ECE 5730
Memory Systems
Spring 2009

Hard Disk Drives

Cornell University

Announcements

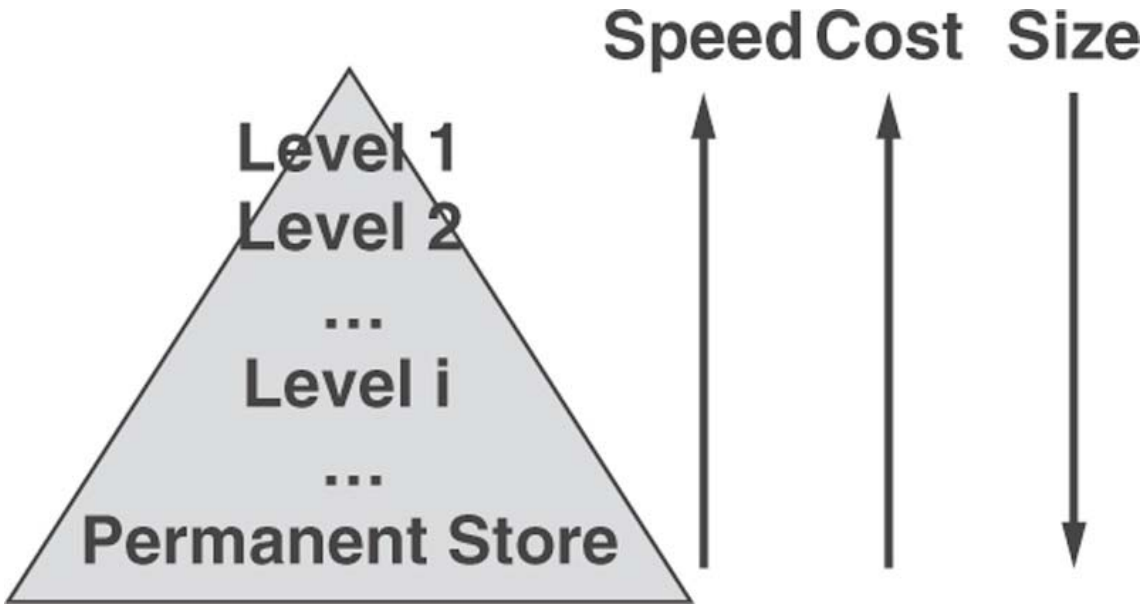
- **Quiz averages**
  - **Quiz 10 average = 8.1**
  - **Quiz 11 average = 10.0**
- **Quiz 12 on Tuesday**
- **Status report due tomorrow, 5pm EDT**
- **Classes on 4/22 and 4/30, 6:00-7:15pm**
  - **Pizza**
  - **No material will show up on a quiz or the exam**

Announcements

- **Exam II**
  - **Scheduled for 4/29, 6:30-9:30pm**
  - **Alternative: 5/7, 7:00-10:00pm**

Recall the Memory Hierarchy

- **Multiple levels of memory, each optimized for an appropriate cost/performance design point**



Recall the Memory Hierarchy

- **Multiple levels of memory, each optimized for an appropriate cost/performance design point**

Technology	Bytes per access	Latency per access	Energy per access	Cost per MB
On-chip cache	10	100's of ps	1 nJ	$1-100
Off-chip cache	100	ns	10-100 nJ	$1-10
DRAM	1000 (internally fetched)	10-100 ns	1-100 nJ per device	$0.1
Disk	1000	ms	100-1000 nJ	$0.001

Hard Disk Drives

~1 ton, like a refrigerator

~50 platters, 2ft. diameter

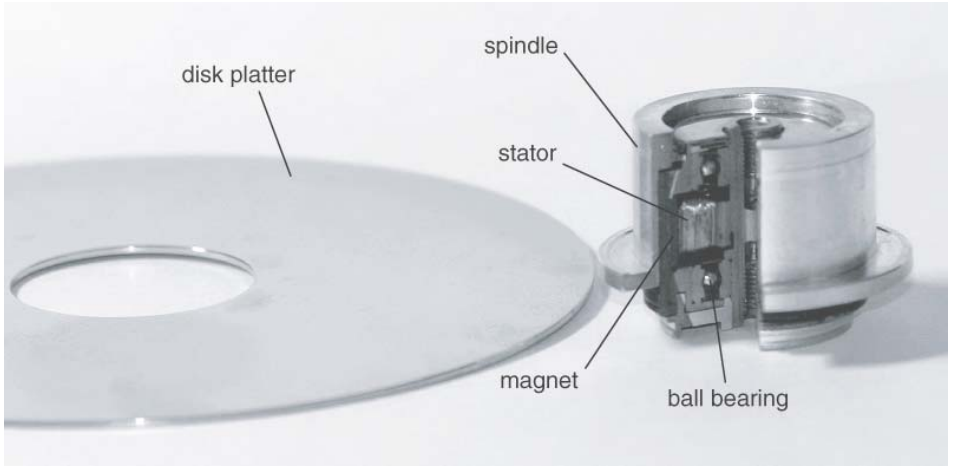
- **Permanent backing storage**
- **Introduced in 1956 in the IBM RAMAC 305** ⇒ accounting machine
- **Incorporated into original PCs**
- **Many applications**
  - **Servers, PCs, laptops**
  - **DVRs, video consoles, network routers**
  - **MP3 players, digital cameras, cell phones**
    - **Now flash is taking over**

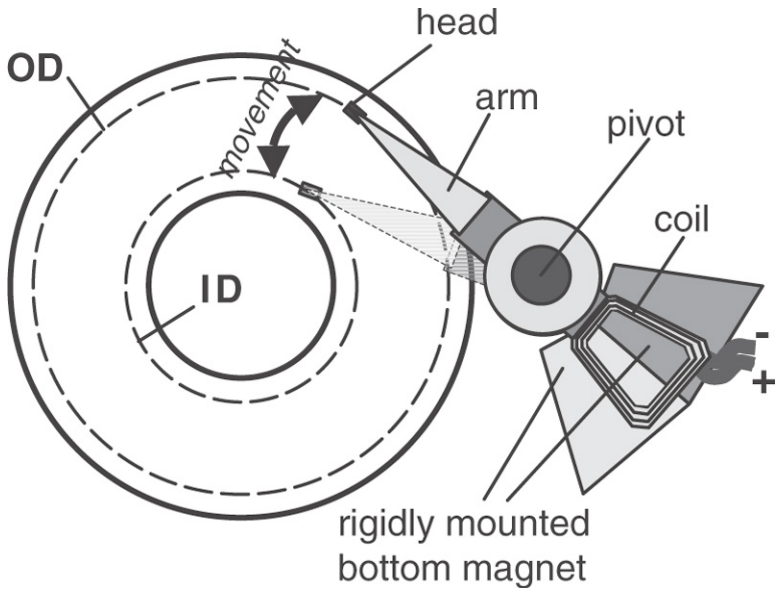
IBM RAMAC 305

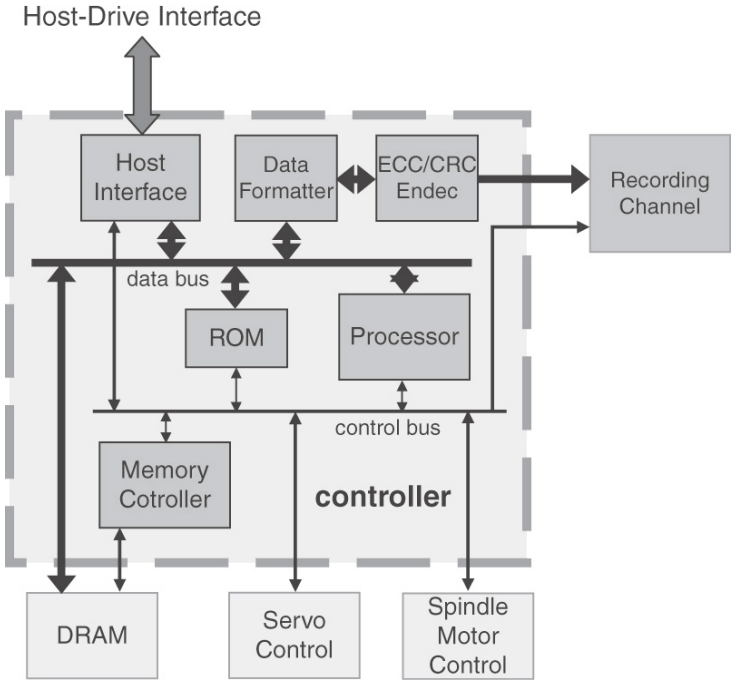
Major Hard Drive Components

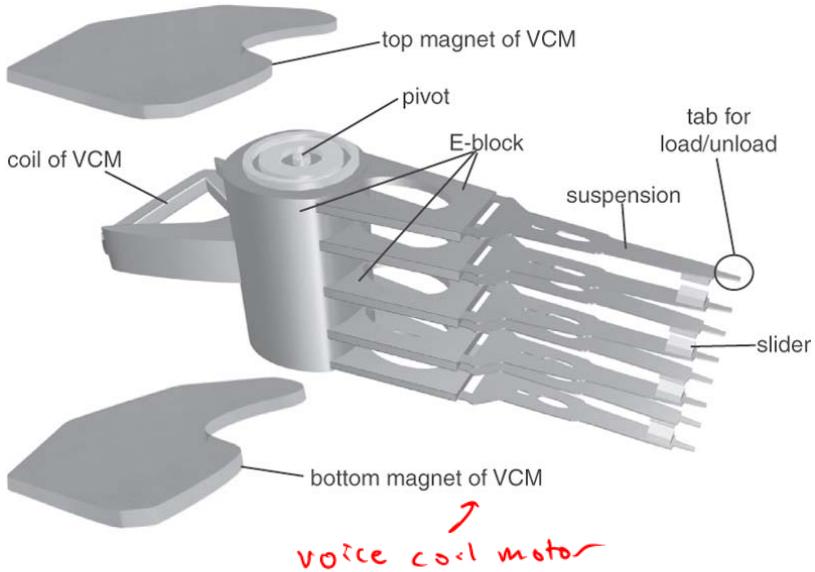
- **Platters: Media that holds the recording material**
- **Spindle motor: Rotates the platters to the desired disk position**
- **Head assembly: Arm with transducers (*heads*) that convert media signals to electrical signals**
- **Controller: Receives commands from the drive interface and coordinates disk actions**

