W4118: disks

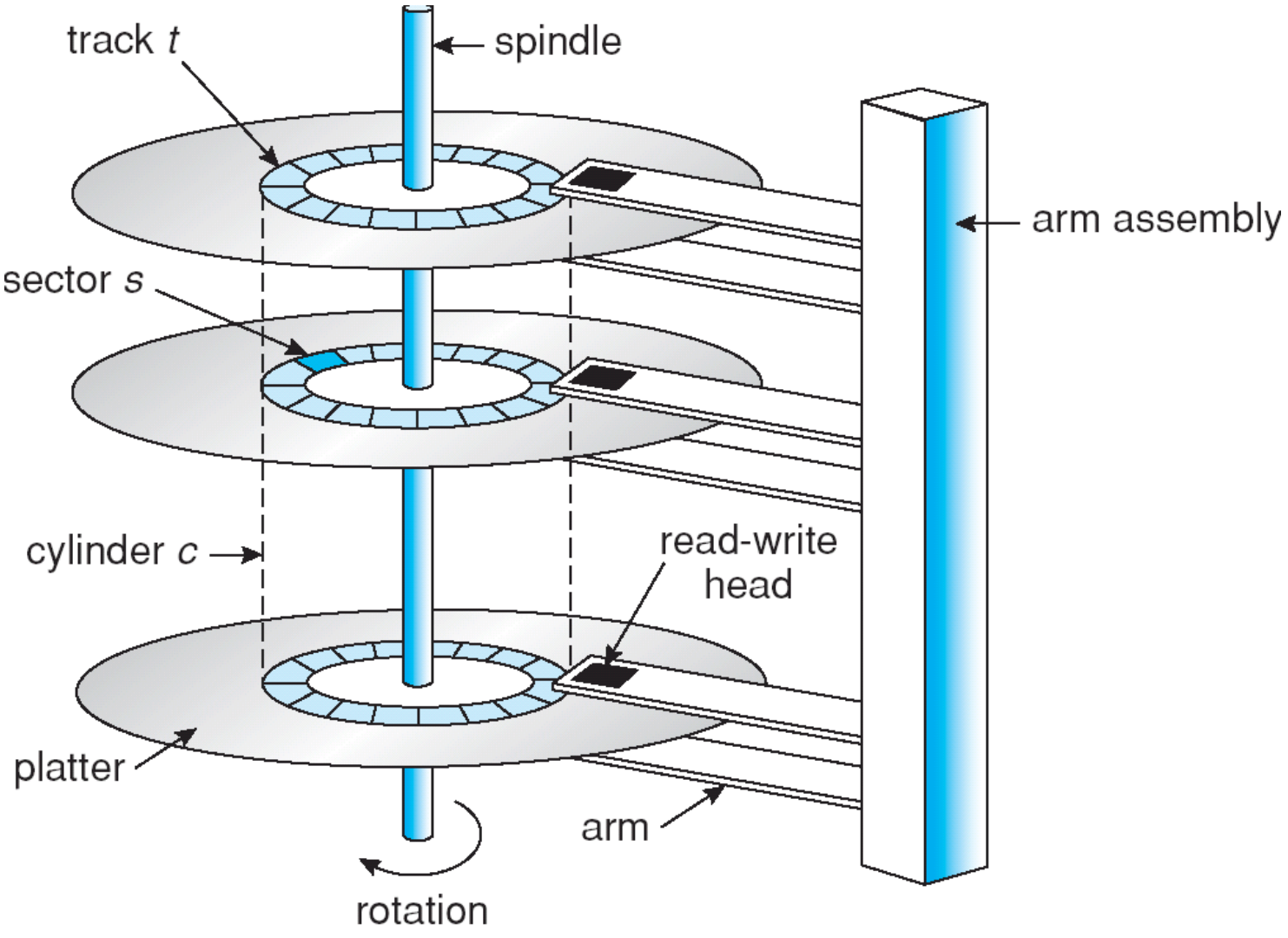
Instructor: Junfeng Yang

References: Modern Operating Systems (3rd edition), Operating Systems Concepts (8th edition), previous W4118, and OS at MIT, Stanford, and UWisc

