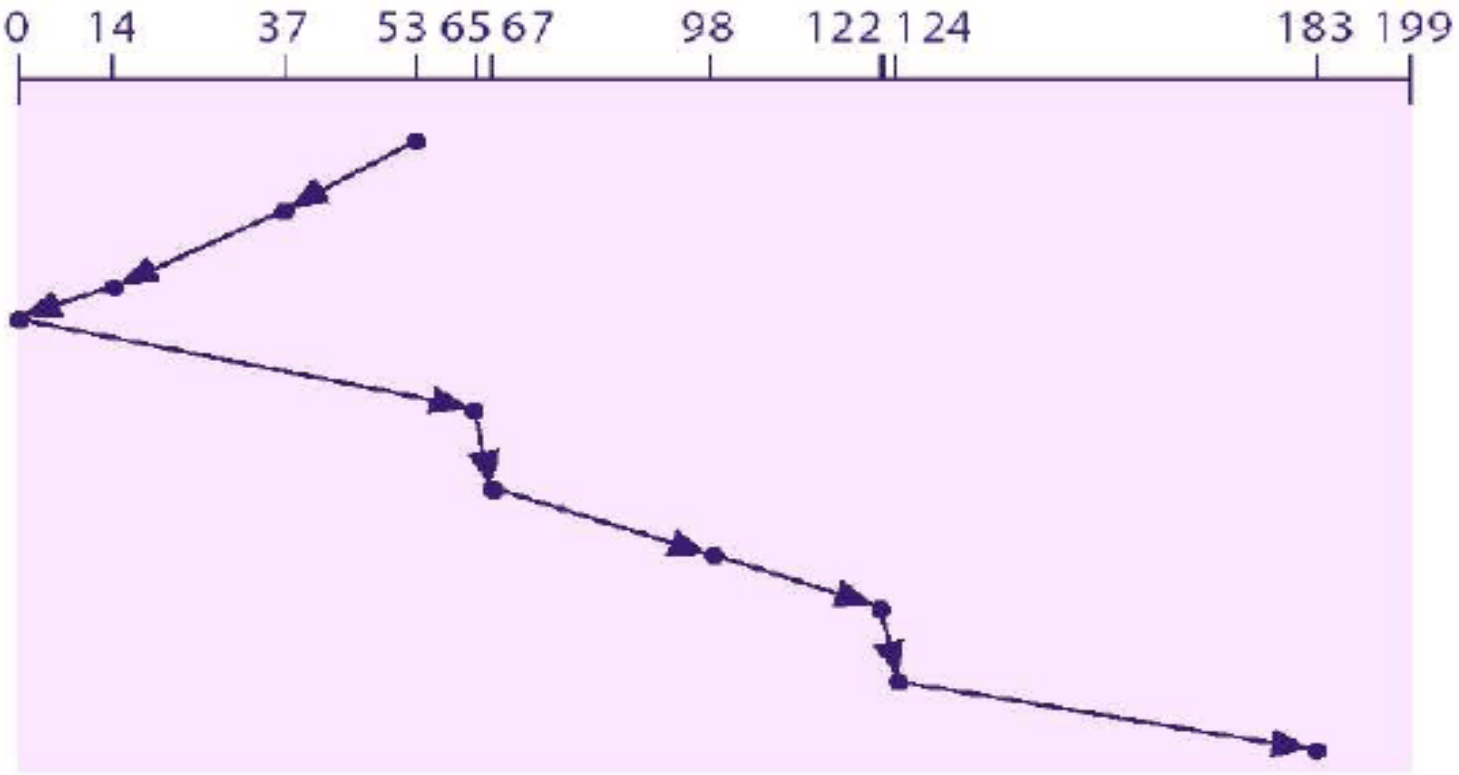
The disk arm starts at one end of the disk, and moves toward the other end, servicing requests until it gets to the other end of the disk, where the head movement is reversed and servicing continues.