Major Hard Drive Components









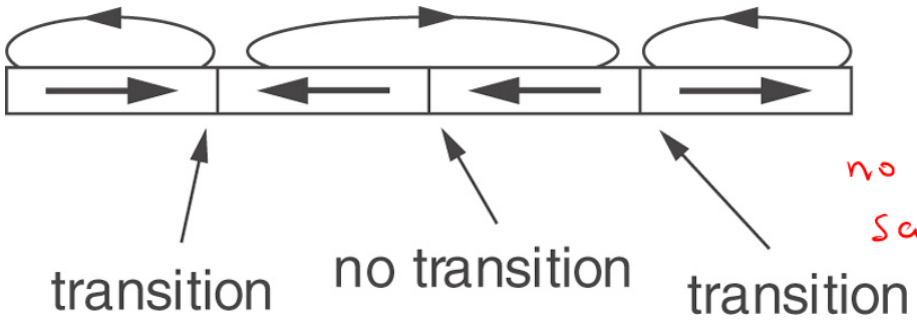
[17.10,17.23,17.24,17.30]

Writing Data on a Disk

- **Disk surface is coated with a ferromagnetic material (can be permanently magnetized)**

- **Data pattern is formed by storing magnetic charges of different polarities**

- **Binary value represented by presence or absence of *transition* to different polarity**

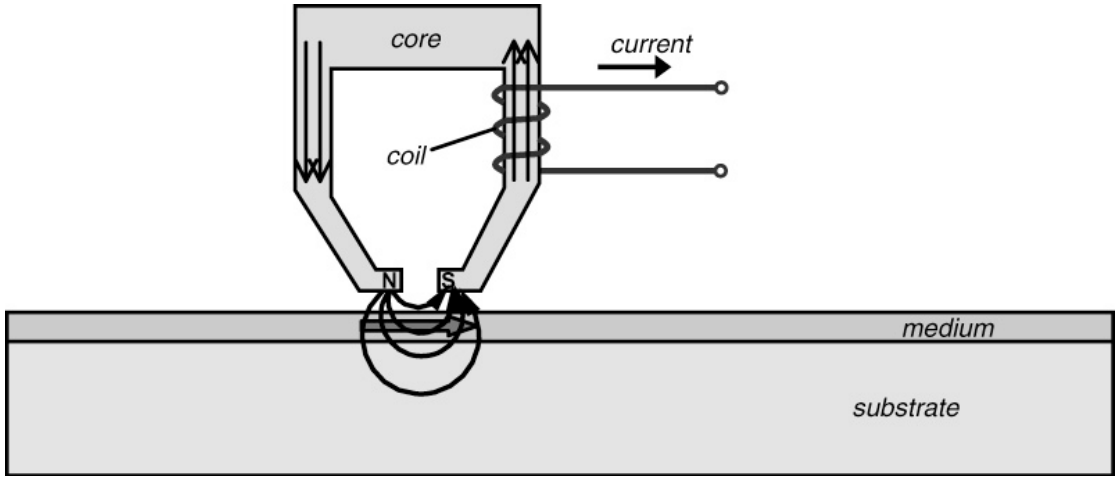


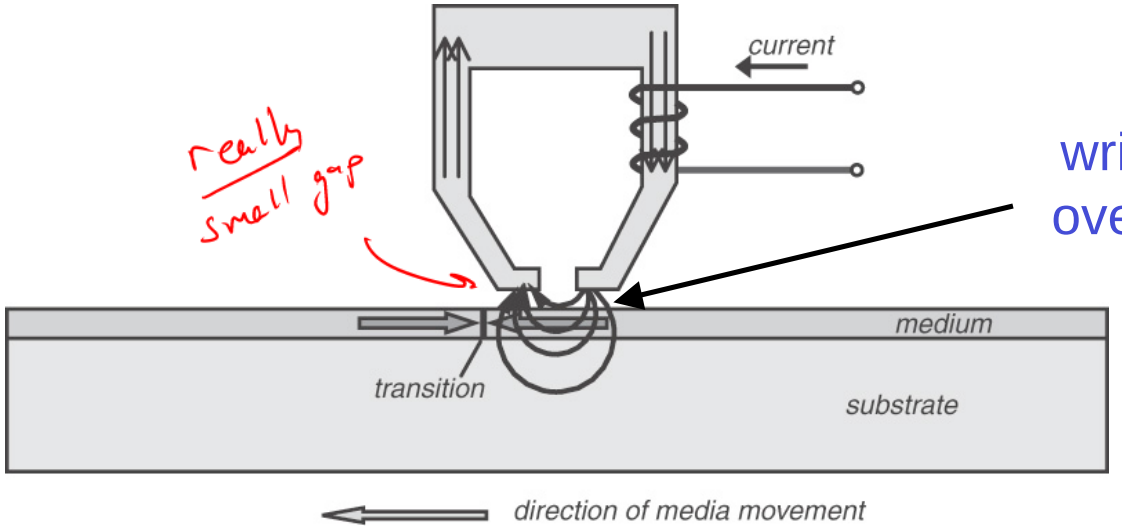
binary value = 101

no real demarcations. Samples at periodic intervals to detect presence or absence of transitions

[17.6,17.7]

Writing Data on a Disk





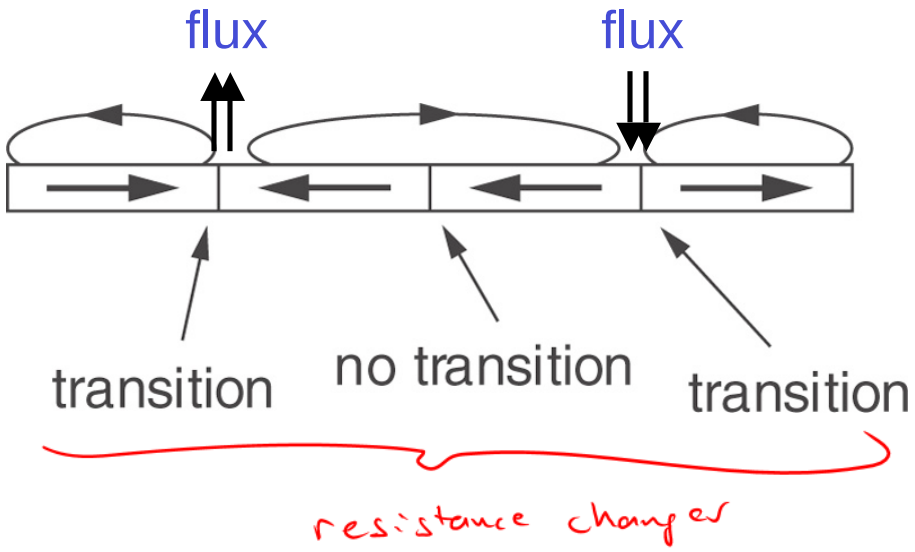
write head *flying* over disk surface

[17.12,17.13]

Reading Data from a Disk

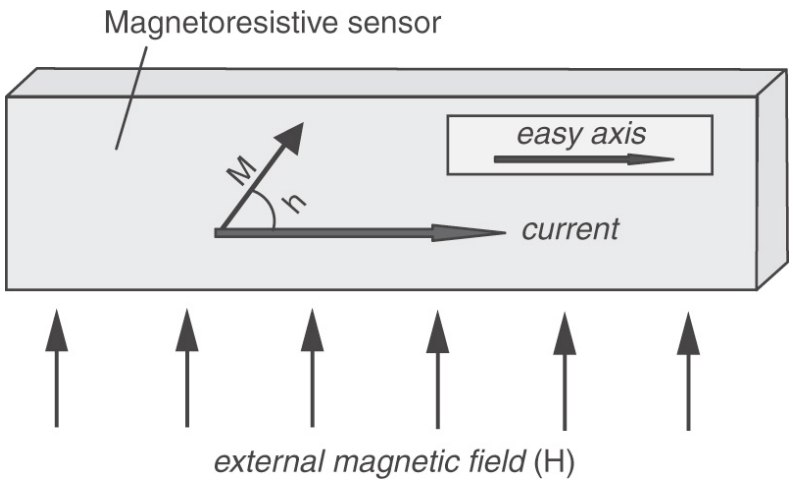
- **Magnetoresistive materials change their resistance depending on magnetic flux**

