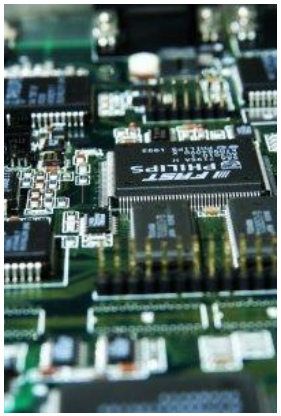
CSE 401

Computer Engineering (2)

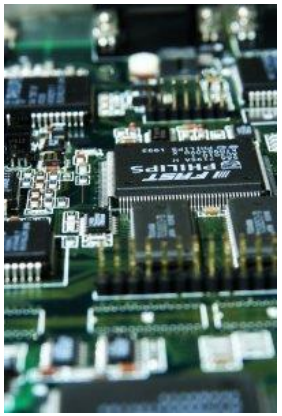
هندسة الحاسبات (2)



4th year, Comm. Engineering

Winter 2016

Lecture #4



Dr. Hazem Ibrahim Shehata

Dept. of Computer & Systems Engineering

Credits to Dr. Ahmed Abdul-Monem Ahmed for the slides

Adminstrivia

- Website:
  - —Moved to a new server (GitHub)!

    http://hshehata.github.io/courses/zu/cse401/
- Assignment #1:
  - —Released last week.
  - —Due: **Thursday, March 10, 2016**.

Website: http://hshehata.github.io/courses/zu/cse401/
Office hours: Monday 11:30am – 12:30pm

Chapter 6. External Memory (*Cont.*)

