16th Edition

Understanding Computers

Today and Tomorrow

Comprehensive

Chapter 2
The System Unit, Processing, and Memory

Deborah Morley
Charles S. Parker

Overview

- This chapter covers:
  - How computers represent data and program instructions
  - How the central processing unit (CPU) and memory are arranged with other components inside the system unit
  - How a CPU performs processing tasks
  - Several strategies are used today to create faster and better computers today and in the future

Data and Program Representation

- Digital data representation is the process of representing data in digital form so it can be understood by a computer
  - Coding systems are used to represent data and programs in a manner understood by the computer
  - Digital computers can only understand two states, off and on (0 and 1)

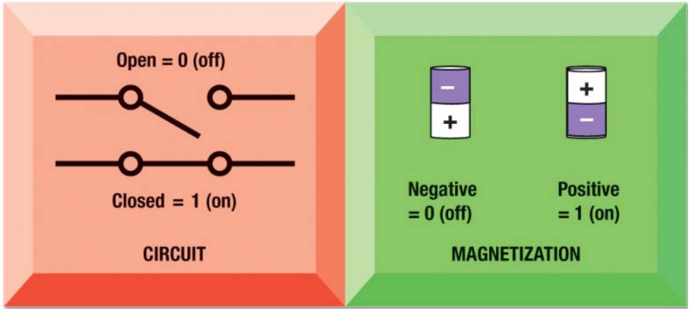
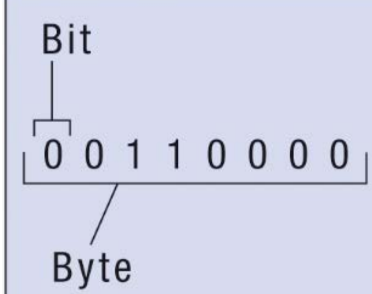


FIGURE 2-1
Ways of representing 0 and 1. Binary computers recognize only two states—off and on—usually represented by 0 and 1.

Digital Data Representation

- A **bit** is the smallest unit of data that a binary computer can recognize (a single 1 or 0)
  - Can be interpreted as 0/1 or True/False
    - True/False is Boolean value
  - Denoted w/lowercase b
- A **byte** contains 8 bits
  - Byte terminology used to express the size of files (documents, programs, etc.)
    - Byte is smallest addressable unit in Memory
      - Contents
      - Address
    - Denoted w/uppercase B
    - Be aware of files measured in Bytes (B) and Network speeds measured in b/s



Abbreviation	Approximate Size
KB	1 thousand bytes
MB	1 million bytes
GB	1 billion bytes
TB	1 trillion bytes
PB	1,000 terabytes
EB	1,000 petabtyes
ZB	1,000 exabytes
YB	1,000 zettabytes

FIGURE 2-2
Bits and bytes.
Document size, storage capacity, and memory capacity are all measured in bytes.

Bytes & Sizes

- Prefixes are often used to express larger quantities of bytes:
  - **1000 = kilobyte (KB)**
  - **1 Million = megabyte (MB)**
  - **1 Billion = gigabyte (GB)**
  - **1 Trillion terabyte (TB)**, **petabyte (PB)**, **exabyte (EB)**, **zettabyte (ZB)**, **yottabyte (YB)**

Representing Numerical Data: The Binary Numbering System

- The **binary numbering system** uses only two symbols (1 and 0) to represent all possible numbers
  - The numbering system computers use (Von Neumann)
- The **decimal numbering system** uses 10 symbols (range 0-9)
  - The numbering system people use
- In both systems, the position of the digits determine the power to which the base number (such as 10 or 2) is raised
- Please read ciss100.com numbering systems introduction:
  - http://www.ciss100.com/lecture-topics-modules/architecture-hardware/numbering-systems/ and textbook ASCII coding charts

