COA18 Computer Organization and Architecture

Lecture 08

RAID

Tongwei Ren

Software Institute, Nanjing University

Oct. 8, 2018

Review

- Error correction
- Parity checking
- Hamming code
- Cyclic redundancy check

RAID

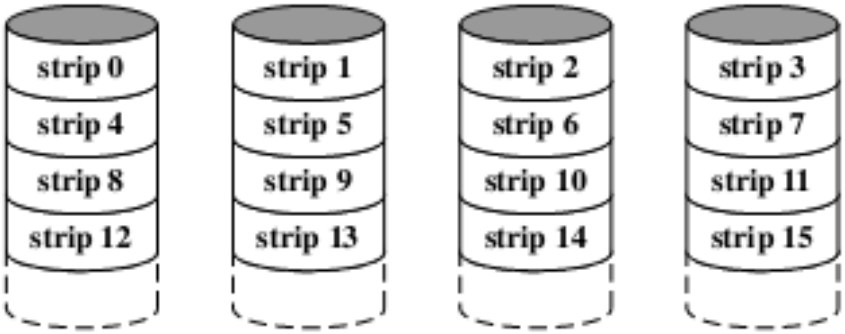
- Redundant Array of Independent Disks
- Basic idea
  - With multiple disks, separate I/O requests, even a single I/O request, can be handled in parallel, as long as the data required reside on separate disks
- Characteristic
  - RAID is a set of physical disk drives viewed as a single logical drive
  - Data are distributed across the physical drives
  - Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure

RAID Level

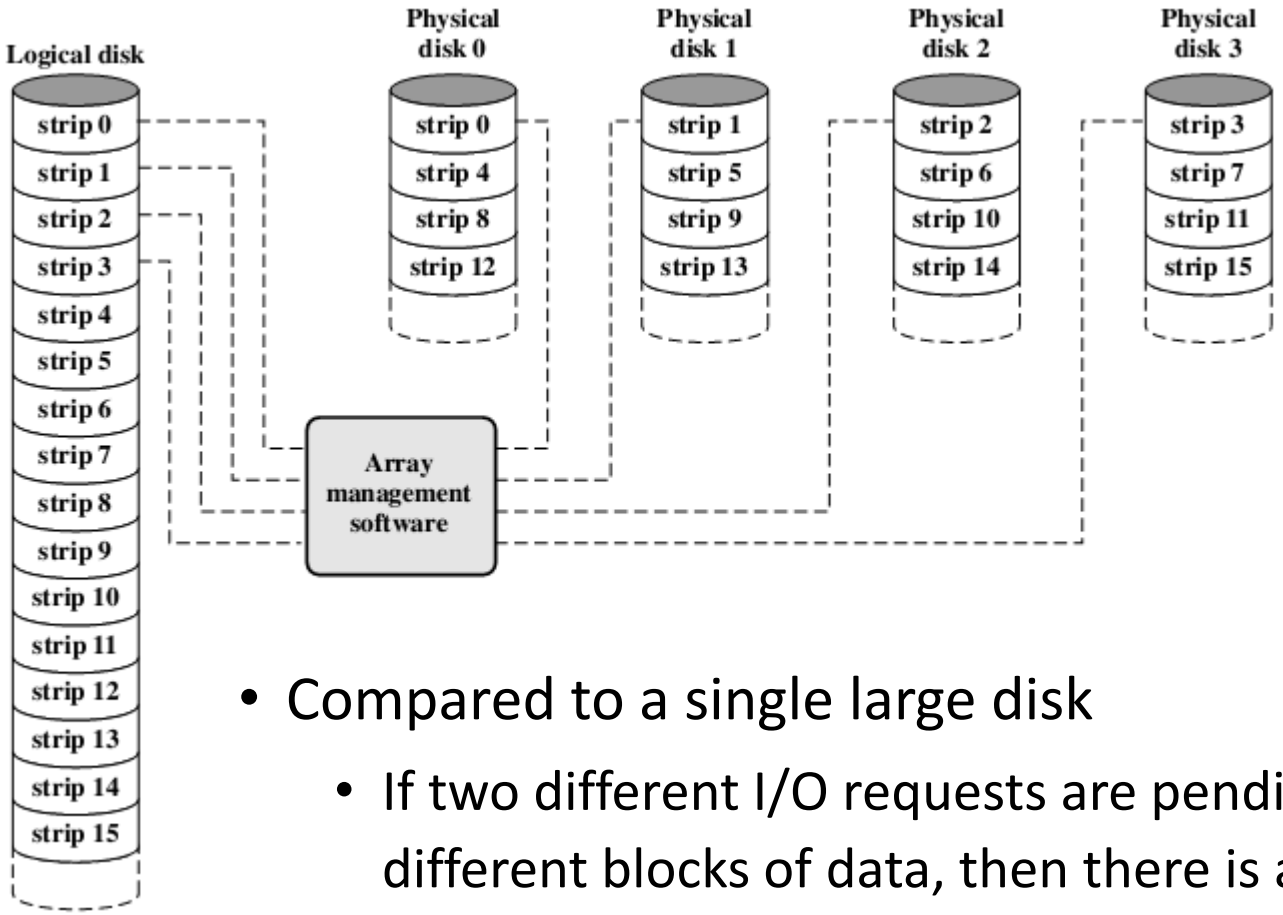
Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code	$N + m$	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N + 2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