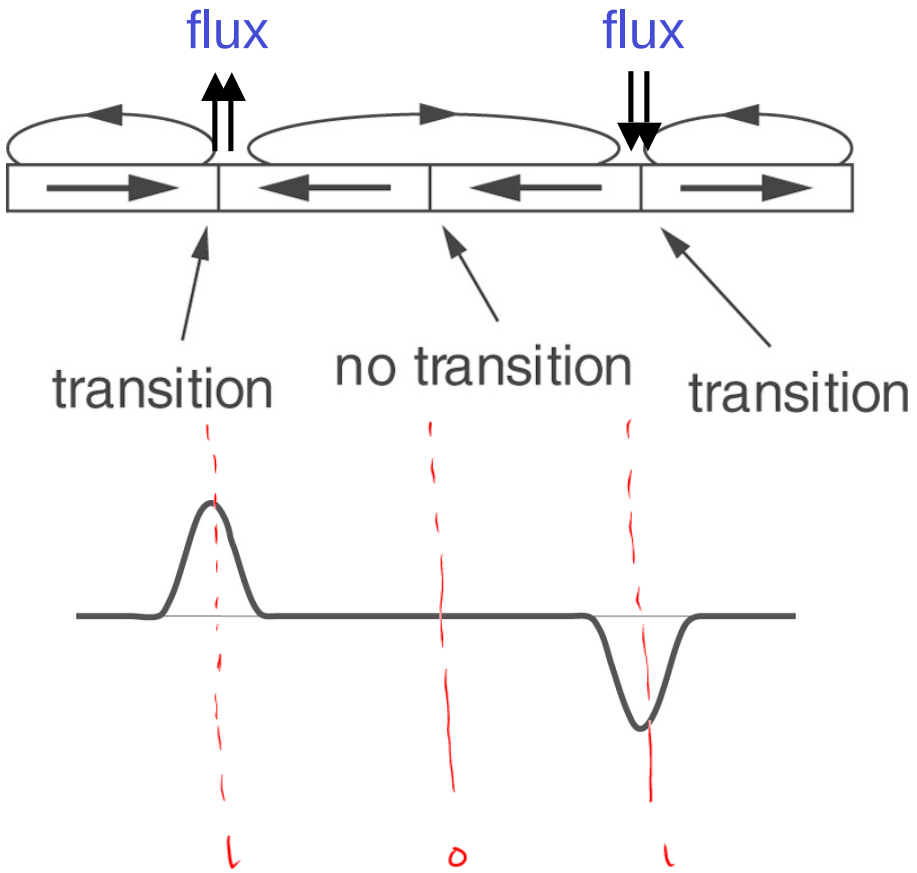
- **Magnetic flux is strong at transitions**



Reading Data from a Disk

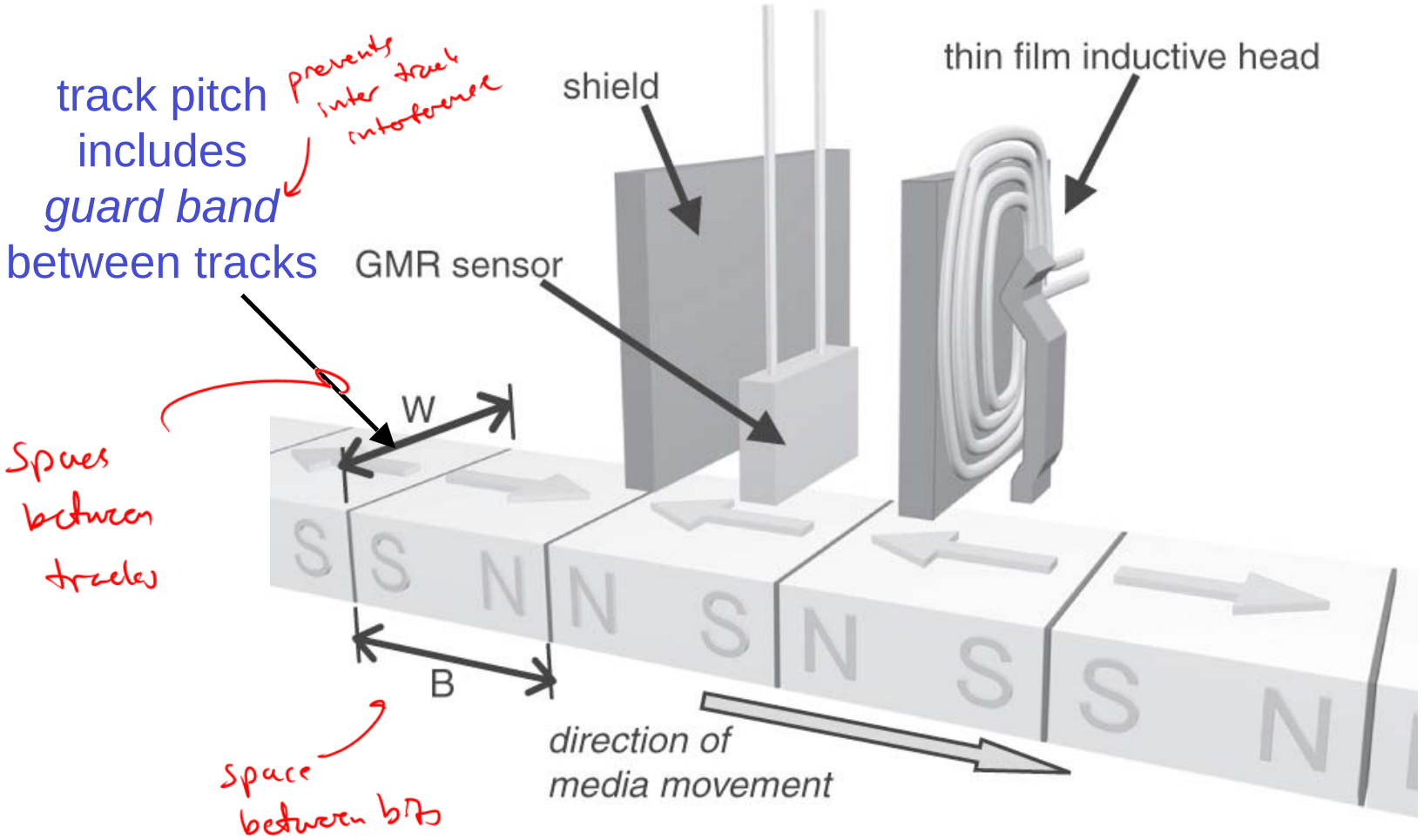
- **By driving current through the material, can detect voltage changes at transitions**





[17.7,17.16,17.17]