Example of the Decimal and Binary Numbering Systems

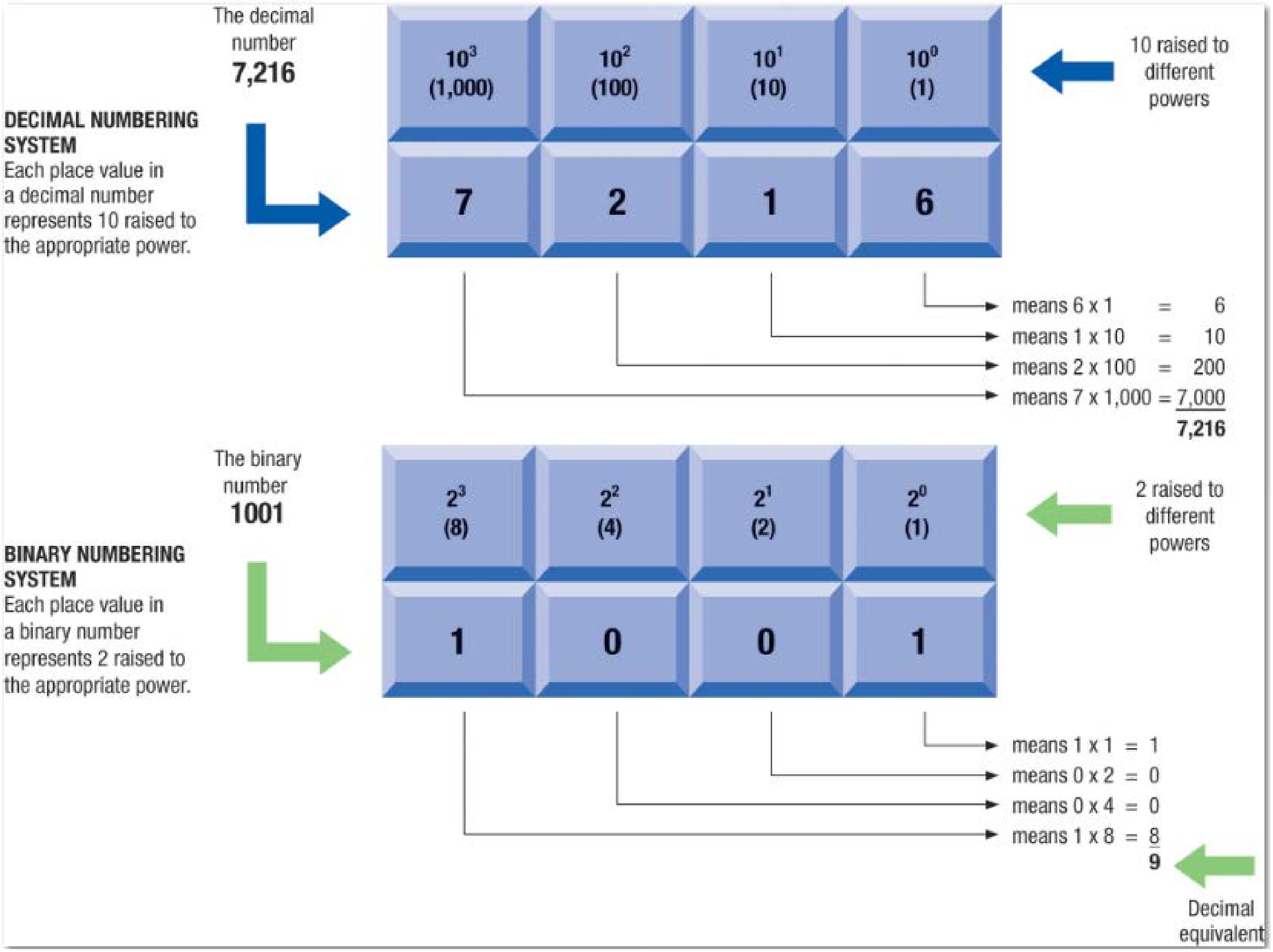


FIGURE 2-3
Examples of using the decimal and binary numbering systems.

Symbolic Representation

- We learned we can represent:
  - Numbers (Binary)
  - Truth values (Boolean => True/False)
- We need to represent:
  - Text
  - Graphics
  - Audio
  - Objects => programming
  - Processes => programming

Coding Systems for Text-Based Data: ASCII and EBCDIC

- **ASCII (American Standard Code for Information Interchange)** is the coding system traditionally used with personal computers
  - Most commonly uses 8 bits
- EBCDIC (Extended Binary-Coded Decimal Interchange Code)
  - Developed by IBM, primarily for mainframes

CHARACTER	ASCII
0	00110000
1	00110001
2	00110010
3	00110011
4	00110100
5	00110101
A	01000001
B	01000010
C	01000011
D	01000100
E	01000101
F	01000110
+	00101011
!	00100001
#	00100011

FIGURE 2-4
Some extended ASCII code examples.

