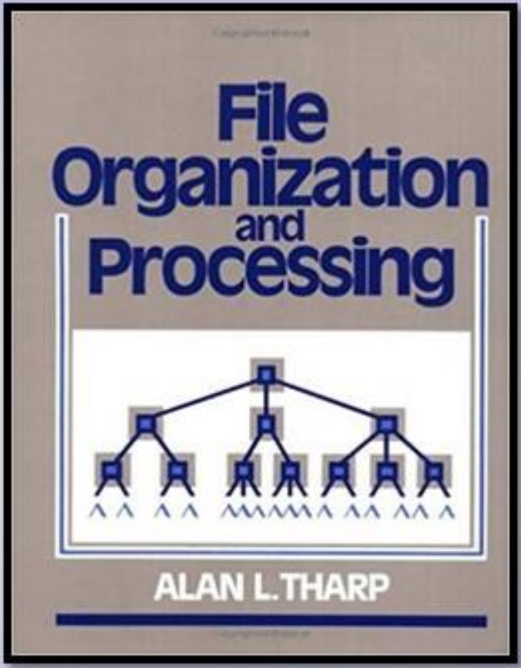
File Organization & Processing

CS2202

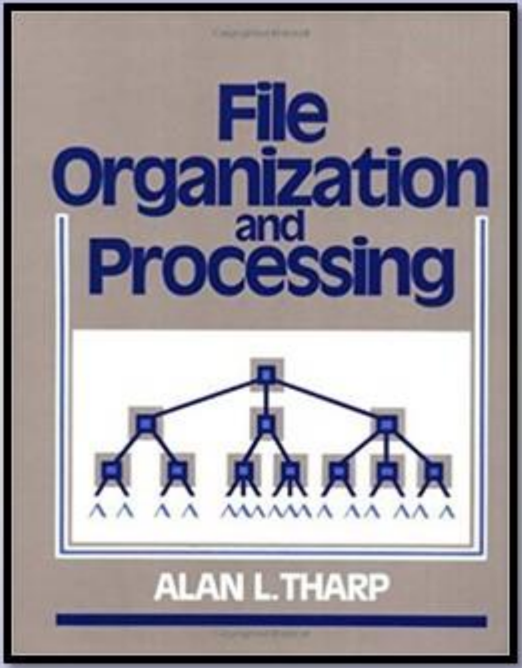
Instructor:

Dr. Mohamed Hassan



Chapter 2:

- Secondary Storage



Secondary Storage Devices

- Two major types of secondary storage devices:

1. **Direct Access Storage Devices (DASDs)**
   - **Magnetic Discs**
     Hard disks (high capacity, low cost, fast)
     Floppy disks (low capacity, lower cost, slow)
   - **Optical Disks**
     CD-ROM = (Compact disc, read-only memory
2. **Serial Devices**
   - **Magnetic tapes** (very fast sequential access)

Storage and Files

- **Storage has major implications for DBMS design!**
  - **READ**: transfer data from disk to main memory (RAM).
  - **WRITE**: transfer data from RAM to disk.
  - Both operations are **high-cost** operations, relative to in-memory operations, so DB must be planned carefully!

Storage and Files

- **Why Not Store Everything in Main Memory?**
  - **Costs** too much: Cost of RAM about 100 times the cost of the same amount of disk space, so relatively small size.
  - Main memory is **volatile**.
  - **Typical storage hierarchy:**
    - **Main memory** (RAM) (**primary** storage) **for currently used data.**
    - **Disk** for **the main database** (**secondary** storage).
    - **Tapes** for **archiving older versions** of the data (**tertiary** storage).

Storage Hierarchy

- **Primary storage** : random access memory (RAM)
  - typical **capacity** a number of **GB**
  - **cost** per **MB $2-3.00**
  - typical **access time 5ns to 60ns**
- **Secondary storage**: magnetic disk/ optical devices/ tape systems
  - typical **capacity** a number of **100GB** for fixed media; ∞ for removable
  - **cost** per **MB $0.01** for fixed media, more for removable
  - typical **access time** 8ms to **12ms** for fixed media, larger for removable

Units of Measurement

Spatial units:

- **byte**: 8 bits
- **kilobyte** (KB): 1024 or 2^{10} bytes
- **megabyte** (MB): 1024 kilobytes or 2^{20} bytes
- **gigabyte** (GB): 1024 megabytes or 2^{30} bytes

Time units:

- **nanosecond** (ns) one- billionth (10^{-9}) of a second
- **microsecond** (μs) one- millionth (10^{-6}) of a second
- **millisecond** (ms) one- thousandth (10^{-3}) of a second

Primary versus Secondary Storage

- Primary storage costs several hundred times as much per unit as secondary storage,
- but has access times that are 250,000 to 1,000,000 times faster than secondary storage.

Memory Hierarchy

- At the **primary storage level**, the memory hierarchy includes, at **the most expensive** end' **cache memory**, which is a static RAM (Random Access Memory).
- The **next level of primary storage** is **DRAM (Dynamic RAM),** The advantage of DRAM is its **low cost, lower speed compared with static RAM.**
- Programs normally reside and execute in DRAM.

