SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 23CST203 - DATABASE MANAGEMENT SYSTEMS

II YEAR / IV SEMESTER

Unit 2- Physical Storage and MongoDB
Topic 2 : RAID

The Motivation for RAID

- Computing speeds double every 3 years

- Disk speeds can't keep up

- Data needs higher MTBF than any component in system

-       IO Performance and Availability Issues!

RAID to the Rescue!

- PERFORMANCE
  - ◆ Parallelism
  - ◆ Load Balancing
- AVAILABILITY
  - ◆ Redundancy: Mirroring, or Striping with Parity
- FLEXIBILITY
  - ◆ Selectable Performance/Availability/Cost

What is RAID Technology?

- **R**edundant **A**rray of *Independent* **D**isks (a.k.a. "Disk Array")
  - ◆ *Multiple* drives, *single* host disk unit
- Provides opportunity to increase:
  - ◆ Performance via Parallelism
  - ◆ Data Availability via Redundancy
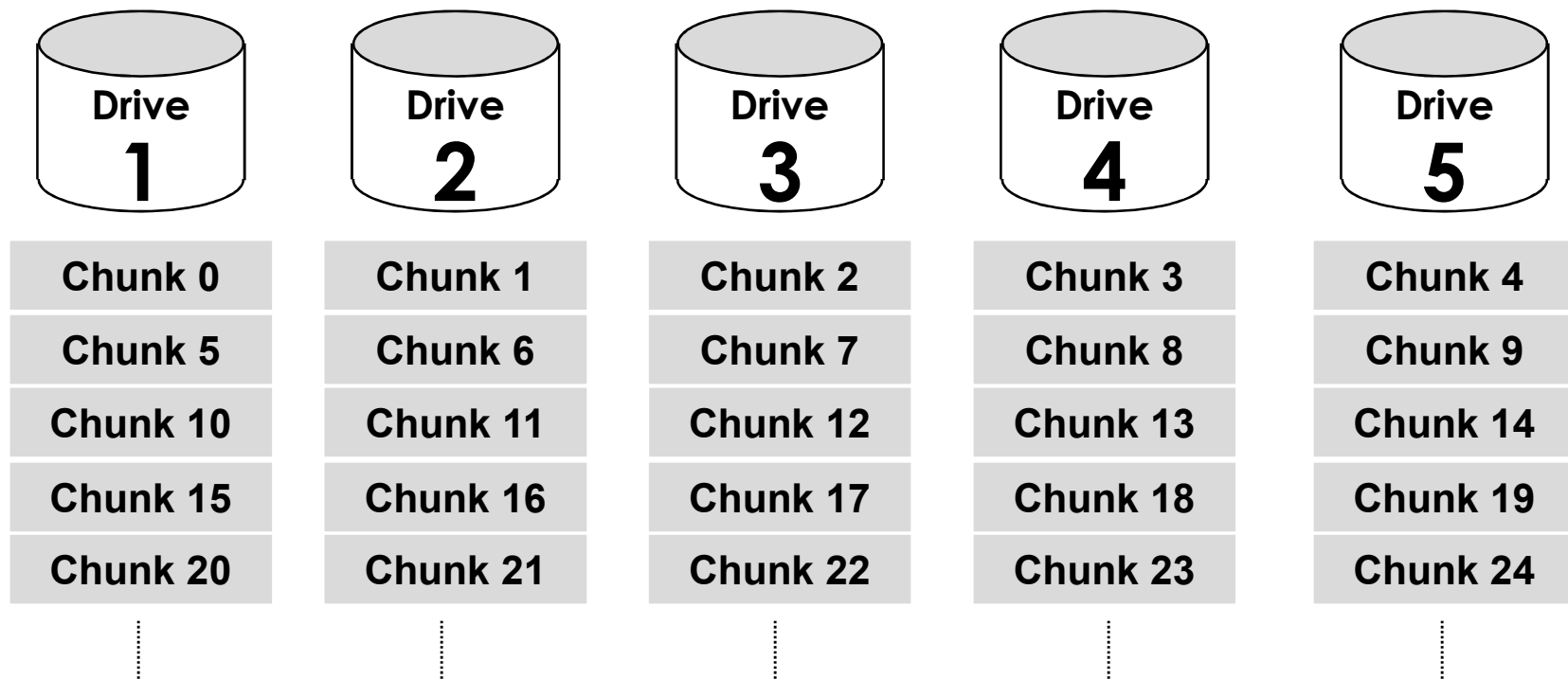- Cheap cost via commodity disks

