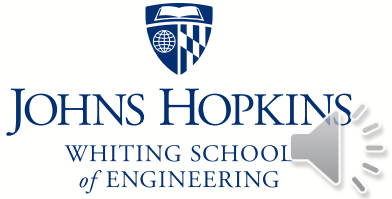
Johns Hopkins Engineering

Module 12: Memory & Storage

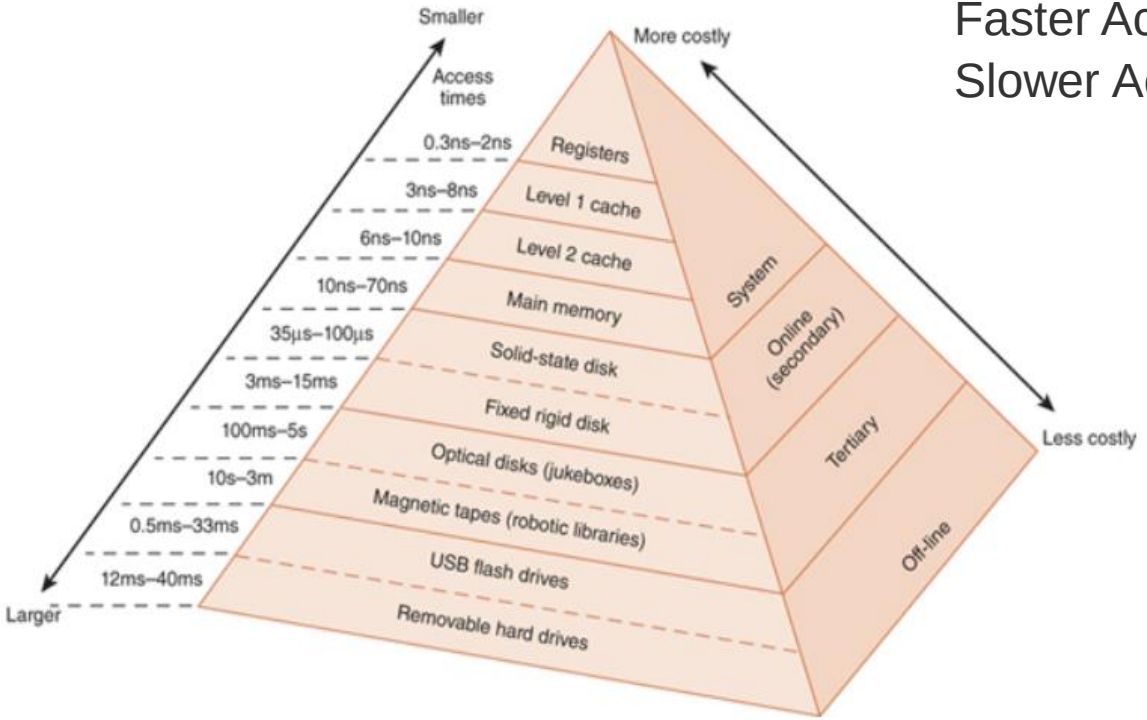
EN605.204: Computer Organization



Memory & Storage

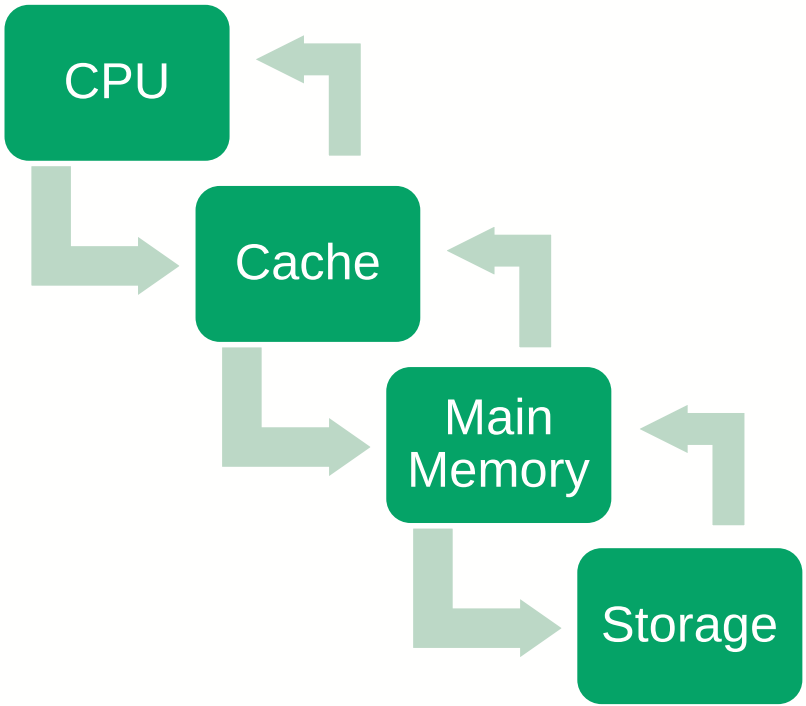
- Memory types
  - RAM
    - SRAM
    - DRAM
  - ROM
  - Cache
- Storage types
  - Fixed Disk
  - SSD
  - RAID