Track Pitch and Bit Pitch

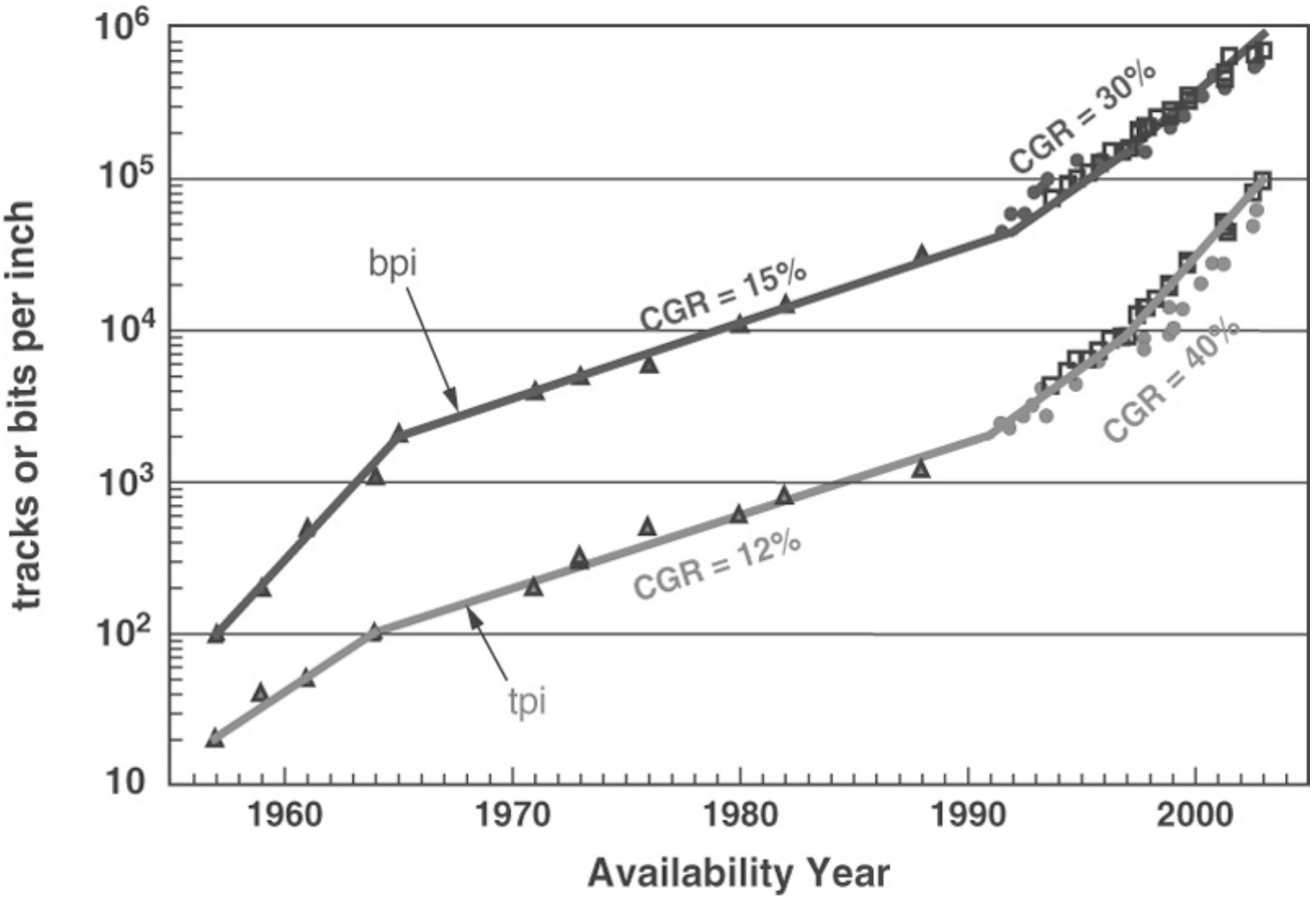


Areal density depends on *tracks per inch* and *bits per inch*

data / in²

TPI and BPI Growth Trends

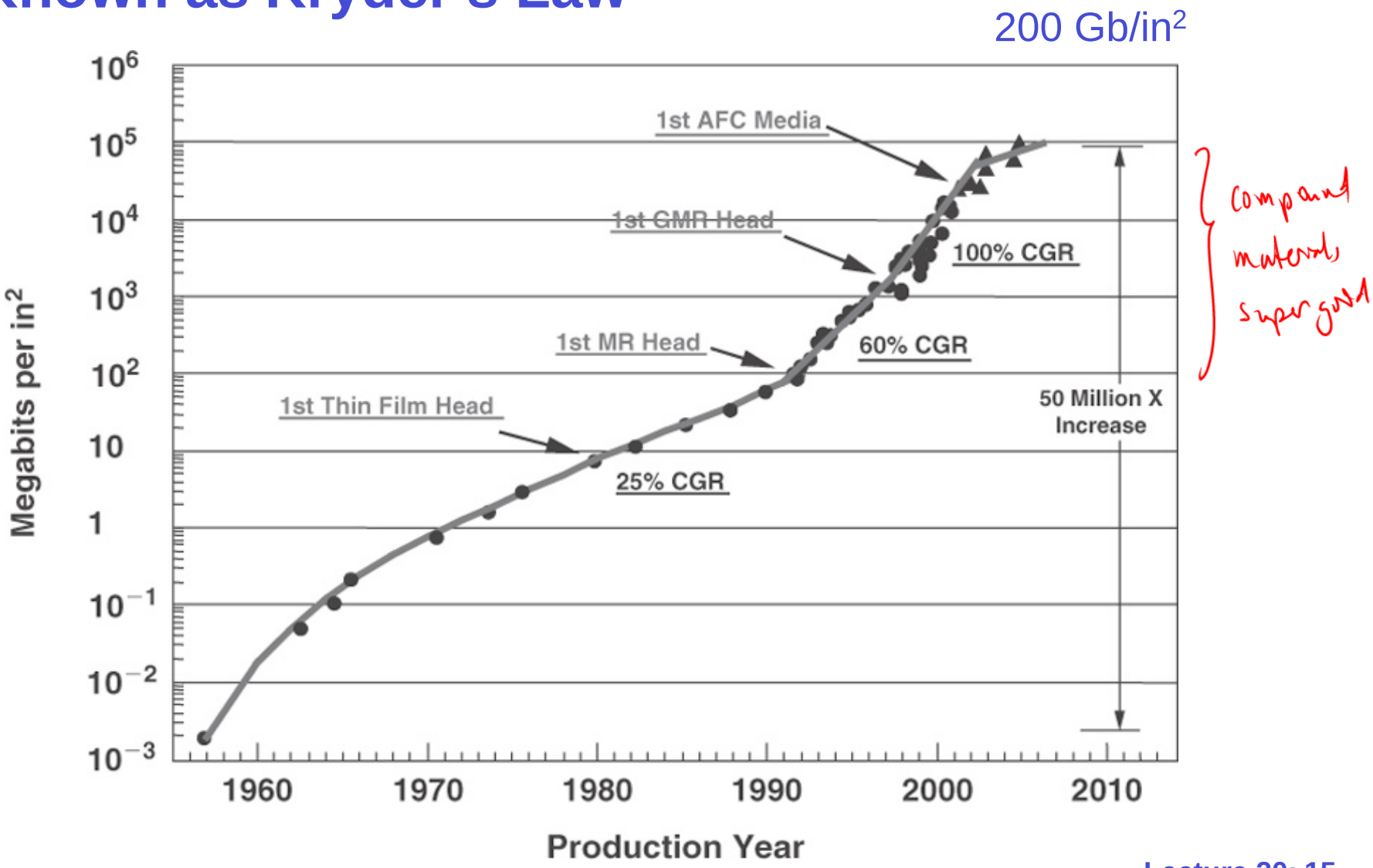
- **Exponential like Moore's Law!**



[16.5]

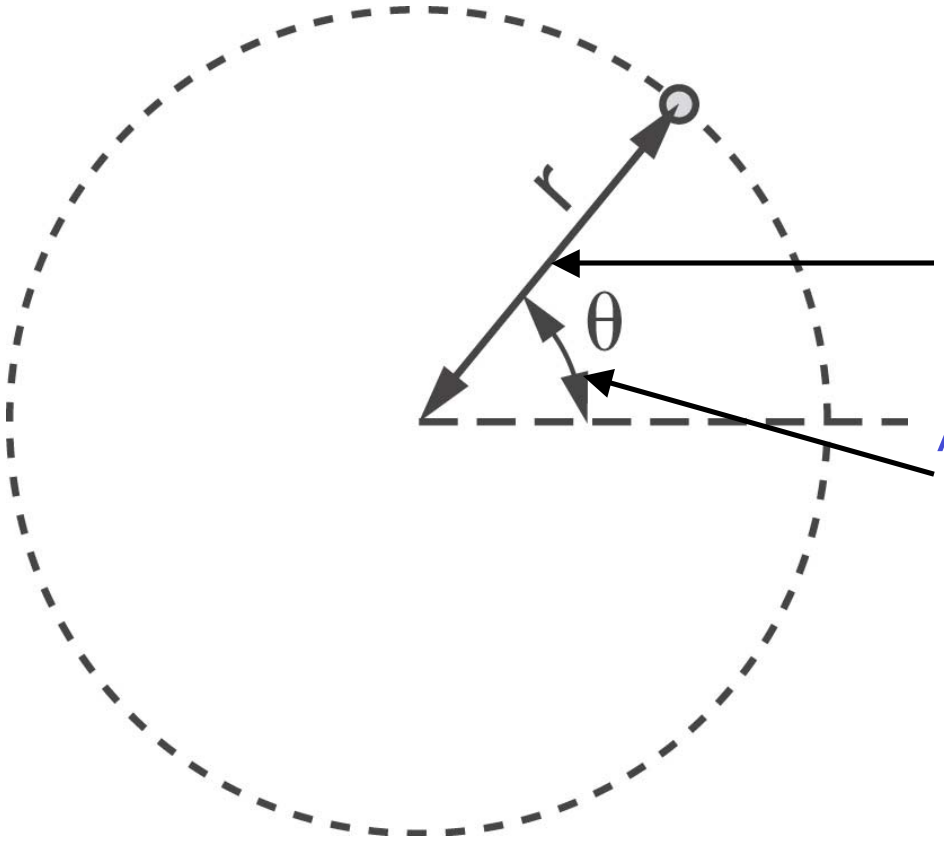
Areal Density Growth Trend

- **Known as Kryder's Law**



[16.4]

Locating Data on a Disk Surface



radial distance from center

Angular position from fixed reference axis

Radial Coordinate System

Different ways to Increase Capacity

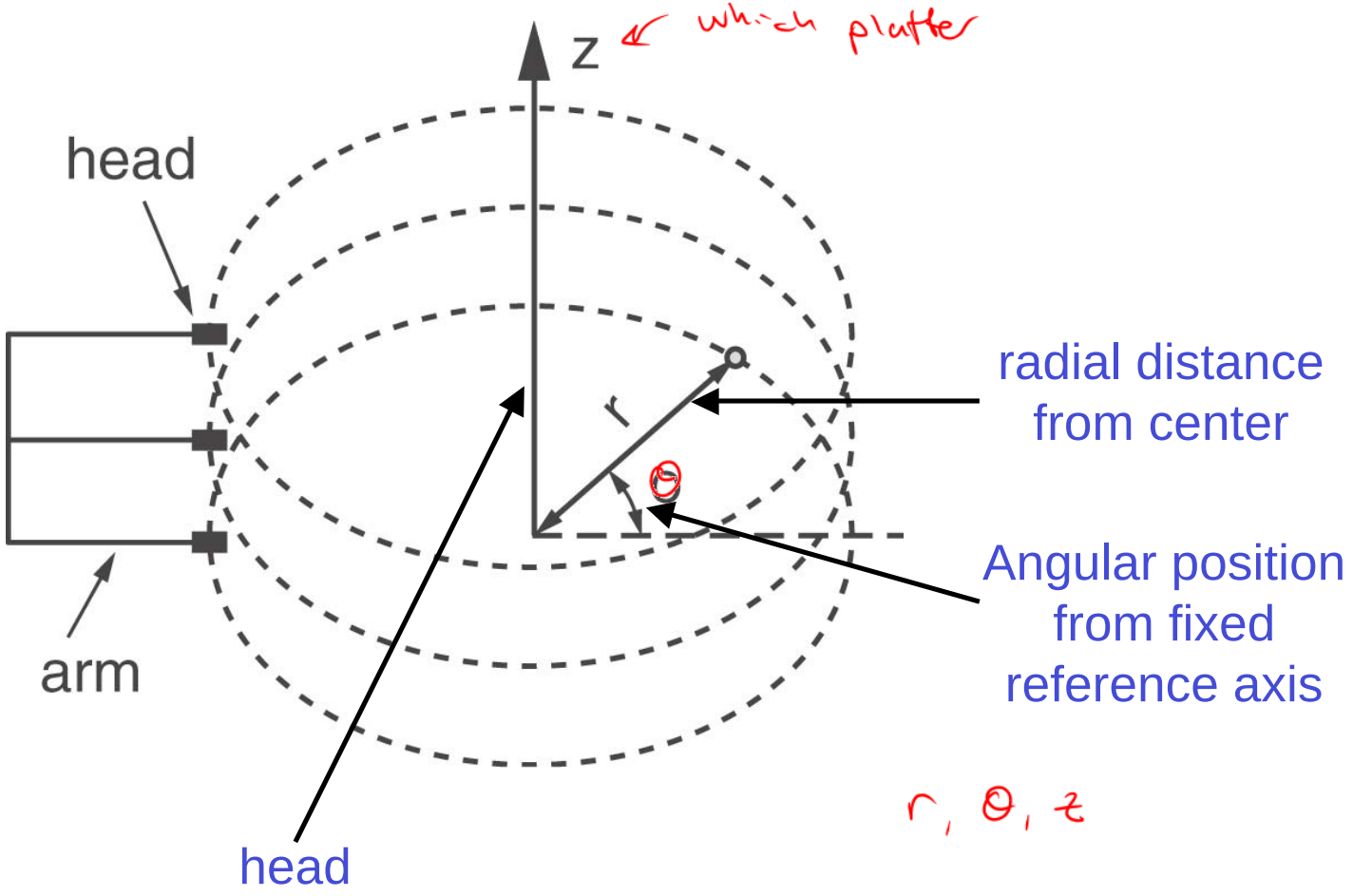
Multiple Platters

- **Increasing disk diameter to increase capacity**
  - **Longer seek distance requires thicker arms**
  - **More air friction with larger surface**
  - **Disk must be thicker to obtain necessary stiffness**
  - **⇒ Longer seek time or more required power**

