Computer Organization and Software Systems

Contact Session 3

Dr. Lucy J. Gudino

Last Class

Contact Hour	List of Topic Title	Text/Ref Book/external resource
3-4	Memory Organization - Internal Memory - External Memory (HDD)	T1, R2

Today’s Session

Contact Hour	List of Topic Title	Text/Ref Book/external resource
5-6	**Memory Organization** - External Memory (RAID, SSD) Cache Memory Organization - Locality - Locality of Reference to Program Data - Locality of instruction fetches	T1, R2

Types of External memory

- Magnetic Disk
  - RAID Memories
  - Removable Disks
- Optical
  - CD-ROM
  - CD-Recordable (CD-R)
  - CD-R/W
  - DVD
- Magnetic Tape

Magnetic Disk Drive

1.8 - 5.25 inch

5400 rpm - 15000 rpm

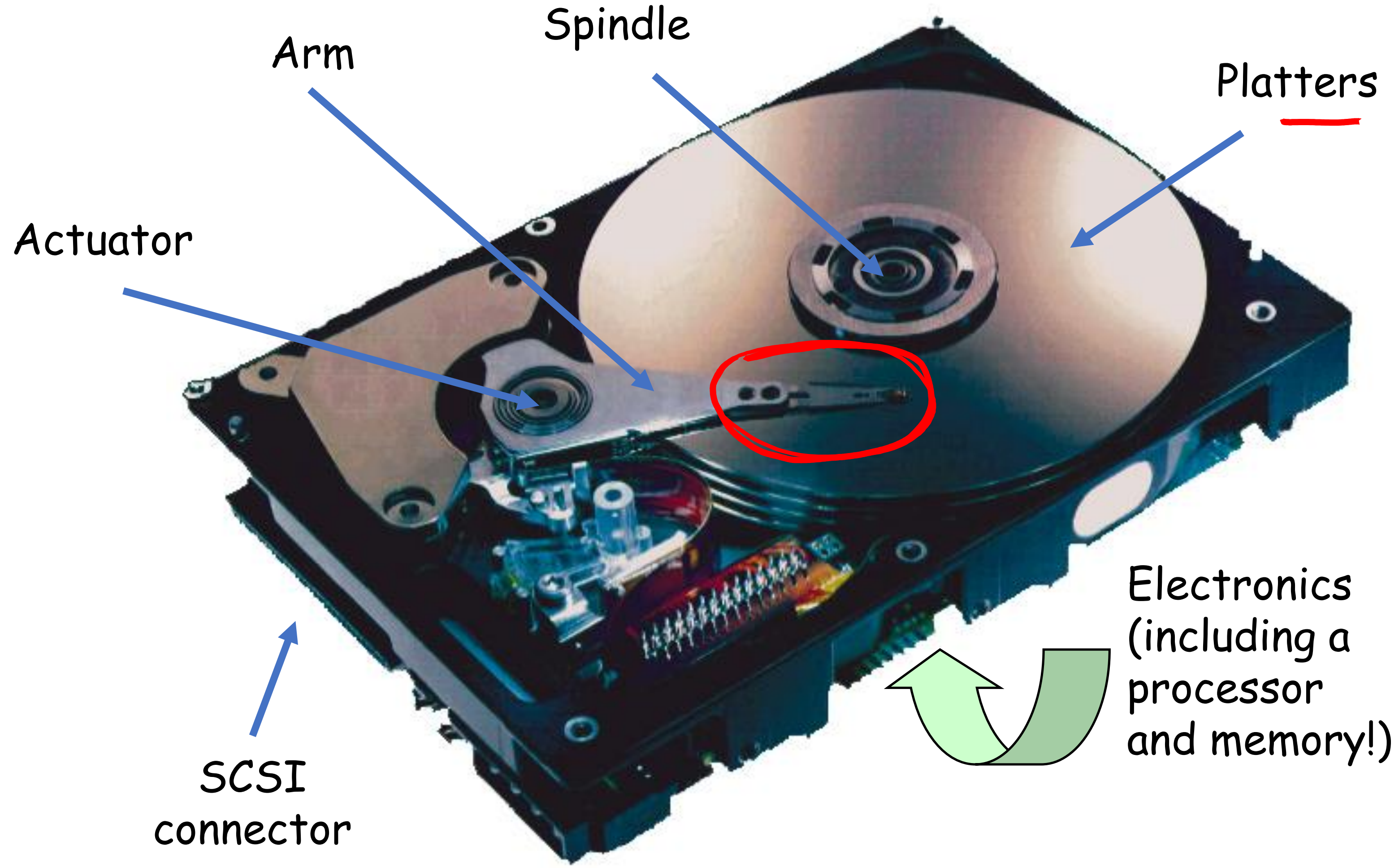


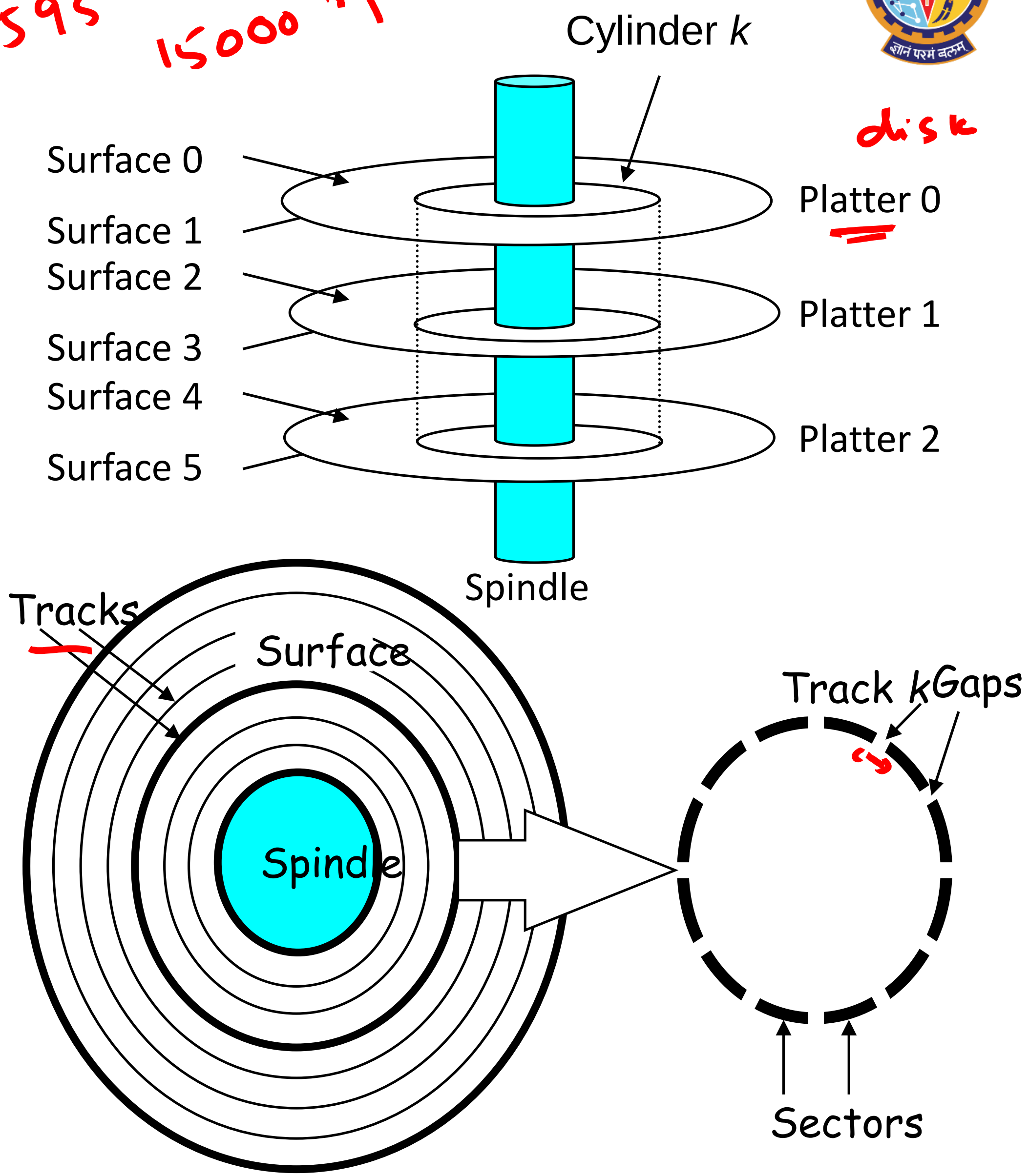
Image courtesy of Seagate Technology

Disk Geometry

- Disks consist of platters, each with two surfaces.
- Each surface consists of concentric rings called tracks
- Aligned tracks form a cylinder
- Each track consists of sectors separated by gaps