Performance: Disk Striping

■ Chunk size tuneable for BW vs Thruput tradeoffs

◆ Large Chunk ⟹ High **Throughput** (IO/sec)

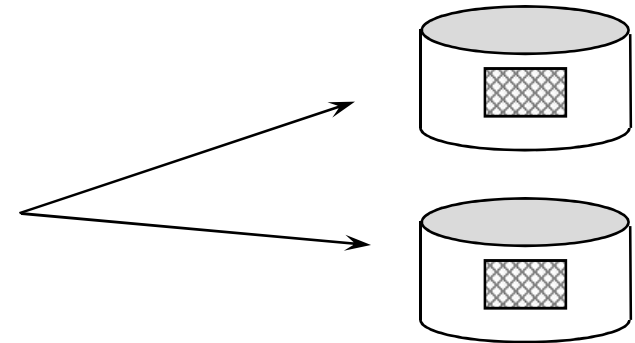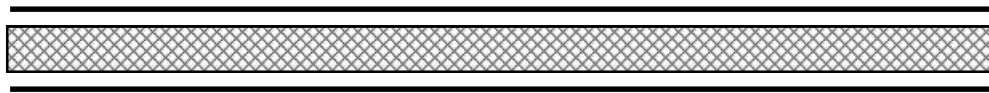◆ Small Chunk ⟹ High **Bandwidth** (MB/sec)

Drive 1	Drive 2	Drive 3	Drive 4	Drive 5
Chunk 0	Chunk 1	Chunk 2	Chunk 3	Chunk 4
Chunk 5	Chunk 6	Chunk 7	Chunk 8	Chunk 9
Chunk 10	Chunk 11	Chunk 12	Chunk 13	Chunk 14
Chunk 15	Chunk 16	Chunk 17	Chunk 18	Chunk 19
Chunk 20	Chunk 21	Chunk 22	Chunk 23	Chunk 24

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

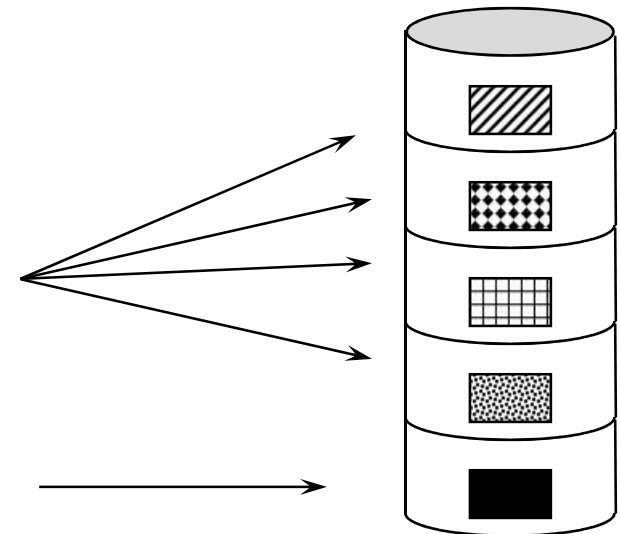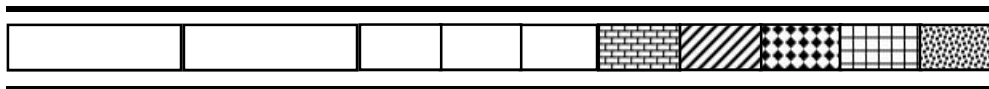4/13/2025

Availability: Redundancy

Mirroring

Same data written to both disks.