RAID Level 0

- The data are striped across the available disks
- Not include redundancy to improve performance (not a true member of the RAID family)
- Usage
  - High rate data transfer
  - High I/O request rate



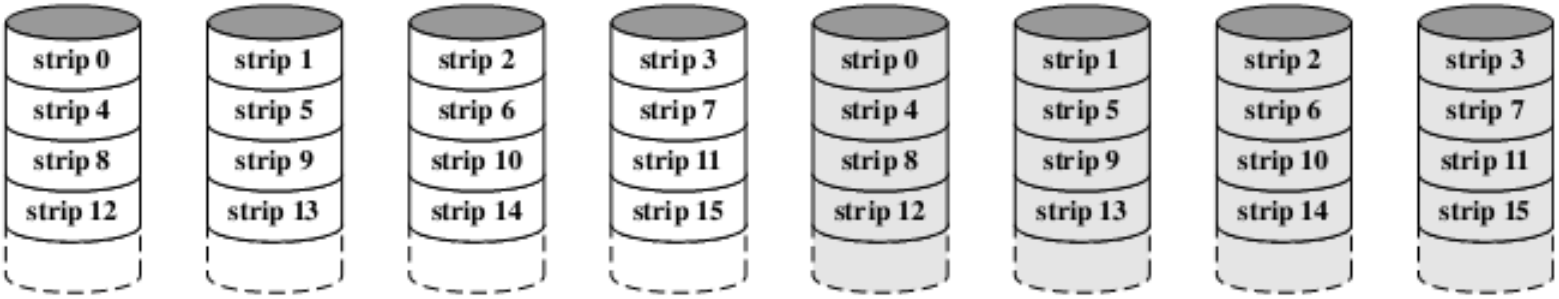
RAID Level 0 (cont.)



- Compared to a single large disk
  - If two different I/O requests are pending for two different blocks of data, then there is a good chance that the requested blocks are on different disks

RAID Level 1

- Redundancy is achieved by the simple expedient of duplicating all the data
- Data striping is used



RAID Level 1 (cont.)

