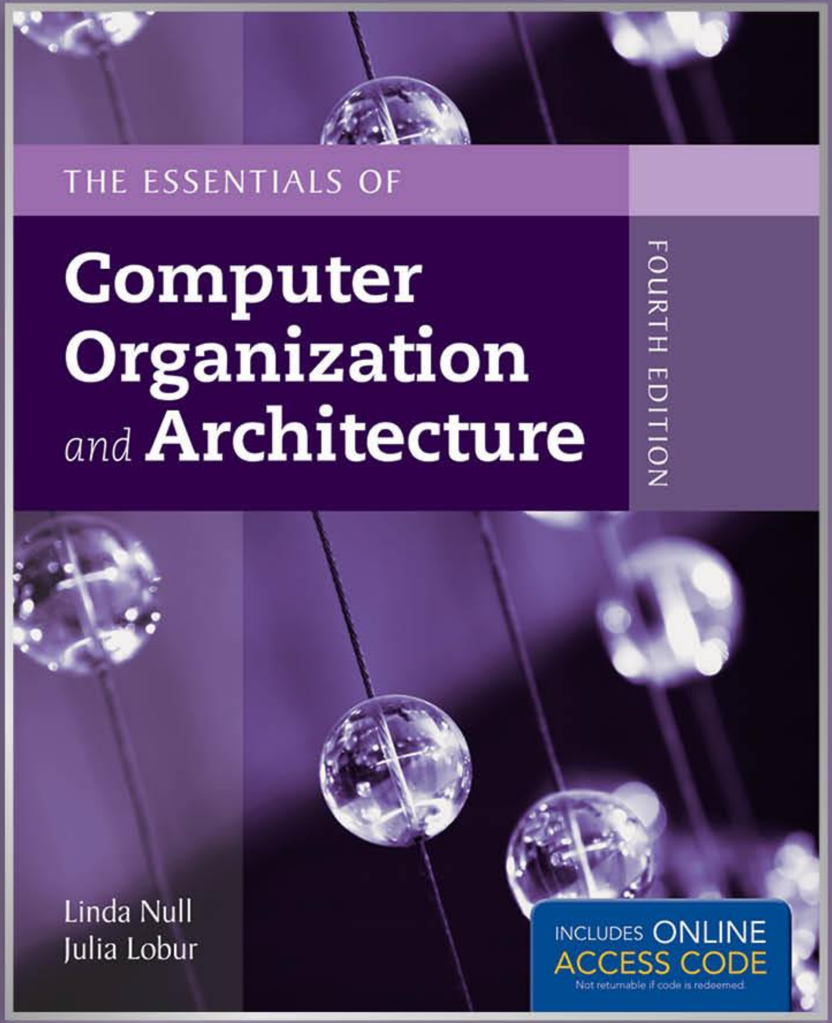
Chapter 7

Input/Output and Storage Systems



Chapter 7 Objectives

- Understand how I/O systems work, including I/O methods and architectures.
- Become familiar with storage media, and the differences in their respective formats.
- Understand how RAID improves disk performance and reliability, and which RAID systems are most useful today.
- Be familiar with emerging data storage technologies and the barriers that remain to be overcome.

7.1 Introduction

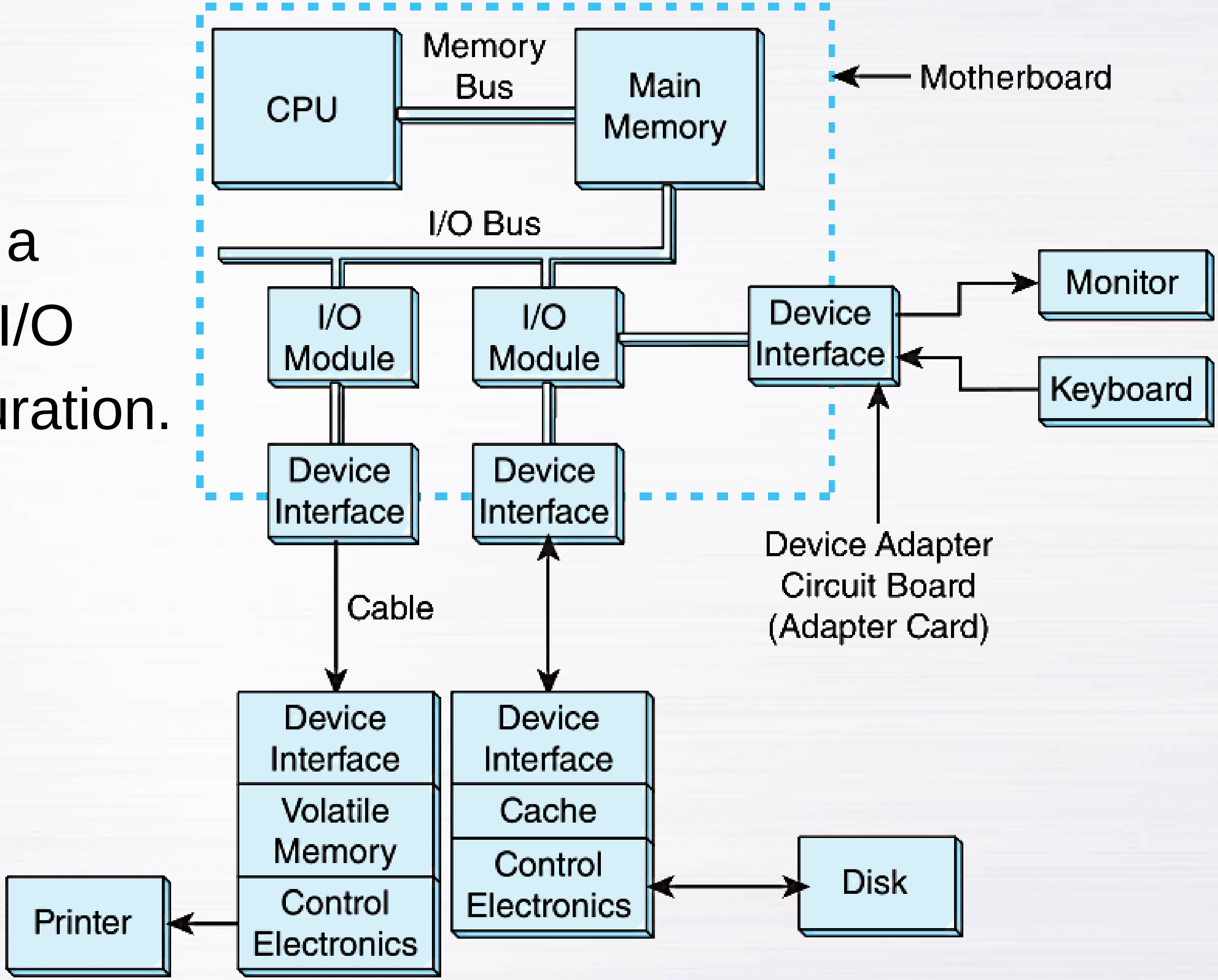
- Data storage and retrieval is one of the primary functions of computer systems.
  - One could easily make the argument that computers are more useful to us as data storage and retrieval devices than they are as computational machines.
- All computers have I/O devices connected to them, and to achieve good performance I/O should be kept to a minimum!
- In studying I/O, we seek to understand the different types of I/O devices as well as how they work.

7.4 I/O Architectures

- We define input/output as a subsystem of components that moves coded data between external devices and a host system.
- I/O subsystems include:
  - Blocks of main memory that are devoted to I/O functions.
  - Buses that move data into and out of the system.
  - Control modules in the host and in peripheral devices
  - Interfaces to external components such as keyboards and disks.
  - Cabling or communications links between the host system and its peripherals.

7.4 I/O Architectures

This is a model I/O configuration.



7.4 I/O Architectures

- I/O can be controlled in five general ways.
  - *Programmed I/O* reserves a register for each I/O device. Each register is continually polled to detect data arrival.
  - *Interrupt-Driven I/O* allows the CPU to do other things until I/O is requested.
  - *Memory-Mapped I/O* shares memory address space between I/O devices and program memory.
  - *Direct Memory Access* (*DMA*) offloads I/O processing to a special-purpose chip that takes care of the details.
  - *Channel I/O* uses dedicated I/O processors.

Programmed I/O

- CPU has direct control over I/O
- After issued command, CPU must wait for I/O module to complete operation
- If processor is faster than the I/O, Wastes CPU time

