COA

Chapter 06: | RAID

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

AGENDA

- What is RAID?
- Why RAID?
- Different Types of RAID

Intro…

- Disk drives have continued to get **smaller and cheaper**
- Economically feasible to **attach many disks** to a computer system
  - Reliability
  - Read/Write rate

RAID

- **R**edundant **A**rray of **I**ndependent **D**isks (**RAID**s)
- **Coined** in a paper - University of California at **Berkeley**
- To address **performance and reliability** issues
- RAID
  - **Past**: **I**nexpensive
  - **Now**: **I**ndependent

RAID

- **With multiple disks**, separate I/O requests can be handled in parallel, as long as the data required reside on separate disks
- Different design architectures that share **three common characteristics**
  - **Single logical drive**
  - **Striping**
  - **Parity information –** data **recoverability**

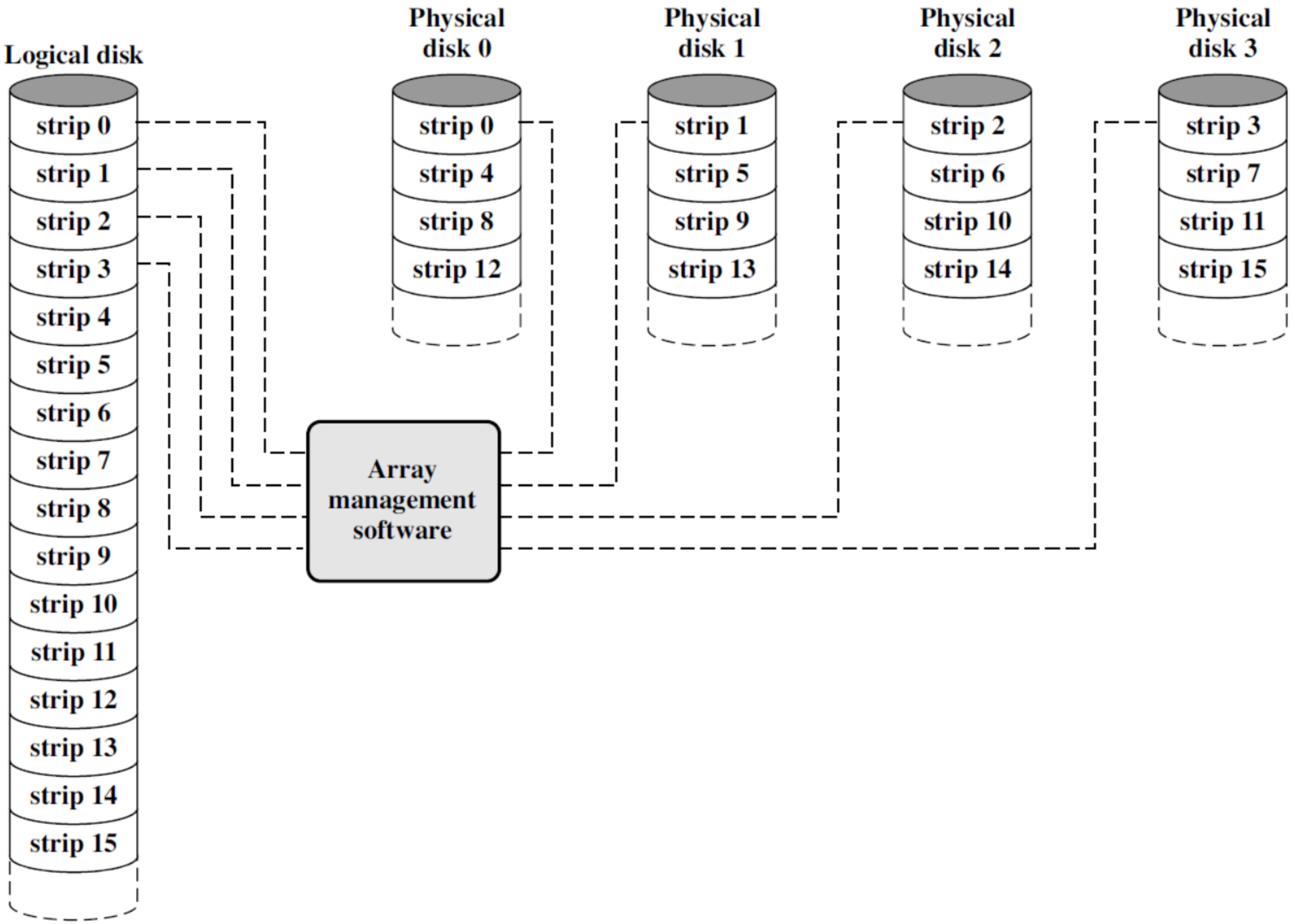
RAID Levels

- **Mirroring** provides high **reliability**, but it is **expensive**