Cheetah 450GB - 512B
595 848 tracks
15000 rpm

Toshiba 5TB
4096B
32 79 583
7199 rpm



disk

Disk Capacity

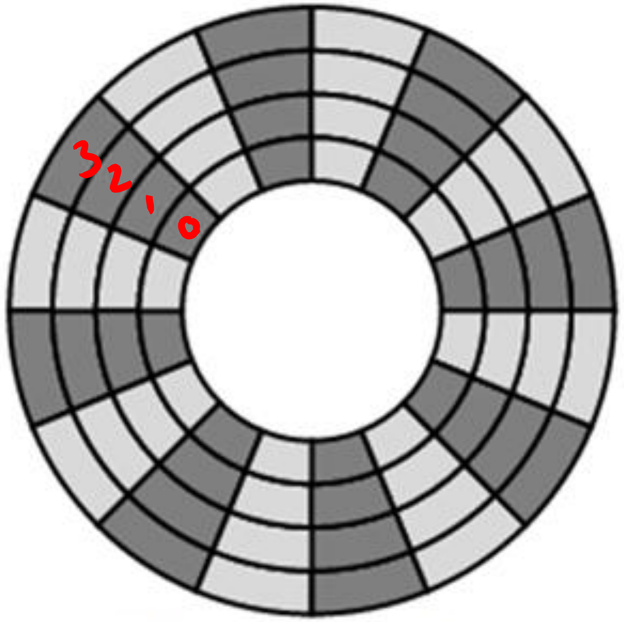
GB or TB

- Capacity: maximum number of bits that can be stored.
  - Vendors express capacity in units of gigabytes (GB /TB), where 1 GB = 2^{30} Bytes, 1 TB = 2^{40} Bytes,
- Capacity is determined by these technology factors:
  - Recording density (bits/in): number of bits that can be squeezed into a 1 inch segment of a track.
  - Track density (tracks/in): number of tracks that can be squeezed into a 1 inch radial segment.
  - Areal density (bits/in2): product of recording and track density.

Recording zones

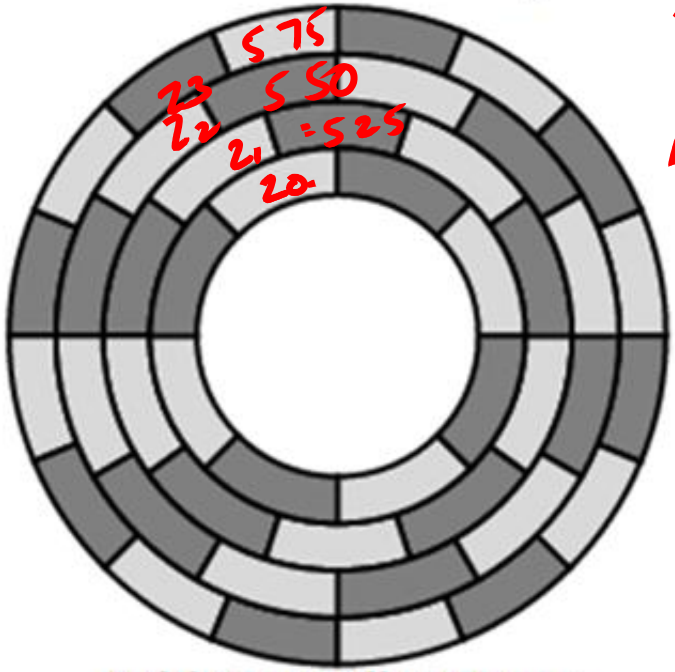
Zones

- Modern disks partition tracks into disjoint subsets called recording zones
  - Each track in a zone has the same number of sectors, determined by the circumference of innermost track.
  - Each zone has a different number of sectors/track, outer zones have more sectors/track than inner zones.
  - So we use **average** number of sectors/track when computing capacity.



Without Recording Zones

40 tracks



10 tracks

⇓

500

With Recording Zones

Computing Disk Capacity

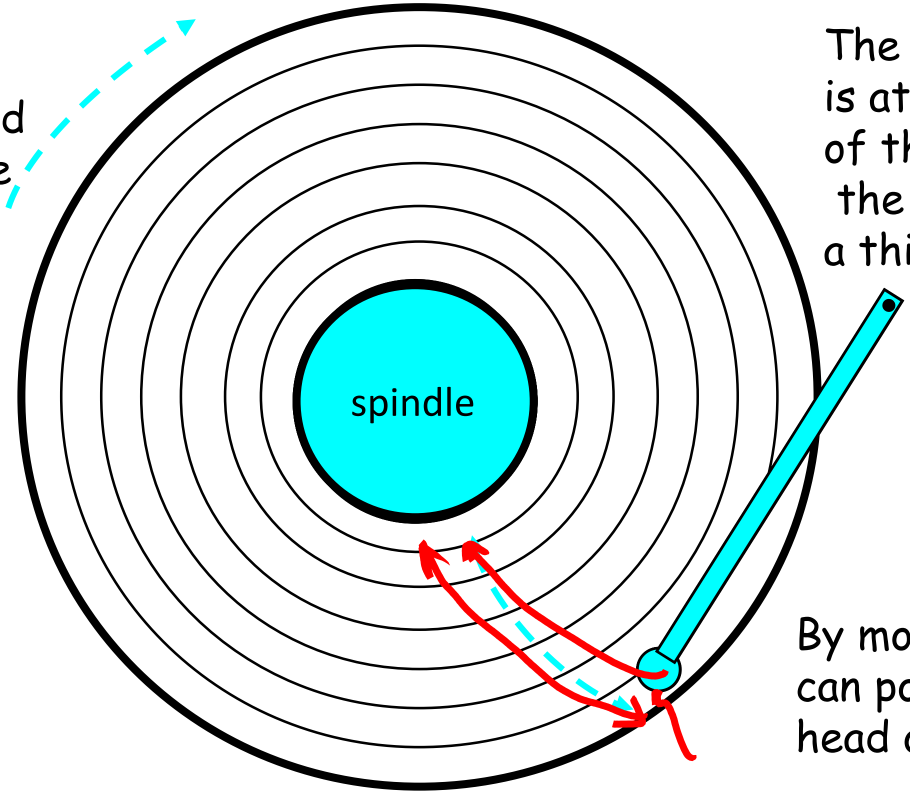
- Capacity = (# bytes/sector) x (avg. # sectors/track) x (# tracks/surface) x (# surfaces/platter) x (# platters/disk)
- Example:
  - 512 bytes/sector
  - 300 sectors/track (on average)
  - 20,000 tracks/surface
  - 2 surfaces/platter
  - 5 platters/disk

- Capacity = 512 x 300 x 20000 x 2 x 5
  = 30,720,000,000
  = 28.61 GB

10^{+12}

Disk Operation (Single-Platter View)

