Physical Storage Systems 1

Instructor: Beom Heyn Kim

beomheynkim@hanyang.ac.kr

Department of Computer Science

Overview

- Storage Hierarchy
- Magnetic Disks
- Assignments

Classification of Physical Storage Media

- Can differentiate storage into:
  - **volatile storage:** loses contents when power is switched off
  - **non-volatile storage**:
    - Contents persist even when power is switched off.
    - Includes secondary and tertiary storage, as well as batter-backed up main-memory.
- Factors affecting choice of storage media include
  - Speed with which data can be accessed
  - Cost per unit of data
  - Reliability

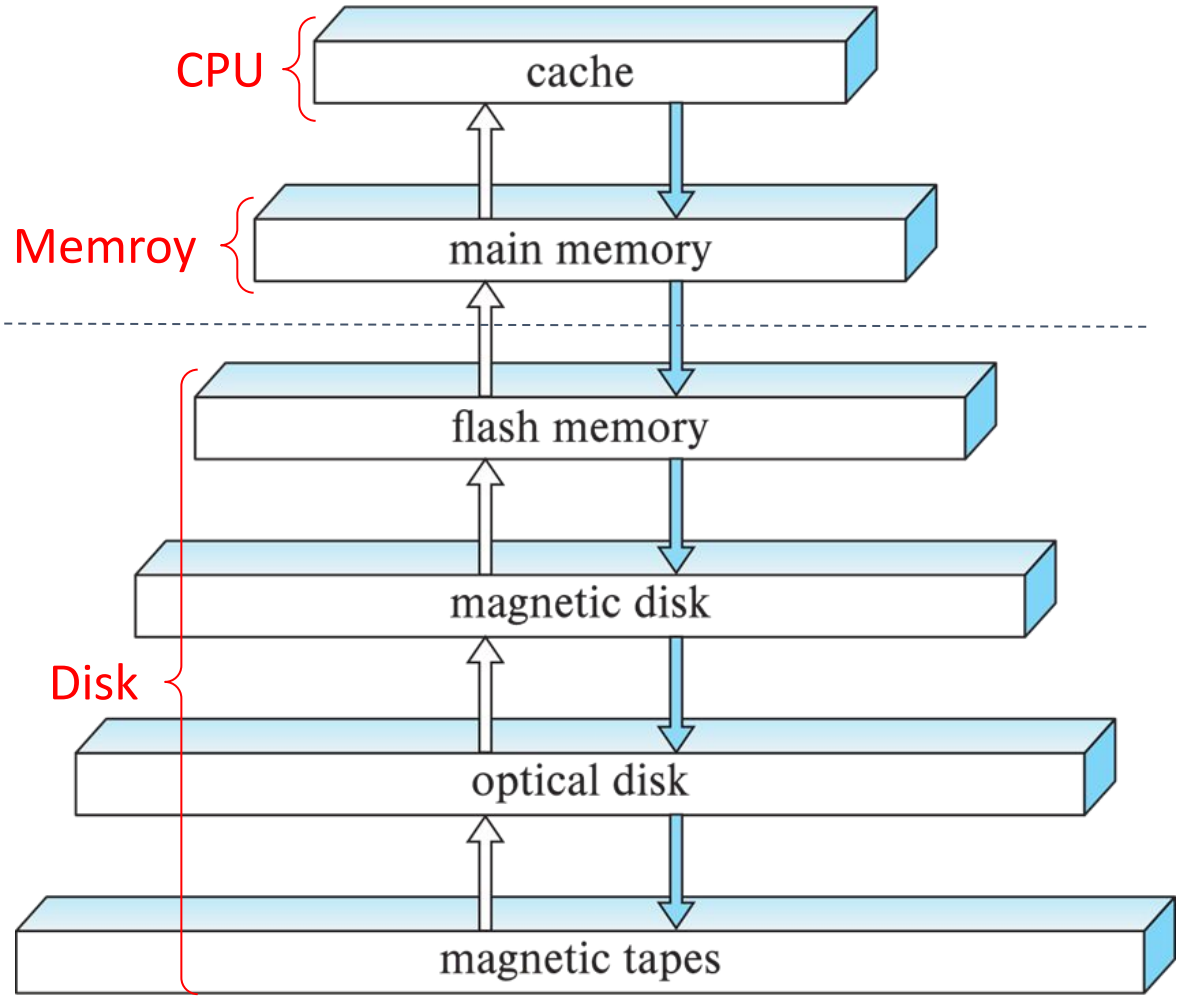
Storage Hierarchy

Volatile

Random Access
Byte-Addressable

Non-Volatile

Sequential Access
Block-Addressable



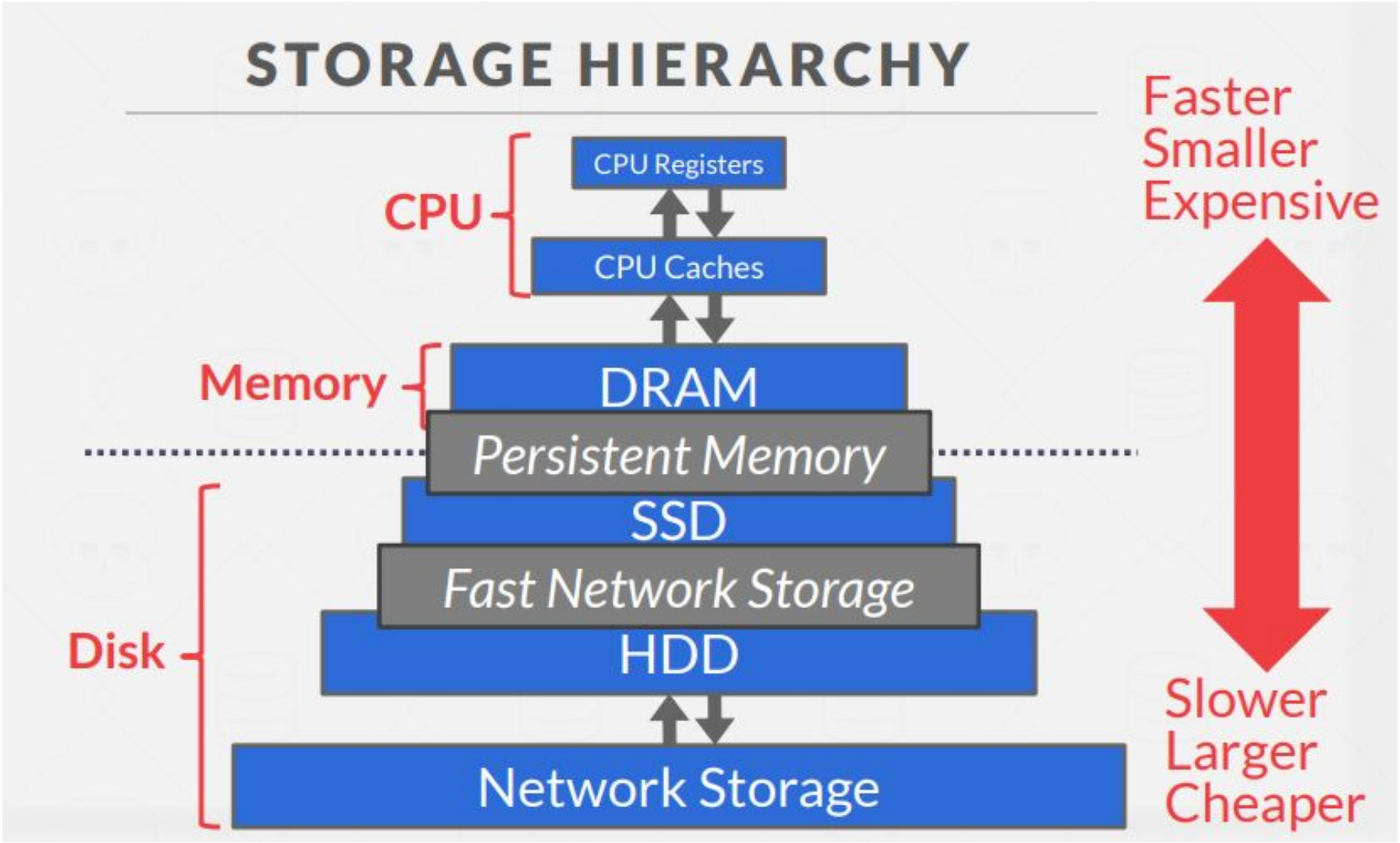
Faster
Smaller
Expensive

Slower
Larger
Cheaper

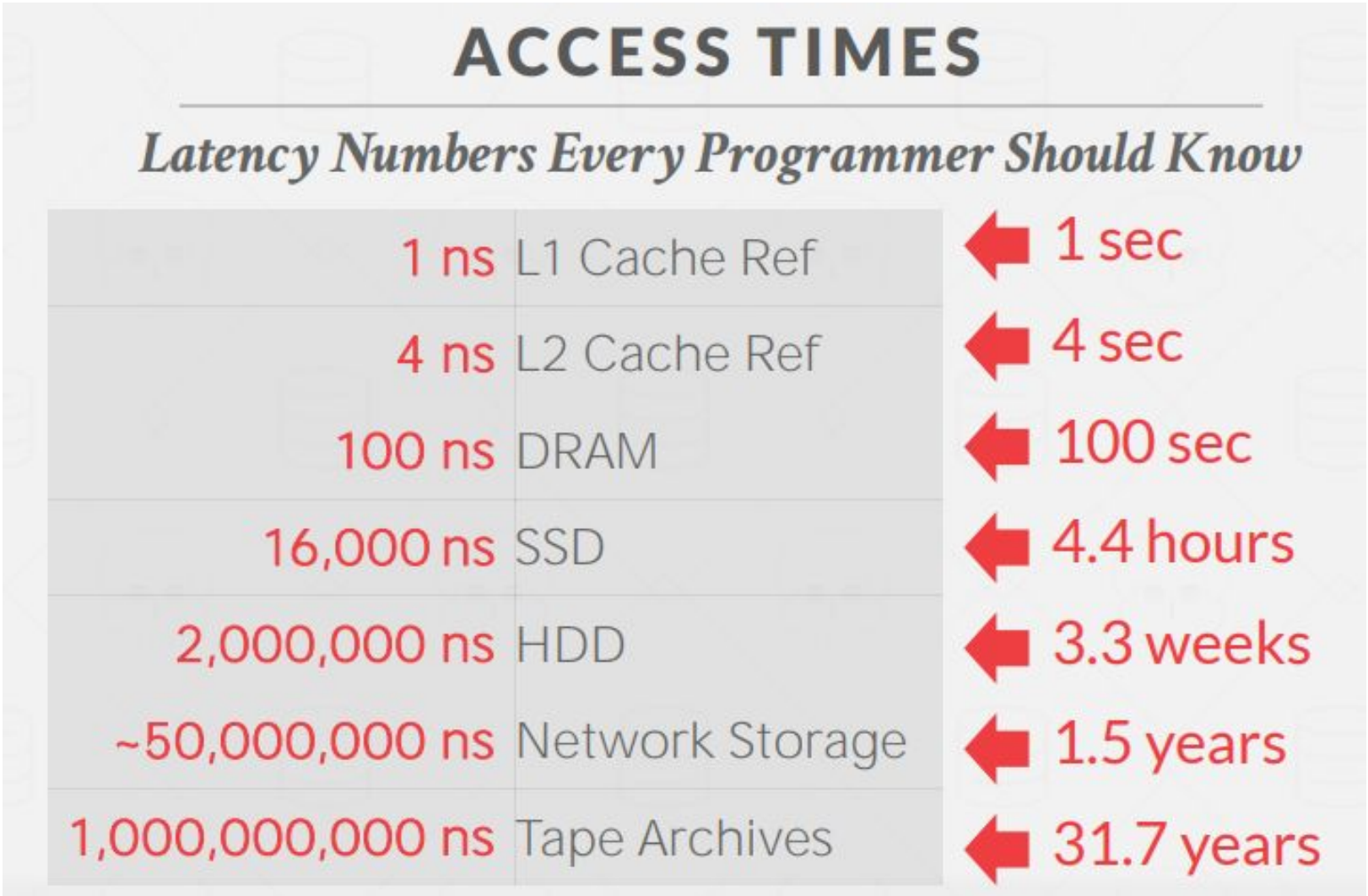
Storage Hierarchy (Cont.)