- **Striping** provides **high data-transfer rates**, but it does **not** improve **reliability**
- **Numerous schemes** to provide redundancy at lower cost by using disk striping combined with parity bits have been proposed
- These schemes have different **cost-performance trade-offs** and are classified according to **levels called RAID Levels**

RAID Level 0

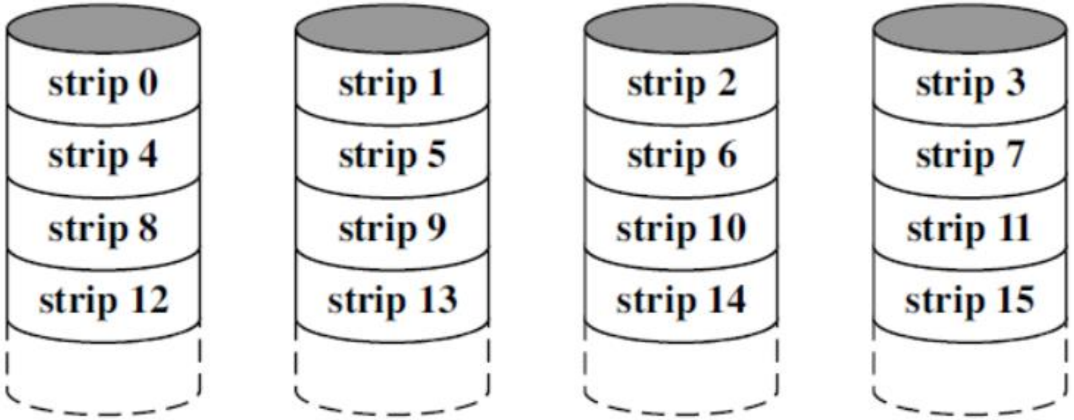
- No redundancy – Not a true member of the RAID Family
- Data **striped** across all disks
- Round Robin striping
- Increase speed
  - Multiple data requests probably not on same disk
  - Disks seek in parallel
  - A set of data is likely to be striped across multiple disks

RAID Level 0


Logical disk
strip 0
strip 1
strip 2
strip 3
strip 4
strip 5
strip 6
strip 7
strip 8
strip 9
strip 10
strip 11
strip 12
strip 13
strip 14
strip 15
Physical disk 0
strip 0
strip 4
strip 8
strip 12
Physical disk 1
strip 1
strip 5
strip 9
strip 13
Physical disk 2
strip 2
strip 6
strip 10
strip 14
Physical disk 3
strip 3
strip 7
strip 11
strip 15
Array management software

