ECE 5730
Memory Systems
Spring 2009

Disk Drive Performance Issues and Design Tradeoffs

Cornell University

Announcements

- **Quiz 12**
  - **Average = 8.5**

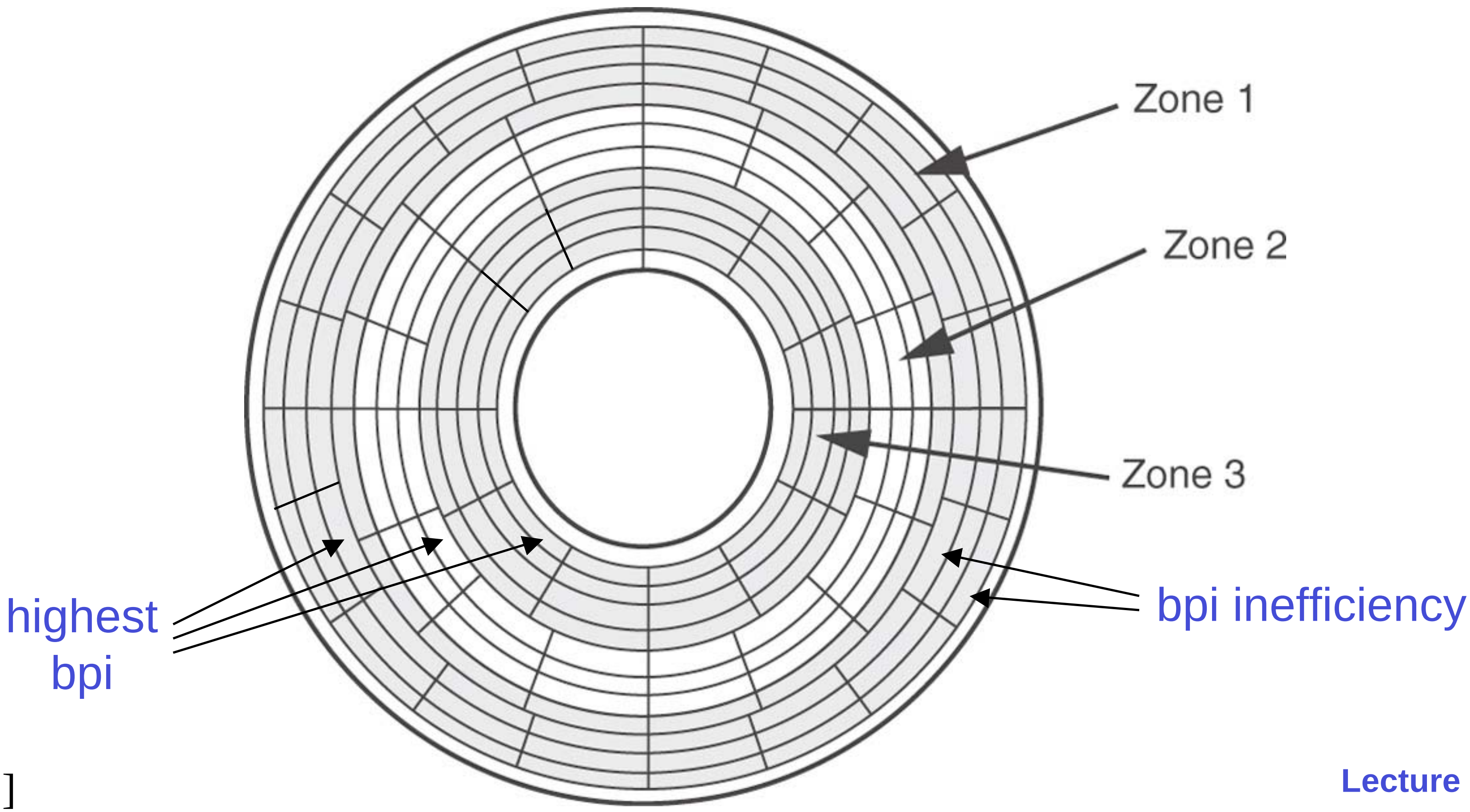
- **Quiz 13 (last one!) on Tuesday**

- **Exam II**
  - **May 7, 7:00-10:00pm, Hollister 314**
  - **Covers material from 3/10-4/28 but excluding 4/22 (Lectures 14-21, 23-24)**
  - **Let me know immediately if you require a make-up**

Project Report Due May 1st

Zoned-Bit Recording (ZBR)

- **To get max density, use max bpi for all tracks**
  - **Different recording rates for 100's of 1000's of tracks**
- **Simplification: Zones with same sectors/track**



[18.11]

Quantization Effect

- **Integer number of sectors per concentric track causes lost bytes in every track**

⇒ could store non-integer # of sectors, but can't address it

- **Example: 512KB/track, 600 bytes/sector**
  - **416 lost bytes/track**
  - **300,000 tracks, 120MB wasted**

- **ZBR causes additional wasted space due to bpi inefficiency**

CLV and CAV

- **Constant Linear Velocity**
  - Variable rotational speed, constant data rate
  - No wasted space, good sequential performance
  - Audio CDs, Blu-Ray movies
- **Constant Angular Velocity**
  - Constant rotational speed, variable data rate
  - Good random access performance
  - Hard drives, computer CDs & DVDs
- **ZBR can be used in both CLV and CAV**

→ "combination" of CLV, CAV

[18.2]