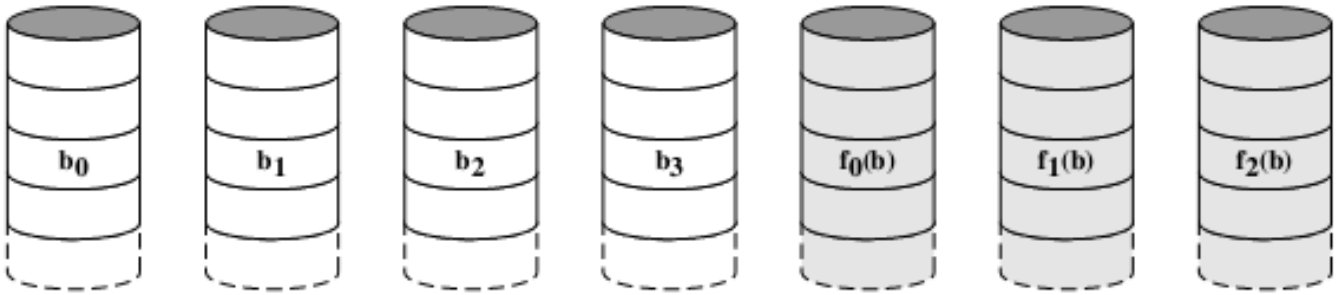
- Advantage
  - A read request can be serviced by either of the two disks that contains the requested data, whichever one involves the minimum seek time plus rotational latency
  - A write request requires that both corresponding strips be updated, but this can be done in parallel
  - Recovery from a failure is simple
- Disadvantage
  - High cost

RAID Level 1 (cont.)

- Usage
  - Limited to drives that store system software and data and other highly critical files
- Compared to RAID 0
  - RAID 1 can achieve high I/O request rates if the bulk of the requests are reads, in which the performance of RAID 1 can approach double of that of RAID 0
  - If a substantial fraction of the I/O requests are write requests, then there may be no significant performance gain over RAID 0

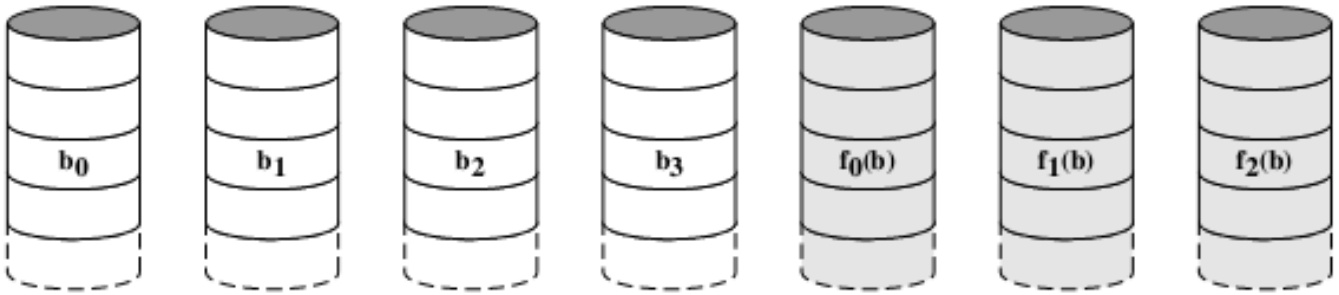
RAID Level 2

- Use of a parallel access technique
  - All member disks participate in the execution of every I/O request
  - the spindles of the individual drives are synchronized so that each disk head is in the same position on each disk at any given time
- Data striping is used
  - The strips are very small, often a single byte or word



RAID Level 2 (cont.)

- An error-correcting code is calculated across corresponding bits on each data disk, and the bits of the code are stored in the corresponding bit positions on multiple parity disks
  - Hamming code is typically used
- Read
  - The requested data and the error-correcting code are fetched
- Write
  - All data disks and parity disks must be accessed



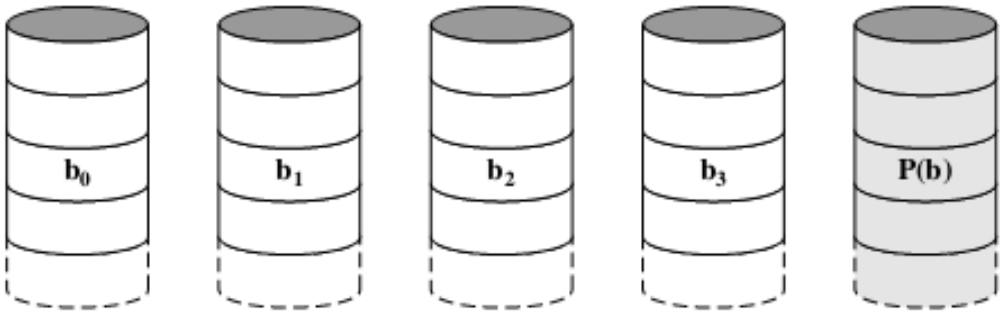
RAID Level 2 (cont.)