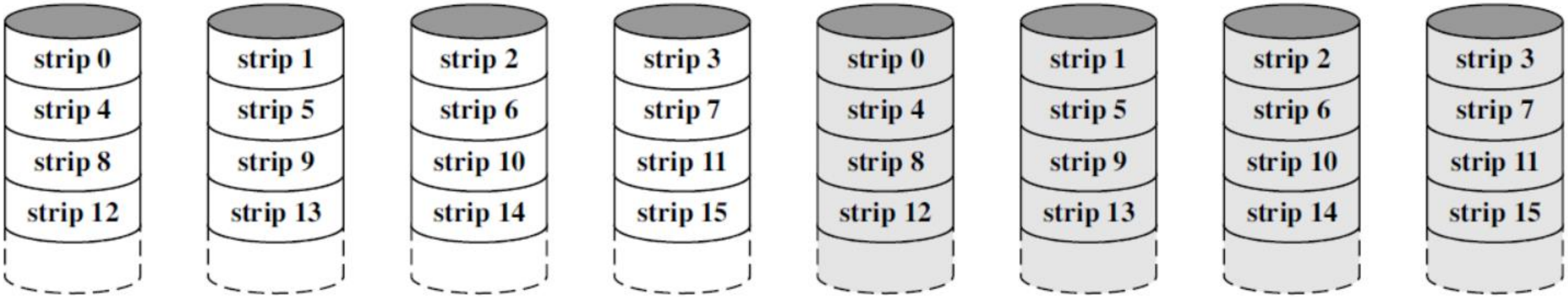
RAID Level 0

- The data are striped across the available disks
- All of the user and system data are viewed as being stored on a logical disk
- The logical disk is divided into **strips**; these strips may be physical **blocks**, **sectors**, or some **other unit**



RAID Level 1

- **Mirrored** Disks – **Duplicating** all the data
- Data is **striped** across disks
- 2 copies of each stripe on separate disks

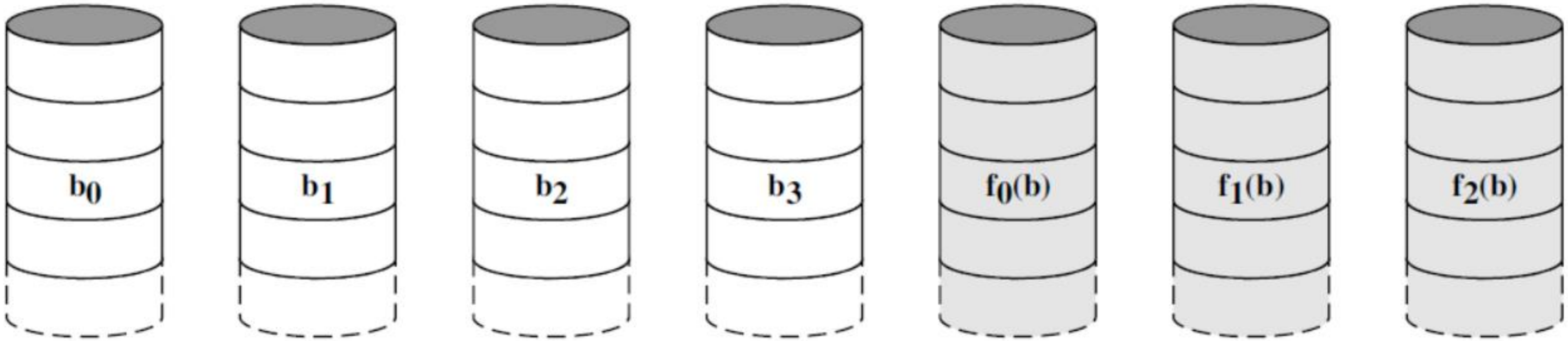


RAID Level 1

- Read from either
  - One with the minimum seek time plus rotational latency
- Write to both
  - In Parallel
  - dictated by the slower of the two writes
- Recovery is simple
  - Swap faulty disk & re-mirror
  - No down time
- Expensive

