CS 4410
Operating Systems

RAID

Summer 2016

Cornell University

Today

- Performance and reliability using RAID.

Need for performance

- Disks are improving, but not as fast as CPUs.
  - 1970s seek time: 50-100 ms.
  - 2000s seek time: <5 ms.
  - Factor of 20 improvement in 3 decades.
- We can use multiple disks for improving performance.
- By <u>striping</u> files across multiple disks (placing parts of each file on a different disk), parallel I/O can improve access time.

Need for reliability

- Striping reduces reliability.
  - 100 disks have 1/100th mean time between failures of one disk.
- Improve reliability with <u>redundancy</u>.
  - Add redundant data to disks.
  - Lost data can be retrieved from redundant data.

RAID Structure

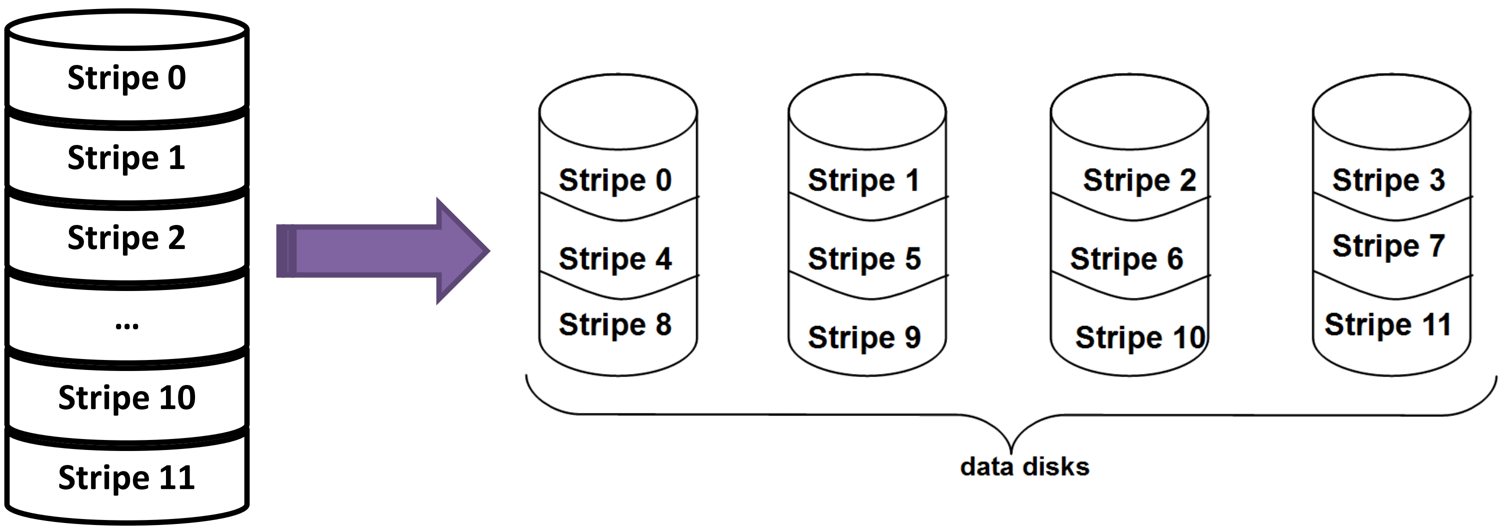
- RAID: Redundant Array of Independent Disks
- Disks are small and cheap, so it’s easy to put lots of disks in one box for increased performance and reliability.

Raid Level 0

- Files are striped across disks.
- No redundant data.
  - Any disk failure results in data loss.
- High read throughput.
- Best write throughput (no redundant data to write).