- Disadvantage
  - Still rather costly
  - Only be an effective choice in an environment in which many disk errors occur, which is overkill and is not implemented to individual disks and disk drives with the high reliability

RAID Level 3

- Use of a parallel access technique
- Use very small data striping
- A simple parity bit is computed for the set of individual bits in the same position on all of the data disks
  - Can be used to reconstruct data when a drive fails

$$b_0 = P(b) \oplus b_1 \oplus b_2 \oplus b_3$$

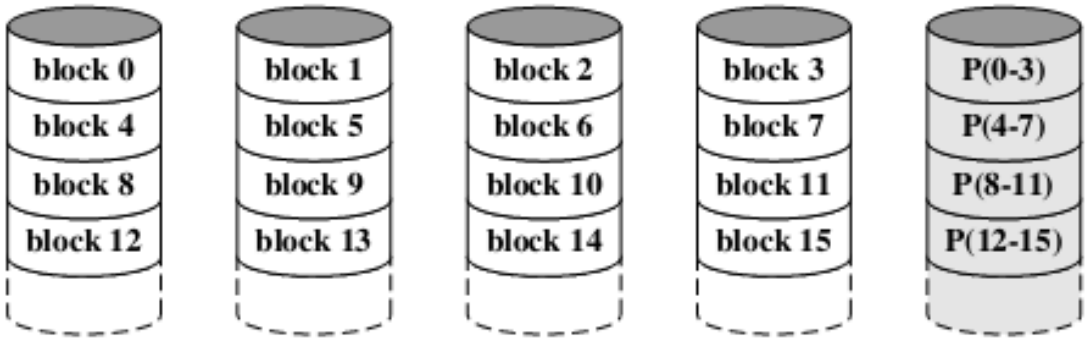


RAID Level 3 (cont.)

- Performance
  - Achieve very high data transfer rates
    - For large transfers, the performance improvement is especially noticeable
  - Only one I/O request can be executed at a time
    - In a transaction-oriented environment, performance suffers

RAID Level 4

- Use of an independent access technique
  - Each member disk operates independently, so that separate I/O requests can be satisfied in parallel
- Data striping is used
- A bit-by-bit parity strip is calculated across corresponding strips on each data disk, and the parity bits are stored in the corresponding strip on the parity disk



RAID Level 4 (cont.)

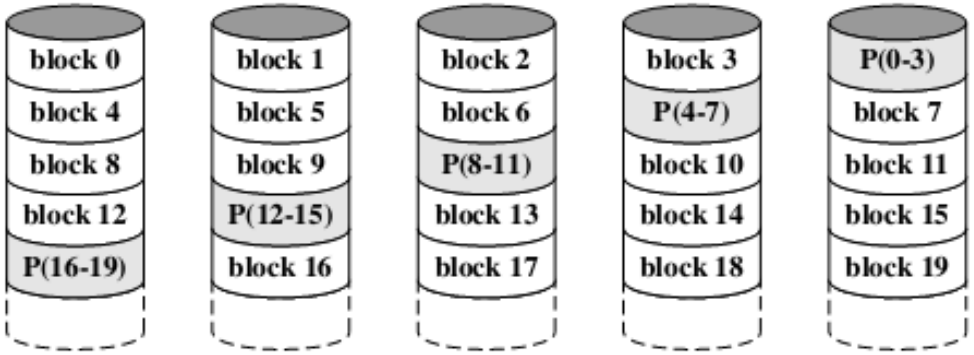
- Performance
  - RAID 4 involves a write penalty when an I/O write request of small size is performed
    - Each time that a write occurs, the array management software must update not only the user data but also the corresponding parity bits

$$P' = P \oplus b_0 \oplus b_0{}'$$

  - In the case of a larger size I/O write that involves strips on all disk drives, parity is easily computed by calculation using only the new data bits
  - Every write operation must involve the parity disk, which therefore can become a bottleneck