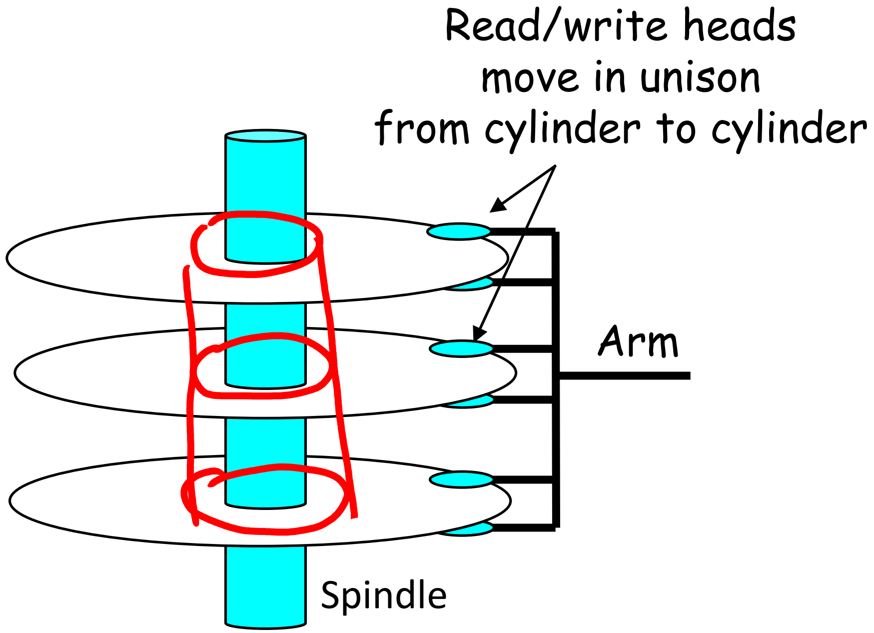
The disk surface spins at a fixed rotational rate



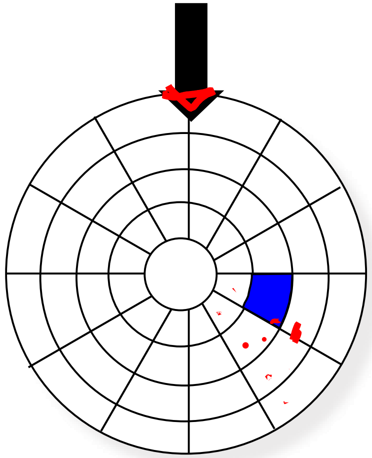
The read/write *head* is attached to the end of the *arm* and flies over the disk surface on a thin cushion of air.

By moving radially, the arm can position the read/write head over any track.

Disk Operation (Multi-Platter View)

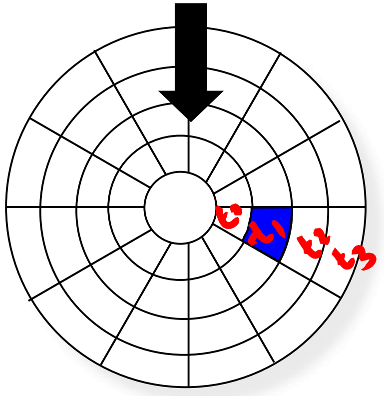


Disk Access



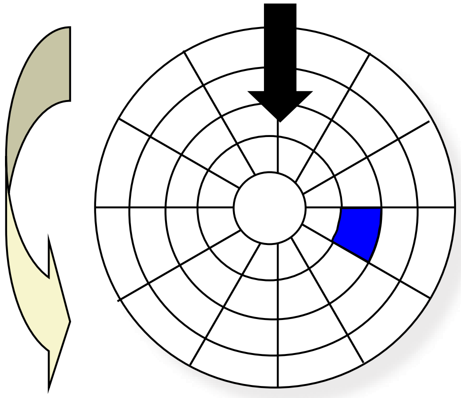
Need to access a sector colored in blue

Disk Access



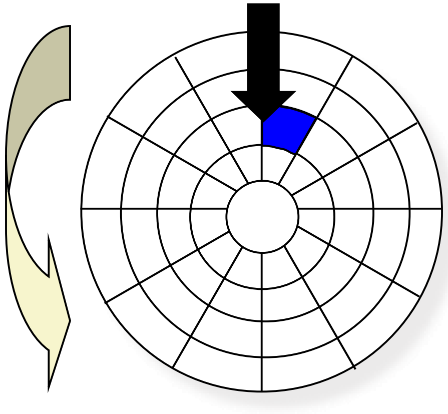
Head in position above a track

Disk Access



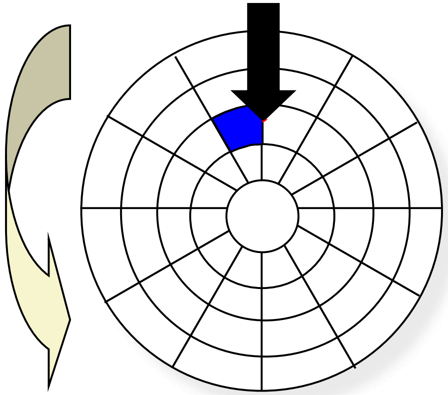
Rotate the platter in counter-clockwise direction

