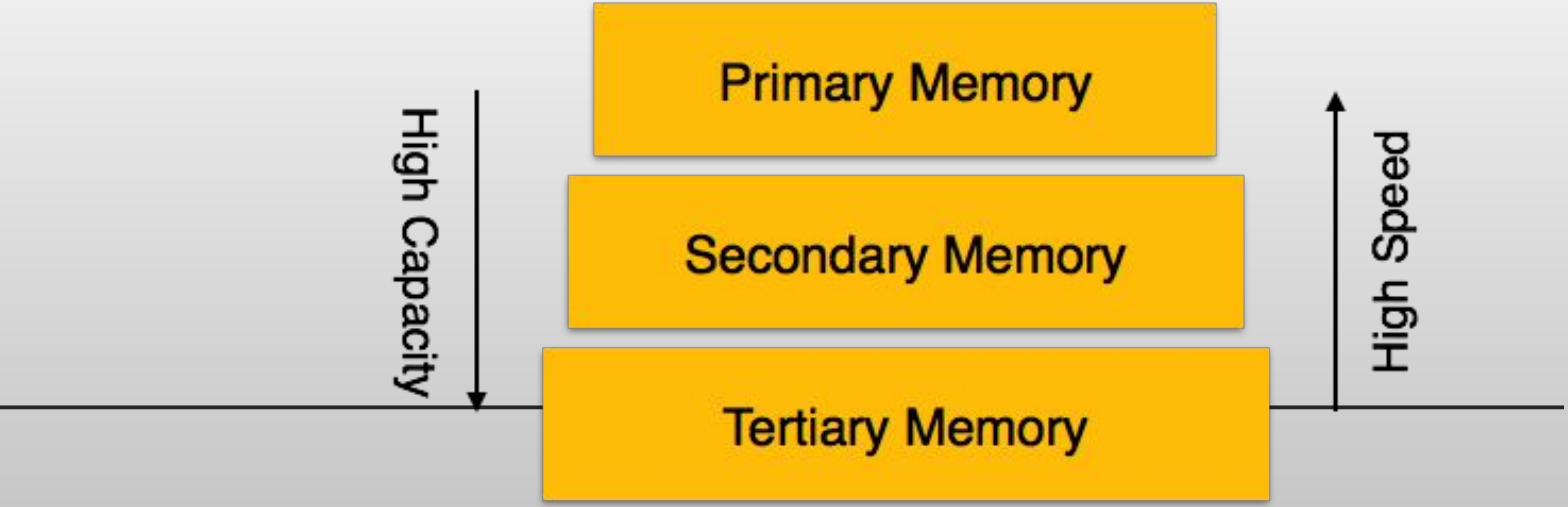
DISKS & STORAGE

DATABASE MANAGEMENT SYSTEM

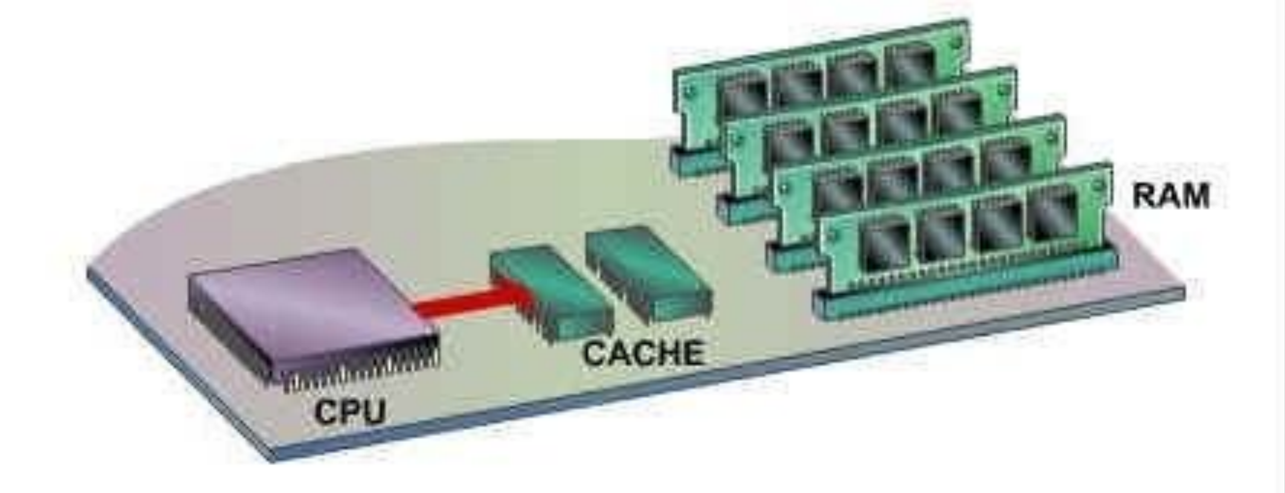
Sujan Tamrakar

- Storage media are classified by
  - speed with which data can be accessed,
  - cost per unit of data to buy the medium and
  - medium’s reliability.
- At physical level, the actual data is stored in electromagnetic format on some device. These storage devices can be broadly categorized into three types:



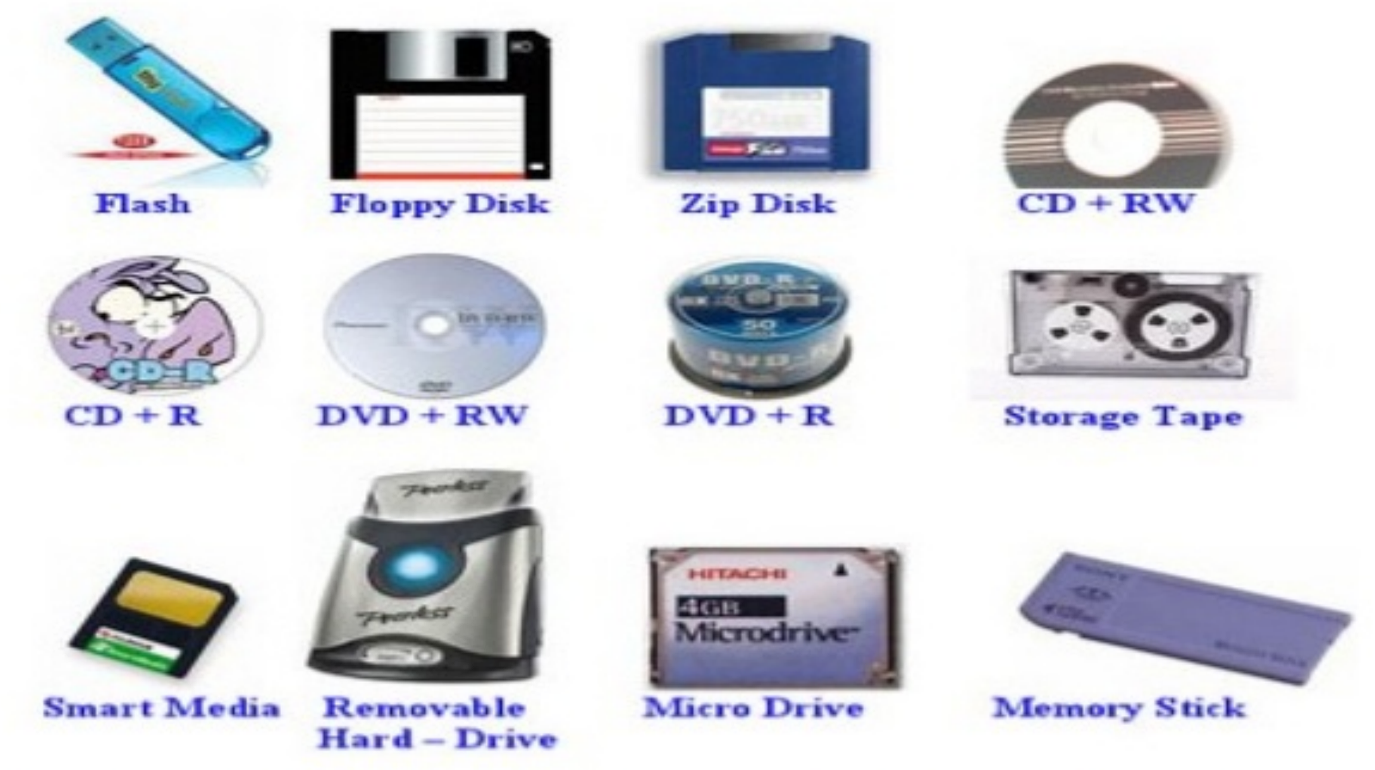
❑ Primary Storage

- The memory storage that is directly accessible to the CPU comes under this category.
- CPU's internal memory (registers), fast memory (cache), and main memory (RAM) are directly accessible to the CPU, as they are all placed on the motherboard or CPU chipset. This storage is typically very small, ultra-fast, and volatile.
- Primary storage requires continuous power supply in order to maintain its state. In case of a power failure, all its data is lost.