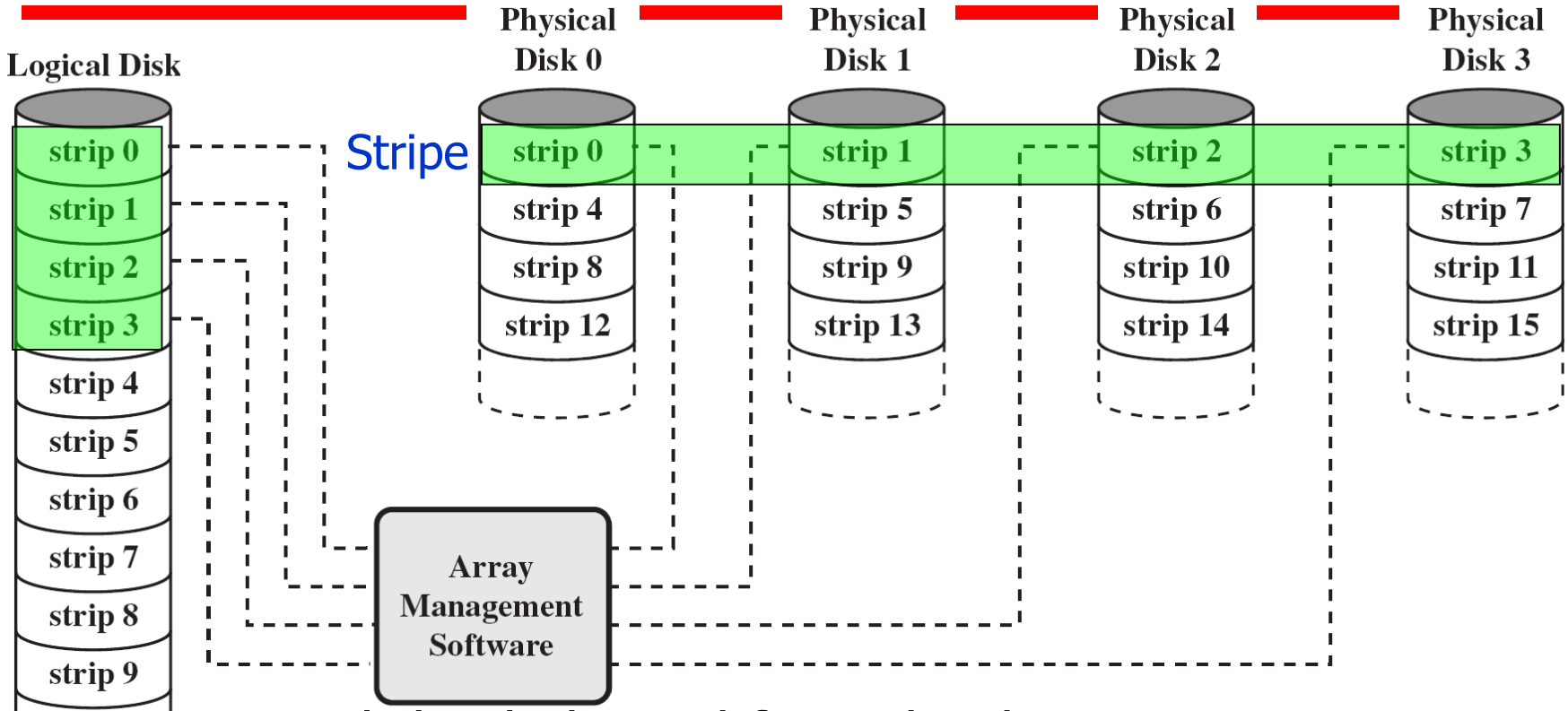
Types of External Memory

- Magnetic Disk
- Redundant Array of Independent Disks (RAID)
- Solid-State Drive (SSD)
- Optical Disk
- Magnetic Tape

RAID

- **Problem**: Improvement rate in secondary storage < rate for CPU and MM.
- **Solution**: Can't improve one-disk perf. ➔ use multiple in parallel!
- Array of disks
  - —Operate independently and in parallel.
  - —Single I/O request can be handled in parallel if the block is distributed across multiple disks.
  - —Separate I/O requests can be handled in parallel.
  - —Performance metrics: depend on request patterns & data layout.
    - I/O data transfer rate.
    - I/O request rate (response time).
    - Recovery from errors & disk failure.
- RAID: Redundant Array of Independent Disks.
  - —7 levels in common use, not a hierarchy.
  - —Set of physical disks viewed by OS as single logical drive.
  - —Data distributed across physical drives.
  - —Can use redundant capacity to store parity information.

RAID 0 - Stripping without Mirroring or Parity



- No disks dedicated for redundancy or parity!
- Data is striped and distributed across disks.
- Strip: physical block, sector, or another unit.
- Strips are mapped round robin to consecutive disks.
- Stripe: Set of logically consecutive strips that maps exactly one strip to each array member.

RAID 0 - Performance

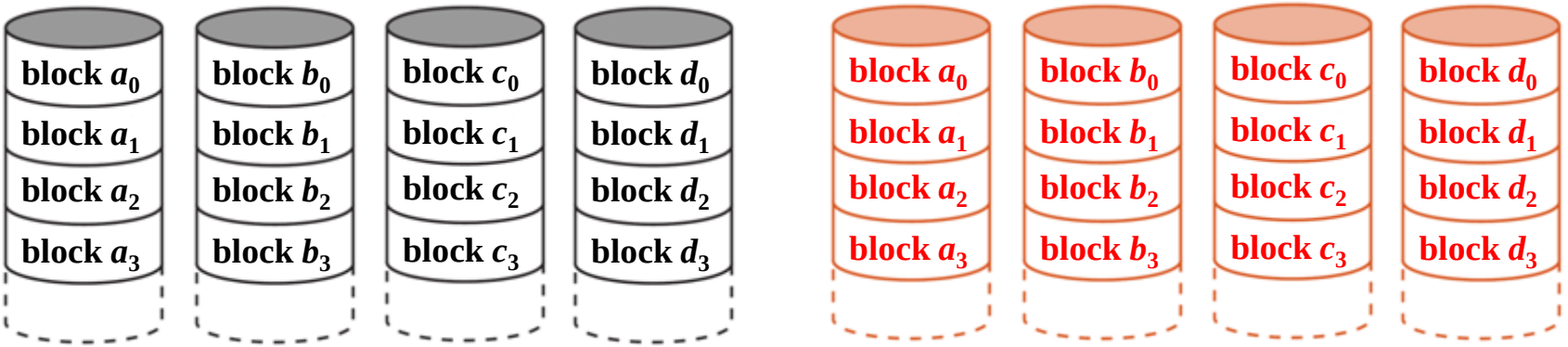
- RAID 0 for high I/O data transfer rate
  - System is configured s. t. each (single) I/O request can be processed by multiple disks in parallel ➔ Less tr. time ➔ Higher tr. rate!
  - Two requirements (to experience high data transfer rate):
    1. High transfer capacity between memory and disks (internal controller buses, I/O buses, memory buses, *etc.*).
    2. **Small strip** ➔ Higher chance that any single I/O request requires data from multiple strips located on different disks.
- RAID 0 for high I/O request rate
  - System configured s. t. multiple I/O requests can be processed by multiple disks in parallel ➔ Higher I/O req. rate!
  - There are typically hundreds of I/O requests per second by multiple independent applications, or single one.
  - Balance I/O load across multiple disks ➔ Achieve high I/O req. rate.
  - **Large strip** ➔ few seeks/disk per request ➔ less I/O queuing time.