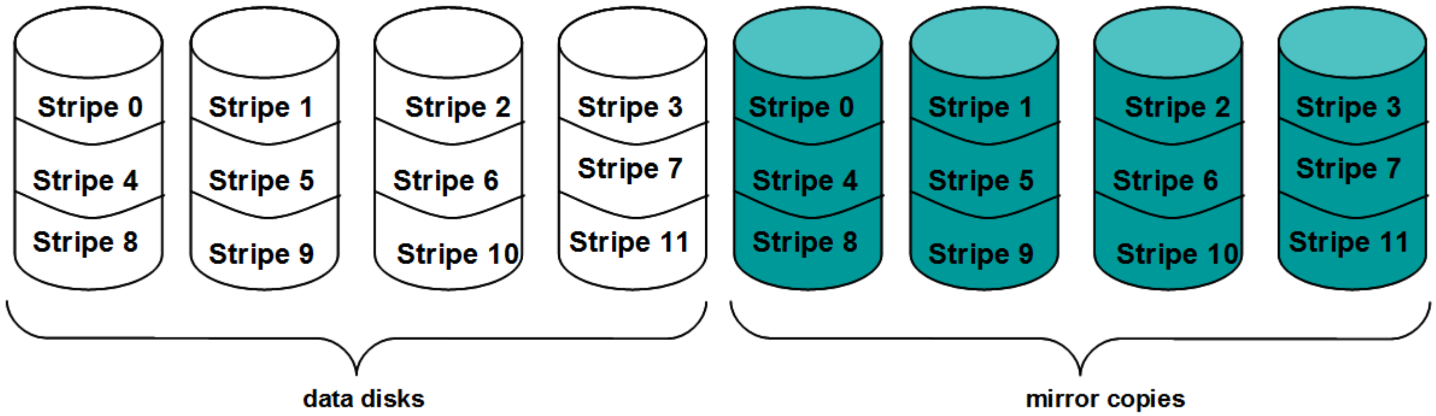
Logical representation of stored data

Physical representation of RAID 0



Raid Level 1

- Mirrored Disks
- Data is written to two places.
  - On failure, just use surviving disk.
- Write performance is same as single drive.
- Read performance is 2x better

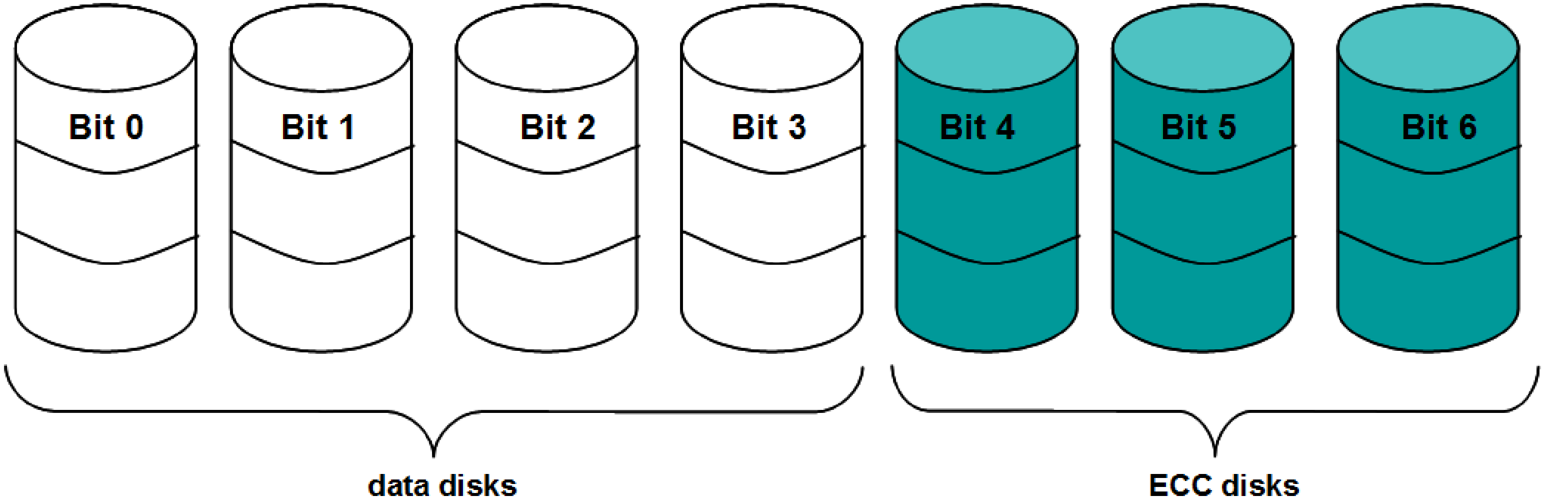


Reliability with less redundancy

- RAID1: For every byte in the data there is a mirror byte.
  - Even if the entire byte is lost/corrupted, it can be recovered by the mirror byte.
- Usually, a few bits of a byte are flipped and need to be recovered.
  - Less redundant bits are needed for recovery.
- There is a pair of functions F, H such that:
  - F takes as input a string s of n bits and produce a string ecc=F(s) of $m \leq n$ bits.
  - If (at most k) bits of s are flipped, resulting to string s', then $F(s') \neq F(s)$.
    - Error detection.
  - If (at most l) bits of s are flipped, resulting to string s', then H(s',ecc)=s.
    - Error correction.
  - k and l determine the strength of F,H to detect and recover flipped bits.
  - ecc is called error correction code.

