Secondary Storage

Nonvolatile bulk memory

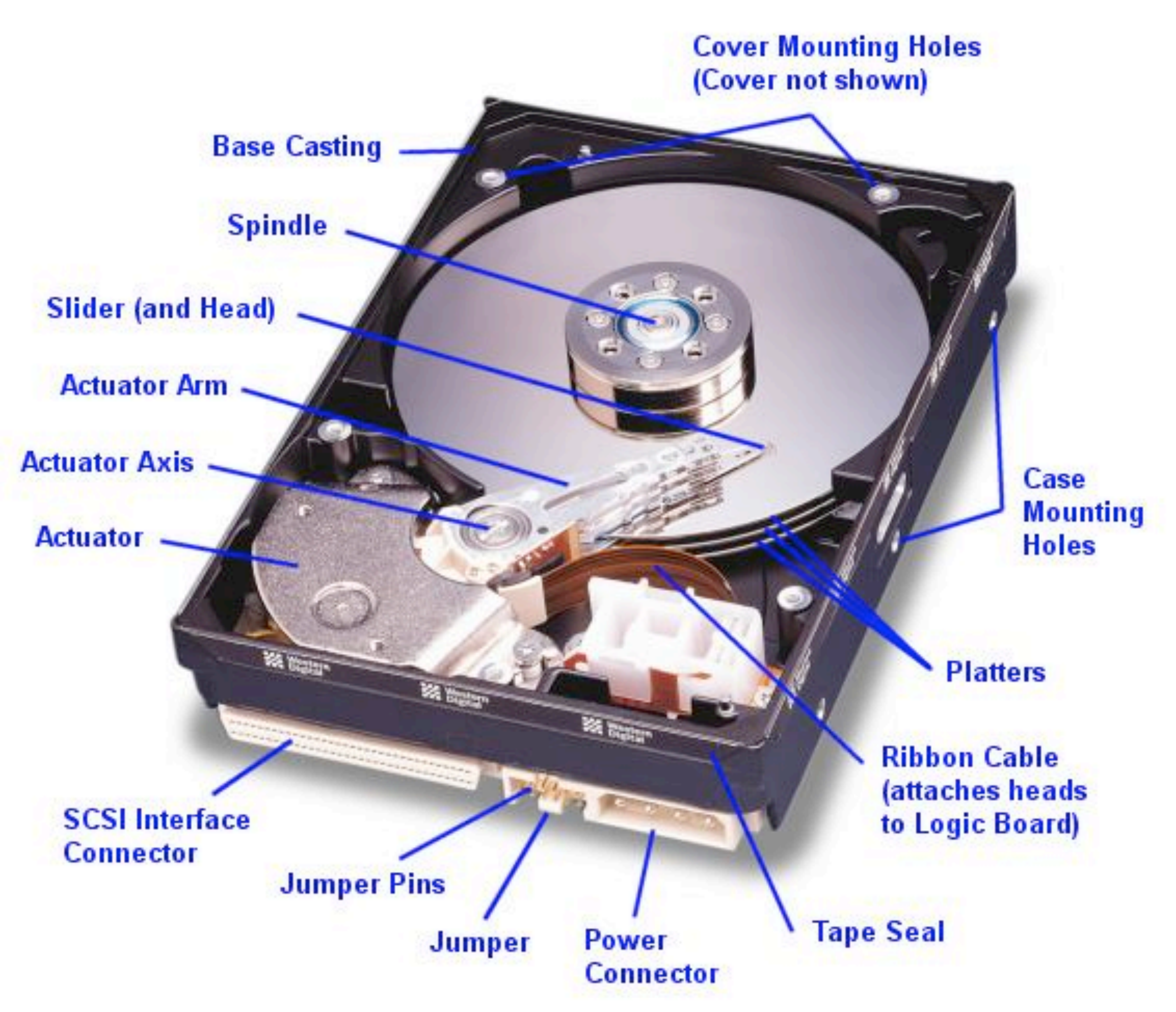
Variations

- Hard disk (HDD)
- Flash (SSD)
- Removable media (DVD, flash)
- Cloud storage (networked HDD, SSD)

Basic Concepts

- Rotating platters
- Moving heads on arms
- Uniform magnetic surface
- Data written as magnetic spots

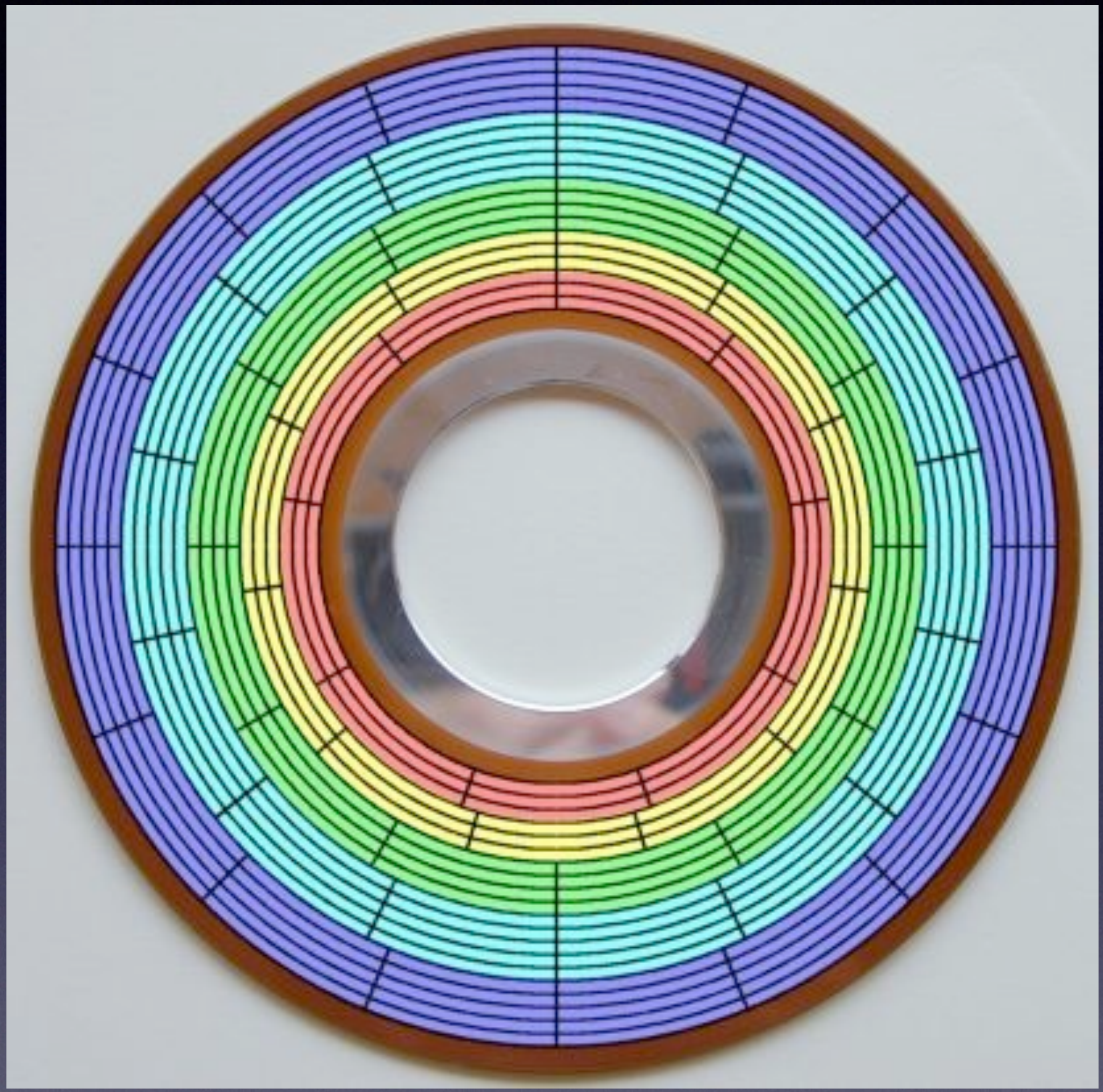
Structure



Data organized in tracks and cylinders

Zoned Bit Recording

- Textbooks refer to tracks with fixed number of sectors
- Modern disks use variable size sectors
- Pack more data on outer, faster-moving tracks
- Disk controller performs logical mapping of fixed sectors to ZBR