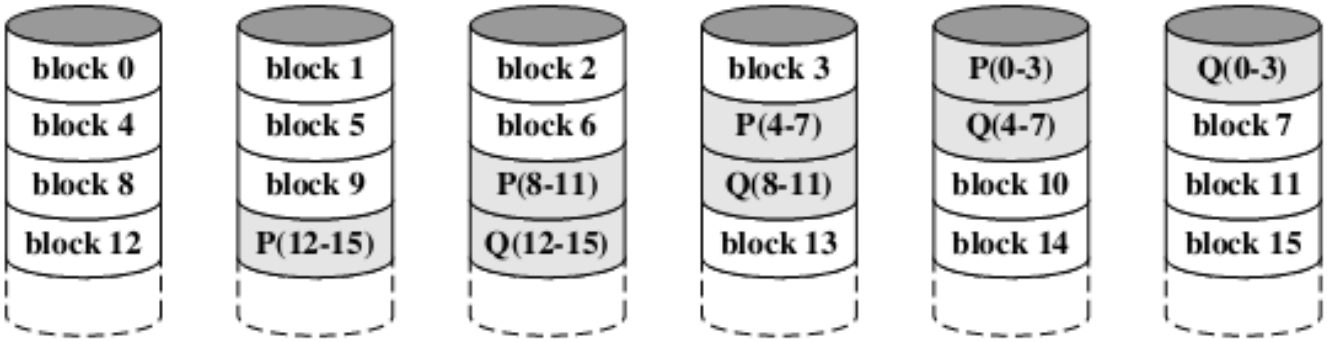
RAID Level 5

- Similar to RAID level 4
- Distributes the parity strips across all disks
  - Avoid the potential I/O bottleneck



RAID Level 6

- Two different parity calculations are carried out and stored in separate blocks on different disks
- Advantage
  - Provide extremely high data availability: three disks would have to fail within the MTTR interval to cause data to be lost
- Disadvantage
  - Write penalty: each write affects two parity blocks



RAID Comparison

Level	Advantages	Disadvantages	Applications
0	I/O performance is greatly improved by spreading the I/O load across many channels and drives No parity calculation overhead is involved Very simple design Easy to implement	The failure of just one drive will result in all data in an array being lost	Video production and editing Image Editing Pre-press applications Any application requiring high bandwidth
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability

RAID Comparison (cont.)

Level	Advantages	Disadvantages	Applications
2	Extremely high data transfer rates possible The higher the data transfer rate required, the better the ratio of data disks to ECC disks Relatively simple controller design compared to RAID levels 3, 4 & 5	Very high ratio of ECC disks to data disks with smaller word sizes—inefficient Entry level cost very high—requires very high transfer rate requirement to justify	No commercial implementations exist/ not commercially viable
3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput

RAID Comparison (cont.)

Level	Advantages	Disadvantages	Applications
4	Very high Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency	Quite complex controller design Worst write transaction rate and Write aggregate transfer rate Difficult and inefficient data rebuild in the event of disk failure	No commercial implementations exist/ not commercially viable
5	Highest Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency Good aggregate transfer rate	Most complex controller design Difficult to rebuild in the event of a disk failure (as compared to RAID level 1)	File and application servers Database servers Web, e-mail, and news servers Intranet servers Most versatile RAID level
6	Provides for an extremely high data fault tolerance and can sustain multi-ple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications

Summary

- RAID 0
- RAID 1
- RAID 2
- RAID 3
- RAID 4
- RAID 5
- RAID 6

Thank You