Hierarchy of Memory & Storage



Faster Access == Costly
Slower Access == Less Costly

Locality



- Lower levels respond to requests from higher levels for a piece of data
- Data block containing smaller requested data returned to higher levels
- Cached data block “probably” will be used in the future

Memory

- There are a number of memory technologies available for use
- At the core… basically two underlying types
  - RAM (random-access memory)
  - ROM (read-only memory)

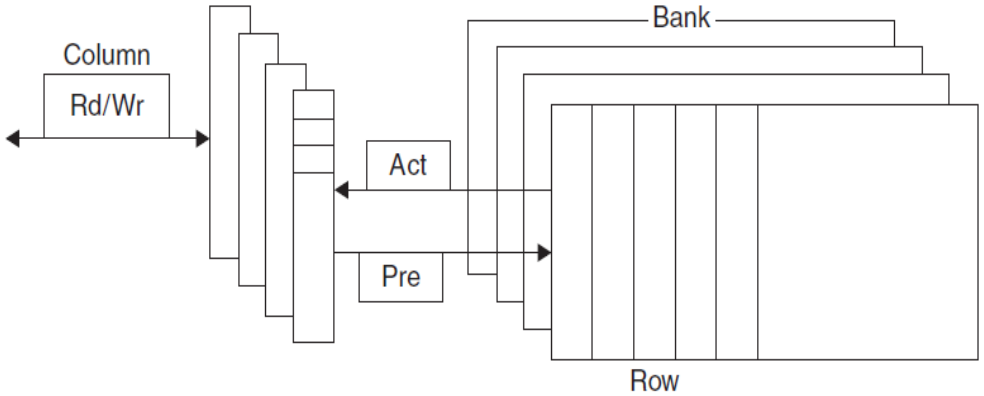
RAM (Random-Access Memory

- Commonly referred to as main memory
- Volatile! Needs power to retain data
- Two types of RAM
  - DRAM (Dynamic RAM)
  - SRAM (Static RAM)

DRAM (Dynamic Random-Access Memory)

- Constructed of tiny capacitors that leak electricity
- Must recharge every few milliseconds to maintain its data
- Memory cell refresh consists of reading its contents and writing it back
- Dynamic RAM is optimized to combine these cell reads and writes across an entire row of memory.
- Dynamic RAM offers higher storage capacity, less power use, and less heat than Static RAM

DRAM (Dynamic Random-Access Memory)