- **Better choice is multiple smaller platters**
  - **Shorter seek distance**
  - **Less weight increase from multiple thinner platters than one thick platter**
  - **But requires more heads (expensive component)**

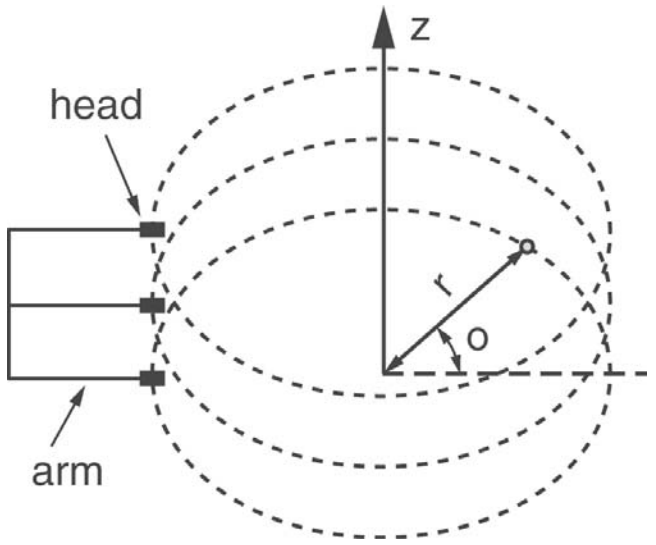
Multiple Platters

- **Data maps to a three dimensional space**



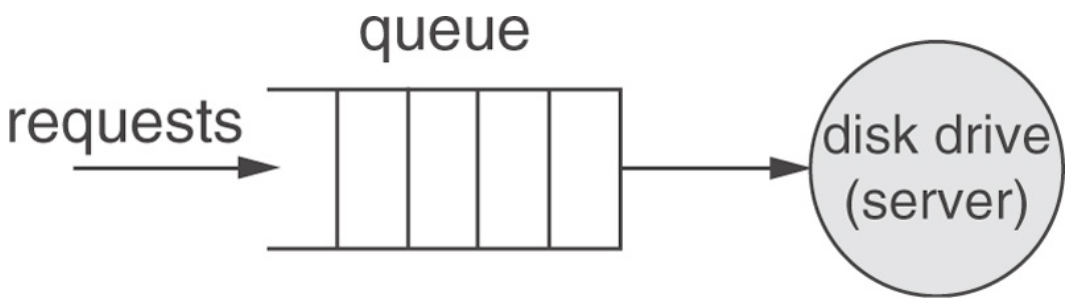
Reading Data from a Disk Drive

- Host sends command over interface
- Controller orchestrates operation
- Head moved to the radial position (*seek*)
- Electric motor rotates the disk platter, passing the head over the desired data
- Data is sensed, converted, and passed to the controller
- Controller delivers data over the interface

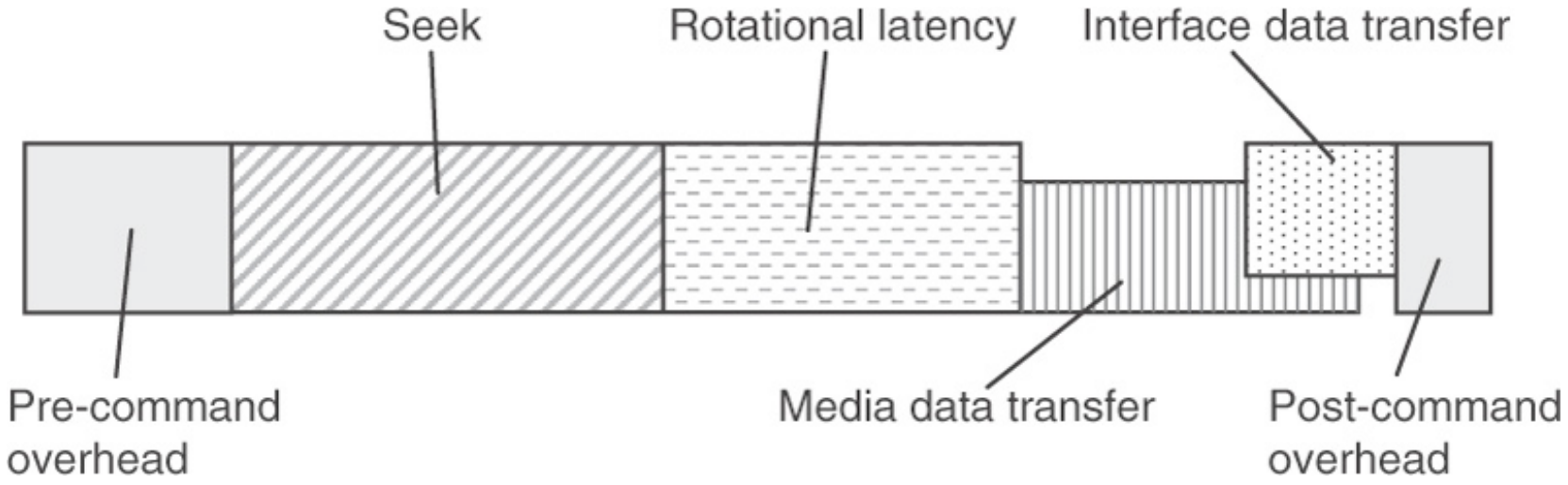


Disk Performance Overview

- **Simple disk drive model**



- **Time components of a read**



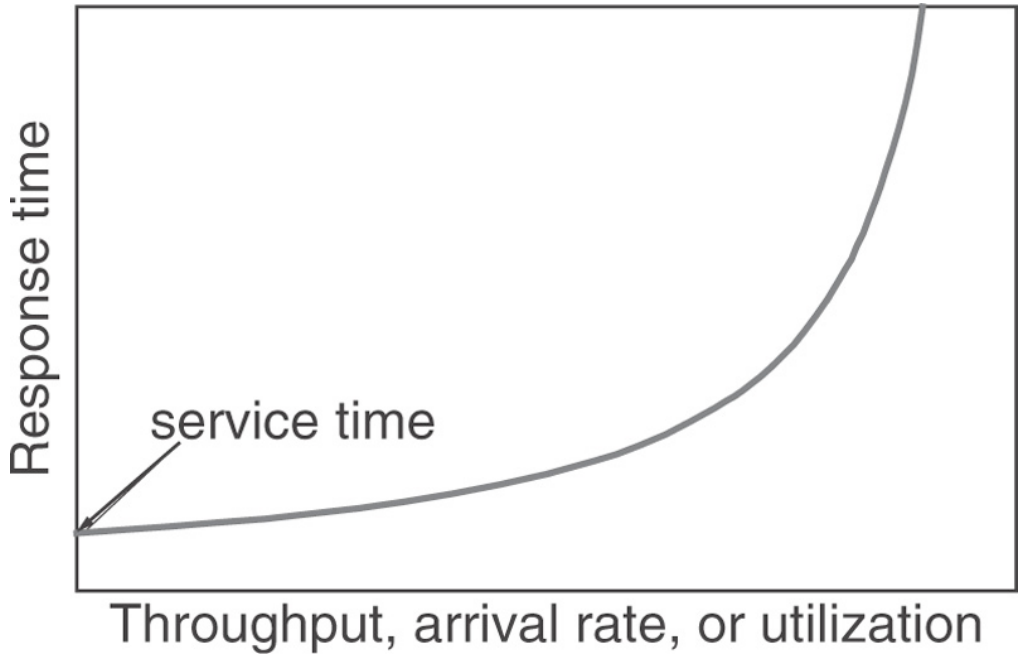
[16.9,19.1]

Disk Performance Overview

- **Factors influencing disk performance**
  - **Access patterns (sequential, random, streams)**
  - **Command arrival rate**
  - **Read/write mix**
  - **Data footprint**
  - **Block size**
  - **Command queue depth**
  - **Latency and transfer rates of disk components**
  - **Management of disk components**

Disk Performance Overview

- **Performance metrics**
  - ***Response time*: Time between I/O command issue and completion of data transfer**
  - ***Throughput*: rate of data transfer (MB/s) Simple disk drive queuing model**
- **Response time versus disk drive utilization**