Coding Systems for Text-Based Data: Unicode

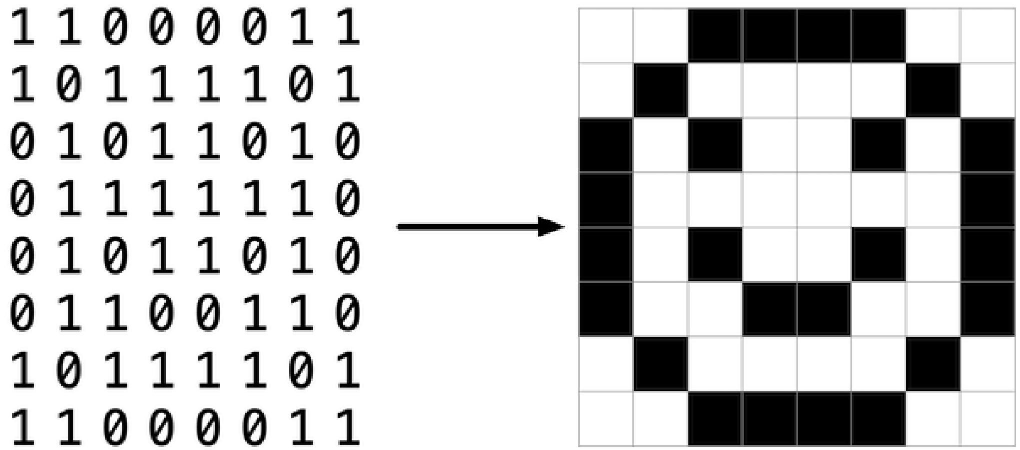
- **Unicode**
  - Newer code consisting of 8 to 32 bits per character
  - Universal coding standard designed to represent text-based data written in any ancient or modern language, as well as thousands of other symbols and signs
  - Replacing ASCII as the primary text-coding system

FIGURE 2-5
Unicode. Many characters, such as these, can be represented by Unicode but not by ASCII or EBCDIC.

Coding Systems for Other Types of Data: Graphics Data

- Graphics data (still images such as photos or drawings) consist of bitmapped images made up of a grid of small dots called pixels
  - Monochrome graphic can only be one of two colors
    - Black or White
    - Requires just one bit for color representation
  - Images with more than two colors
    - Can use 4, 8, or 24 bits to store the color data for each pixel
    - Grayscale (ranges from white to black)
  - The number of bits per pixel depends on the type of image
    - JPEG images taken by most digital cameras today use 24-bit true color images; large images can be compressed
    - Common image formats include TIF, BMP, GIF, and PNG

Bitmapped Image Representation



8 x 8 cells/grid above

- 1 bit/cell/pixel => Black/White => 8 bytes
- 1 Byte/cell/pixel => Grayscale (0 – 255) => 64 bytes
- 3 Bytes/cell/pixel => RBG (1 Byte each for red/green/blue) => 192 bytes
- Add 4th Byte (32 bit) => 256 bytes
  - simulate texture => gradients/shadows/transparencies/etc.