Images from storagereview.com

Low-level Formatting

- Done at factory -- not changeable
- Patterns tracks, sectors, servo marks
- Bad sectors identified
- Spare sectors mapped into their place
- Means different disks with identical data, written in the same order, can have different access times

Error Correction

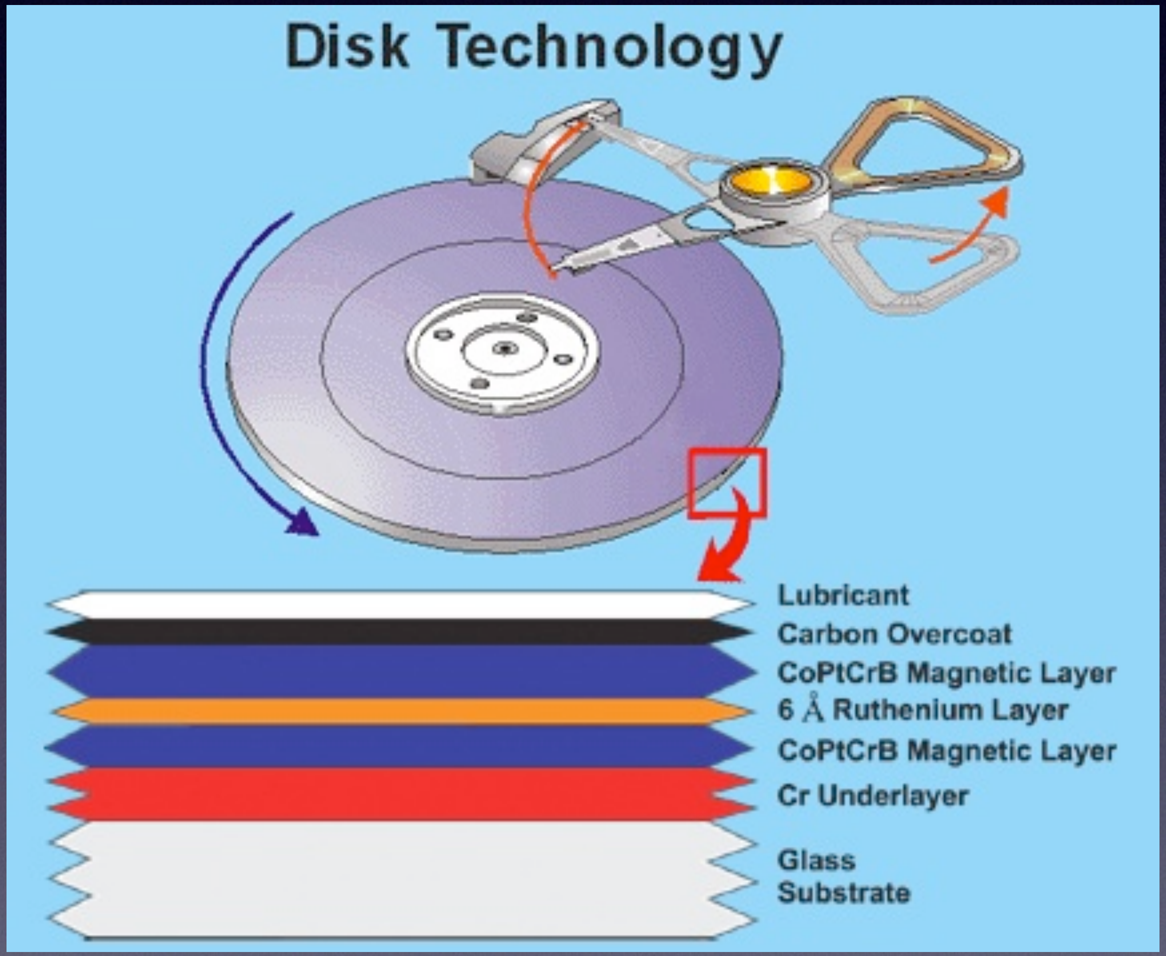
- Read errors are common
- Sectors include error correcting code
- Read and check for error -- if none, good
- If error, apply ECC to fix
- If not fixed, reread, try stronger correction
- If not recoverable, report error

Parameters

- Typically 1 to 10 platters
- 5.25, 3.5, 2.5, 1.8, 1.3, 1.0 inches in diameter
  - Smaller platters: Easier to make, lighter, more rigid, less noise and vibration, faster seek times
- Rotation speed: 7200, 10,000, 15,000 RPM
- Substrate materials: aluminum or glass

Coating

- Early disks used iron oxide or similar coating
  - Relatively thick, easily damaged, low data density
- Modern disks use a thin film with carbon overcoat and lubricant