- **primary storage:** Fastest media but volatile (cache, main memory).
- **secondary storage:** next level in hierarchy, non-volatile, moderately fast access time
  - Also called **on-line storage**
  - E.g., flash memory, magnetic disks
- **tertiary storage:** lowest level in hierarchy, non-volatile, slow access time
  - Also called **off-line storage** and used for **archival storage**
  - e.g., magnetic tape, optical storage
  - Magnetic tape
    - Sequential access, 1 to 12 TB capacity
    - A few drives with many tapes
    - Juke boxes with petabytes (1000's of TB) of storage

(Modern) Storage Hierarchy



Access Times (Latency Numbers)



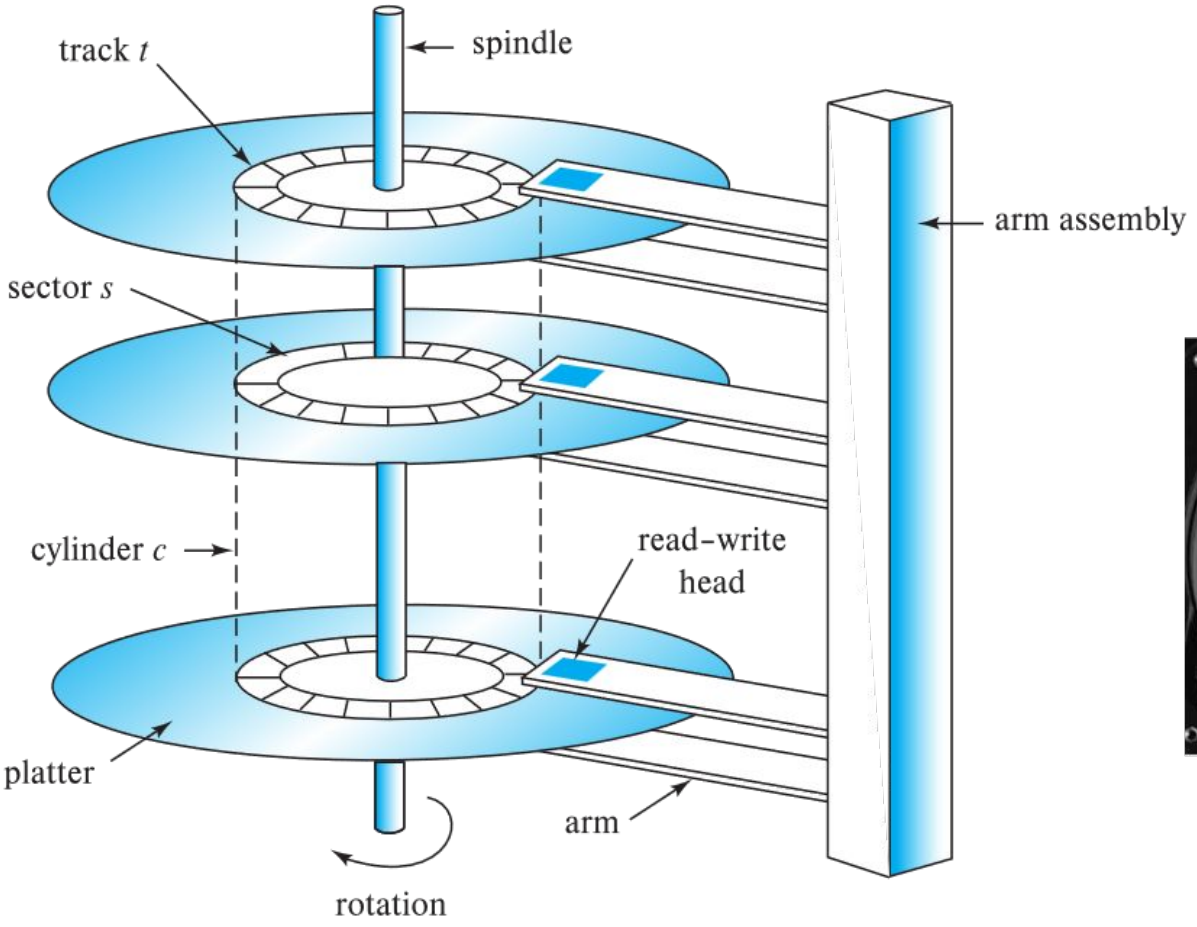
Storage Interfaces

- Disk interface standards families
  - SATA (Serial ATA)
    - SATA 3 supports data transfer speeds of up to 6 gigabits/sec
  - SAS (Serial Attached SCSI)
    - SAS Version 3 supports 12 gigabits/sec
  - NVMe (Non-Volatile Memory Express) interface
    - Works with PCIe connectors to support lower latency and higher transfer rates
    - Supports data transfer rates of up to 24 gigabits/sec
- Disks usually connected directly to computer system
- In **Storage Area Networks (SAN)**, a large number of disks are connected by a high-speed network to a number of servers
- In **Network Attached Storage (NAS)**, networked storage provides a file system interface using networked file system protocol, instead of providing a disk system interface