Disk Access - Read



About to read blue sector

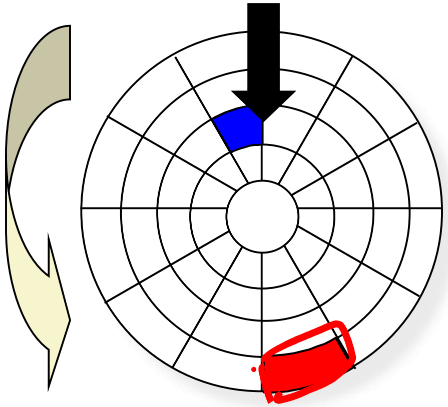
Disk Access - Read



After BLUE read

After reading blue sector

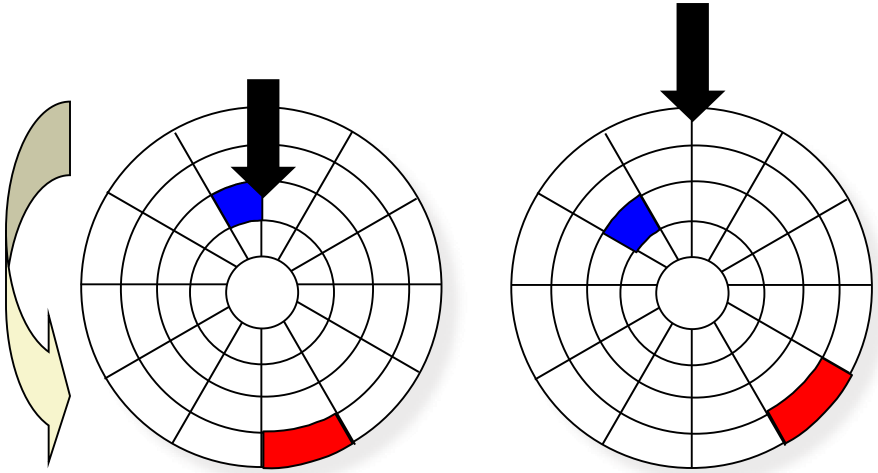
Disk Access - Read



After BLUE read

Red request scheduled next

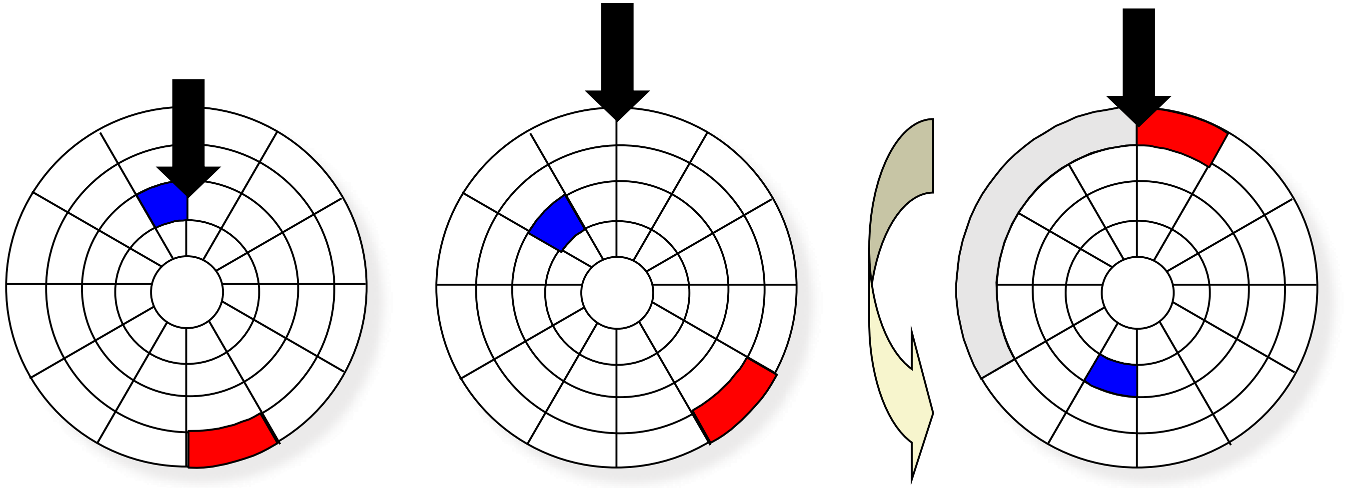
Disk Access - Seek



After BLUE read

Seek for RED

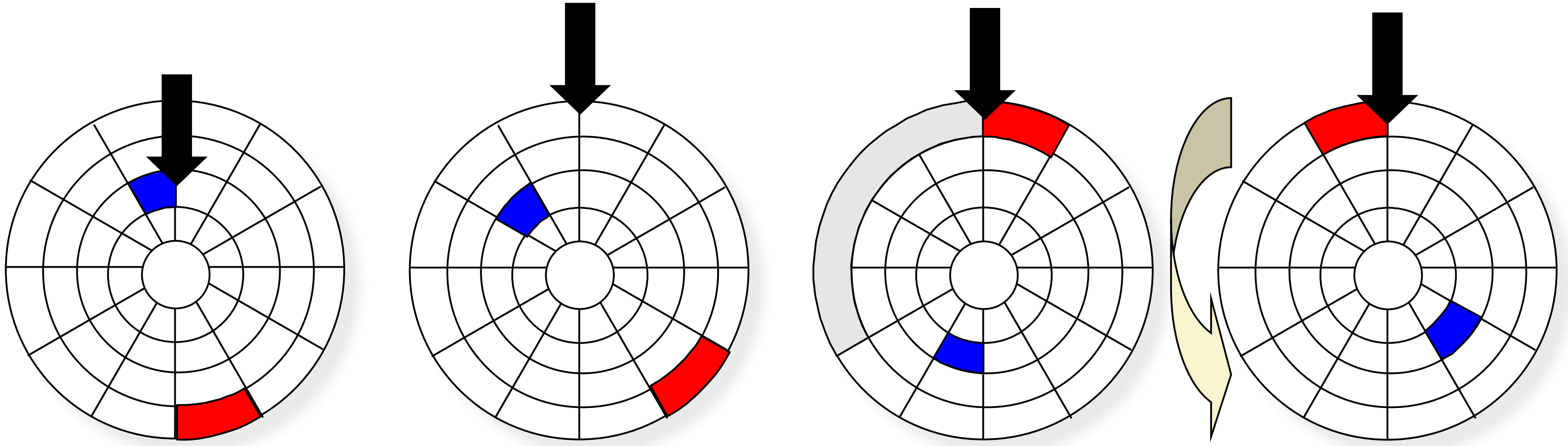
Disk Access – Rotational Latency



After BLUE read Seek for RED Rotational latency

Wait for red sector to rotate around

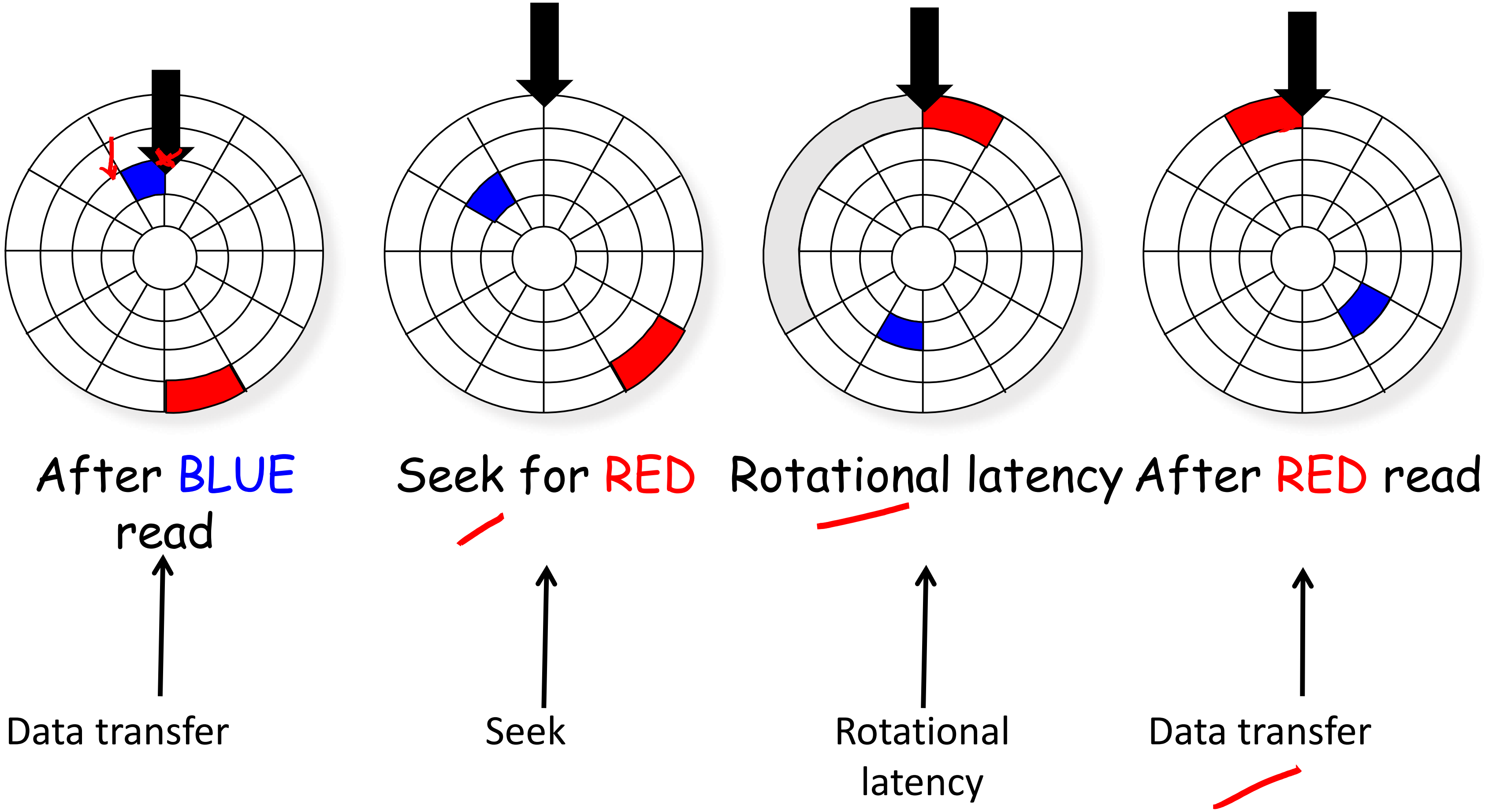
Disk Access – Read



After BLUE read | Seek for RED | Rotational latency | After RED read

Complete read of red

Disk Access – Access Time Components

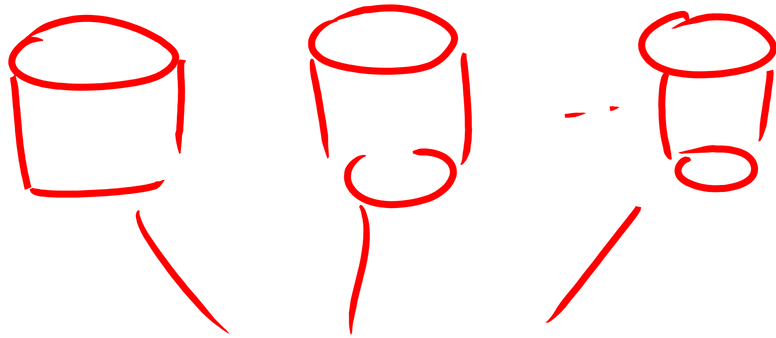


Today’s Class

Contact Hour	List of Topic Title	Text/Ref Book/external resource
5-6	**Memory Organization** - External Memory (RAID, SSD) Cache Memory Organization - Locality - Locality of Reference to Program Data - Locality of instruction fetches	T1, R2

RAID

- RAID - Redundant Array of Independent Disks
- Variety of ways in which the data can be organized
- Need for RAID:
    - Parallel I/O
    - Reliability
    - Redundancy through multiple inexpensive disks



Common Characteristics

1. RAID is a 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.

   byte

   block
3. Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure.