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

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



Total Head Movement = 208 cylinders

C-SCAN

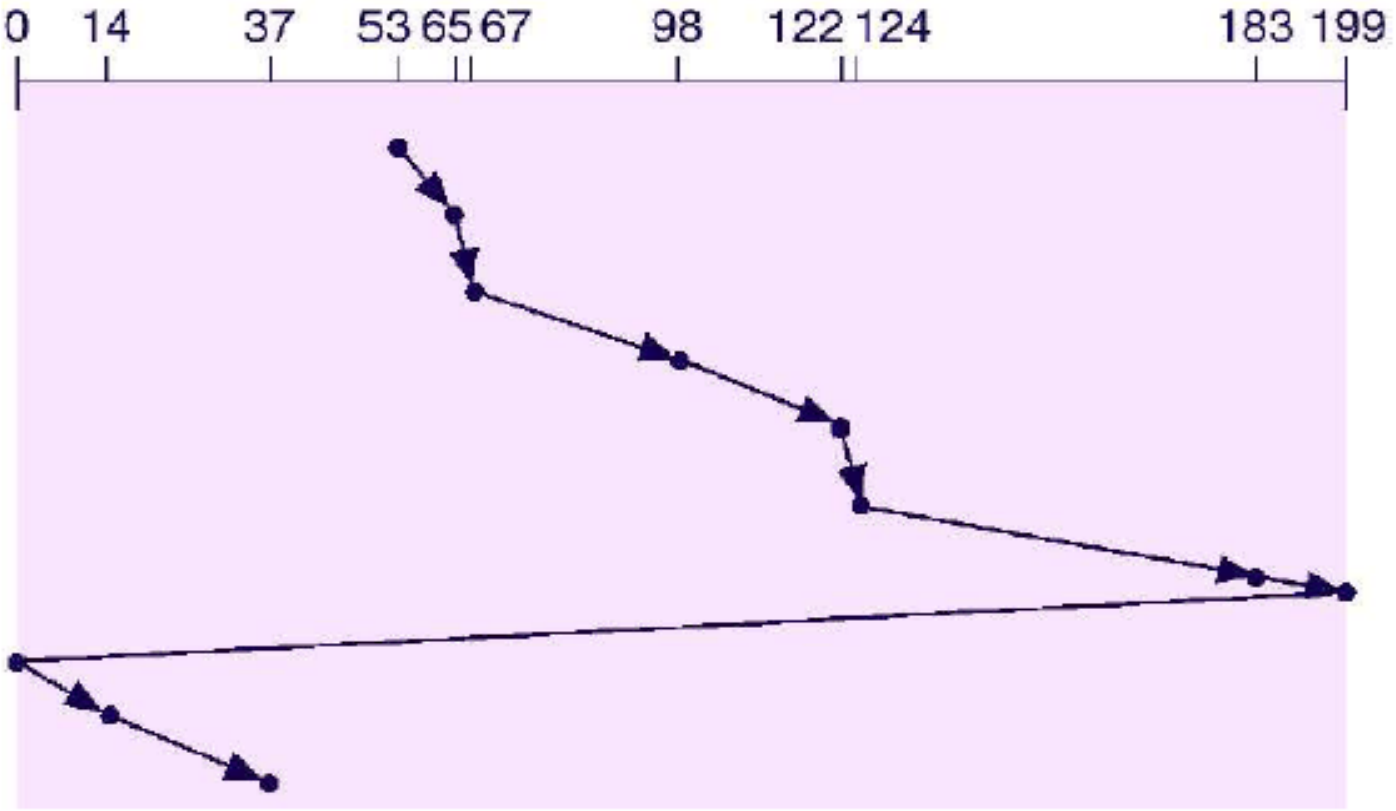
- When a uniform distribution of request for cylinders, only the few request are in extreme cylinders, since these cylinders have recently been serviced.