RAID 0 - Pros and Cons

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

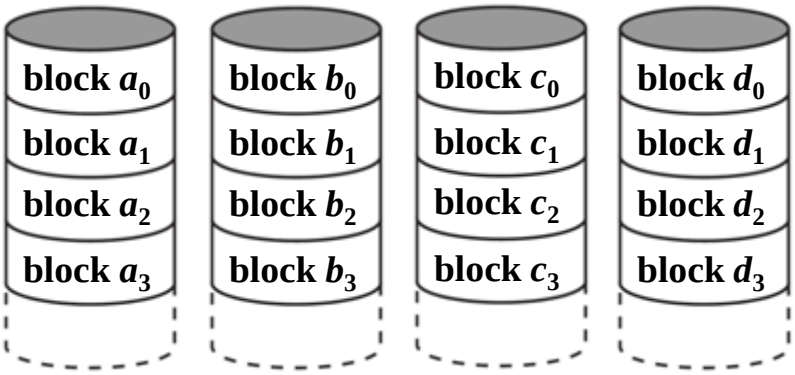
RAID 1 - Mirroring without Stripping or Parity

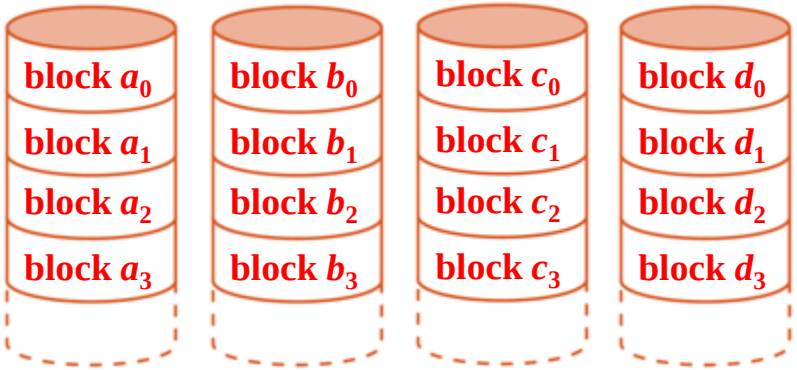
- Duplicate data (without striping) ➔ mirrored disks.
- 2 copies of each block on separate disks.
  - Read from either disk (the one with min. access time).
  - Write to both (in parallel) ➔ Time = larger access time.
- Recovery is simple: swap faulty disk & re-mirror.
- Expensive ➔ used to store system S/W & critical files.
- I/O transfer rate
  - read: > single disk, write: ≈ single disk.
- I/O request rate
  - read: ≈ 2x single disk, write: ≈ single disk.