Outline

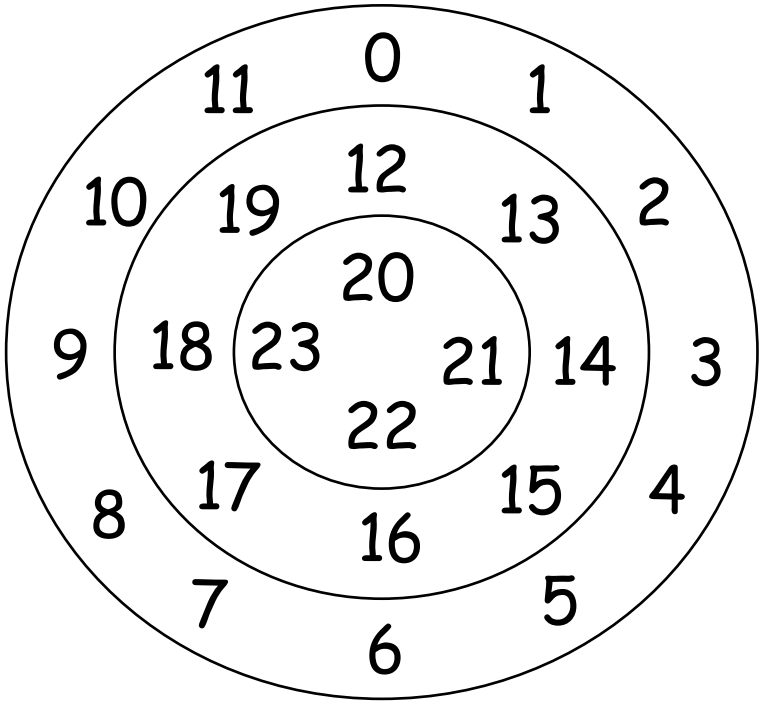
- Disk characteristics
- Disk scheduling

Disk structure



Disk interface

- From FS perspective: disk is addressed as a one dimension array of logical sectors
- Disk controller maps logical sector to physical sector identified by surface #, track #, and sector #