Memory Hierarchy-flash memory

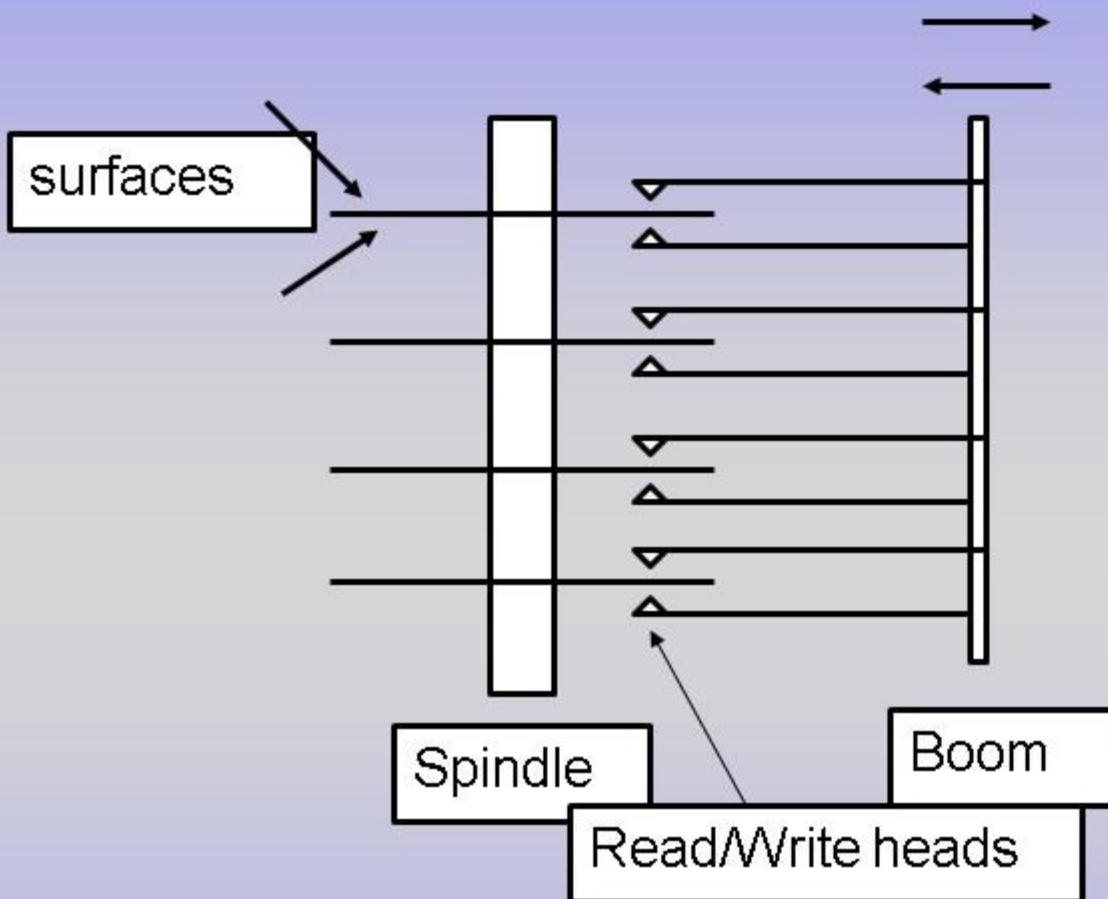
Flash memory,

- Since 1988 it has become common, particularly because it is **nonvolatile,** using **EEPROM** (**Electrically Erasable Programmable Read-Only Memory**) technology.
- Its life is **10,000-1,000,000 times erase**…
- **Read/write** is fast, but **erase** is slow…
- Therefore special arrangements are made for the file system, regarding file delete or update.
- **Capacities** up to **128 GB** has been realized to date.