RAID 1 - Pros and Cons

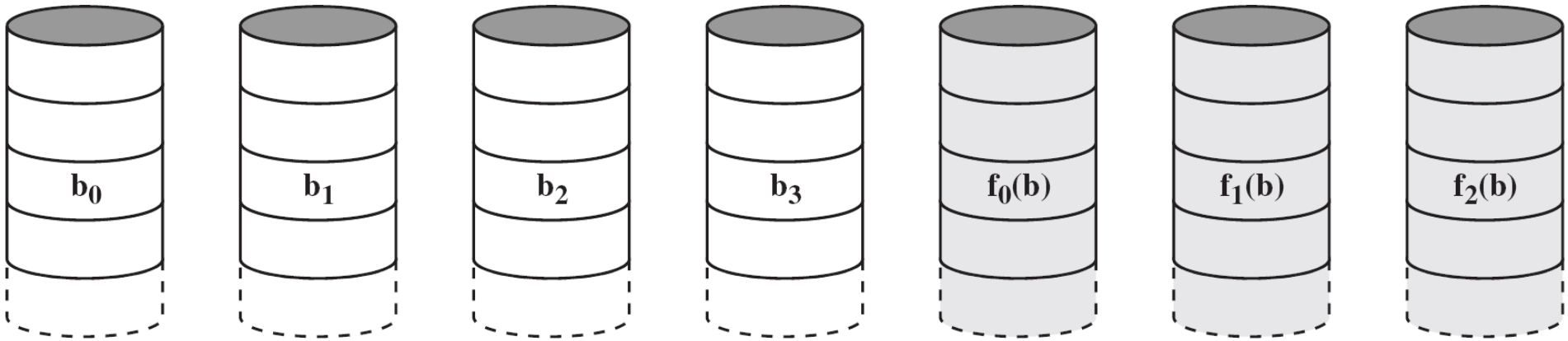
Level	Advantages	Disadvantages	Applications
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability





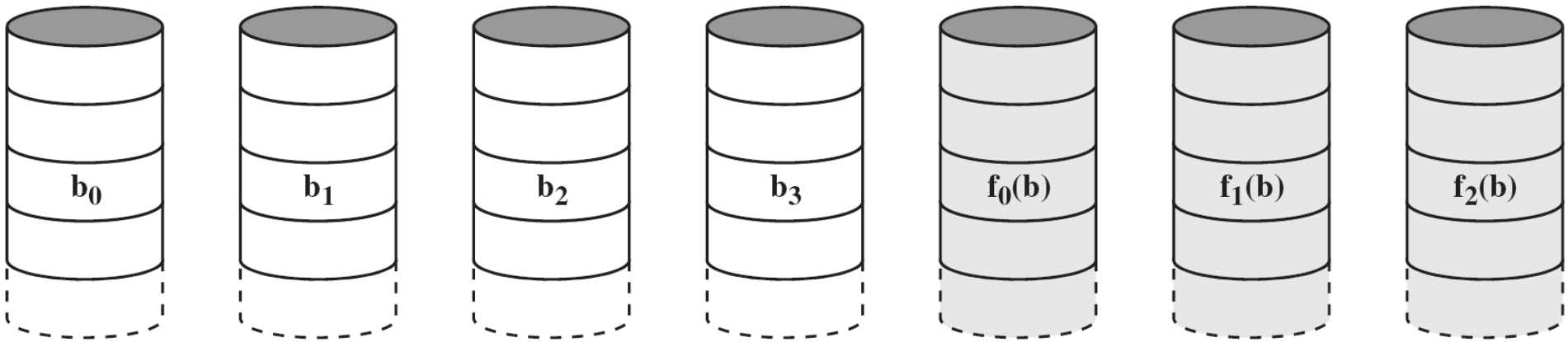
RAID 2 - Bit-Level Stripping with Hamming ECC

- Parallel access: all disks participate in every I/O request.
- Disks are synchronized: all heads in same position at any given time.
- Bit-level stripping: Very small strips ➔ single-bit strips!
- Error correction calculated across corresponding bits on disks.
- Multiple parity disks store Hamming code in corresponding bit positions.
- Number of redundant disks is proportional to **log** number of data disks.
- Read: data & parity delivered to controller ➔ error corrected instantly.
- Write: all data and parity disks accessed.
- Contemporary disks are highly reliable ➔ RAID 2 **not used** in practice!!
- I/O transfer rate: very high due to small strip size.
- I/O request rate: only one at a time ➔ ≈ single disk!