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

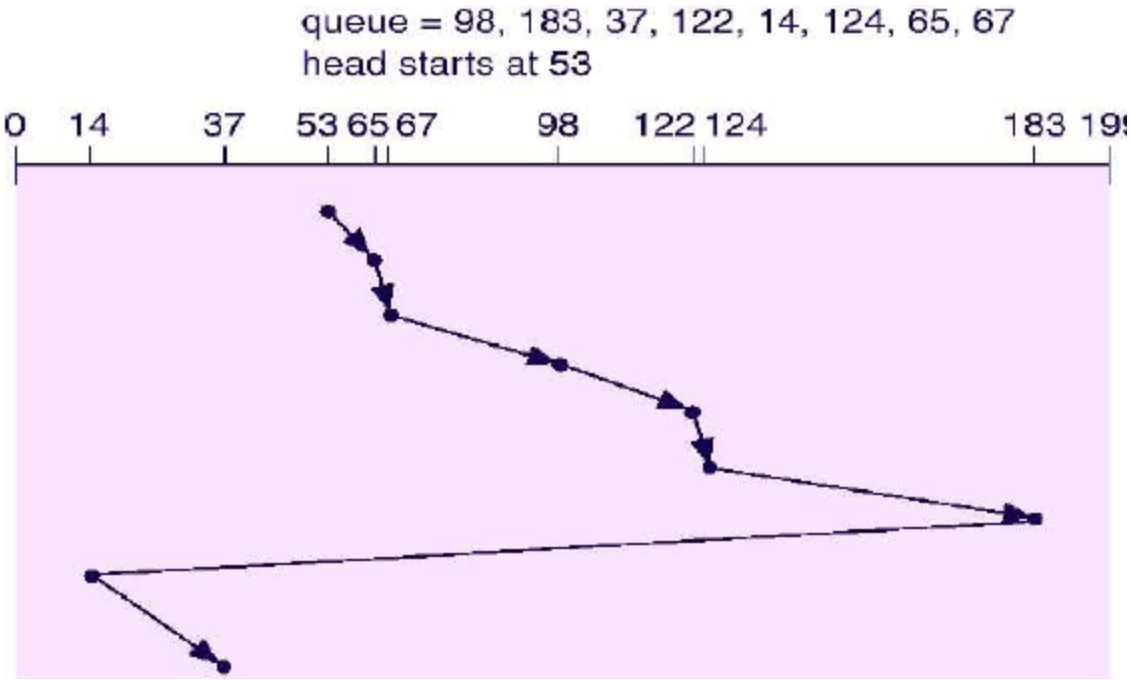
queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53



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.


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

SLED Vs RAID

- SLED
  - Single Large Expensive Disk
- RAID
  - Redundant Array of Inexpensive (Independent) Disks.