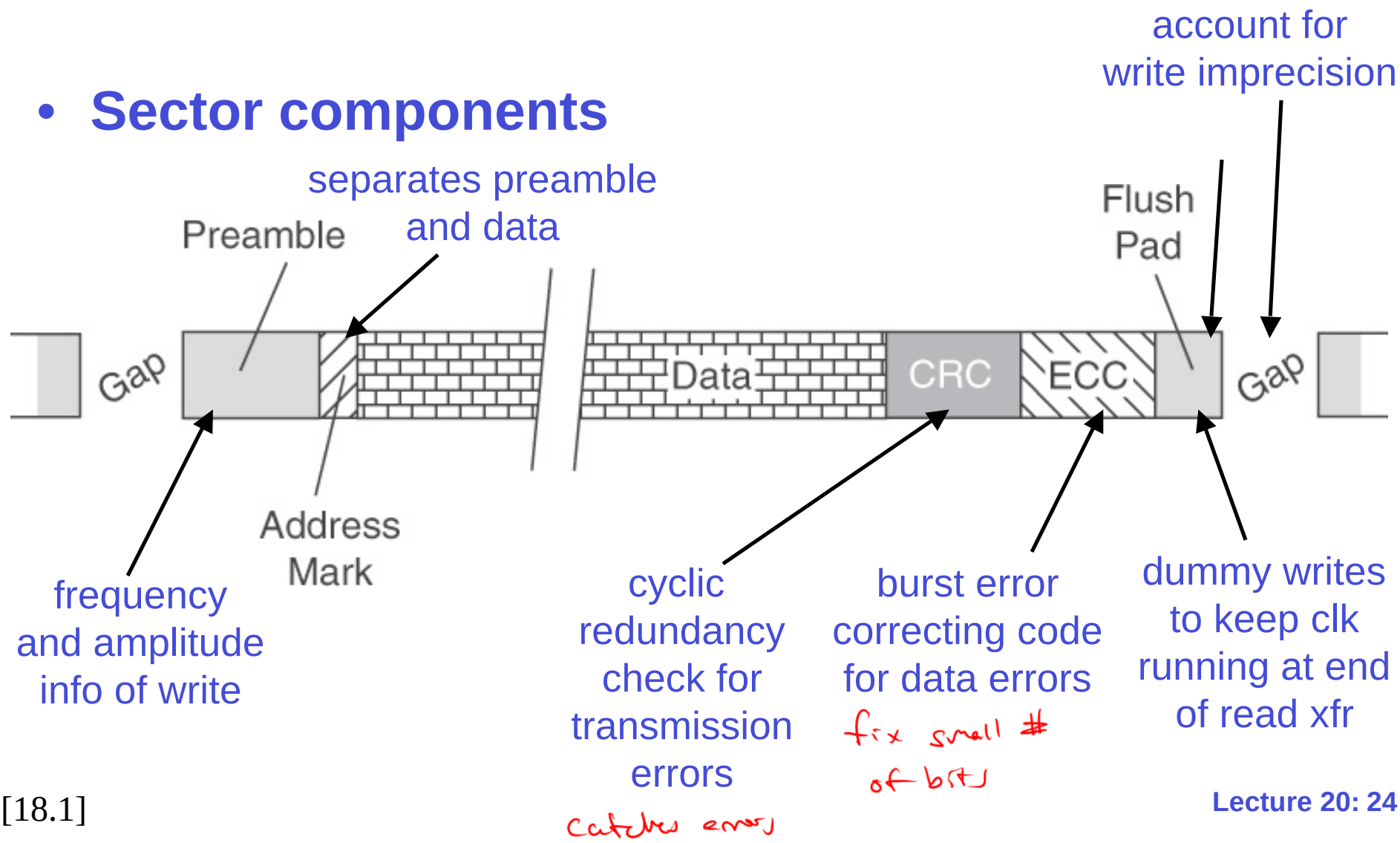


Disk Blocks

- Disk storage space is partitioned into *blocks*
- Most systems use a fixed block size or *sector*
- Block (sector) size tradeoff
  - Smaller blocks have less internal fragmentation for small files
  - Large blocks have better sequentiality for large files
  - Large blocks allow more powerful ECC protection for the same amount of storage overhead
  - Moving to 1-4KB sectors in addition to usual 512B

Sector Organization

- **Disk drives used fixed block size or sector**
- **Sector components**

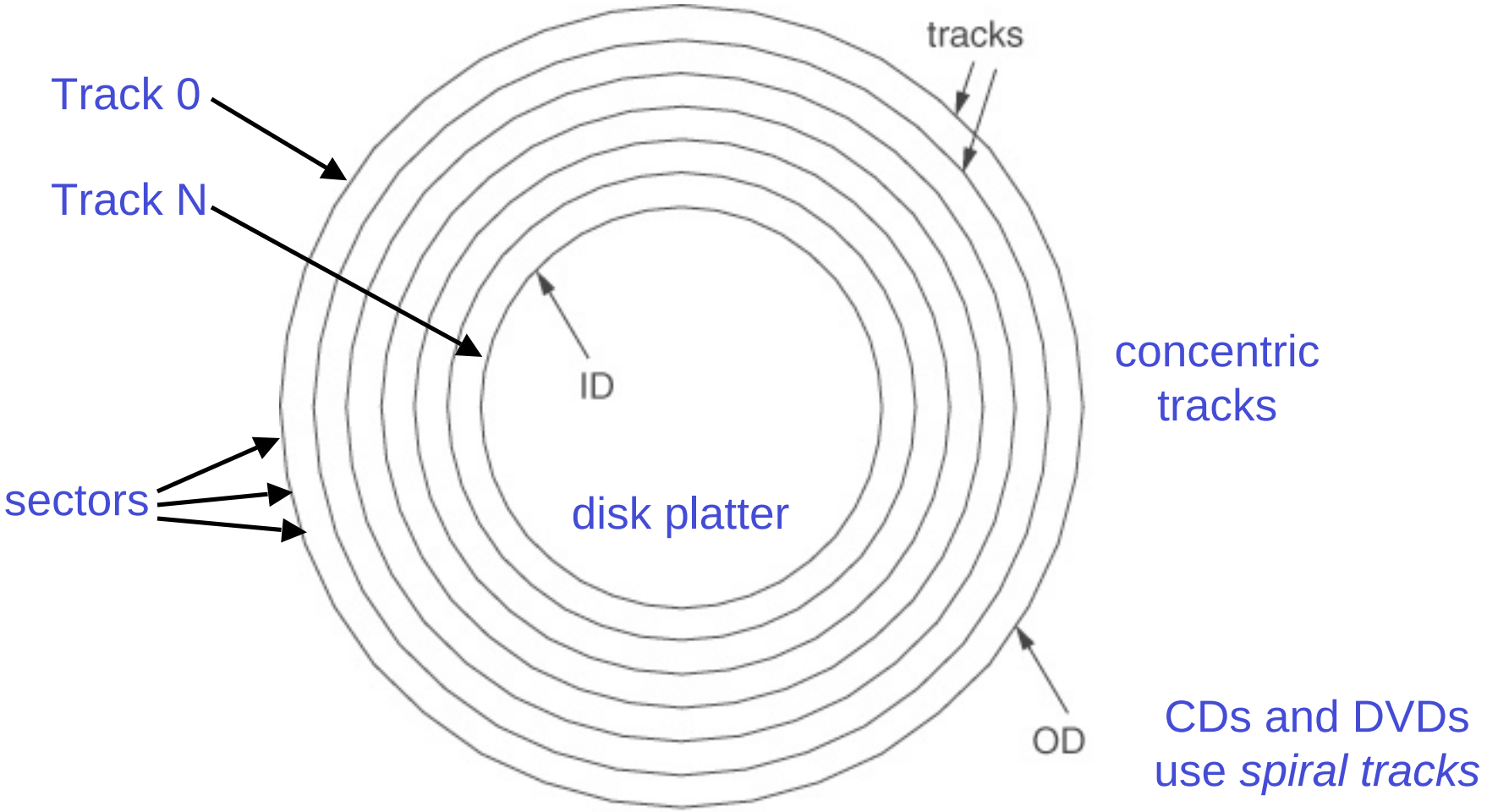


Sector Size Tradeoffs

- **Smaller blocks**
  - Less internal fragmentation for small files
- **Large blocks**
  - Better locality of access for large files
  - Allow more powerful ECC protection for the same amount of check bit overhead
- **512B has been standard sector size for years**
- **Recent OS's allow larger (1-4KB) sectors**

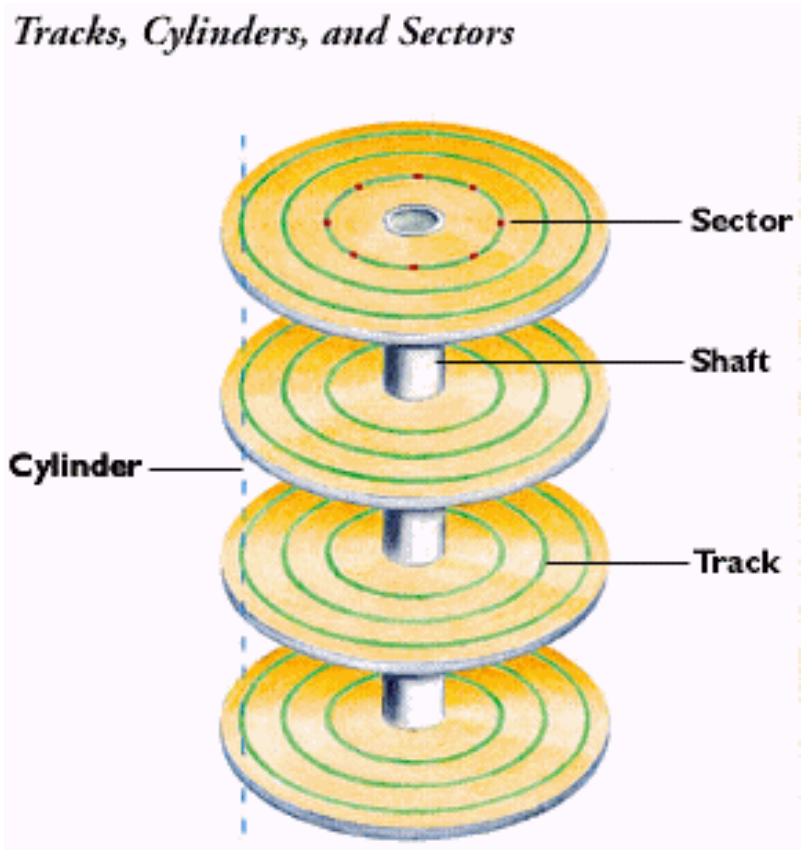
Tracks and Cylinders

- ***Tracks*: Circles containing sectors**



Tracks and Cylinders

- ***Cylinder*: All tracks with same track number on all all disk surfaces**
- **Cylinder 0 is first user cylinder**
- **Drive reserves first n cylinders for drive info**
  - ***Negative cylinders***