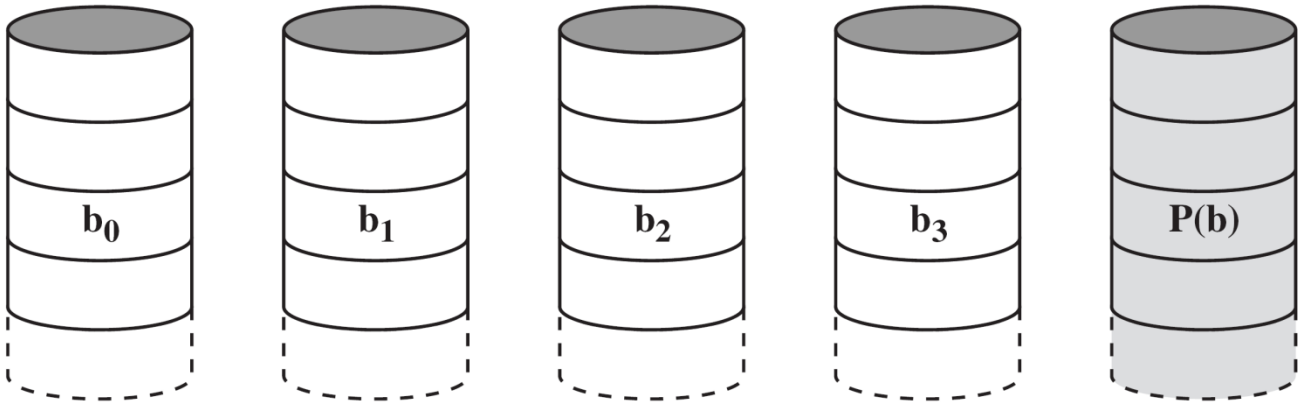
RAID 2 - Pros and Cons

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



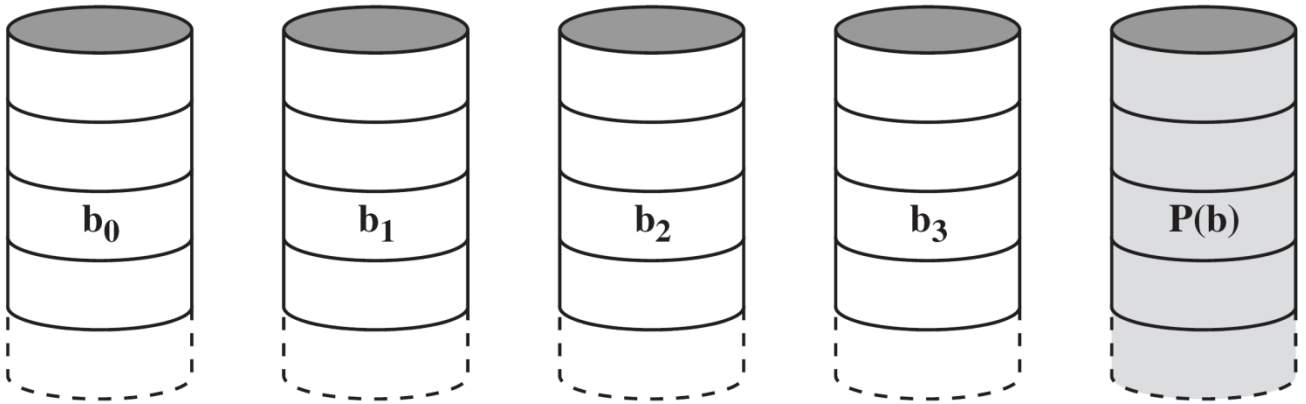
RAID 3 - Byte-Level Stripping with Parity

- Similar to RAID 2: synchronized disks, small strips. Byte-level stripping ➔ single-byte strips.
- Only 1 redundant disk, no matter how large the array.
- Simple parity bit for each set of corresponding bits.
- Drive fails ➔ replace it and reconstruct data from surviving disks and parity info. For instance, in a 5-disk array:
  - Disk #1 fails ➔ $X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$
- I/O transfer rate: very high due to small strip size.
- I/O request rate: only one at a time ➔ ≈ single disk.