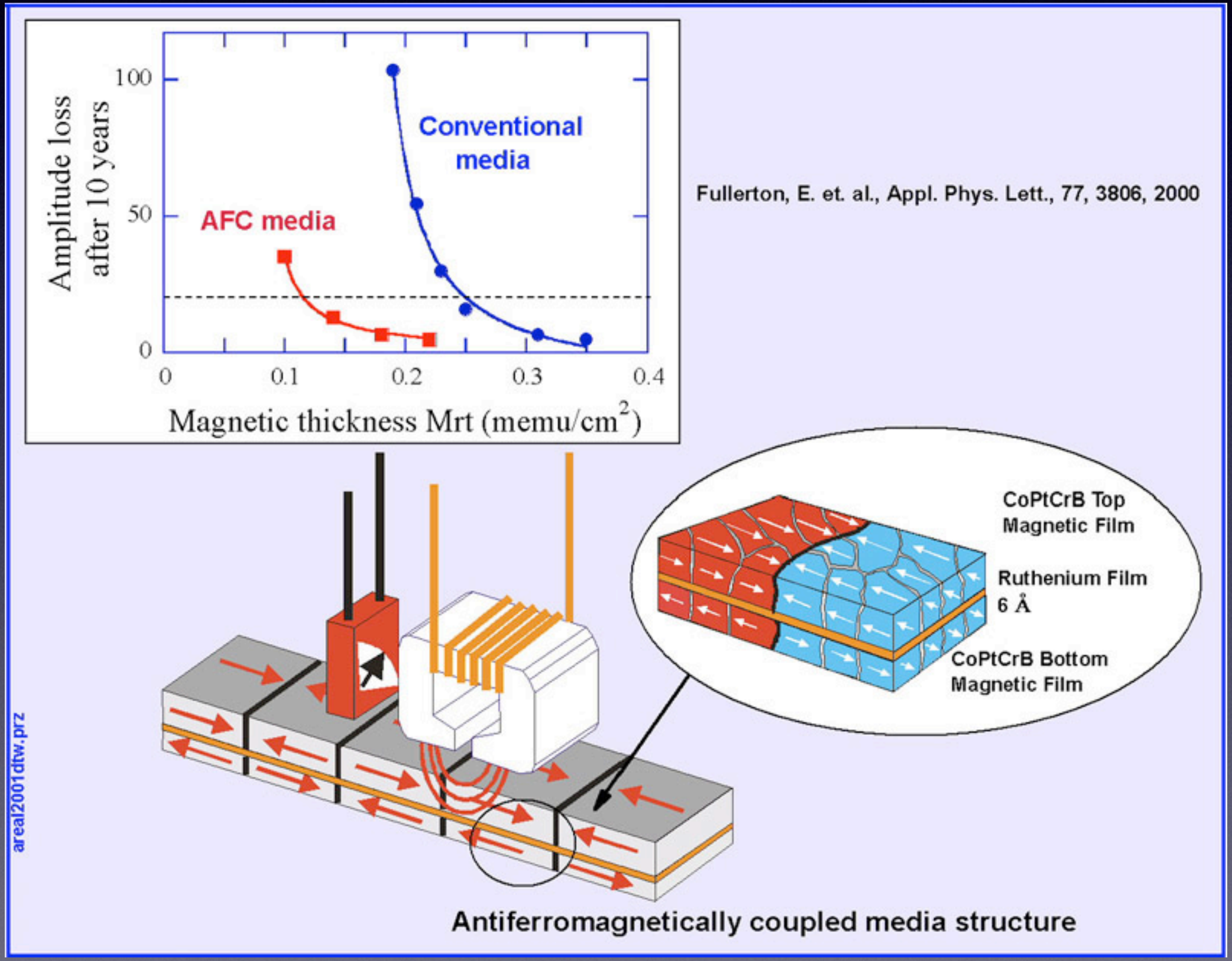


Thin Film

- Thinner enables denser storage -- domains cannot spread out as far
- Grains must be very small
- Must have higher coercivity (resistance to change) and magnetization
- As spot size shrinks, energy to change increases, and approaches thermal limit

Antiferromagnetic Coupling

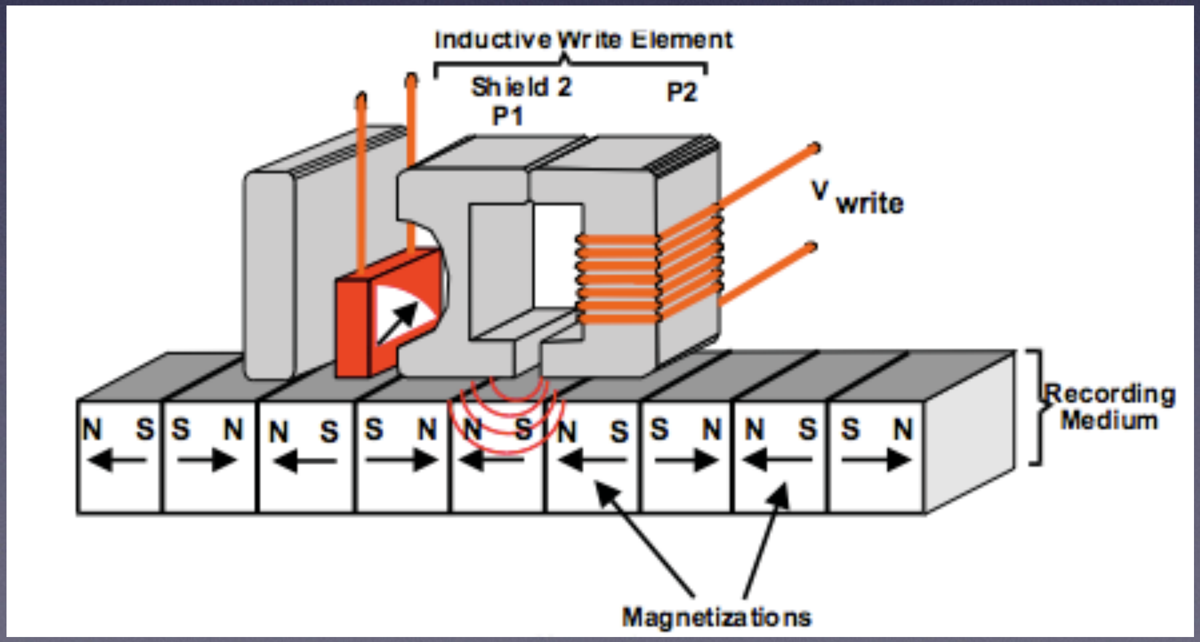
- Coupling layer between magnetic layers
- Effectively makes magnetization layer as thin as coupling layer (a few atoms)
- Allows thicker magnetic layers
- Extends life