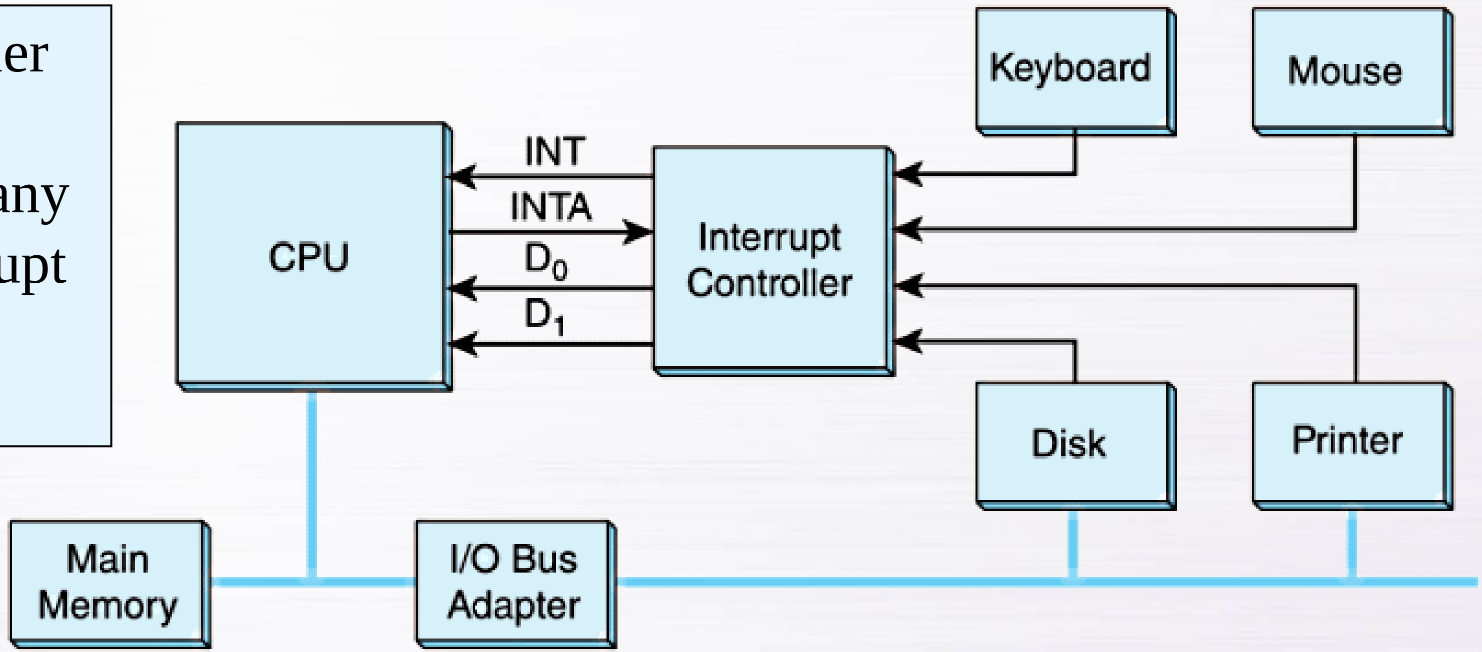
- E.g. legacy mouse controller
  – If the mouse has moved, the command buffer will flag this
  – The CPU can then fetch the mouse data and act accordingly
  – very expensive and wastes processor time
  – Wastes CPU time, since the mouse may not have moved between polling cycles.
- Other devices, i.e. all legacy serial port, parallel ports, midi, joystick, PS/2 keyboards, interval timers

7.4 I/O Architectures: Interrupt-driven

This is an idealized I/O subsystem that uses interrupts.

Each device connects its interrupt line to the interrupt controller.

The controller signals the CPU when any of the interrupt lines are asserted.

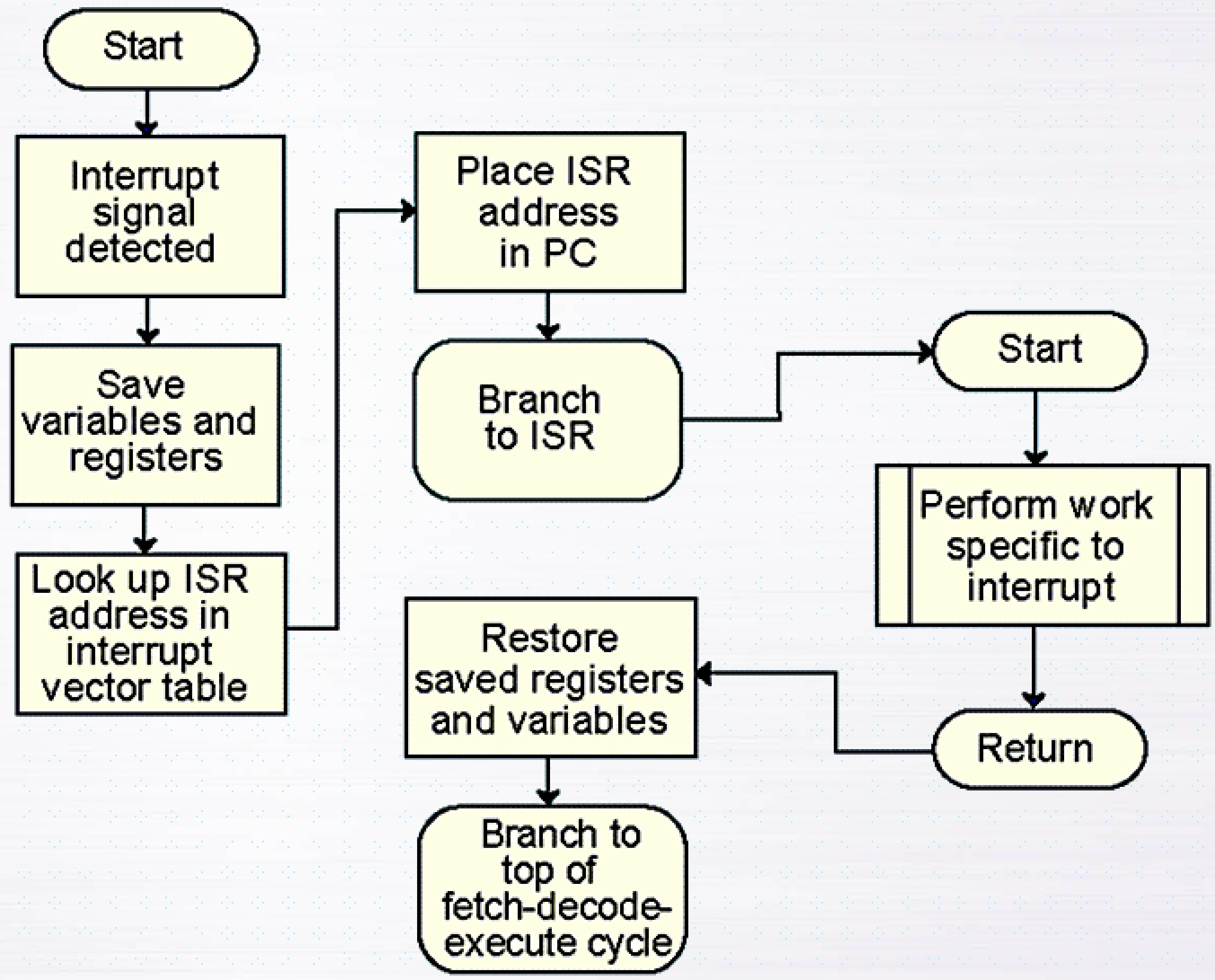


7.4 I/O Architectures: Interrupt-driven

- Recall from Chapter 4 that in a system that uses interrupts, the status of the interrupt signal is checked at the top of the fetch-decode-execute cycle.
- The particular code that is executed whenever an interrupt occurs is determined by a set of addresses called *interrupt vectors* that are stored in low memory.
- The system state is saved before the interrupt service routine is executed and is restored afterward.

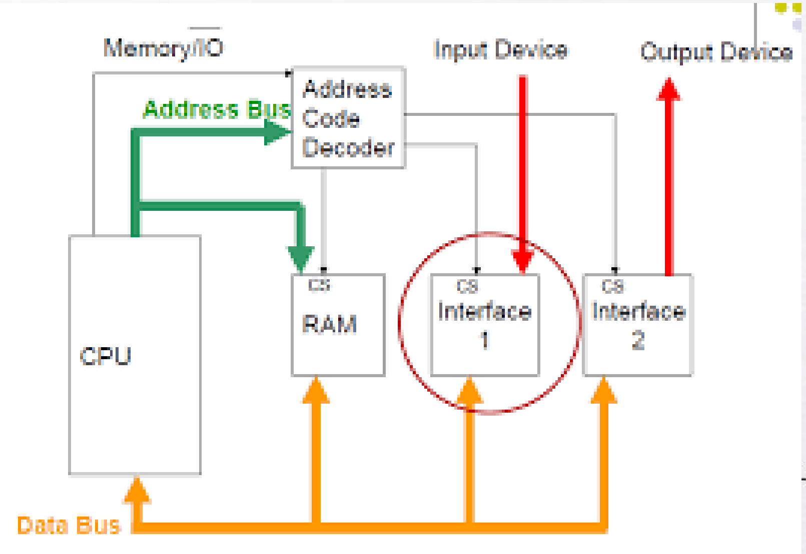
We provide a flowchart on the next slide.

7.4 I/O Architectures: Interrupt driven



7.4 I/O Architectures: Memory-mapped

- In memory-mapped I/O devices and main memory share the same address space.
  - Each I/O device has its own reserved block of memory.
  - Memory-mapped I/O therefore looks just like a memory access from the point of view of the CPU.
  - Thus the same instructions to move data to and from both I/O and memory, greatly simplifying system design.
- In small systems the low-level details of the data transfers are offloaded to the I/O controllers built into the I/O devices.