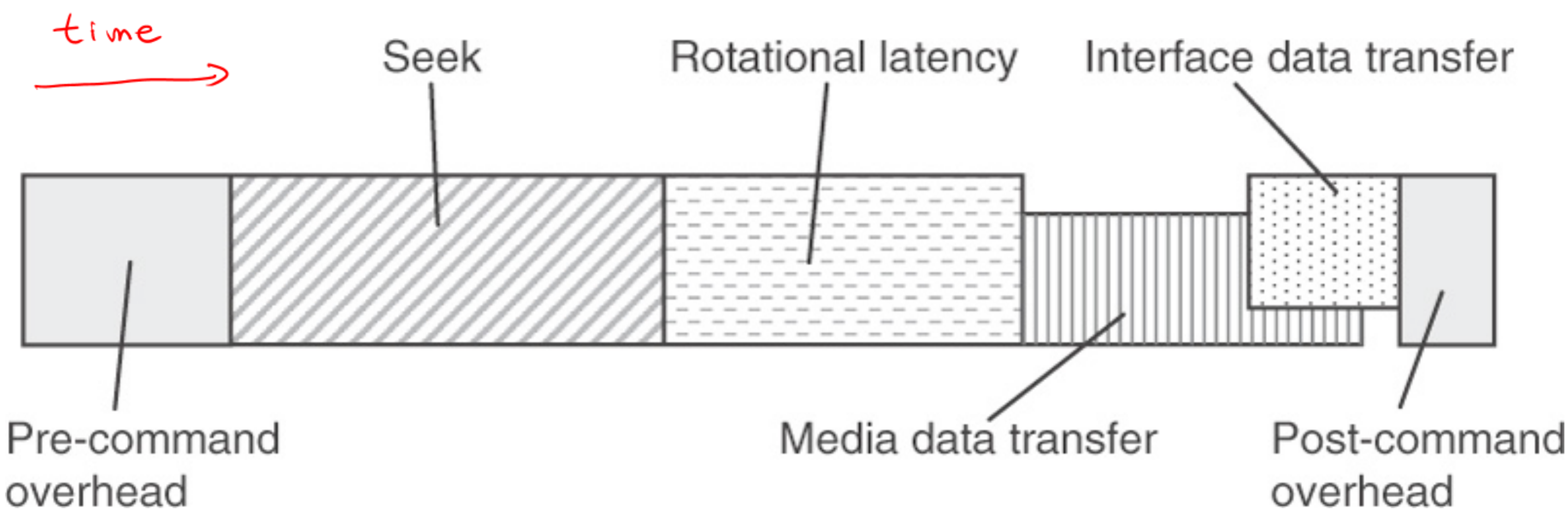
SAS and SATA Compatibility

- SAS supports SATA drives through the *Serial ATA Tunneling Protocol*
- Interfaces are pin compatible but voltage levels and protocols are different
- SAS controller recognizes type of drive and uses appropriate voltages and commands

Disk Drive Performance

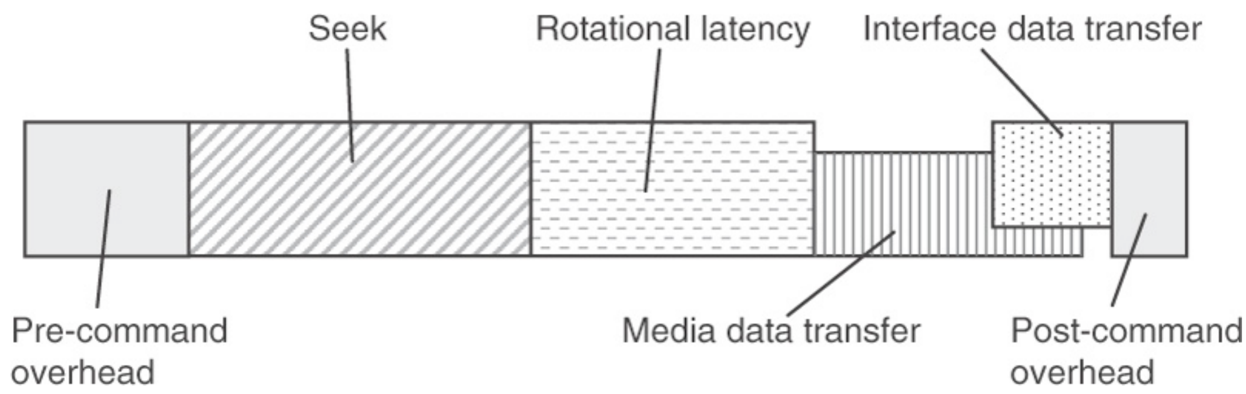
- **Desire fast response time and high throughput**

- **Time components of an I/O read operation**



[19.1]