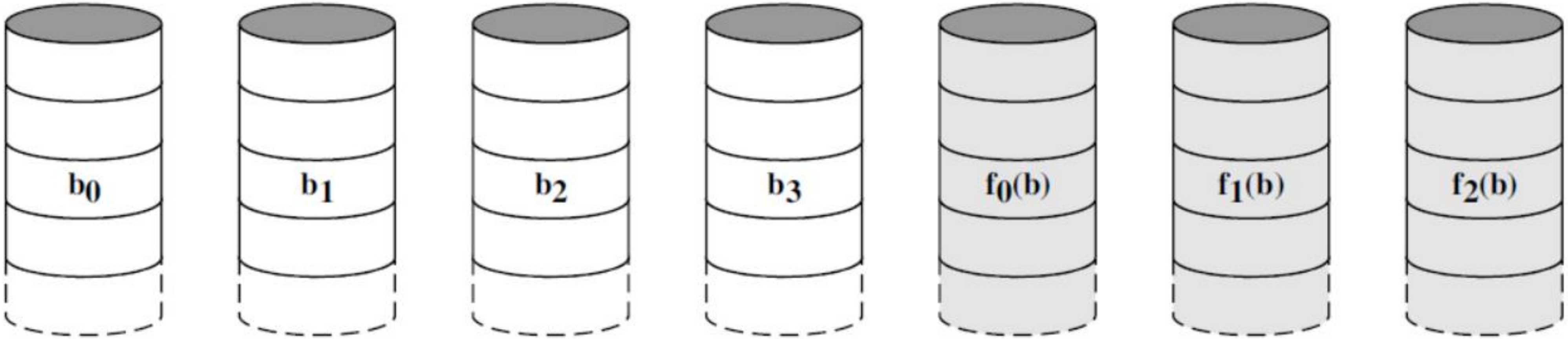
RAID Level 2

- Disks are synchronized
  - **Parallel** access array
  - all **member disks participate** in the execution of every I/O request
- Very **small** stripes
- **Error correction** calculated across corresponding bits on disks



RAID Level 2

- Multiple parity disks store **Hamming code** error correction in corresponding positions
  - correct single-bit errors and detect double-bit errors
- Lots of redundancy
- Expensive
  - # **redundant** disks is proportional to the **log** of the number of data disks