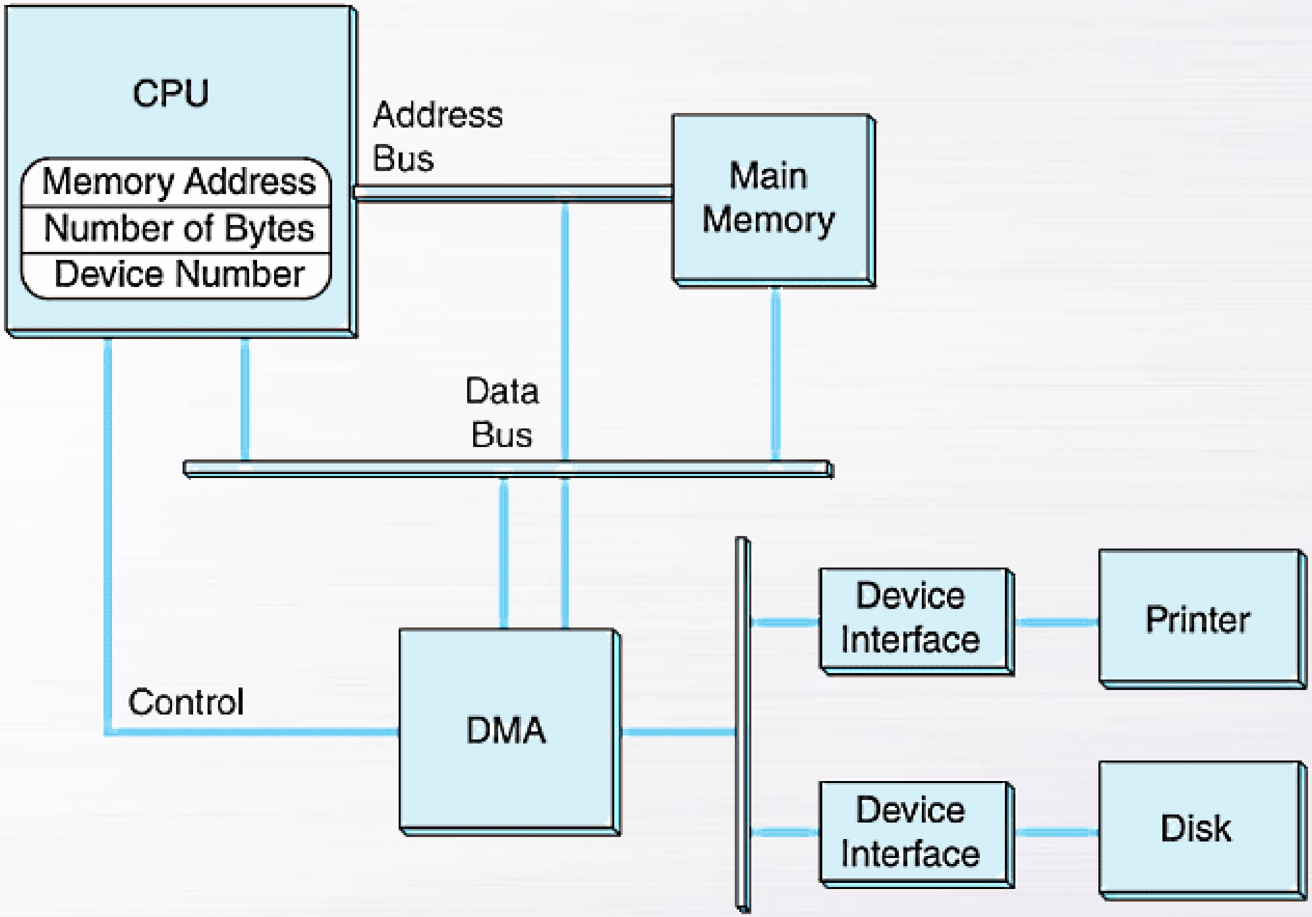


7.4 I/O Architectures: Direct Memory Access

This is a DMA configuration.

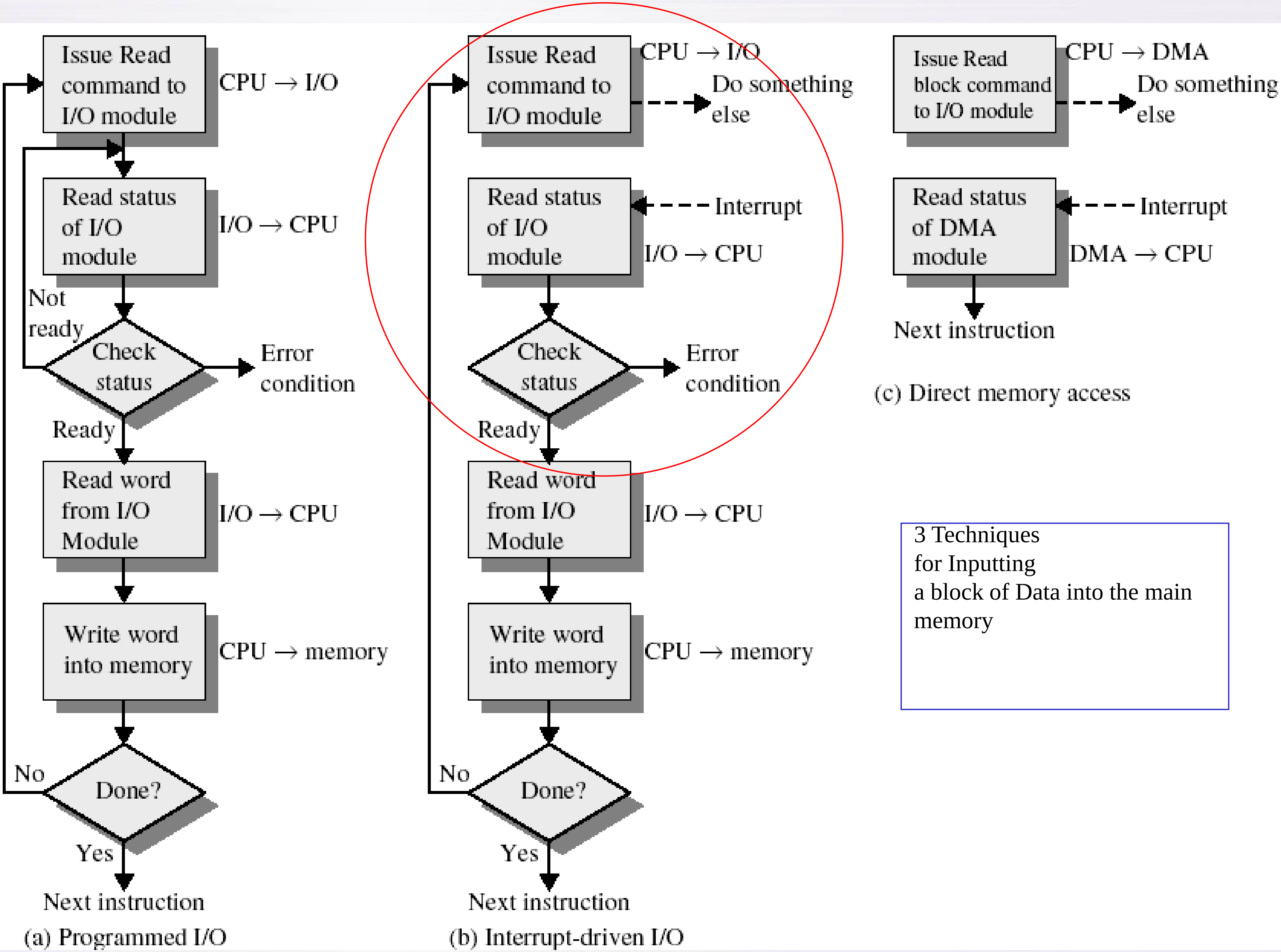
Notice that the DMA and the CPU share the bus.

The DMA runs at a higher priority and steals memory cycles from the CPU.



7.4 I/O Architectures: Channel I/O

- Very large systems employ channel I/O.
- Channel I/O processors (IOPs) I/O consists of one or more that control various channel paths.
- Slower devices such as terminals and printers are combined (*multiplexed*) into a single faster channel.
- On IBM mainframes, multiplexed channels are called *multiplexor channels*, the faster ones are called *selector channels*.



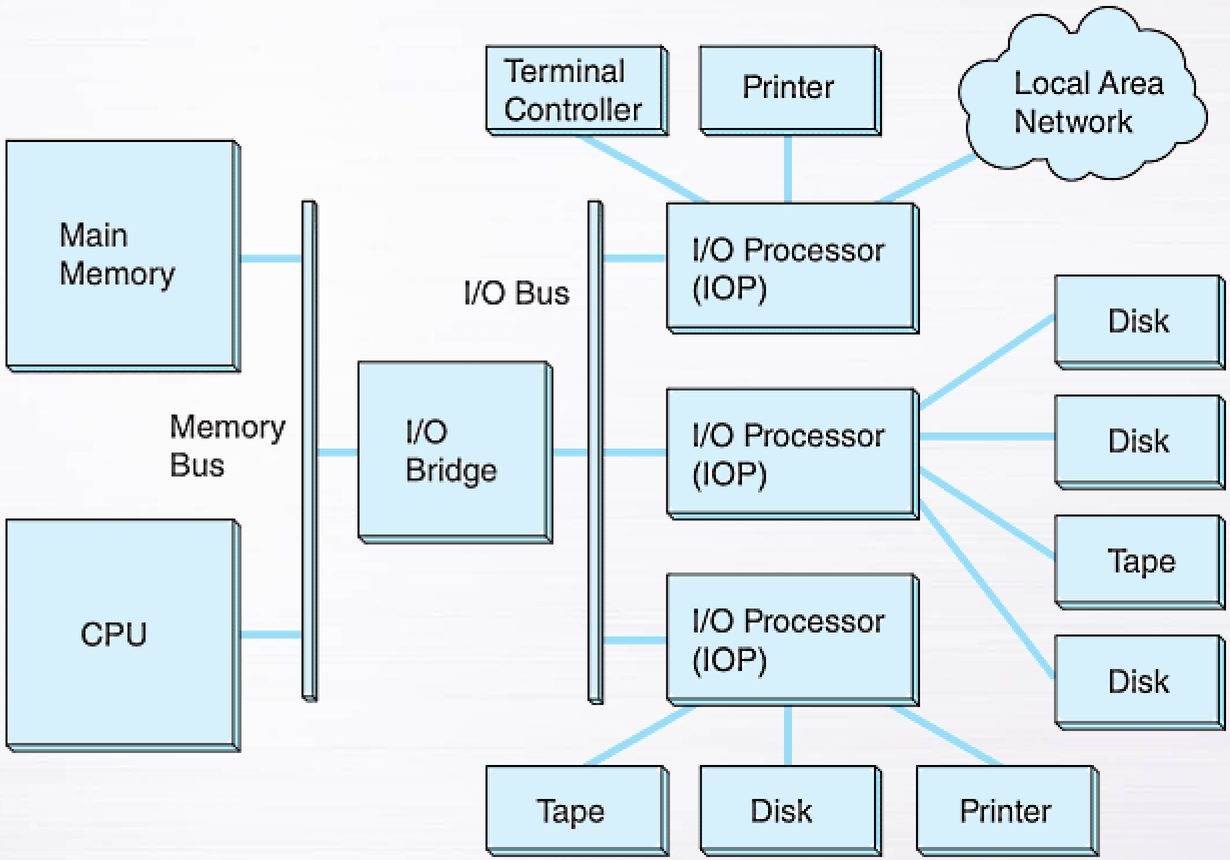
3 Techniques
for Inputting
a block of Data into the main
memory

7.4 I/O Architectures: Channel I/O

- Channel I/O is distinguished from DMA by the intelligence of the IOPs.
- The IOP negotiates protocols, issues device commands, translates storage coding to memory coding, and can transfer entire files or groups of files independent of the host CPU.
- The host has only to create the program instructions for the I/O operation and tell the IOP where to find them.