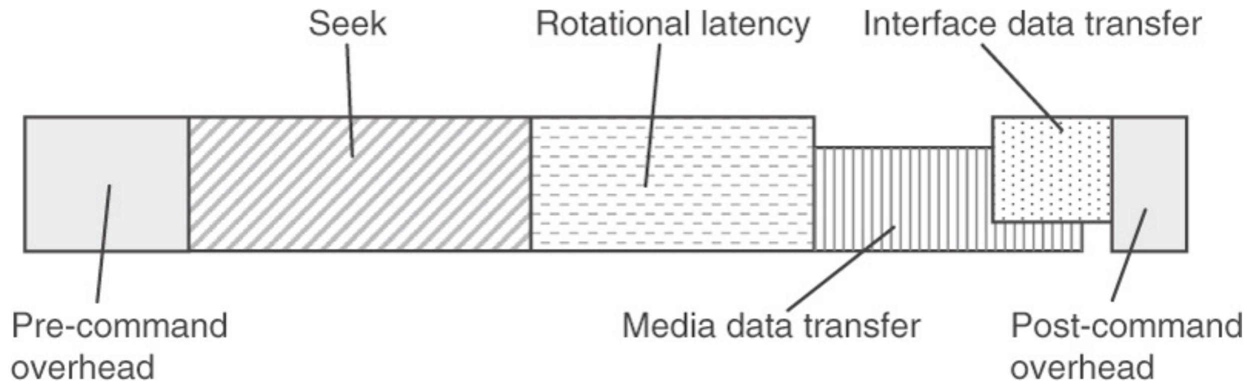
Time Components of an I/O Read



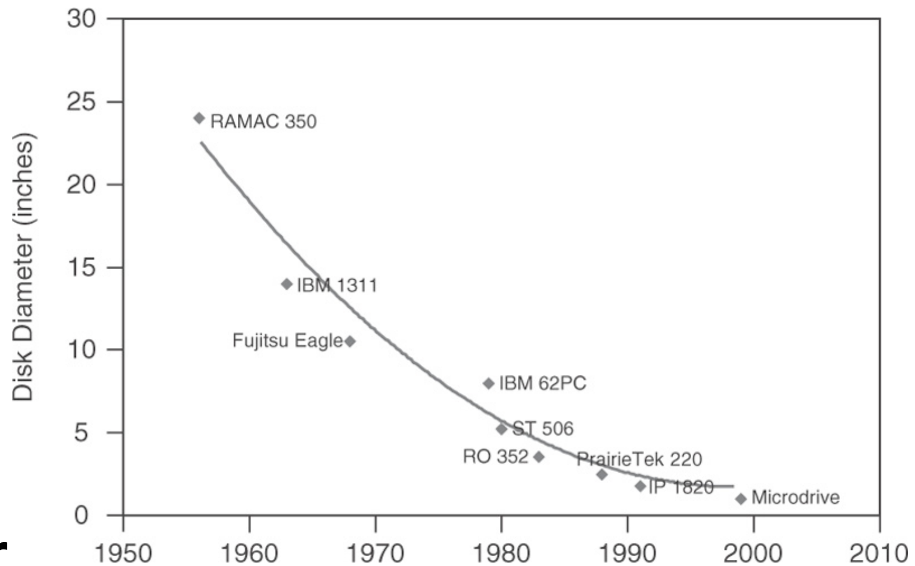
- **Pre-command overhead**
  - **Receive, decode, buffer command**
  - **Disk cache lookup** (to see if it's in there)
  - **Start up the operation** (actually start doing stuff)

[19.1]

Time Components of an I/O Read

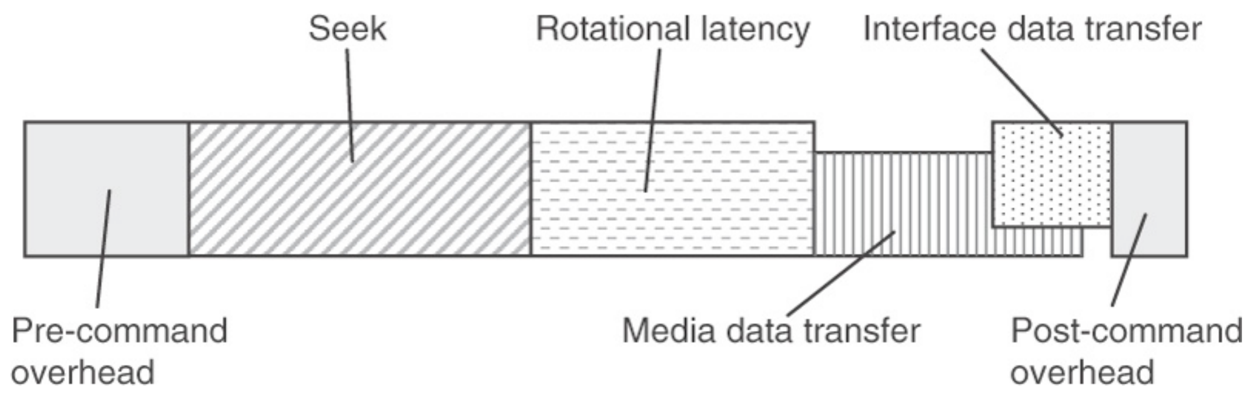


- **Seek time**
  - Move head from current cylinder to target cylinder
  - One of the largest time overheads (~5-10ms)
  - Travel time + settle time
  - Lower travel time with reduction in disk diameter