RAID

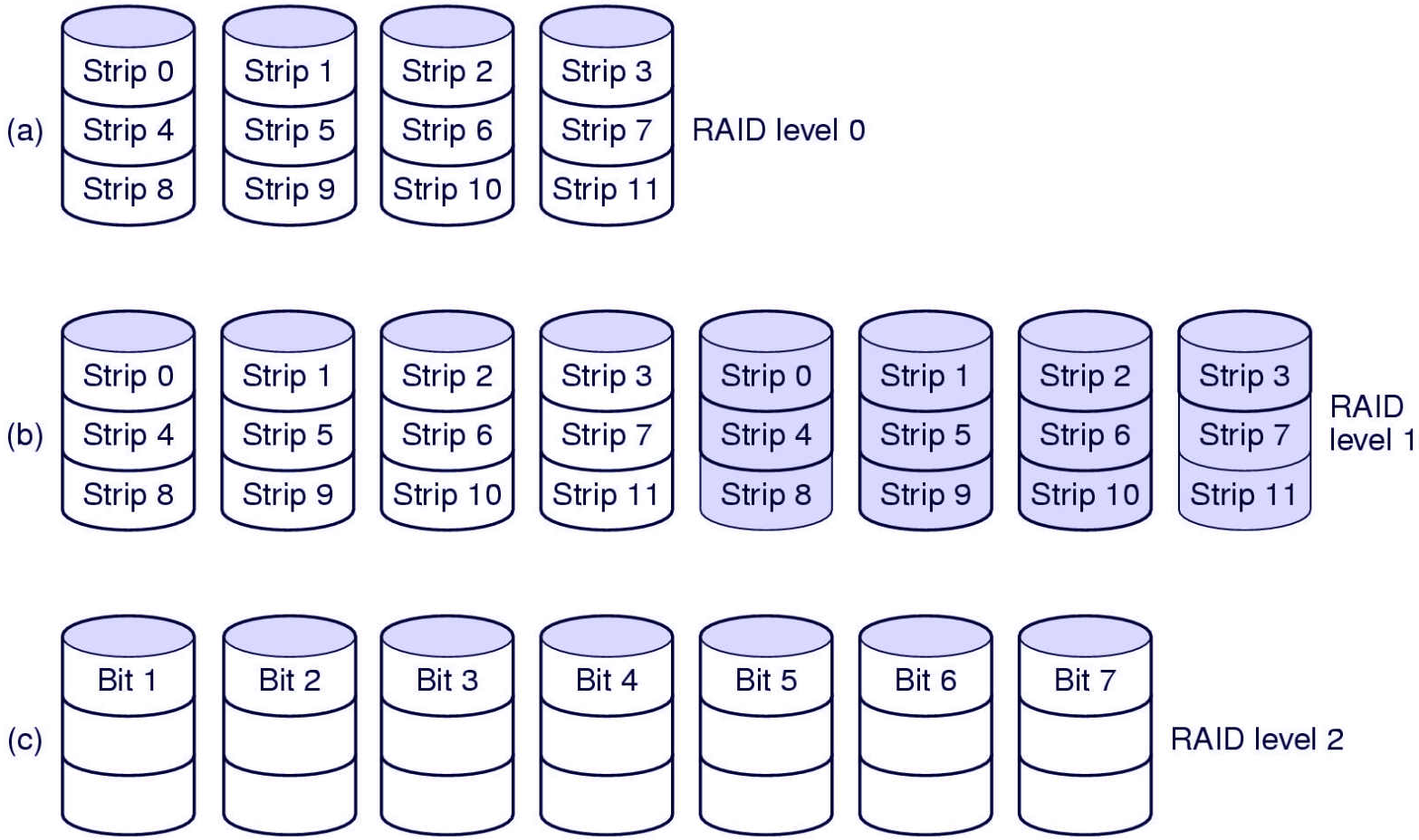
Redundant Array of Inexpensive (Independent) Disks.

CPU performance has been increasing exponentially over the past decade, roughly doubling every 18 months. Not so with disk performance. In the 1970s, average seek times on minicomputer disks were 50 to 100 msec

- **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 organization of RAID called RAID levels



RAID Level 0

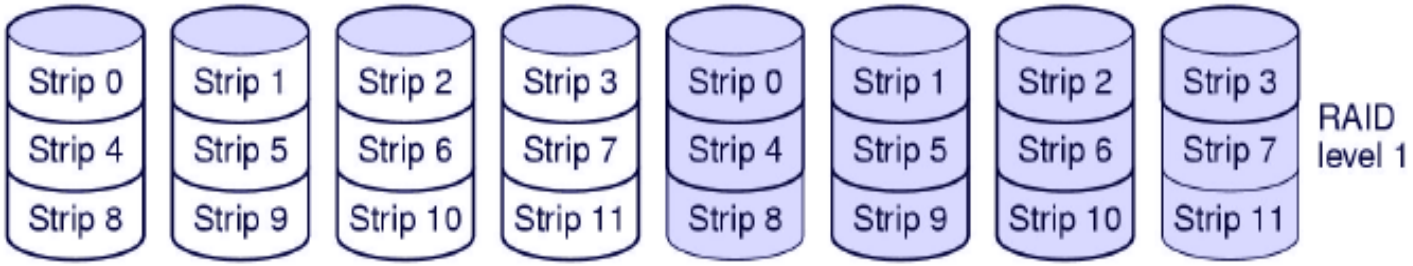
- RAID level 0 creates one large virtual disk from a number of smaller disks.
- Storage is grouped into logical units called strips (It is a collection of sectors) with the size of a strip being some multiple of sector size.
- The virtual storage is sequence of strips interleaved among the disks in the array.
- Distributing data over multiple drives like this is called **striping**.