Magnetic Disks

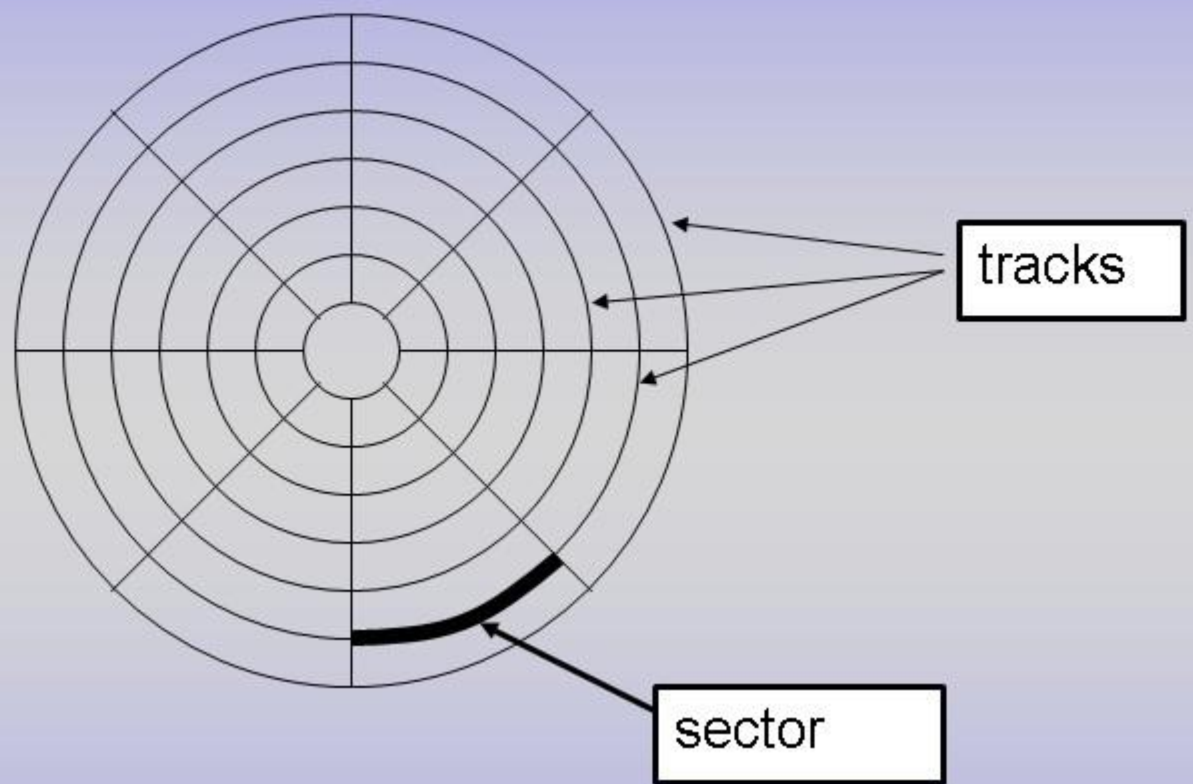
- Bits of data (0's and 1's) are stored on **circular magnetic platters** called **disks**.
- A disk **rotates rapidly** (& **never stops**).
- A **disk head** reads and writes bits of data as they pass under the head.
- Often, **several platters are organized** into a **disk pack (or disk drive).**

A Disk Drive



Disk drive with 4 platters and 8 surfaces and 8 RW heads

Looking at a surface



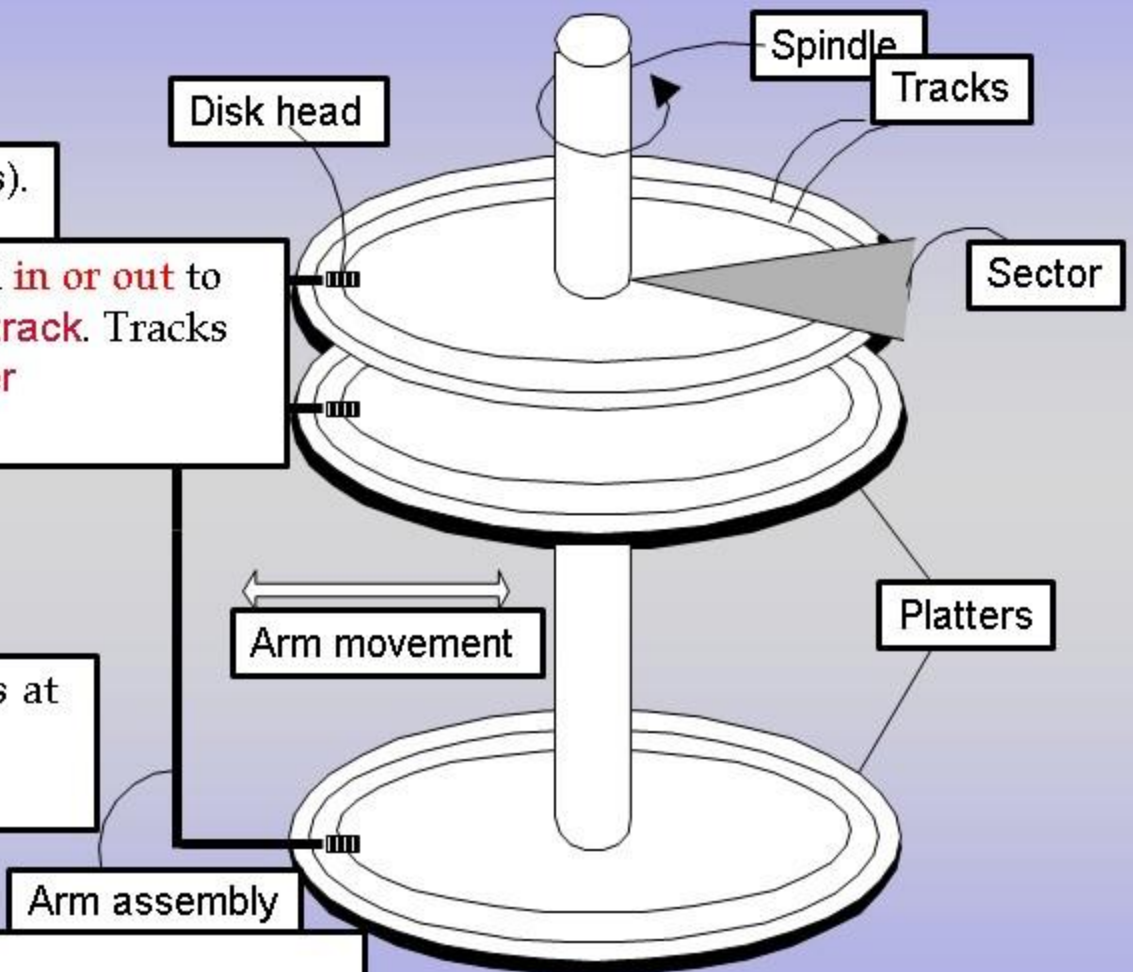
Surface of disk showing tracks and sectors