Overview

- Storage Hierarchy
- Magnetic Disks

Magnetic Hard Disk Mechanism



Schematic diagram of magnetic disk drive

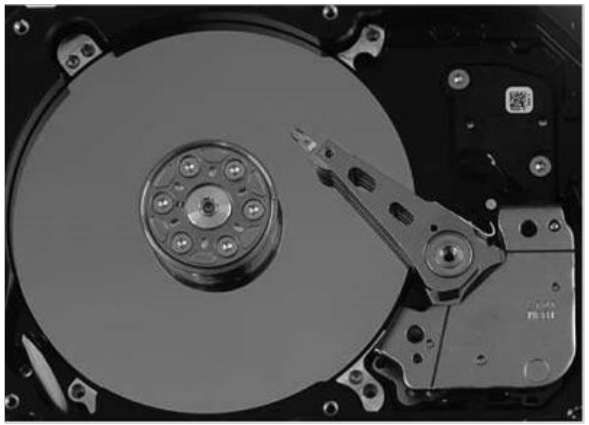


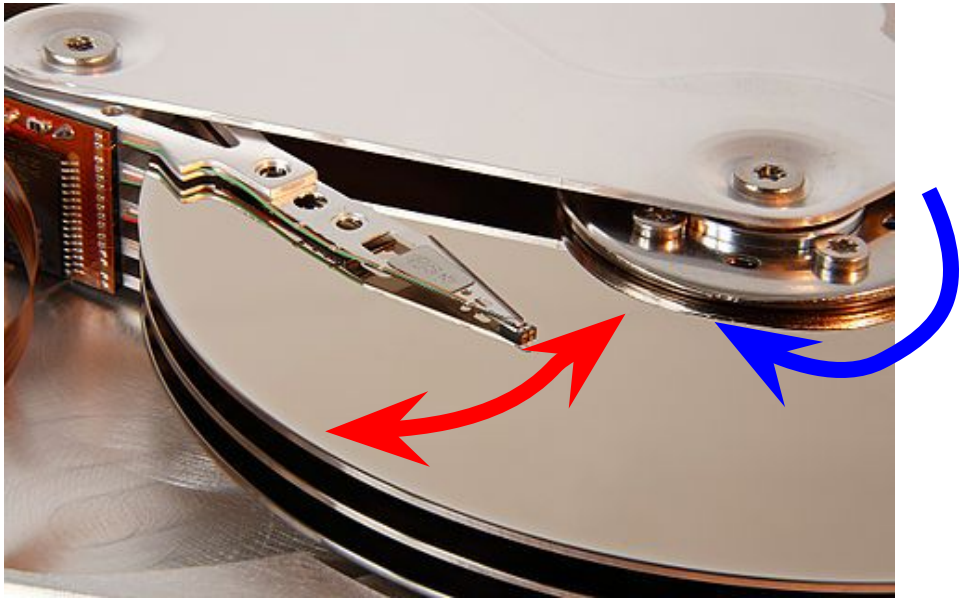
Photo of magnetic disk drive

Magnetic Disks

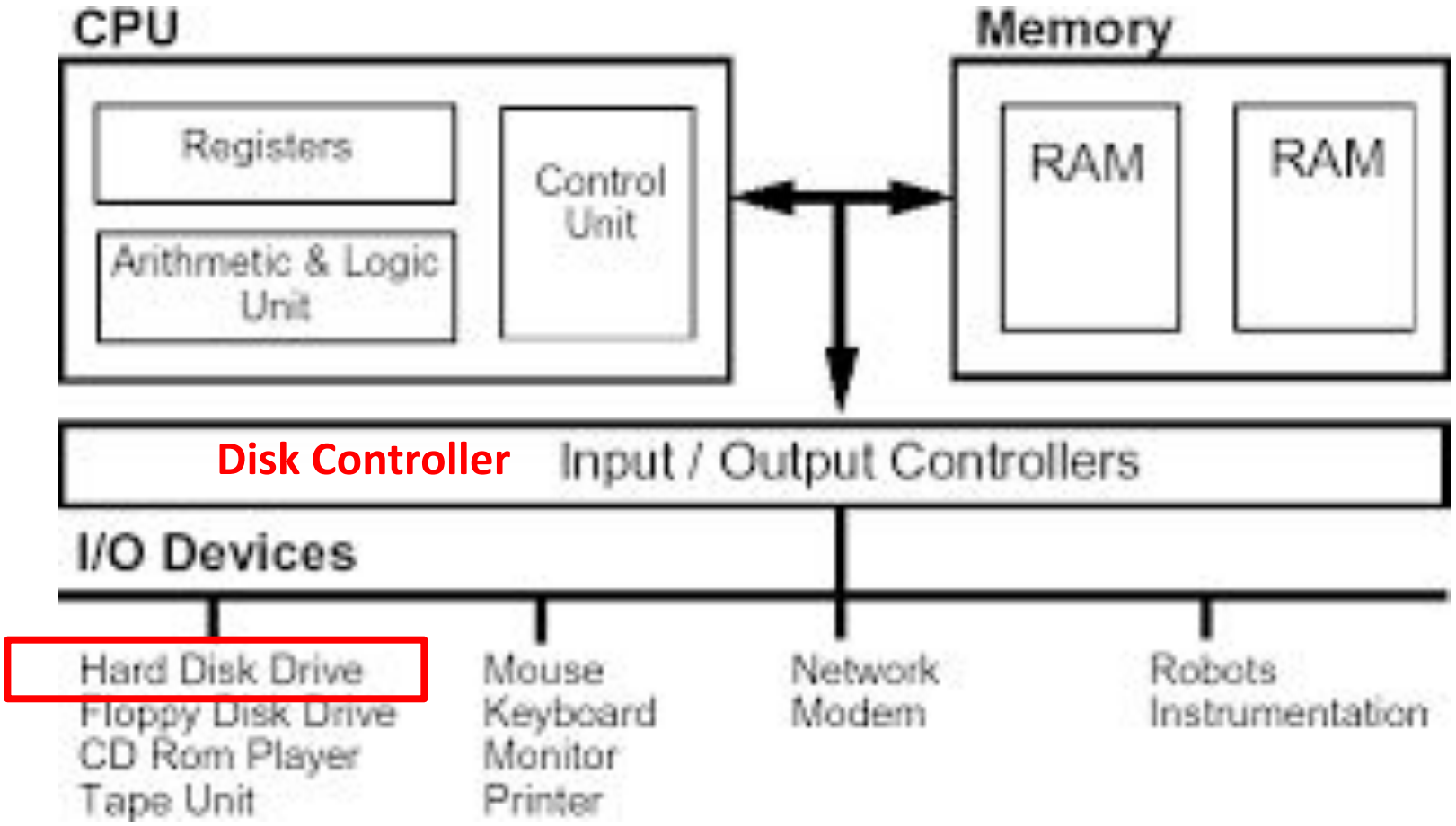
- Surface of platter divided into circular **tracks**
  - Over 50K-100K tracks per platter on typical hard disks
- Each track is divided into **sectors.**
  - A sector is the smallest unit of data that can be read or written.
  - Sector size typically 512 bytes
  - Typical sectors per track: 500 to 1000 (on inner tracks) to 1000 to 2000 (on outer tracks)
- Head-disk assemblies
  - multiple disk platters on a single spindle (1 to 5 usually)
  - one head per platter, mounted on a common arm
- **Cylinder** *i* consists of *i*-th track of all the platters