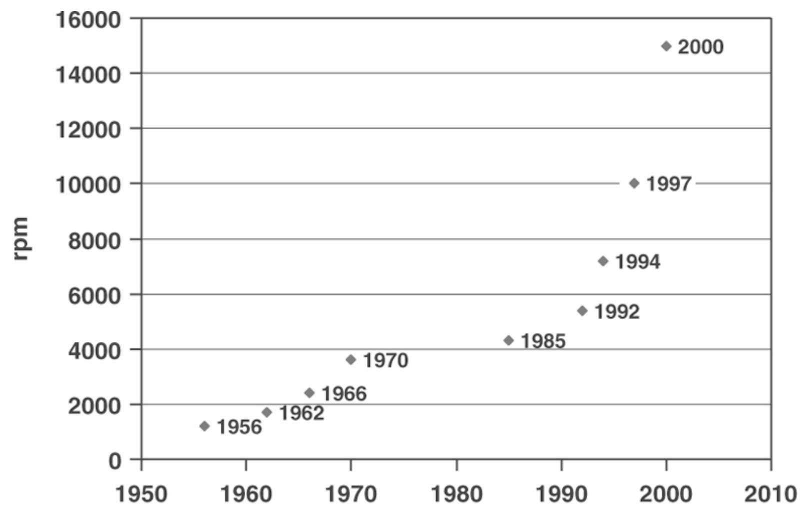


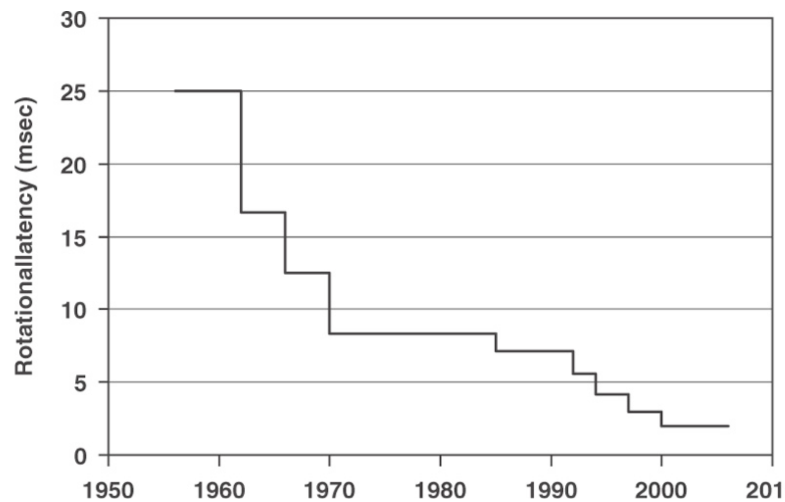
[19.1,16.3]

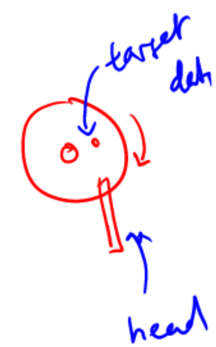
Time Components of an I/O Read



- **Rotational latency** (how long it takes for the data you want to rotate into position)
  - **Time to bring start of target sector to the head**
    - **Avg RL = 1/2 time to do complete revolution**





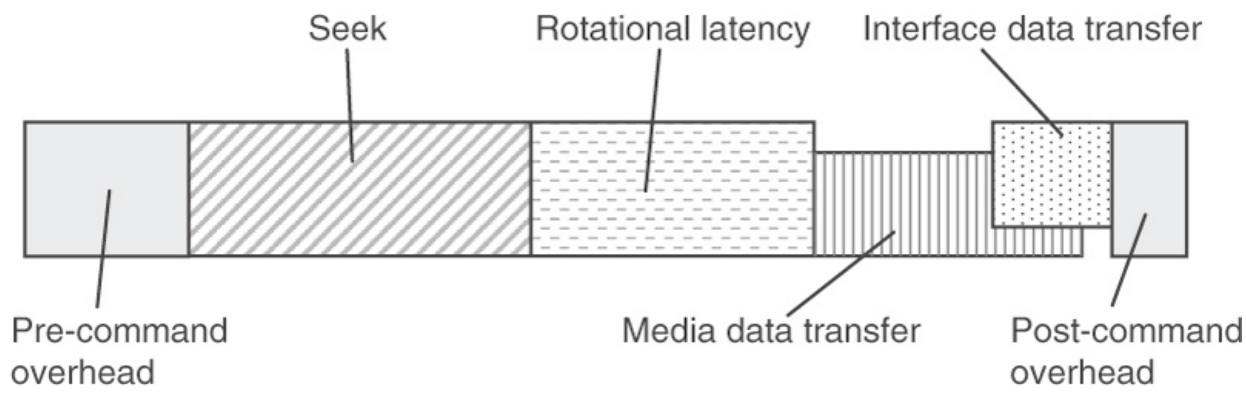


[19.1,19.2]

year when particular rpm and RL were *introduced*

heavily drive dependent

Time Components of an I/O Read



- **Media data transfer**
  - **Time to transfer data from media (disk)**
  - **Calculated as *sustained data rate* (SDR)**
  - **Depends on**
    - **Rotational speed (60/rpm)**
    - **Track density, measured as sectors per track (SPT)**
    - **Number of heads (N)**
    - **Cylinder skew: time to switch between cylinders (T_{CS})**
    - **Head switch time: time to switch between tracks (T_{HS})**

[19.1]

no overhead → $SDR = \frac{1000 \times 512}{60/7200} \approx 61$ MB/s

Time Components of an I/O Read

- **Media data transfer**
  - **For serpentine formatting**

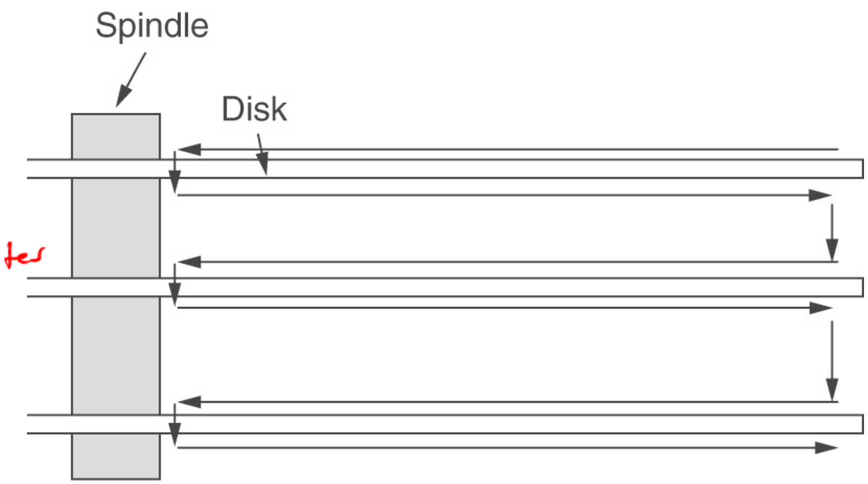
$$SDR = \frac{SPT \times 512}{60/rpm + T_{CS}}$$