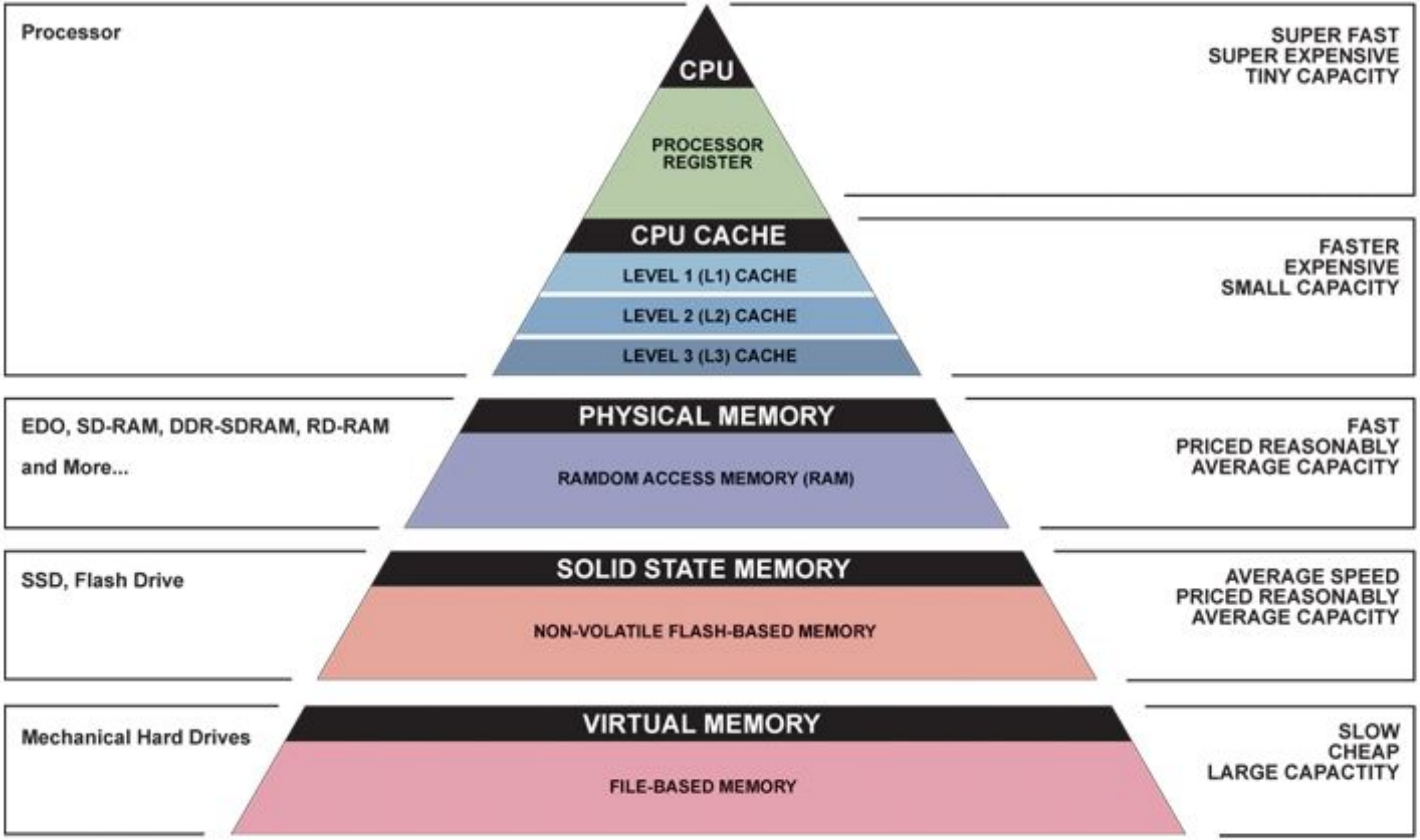
❑ Secondary Storage

- Secondary storage devices are used to store data for future use or as backup.
- Secondary storage includes memory devices that are not a part of the CPU chipset or motherboard, for example, magnetic disks, optical disks (DVD, CD, etc.), hard disks, flash drives, and magnetic tapes.



❑ Tertiary Storage

- Tertiary storage is used to store huge volumes of data.
- Since such storage devices are external to the computer system, they are the slowest in speed.
- These storage devices are mostly used to take the back up of an entire system. Optical disks and magnetic tapes are widely used as tertiary storage.