Representing Graphics Data

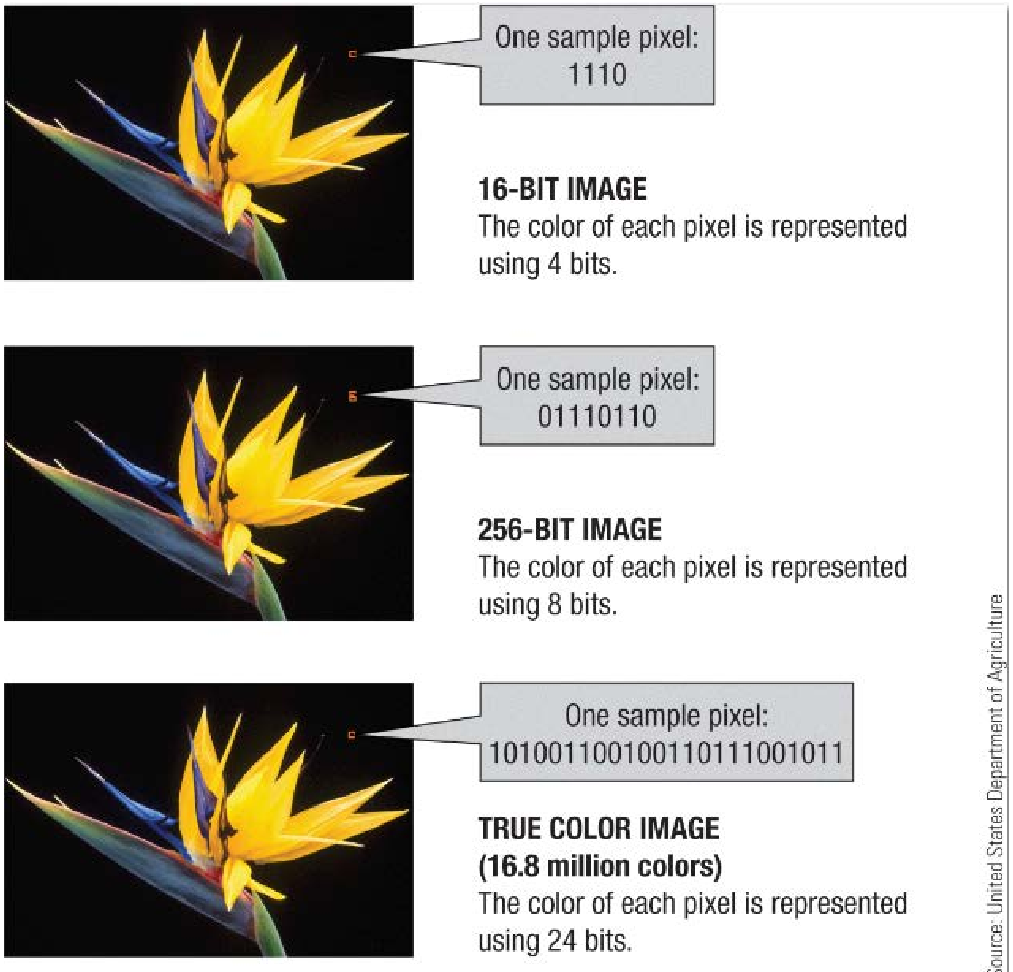


FIGURE 2-6

Representing graphics data. With bitmapped images, the color of each pixel is represented by bits; the more bits used, the better the image quality.

Vector Graphics (not in text)

- Mathematical Vector is magnitude and direction
- Vector Graphics are points on Cartesian plane
  - Connected lines and curves => polygons
  - Scale up/down without aliasing
    - => no loss in quality/no pixelation