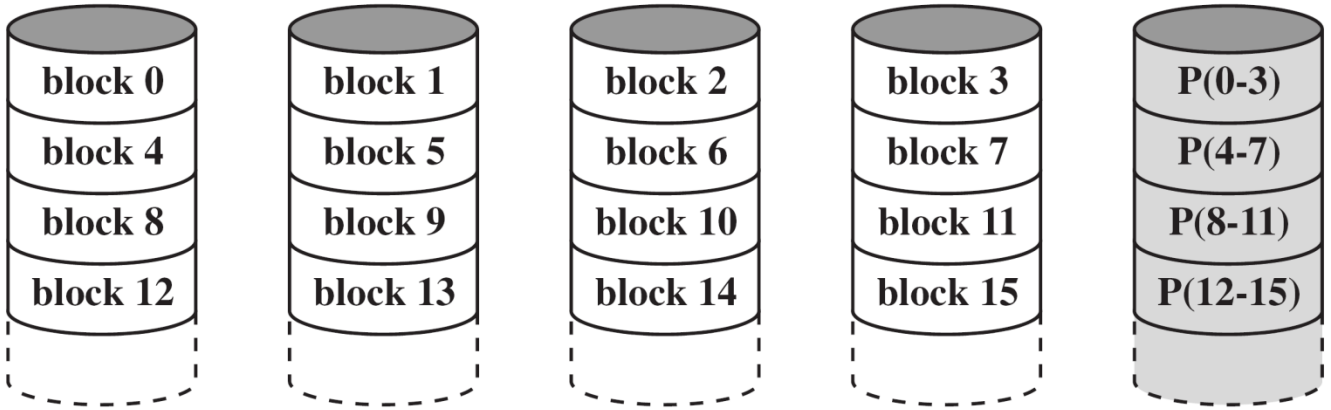
RAID 3 - Pros and Cons

Level	Advantages	Disadvantages	Applications
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



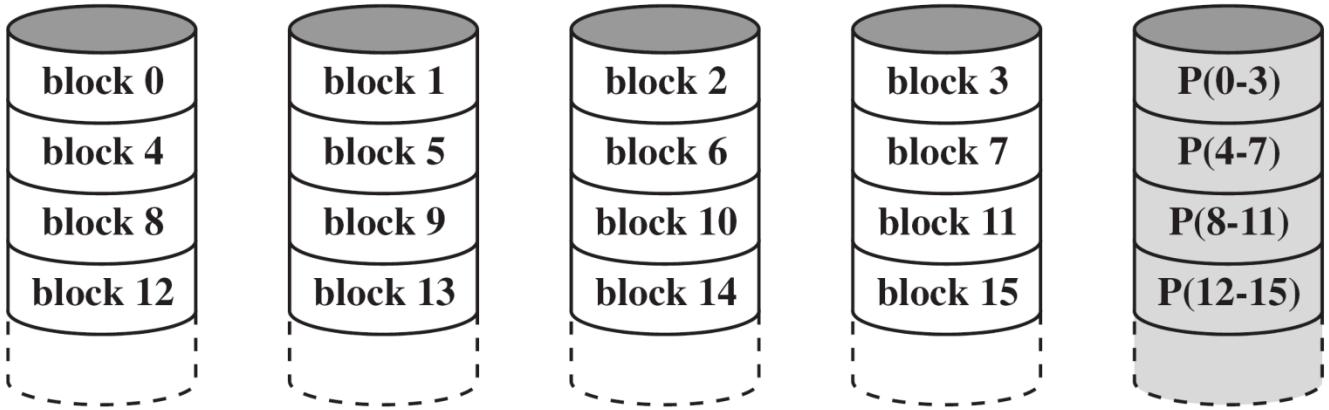
RAID 4 - Block-Level Stripping with Parity

- Independent access: each disk operates independently (not synchronized) ➔ separate I/O requests in parallel.
- Relatively large strips (block-level).
- Bit-by-bit parity calculated across strips on each disk.
- Parity stored on parity disk.
- Write: read old data and old parity, update both.
  - Write to disk #1 ➔ X4′(i) = X4(i) ⊕ X1(i) ⊕ X1′(i)
  - Not the case in RAID 2 and RAID 3 due to parallel access.
- I/O transfer rate: read: ≈ RAID 0, write: < single disk.
- I/O request rate: read: ≈ RAID 0, write: < single disk.