← sector size in Bytes

cylinder switch time

~ 2ms

$SDR = \frac{1000 \times 512}{60/7200 + 2} \approx 50$ MB/s



  - **For cylinder formatting**

$$SDR = \frac{N \times SPT \times 512}{(N \times 60)/rpm + \underbrace{(N-1)} \times T_{HS} + T_{CS}}$$

number of heads

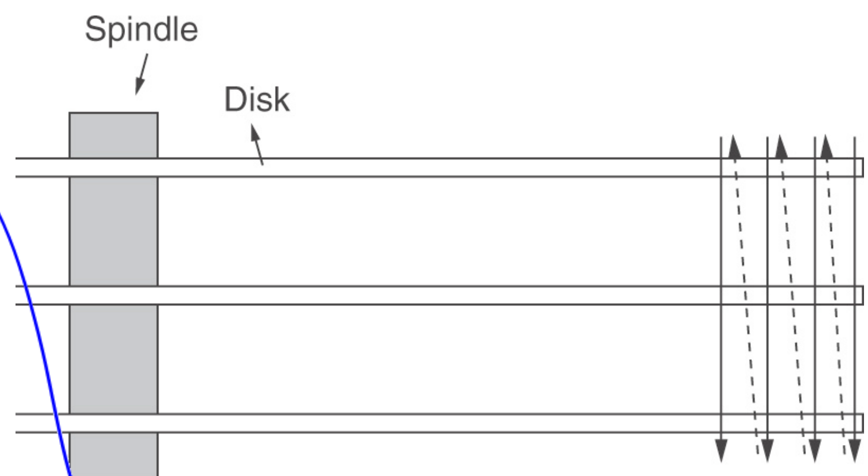
head switch time

~ 1ms

cylinder switch time

~ 2ms

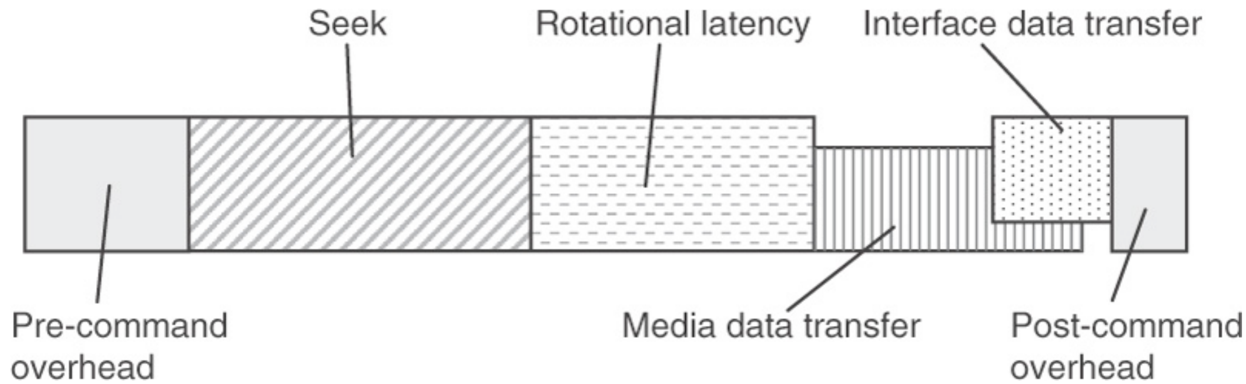
\# of times to switch between heads



$SDR = \frac{6 \times 1000 \times 512}{6 \times 60/7200 + 5(1ms) + 2} \approx 54$ MB/s

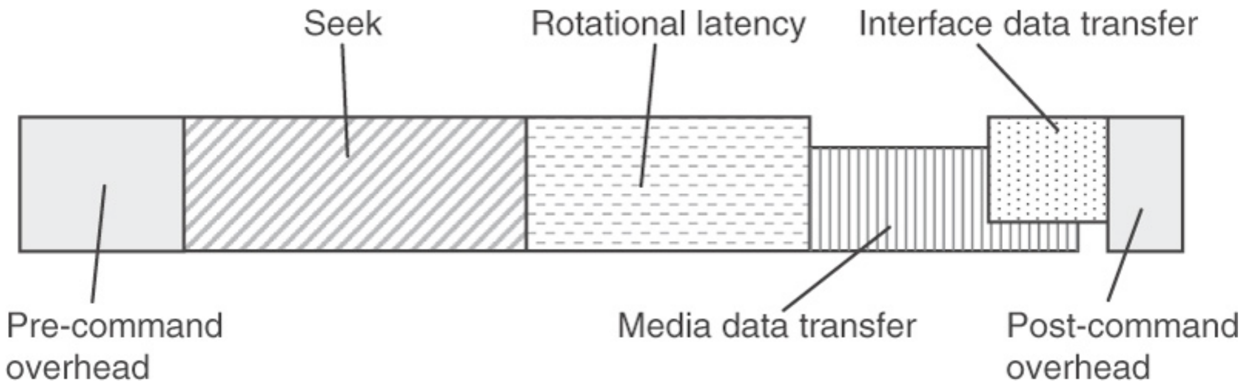
[18.5,18.3]