Striping with Parity

Data Stream, OR IO Stream, can be multiplexed across multiple disks, depending on BW vs Thruput. Parity data is also stored on disk.
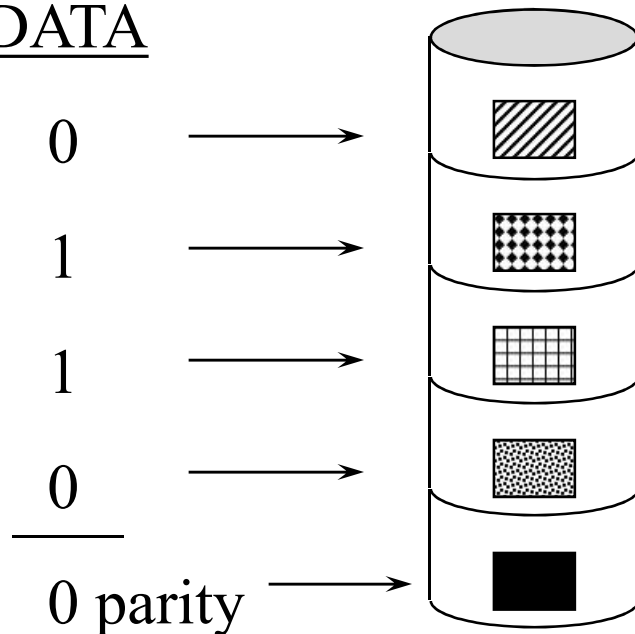
> Add XOR'd parity for increased availability.

Parity Redundancy

- Parity = XOR of data from every disk in the RAID unit

DATA

0   ⟶

1   ⟶

1   ⟶

0   ⟶

____

0 parity ⟶

- Any single disk's data can be recovered by XOR'ing the data of the surviving disks.

RAID Levels

- **Many to choose from**

- **Each offers unique tradeoffs**
  - ◆ Performance
  - ◆ Availability
  - ◆ Costs

- **We offer levels 0, 1, 3, 5, 10**

RAID 0

Disk Striping with No Redundancy

- High Performance; Low Availability
- Data Striped on Multiple Disks
- Multi-threaded Access

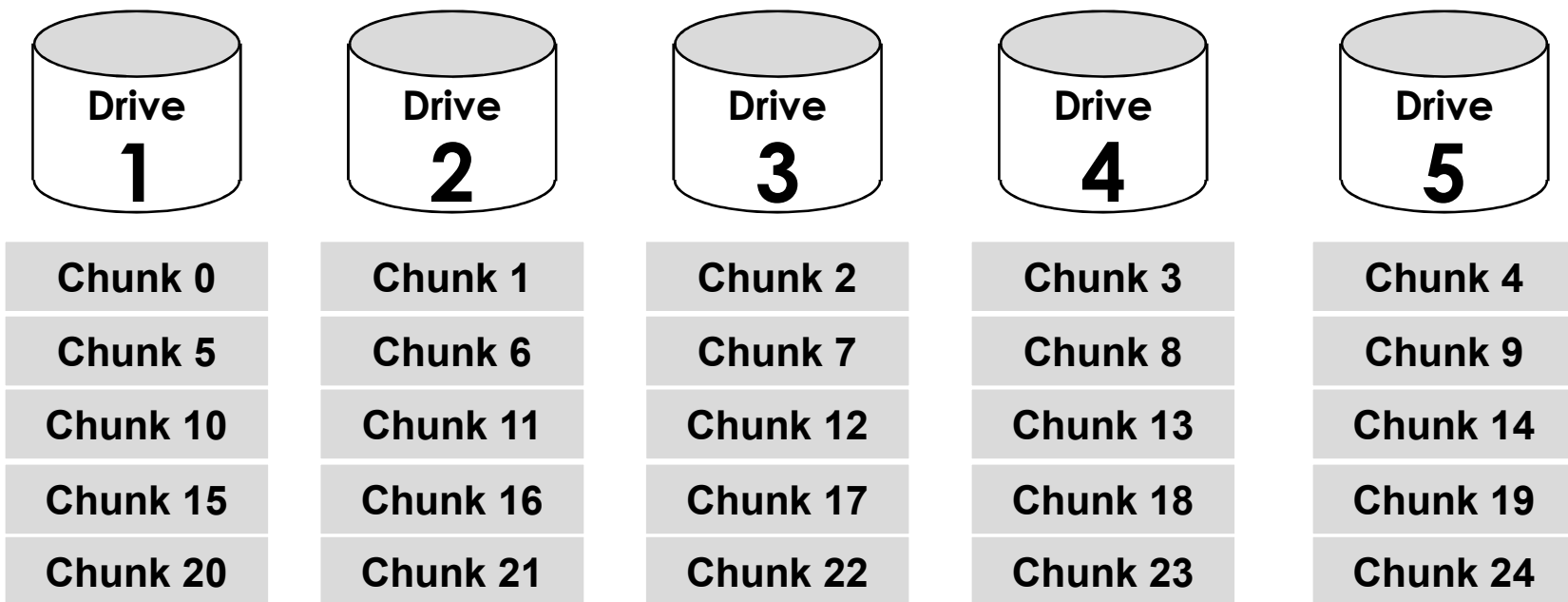_Data Stream, OR IO Stream, can be Striped across multiple disks (BW vs Thruput)._