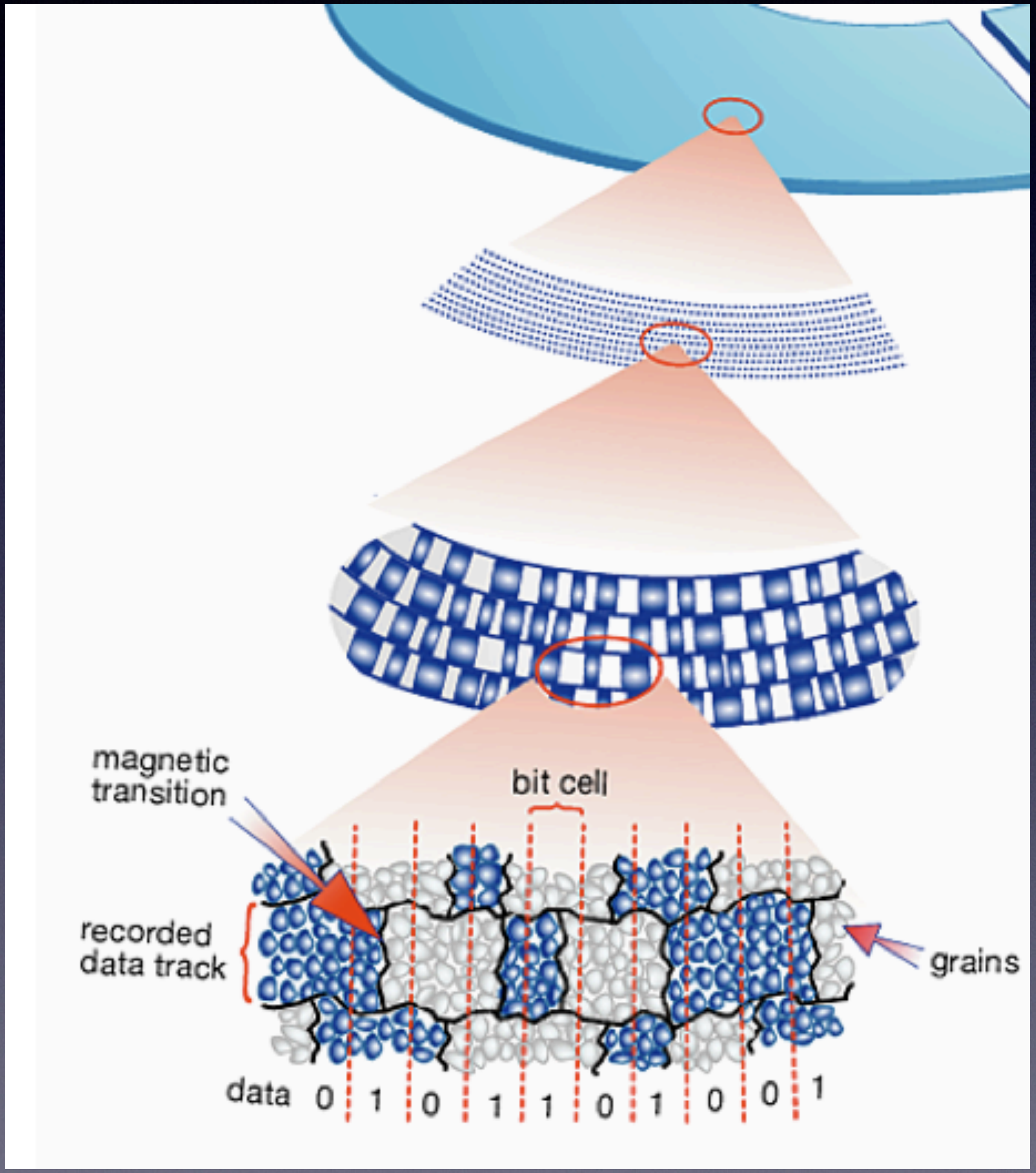


Figures from Hitachi Global Storage Technologies

Longitudinal Recording

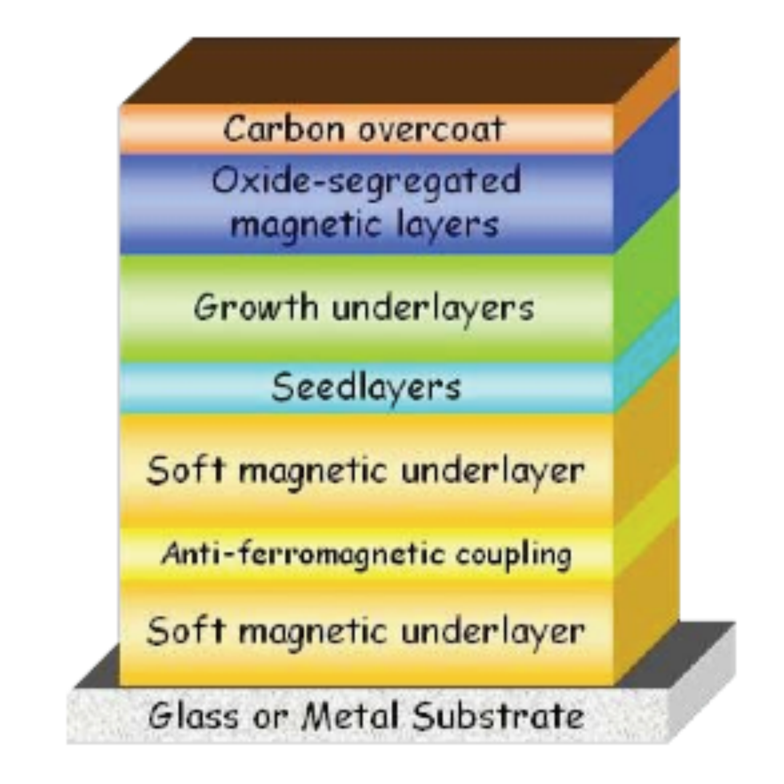
- Spots with same magnetic orientation = 0
- When orientation changes within spot = 1