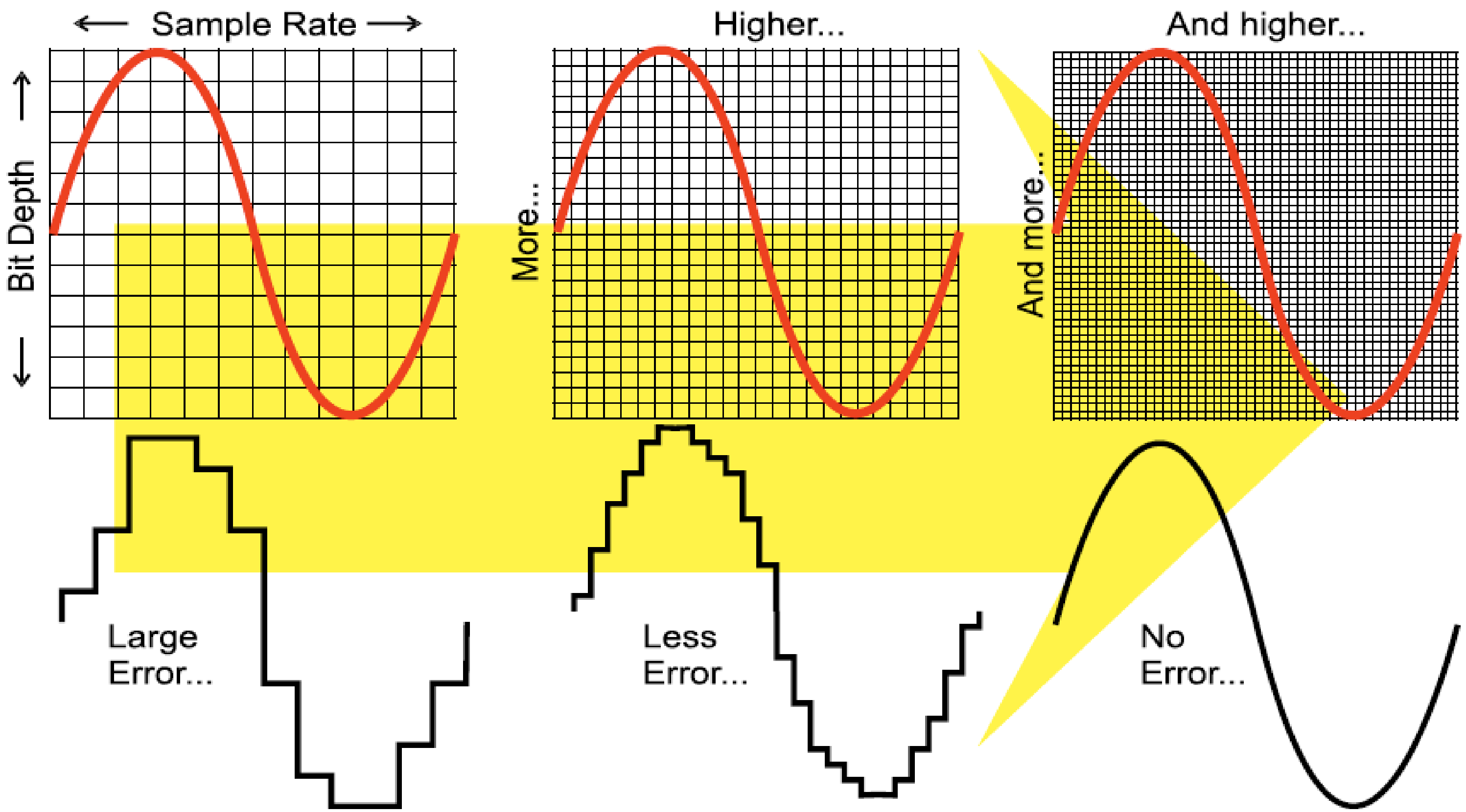
Encoding Audio Data

- Audio data must be in digital form in order to be stored, processed or communicated (network) by a computer
  - Often compressed when sent over the Internet
    - MP3 files are 10 times smaller than their uncompressed digital versions
    - Compressed files download more quickly and take up less storage space
- Lossy vs. Lossless Compression
  - Lossy => Human Perception
  - Lossless => Data

Audio Digitization

MP3 BitRate = Bit Depth (vertical) x Sampling Rate (horizontal)

- **Digitization inherently lossy**



Encoding Video Data

- Video data is displayed using a collection of frames, each frame contains a still image
  - Typically 24 frames per second
  - Amount of data can be substantial, but can be compressed
- Video digitization inherently lossy