[pcguide.com]

Address Mapping

- **A disk drive is *internally* addressed using a *physical block address (PBA)***

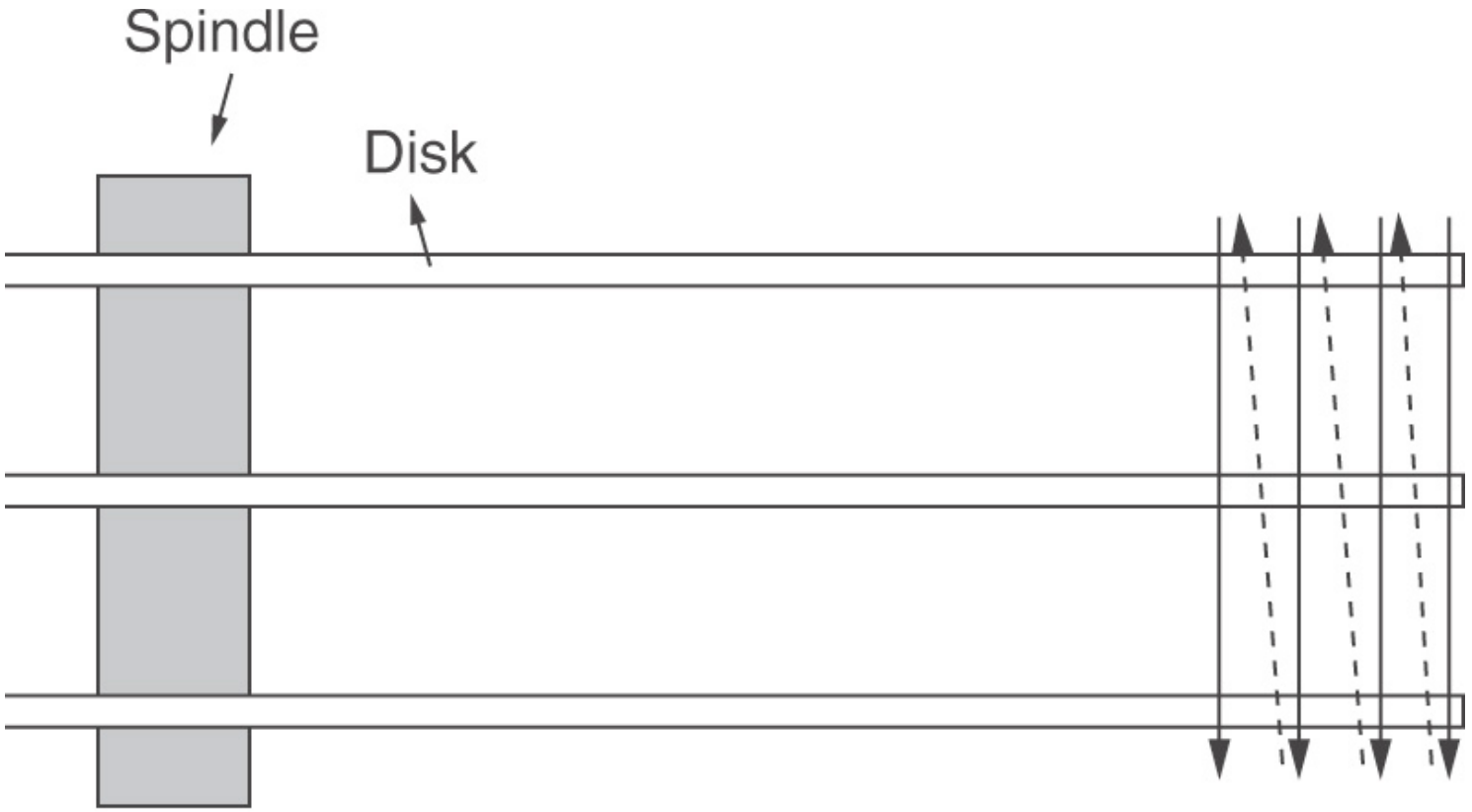
- **PBA consists of the cylinder, head, and sector**
  - ***CHS addressing***
  - **Location of the sector in three dimensional space**

Address Mapping

- External Logical Block Address (LBA) from the host gets mapped into the PBA
  - Necessary due to presence of defective sectors
- Logically sequential blocks are laid out physically sequential on a track
- Where get the next block when reach the end of a track?
  - Cylinder mode
  - Serpentine format