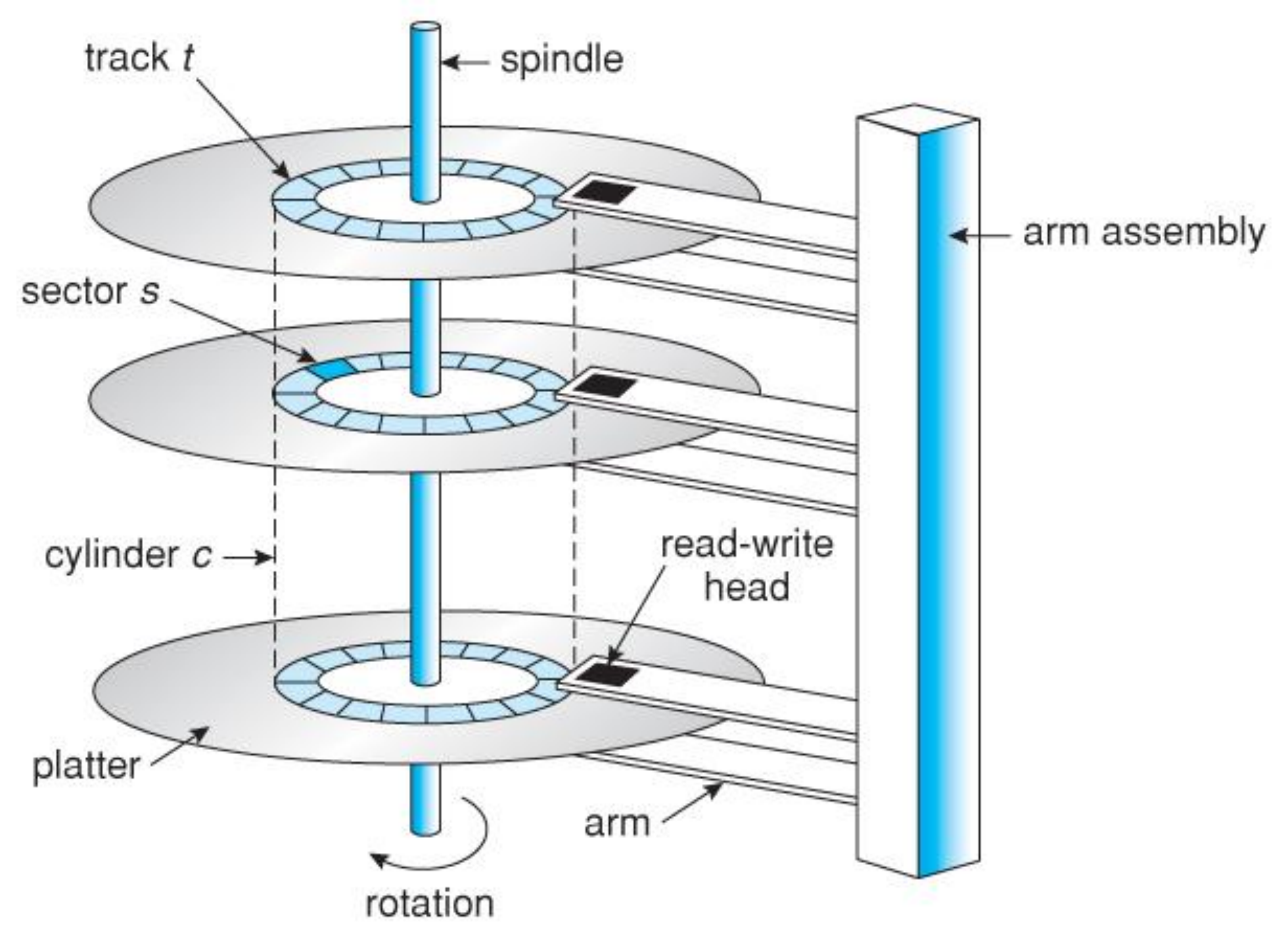


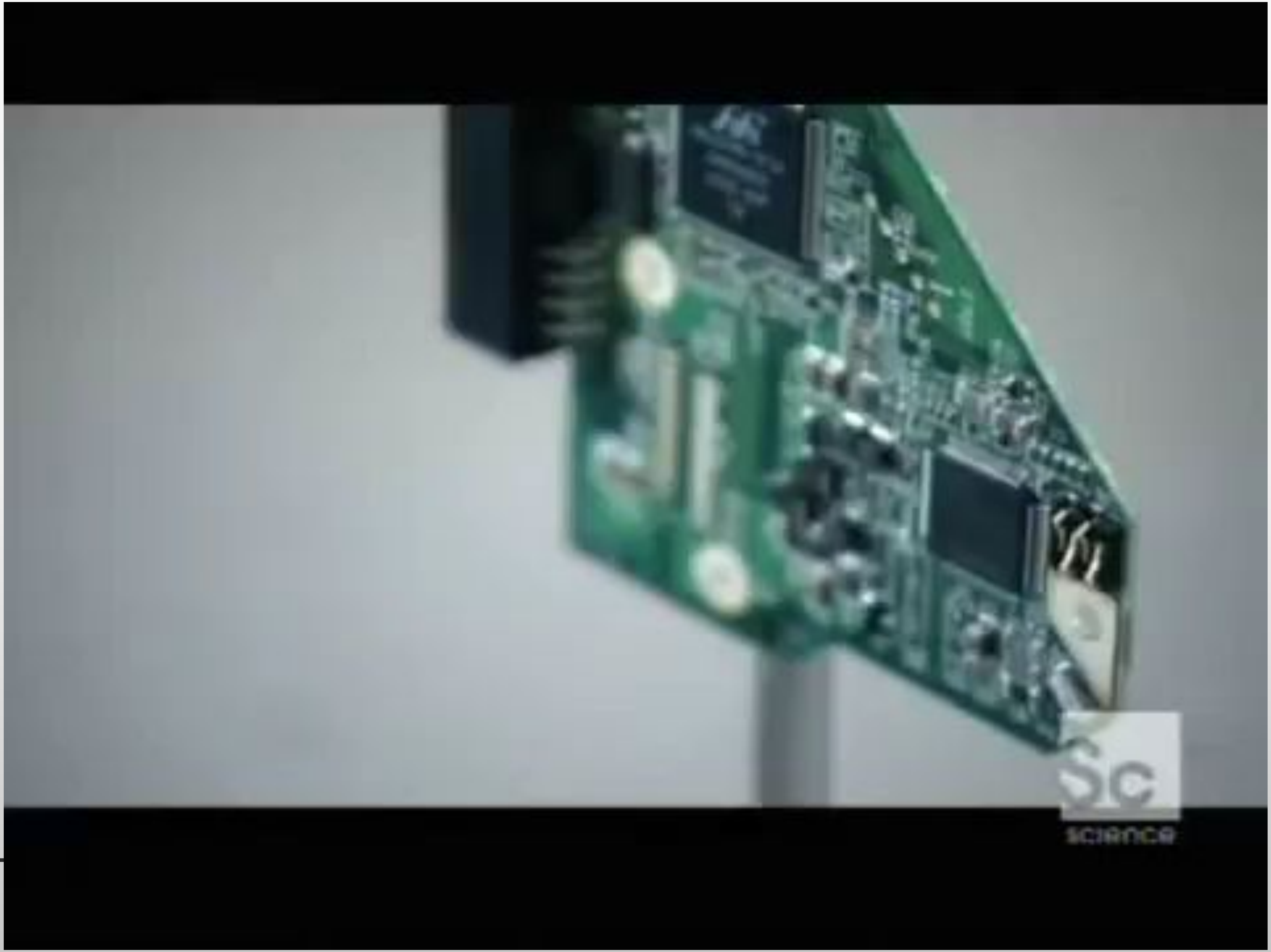
▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