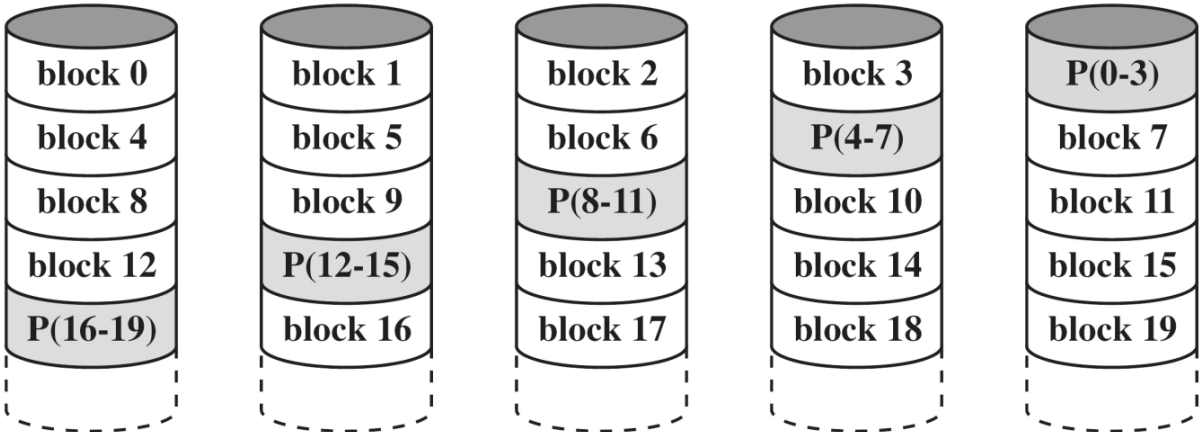
RAID 4 - Pros and Cons

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



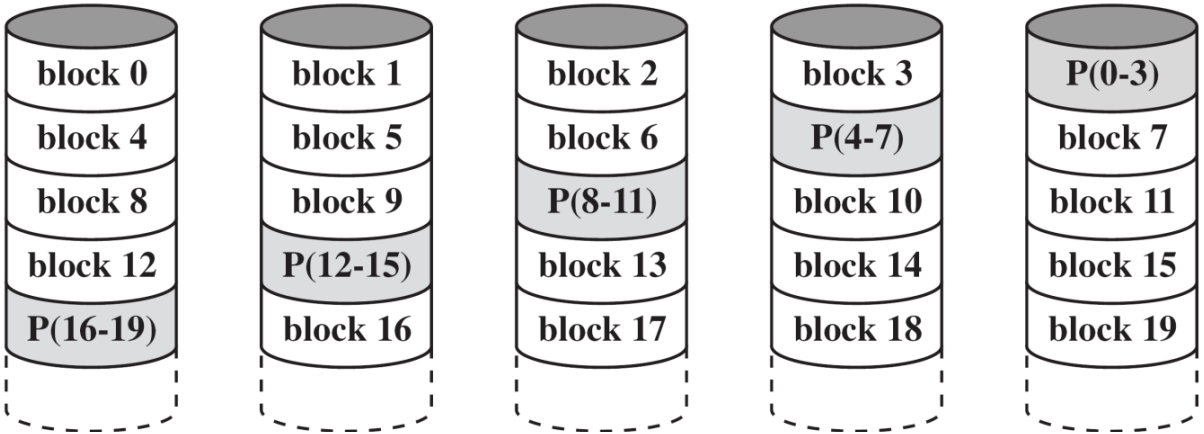
RAID 5 - Block-Level Stripping with Distributed Parity

- Independent access, relatively large strips (block-level).
- Like RAID 4, except parity distributed across all disks.
- Round robin allocation for parity strips.
  - —n-disk array: parity strip is on a different disk for the first n stripes, and the pattern repeats.
- Avoids RAID 4 bottleneck at parity disk.
- Commonly used in network servers.
- N.B. Does not mean 5 disks!
- I/O transfer rate: read: ≈ RAID 0, write: < single disk.
- I/O request rate: read: ≈ RAID 0, write: < single disk.