Categories

- RAID category
  - Striping (Level 0)
  - Mirroring (Level 1)
  - Parallel access (Level 2,3)
  - Independent access (Level 4,5,6)

RAID Level 0

- Is not a true member of RAID family – No Redundancy
- The data are striped across the available disks
- Data are distributed across all of the disks in the array
- IO requests can be issued in Parallel
- Reduce IO queuing time

file 1 : 600 bits / 8
file 2
file 3
Strip < 8 bits

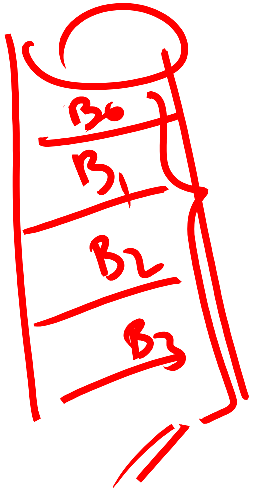
data

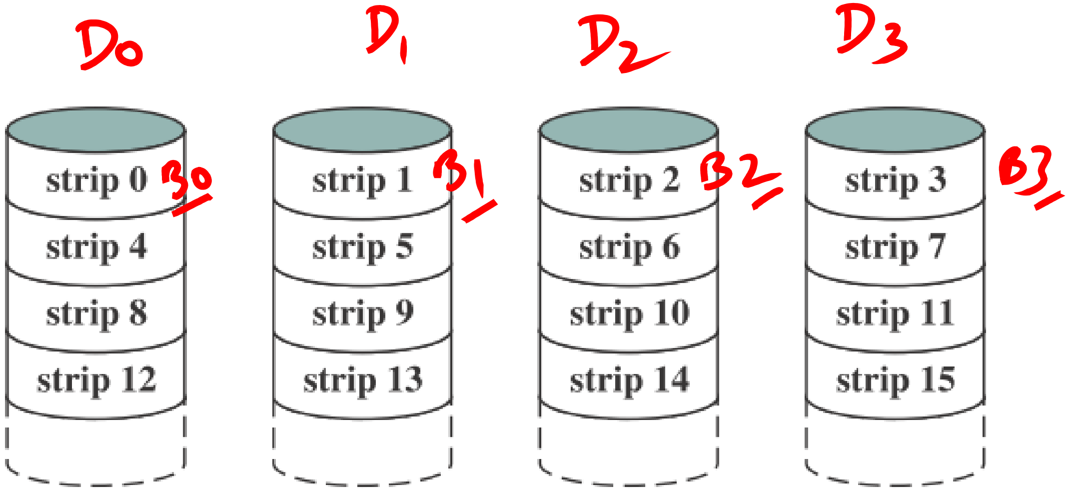
32 bits

block size = 8 bits

blocks = 32/8
= 4 block

B0 - 0-7
B1 - 8-15
B2 = 16-23
B3 = 24-31



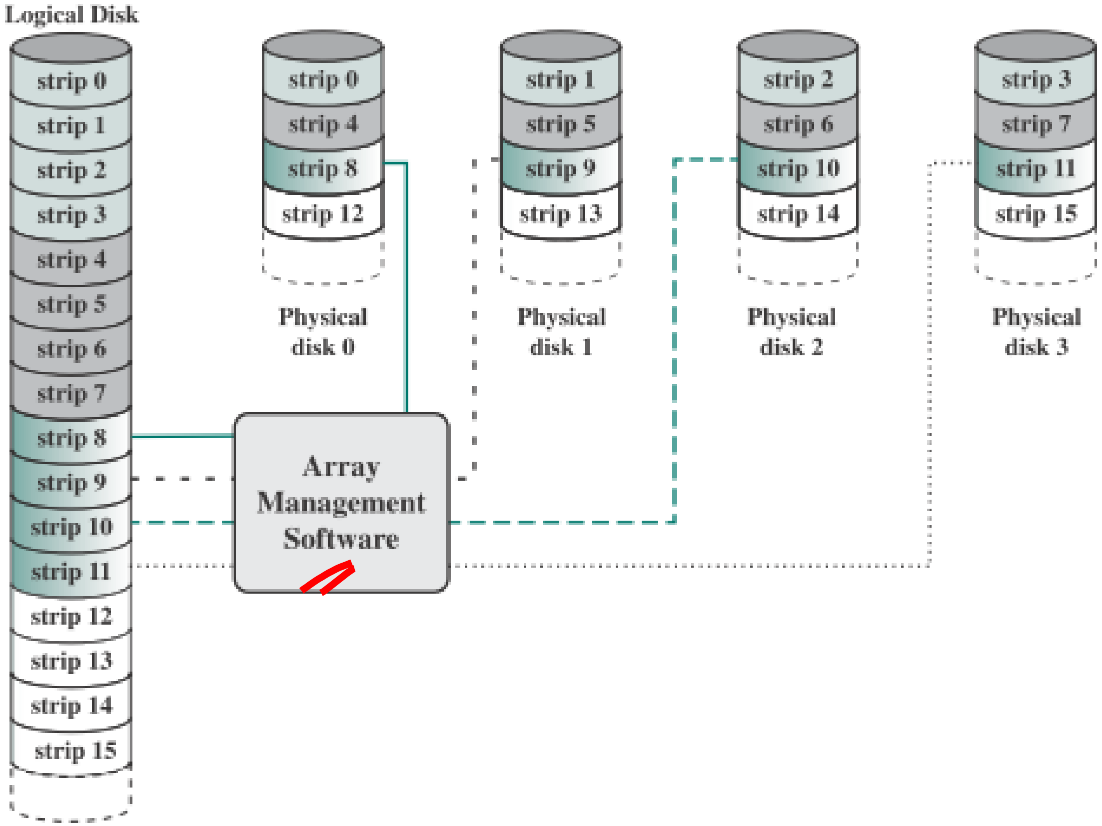


(a) RAID 0 (Nonredundant)

RAID level 0 configuration

cpu – logical } disk
Adm – physical ⇓ OS

- RAID 0 for
  - high data transfer capacity
  - high I/O request rate



RAID Level 0

Level	Advantages	Disadvantages	Applications
0	• Very simple design • Easy to implement • I/O performance is greatly improved by spreading the I/O load across many drives • No parity calculation overhead is involved	The failure of just one drive will result in all data in an array being lost	• Video production and editing • Image Editing • Any application requiring high bandwidth

RAID Level 1 : disk mirroring.