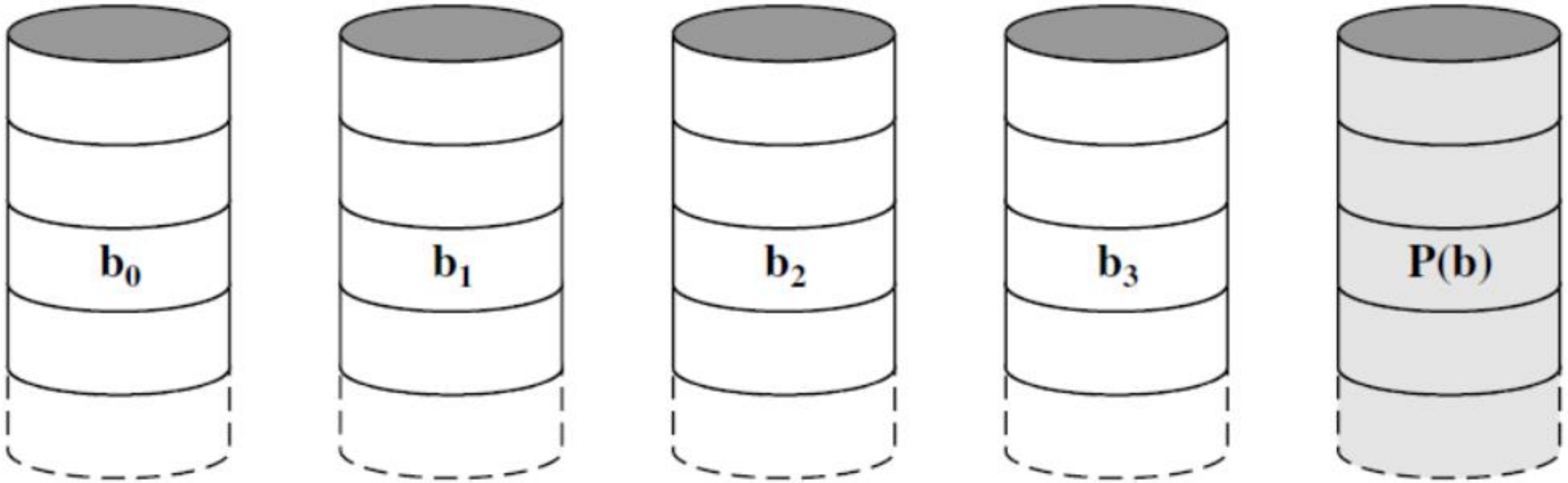


RAID Level 2

- **Read**
  - all disks are **simultaneously accessed**. The **requested data** and the **associated error-correcting code** are delivered to the array controller.
  - If there is a **single-bit error**, the controller can **recognize** and **correct** the error instantly, so that the read access time is not slowed.
- **Write**
  - all **data disks and parity disks** must be accessed for the write operation

RAID Level 3

- Similar to RAID 2
- Only **one redundant disk**, no matter how large the array
- Simple **parity bit** for each set of corresponding bits
- Data on failed drive can be reconstructed from surviving data and parity info
- Very high transfer rates



(d) RAID 3 (Bit-interleaved parity)

RAID Level 3

- Data reconstruction is simple
- X0 through X3 contain data and X4 is the parity disk

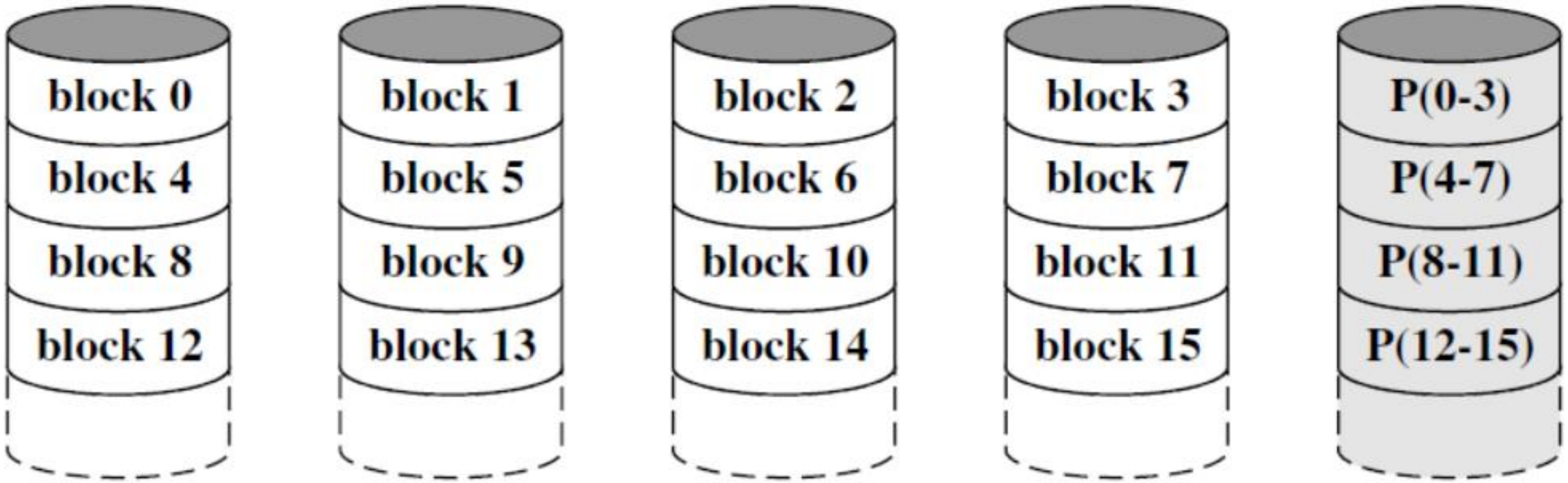
$$X4(i) \ = \ X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