Time Components of an I/O Read



- **Interface data transfer**
  - **Time to transfer data over the interface**
  - **Faster than media data rate for SATA, SAS, USB 3.0**
  - **Transfer of all but last sector can be overlapped with media data transfer**
    - **Each sector is buffered in its entirety, and then ECC checked**

Time Components of an I/O Read



- **Post-command overhead**
  - **Signal completion back to the host**

[19.1]

Random Read Time

- Time to perform one random read

  overhead + avg seek time + 1/2 RL + media xfr + IF xfr

- Assume
  - 4KB transfer ↙ amount of data we want
  - fancy seagate cheetah! ⤳ 10K rpm drive ⇒ 1/2 RL = 3ms
  - Average seek time = 4.5ms
  - 50MB/s media xfr ⇒ 0.08ms to transfer 4KB
  - 100MB/s interface xfr ⇒ 0.005 to transfer last sector
  - Overhead = 0.3ms

    derivation in the book

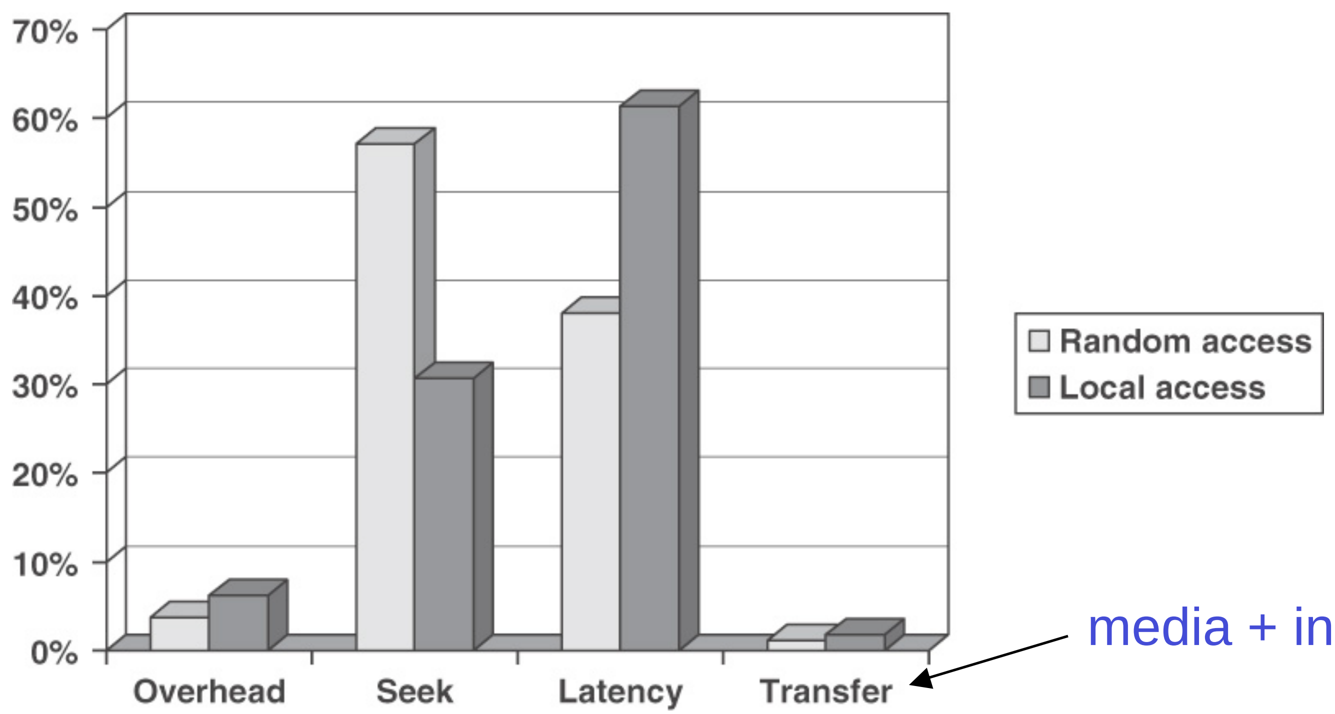
0.3ms + 4.5ms + 3ms + 0.08ms + 0.005 = 7.885ms

Local Read Time

- **Access to user data located in a small area of the disk may have 1/3 the average seek time**

  0.3ms + 1.5ms + 3ms + 0.08ms + 0.005 = 4.885ms

- **Overhead of various read time components**



media + interface

[19.3]

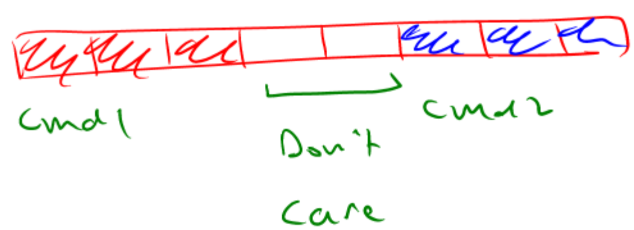
Sequential Read Time

- Address of next read immediately follows that of the prior read
  - seek + rotational latency = 0
  - 0.3ms + 0.08ms + 0.005 = 0.385ms
  - Media data rate has larger impact

overhead is now a really big contributor

- ***Near sequential***: Two otherwise sequential cmds are separated by a few other cmds