Discussion : RAID level 0

- if the software issues a command to read a data block consisting of four consecutive strips starting at a strip boundary, the RAID controller will break this command up into four separate commands, one for each of the four disks, and have them operate in parallel. Thus we have parallel I/O without the software knowing about it.
- RAID level 0 works worst with operating systems that habitually ask for data one sector at a time. The results will be correct, but there is no parallelism and hence no performance gain.
- Another disadvantage of this organization is that the reliability is potentially worse than having a SLED.
- Because no redundancy is present in this design, it is not really a true RAID.

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

Discussion : RAID level 1

- write performance is no better than for a single drive, but read performance can be up to twice as good.
- Fault tolerance is excellent: if a drive crashes, the copy is simply used instead.
- Recovery consists of simply installing a new drive and copying the entire backup drive to it.
- Double space !!!

RAID Level 2

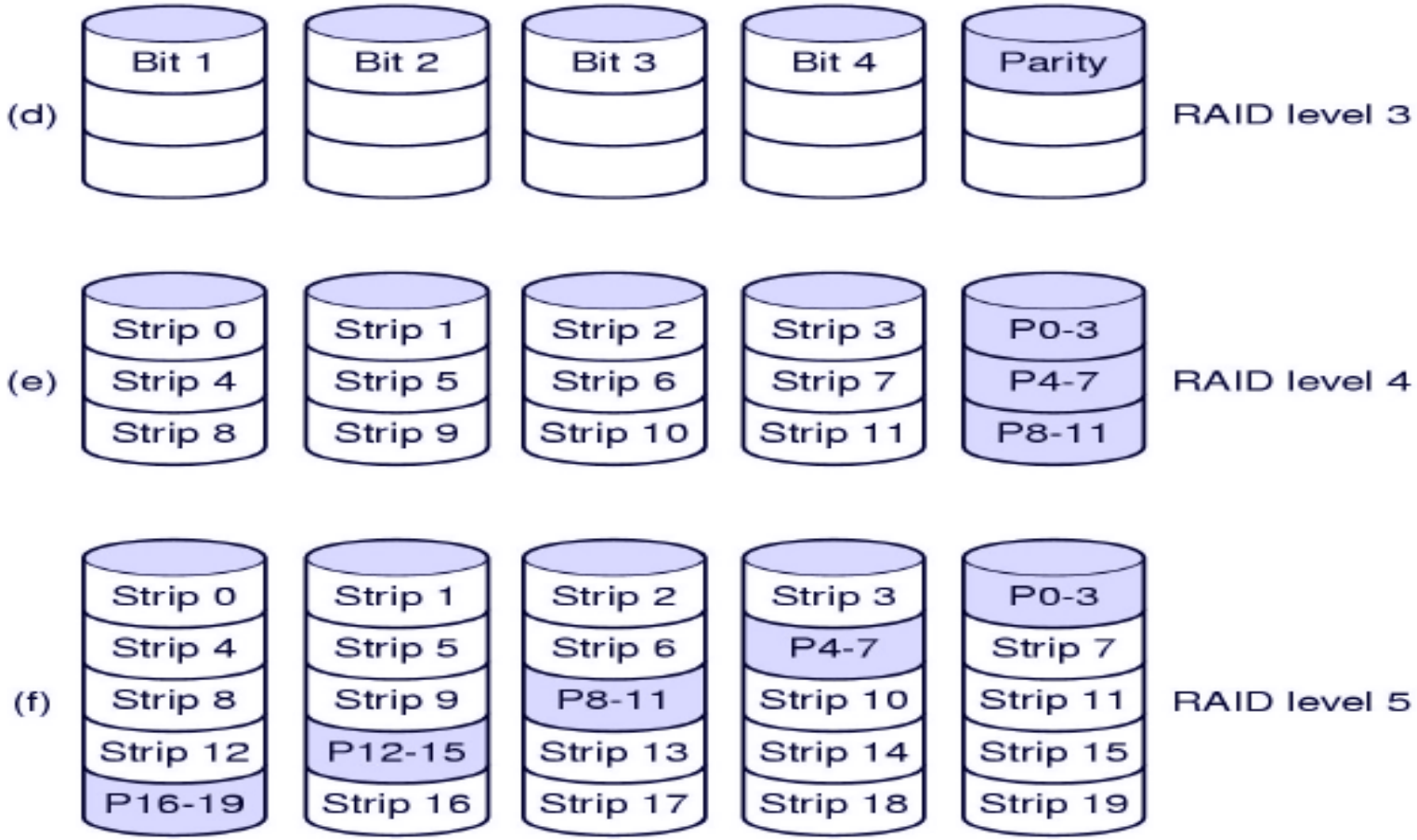
- An error-correcting code is used for corresponding bits on each data disks.
- Error-correcting scheme store two or more extra bits, and can reconstruct the data if a single bit get damaged
- For Example, the first bit of each byte is stored in disk 1, second bit in disk 2, and until eight bit in disk 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.