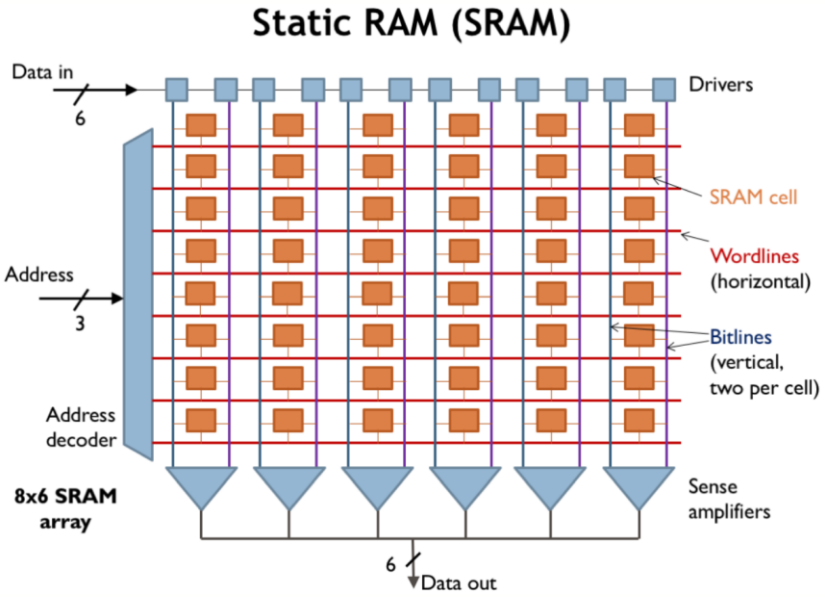
Internal organization of Dynamic RAM

- Organized in banks
- Banks consists of a series of rows
- Sending a Pre (precharge) command opens or closes a bank for access
- Row address is sent with an Act (activate), which causes the row to transfer to a buffer
- When the row is in the buffer, it can be transferred by successive column addresses or by specifying a block transfer and the starting column address
- Rows are buffered for repeated access

SRAM (Static Random-Access Memory)

- Requires constant recharge
- Can hold its contents providing power is available
- Much faster and more expensive than Dynamic RAM

SRAM (Static Random-Access Memory)



Internal organization of Static RAM

- Data flow…One port in / One port out
- One row for each memory location and one column for each bit in a location.
- Drivers charge with 0 or 1 down vertical bitlines for read and write operations
- Address decoder activates a wordline
- Data is pulled down through the matrix from the wordline to data out port
- Reads and writes are essentially analog operations performed by way of the vertical bitlines

ROM (Read-Only Memory)

- Most computers will contain a small amount of ROM
- ROM stores critical information necessary to operate the system
- ROM is not volatile and always retains its data

Types of ROM (Read-Only Memory)

- Basic types of ROM include ROM, PROM, EPROM, EEPROM
  - PROM (Programmable Read-Only Memory
    - Require special equipment to program
    - Data and instructions cannot be changed once programmed
  - EPROM (Erasable PROM)
    - Can be reprogrammed after erasing the entire chip
    - Erasing requires special tool that emits ultraviolet light
  - EEPROM (Electrically Erasable PROM)
    - No special tools needed to reprogram
    - Can erase only portions of the chip one byte at a time
    - Flash memory is a form of EEPROM

Cache Memory

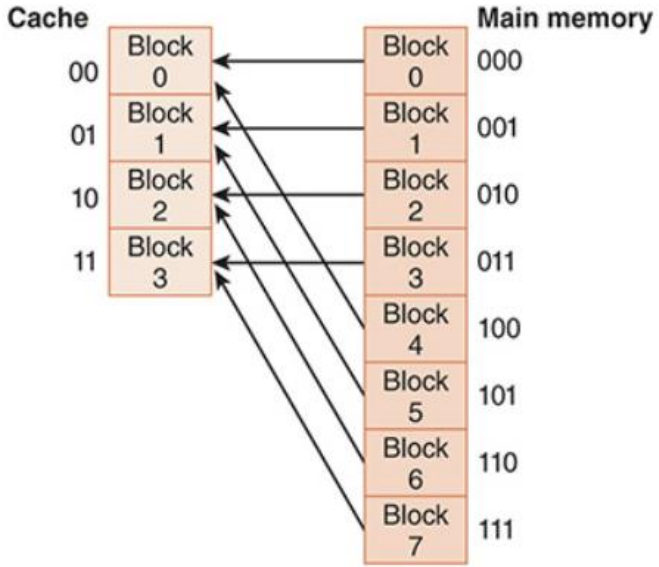
- Level 1 (L1) cache
  - resident on the CPU chip
  - fastest, smallest cache usually around 8KB to 64 KB.

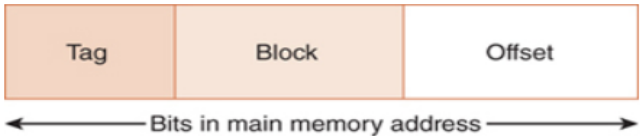
- Level 2 (L2) cache
  - located outside of the CPU
  - found on the motherboard, separate chip, or expansion board.
  - larger than L1 cache…usually around 64KB to 2MB
  - slower access speeds than L1 cache.

- Level 3 (L3) cache
  - extra cache between the processor and memory on processors that include L2 cache as part of their architecture
  - typically range from 8MB to 256MB.