N → data storage

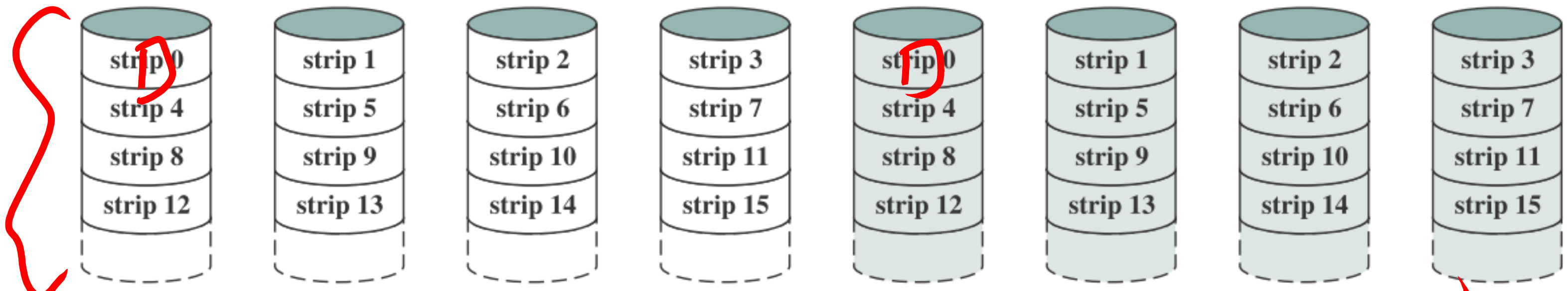
N → redundant data

- Mirroring: redundancy is achieved by duplicating all the data
- Each logical strip is mapped to two separate physical disks
- Every disk in the array has a mirror disk that contains the same data

4

4

Mirror image



(b) RAID 1 (Mirrored)

RAID Level 1

Level	Advantages	Disadvantages	Applications
1	• Simplest RAID storage subsystem design • 100% redundancy of data • RAID 1 can sustain multiple simultaneous drive failures	• Highest disk overhead of all RAID types • inefficient	• Accounting • Payroll • Financial • Any application requiring very high availability

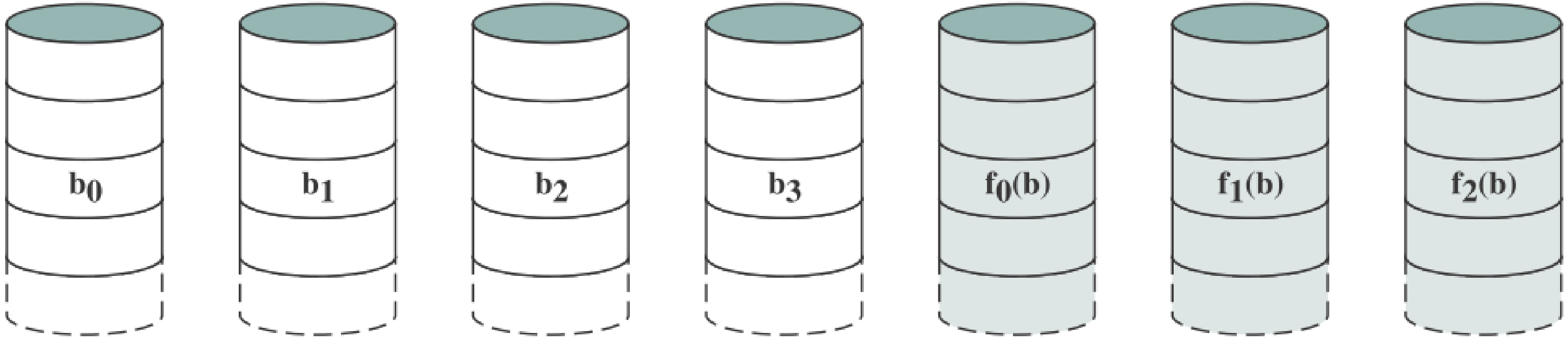
RAID Level 2

K = Check bits disk
M = data bits → no data disk

$2^K - 1 >= M + K$

- Use 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

K = 1 $2^1 - 1 \geq 4 + 1 \Rightarrow F$
K = 2 $2^2 - 1 \geq 4 + 2 \Rightarrow F$
K = 3 $2^3 - 1 \geq 4 + 3 \Rightarrow T$



(c) RAID 2 (Redundancy through Hamming code)

RAID Level 2

4 3

4 4

Level	Advantages	Disadvantages	Applications
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

RAID Level 3

- RAID 3 is organized in a similar fashion to RAID 2
- RAID 3 requires only a single redundant disk
- RAID 3 employs parallel access, with data distributed in small strips
- Parity bit is computed for the set of individual bits in the same position on all of the data disks

d_5 $\underbrace{d_4\ d_3\ d_2\ d_1\ d_0}_{\text{data bits}}$ even parity

1 1 0 0 1 1

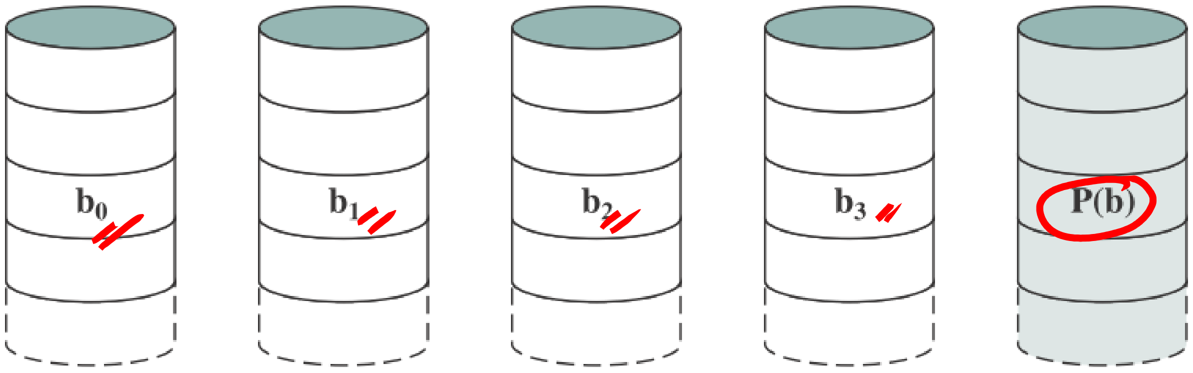
Parity

1 1 1 0 0 1

0 1 1 0 0 0

ASCII - 7bit } even parity

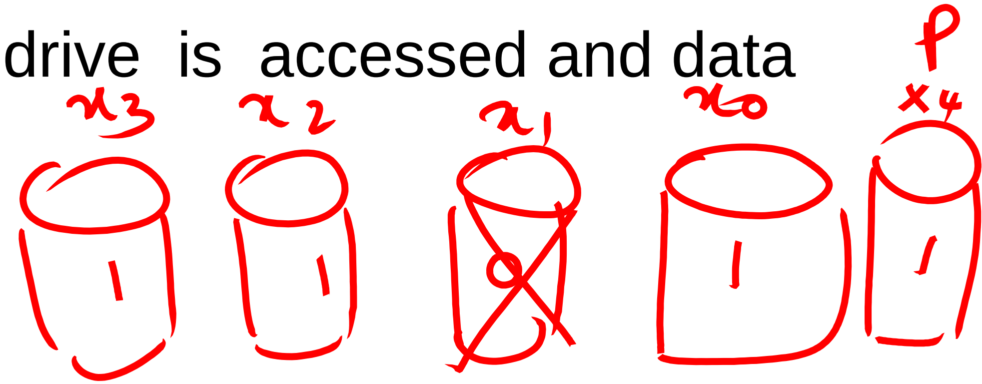
8th - parity



(d) RAID 3 (Bit-interleaved parity)

RAID Level 3

- Redundancy: Event of a drive failure - parity drive is accessed and data is reconstructed from the remaining devices
    - Missing data can be restored on the new drive



