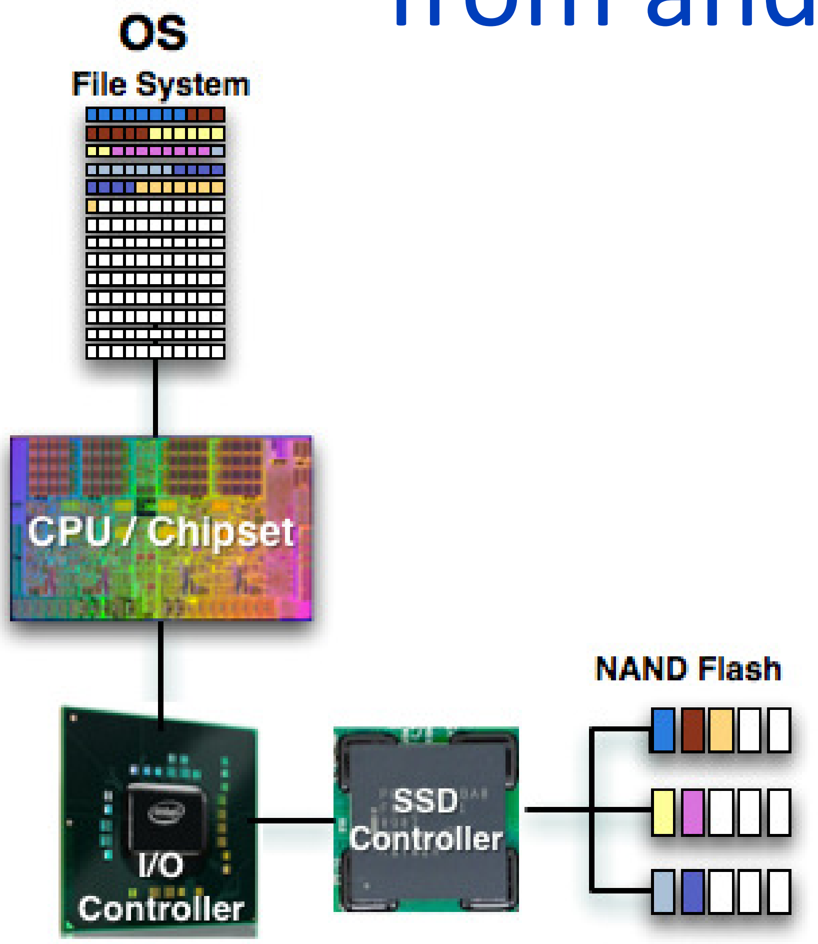
CSCU9V4
Systems

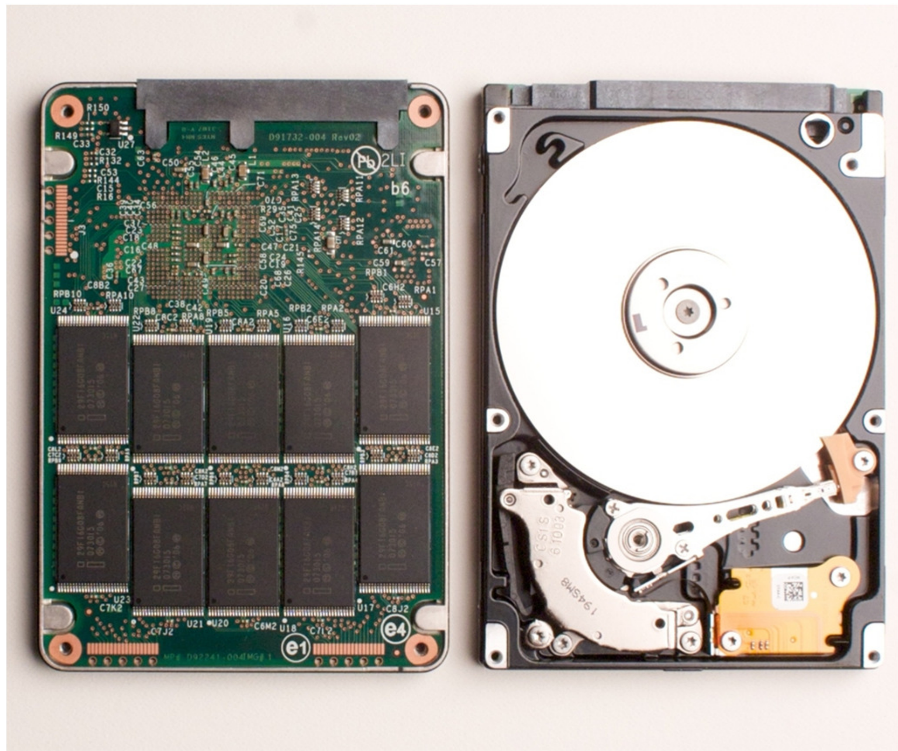
Systems lecture 13

Computer Organisation

Peripherals 2 – SSDs and Optical Disks

Storage: Where it's come from and is going to

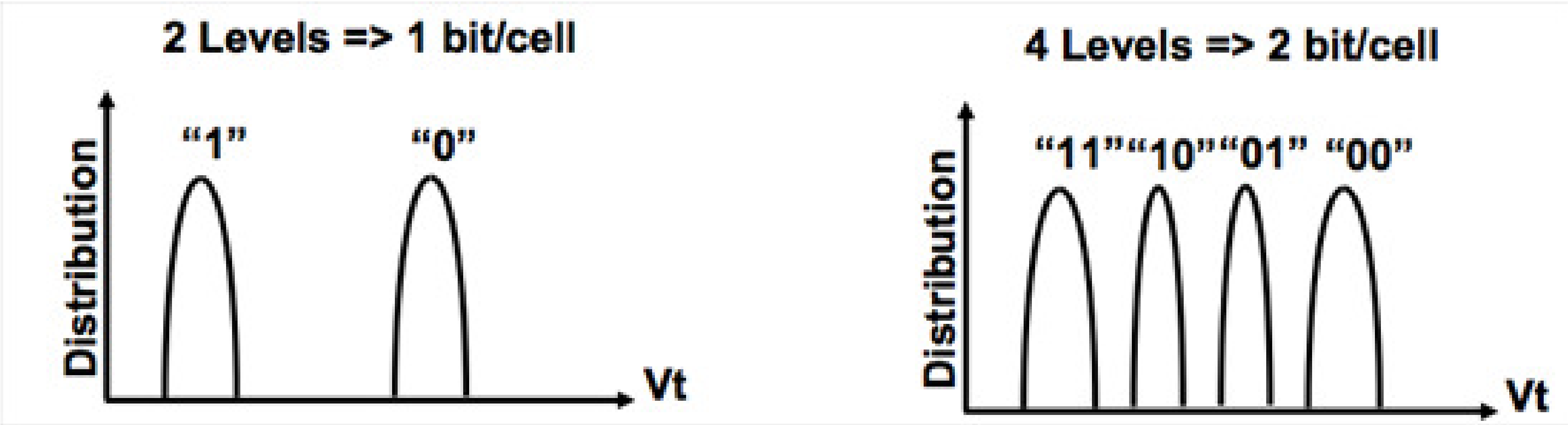




History

- First SSD not in 2007, more like 1976…
  – Only had capacity of 2 MB
- Fast-forward to 2007:
  – NAND Flash (Invented in 1980)
    - Non-Volatile Storage System
    - Used in USB Flash Drives, MP3 Players & SSDs
    - Multiple types: Single Level Cell(SLC) & Multi Level Cell(MLC)
    - Guess and Check reading and writing

Reading & Writing



- SLC Flash MLC Flash

- Apply Voltage
- Wait for Reaction
  - Return Result or
  - Apply More Voltage, Repeat

Performance of SLC vs. MLC