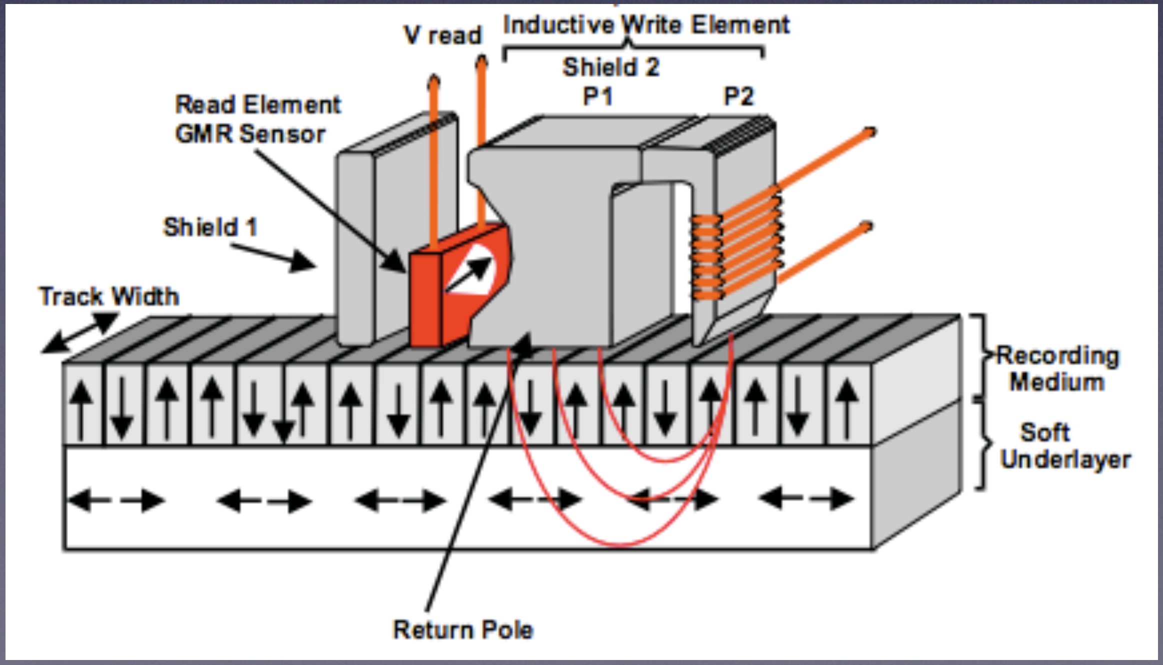




Perpendicular Recording

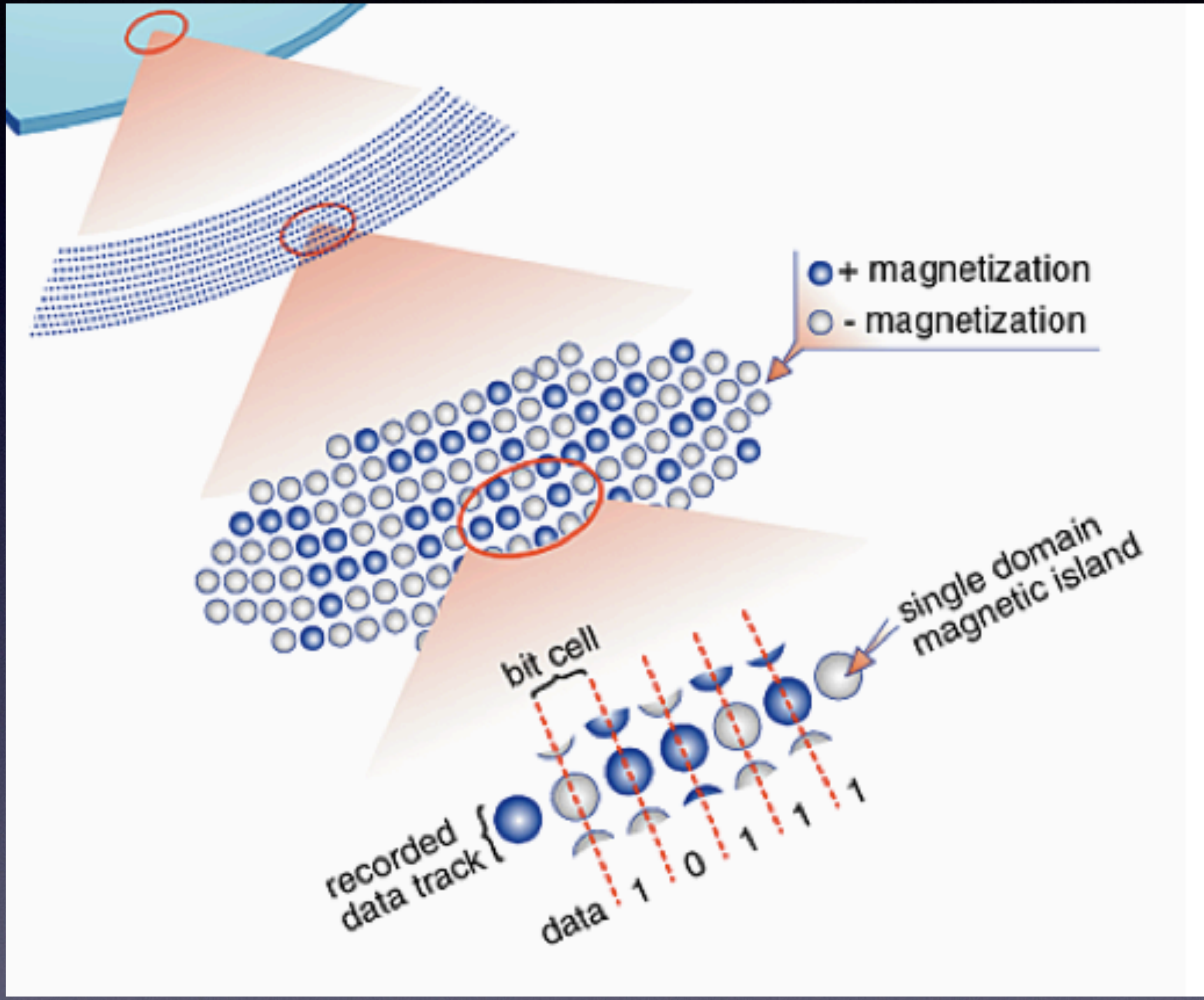
- New film layering with soft underlayer
- New form of write head
- Increases density without reaching thermal limit
- Density will eventually reach point that adjacent domains flip each other

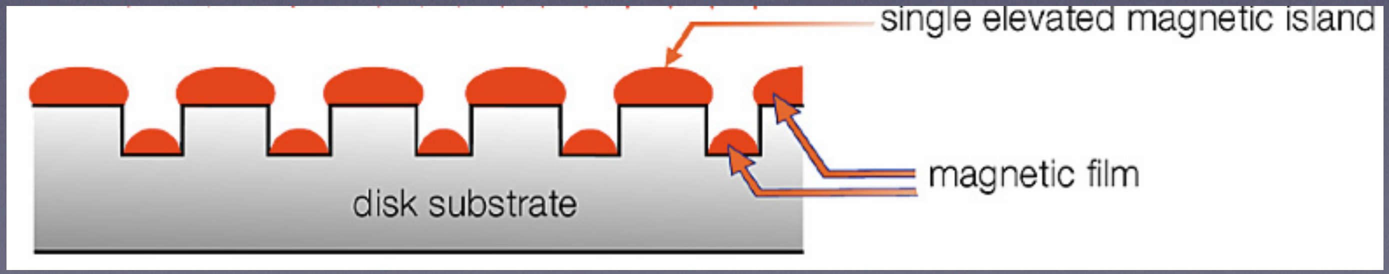




Patterned Recording

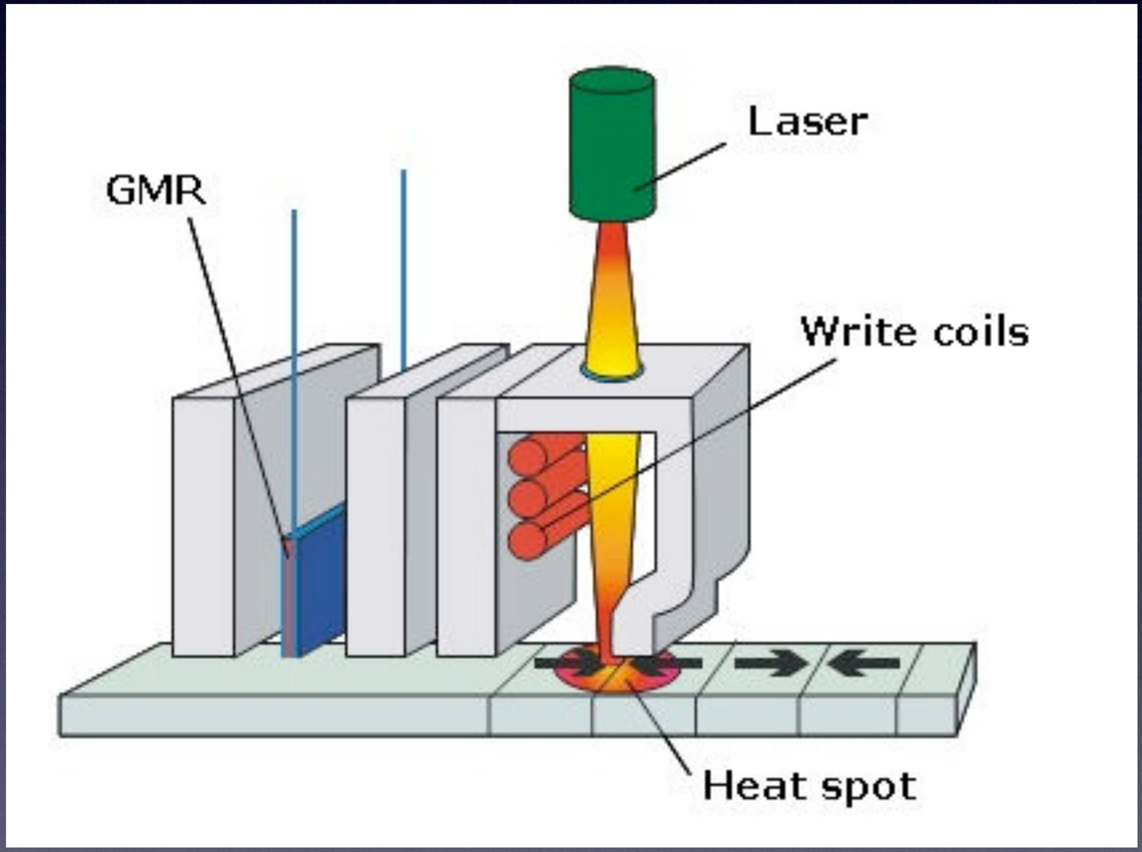
- Use lithography to texture surface for application of film
- Separates domains to avoid interference
- Creates rough surface
- More fabrication steps