7.4 I/O Architectures: Channel I/O

- This is a channel I/O configuration.



7.4 I/O Architectures

- Character I/O devices process one byte (or character) at a time.
  - Examples include modems, keyboards, and mice.
  - Keyboards are usually connected through an interrupt-driven I/O system.
- Block I/O devices handle bytes in groups.
  - Most mass storage devices (disk and tape) are block I/O devices.
  - Block I/O systems are most efficiently connected through DMA or channel I/O.

7.4 I/O Architectures

- I/O buses, unlike memory buses, operate asynchronously. Requests for bus access must be arbitrated among the devices involved.
- Bus control lines activate the devices when they are needed, raise signals when errors have occurred, and reset devices when necessary.
- The number of data lines is the *width* of the bus.
- A bus clock coordinates activities and provides bit cell boundaries.

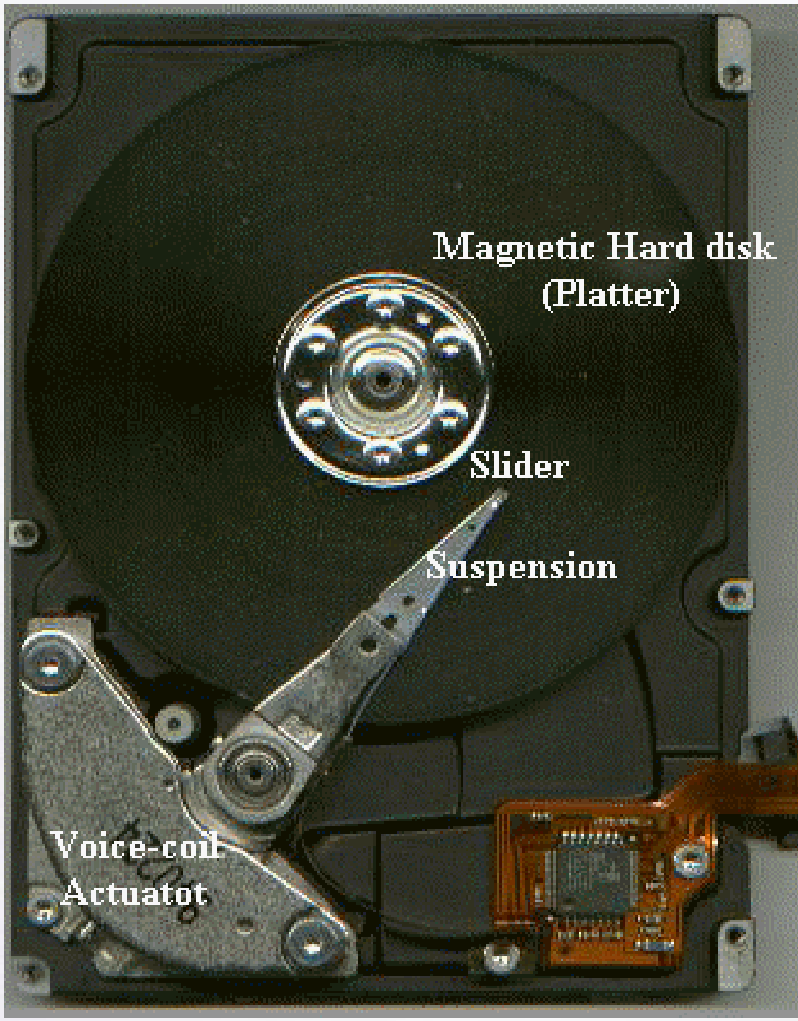
External storage

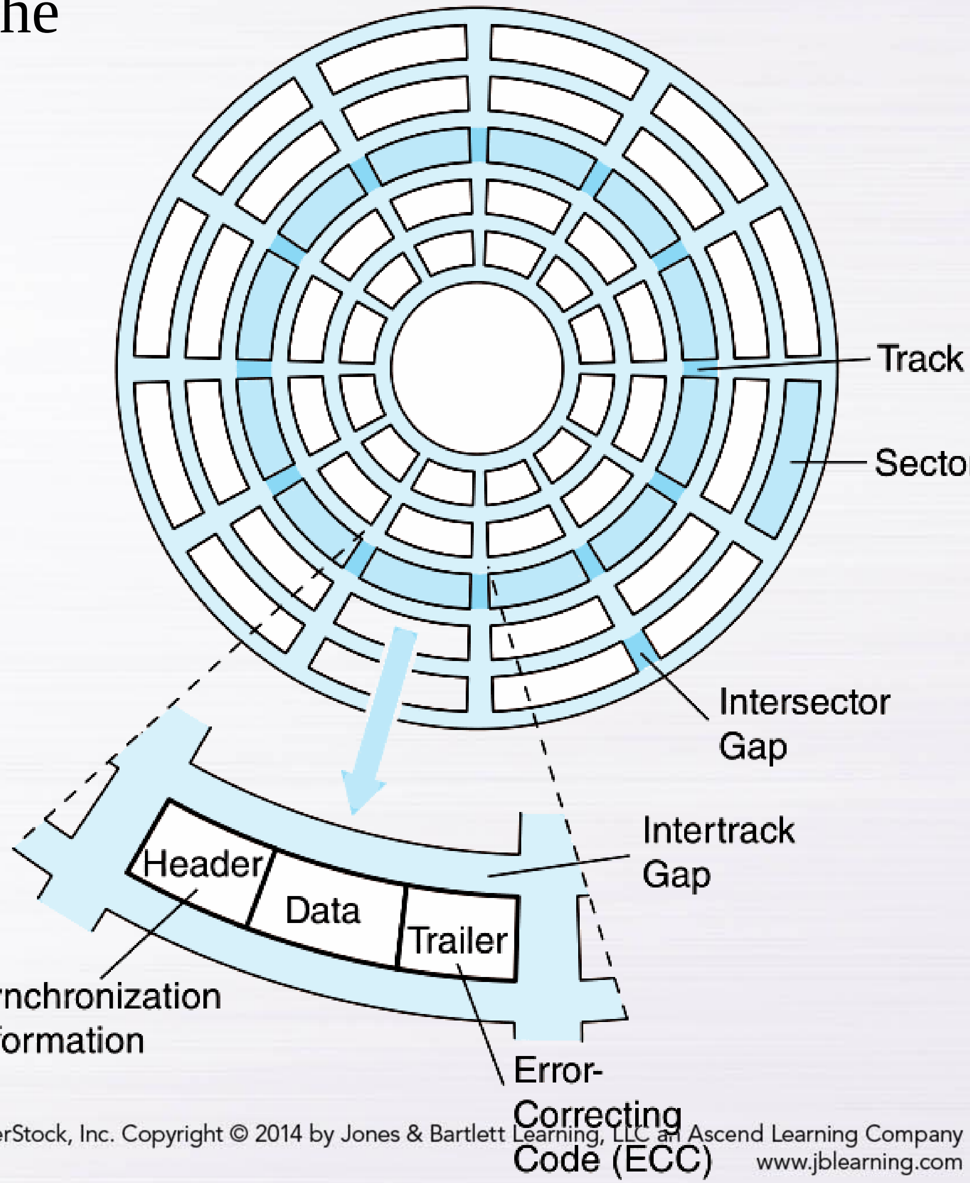
7.6 Magnetic Disk Technology

- Magnetic disks offer large amounts of durable storage that can be accessed quickly.
- Disk drives are called *random* (or *direct*) *access storage devices*, because blocks of data can be accessed according to their location on the disk.
  - This term was coined when all other durable storage (e.g., tape) was sequential.
- Magnetic disk organization is shown on the following slide.