23CST207 DBMS/ K.KARTHIKEYAN/AP-
CSE/SNSCE

RAID 0 Striping

- **Chunk size tuneable for BW or Thruput**
- **No redundancy**

Drive 1	Drive 2	Drive 3	Drive 4	Drive 5
Chunk 0	Chunk 1	Chunk 2	Chunk 3	Chunk 4
Chunk 5	Chunk 6	Chunk 7	Chunk 8	Chunk 9
Chunk 10	Chunk 11	Chunk 12	Chunk 13	Chunk 14
Chunk 15	Chunk 16	Chunk 17	Chunk 18	Chunk 19
Chunk 20	Chunk 21	Chunk 22	Chunk 23	Chunk 24

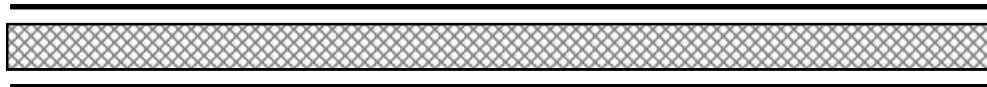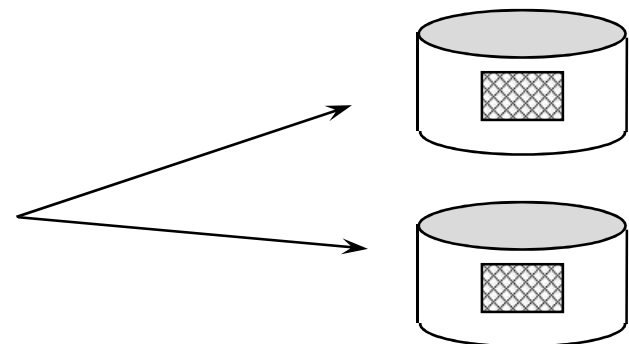23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

4/13/2025

RAID 1

Disk Mirroring

- Single-disk Performance; Expensive Availability

- Data 100% duplicated across both spindles.

- Single-threaded access

Single Host Unit

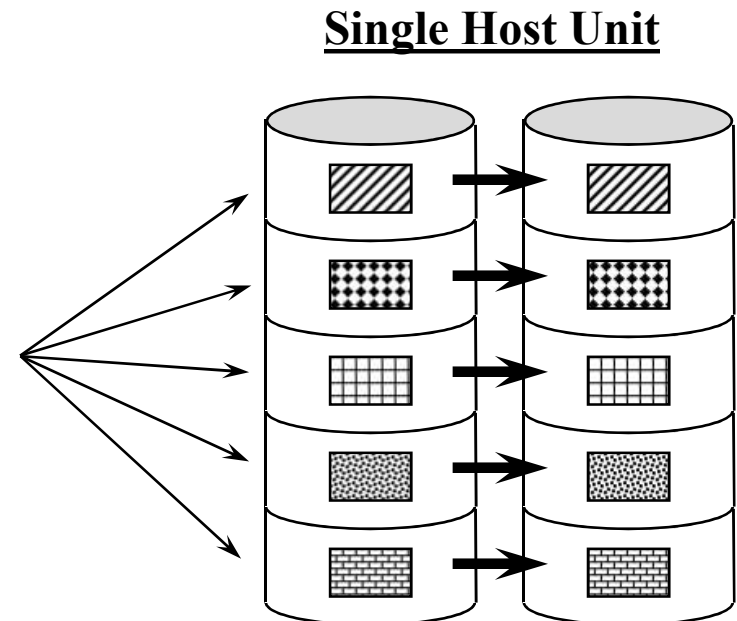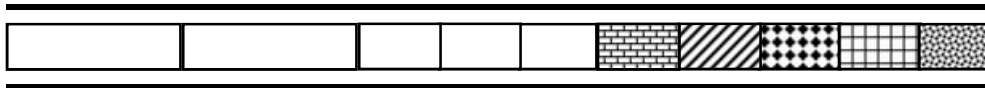*SAME data written to BOTH disks -- no segmenting.*