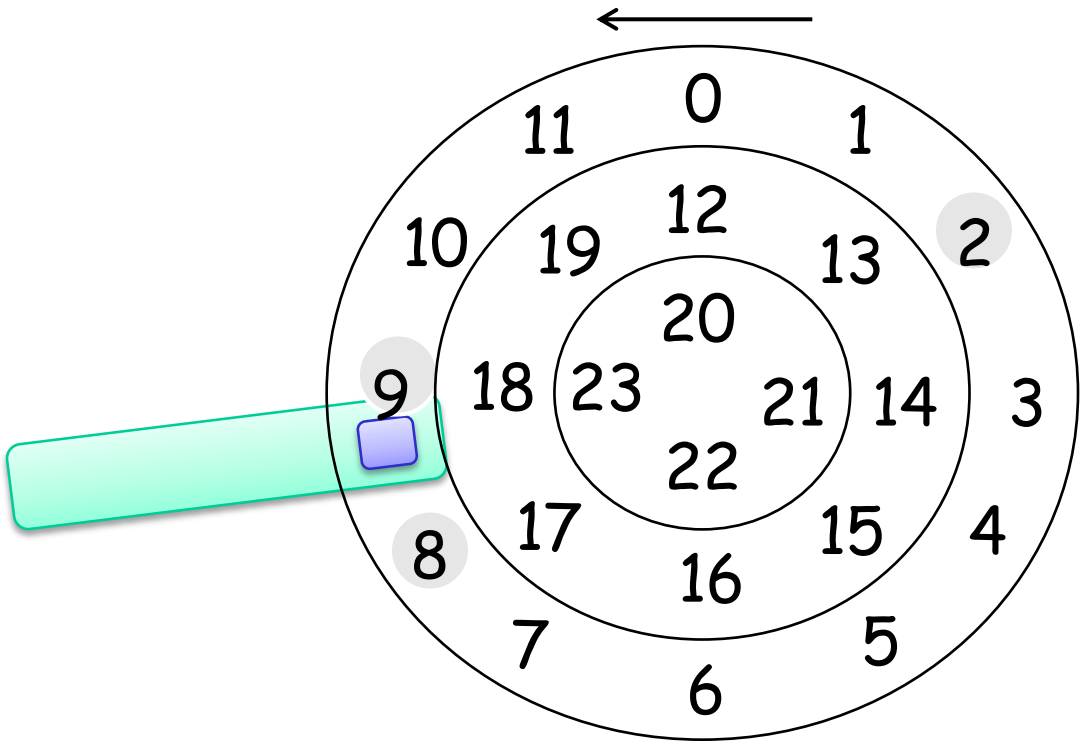


Disk latencies

- Rotational delay: rotate disk to get to the right sector
- Seek time: move disk arm to get to the right track
- Transfer time: get bits off the disk

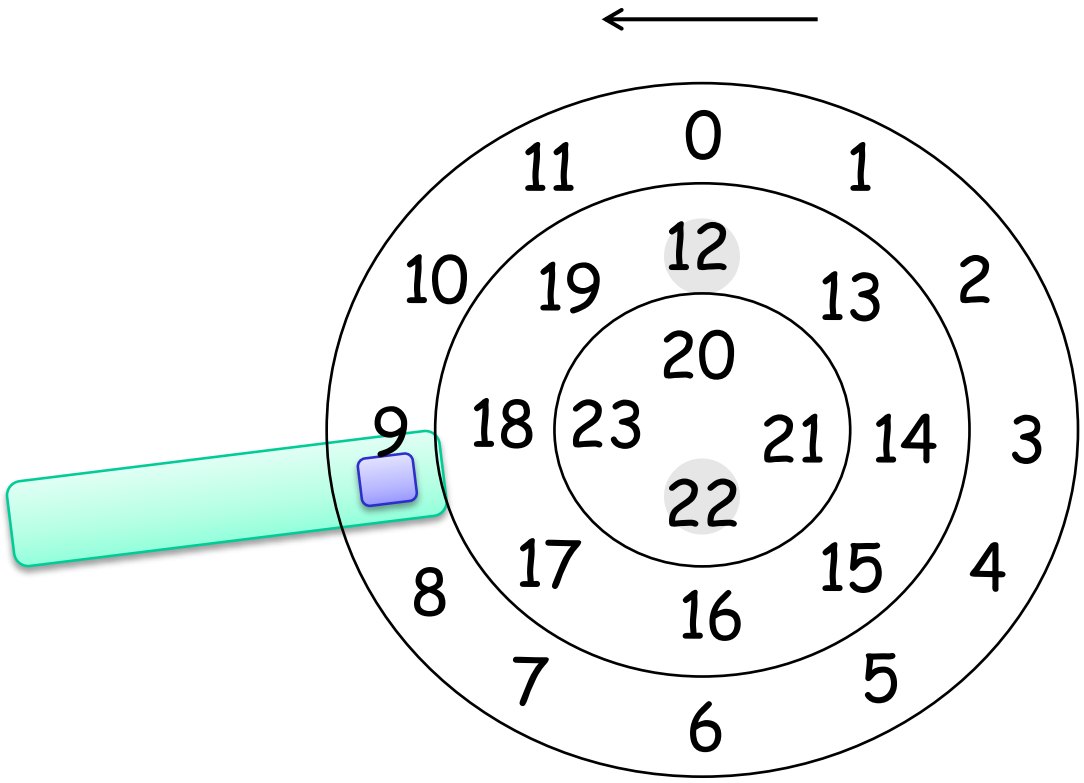
Rotational delay

- Full rotation time: e.g., 4-8ms
- Average rotational delay: half of full rotation time



