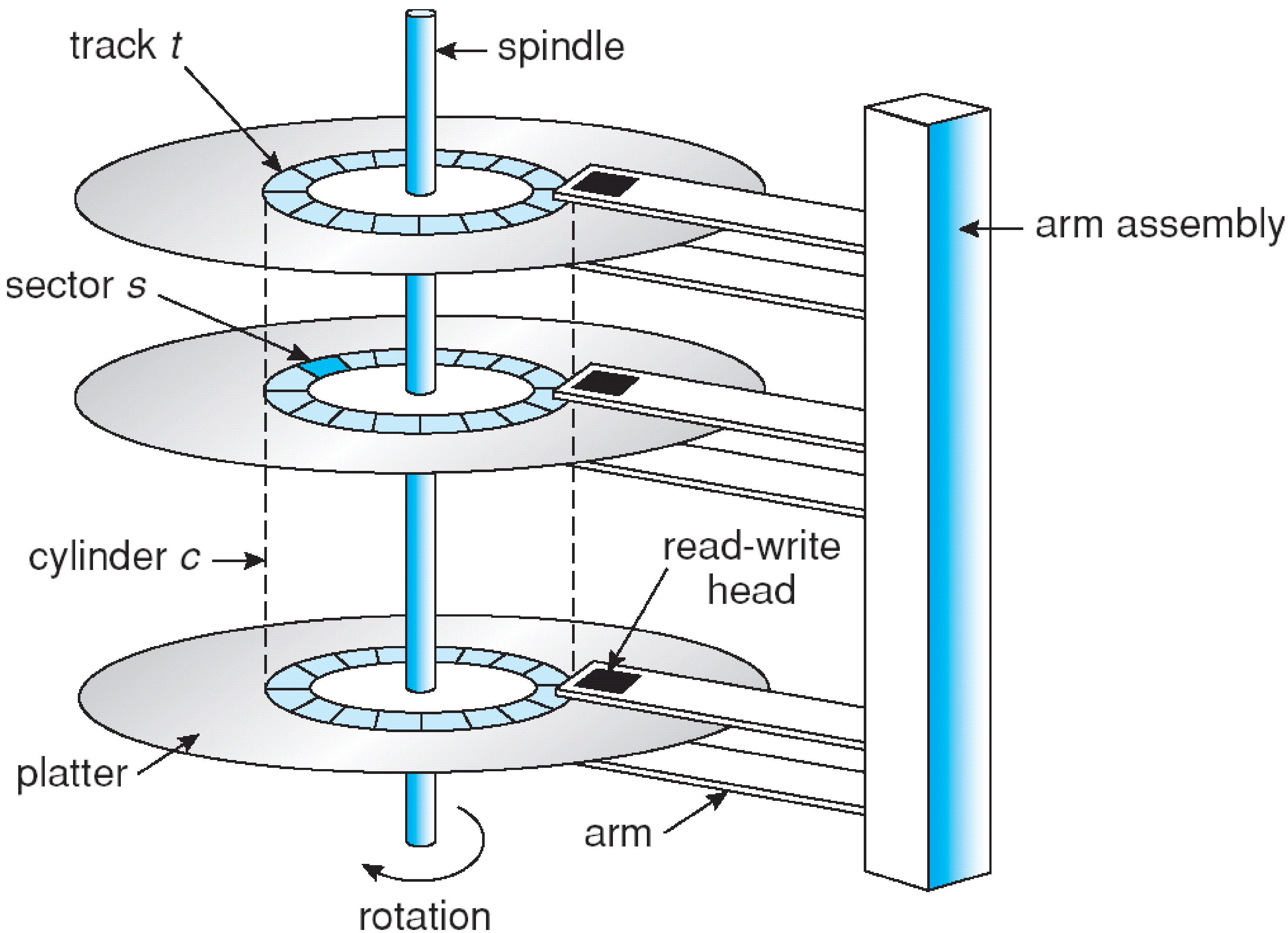
W4118 Operating Systems

Instructor: Junfeng Yang

Outline

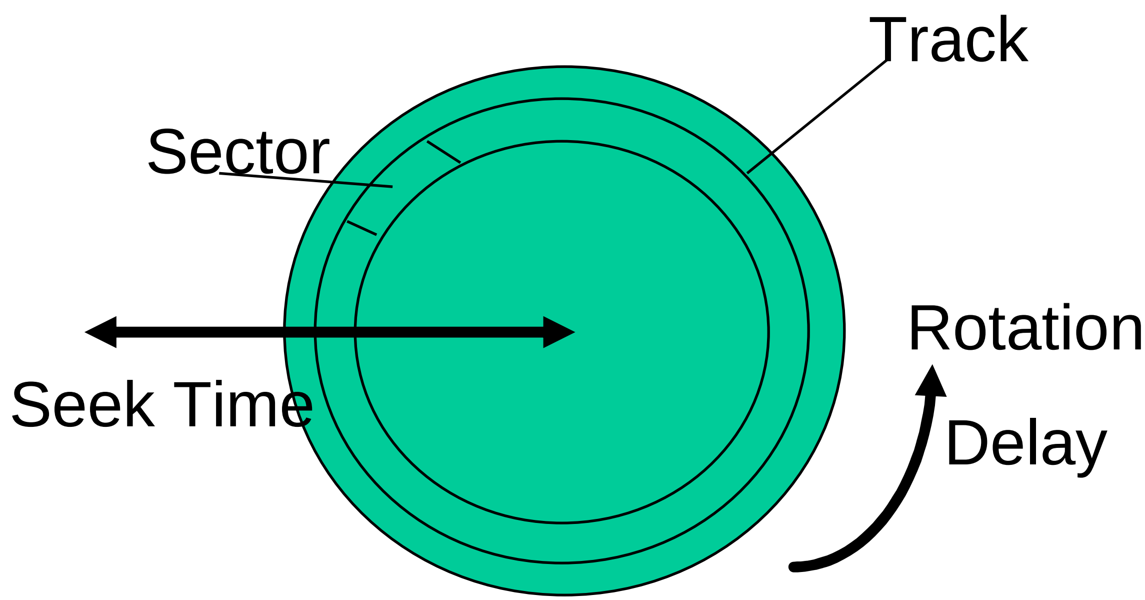
- Disk
- Redundant Arrays of Inexpensive Disks (RAID)

Disk structure



Disk latencies

- Latency includes:
  - Rotational delay: to get to the sector
  - Seek time: to get to the track
  - Transfer time: get bits off the disk



Disk parameters

	Barracuda 180	Cheetah X15 36LP
Capacity	181GB	36.7GB
Platters/Heads	12/24	4/8
Cylinders	24,247	18,479
Sectors/track	~609	~485
Rotational speed	7200 RPM	15000 RPM
Rotational latency (ms)	4.17	2.0
Avg seek (ms)	7.4	3.6
Track-2-track(ms)	0.8	0.3

Disk interface

- From FS perspective: disk is addressed as a one dimension array of logical sectors
- Disk controller maps logical sector to physical sector identified by surface #, track #, and sector #
- Default mapping: sequential
  - Logical sector 0 is the first sector of the first (outermost) track of the first surface
  - Logical sector address incremented within track, then tracks within cylinder, then across cylinders, from outermost to innermost

Sequential v.s. random access latency

- Sequential access latency
  - Seek to the right track
  - Rotate to the right sector
  - Transfer
- Random access latency
  - Seek to the right track
  - Rotate to the right sector
  - Transfer
  - Repeat
- Sequential access is faster than random access because it needs less seek and rotation

Pros and cons of disk interface

- Pros
  - Simple to program
  - Default mapping reduces seek time for sequential access
- Cons
  - FS can't precisely see mapping
  - Reverse-engineer mapping in OS is difficult
    - # of sectors per track changes
    - Disk silently remaps bad sectors

Disk technology trends

- Data → more dense
  - More bits per square inch
  - Disk head closer to surface
  - Create smaller disk with same capacity
- Disk geometry → smaller
  - Spin faster → Increase b/w, reduce rotational delay
  - Faster seek
  - Lighter weight
- Disk price → cheaper
- Density improving more than speed (mechanical limitations)

New mass storage technologies

- New memory-based mass storage technologies avoid seek time and rotational delay
  - NAND Flash
  - Battery-backed DRAM (NVRAM)
- Disadvantages
  - Price: more expensive than same capacity disk
  - Reliability: more likely to lose data
- Open research question: how to effectively use flash in commercial storage systems

Outline

- Disk
- Redundant Arrays of Inexpensive Disks (RAID)