Magnetic Disks (Cont.)

- **Read-write head** reads or writes data of the sector placed under it
  - To read/write a sector
    - disk arm swings to position head on right track
    - platter spins continually
    - data is read/written as sector passes under head



Magnetic Disks (Cont.)

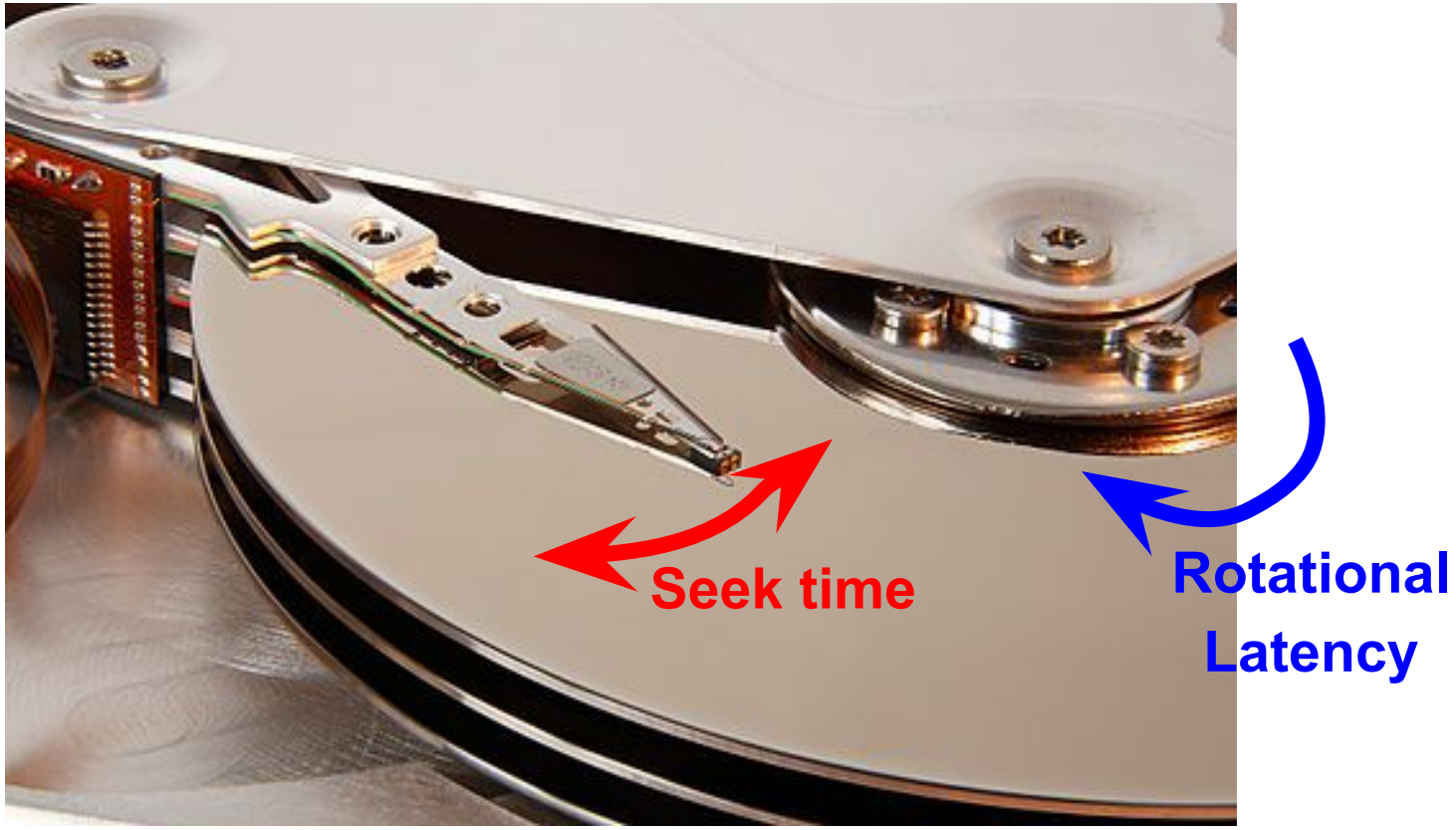


Magnetic Disks (Cont.)

- **Disk controller** – interfaces between the computer system and the disk drive hardware.
  - accepts high-level commands to read or write a sector
  - initiates actions such as moving the disk arm to the right track and actually reading or writing the data
  - computes and attaches **checksums** to each sector to verify that data is read back correctly
    - If data is corrupted, with very high probability stored checksum won't match recomputed checksum
  - Ensures successful writing by reading back sector after writing it
  - Performs **remapping of bad sectors**

Performance Measures of Disks

- **Access time = Seek time + Rotational latency**



Performance Measures of Disks (Cont.)

- **Access time** – the time it takes from when a read or write request is issued to when data transfer begins. Consists of:
  - **Seek time** – time it takes to reposition the arm over the correct track.
  - **Rotational latency** – time it takes for the sector to be accessed to appear under the head.
  - Overall latency is 5 to 20 ms depending on disk model!
- **Data-transfer rate** – the rate at which data can be retrieved from or stored to the disk.

Performance Measures of Disks (Cont.)

- Requests for disk I/O are typically generated by the file system but can be generated directly by the database system