7.6 Magnetic Disk Technology

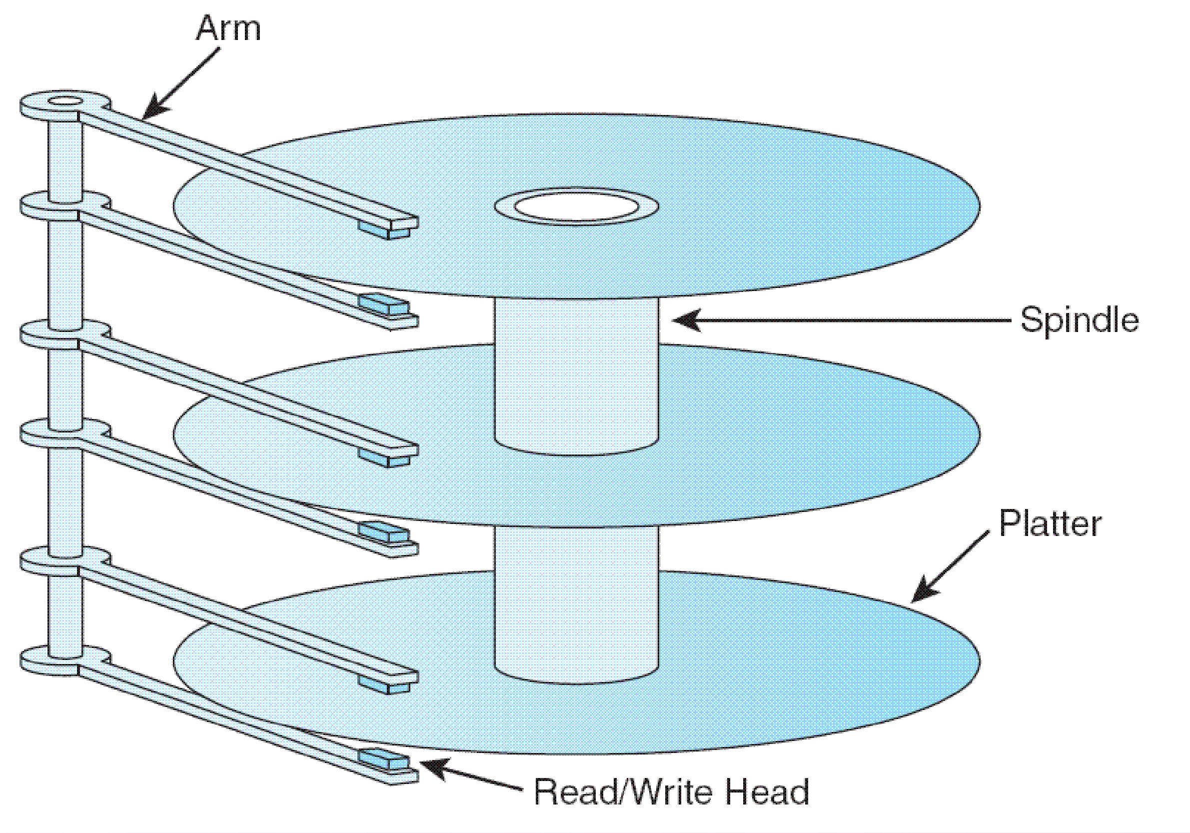
Disk tracks are numbered from the outside edge, starting with zero.





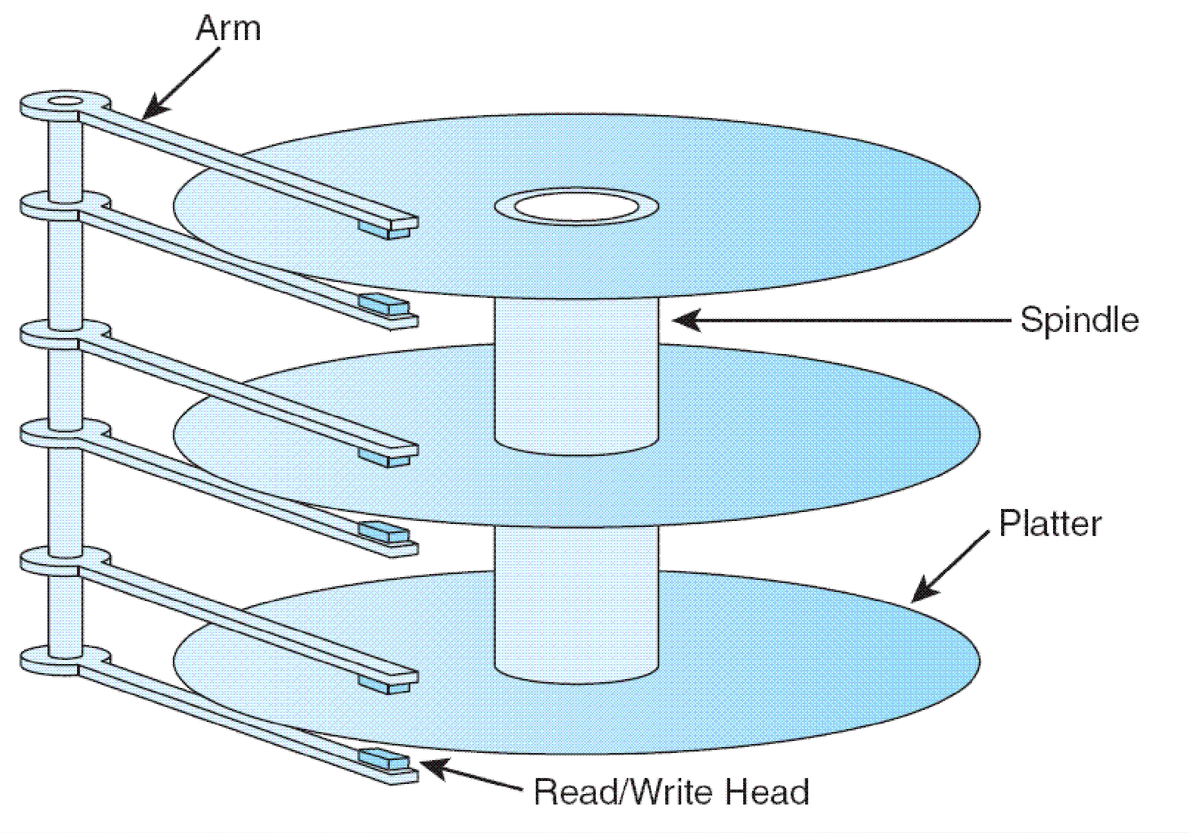
7.6 Magnetic Disk Technology

- Hard disk platters are mounted on spindles.
- Read/write heads are mounted on a comb that swings radially to read the disk.



7.6 Magnetic Disk Technology

- The rotating disk forms a logical cylinder beneath the read/write heads.
- Data blocks are addressed by their cylinder, surface, and sector.



7.6 Magnetic Disk Technology

- There are a number of electromechanical properties of hard disk drives that determine how fast its data can be accessed.
- *Seek time* is the time that it takes for a disk arm to move into position over the desired cylinder.
- *Rotational delay* is the time that it takes for the desired sector to move into position beneath the read/write head.
- Seek time + rotational delay = ***access time***.

7.6 Magnetic Disk Technology

- *Transfer rate* gives us the rate at which data can be read from the disk.
- *Average latency* is a function of the rotational speed:

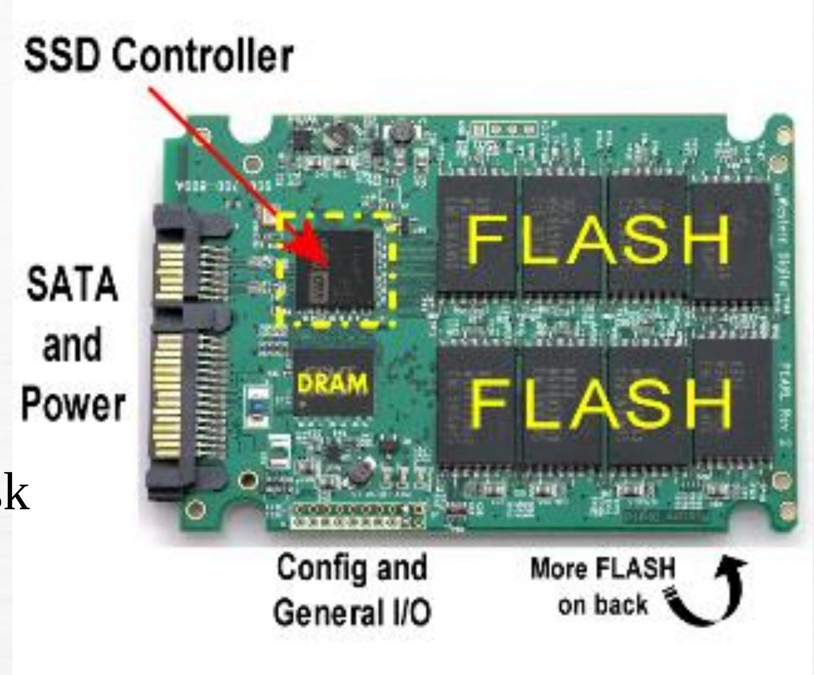
$$\frac{\dfrac{60 \text{ seconds}}{\text{disk rotation speed}} \times \dfrac{1000 \text{ ms}}{\text{second}}}{2}$$

- *Mean Time To Failure (MTTF)* is a statistically-determined value often calculated experimentally.
  – It usually doesn't tell us much about the actual expected life of the disk. *Design life* is usually more realistic.

Figure 7.15 in the text shows a sample disk specification.

7.6 Magnetic Disk Technology

- Low cost is the major advantage of hard disks.
- But their limitations include:
  - Very slow compared to main memory
  - Fragility
  - Moving parts wear out
- Reductions in memory cost enable the widespread adoption of *solid state drives*, *SSDs*.
  - Computers "see" SSDs as jut another disk drive, but they store data in non-volatile *flash* memory circuits.
  - Flash memory is also found in memory sticks and MP3 players.



7.6 Magnetic Disk Technology

- SSD access time and transfer rates are *typically* 100 times faster than magnetic disk, but slower than onboard RAM by a factor of 100,000.
  - There numbers vary widely among manufacturers and interface methods.
- Unlike RAM, flash is block-addressable (like disk drives).
  - The duty cycle of flash is between 30,000 and 1,000,000 updates to a block.
  - Updates are spread over the entire medium through *wear leveling* to prolong the life of the SSD.

7.7 Optical Disks

- Optical disks provide large storage capacities very inexpensively.
- They come in a number of varieties including CD-ROM, DVD, and WORM.
- Many large computer installations produce document output on optical disk rather than on paper. This idea is called COLD-- *Computer Output Laser Disk*
- It is estimated that optical disks can endure for a hundred years. Other media are good for only a decade-- at best.

7.7 Optical Disks

- CD-ROMs were designed by the music industry in the 1980s, and later adapted to data.
- This history is reflected by the fact that data is recorded in a single spiral track, starting from the center of the disk and spanning outward.
- Binary ones and zeros are delineated by bumps in the polycarbonate disk substrate. The transitions between pits and lands define binary ones.
- If you could unravel a full CD-ROM track, it would be nearly five miles (~8 km) long!

7.7 Optical Disks

- The logical data format for a CD-ROM is much more complex than that of a magnetic disk.
- Different formats are provided for data and music.
- Two levels of error correction are provided for the data format.
- Because of this, a CD holds at most 650MB of data, but can contain as much as 742MB of music.