- Data reconstruction is simple

$$X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

$= 1 \oplus 1 \oplus 0 \oplus 1$

$= 1$

1 bit

- Suppose drive X1 fails

EXOR

$$X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$$

$= 1 \oplus 1 \oplus 1 \oplus 1$

$= 0 \oplus 0 = 0$

RAID Level 3

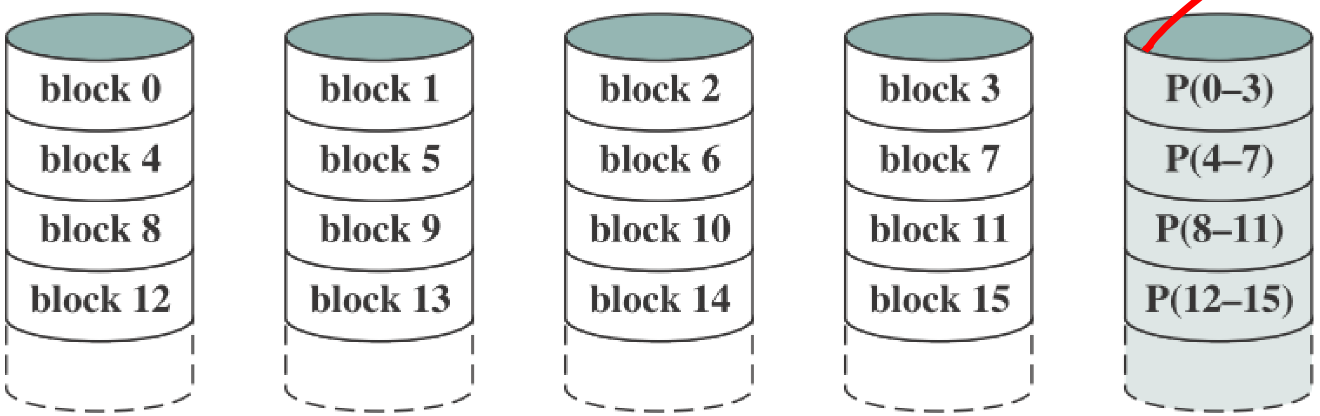
<table>
<tr><th>Level</th><th>Advantages</th><th>Disadvantages</th><th>Applications</th></tr>
<tr><td>3</td><td>Very high read data transfer rate
Very high write data transfer rate
Disk failure has an insignificant impact on throughput
Low ratio of ECC (parity) disks to data disks means high efficiency</td><td>In case of small size files it performs slowly.
Controller design is fairly Complex</td><td>Video production and live streaming
Image editing
Video editing
Prepress applications
Any application requiring high throughput</td></tr>
</table>

RAID Level 4 , 5, 6

Raid 3 → byte level striping
= Raid 4 → block level

- RAID levels 4 through 6 make use of an independent access technique
- Each member disk operates independently, so that separate I/O requests run in parallel

4



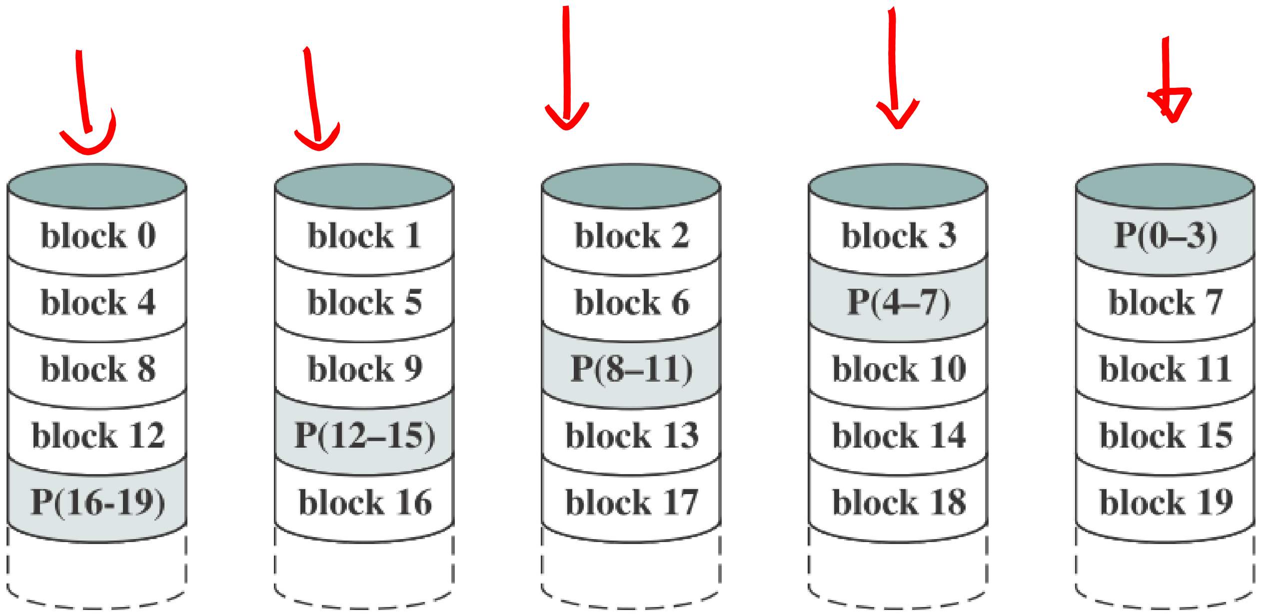
(e) RAID 4 (Block-level parity)

RAID Level 4

Level	Advantages	Disadvantages	Applications
4	• Low ratio of ECC (parity) disks to data disks means high efficiency • Very high Read data transaction rate	• Worst write transaction rate and Write aggregate transfer rate • Quite complex controller design • Difficult and inefficient data rebuild in the event of disk failure	• No commercial implementations exist/not commercially viable

RAID Level 5

- RAID 5 is organized in a similar fashion to RAID 4.
- The difference is that RAID 5 distributes the parity strips across all disks



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

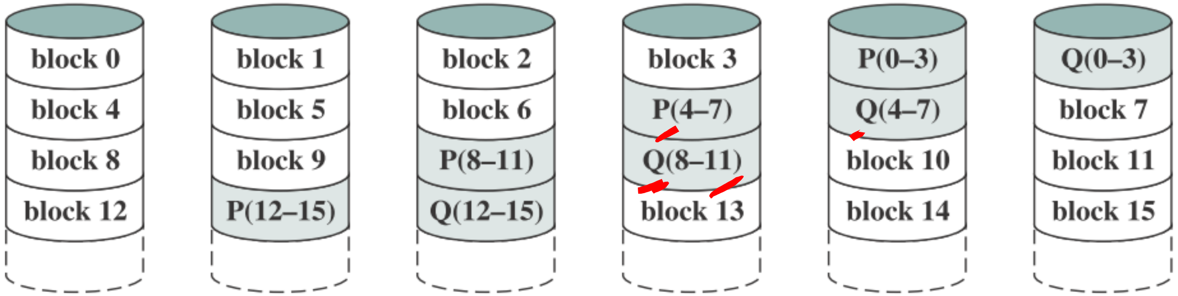
RAID Level 5

Level	Advantages	Disadvantages	Applications
5	• Highest Read data transaction rate • Low ratio of ECC (parity) disks to data disks means high efficiency • Better reliability	• 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

RAID Level 6

- Two different parity calculations are carried out and stored in separate blocks on different disks
- RAID 6 array whose user data require N disks consists of N + 2 disks
- One of the two is the exclusive- OR calculation used in RAID 4 and 5
- Other is an independent data check algorithm
- Possible to regenerate data even if two disks containing user data fail

[Reed soloman algorithm



(g) RAID 6 (Dual redundancy)

RAID Level 6

4 - data disk

+

Level	Advantages	Disadvantages	Applications
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 is extremely high	• Perfect solution for mission critical applications

Important Notice

There will be class at 11.00am to 1 pm tomorrow.

Assignment 1 UP

Solid state drives

flash memory ⇓ EEPROM

- SSDs have the following advantages over HDDs:
  - High- performance input/output operations per second (IOPS): Significantly increases performance I/O subsystems.
  - Lower access times and latency rates: Over 10 times faster than the spinning disks in an HDD.
  - Lower power consumption: SSDs use considerably less power than comparable- size HDDs.
  - Durability: Less susceptible to physical shock and vibration.
  - Longer lifespan: SSDs are not susceptible to mechanical wear.
  - Quieter and cooler running capabilities: Less space required, lower energy costs, and a greener enterprise.