  - Each request specifies the address on the disk to be referenced in the form of a *block number*.
- **Disk block** is a logical unit for storage allocation and retrieval
  - 4 to 16 kilobytes typically
    - Smaller blocks: more transfers from disk
    - Larger blocks: more space wasted due to partially filled blocks
  - Data are transferred between disk and main memory in units of blocks

Performance Measures of Disks (Cont.)

- **Sequential access pattern**
  - Successive requests are for successive disk blocks
  - Disk seek required only for first block
- **Random access pattern**
  - Successive requests are for blocks that can be anywhere on disk
  - Each access requires a seek
  - Transfer rates are low since a lot of time is wasted in seeks
- **I/O operations per second (IOPS)**
  - Number of random block reads that a disk can support per second
  - 50 to 200 IOPS on current generation magnetic disks

Performance Measures of Disks (Cont.)

- **Mean time to failure (MTTF)** – the average time the disk is expected to run continuously without any failure.
  - Typically 3 to 5 years
  - Probability of failure of new disks is quite low, corresponding to a "theoretical MTTF" of 500,000 to 1,200,000 hours for a new disk – about 57 to 136 years
    - E.g., an MTTF of 1,200,000 hours for a new disk means that given 1000 relatively new disks, on an average one will fail every 1200 hours
  - MTTF decreases as disk ages

Assignments

- Reading: Ch12.1-12.3
- Practice Excercises: 12.1, 12.2

Solutions to the Practice Excercises:
https://www.db-book.com/Practice-Exercises/index-solu.html

The End