RAID motivation

- Performance
  - Disks are slow compared to CPU
  - Disk speed improves slowly compared to CPU
- Reliability
  - In single disk systems, one disk failure → data loss
- Cost
  - A single fast, reliable disk is expensive

RAID idea

- RAID idea: use redundancy to improve performance and reliability
  - Redundant array of cheap disks as one storage unit
  - Fast: simultaneous read and write disks in the array
  - Reliable: use parity to detect and correct errors

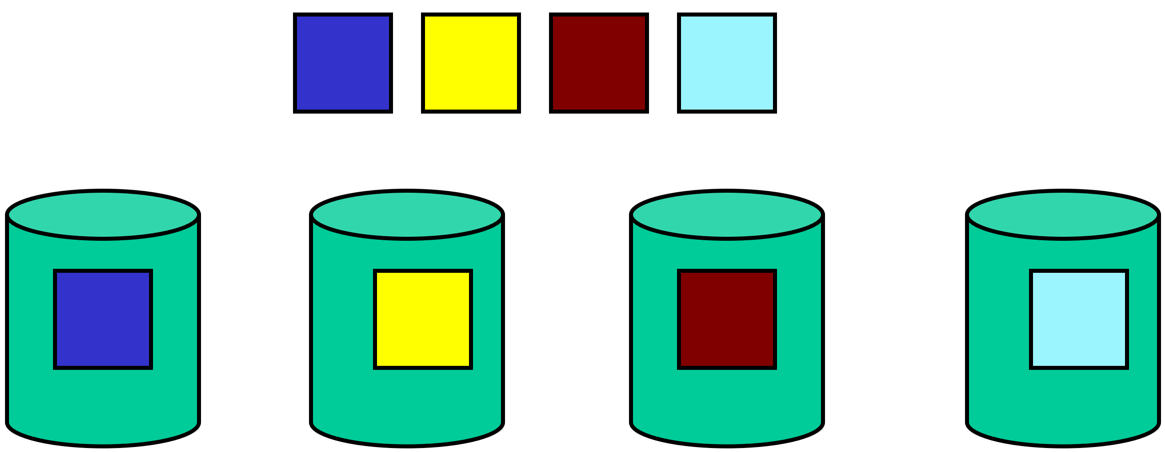
- RAID can have different redundancy levels, achieving different performance and reliability
  - Seven different RAID levels (0-6)

Evaluating RAID

- Performance
  - (Large) sequential read, write, read-modify-write
  - (Small) random read, write, read-modify-write
- Reliability
  - Tolerance of disk failures
- Cost
  - Storage utilization: data capacity / total capacity

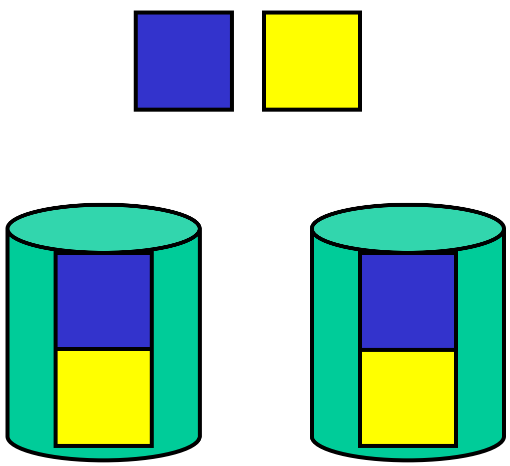
RAID 0: non-redundant striping

- Structure
  - Data striped across all disks in an array
  - No parity
- Advantages:
  - Good performance: with N disks, speed up N times
- Disadvantages:
  - Poor reliability: one disk failure → data loss