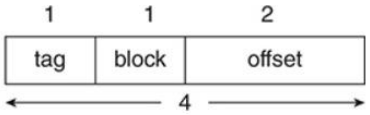
Cache Memory Mapping

Direct-Mapped Cache

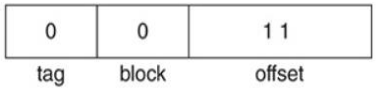




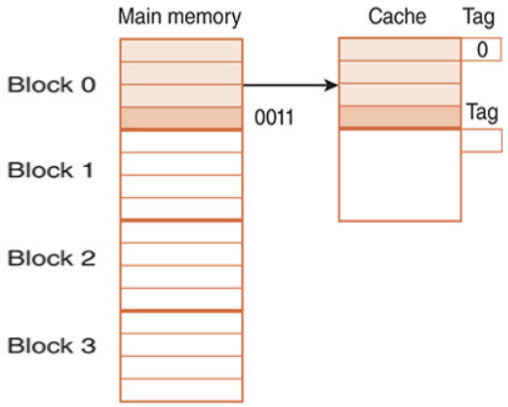
Direct-Mapped Cache



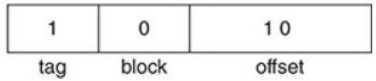
Main memory format



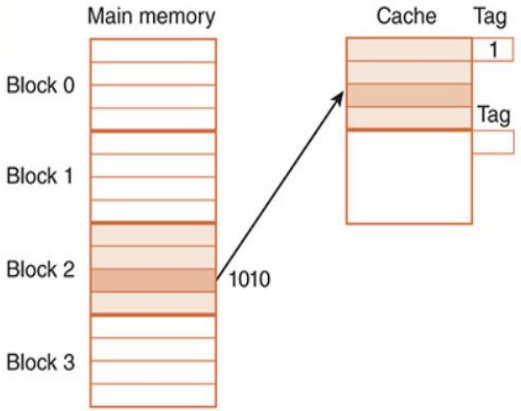
The address 0011 partitioned into fields



Mapping of block containing
Address 0011 = 0x3



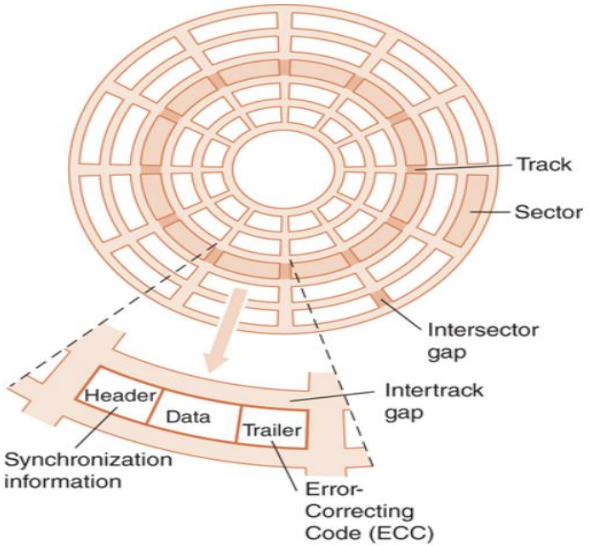
The address 1010 partitioned into fields



Mapping of Block Containing
Address 1010 = 0xA

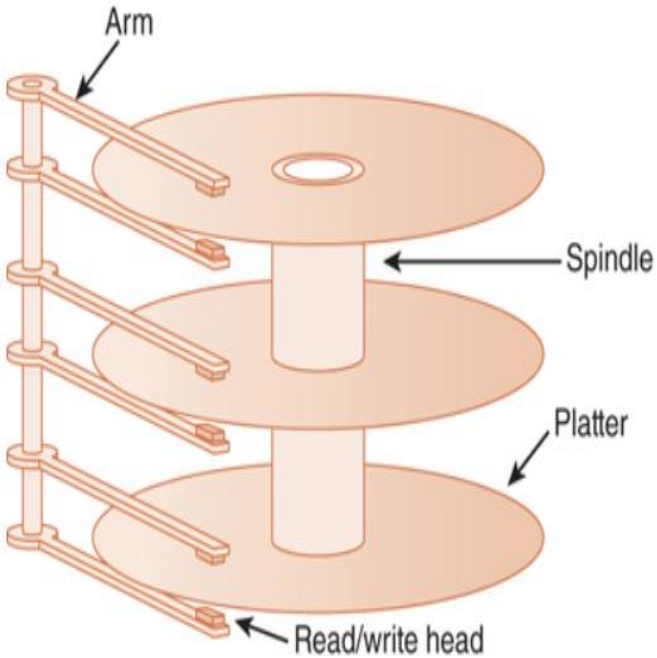
Fixed Disk Storage

- Fixed == rigid == hard disk
- Consist of one or more metal or glass disk with magnetic bonded surface for data storage
- Disk surface divided into tracks
- Tracks are divided into sectors
- Sectors contain header, data and error-correcting code