- Difference?
  - Since there are only 2 possible values of SLC it only takes 1 voltage to return a 0 or 1
  - MLC requires a maximum of 3 different voltages to assure that the value will be found.

  - Random Read Speeds:
    - SLC: 25 µs MLC: 50 µs
  - Random Write Speeds:
    - SLC: 250 µs MLC: 900 µs

What makes an SSD special?

- We’ve all used SSDs before, the same technology is used in USB **Flash** Drives and they only function around 5-40MB/s.

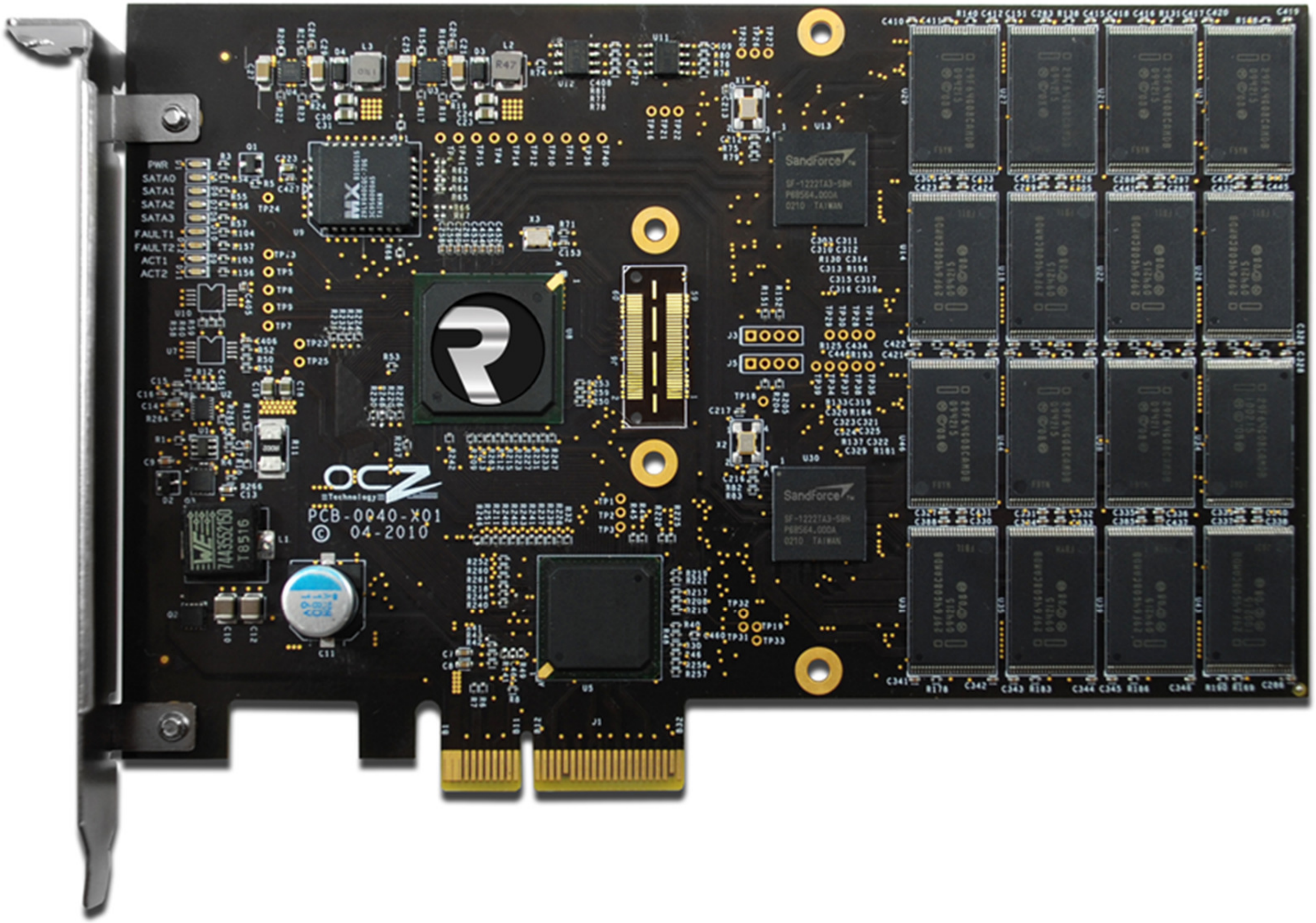
- What if we use 10 USB Flash Drives in raid as a HDD?
    - Anywhere between 40GB – 320GB
    - Possible access speed of 50-400MB/s

Structure on the Board

- Each SSD board contains any number of NAND ICs, chip used to store data, all depending on how expensive the board is.