- ***Skip sequential***: Two cmds are separated by a few sectors (in the positive rotational direction)



Performance Impact of Disk Parameters

- **Track capacity (bpi)**

- **Track density (tpi)**

- **Cylinder capacity**

- **Number of heads**

Track Capacity (bpi)

- ↑ SPT, ↑ in media data rate → more data / inch, same linear speed

$$SDR_{Serp} = \frac{SPT \times 512}{60/rpm + T_{CS}} \qquad SDR_{Cyl} = \frac{N \times SPT \times 512}{(N \times 60)/rpm + (N-1) \times T_{HS} + T_{CS}}$$

- ↑ SPT, ↓ track switching overhead → more data, less likely to switch tracks

number of requested sectors

$$T_{S_{Serp}} = \left(\frac{K-1}{SPT}\right) \times T_{CS} \qquad T_{S_{Cyl}} = \left(\frac{K-1}{SPT}\right) \times \left(\frac{(N-1) \times T_{HS} + T_{CS}}{N}\right)$$

probability of crossing a track boundary

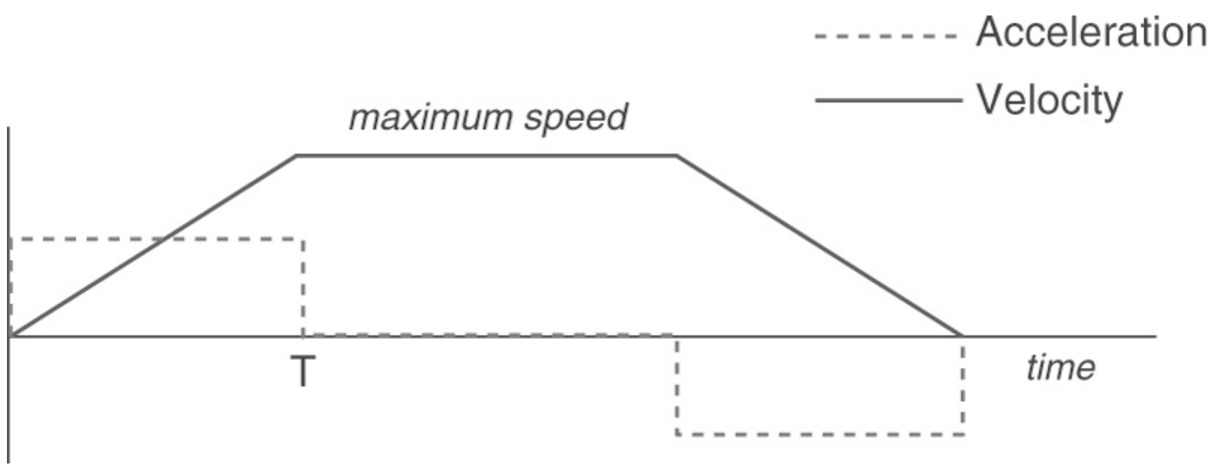
cylinder switch time (~2ms)

number of heads

head switch time (~1ms)

Track Density (tpi)

- ↑ tpi, ↓ seek *travel time*
  - Seek distance = (number of cylinders of seek)/tpi
  - Actuator accelerates to max speed, coasts for a while, and then decelerates



$$travel\ time = max\ speed / acceleration + seek\ distance / max\ speed$$

[19.5]

Track Density (tpi)

- ↑ tpi, ↑ seek *settle time*

- Actuator overshoots target track and wobbles back and forth (settles)

- Write settling more stringent than read settling

  → you're making changes, so be careful!

- More densely packed tracks, higher settle time
  - Settle time goes up logarithmically with tpi

