1 Storage Devices

Disks

- Main components – Arm, head, track/cylinder, platter, sector, controller.
- Typically 2-20 surfaces, 1000-2000 tracks per surface, 32-64 sectors per track, 1024-8192 bytes per sector.
- Currently, 1 TByte disk costs less than $100.
- Sectors can be read and written individually, or in adjacent groups:
  - *Seek*: move heads to correct track.
  - Select desired read/write head.
  - Wait for disk to rotate desired sector into position.
  - Read or write sector while it spins by.
- Seek time = 5-50 ms, rotational latency is 0-16 ms (drive spins at 3600/7200/10000) RPM).
- Technology advances lately mostly in miniaturization. In 1975, 40 Mbytes took up space the size of a washing machine.

Hardware evolution – capacity grows much faster than speed (it takes longer and longer to read an entire hard disk). As a consequence disks throughput is likely to become an important bottleneck, and the operating system should work to mitigate this.

A formatted sector includes a preamble to detect beginning and an error correcting code (ECC) that corrects a small number of bit flips and detects up to a certain larger number of bit flips.

The first sectors of different tracks are not aligned. A track differential is required for continuous scanning of consecutive sectors. E.g., suppose a disk with 300 sectors/track, a seek time between adjacent tracks of 0.8 ms and that rotates at $10,000$ rpm. How many sectors should the track differential be?

Disk access time = seek time + rotational latency + transfer time.
OS can only try to minimize the first component (see next lecture).

Solid State Drives (SSDs)

- solid state memory, NAND-flash based (other technologies: M-RAM, PCM).
- No mechanically moving parts
- Multiple flash packages + volatile memory for controller
- fast read access (less than $.01ms$).
- still considerably more expensive than disk drives ($3.5 versus $.2 per Gigabyte in 2008).
- capacity catching up quickly with disk drives.
- no mechanical seek, no rotational latency.

Internal organization:

- 1000s of blocks
- Each block has several pages:
  - 4 KB data

  - 128 bytes for metadata (ECC,)
- Interface allow reading and writing at page granularity
- Atomically access data + metadata

Access characteristics:

- Fast random-access reads (sub-millisecond)
- Small in-place updates are inefficient
  - Need to erase block before overwriting
  - Costly (1-2 ms)

Solution: Flash Translation Layer (FTL)

- In-memory remap table
- Map logical pages to physical ones
- Write to new page + update mapping
- What happens upon reboot?
- Need to reconstruct remap table. How?
- Store identity and version number of logical page in metadata part
- Problem: obsolete pages within blocks
- Need GC procedure: Select block, copy valid pages out, erase, add it to free block list

Another limitation: wear

- Reliability degrades after many write-erase cycles (100,000s)
- Can be alleviated using wear-leveling algorithms
- Maximize device lifetime as a whole by shifting blocks around to avoid permanent blocks from never being written while "hot-spots" degrade quickly.

Comparison: Disks vs SSDs

SSDs have:

- Low latency
- Higher read+write throughput (slightly higher in sequential access, much higher for random access)
- Lower energy consumption

Disks have:

- Better (lower) cost / gigabyte
- No write-wear

Different market segments: Disks are still the main choice for enterprise storage (and likely to continue so in the future) while SSDs are used when energy and lack of mechanical components matter (portable devices).