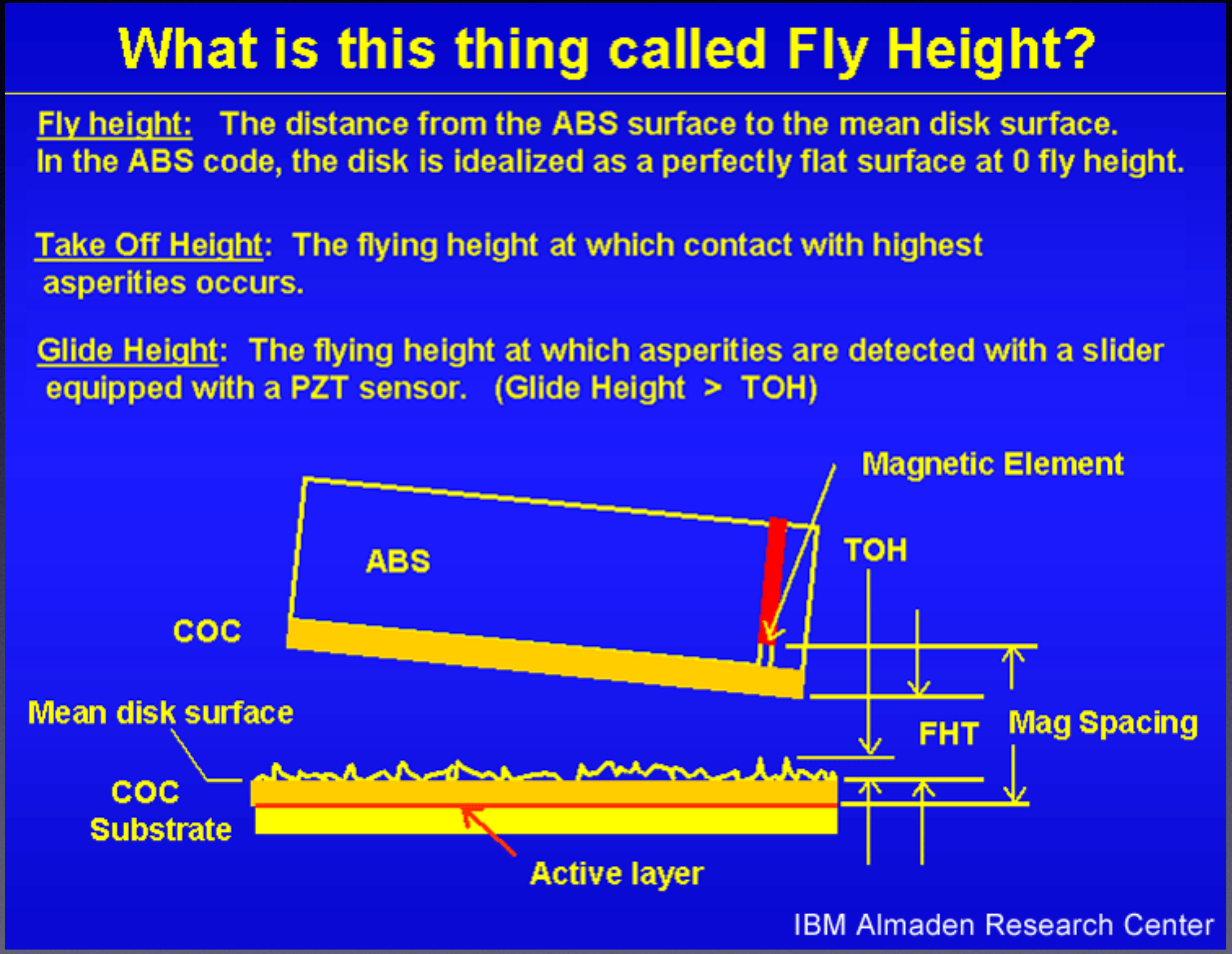
Thermally Assisted Recording

- Use more stable material
- Heat with laser to make temporarily unstable
- Use perpendicular recording to control magnetization before the spot cools