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

4/13/2025

RAID 1/0

Striped Mirrors

- Highest Performance; Most Expensive Availability
- Multi-threaded Access

Single Host Unit

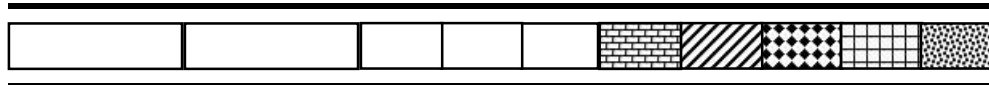*Data Stream, OR IO Stream, can be Striped across multiple disks (BW vs Thruput).*
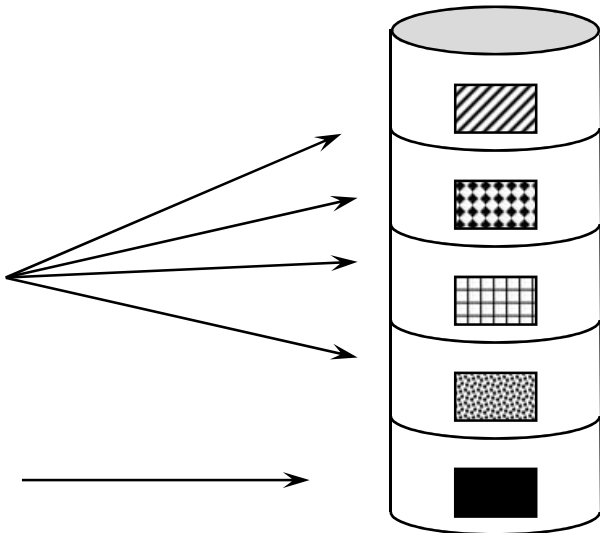
RAID 3

Disk Striping with dedicated parity drive

- High BW Performance; Cheap Availability
- Sector-granular data striping
- Single-threaded Access

Data Stream is Striped across N-1 disks for high bandwidth.
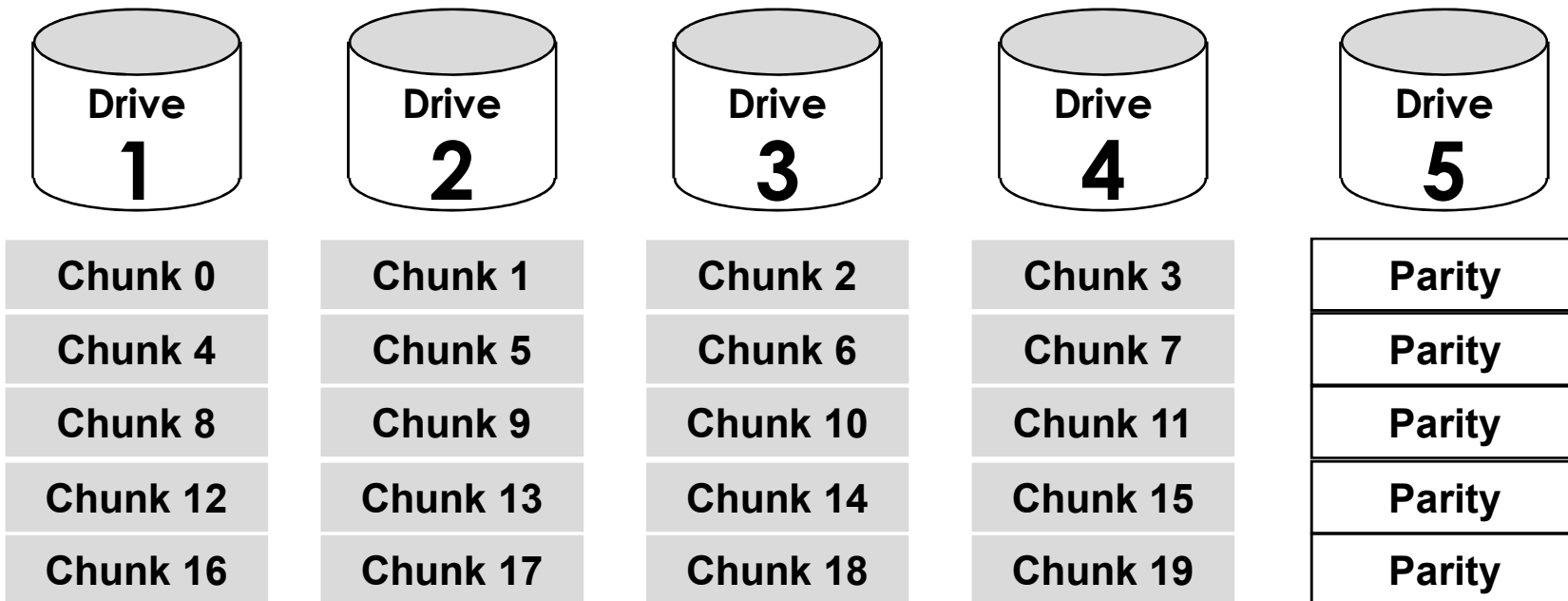
XOR Parity Data