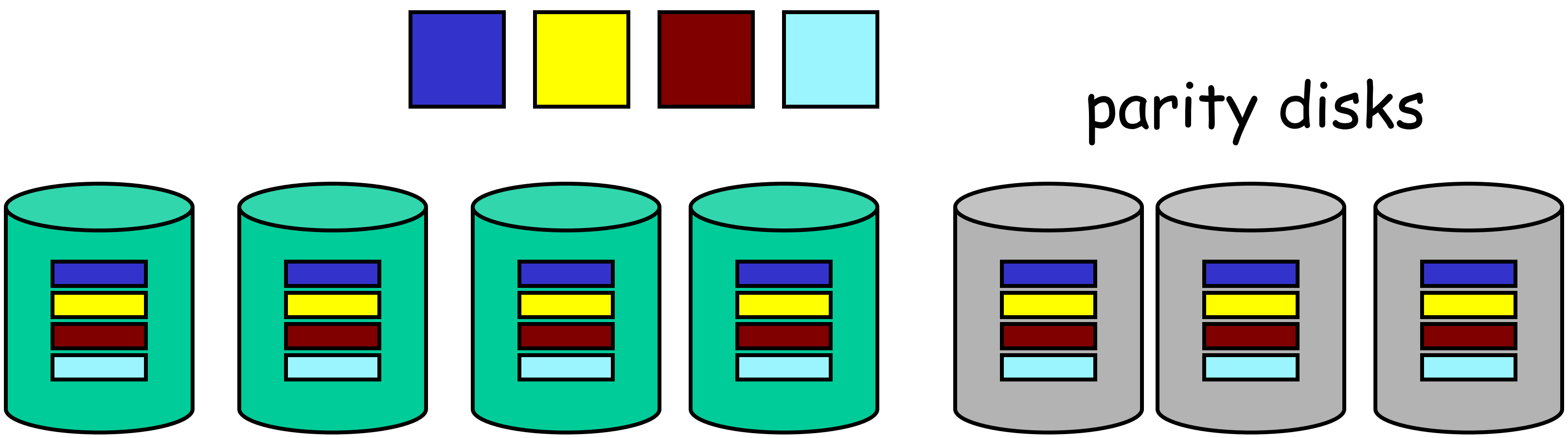
RAID 1: mirroring

- Structure
  - Keep a mirrored (shadow) copy of data
- Advantages
  - Good reliability: one disk failure OK
  - Good read performance
- Disadvantage
  - High cost: one data disk requires one parity disk



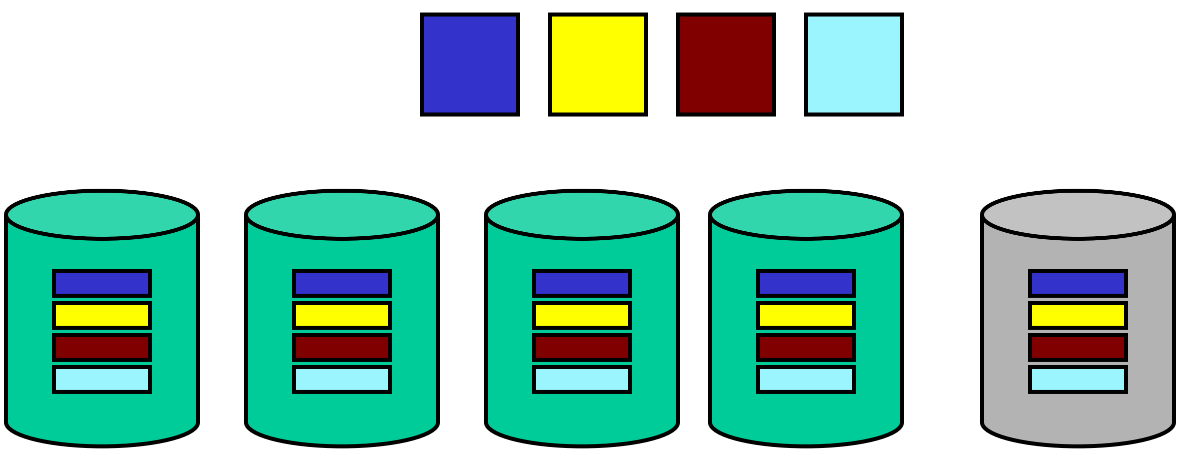
RAID2: error-correction parity

- Structure
  - A data sector striped across data disks
  - Compute error-correcting parity and store in parity disks
- Advantages
  - Good reliability with higher storage utilization than mirroring
- Disadvantages
  - Unnecessary cost: disk can already detect failure
  - Poor random performance