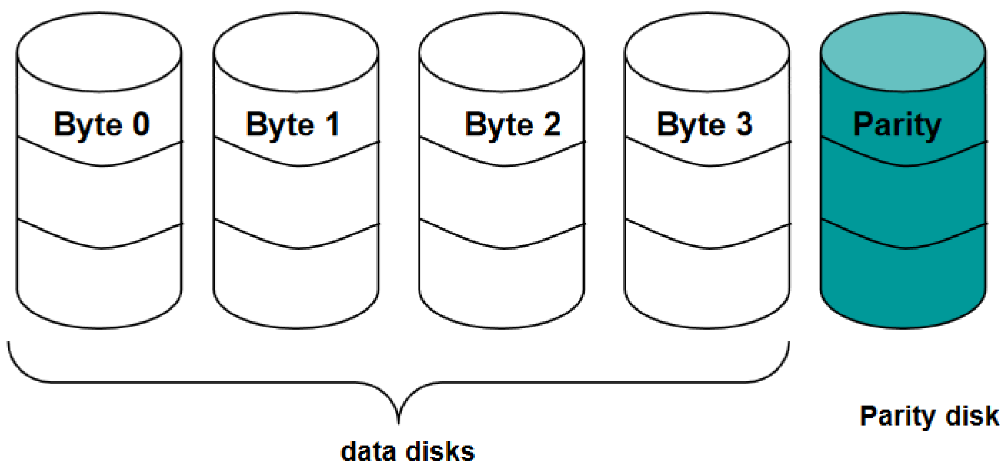
Raid Level 2

- Bit-level striping with error correction codes.
- Single access at a time.
- In the example:
  - F(Bit 0, Bit 1, Bit 2, Bit 3) = Bit 4, Bit 5, Bit 6
  - At most 2 bit errors can be detected.
  - At most 1 bit error can be corrected.



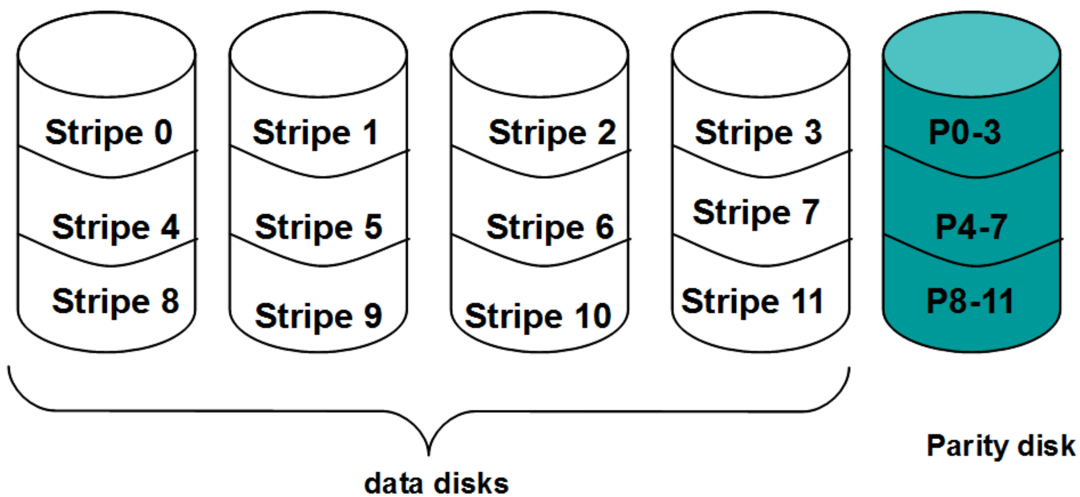
Raid Level 3

- Byte-level striping with parity disk.
  - F(Byte 0, Byte 1, Byte 2, Byte 3) = Byte 0 XOR Byte 1 XOR Byte 2 XOR Byte 3
  - At most 1 byte can be corrected.
- An external mechanism detects with disk has failed, and thus which bit has been corrupted.