Organization of Disks

- Disk contains concentric **tracks**.
- Tracks are divided into **sectors**
- A **sector** is the smallest addressable unit in a disk.

Components of a Disk

- The platters spin (say, 90rps).
- The arm assembly is moved in or out to position a head on a desired track. Tracks under heads make a cylinder (imaginary!).
- Only one head reads/writes at any one time.



- Block size is a multiple of sector size (which is often fixed).

Disk Controller

- Disk controllers: typically embedded in the disk drive, which **acts as an interface between the CPU and the disk hardware**.

- The controller has **an internal cache** (typically a number of MBs) that it uses to **buffer data for read/write requests.**

Accessing Data

- When **a program reads a byte from the disk**, **the operating system locates the surface, track and sector containing that byte**, and **reads the entire sector into a special area in main memory** called **buffer**.

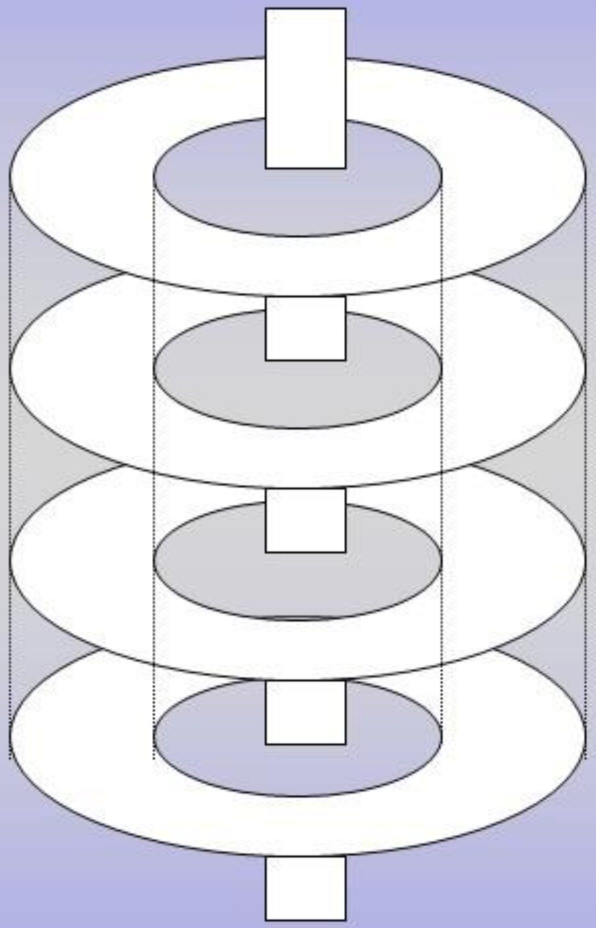
- The **bottleneck** of a disk access is **moving the read/write arm.**

- So it makes sense to **store a file in tracks that are below/above each other on different surfaces**, **rather than in several tracks on the same surface.**

Cylinders

- A **cylinder** **is the set of tracks at a given radius of a disk pack.**
  - i.e. a cylinder is the set of tracks that can be accessed without moving the disk arm.
- All the information on a cylinder can be accessed without moving the read/write arm.

Cylinders



Estimating Capacities

- **Track capacity** = # of sectors/track * bytes/sector
- **Cylinder capacity** = # of tracks/cylinder * track capacity
- **Drive capacity** = # of cylinders * cylinder capacity
- **Number of cylinders** = # of tracks in a surface

Exercise

- Store a file of 20000 records on a disk with the following characteristics:

  # of bytes per sector = 512

  # of sectors per track = 40

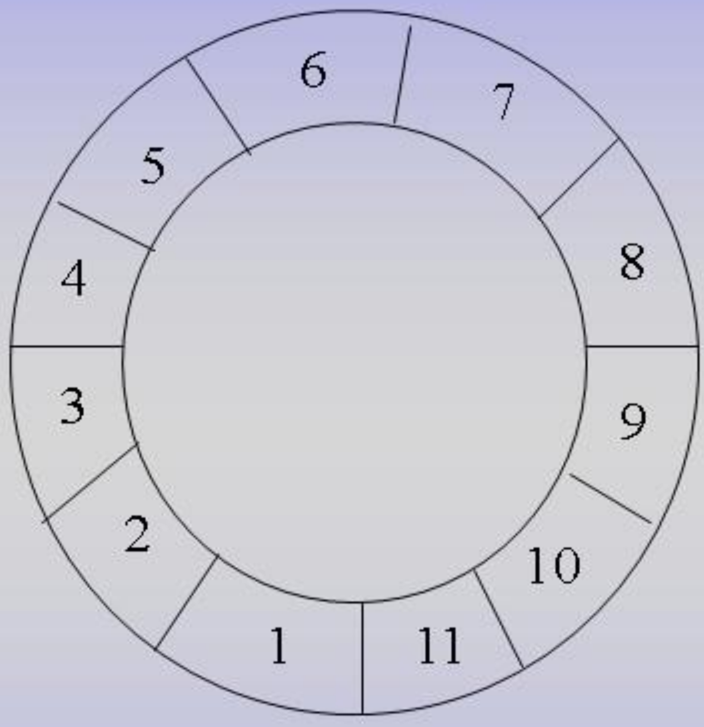
  # of tracks per cylinder = 11

  # of cylinders = 1331

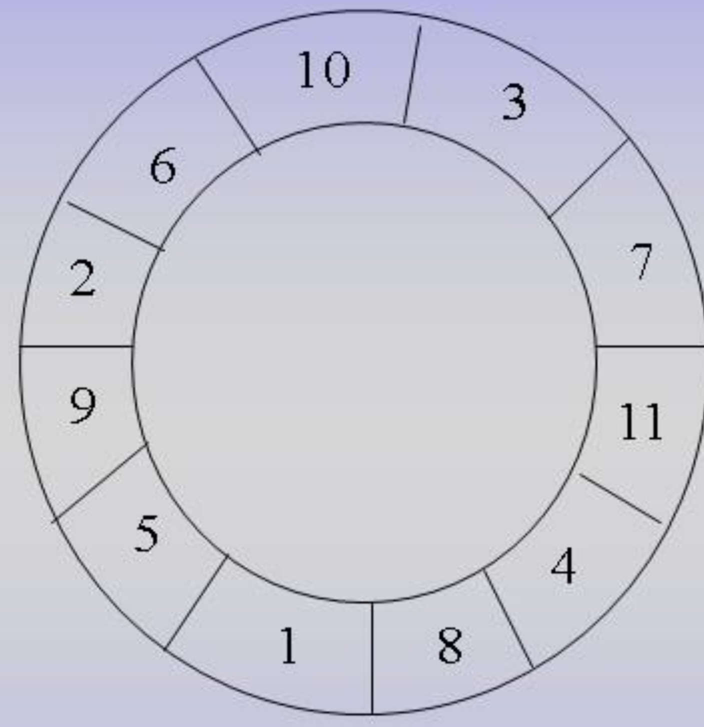
Q1. How many cylinders does the file require if each data record requires 256 bytes?

Q2. What is the total capacity of the disk?

Organizing Tracks by sector



Physically adjacent sectors



Sectors with 3:1 interleaving

Clusters

- Usually **File manager**, under the operating system, **maintains the logical view of a file.**
- **File manager views the file as a series of clusters**, each of a **number of sectors**. The clusters are **ordered by their logical order.**
- **Files** can be seen **in the form of logical sectors or blocks**, which **needs to be mapped** to physical clusters.
- **File manager** uses a **file allocation table (FAT)** to **map logical sectors of the file to the physical clusters.**

Extents

- If there is a lot of room on a disk, it may be possible to make **a file consist entirely of contiguous clusters**. Then we say that the file is one **extent**. (very good for sequential processing)
- If there isn't enough contiguous space available to contain an entire file, the file is divided **into two or more noncontiguous** parts. **Each part is a separate extent**.

Internal Fragmentation

- **Internal fragmentation**: loss of space within a sector or a cluster.

1) **Due to records not fitting exactly in a sector**: **e.g.** Sector size is 512 and record size is 300 bytes. Either
   - store one record per sector, or
   - allow records *span* sectors…
2) **Due to the use of clusters**: If the file size is not a multiple of the cluster size, then the last cluster will be partially used.

Organizing Tracks by Block

- **Disk tracks may be divided into user-defined blocks rather than into sectors**.
- Blocks can be fixed or variable length.
- A block is usually organized to hold an integral number of logical records.
- **Blocking Factor** = number of records stored in a block.
- No internal fragmentation, no record spanning over two blocks.
- In block-addressing scheme each block of data may be accompanied by one or more ***subblocks*** containing extra information about the block: record count, last record key on the block…

Non-data Overhead

- **Both blocks and sectors require non-data overhead** (written during formatting)
- **On sector addressable disks**, this information involves **sector address, track address, and condition (usable/defective)**. Also pre-formatting involves placing gaps and synchronization marks between the sectors.
- **On block-organized disk**, **where a block may be of any size**, **more information is needed and the programmer should be aware of some of this information to utilize it for better efficiency**…

The Cost of a Disk Access

- The time to access a sector in a track on a surface is divided into 3 components:

Time Component	Action
Seek Time	Time to move the read/write arm to the correct cylinder
Rotational delay (or latency)	Time it takes for the disk to rotate so that the desired sector is under the read/write head
Transfer time	Once the read/write head is positioned over the data, this is the time it takes for transferring data

Seek time

- **Seek time** is **the time required to move the arm to the correct cylinder**.
- **Largest in cost**.