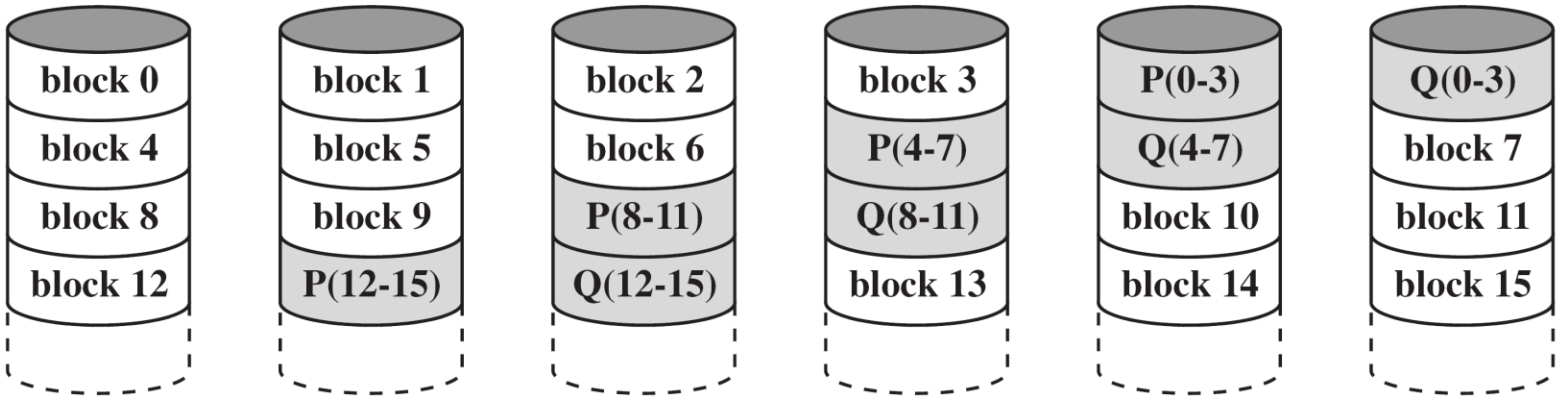
RAID 5 - Pros and Cons

Level	Advantages	Disadvantages	Applications
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



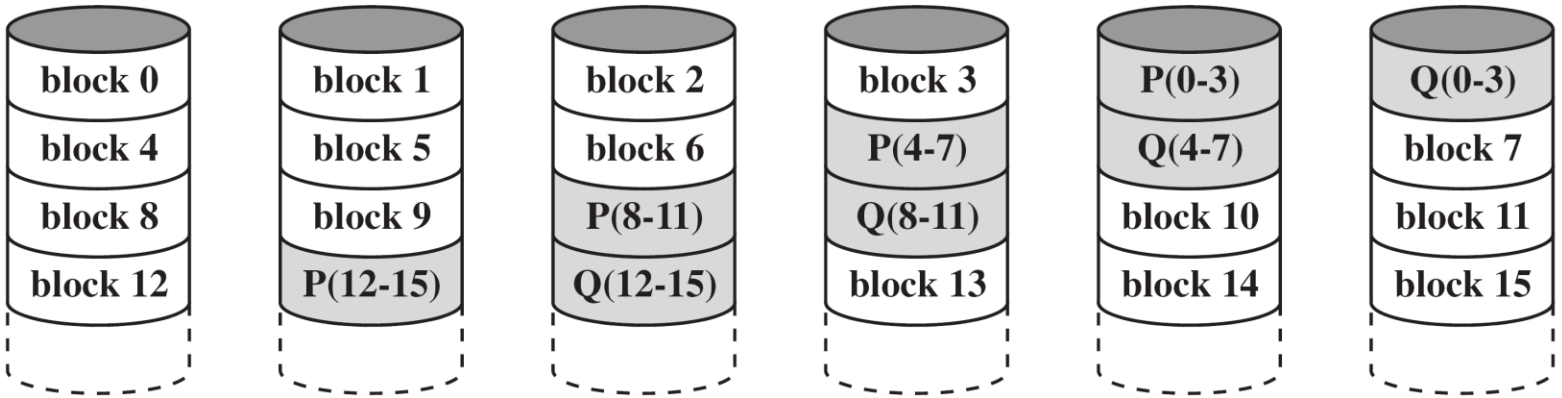
RAID 6 - Block-level striping with double distributed parity

- Independent access, relatively large strips (block-level).
- Two parity calculations.
  - —P and Q are two different data check algorithms.
- Stored in separate blocks on different disks.
- N data disks ➔ N+2 disks required to build the array.
- High data availability
  - —Three disks need to fail for data loss.
  - —Significant write penalty.
- I/O transfer rate: read: ≈ RAID 0, write: < RAID 5.
- I/O request rate: read: ≈ RAID 0, write: < RAID 5.



RAID 6 - Pros and Cons

Level	Advantages	Disadvantages	Applications
6	Provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications



RAID Levels - Summary

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	**Similar to 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	**Similar to 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

*** N = number of data disks. N must be greater than 1 in all RAID configurations except RAID 1 where N could be equal to 1.**

*** m = number of ECC disks. N is proportional to log m.**

Reading Material

- Stallings, Chapter 6:
  - Pages 195 – 205