Magnetic Disks

- Hard disk drives are the most common secondary storage devices in present computer systems.
- These are called magnetic disks because they use the concept of magnetization to store information.
- Hard disks consist of metal disks coated with magnetisable material.
- These disks are placed vertically on a spindle.
- A read/write head moves in between the disks and is used to magnetize or de-magnetize the spot under it.
- A magnetized spot can be recognized as 0 (zero) or 1 (one).
- Hard disks are formatted in a well-defined order to store data efficiently. A hard disk plate has many concentric circles on it, called tracks.
- Every track is further divided into sectors. A sector on a hard disk typically stores 512 bytes of data.

Magnetic Disks



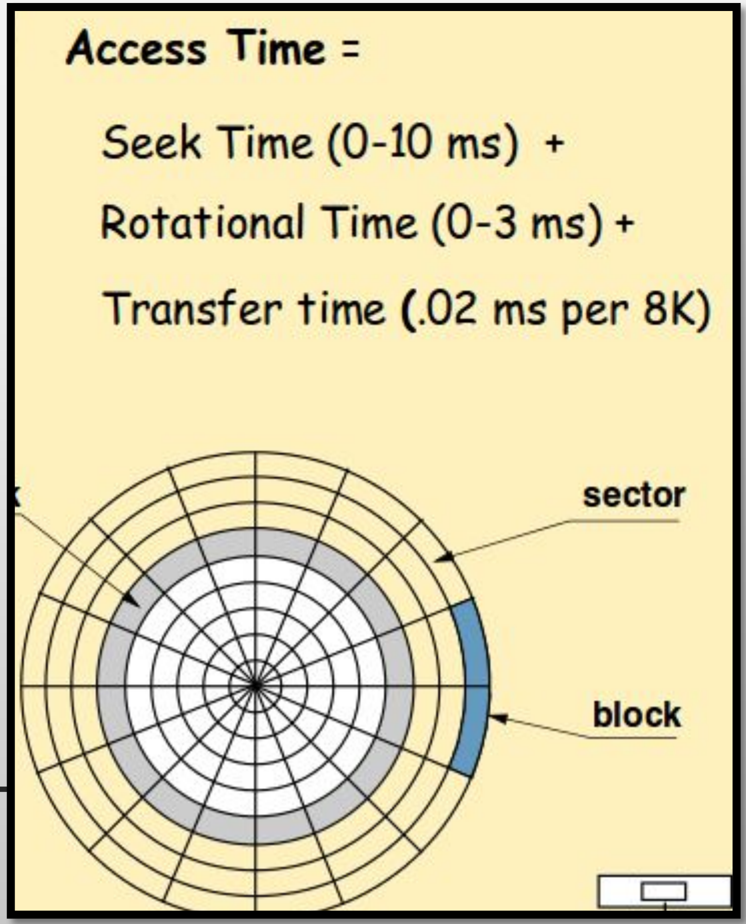

Sc
science

Performance measures of Magnetic Disks

- Access time is the time from when a read or write request is issued to when data transfer begins.
- To access data on given sector of a disk, the arm must first move so that it is positioned over the correct track & must wait for sector to appear under it as disk rotates.
- Time for repositioning the arm is called seek time & it increases with distance that the arm must move.
- Smaller disk (less diameter of platter) has lower seek times since head has to travel less distance.
- Once head reaches the desired track, the time spent waiting for the sector to be accessed to appear under the head is called rotational latency time.
- The data transfer rate is the rate at which data can be retrieved from or stored to the disk.
- Mean time to failure (MTTF) is a measure of the reliability of the disk. It is the amount of time that on average we can expect the system to run continuously without any failure.