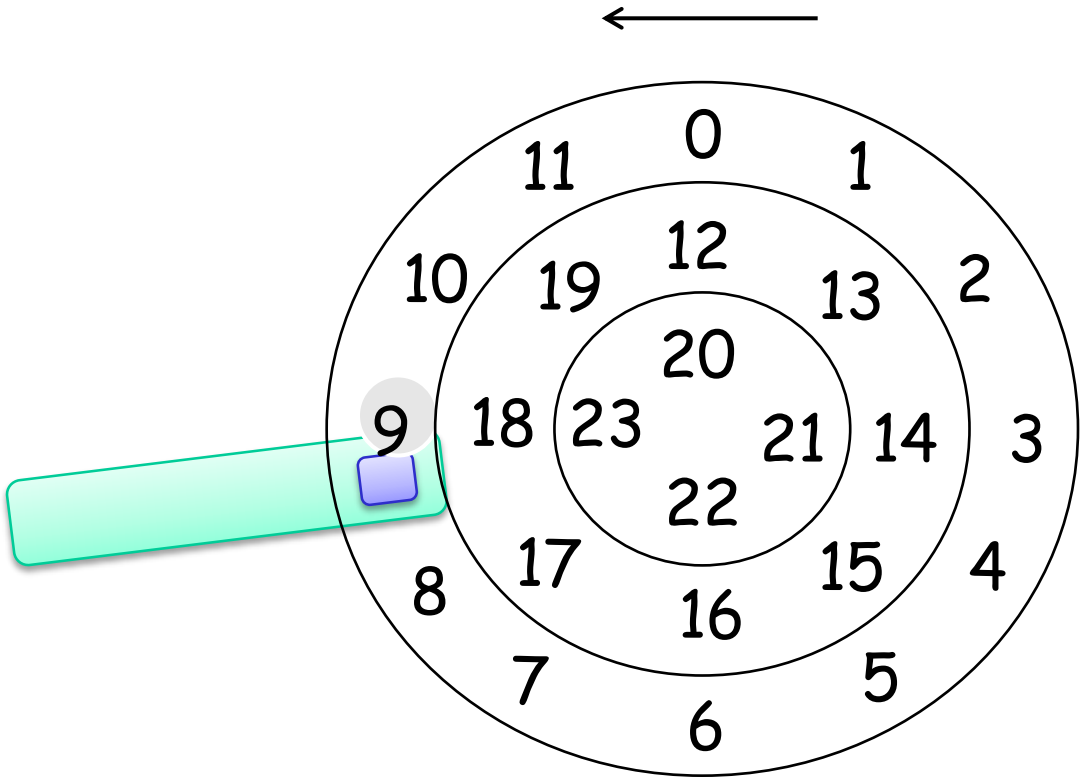
Seek time

- Must move arm to the right track
- Can take a while (e.g., 0.5 – 2ms)



Transfer time

- Transfer bits out of disk
- Actually pretty fast (e.g., 125MB/s)



I/O time (T) and rate (R)

- T = Rotational delay + seek time + txfer time
- R = Size of transfer / T
- Workload 1: large sequential accesses?
- Workload 2: small random accesses?

Example

	Barracuda	Cheetah 15K.5
Capacity	1TB	300GB
Rotational speed	7200 RPM	15000 RPM
Rotational latency (ms)	4.2	2.0
Avg seek (ms)	9	4
Max Transfer	105 MB/s	125 MB/s
Platters	4	4
Connects via	SATA	SCSI

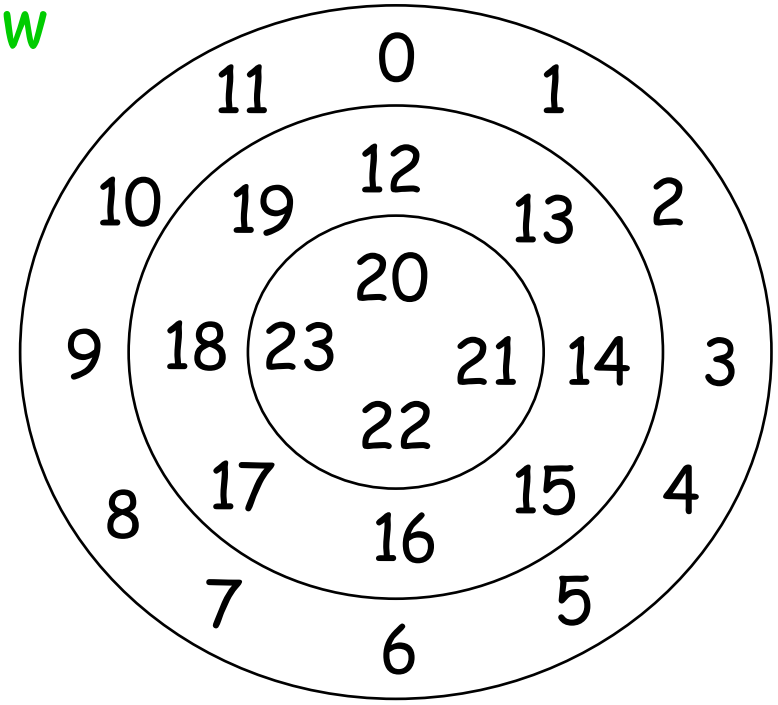
- Random 4KB read
  - Barracuda: T = 13.2ms, R = 0.31MB/s
  - Cheetah: T = 6ms, R = 0.66MB/s
- Sequential 100 MB read
  - Barracuda: T = 950ms, R = 105 MB/s
  - Cheetah: T = 800ms, R = 125 MB/s