RAID level 2

RAID level 2

- Advantages: Total parallelism at bit level.
- Disadvantages: Requires substantial number of drives. Out of 7 drives, 3 are for error correcting code
- Hamming code of 7 bit= 4 bit original +3 bit error correcting code

RAID levels


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

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 Level 4

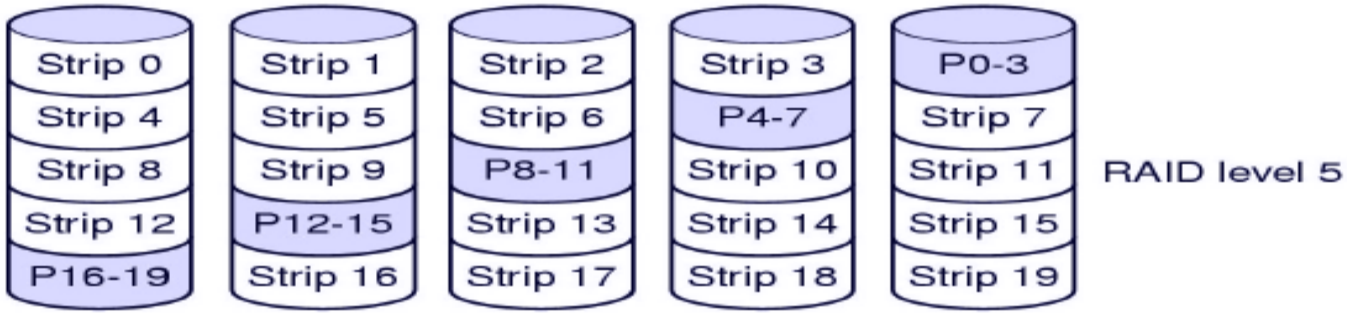
- It uses block-level striping, as in Level 0, and in addition keeps a parity block on separate disk for corresponding blocks from other disks.
- If one of the disks fails, the parity block can be used with the corresponding blocks from other disks to restore the blocks of the fail disks.
- The transfer 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.

RAID level 4

- This design protects against the loss of a drive but performs poorly for small updates. If one sector is changed, it is necessary to read all the drives in order to recalculate the parity, which must then be rewritten.
- As a consequence of the heavy load on the parity drive, it may become a bottleneck.