Optimization of Disk Block Access

- Each request in system specifies the address on the disk to be referenced; that address is in the form of a block number.
- A block is a logical unit consisting of a fixed number of contiguous (together in sequence) sectors.
- Block sizes range from 512 bytes to several KB.
- Data are transferred between disk & main memory in units of blocks.
- Access to data in disk is slower than in main memory so, number of techniques have been developed for increasing speed to access the block in disk.



Access Time =

Seek Time (0-10 ms) +

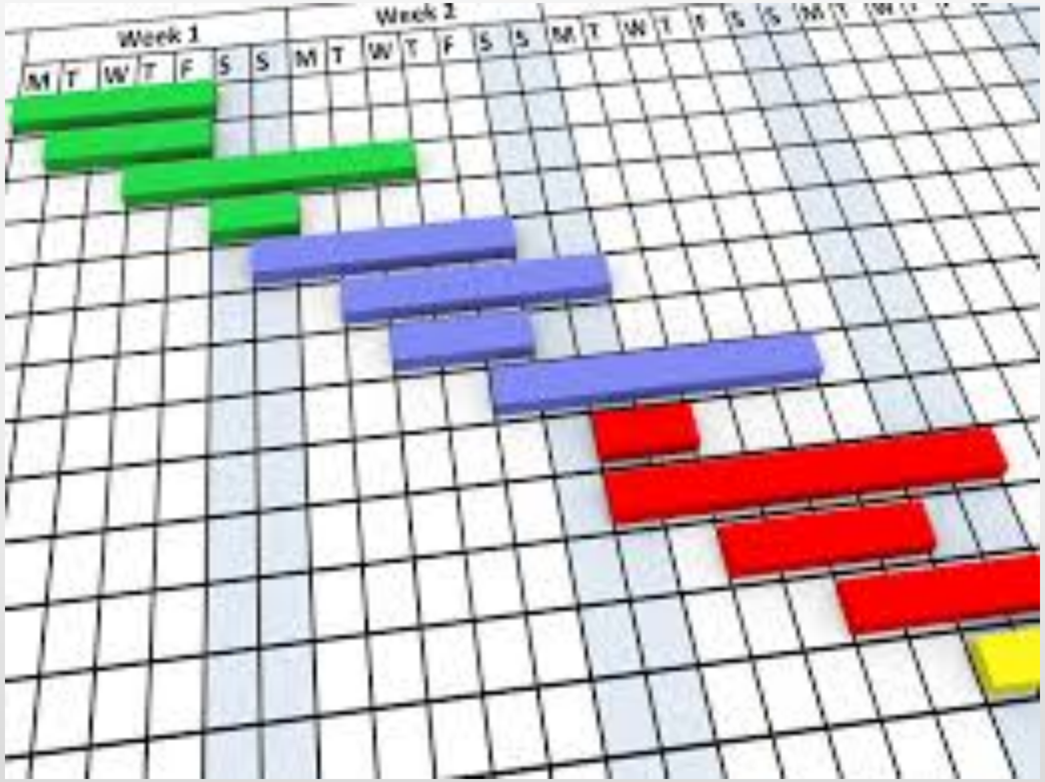
Rotational Time (0-3 ms) +

Transfer time (.02 ms per 8K)