7.7 Optical Disks

- DVDs can be thought of as quad-density CDs.
  - Varieties include single sided, single layer, single sided double layer, double sided double layer, and double sided double layer.
- Where a CD-ROM can hold at most 650MB of data, DVDs can hold as much as 17GB.
- One of the reasons for this is that DVD employs a laser that has a shorter wavelength than the CD's laser.
- This allows pits and lands to be closer together and the spiral track to be wound tighter.

7.7 Optical Disks

- A shorter wavelength light can read and write bytes in greater densities than can be done by a longer wavelength laser.
- This is one reason that DVD's density is greater than that of CD.
- The 405 nm wavelength of blue-violet light is much shorter than either red (750 nm) or orange (650 nm).
- The manufacture of blue-violet lasers can now be done economically, bringing about the next generation of laser disks.

7.7 Optical Disks

- The Blu-Ray disc format won market dominance over HD-CD owing mainly to the influence of Sony.
  - HD-CDs are backward compatible with DVD, but hold less data.
- Blu-Ray was developed by a consortium of nine companies that includes Sony, Samsung, and Pioneer.
  - Maximum capacity of a single layer Blu-Ray disk is 25GB.
  - Multiple layers can be "stacked" up to six deep.
  - Only double-layer disks are available for home use.

7.7 Optical Disks

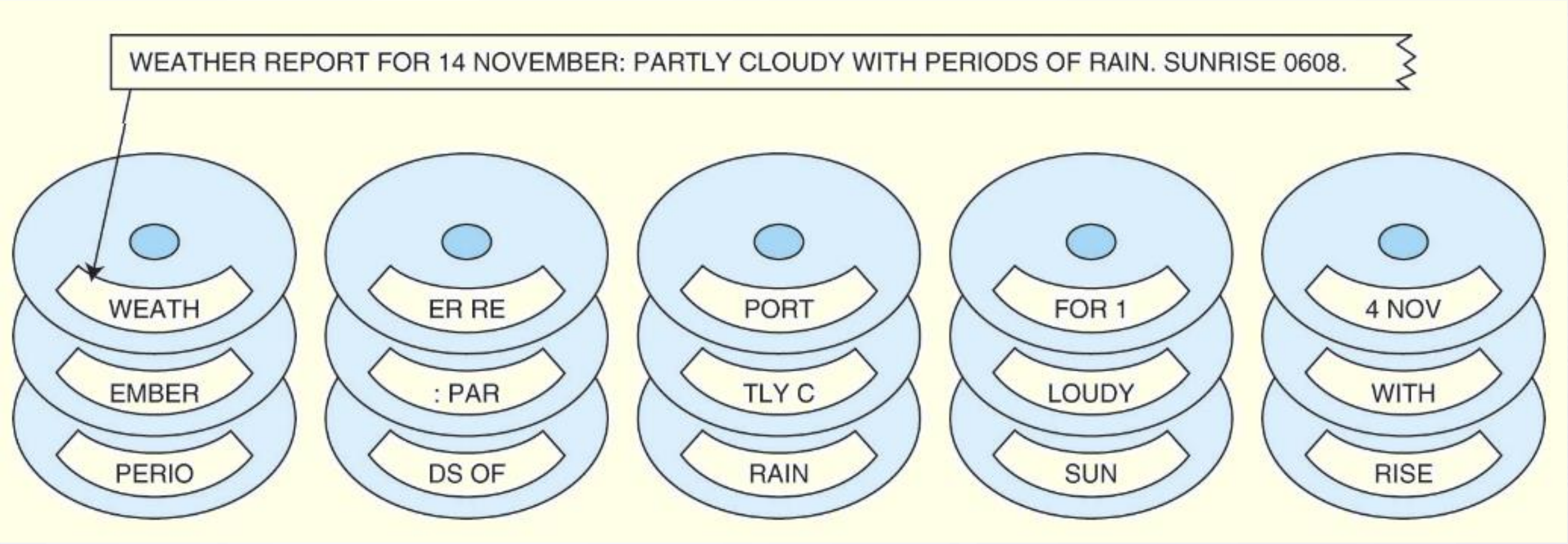
- Blue-violet laser disks are also used in the data center.
- The intention is to provide a means for long term data storage and retrieval.
- Two types are now dominant:
  - Sony’s Professional Disk for Data (PDD) that can store 23GB on one disk and
  - Plasmon’s Ultra Density Optical (UDO) that can hold up to 30GB.
- It is too soon to tell which of these technologies will emerge as the winner.

7.9 RAID

- RAID, an acronym for *Redundant Array of Independent Disks* was invented to address problems of disk reliability, cost, and performance.
- In RAID, data is stored across many disks, with extra disks added to the array to provide error correction (redundancy).
- The inventors of RAID, David Patterson, Garth Gibson, and Randy Katz, provided a RAID taxonomy that has persisted for a quarter of a century, despite many efforts to redefine it.

7.9 RAID

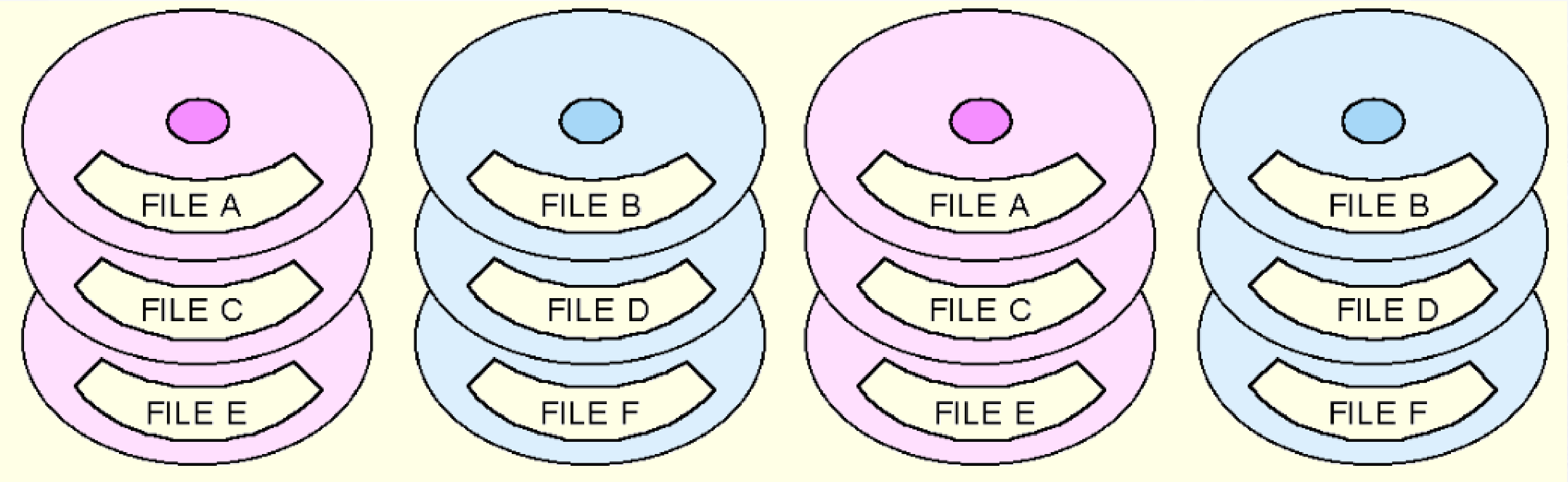
- RAID Level 0, also known as *drive spanning*, provides improved performance, but no redundancy.
  - Data is written in blocks across the entire array



  - The disadvantage of RAID 0 is in its low reliability.

7.9 RAID

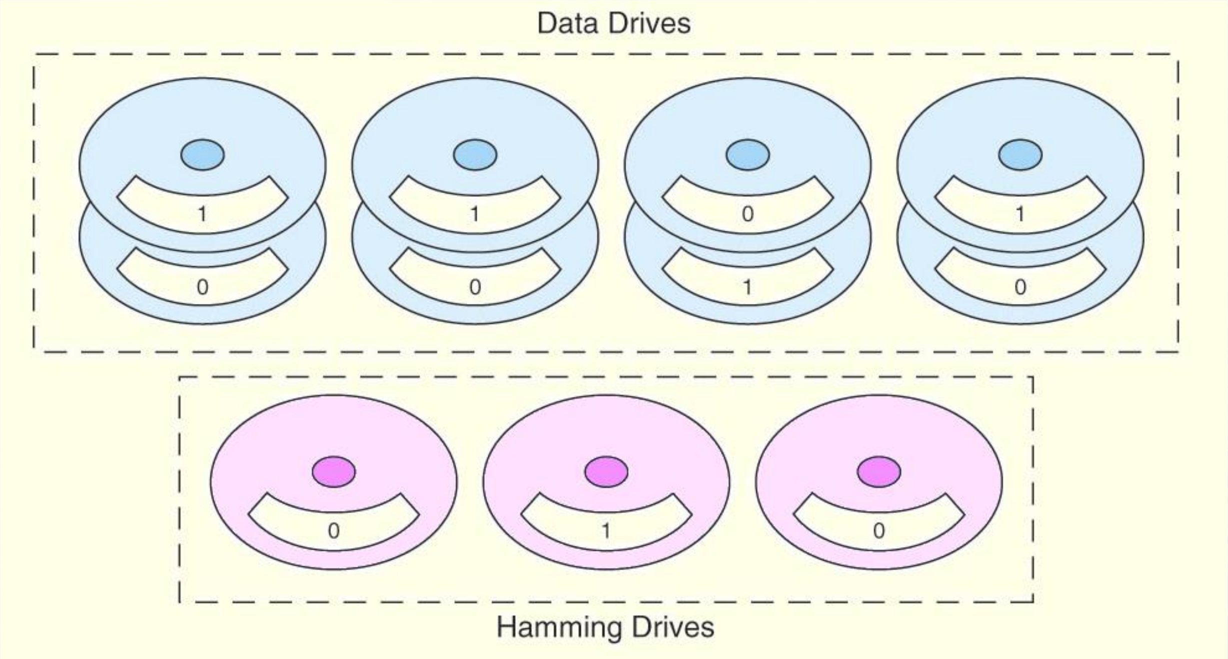
- RAID Level 1, also known as *disk mirroring*, provides 100% redundancy, and good performance.
  - Two matched sets of disks contain the same data.