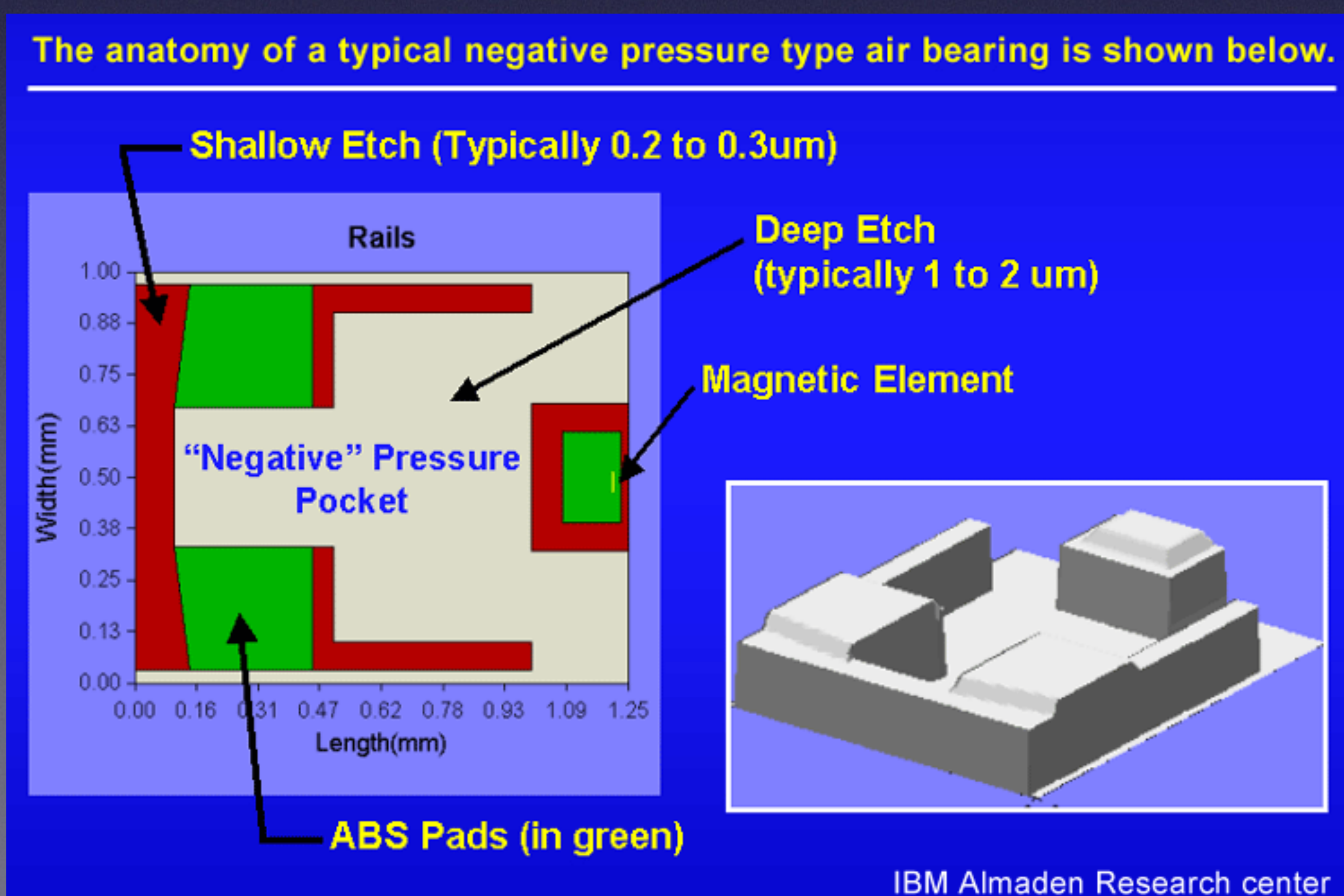
Read Head

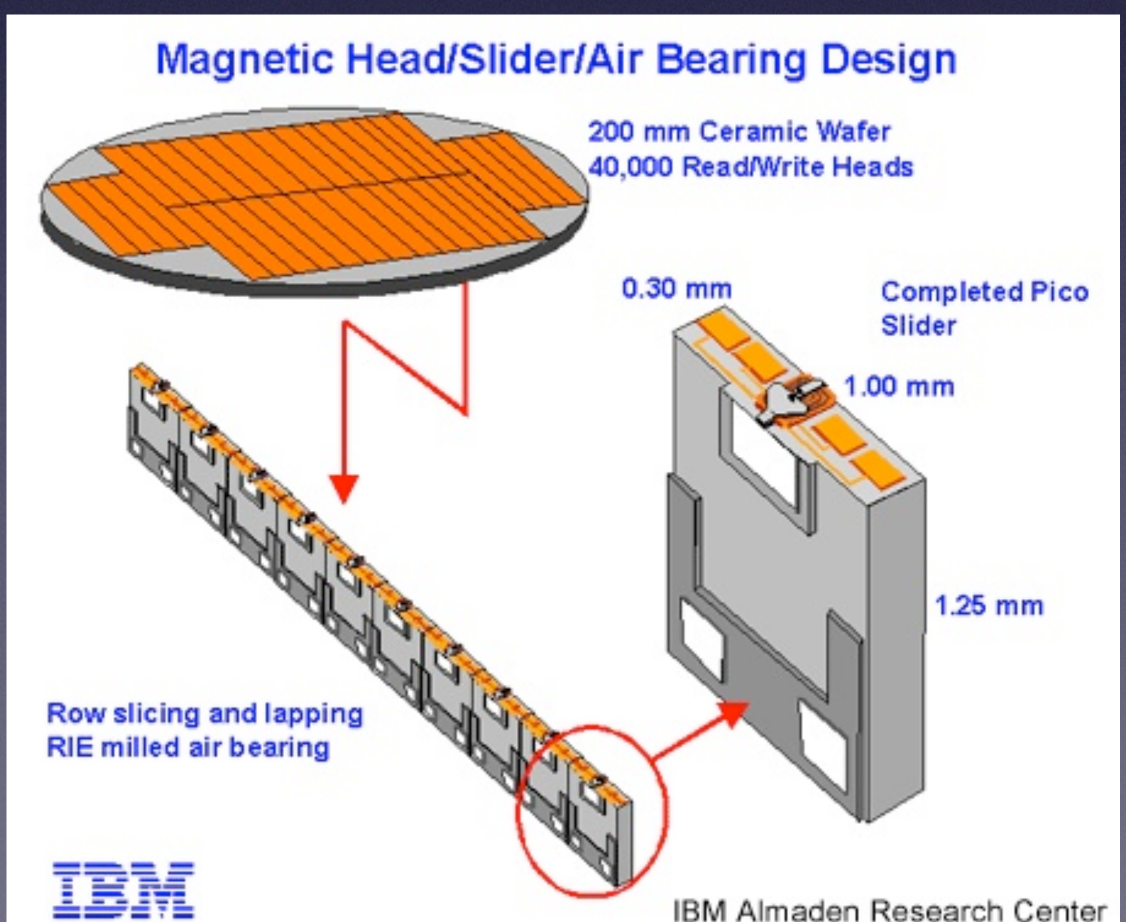
- Flies above spinning surface
- Disk creates airflow
- Lifts head against pressure
- Disk has landing zone for spin-down



Slider

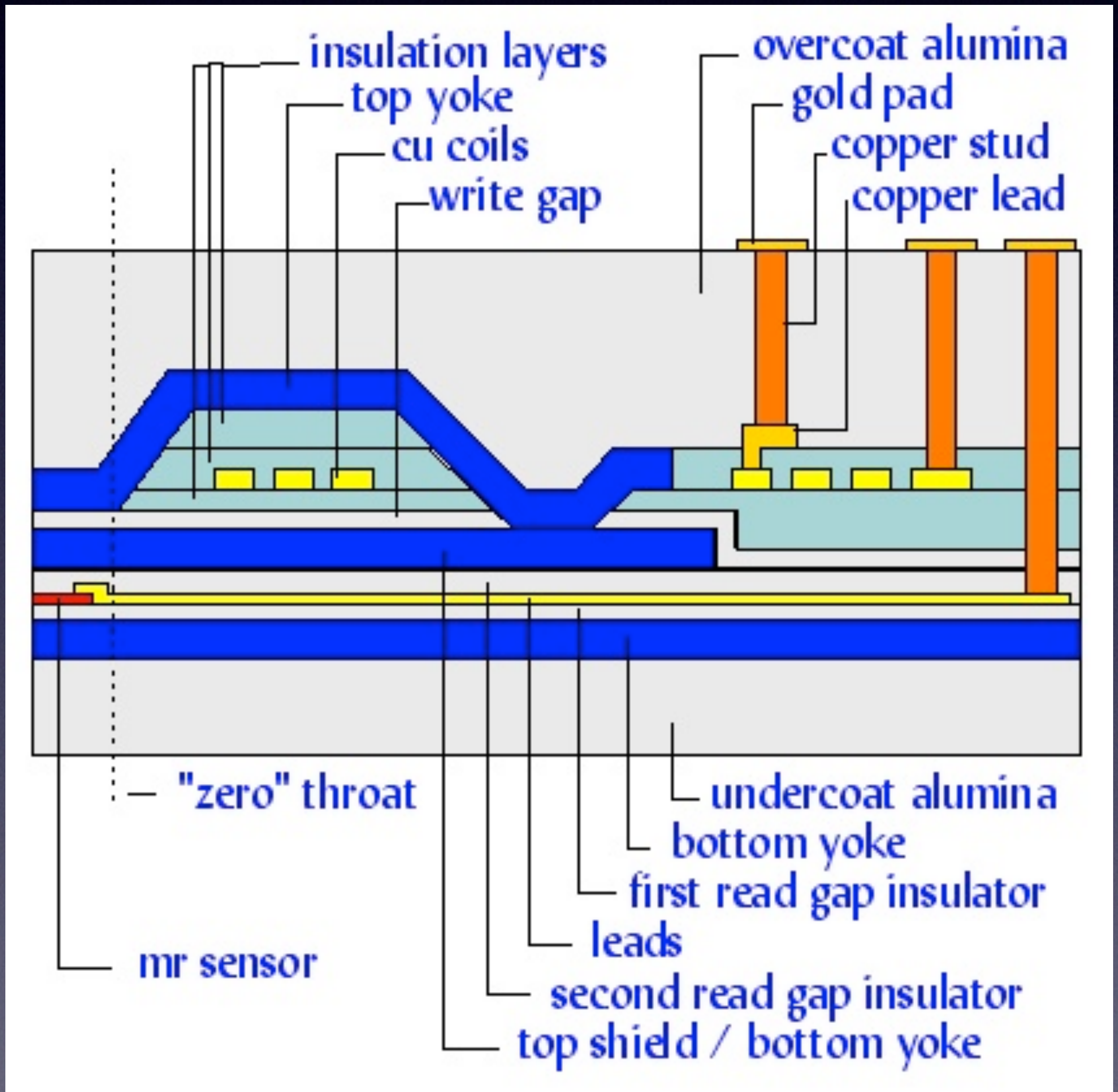
- Aerodynamic shape etched into underside of head to create proper lift and angle
- Electromagnet head attached to edge





Thin Film Head Construction

- Created with lithographic processes
- Copper coils to induce field
- Yoke to concentrate
- Connections to outside