Optimization of Disk Block Access

1. Scheduling:

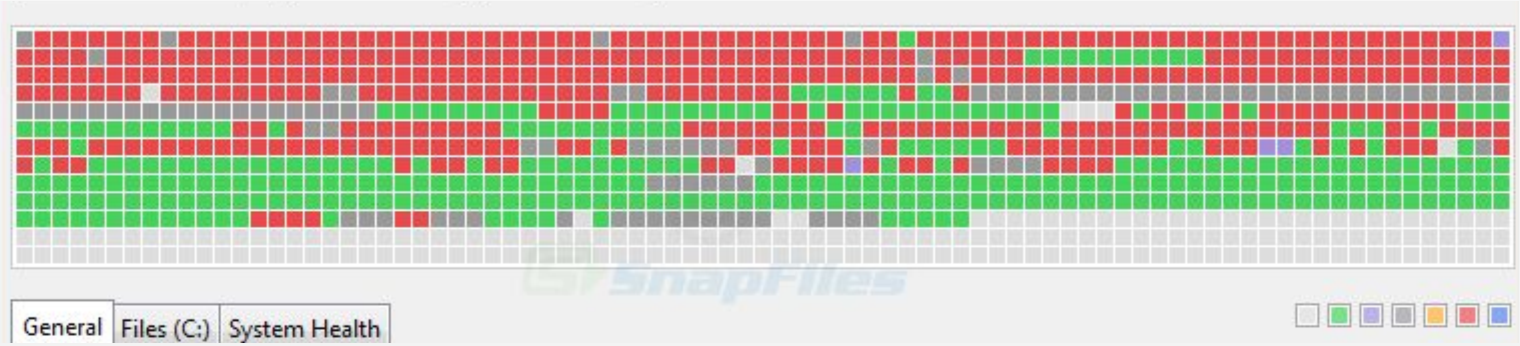
- May be able to save access time by requesting blocks in order in which they will pass under the heads. If all blocks are in same cylinder, then time is saved.
- If desired blocks are in different cylinder, it is better to request blocks in such an order that minimizes disk-arm movement.
- Elevator algorithm is used (concept of elevator/lift)
    - Arm moving from innermost track to outermost track
    - Finishes the service (read/write) in inner track & then moves to outer track until all desired tracks are visited.
    - Now arm changes direction & moves toward inside doing same operation of executing service. Then, it reverses direction & starts a new cycle.
- Disk controller performs tasks of re-ordering read requests to improve performance.