Typically:

- **5 ms** (miliseconds) to **move from one track to the next (track-to-track)**
- **50 ms maximum** (**from inside track to outside track)**
- **30 ms average** (**from one random track to another random track**)

Latency (Rotational Latency)-1

- **Latency is the time needed for the disk to rotate so the sector we want is under the read/write head**.
- **Hard disks** usually **rotate** at about **5000-7000 rpm**,
  - **12-8** msec per **revolution**.
- Note:
  - **Min latency** = 0
  - **Max latency** = Time for one disk revolution
  - **Average latency (r)** = (min + max) / 2

    = max / 2

    = time for ½ disk revolution
    - **Typically 6 – 4 ms, at average**

Transfer Time-1

- **Transfer time** **is the time for the read/write head to pass over a block**.
- The transfer time is given by the formula:

$$\text{Transfer time} = \frac{\text{number of sectors}}{\text{track capacity in number of sectors}} \times \text{rotation time}$$

  - e.g. if there are S_t sectors per track, the time to transfer one sector would be $1/S_t$ of a revolution.

Transfer Time-2

- The transfer time depends only on the speed at which the spindle rotates, and the number of sectors that must be read.

  - **Sequential Reading**

- Given the following disk:
  - Avg. Seek time s = 16 ms
  - Avg. Rot. Latency r = 8.3 ms
  - Block transfer time = 8.4 ms

a) Calculate the time to read 10 sequential blocks, on the same track.

a) 16+8.3+8.4*10

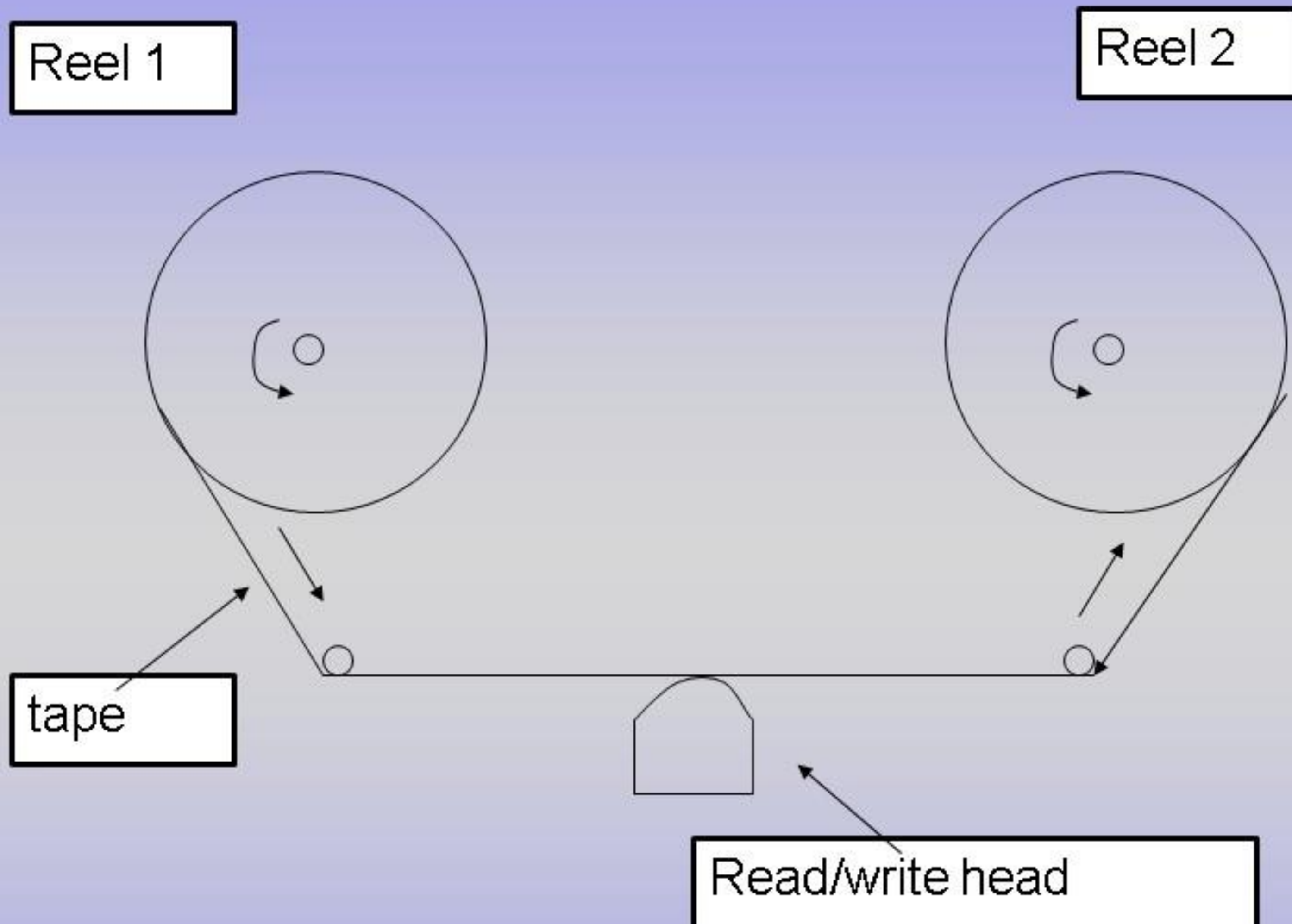
Secondary Storage Devices: Magnetic Tapes

Characteristics

- **No direct access**, but **very fast sequential access**.
- Resistant to different environmental conditions.
- Easy to **transport, store, cheaper than disk**.
- Before it was widely used to store application data; nowadays, **it's mostly used for backups or archives**.