Solid State Disks (SSDs)

- Pages: 512KB to 4KB, Blocks: 32 to 128 pages
- Data read/written in units of pages.
- Page can be written only after its block has been erased
- A block wears out after about 100,000 repeated writes

$\underbrace{B_0 \text{ to } B-1}_{B \text{ number of block}}$

P => pages

Size = 512B - 4KB

pages in a block : 32 - 128 pages

Block 0

Page 0 | Page 1 | . . . | Page P-1

. . .

Block B-1

Page 0 | Page 1 | . . . | Page P-1

SSD Organization

- SSD contains the following components:
  - Controller: Provides SSD device level interfacing and firmware execution.
  - Addressing: Logic that performs the selection function across the flash memory components.
  - Data buffer/cache: High speed RAM memory components used for speed matching and to increased data throughput.
  - Error correction: Logic for error detection and correction.
  - Flash memory components: Individual NAND flash chips.

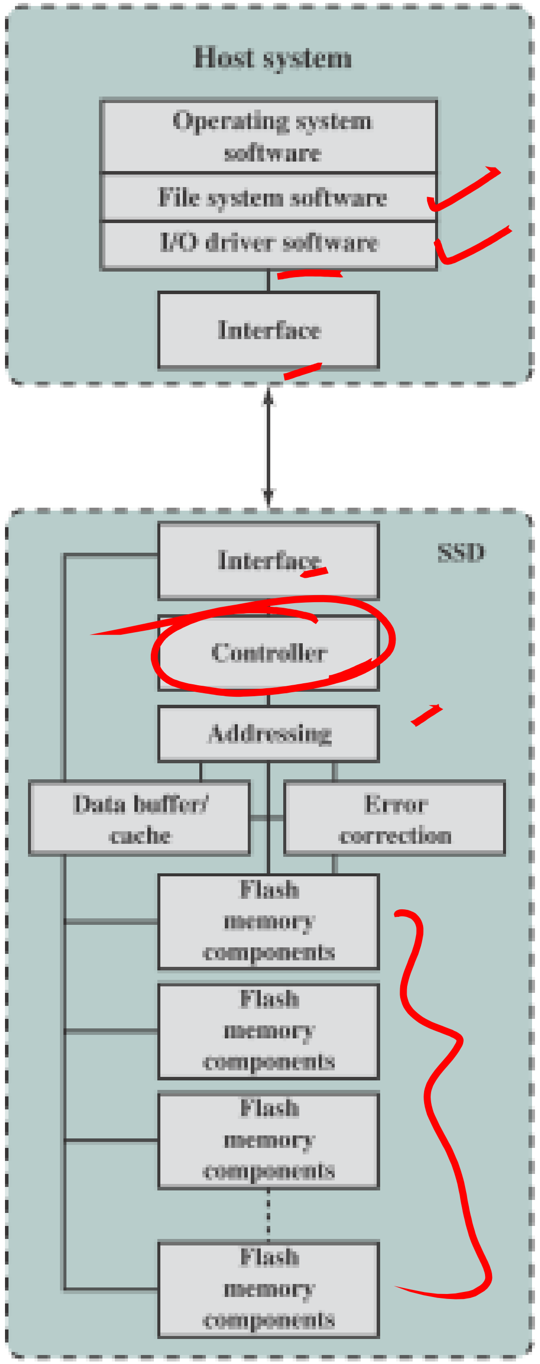


Figure 6.8 Solid State Drive Architecture

Issues with SSD

- Two practical issues:
  - SSD performance has a tendency to slow down as the device is used
  - Memory becomes unusable after a certain number of writes

First Issue

- Files are stored on disk as a set of pages, typically 4 KB in length
- Pages are not necessarily stored as a contiguous set of pages on the disk
- Flash memory is accessed in blocks, block size of 512 KB, (128 pages per block)
- Write a page onto a flash memory:

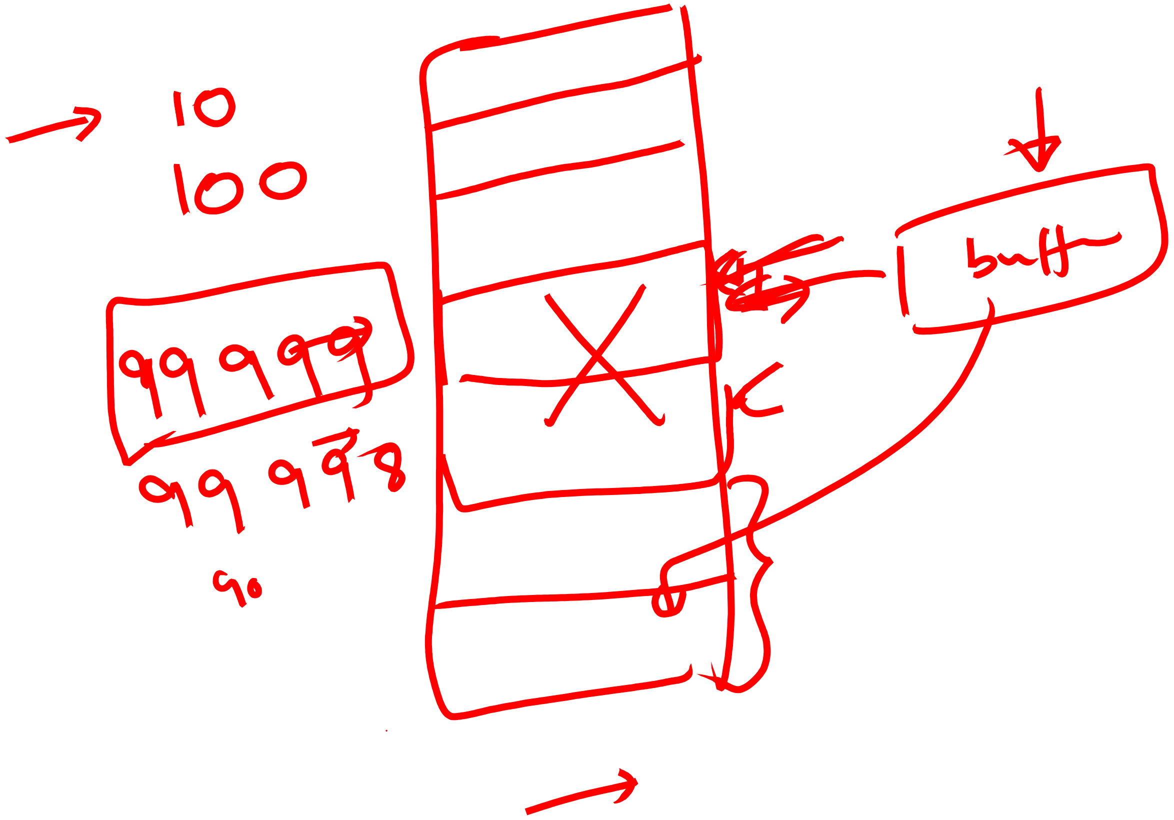
  1. The entire block must be read from the flash memory and placed in a RAM buffer. Then the appropriate page in the RAM buffer is updated.

  2. Before the block can be written back to flash memory, the entire block of flash memory must be erased— it is not possible to erase just one page of the flash memory.

  3. The entire block from the buffer is now written back to the flash memory.

Second Issue

- Flash memory becomes unusable after a certain number of writes.
- Typical limit is 100,000 writes



Comparison of Solid State Drives and Disk Drives

	NAND Flash Drives	Seagate Laptop Internal HDD
File copy/write speed	200–550 Mbps	50–120 Mbps
Power draw/battery life	Less power draw, averages 2–3 watts, resulting in 30+ minute battery boost	More power draw, averages 6–7 watts and therefore uses more battery
Storage capacity	Typically not larger than 512 GB for notebook size drives; 1 TB max for Desktops	Typically around 500 GB and 2 TB max for notebook size drives; 4 TB max for desktops
Cost	Approx. $0.50 per GB for a 1-TB drive	Approx. $0.15 per GB for a 4-TB drive