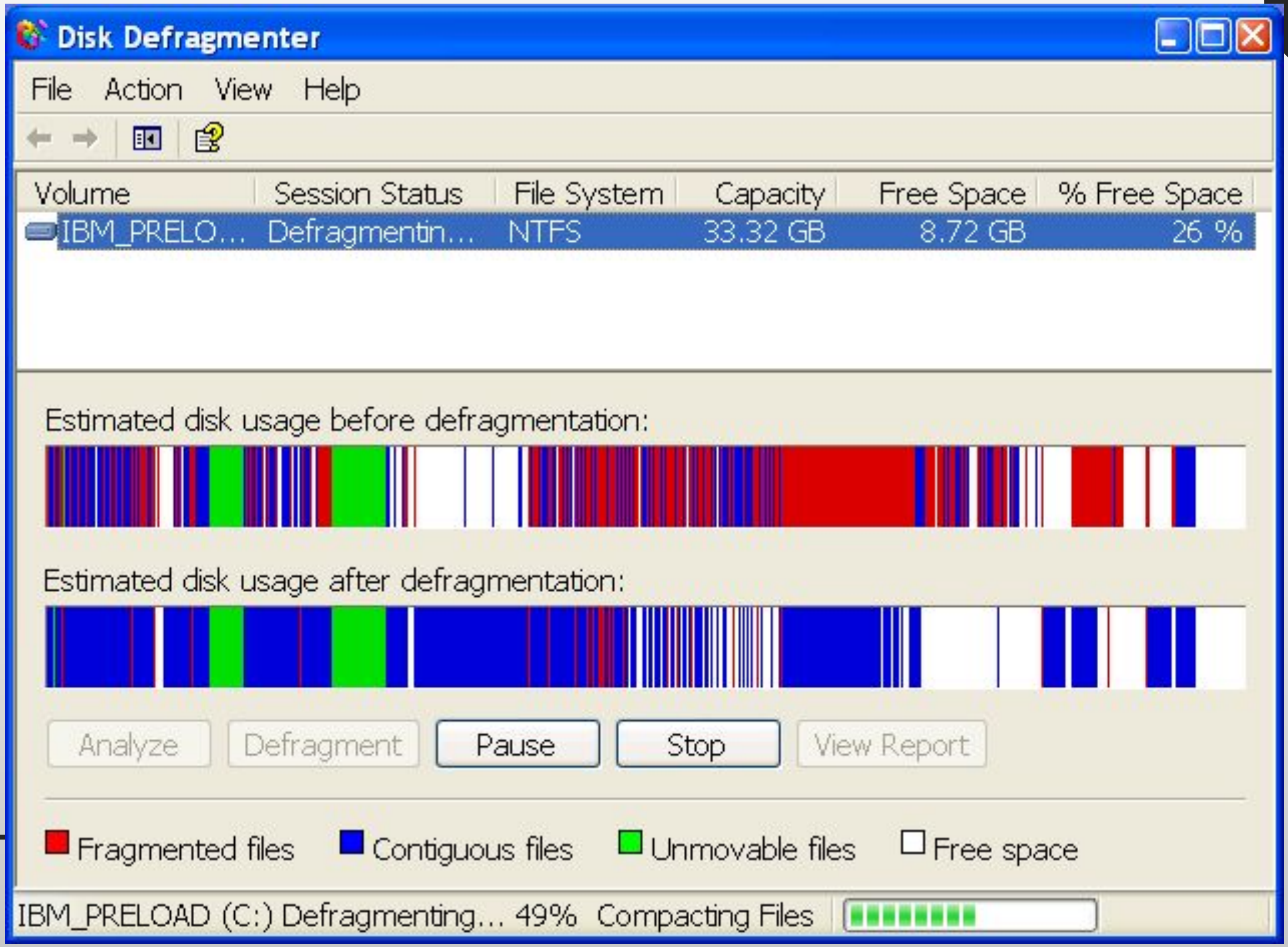

Week 1
Week 2
M
T
W
T
F
S
S
M
T
W
T
F
S
S
M
T
W
T
F
S
S

Optimization of Disk Block Access

2. File Organization/Defragmentation:

    - To reduce block access time, we can organize blocks on disk in a way that corresponds closely to the way we expect data to be accessed.
    - If we expect a file to be accessed sequentially then we should ideally keep all the blocks of file sequentially on adjacent cylinders.
    - Files may stay scattered in a disk (fragmented).
    - To reduce fragmentation, system can make a backup copy of data on disk & restore entire disk. Restore operation writes back the blocks of each file contiguously.
    - Some systems have utilities to scan the disk & move blocks to decrease fragmentation.





Volume	Session Status	File System	Capacity	Free Space	% Free Space
IBM_PRELO...	Defragmentin...	NTFS	33.32 GB	8.72 GB	26 %

Estimated disk usage before defragmentation:

Estimated disk usage after defragmentation:

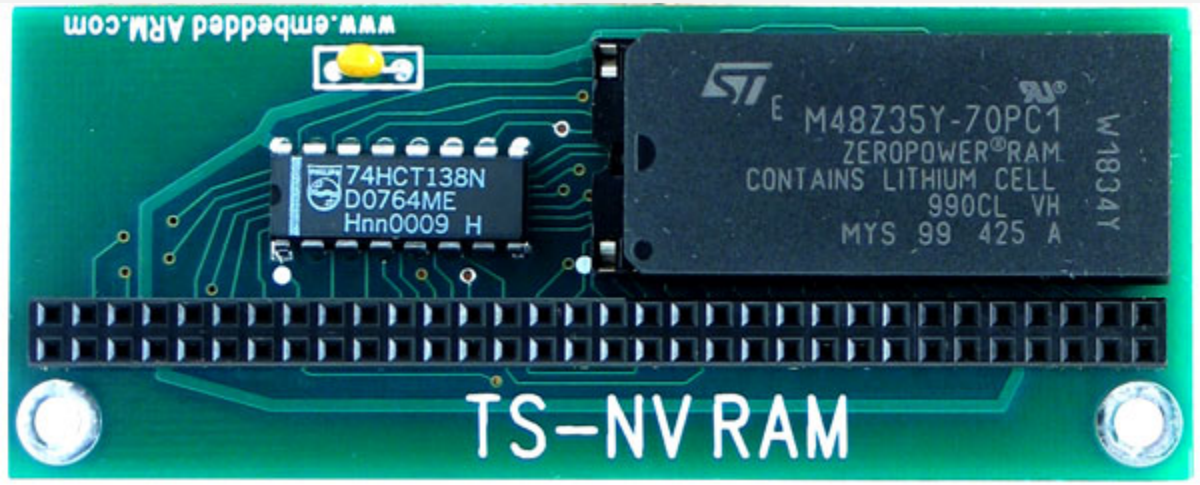
🟥 Fragmented files 🟦 Contiguous files 🟩 Unmovable files □ Free space

IBM_PRELOAD (C:) Defragmenting... 49% Compacting Files

Optimization of Disk Block Access

3. Non-volatile write buffers:

   - Use of NVRAM to speed up disk writes drastically.
   - Contents are not lost in power failure.
   - Common way to implement NVRAM is to use battery backed-up RAM.
   - When write command is issued, disk controller writes block to NVRAM & notifies to OS about successful writing. When disk is free of operation or NVRAM buffer is full, disk controller writes the block to the disk.
   - On recovery from system crash, any pending buffered writes in NVRAM are written to disk.



www.embeddedARM.com
74HCT138N
D0764ME
Hnn0009 H
M48Z35Y-70PC1
ZEROPOWER®RAM
CONTAINS LITHIUM CELL
990CL VH
MYS 99 425 A
W1834Y
TS–NV RAM

Optimization of Disk Block Access

4. Log disk:

- A disk devoted to writing a sequential log.
- Access to log disk is sequential, eliminating seek time. Several consecutive blocks can be written at once, making writing to log disk faster than random writes.
- Log disk can write to disk later without the system having to wait for real write in hard disk.
- Log disk can re-order writes to minimize disk-ark movement
- On system crash while or before writing to hard disk- after system recovers, it reads the disk log for unwritten blocks & writes to hard disk.
- File system that support log disks are called Journaling File System.

THANK YOU!