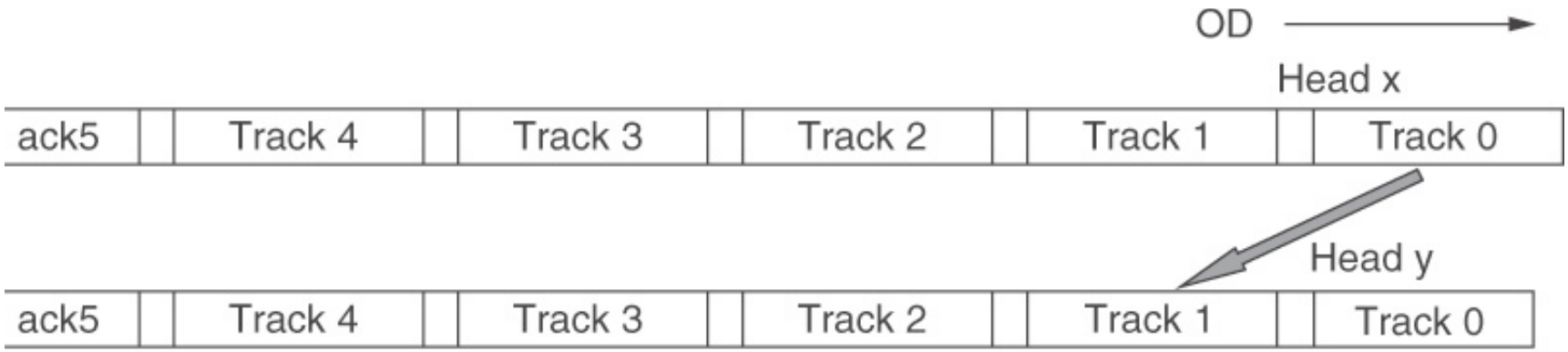
Cylinder Mode

- **Move in the z-axis direction**

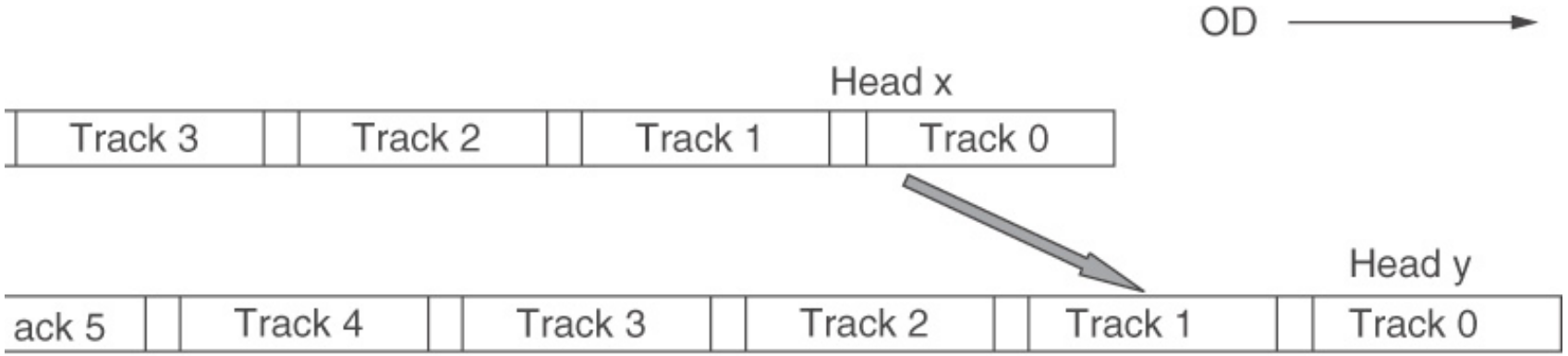


Cylinder Mode

- **Works well in theory if tracks are aligned**



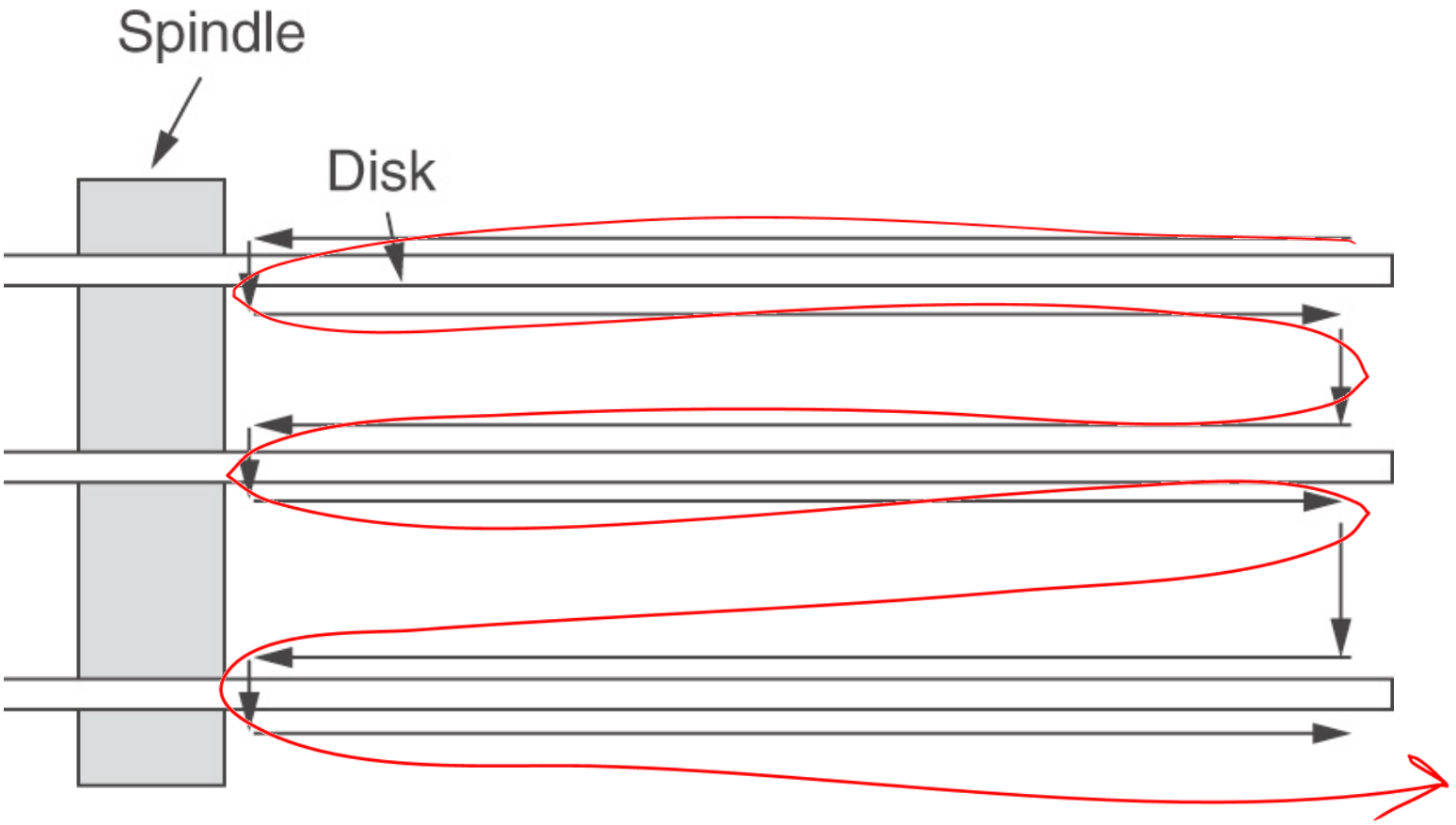
- **Situation with current high density drives**



difficult to align heads with high density drives

Serpentine Format

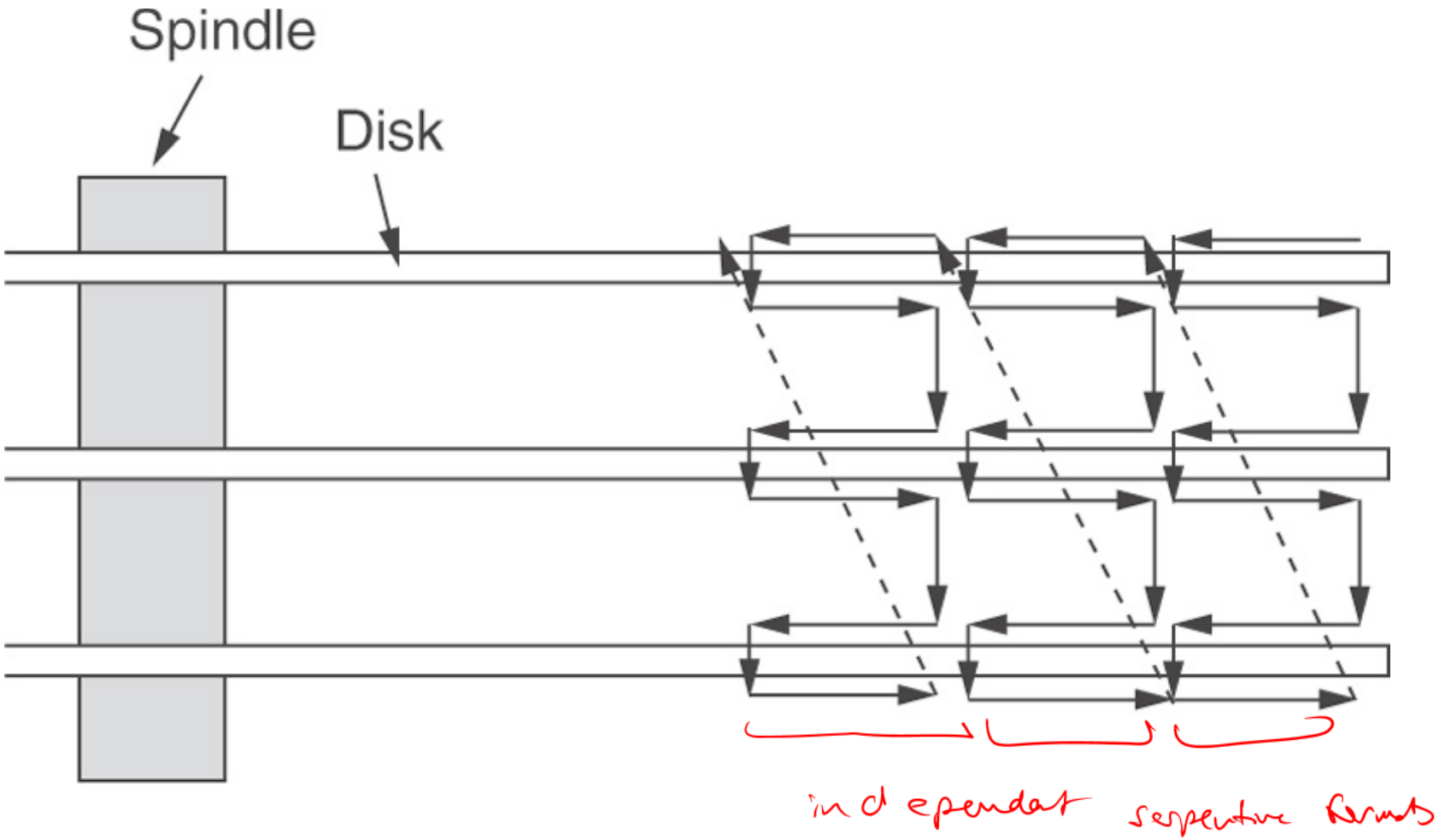
- **Move in radial direction**



May have to wait a full disk rotation before we can continue this pattern

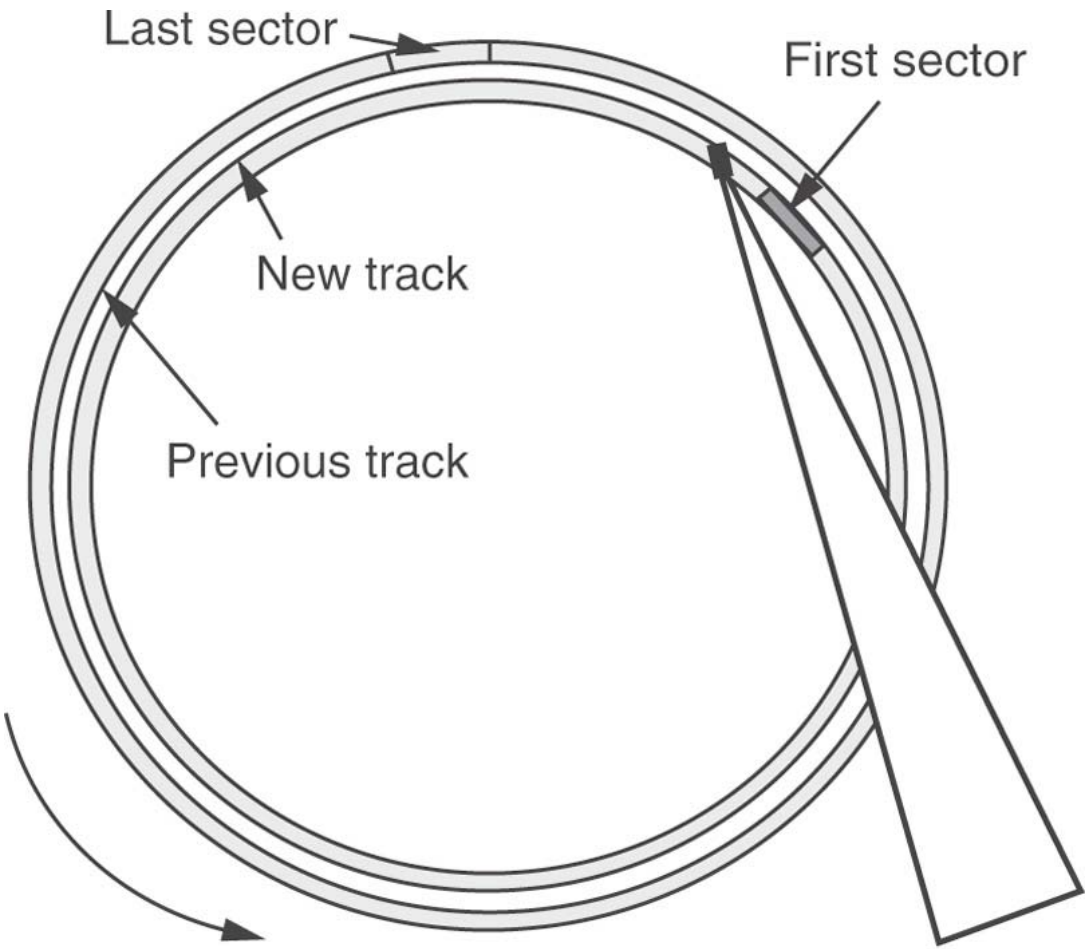
Banded Serpentine Format

- **Reduces seek distance for contiguous data**



Track Skew

- **Accounts for time to move head to new track**



as disk rotates

Next Time

Defeet

Defet Management

Drive Interfaces