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

4/13/2025

RAID 3 Striping

- **Chunk size = single sector (pure RAID 3 would be single _byte_)**
- **All parity data on same spindle**

| Drive 1	Drive 2	Drive 3	Drive 4	Drive 5
Chunk 0	Chunk 1	Chunk 2	Chunk 3	Parity
Chunk 4	Chunk 5	Chunk 6	Chunk 7	Parity
Chunk 8	Chunk 9	Chunk 10	Chunk 11	Parity
Chunk 12	Chunk 13	Chunk 14	Chunk 15	Parity
Chunk 16	Chunk 17	Chunk 18	Chunk 19	Parity

RAID 5

Disk Striping with rotating parity drive

- High Read Performance, expensive Write performance; Cheap Availability
- Tuneable Stripe granularity
- Optimized for multi-thread access

IO Stream is Striped across N-1 disks for high IOs per second (thruput).
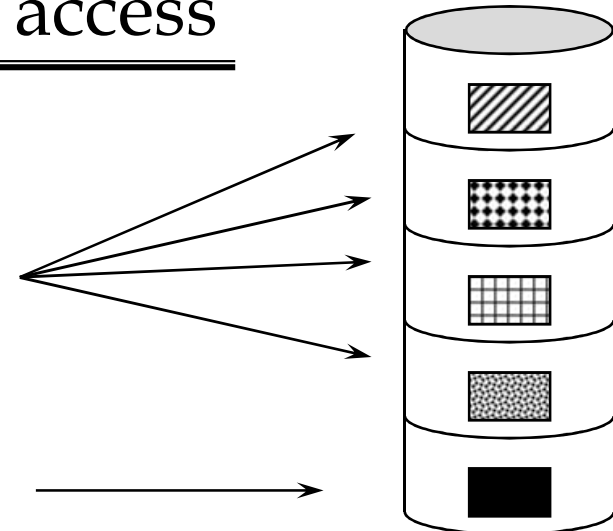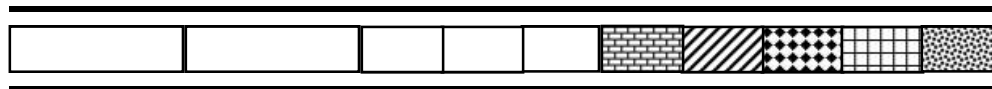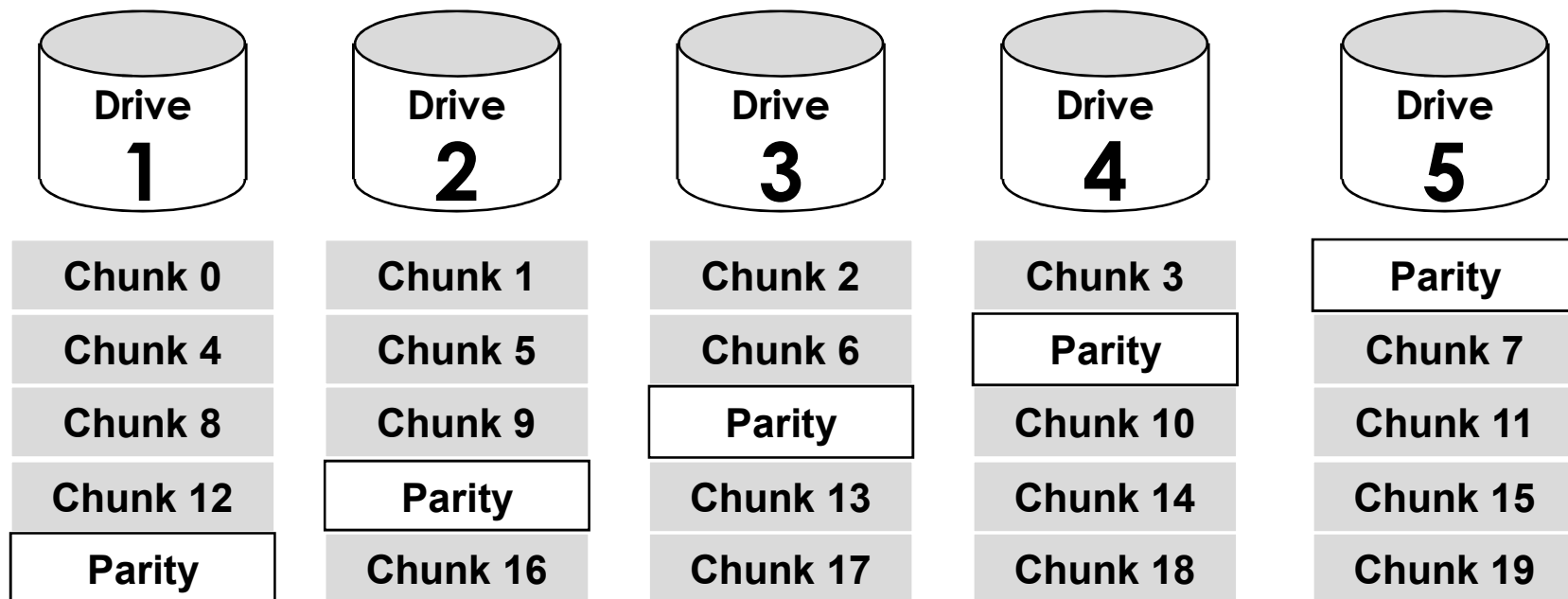
XOR Parity Data

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

RAID 5 Striping

- **Chunk size is tuned such that typical IO aligns on _single_ disk.**
- **Parity rotates amongst disks to avoid write bottleneck**

Drive 1	Drive 2	Drive 3	Drive 4	Drive 5
Chunk 0	Chunk 1	Chunk 2	Chunk 3	Parity
Chunk 4	Chunk 5	Chunk 6	Parity	Chunk 7
Chunk 8	Chunk 9	Parity	Chunk 10	Chunk 11
Chunk 12	Parity	Chunk 13	Chunk 14	Chunk 15
Parity	Chunk 16	Chunk 17	Chunk 18	Chunk 19