- Each SDD board also contains any number of support channels, normally one per NAND IC, which allows the controller to communicate to each NAND IC.

OCZ Revodrive



Parallelism!

- The sheer speed of a SSD comes from the fact that it can access each of its NAND IC at exactly the same time.

- While Platter HDDs like to have reads/writes to be in the same location for future access, SSDs would rather have the data spread evenly across all of its NAND ICs for maximum accessibility.

A Holy Grail?

- SSDs seem to have it all:
  - Almost instantaneous read and write times
  - The ability to read or write in multiple locations at once
  - The speed of the drive scales extremely well with the number of NAND ICs on board

- But what is the cost?

What about Erasing?

- To erase the value in flash memory the original voltage must be reset to neutral before a new voltage can be applied, known as write amplification.

- Random Erase Speed:
  - SLC 2ms per block   MLC 2ms per block

- Blocks?

Memory Structure

- 1 or 2 bits does us no good
- Pages!!!
  - 1 Page = 4KB (we'll revisit `downstream')
- Block = 128 Pages = 512 KB
- Plane = 1024 Blocks = 512 MB
  - Depending on the board the combining keeps going up until you get a single chip, NAND IC, on the board

The Achilles' Heel of SSDs

- Caveat: write and read to a single page of data from a SSD but entire blocks must be deleted to release the page?
- Maybe acceptable since enables single-page read/writes.
  - Except: the page has to be empty before we can write to it!
- To make matters worse, a standard MLC can only be erased 10,000 times before it goes bad.
- Solution: Lets not actually delete files when they are deleted on the OS, much like a platter drive.

The Original Solution

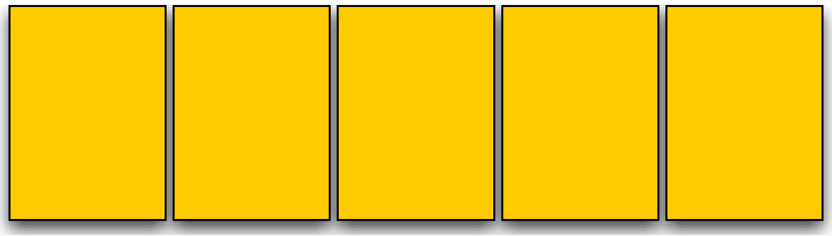
- Rather then deleting the block and writing the modified block with the new page back in the original location just write the modified block to another location in memory.

- Believe it or not drives were actually shipped with this solution, never thinking about what happened when the drive filled up.

- After the drive filled up the amount of time to write a block of data went from 250 μs to 250 μs + 2ms since it also had to delete a block. This actually made the SSDs slower then a regular platter drive when writing.

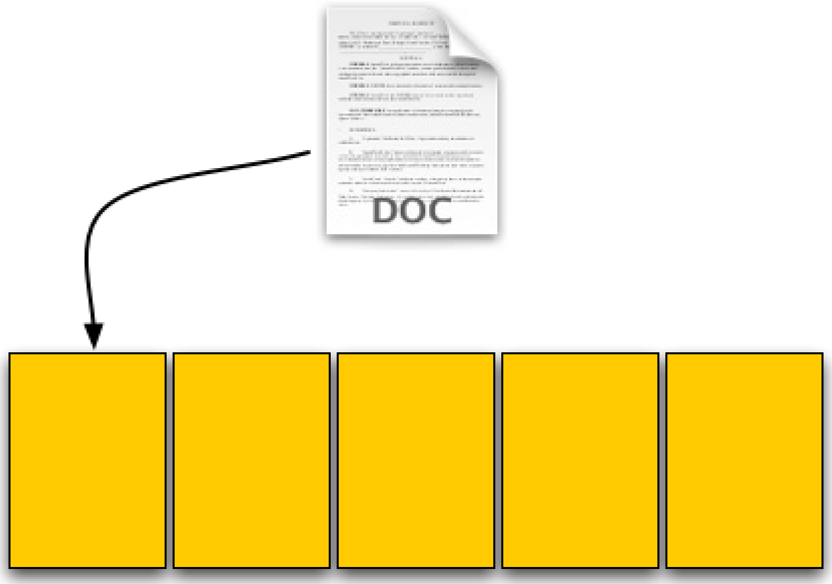
Example:

– Hypothetical SSD:
  - Page Size: 4KB
  - Block Size: 5 Pages
  - Drive Size: 1 Block
  - Read Speed: 2 KB/s
  - Write Speed: 1 KB/s



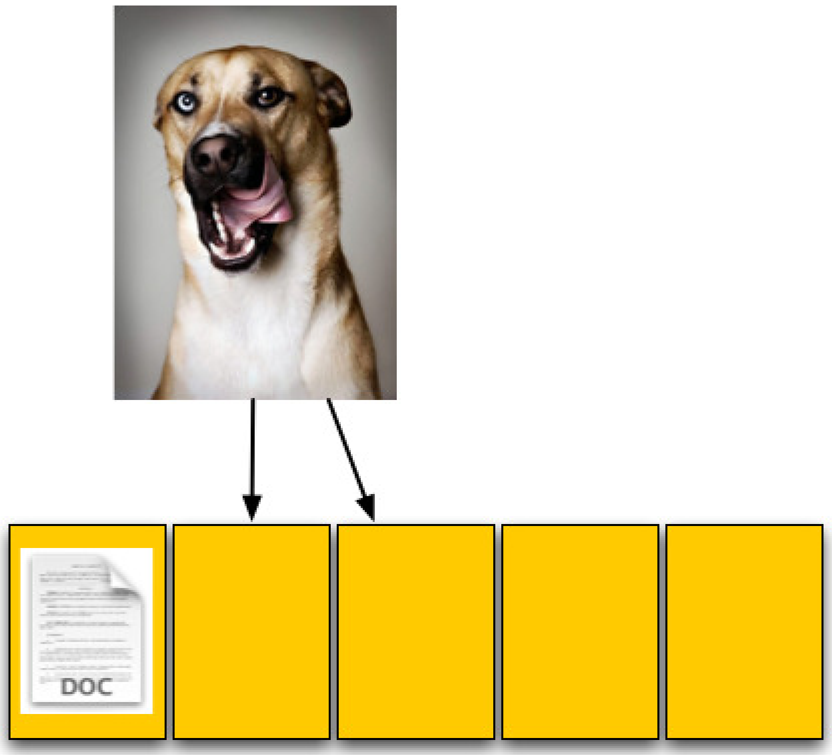
Example Cont.