MT Characteristics-2

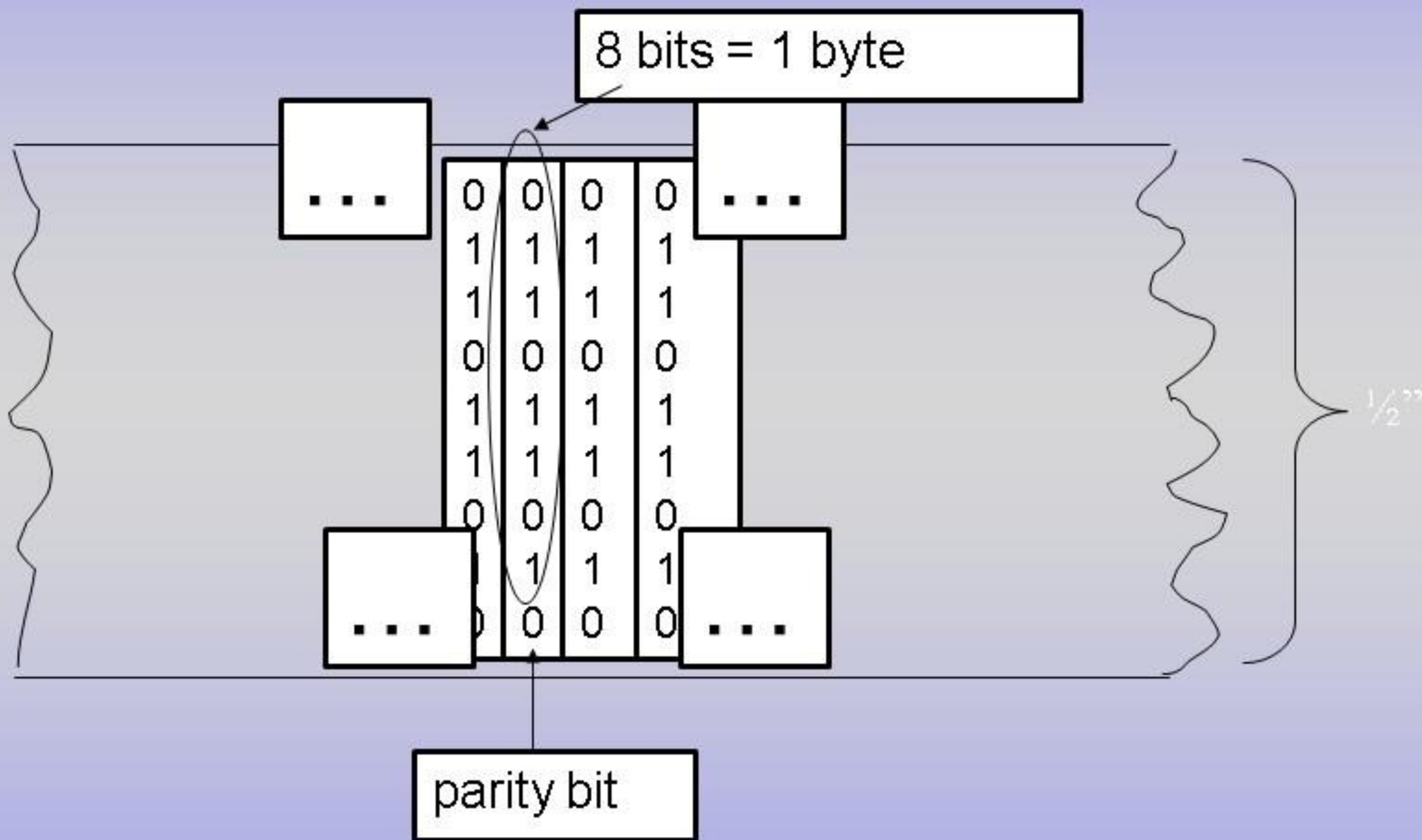
- **A sequence of bits** are **stored on magnetic tape.**
- **For storage**, **the tape is wound on a reel.**
- **To access the data**, the **tape is unwound from one reel to another.**
- **As the tape passes the head**, **bits of data are read from or written onto the tape.**