- Suppose that drive X1 has failed

$$X1(i) \ = \ X4(i) \oplus X3(i) \oplus X2(i) \oplus X1(i)$$

RAID Level 4

- Each disk operates **independently**
  - Separate I/O requests can be satisfied in parallel
- Good for high I/O request rate



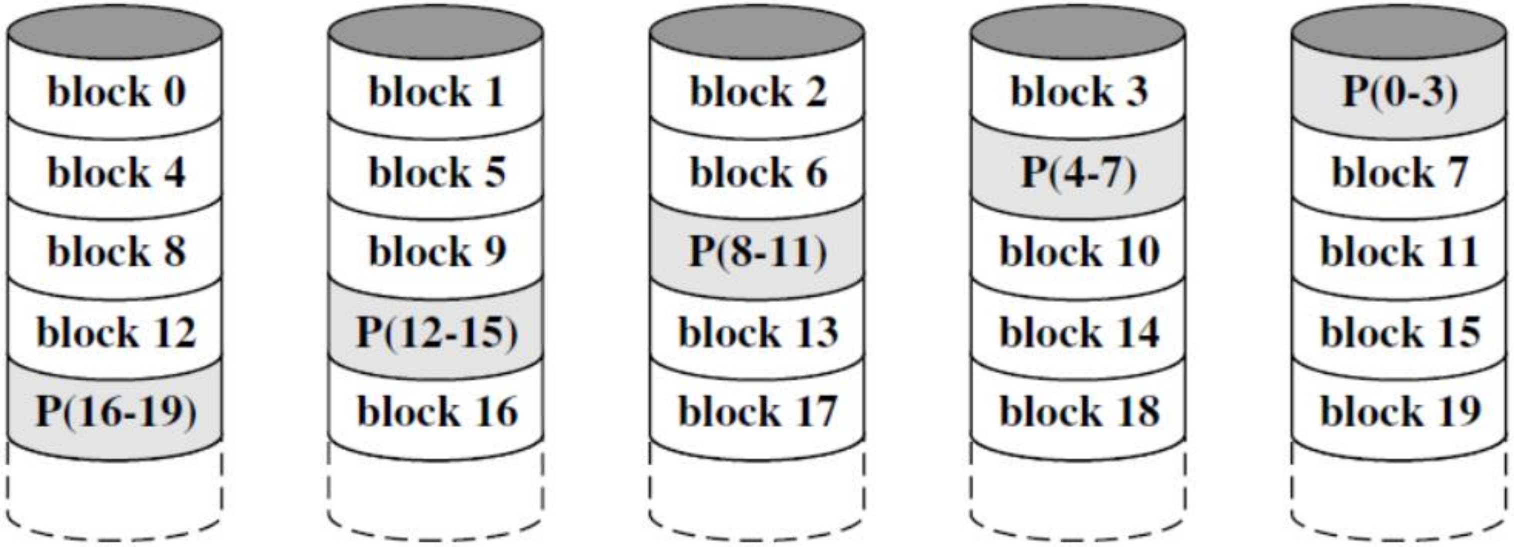
(e) RAID 4 (Block-level parity)

RAID Level 4

- Large stripes
- Bit by bit parity calculated across stripes on each disk
- Parity stored on parity disk
- Consider an **array of five drives in** which X0 through X3 contain data and **X4 is the parity disk**. Suppose that a write is performed that only involves a strip on disk X1
- $X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$
- $X\acute{4}\,(i) = X3(i) \oplus X2(i) \oplus X\acute{1}(i) \oplus X0(i)$
- $X\acute{4}\,(i) = X4(i) \oplus X1(i) \oplus X\acute{1} \oplus$