- Lets write a 4kb text file to the brand new SSD.



Example Cont.

- Now lets write an 8kb pic file to the almost brand new SSD, thankfully there's space.

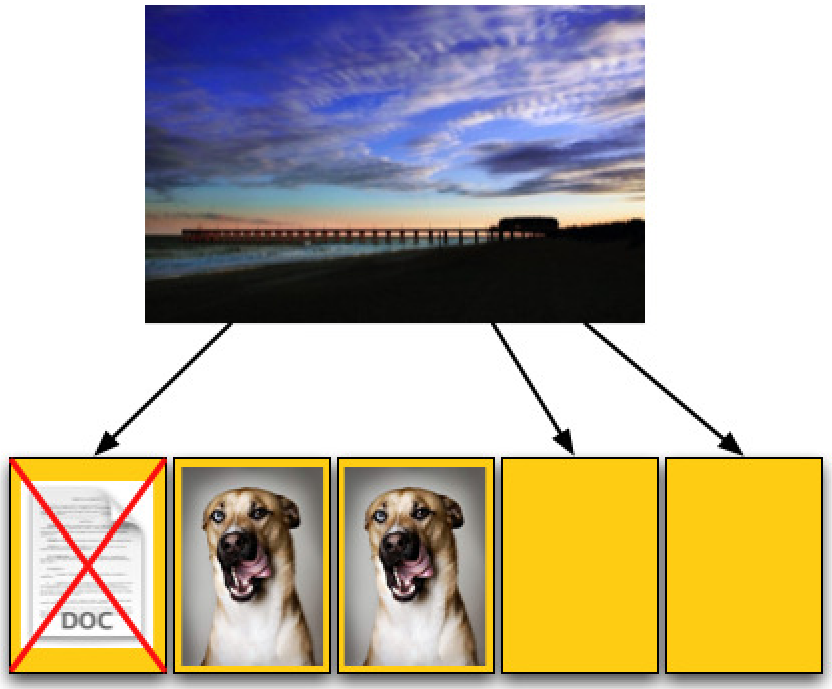


Example Cont.

- Now lets delete the txt file in the first page.

Example Cont.

- Finally lets write a 12kb pic to the SSD. How long should it take? 1 kb/s write speed



Example Cont.

- What’s wrong here?
  - The OS is told there are 3 open pages on the SSD when there are only 2 available.
  - Time for the SSD to do something fancy to open up the space.
    - Banking on the quality of the SSD hopefully it has an onboard cache otherwise it has to use ram, taking much longer and consuming CPU cycles.

Example Cont.

- Step 1: Read block into cache
- Step 2: Delete page from cache
- Step 3: Write new pic into cache
- Step 4: Delete the old block on SSD
- Step 5: Write cache to SSD

What’s happening?

- The OS only thought it was writing 12 KBs of data when in fact the SSD had to read 12 KBs and then write 20KBs, the entire block.
- Since the SSD is quite slow the operation should have taken 12 secs but actually took 26 seconds, resulting in a write speed of .46KB/s not 1KB/s
- Can the problem fixed, thereby saving SSDs from an early grave?

Trim to the Rescue in 2009

- One Idea: delete the file when it is deleted from the OS, or a relatively short time afterwards, and clear out the page in the block.
- Actually this fixes the problem of running out of space, but what happens when we try to overwrite a file, ie saving an updated word document?
- Unfortunately there is no way around having to read the block containing the original file into cache and deleting it, however there is a choice to do it before or after the write. Which is better?

Strange Side Effects

- So what happens when I want to do a fresh install of my OS on the HDD?
- Should I just follow the standard reformat option and install like normal?

There’s a Command for That

- Believe it or not there is actually a command for “trimming” the entire drive so it appears brand new, except for the fact that the life span of each NAND cell has been decreased by one.
  - Intel and many of the SSD manufactures supports a command called HDD ERASE that permanently deletes the data on the drive.

Additional Food for Thought

- Data Recovery?
  - On a standard HDD data recovery of deleted files is quite easy because the actual bits are still on the HDD since the HDD doesn't actually delete them.
  - What happens with TRIM?
    - Severely reduces the possibility of locating deleted files on the drive, making computer forensics impossible.

Conclusion on SSDs

- Pros:
  - SSDs are extremely fast
  - SSDs are easy to use
  - SSDs are the `future’ of media storage

- Cons:
  - SSDs are expensive
  - SSDs are constantly upgrading
  - SSDs are complicated to understand
  - (arguably) lower life span.