Fixed Disk Storage

- Steps to access data from spinning disk
  1. Position head over the proper track (seek)
  2. Wait for desired sector to rotate under the r/w head
  3. Transfer block of bits



SSD (Solid State Drives)

- **SSD or Solid-state drives** are a type of memory based drive
- Consist of a microcontroller and flash memory
- 100x faster than hard drives
- Used in desktops, laptops and applications requiring high performance

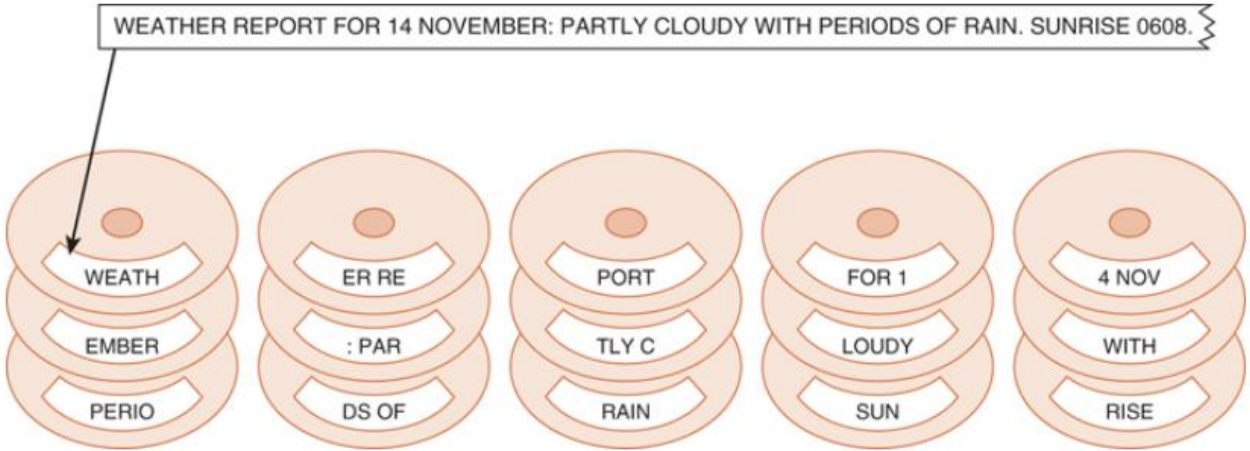
RAID (Redundant Array of Independent Disks)

RAID Level	Description	Reliability	Throughput	Pros and Cons
0	Block interleave data striping	Worse than single disk	Very good	Least cost, no protection
1	Data mirrored on second identical set	Excellent	Better than single disk on reads, worse on writes	Excellent protection, high cost
2	Bit interleave data striping with Hamming code	Good	Very good	Good performance, high cost, not used in practice
3	Bit interleave data striping with parity disk	Good	Very good	Good performance, reasonable cost
4	Block interleave data striping with one parity disk	Very good	Much worse on writes as single disk, very good on reads	Reasonable cost, poor performance, not used in practice

RAID Level	Description	Reliability	Throughput	Pros and Cons
5	Block interleave data striping with distributed parity	Very good	On writes not as good as single disk, very good on reads	Good performance, reasonable cost
6	Block interleave data striping with dual error protection	Excellent	On writes much worse than single disk, very good on reads	Good performance, reasonable cost, complex to implement
10	Mirrored disk striping	Excellent	Better than single disk on reads, not as good as single disk on writes	Good performance, high cost, excellent protection
50	Parity with striping	Excellent	Excellent read performance. Better than RAID-5; not as good as RAID 10	Good performance, high cost, good protection