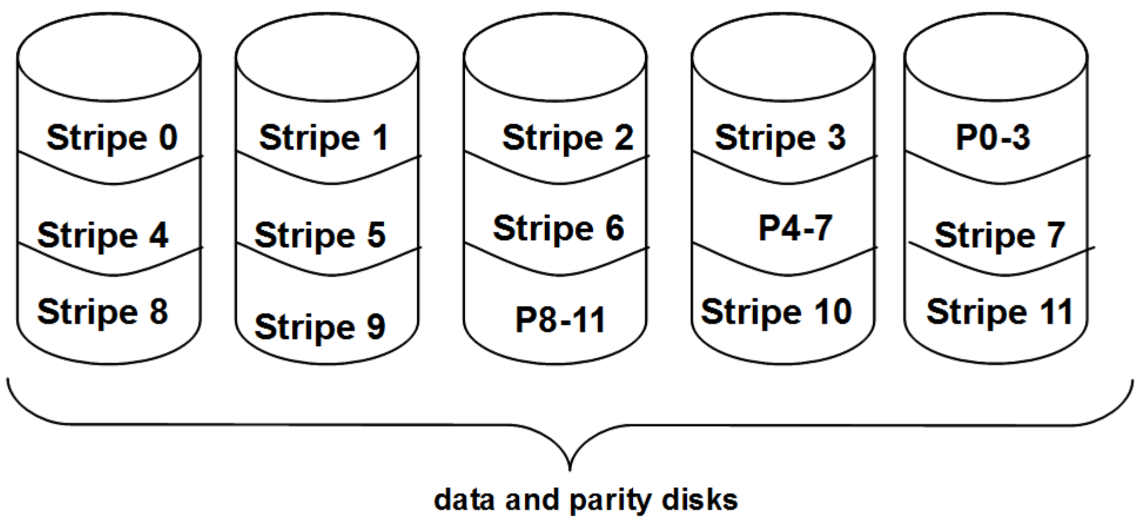
Raid Level 4

- Combines Level 0 and 3 – block-level parity with stripes.
- A large read can access all the data disks.
- A large write can access all data disks plus the parity disk.
- Heavy load on the parity disk.

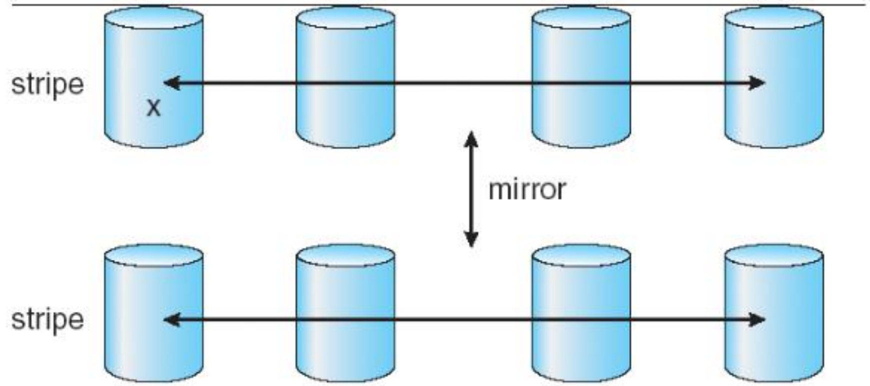


Raid Level 5

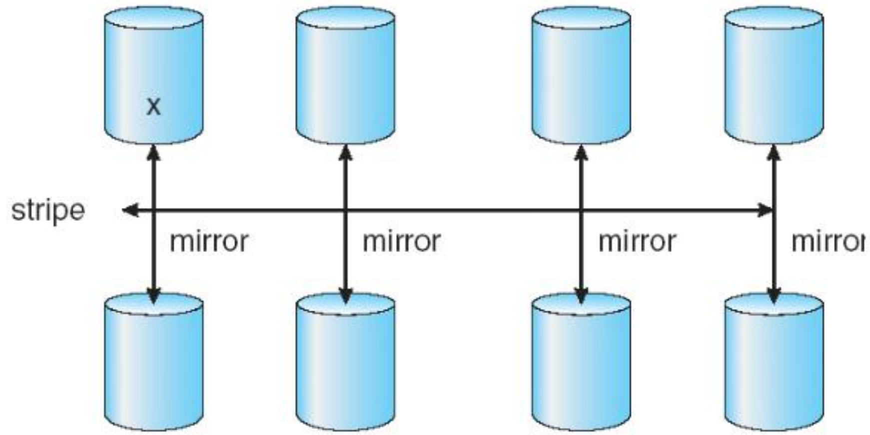
- Block Interleaved Distributed Parity
- Like parity scheme, but distribute the parity info over all disks (as well as data over all disks).



RAID 01 and RAID 10



a) RAID 0 + 1 with a single disk failure.



b) RAID 1 + 0 with a single disk failure.

Today

- Performance and reliability using RAID.

Coming up…

- Next lecture: file system implementation
- HW4: ex 1,2,3,4
- Office hours:
  - Tuesday 10-11am, instead of Monday