Optical secondary storage

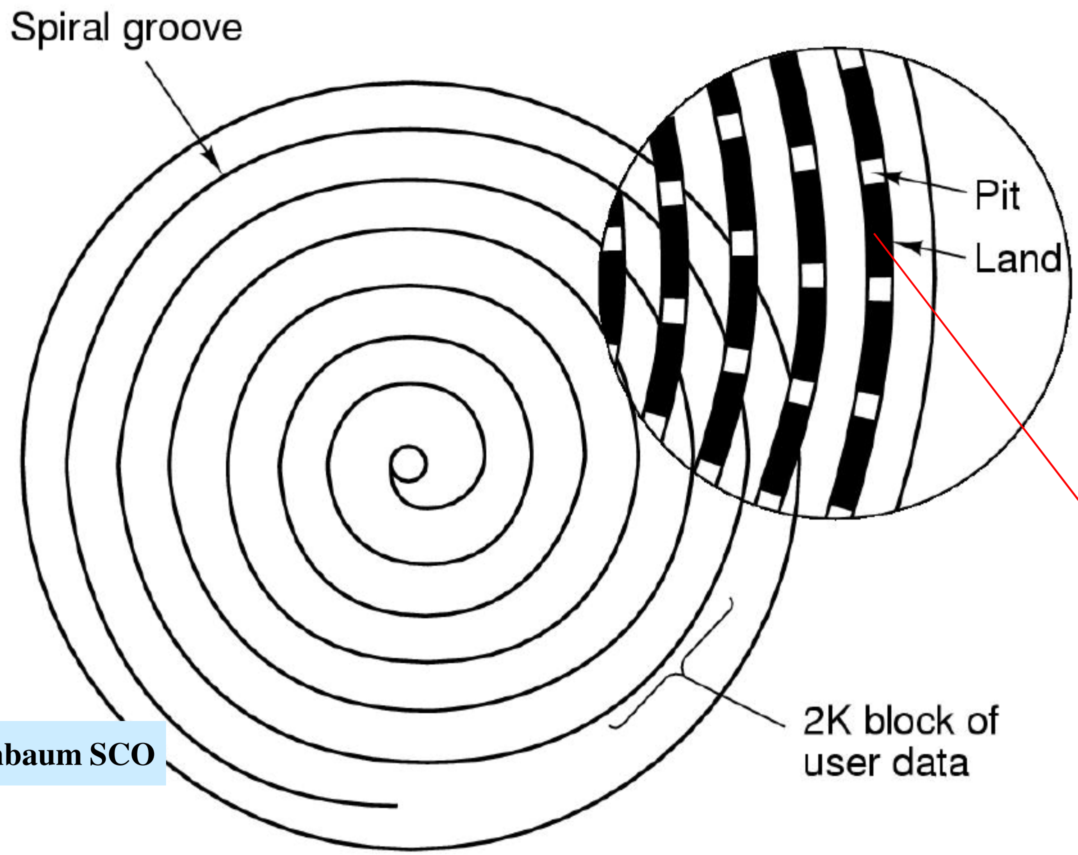
- CD-ROM
  - Read Only Memory
- CD-Rs
  - recordable CDs,
  - used extensively for data distribution and archival data.
- CD-RWs
  - rewritable CDs
    - relatively expensive.
- DVD
  - digital video/versatile disks
  - used for archiving/data distribution
  - even more data!
    - 4 Gbytes or more.
- DVD-Rs
  - Writable DVDs: not as well standardised.

- Bluray

CD-ROMs: *intro*

- CDs have been around since about 1980.
  - see Tanenbaum Section 2.3.7.
- originally used for recording music.
- used now also for recording data, using the same technology
  - mature technology
  - standard
    - 12cm across
    - 1.2mm thick
    - 15mm hole
    - standard encoding
- mass produced players
- data is recorded in a single continuous spiral
  - like an old-fashioned gramophone record (LP)
  - on the spiral there are pits and lands
    - pits are depressions
    - lands are the areas between the pits

CD-ROMs: *reading*



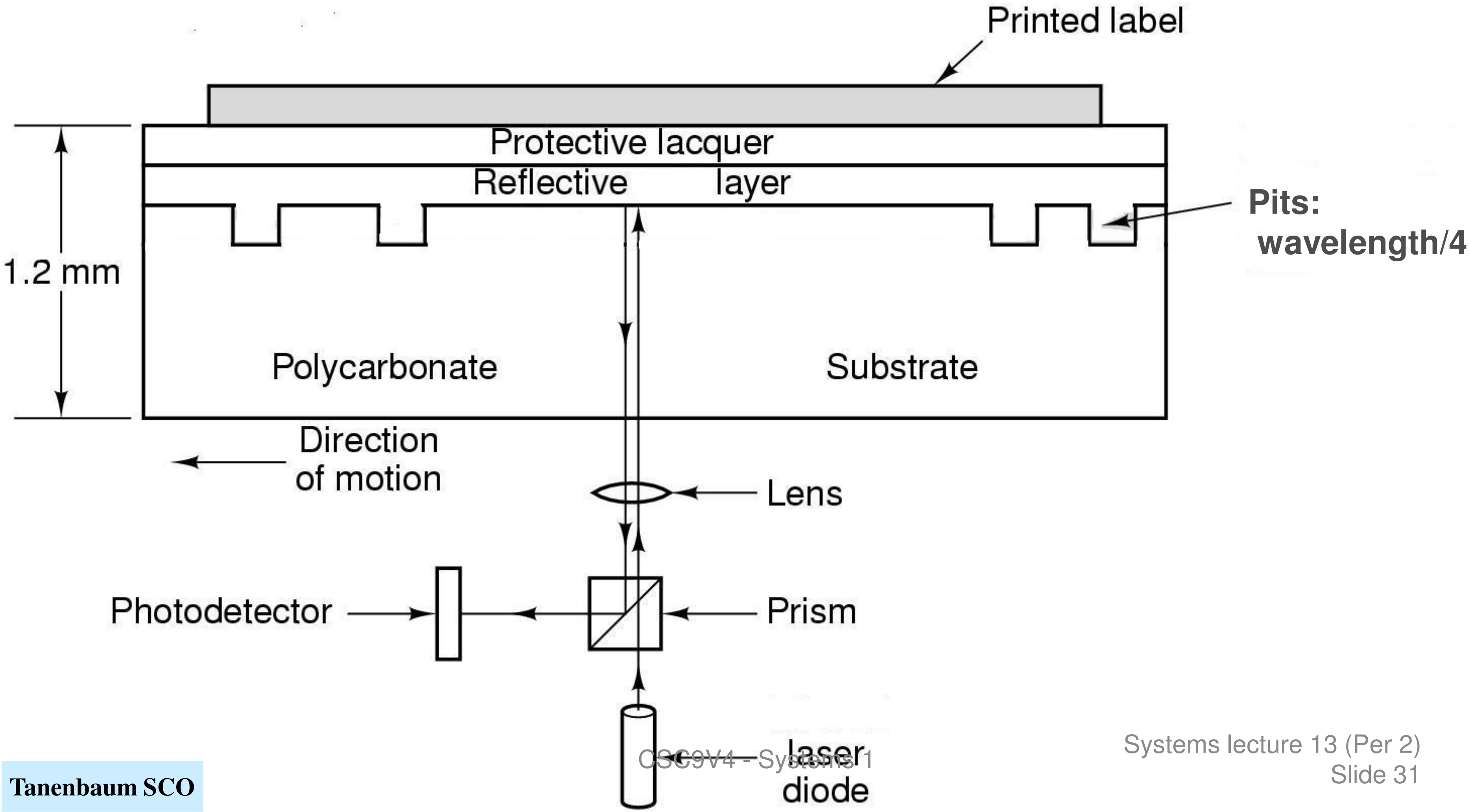
Tanenbaum SCO

- **spiral, unlike magnetic HD**
- **constant *linear* speed, so a variable *rotational* speed**
- **120cm per second**
- **530rpm on the inside**
- **200rpm at the outer edge**
- **600 spirals per mm**
- **full track is 5.6 km long**
- **650 Mbytes approx.**

a bit is the land-pit *transition*

CD-ROMs: *reading*

- surface is with read by a 0.78 micron wavelength laser and a photodiode
- read from polycarbonate side … pits therefore stick "out" (bumps)
- if detector has "bright" reflection it's land; the bumps seem dark