  - The disadvantage of RAID 1 is cost.

7.9 RAID

- A RAID Level 2 configuration consists of a set of data drives, and a set of Hamming code drives.
  – Hamming code drives provide error correction for the data drives.

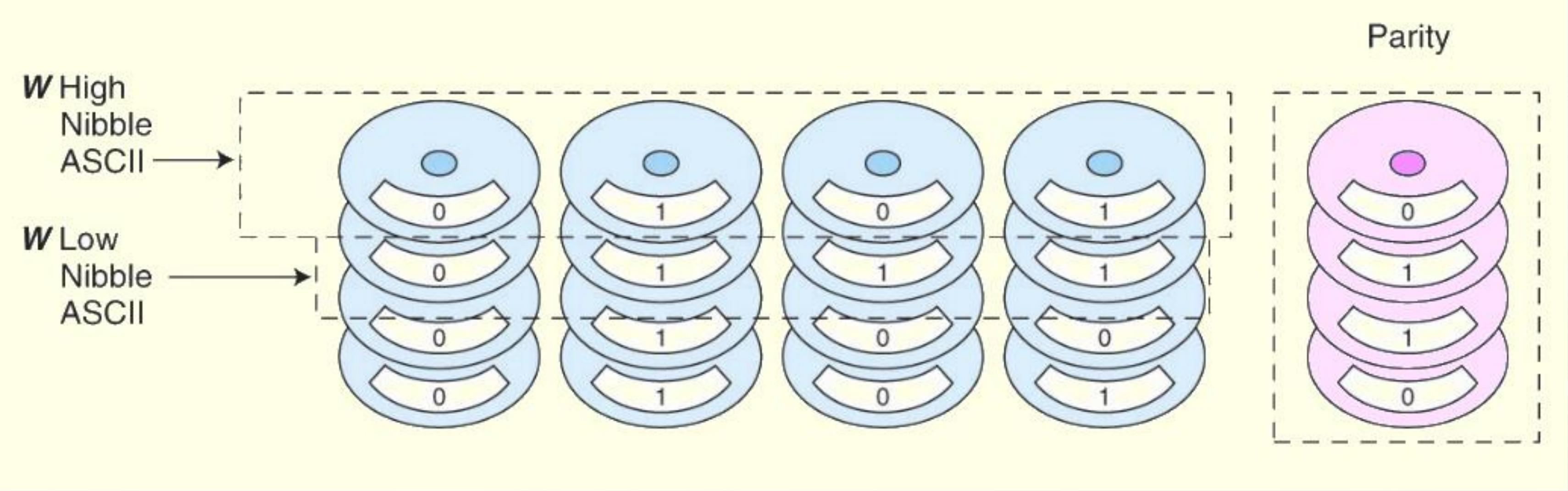


  – RAID 2 performance is poor and the cost is relatively high.

7.9 RAID

- RAID Level 3 stripes bits across a set of data drives and provides a separate disk for parity.
  - Parity is the XOR of the data bits.

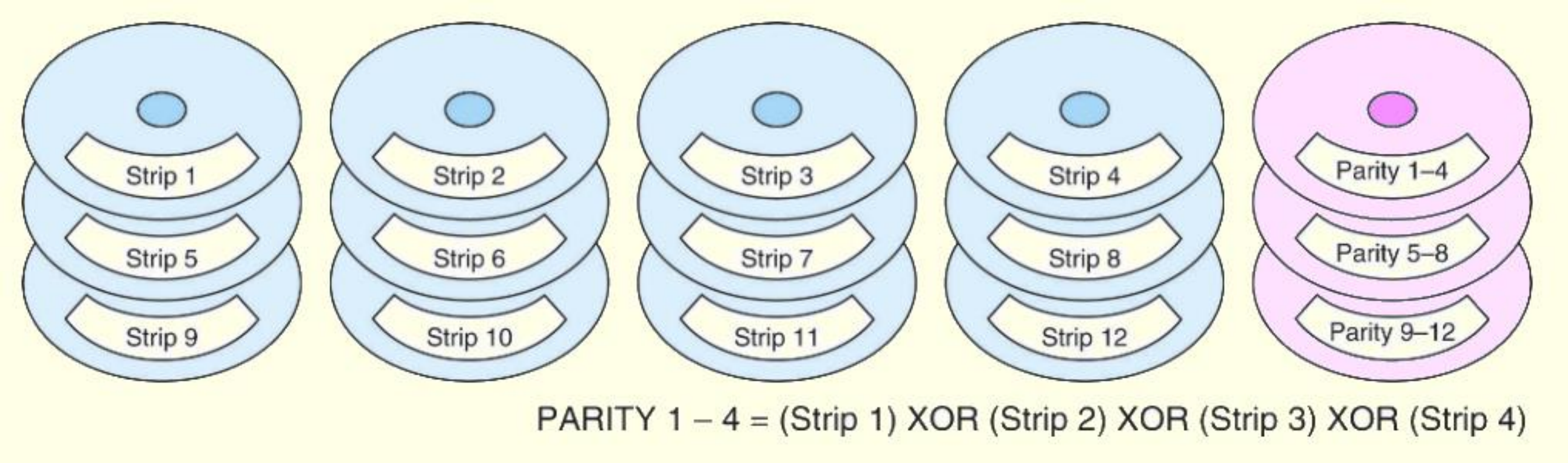
if drive X1 failed,
we get $X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$



-RAID 3 is not suitable for commercial applications, but is good for personal systems.

7.9 RAID

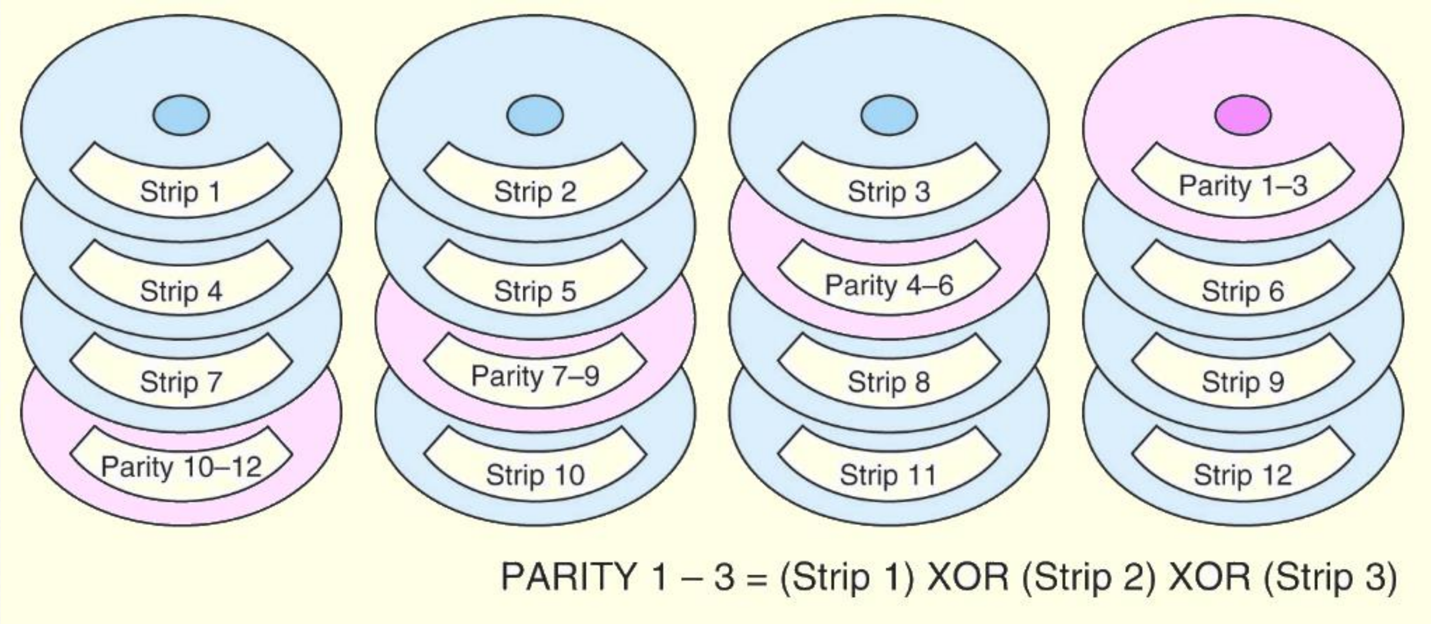
- RAID Level 4 is like adding parity disks to RAID 0.
  - Data is written in blocks across the data disks, and a parity block is written to the redundant drive.



  - RAID 4 would be feasible if all record blocks were the same size.