RAID Level 5

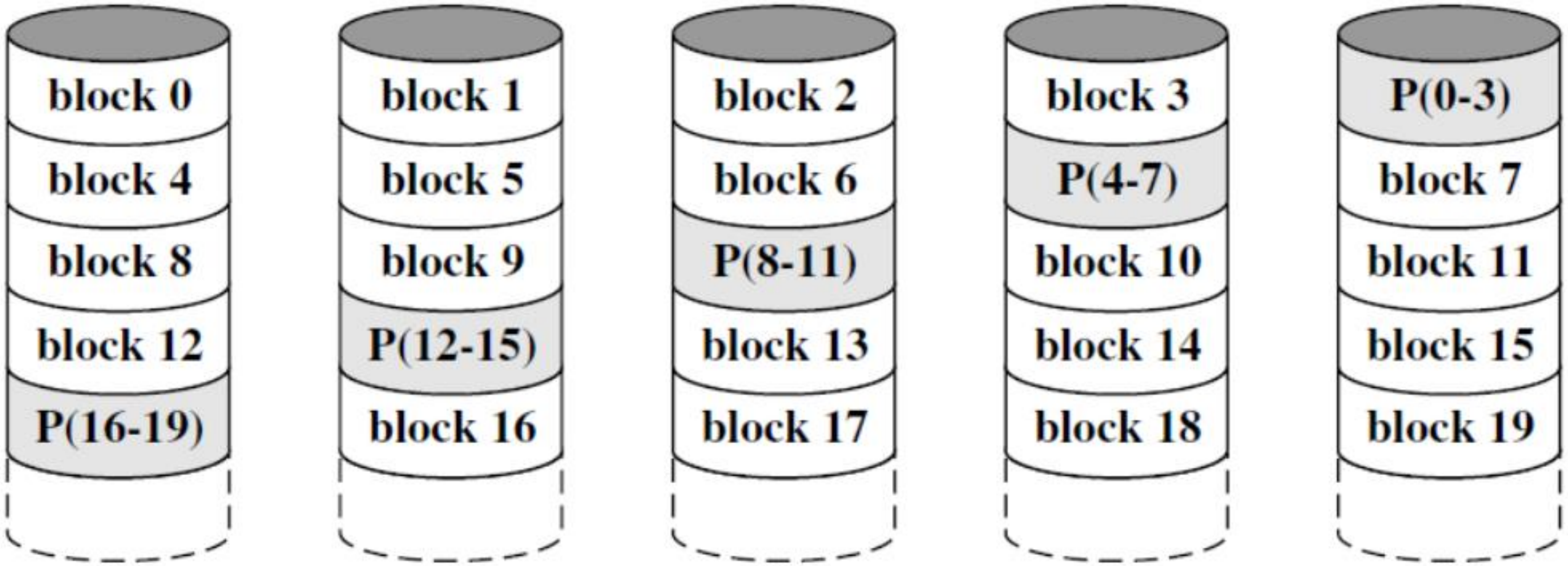
- Like RAID 4
- Parity striped across all disks
- Round robin allocation for parity stripe



(f) RAID 5 (Block-level distributed parity)