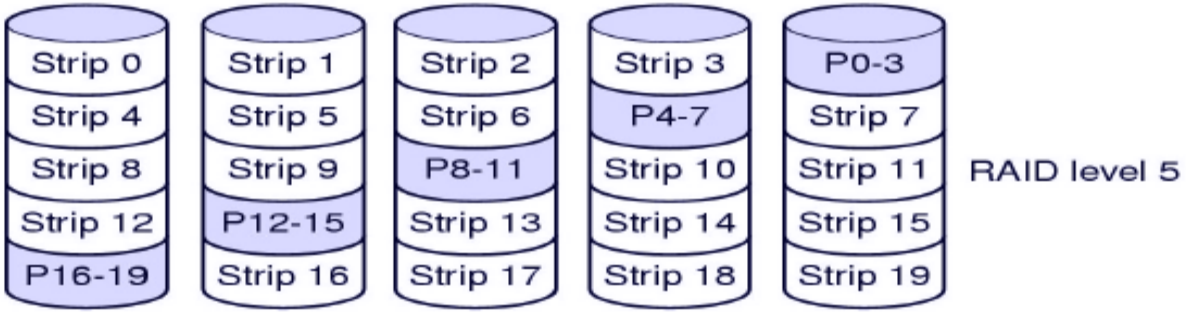
RAID level 5

- The bottleneck of parity is eliminated in RAID level 5 by distributing the parity bits uniformly over all the drives, round robin fashion,



RAID level 5

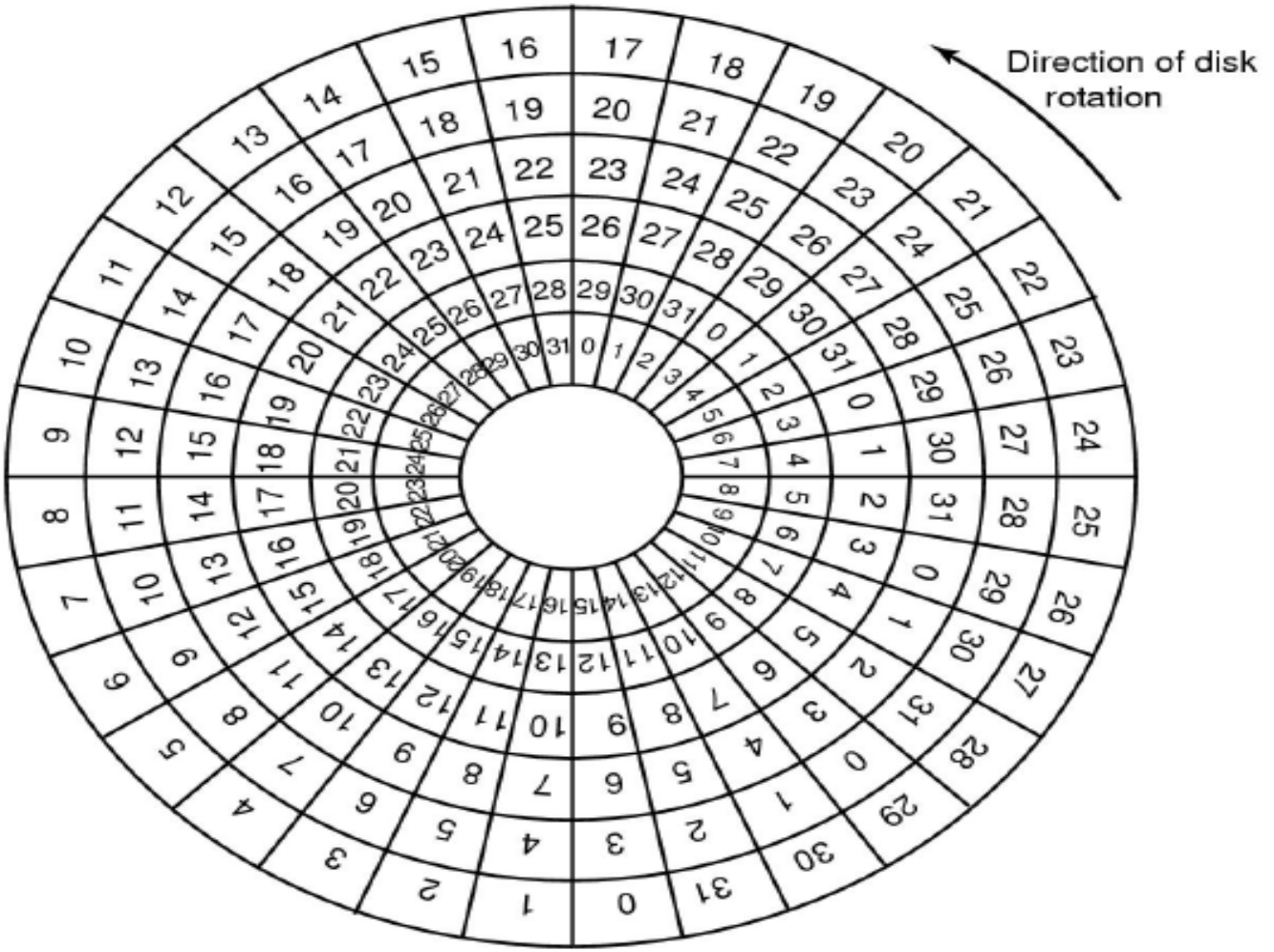
- Similar to level 4 but parity information is distributed in all disks.
- For each block one of the disk stores parity and other stores data.
- For example, with an array of five disks, the parity for nth blocks is stored in disks (n mod 5) + 1; the nth block of the other four disks stores actual data 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 formatting.
- 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.