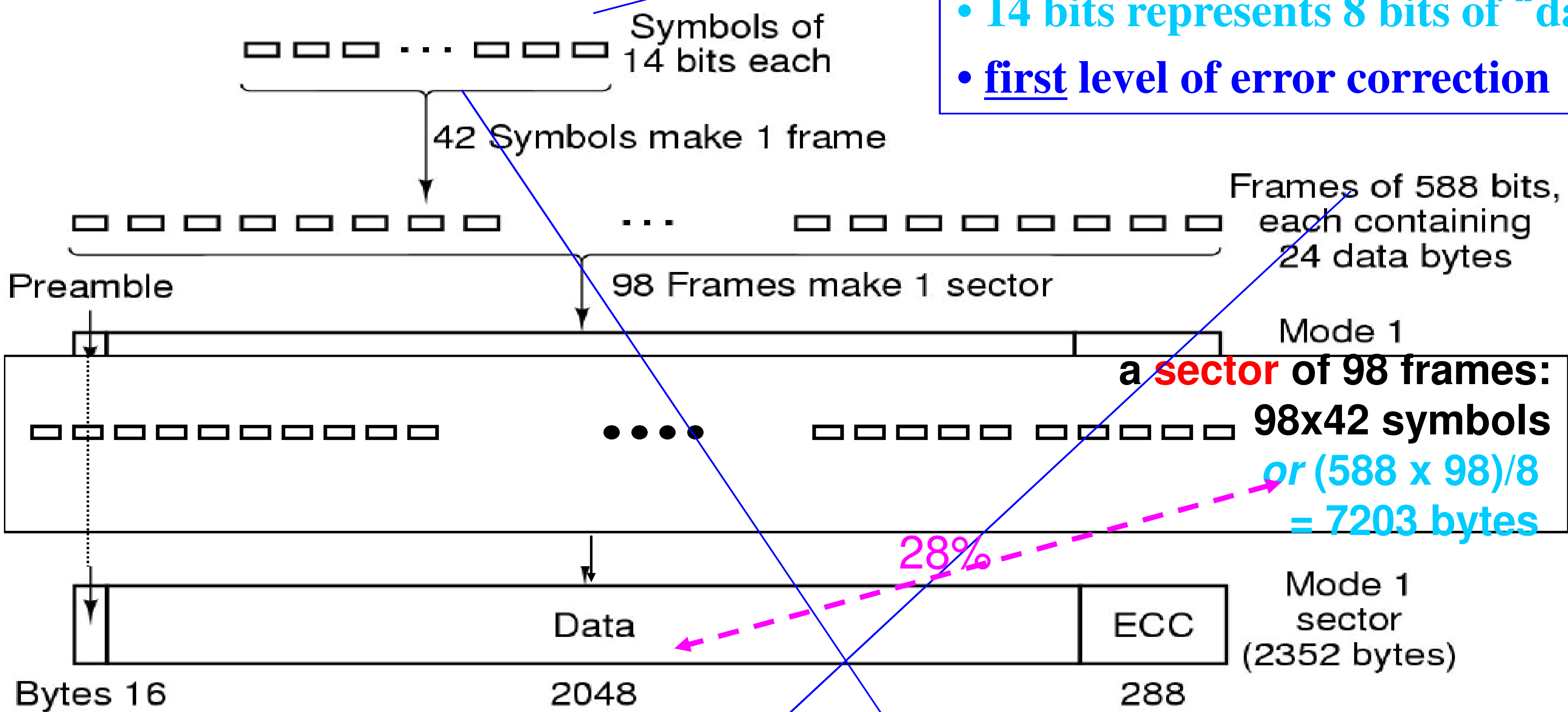


Tanenbaum SCO

Tanenbaum SCO

CD-ROMs: *data encoding (1)*

- a single byte
- 14 bits represents 8 bits of "data"
- first level of error correction