Future

- Projected growth in density of 50% per year (down from 100% per year 10 years ago)
- Superparamagnetic limit probably about 2019
- Current density about 1 Tb/in^2
- Expect growth of 100 before limit is reached
- Will lead to interesting shifts in research focus

Disk Power

- Rotational power proportional to $P * R^{2.8} * D^{4.6}$
- P = platter count
- R = rotational speed (RPM)
- D = diameter of platters
- Head movement small in comparison

Seeking

- Time depends on weight of arm, strength of voice coil, distance to seek
- Speedup phase, coasting phase, slowdown phase, settling phase (servo guidance)
- Moving a few tracks is mostly resettling (more common for smaller platters)
- Moving 10s of tracks is speedup/slowdown
- Moving long distance is mainly coasting
- Controller keeps table of seek impulse quantities

Special Cases

- When moving one track (e.g., data continues on next track), essentially same as settle time
- Does not read from cylinder in parallel -- minor track misalignment. Switch to reading same track on another platter requires settling time
- Reading tries to get data before settling, then use ECC
- Write must wait for settling

Reading

- Signal is weak and noisy
- Must be amplified, converted from analog to digital at higher frequency than data bit rate
- Signal processing applied to extract bits from waveform
- Bits then forwarded to ECC for check/correct

Disk Controller Caching

- RAM, NVRAM buffer for data going to/from disk
- Helps hide latency
- On reading, prefetch extra sectors
- On write, store data until seek/rotation into place
  - Multiple cached writes enable dynamic scheduling

Reliability Factors

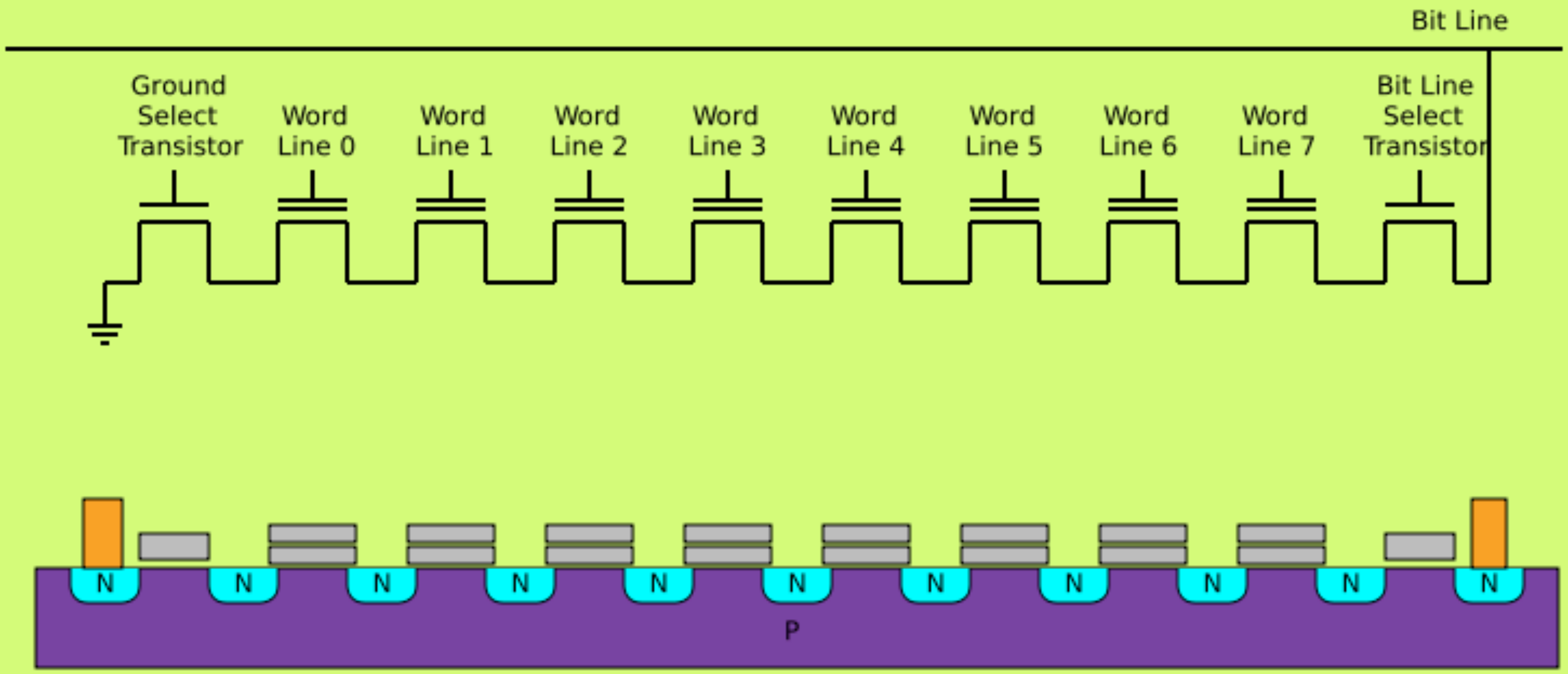
- Vibration
- Rotation speed, mass of platter assembly
- Temperature (15°C increase = 50% lower life)
- Frequency of access
- Power-down after long run time (bearing lubricant)

Flash Storage

SS (Flash) Drives

- Solid state, like RAM (10X slower read, 100X slower write)
- Uses a double-layer transistor with a suspended gate
- Relatively non-Volatile (2 -10 year shelf life)
- Wearout (10K - 100K write cycles)
  - Wear leveling, flash translation layer

NAND Organization



Source: Wikipedia

Flash Organization

- Arranged in planes, with blocks of pages (typically blocks contain 64 to 128 pages at, 2KB to 8KB per page). Planes can operate in parallel
- Whole pages are written at once by setting 1s to 0s
- Can rewrite pages, so data can effectively be stored in smaller units, though there are limits
- Erasure is by whole blocks only (reset to 1s), slower
- Reads are for whole pages

Flash Translation Layer (FTL)

- Indirection table that maps logical to physical addresses
- Hides wear leveling and layout policies
- Also hides buffering, write coalescing, etc.
- Often seen as the point where Flash can be architected

SLC vs. MLC

- Single Level Cell holds a single bit
- Multi Level Cell holds two to four bits
- MLC stores multiple levels of charge
- SLC is faster, more reliable, more expensive
- MLC is slower, less reliable, cheaper, wears out 10X faster, shorter shelf life

Hybrids

- Flash/Hard drive hybrid
  - Most files are written once, rarely accessed
  - Flash caches active files, HD spins less
- RAM/Flash
  - Large RAM buffer (cache) for fast access
  - Power source for flash write on power loss

Future

- OS file system built on disk concepts
- Flash has different characteristics
  - Page write, block erase, fast read, slow write, wear leveling, blurs RAM/disk
- May eventually see new approaches with persistent objects

Demo Day

- Wednesday during class time
- CS 150/151, will be open 1 hour earlier
- Will have 1/2 table to share with someone
- Provide a 1-page description to post on the front of your table: Title for project, your name, description of what it does, how to run the demo, what you learned — they will be collected at the end

Kit Check-In

- At end of demo, restore kit to original condition
- Will be inspected and put on cart with sign-out sheet

Final Exam

- Monday, December 17, 3:30 - 5:30, CS 142
- Basically like sample
- Open book, open notes, calculator
- Will also include questions about virtual memory, buses, secondary storage

Course Evaluation

http://owl.umass.edu/partners/courseEvalSurvey/uma/

Suggestions for how to improve the course are most useful

Please be as specific as possible in terms of topics, materials, exercises, ordering, etc.