RAID Level 5

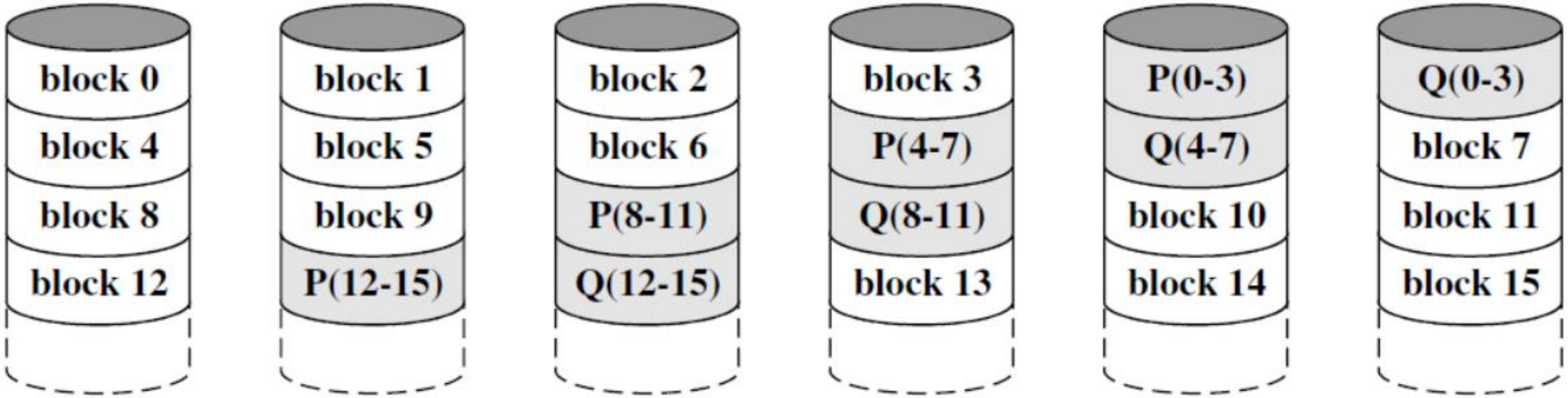
- Avoids RAID 4 bottleneck at parity disk
- Commonly used in network servers



(f) RAID 5 (Block-level distributed parity)

RAID Level 6

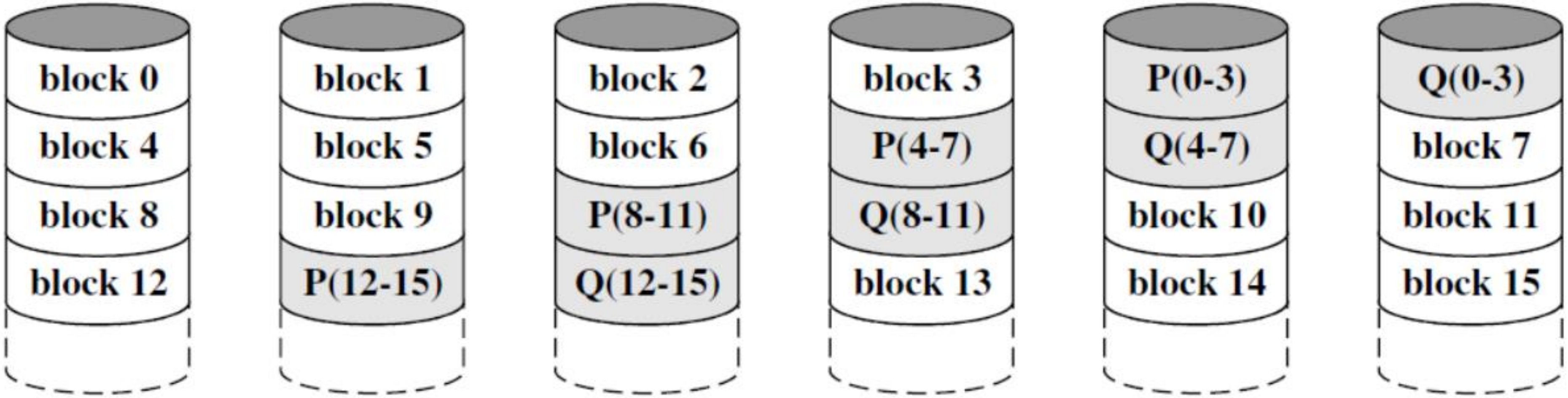
- **Two parity** calculations
  - One of the two is the exclusive-OR calculation.
  - the other is an independent data check algorithm.
- Stored in separate blocks on different disks
- User requirement of N disks needs **N+2**



(g) RAID 6 (Dual redundancy)

RAID Level 6

- High data availability
- Three disks need to fail for data loss
- Significant write penalty



(g) RAID 6 (Dual redundancy)