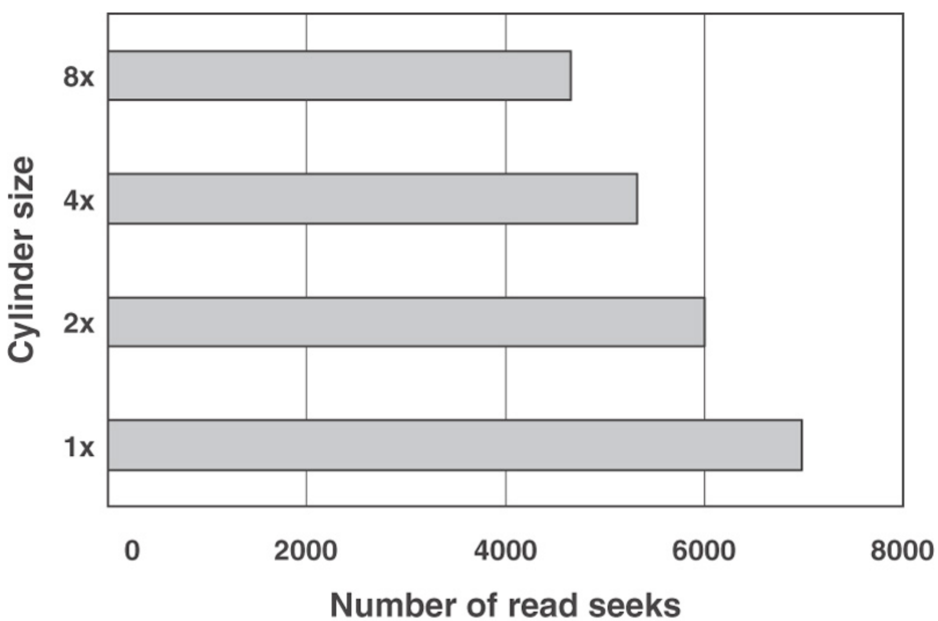
- For short seeks, ↑ tpi, ↑ seek time

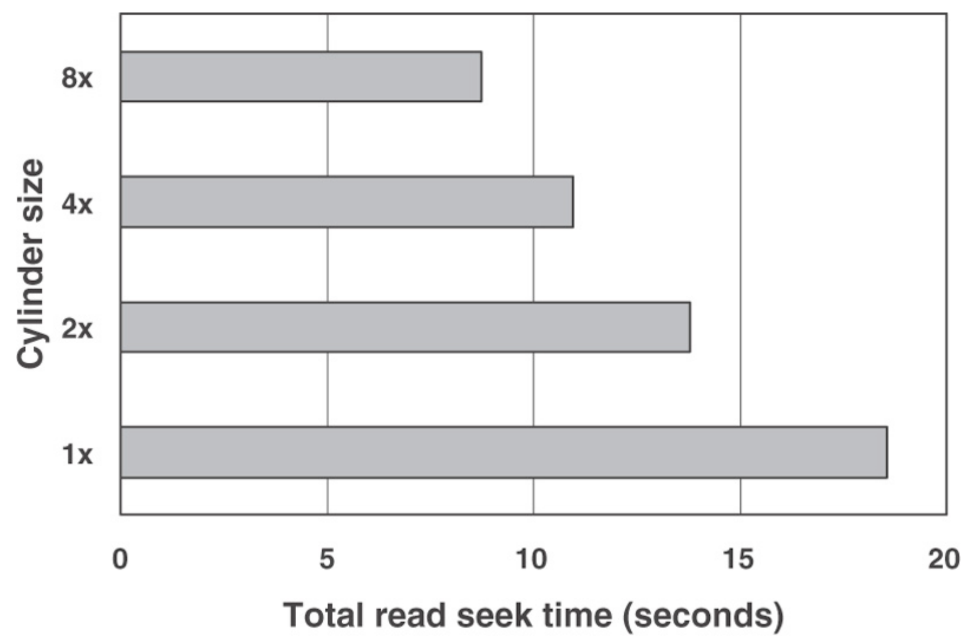
Cylinder Capacity

- ↑ sectors per cylinder, ↓ seek time
  - For cylinder mode (also helps banded serpentine)
- Assuming 64MB user file, probability that two different sectors are located on same cylinder?
  - 4 heads/cylinder, 300KB/track ⇒ 4×300K/64M = 1.9%
  - 8 heads/cylinder, 500KB/track ⇒ 8×500K/64M = 6.2%
  - Also, 67% fewer head switches with 8 heads, 500KB

Cylinder Capacity

- **Simulation results**
  - 900 sectors/track, cylinder mode
  - 10 PC applications with ~15K reads and 30K writes



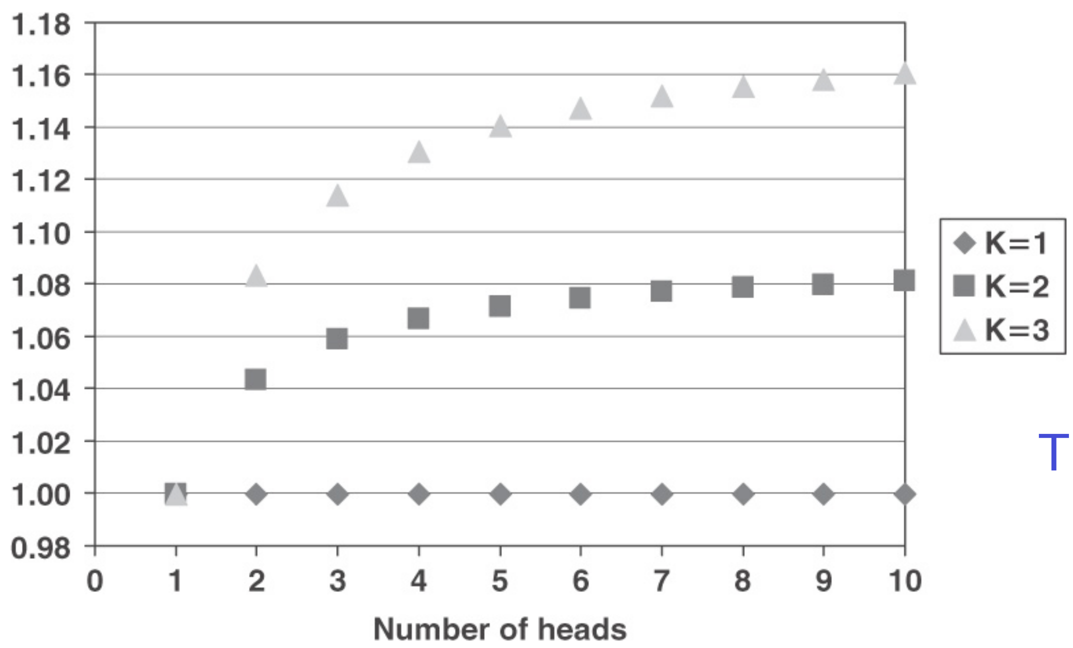


[19.4]

Number of Heads (Cylinder Format)

- **An increase in areal density permits fewer platters and heads (↓ cost) for the same capacity**

- **tpi improvement impact on data rate**
  - **Fewer head switches and more cylinder switches**
  - **Since K = T_{CS}/T_{HS} > 1, lower sustained data rate**



$T_{HS} = .1 \times RL$

[19.7]

Next Time

Command Scheduling
Data Relocation