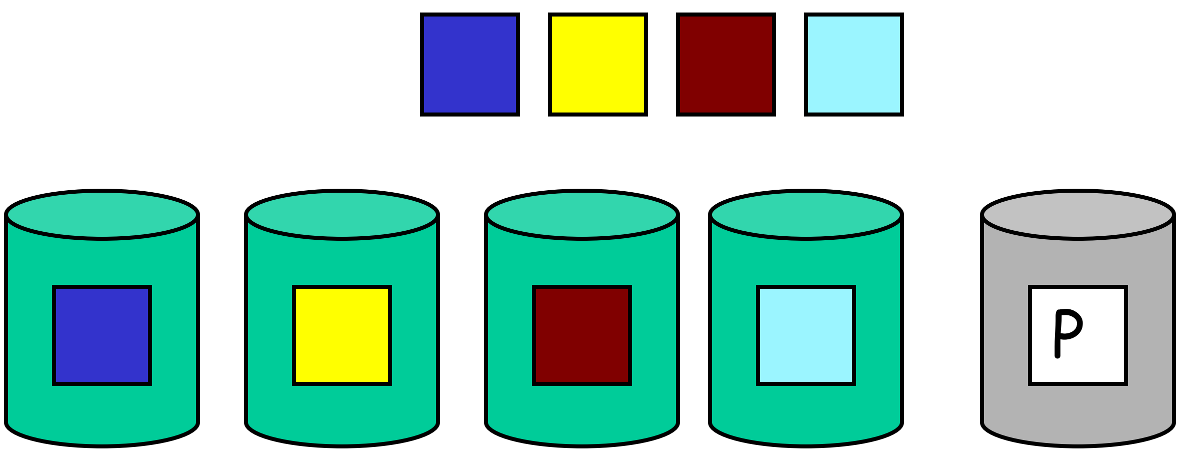
RAID3: bit-interleaved parity

- Structure
  - Single parity disk (XOR of each stripe of a data sector)
- Advantages
  - Same reliability with one disk failure as RAID2 since disk controller can determine what disk fails
  - Higher storage utilization
- Disadvantages
  - Poor random performance