Design tip: use disks sequentially

- Disk performance differs by a factor of 200 or 300 for random v.s. sequential accesses
- When possible, access disks sequentially

Mapping of logical sectors to physical

- Logical sector 0: the first sector of the first (outermost) track of the first surface
- Logical sector address incremented within track, then tracks within cylinder, then across cylinders, from outermost to innermost
- Track skew



Pros and cons of default mapping

- Pros
  - Simple to program
  - Default mapping reduces seek time for sequential access
- Cons
  - FS can't precisely see mapping
  - Reverse-engineer mapping in OS is difficult
    - # of sectors per track changes
    - Disk silently remaps bad sectors

Disk cache

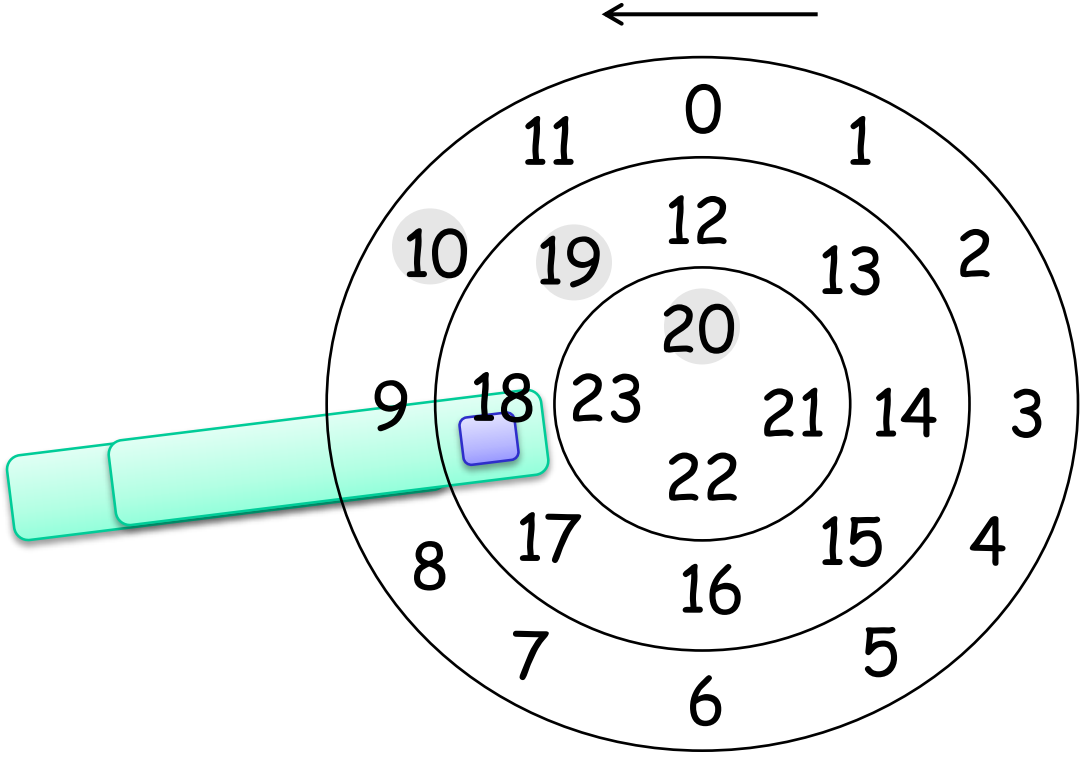
- Internal memory (8MB-32MB) used as cache
- Read-ahead: "track buffer"
  - Read contents of entire track into memory during rotational delay
- Write caching with volatile memory
  - Write back or immediate reporting: claim written to disk when not
    - Faster, but data could be lost on power failure
  - Write through: ack after data written to platter