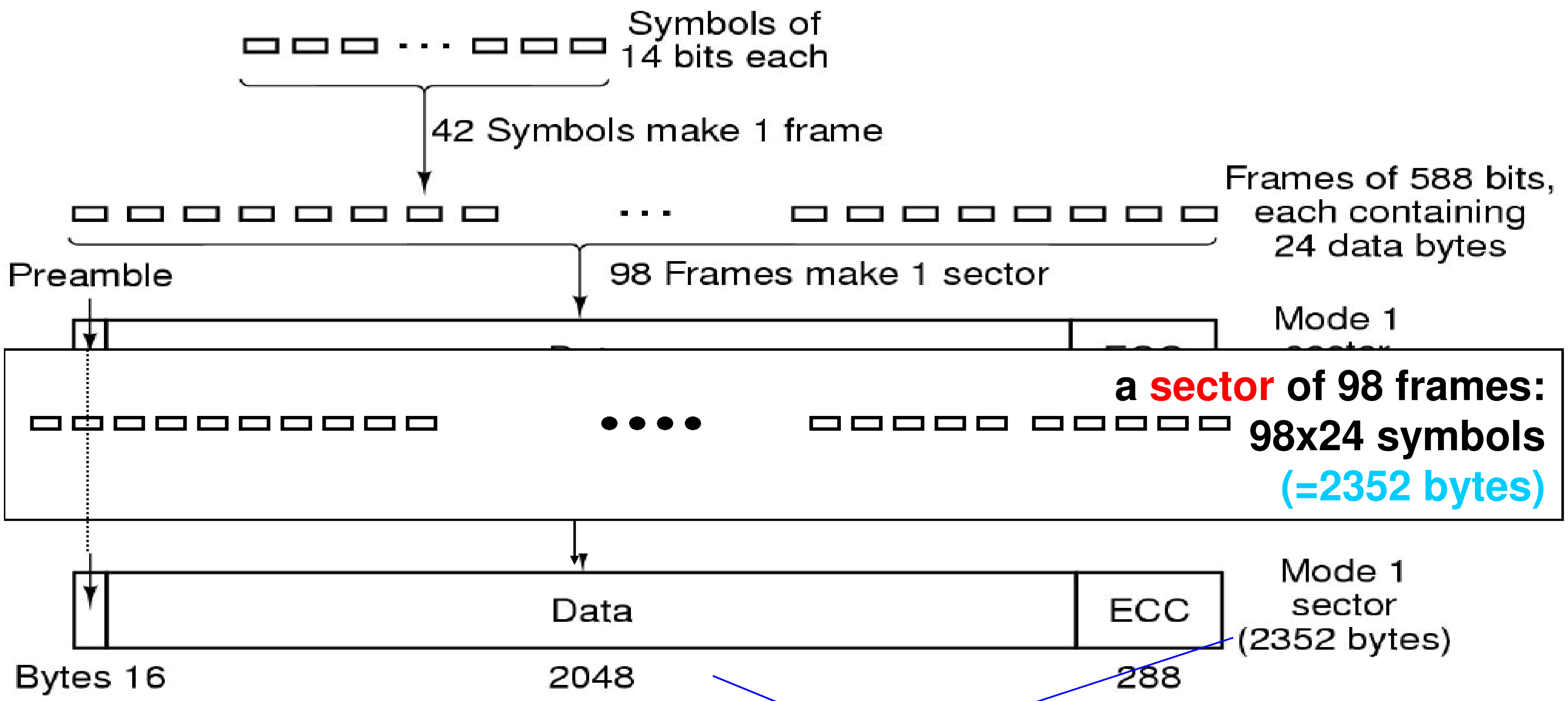


- 42 symbols x 14 bits = 588 bits

- 24 symbols have "data"
- the remainder are for errors
- second level of error correction

- CCITT yellow book (1984)

CD-ROMs: *data encoding (2)*

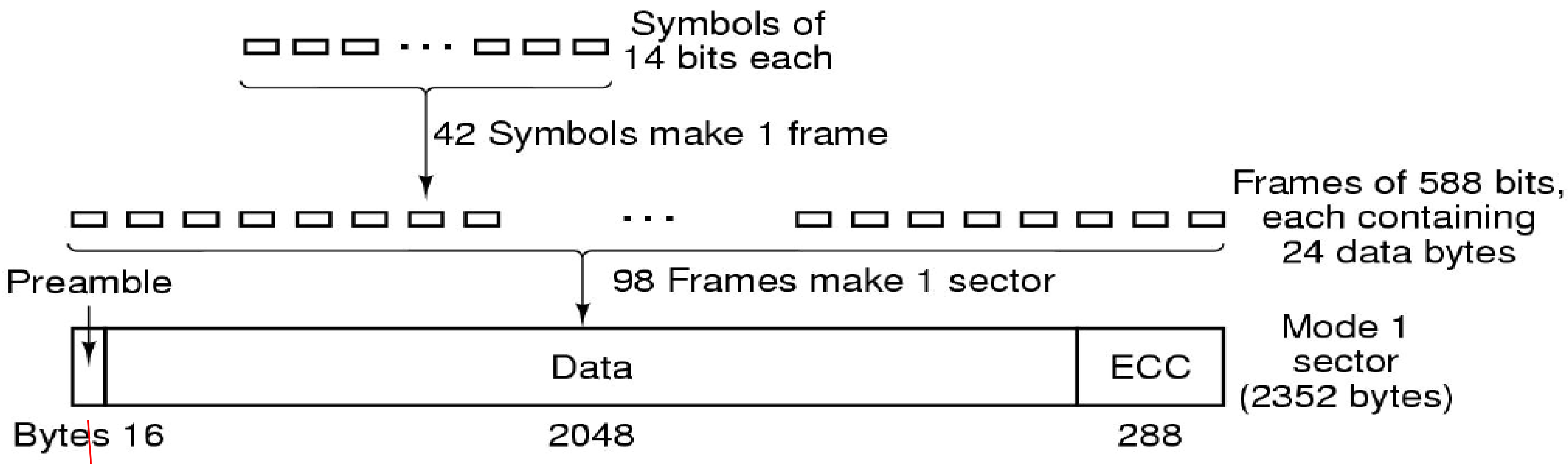


- **24 symbols have "data" in each frame**
- **with 98 frames**
- **gives 24 x 98 = 2352 bytes of "data"**

- **2048 bytes of data**
- **288 are for errors**
- **<u>third</u> level of error correction**