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

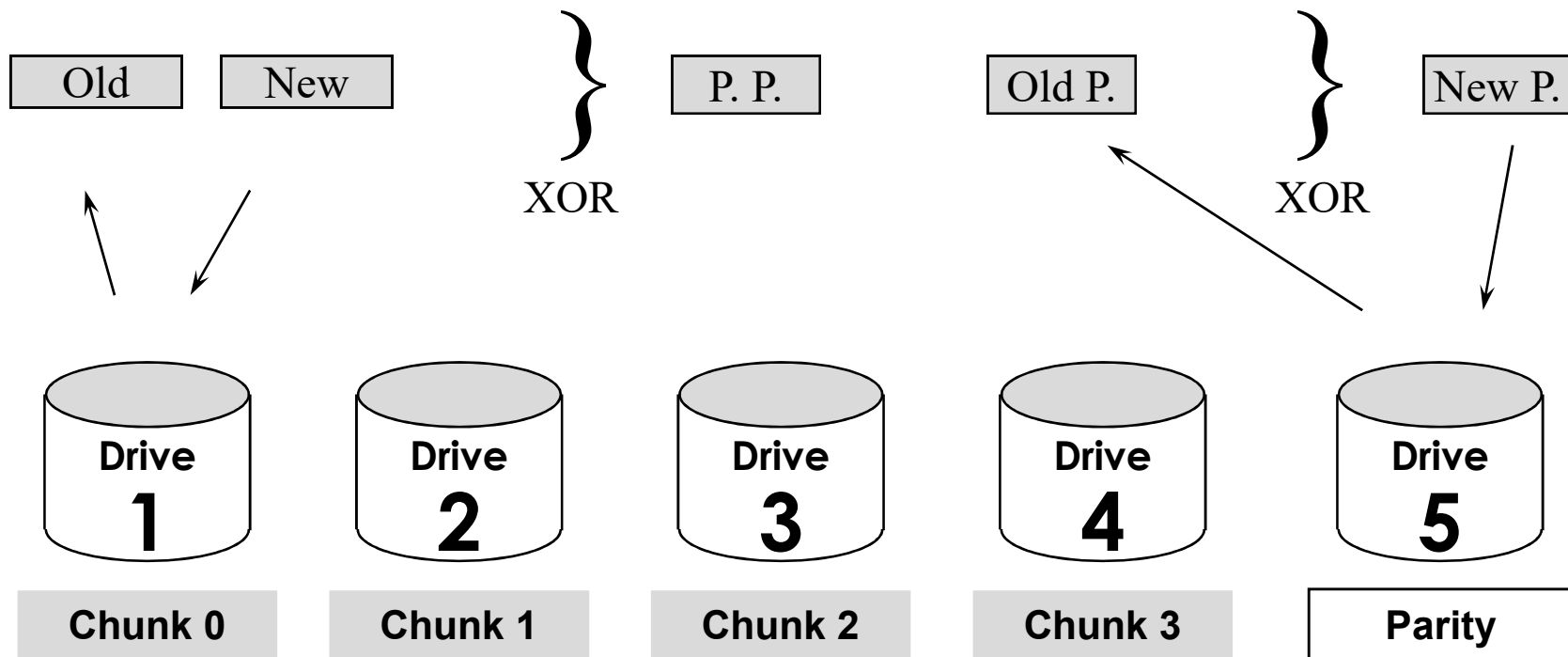4/13/2025

RAID 5 - Write Operation

1. Read old data.

2. Write new data

3. XOR old and new data to create "Partial Product".

4. Read old parity data.

5. Xor old parity with partial product, writing out result as new parity.

Old	New

} P. P.

Old P.

} New P.

XOR

XOR

Drive 1

Drive 2

Drive 3

Drive 4

Drive 5

Chunk 0

Chunk 1

Chunk 2

Chunk 3

Parity

23CST207 DBMS/ K.KARTHIKEYAN/AP-CSE/SNSCE

4/13/2025

RAID Level Review

◆ RAID 0 - Data striping, Non-redundant.

  ◆ _High Performance, Low Availability_

◆ RAID 1 - Mirroring

  ◆ _Moderate Performance, Expensive High Availability_

◆ RAID 1/0 - Striping and Mirroring

  ◆ _High Performance, Expensively High Availability_

◆ RAID 3 - Striping, single parity disk.

  ◆ _High **Bandwidth** Performance, Cheap Availability_

◆ RAID 5 - Striping, rotating parity disk.

  ◆ _High **Thruput** Performance, Cheap Availability_

23CST207 DBMS/ K.KARTHIKEYAN/AP-
CSE/SNSCE

Summary

- **Increasing performance gap between CPU and IO**

- **Data availability a priority**

- **RAID meets the IO challenge:**

  - Performance via parallelism

  - Data Availability via redundancy

  - Flexibility via multiple RAID levels, each offer unique performance/availability/cost tradeoffs

THANK YOU