7.9 RAID

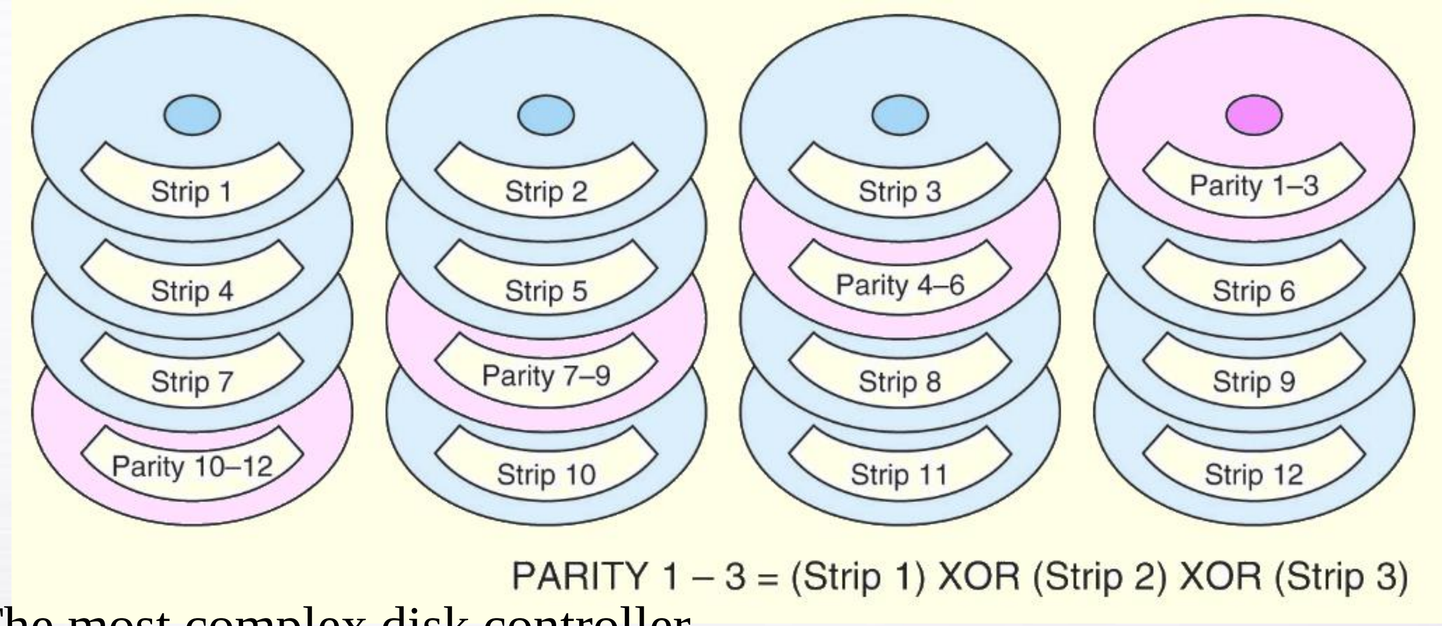
- RAID Level 5 is RAID 4 with distributed parity.
  - With distributed parity, some accesses can be serviced concurrently, giving good performance and high reliability.



  - RAID 5 is used in many commercial systems.

RAID-5

- RAID Level 5 is RAID 4 with distributed parity (= load balancing).
  - With distributed parity, some accesses can be serviced concurrently, giving good performance and high reliability.
    - **E.g. we can write strip 6 and strip 7 at the same time, why?**



  - The most complex disk controller

Updating parity bit in RAID 4 and 5

$$x4(i) = x3(i) \oplus x2(i) \oplus x1(i) \oplus x0(i)$$

to update a change in disk $x1$

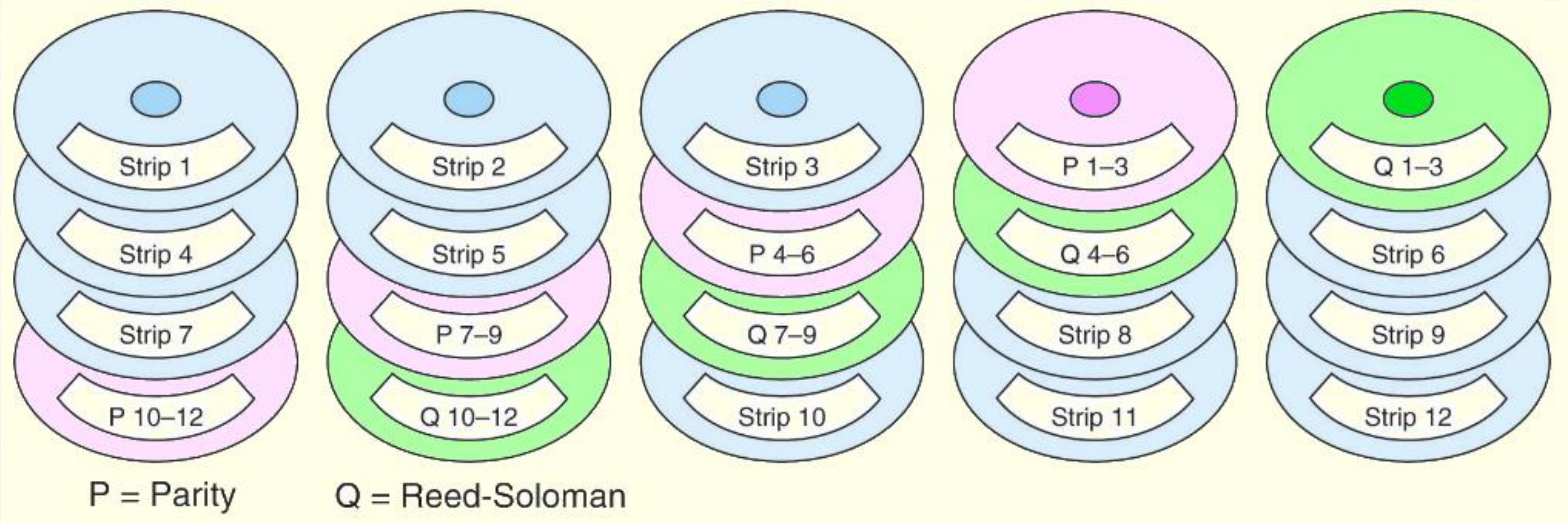
$$\begin{aligned}
x4'(i) &= x3(i) \oplus x2(i) \oplus x1'(i) \oplus x0(i) \\
&= x3(i) \oplus x2(i) \oplus x1'(i) \oplus x0(i) \oplus x1(i) \oplus x1(i) \\
&= \left[x3(i) \oplus x2(i) \oplus x1(i) \oplus x0(i)\right] \oplus x1(i) \oplus x1'(i) \\
&= \boxed{x4(i) \oplus x1(i) \oplus x1'(i)}
\end{aligned}$$

Old parity strip

Old data strip

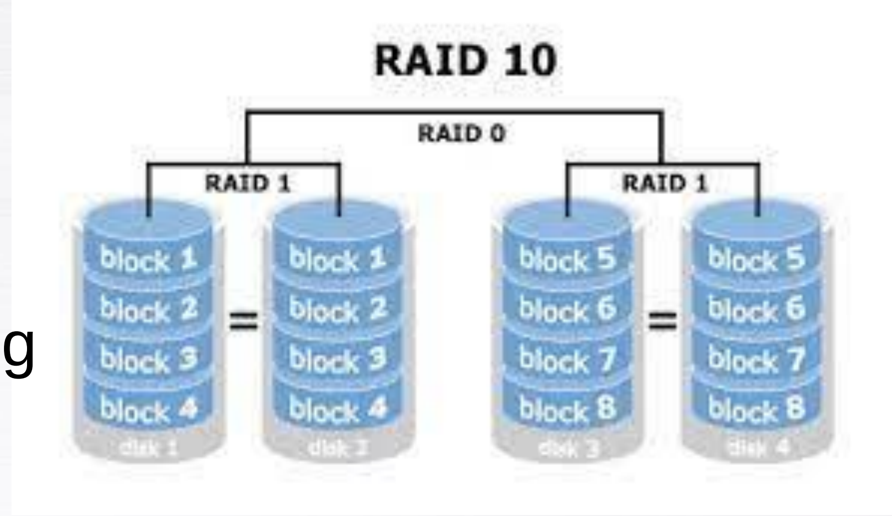
7.9 RAID

- RAID Level 6 carries two levels of error protection over striped data: Reed-Soloman and parity.
  - It can tolerate the loss of two disks.



  - RAID 6 is write-intensive, but highly fault-tolerant.

7.9 RAID



- Large systems may employ various RAID levels, depending on the criticality of the data on the drives.

- Critical, high-throughput files can benefit from combining RAID 0 with RAID 1, called RAID 10.
- RAID 50 combines striping and distributed parity. For good fault tolerance and high capacity.
  - Note: Higher RAID levels do not necessarily mean "better" RAID levels. It all depends upon the needs of the applications that use the disks.

7.10 The Future of Data Storage

- Advances in technology have defied all efforts to define the ultimate upper limit for magnetic disk storage.
  - In the 1970s, the upper limit was thought to be around 2Mb/in^2.
  - Today's disks commonly support 20Gb/in^2.
- Improvements have occurred in several different technologies including:
  - Materials science
  - Magneto-optical recording heads.
  - Error correcting codes.

7.10 The Future of Data Storage

- Present day biological data storage systems combine organic compounds such as proteins or oils with inorganic (magentizable) substances.
- Early prototypes have encouraged the expectation that densities of $1Tb/in^2$ are attainable.
- Of course, the ultimate biological data storage medium is DNA.
  - Trillions of messages can be stored in a tiny strand of DNA.
- Practical DNA-based data storage is most likely decades away.

Chapter 7 Conclusion

- I/O systems are critical to the overall performance of a computer system.
- I/O systems consist of memory blocks, cabling, control circuitry, interfaces, and media.
- I/O control methods include programmed I/O, interrupt-based I/O, DMA, and channel I/O.

Chapter 7 Conclusion

- Magnetic disk is the principal form of durable storage.
- Disk performance metrics include seek time, rotational delay, and reliability estimates.
- Enterprise SSDs save energy and provide improved data access for government and industry.
- Optical disks provide long-term storage for large amounts of data, although access is slow.

Chapter 7 Conclusion

- RAID gives disk systems improved performance and reliability. RAID 3 and RAID 5 are the most common.
- RAID 6 protect against dual disk failure
- Any one of several new technologies including biological may someday replace magnetic disks.
- The hardest part of data storage may be in locating the data after it’s stored.