Reel 1
Reel 2
tape
Read/write head

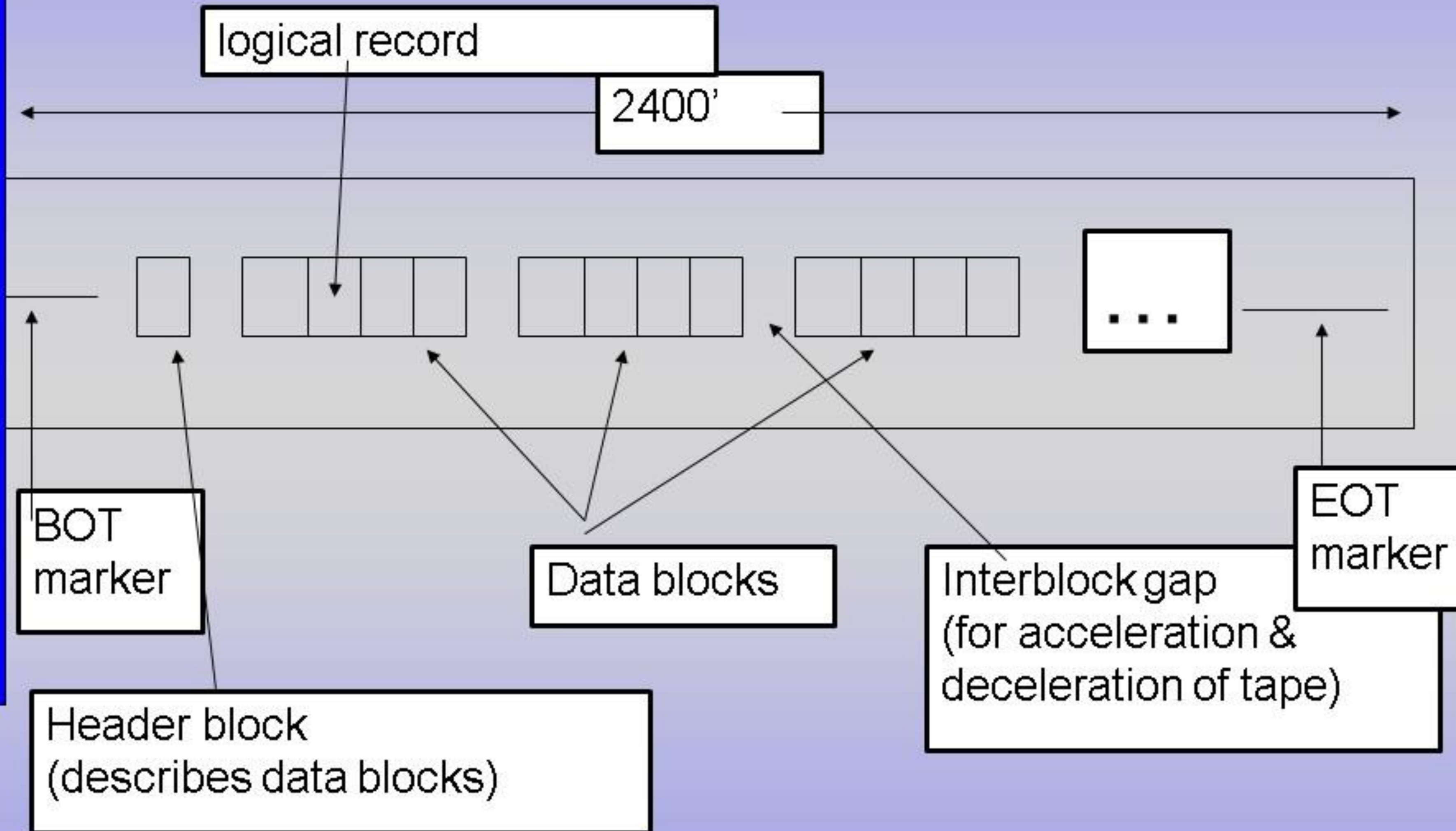
Tracks

- Typically data on tape is **stored** in **9 separate bit streams**, or **<u>tracks</u>**.
- Each track is a sequence of bits.
- **Recording density = # of bits per inch (bpi).** Typically 800 or 1600 bpi. 30000 bpi on some recent devices.

MT recording in detail



Tape Organization



Secondary Storage Devices: CD-ROM

Physical Organization of CD-ROM

- **Compact Disk** – read only memory (write once), R/W is also available.
- Data is encoded and **read optically with a laser**
- Can store around +600MB data
- **Digital data** is represented as a **series of Pits and Lands:**
  - **Pit** = a **little depression, forming a lower level in the track**
  - **Land**_= **the flat part between pits**, or the upper levels in the track

Organization of data

- Reading a CD is done by shining a laser at the disc and detecting changing reflections patterns.
  - **1 = change in height (land to pit or pit to land)**
  - **0 = a "fixed" amount of time between 1's**

```
LAND        PIT      LAND              PIT       LAND
...------+        +--------------+          +---...
         |_______|               |_________|
..0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 ..
```

- Note : **we cannot have two 1's in a row**!

CD-ROM

- **While the speed of CD-ROM readers is relatively higher, such as 24X(24 times CD audio speed), the speed of writing is much slower, as low as 2X**.

  .**The DVD** (**Digital Video Disc or Digital Versatile Disc**) technology is based on CD technology with increased storage density.

- **The DVD technology allows two-side medium**, with a storage capacity of up to 10GB.