CD-ROMs: *data encoding (3)*

Tanenbaum SCO



first 12 bytes: 00FFFFFFFFFFFFFFFFFFFF00

next 3 bytes: sector number

~~**last byte is the mode: first mode is , the second has no ECC**~~

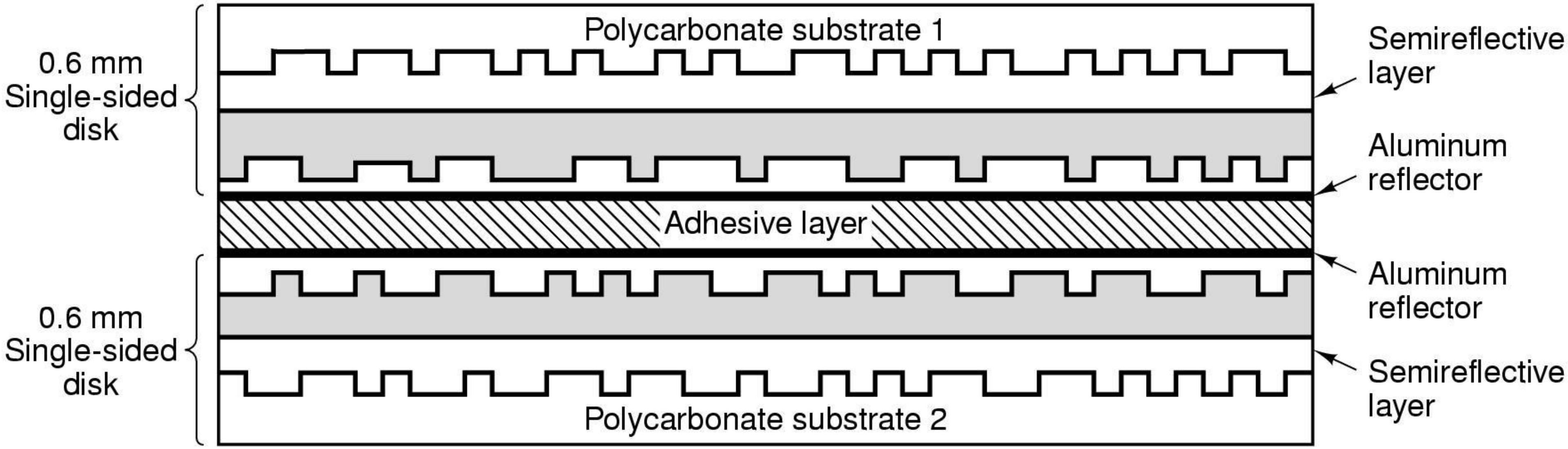
- **CCITT yellow book (1984)**

CD-ROMs: *data encoding (4)*

- designed to be highly resilient to errors
- 14-bit coding for each byte
  - 14 -> 8 coding via look-up table
  - allows error-correction (can fix any 2-bit errors).
- 1 frame = 42 symbols (588 bits) used to form 192 data bits.
  - corrects short burst errors
- (this much is still common between music and data CDs.)
- 98 frames is a CD-ROM sector:
- A sector consists of
  - 16-byte preamble (hex 00FF…FF00, then sector number)
  - 2048 bytes of data
  - 288-byte error-correcting code
  - …and this can detect and correct any remaining errors

DVDs

Tanenbaum SCO



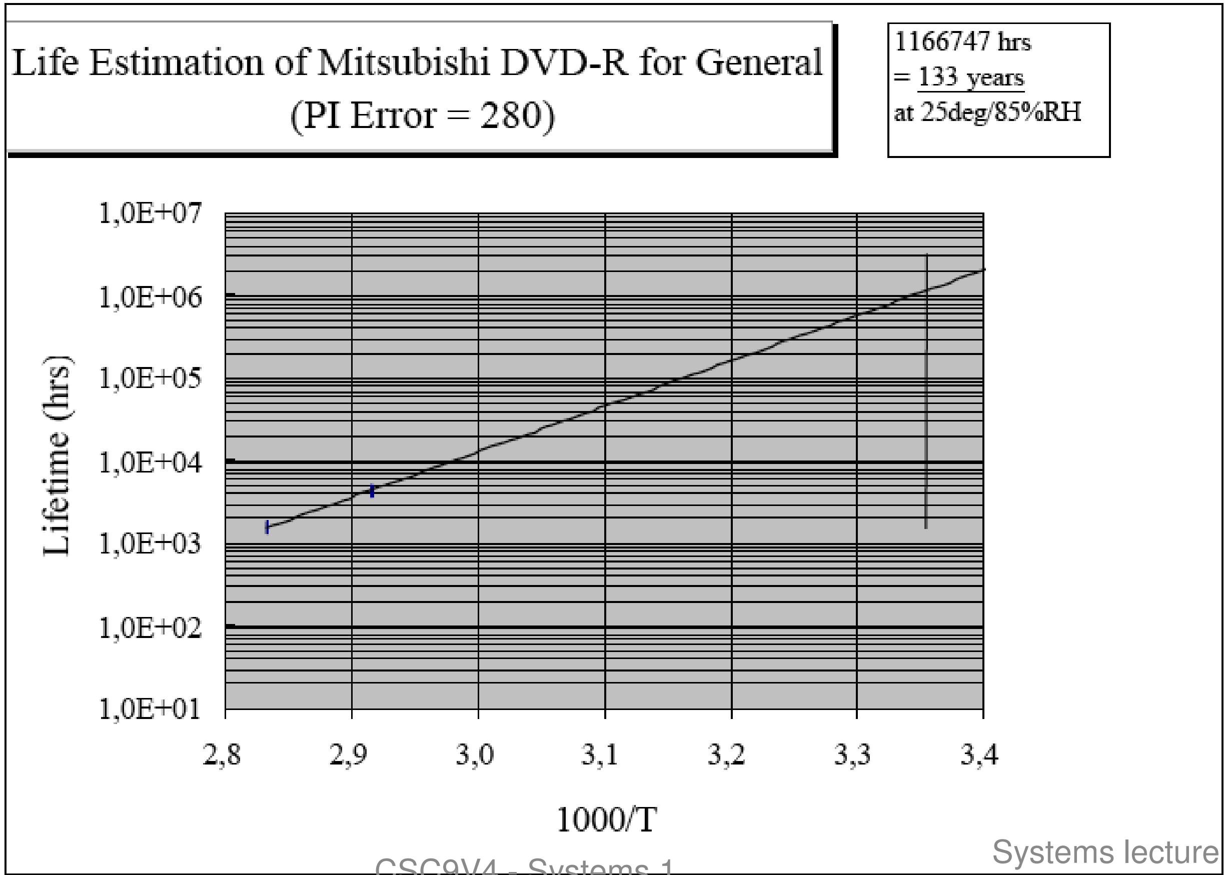
- Digital Video Disk
- Digital Versatile Disk
- smaller features than CDs:
  - 0.4 Vs 0.8 micron
  - more features per unit length
  - more spirals per unit length
  - needs a 0.65 micron laser

- 4 variations:
  - single-sided: 4.7Gb
  - SS & dual layer: 8.5 Gb
  - double-sided: 9.4Gb
  - DS & dual layer: 17Gb

- 2nd layer's features are larger

Issues

- Lifetime
  - How long will the data on a disc last? Depends on temperature it's kept at.
  - Generally less for -RW types than for -R types.
- Format creep
  - Will one be able to find a reader for this data in 10 or 20 years?



End of lecture