Disk Formatting



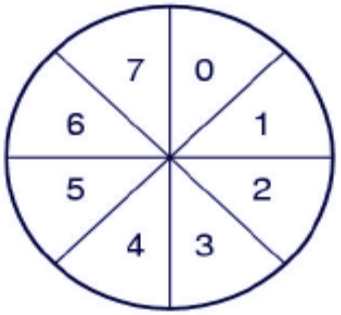
Disk Formatting

- If disk I/O operations are limited to transferring a single sector at a time, it read 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.

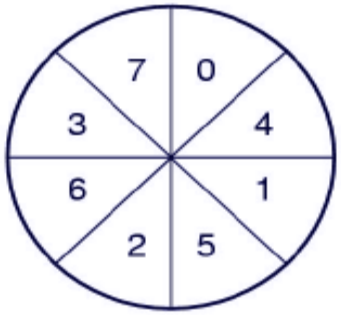
- When transferring completes, the controller will have to wait almost an entire rotation for the second sector to come around again.

Disk Formatting

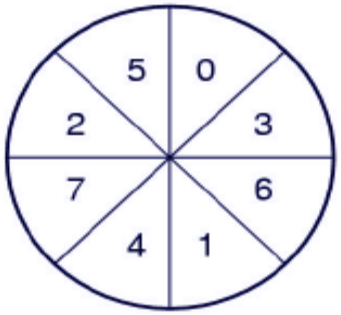
- This problem can be eliminated by numbering the sectors in an interleaved fashion when formatting the disk.
- According to the copying rate, interleaving may be of single or double.



(a)



(b)



(c)

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

RAM Disk

- 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 because main memory is direct access.
- 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.

RAM Disk: Discussion

Disadvantages: cost and volatility.

- To implement RAM disk, OS must reserve a section of main memory for RAM disk use. So additional space should be purchased otherwise memory will be insufficient for use by process.
- It loses its content when power is off. If RAM disk is used to store file system, that file system must be remade each time a system is boot.
- The volatility is solve by providing battery backups.

Exercise

- 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

Exercise

- 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 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