Disk scheduling

- Goal: minimize positioning time
  - Performed by both OS and disk itself
  - Why?
- Schedule requests in order received (FCFS)
  - Advantage: fair
  - Disadvantage: high seek cost and rotation
- Shortest seek time first (SSTF):
  - Handle nearest cylinder next
  - Advantage: reduces arm movement (seek time)
  - Disadvantage: unfair, can starve some requests

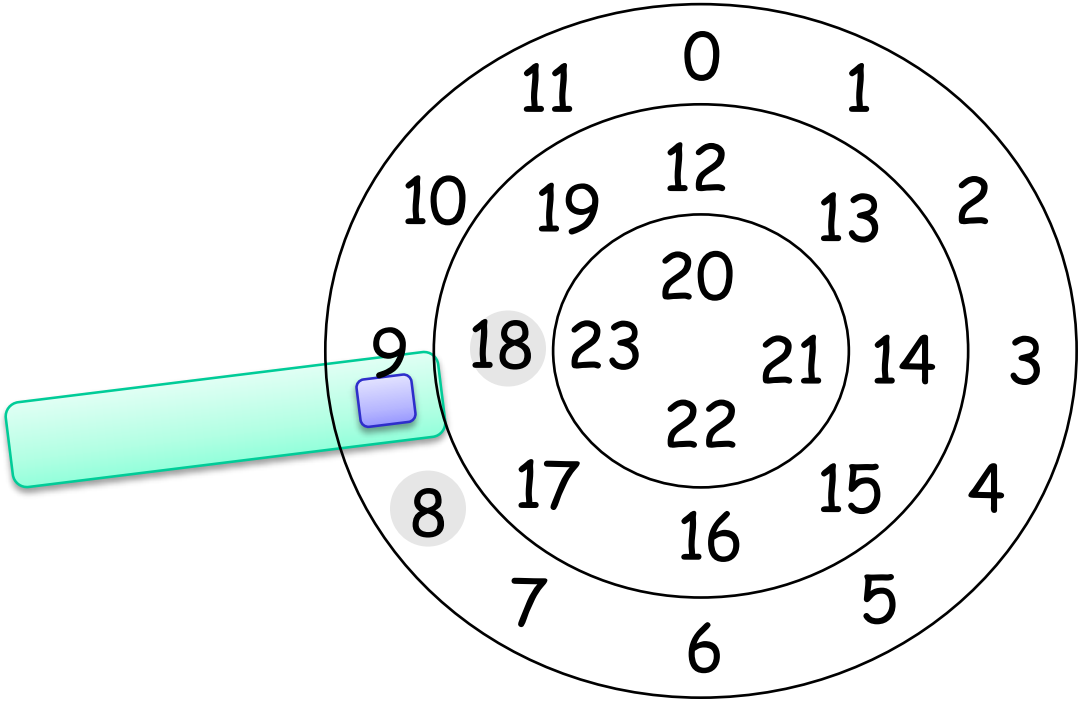
Elevator (aka SCAN or C-SCAN)

- Disk arm sweeps across disk
- If request comes for a block already serviced in this sweep, queue it for next sweep



Modern disk scheduling issues

- Elevator (or SSTF) ignores rotation!
- Shortest positioning time first (SPTF)
- OS + disk work together to implement



Disk technology trends

- Data ➔ more dense
  - More bits per square inch
  - Disk head closer to surface
  - Create smaller disk with same capacity
- Disk geometry ➔ smaller
  - Spin faster ➔ Increase b/w, reduce rotational delay
  - Faster seek
  - Lighter weight
- Disk price ➔ cheaper
- Density improving more than speed (mechanical limitations)

New mass storage technologies

- New memory-based mass storage technologies avoid seek time and rotational delay
  - NAND Flash
  - Battery-backed DRAM (NVRAM)
- Disadvantages
  - Price: more expensive than same capacity disk
  - Reliability: more likely to lose data
- Open research question: how to effectively use flash in commercial storage systems