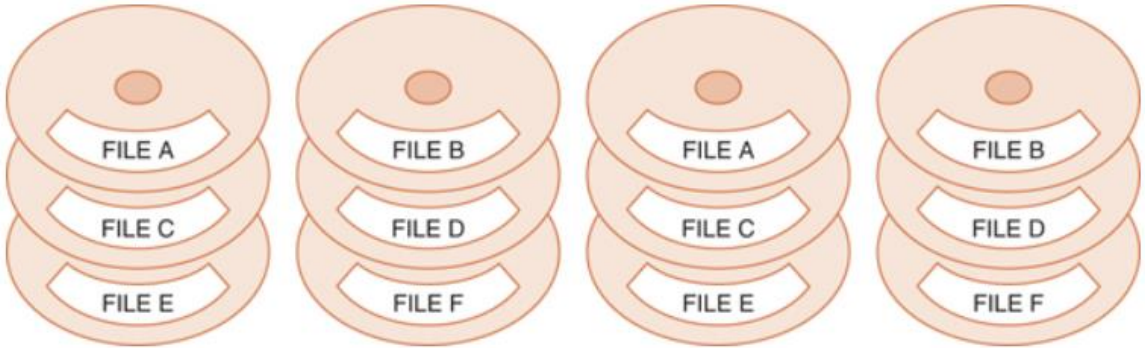
RAID (Redundant Array of Independent Disks)

- **RAID Level 0**
  - Data segmented and spread sequentially across drives (data striping)
  - No redundancy
  - High performance



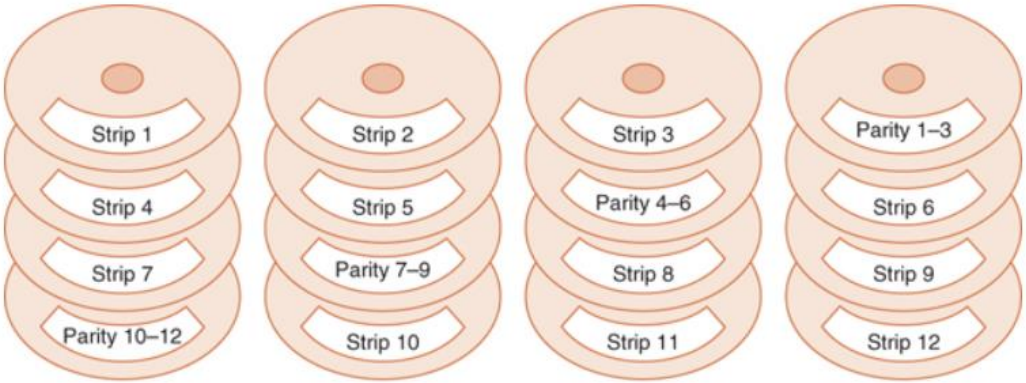
RAID (Redundant Array of Independent Disks)

- **RAID Level 1**
  - Disk mirroring
  - Full redundancy
  - High availability
  - Acceptable performance



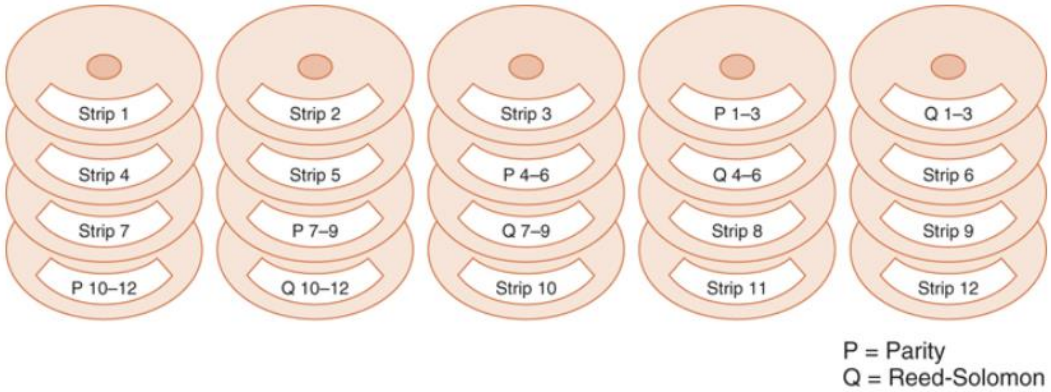
RAID (Redundant Array of Independent Disks)

- **RAID Level 5**
  - XOR parity calculations on data blocks
  - Parity results load balanced across disks
  - Best data protection for cost



RAID (Redundant Array of Independent Disks)

- **RAID Level 6**
  - Duplicate XOR parity calculations on data blocks
  - Duplicate parity results load balanced across disks
  - Extreme high availability
  - Complex to implement



RAID (Redundant Array of Independent Disks)

- **Hybrid RAID Level 10**
  - Combines Raid 0 striping with Raid 1 mirroring
  - Extreme high availability
  - Good performance
  - Extremely expensive to implement