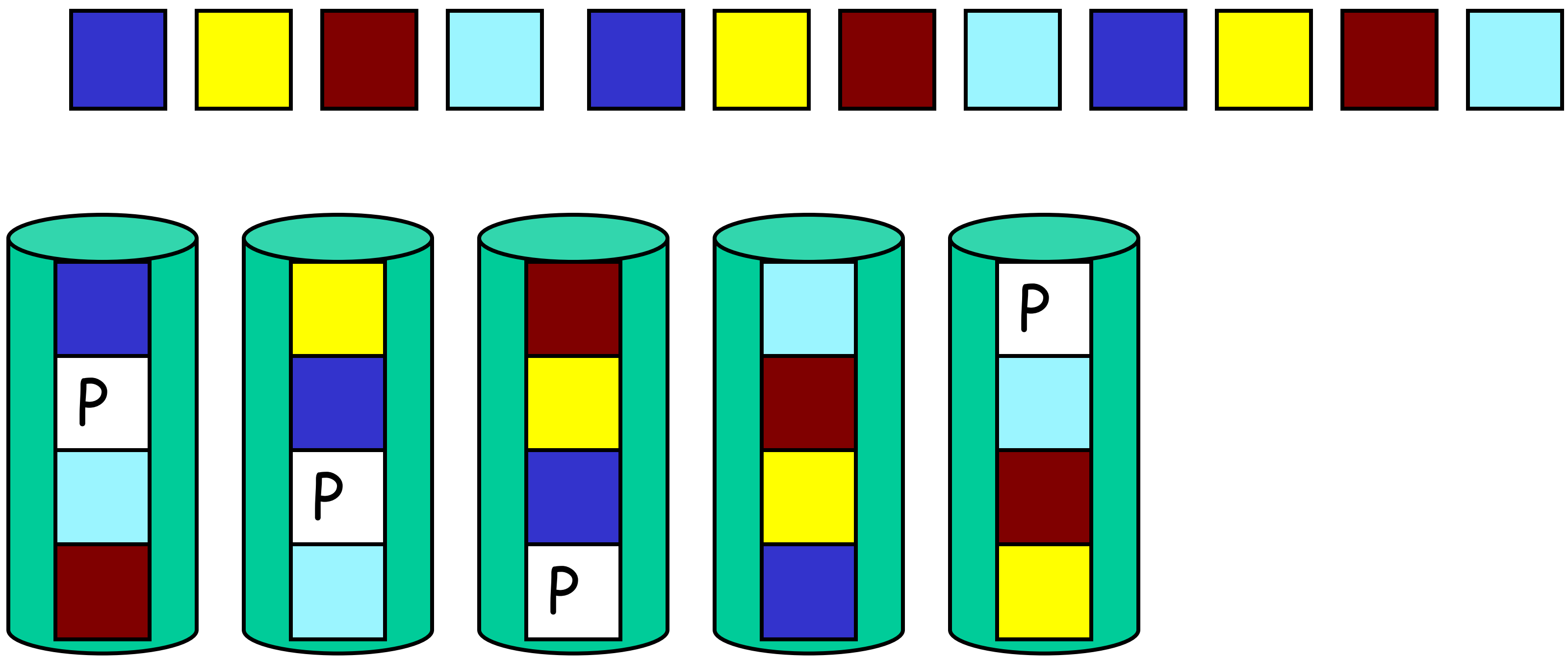
RAID4: block-interleaved parity

- Structure
  - A set of data sectors (parity group) striped across data disks
- Advantages
  - Same reliability as RAID3
  - Good random read performance
- Disadvantages
  - Poor random write and read-modify-write performance



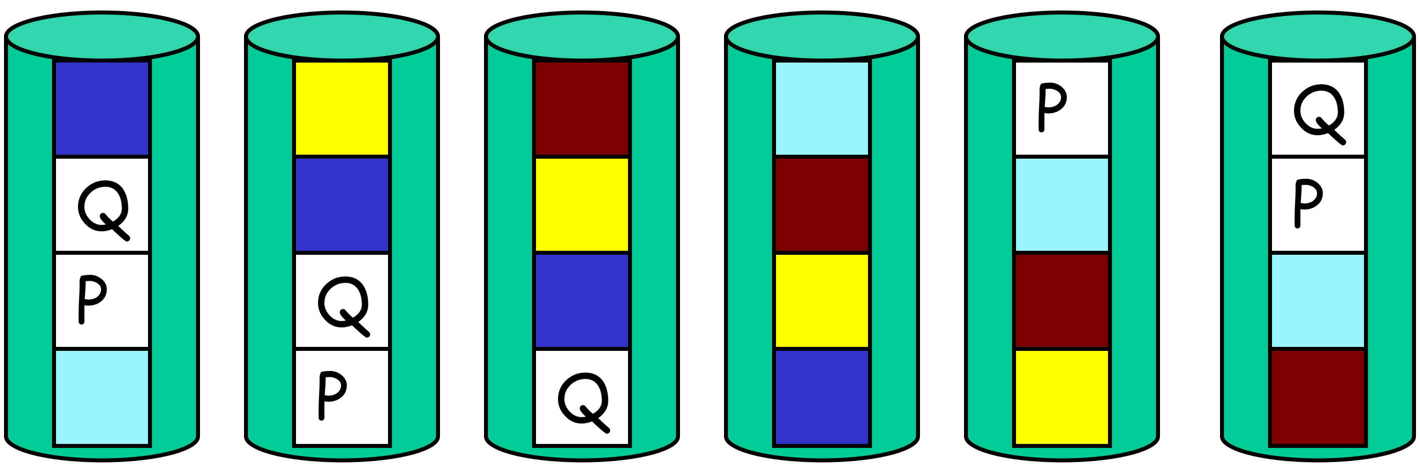
RAID5: block-interleaved distributed parity

- Structure
  - Parity sectors distributed across all disks
- Advantages
  - Good performance

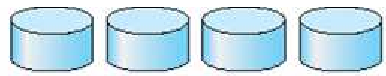


RAID6: P+Q redundancy

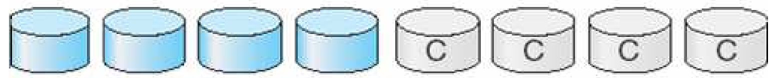
- Structure
  - Same as RAID 5 except using two parity sectors per parity group
- Advantages
  - Can tolerate two disk failures



RAID levels

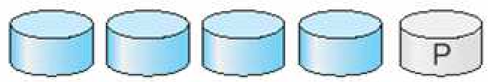


(a) RAID 0: non-redundant striping.

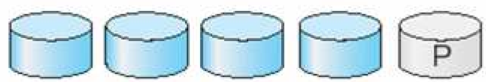


(b) RAID 1: mirrored disks.

(c) RAID 2: memory-style error-correcting codes.



(d) RAID 3: bit-interleaved parity.



(e) RAID 4: block-interleaved parity.

(f) RAID 5: block-interleaved distributed parity.

(g) RAID 6: P + Q redundancy.