Representing Software Programs: Machine Language

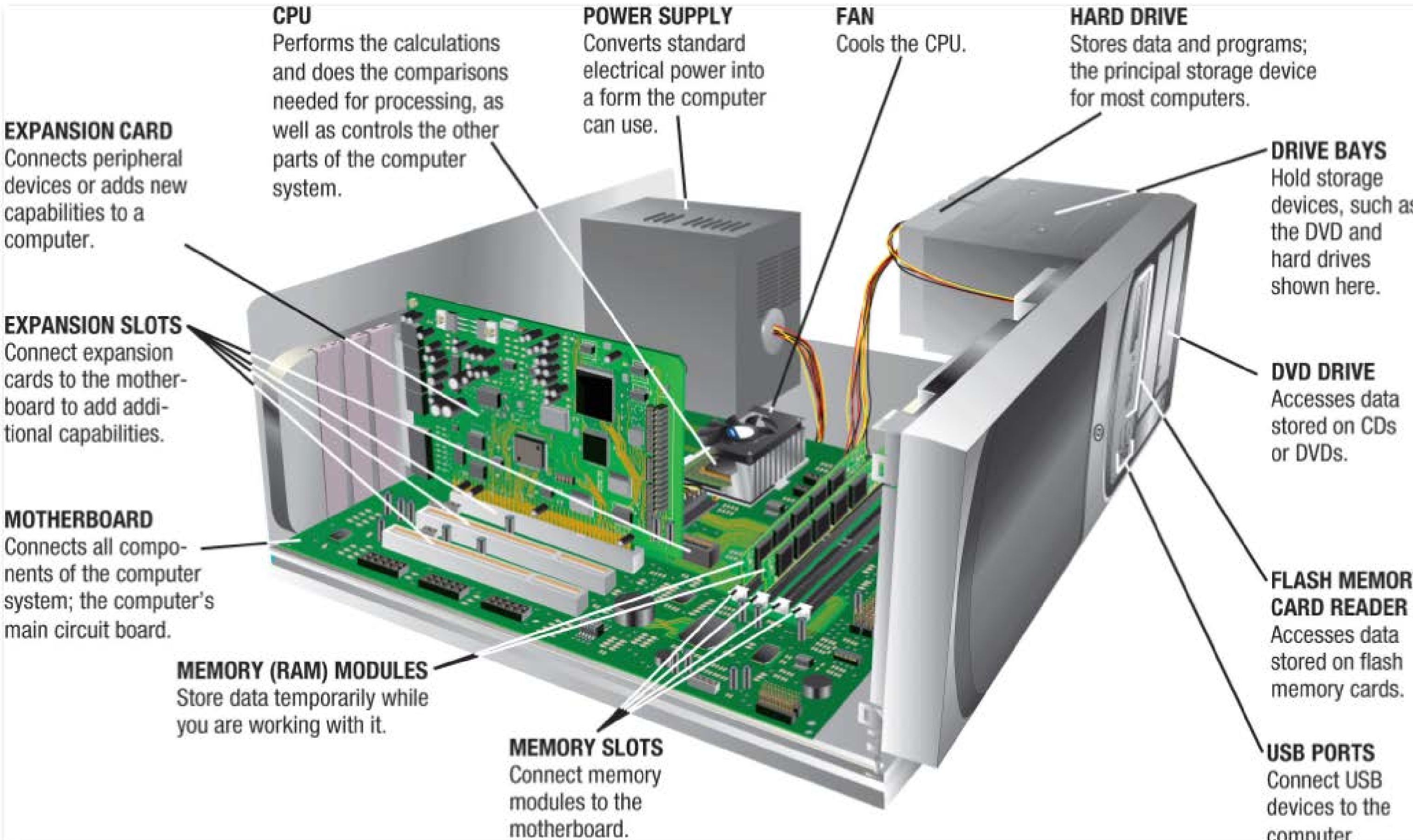
- **Machine language** is a binary-based language for representing computer programs the computer can execute directly (Instruction set => Architecture)
  - Early programs written in machine language (1st Gen)
  - Today's computers programs written in a High Level programming language
    - High Level PL translated by computer into machine language to run on computer
    - Compilation (static – pre-execution)
      - vs
    - Interpretation (dynamic – during execution)

Inside the System Unit

- The **system unit** is the main case of a computer or mobile device
  - Houses processing hardware for a computer
  - Contains other components => storage devices, power supply, cooling hardware, one or more processors, several types of memory, and interfaces to peripheral devices
  - Interconnected through sets of wires called buses on the motherboard

Inside a Desktop System Unit

FIGURE 2-7
Inside a desktop system unit.



The Motherboard

- The **motherboard** is a circuit board consisting of computer chips, also called integrated circuits (ICs)
  - ICs contain interconnected components (such as transistors) to perform particular functions
  - All devices connect via a wired/wireless connection to the motherboard (wired directly or slots)
  - External devices connect by plugging into a port
    - Port is either built directly into the motherboard or created via an expansion card
  - Wireless external devices use either a transceiver or wireless networking technology

The Power Supply and Drive Bays

- The power supply connects to motherboard to deliver electricity
- Portable computers use rechargeable battery pack
  - Built-in batteries more difficult and expensive to replace, resulting in electronic waste (e-waste)
- Drive bays are rectangular metal racks inside the system unit that house storage devices
  - Hard drive, CD/DVD drive, flash memory card reader
  - Storage devices also connect to the motherboard

Processors and the CPU

- **Processor** consists of circuitry and components packaged together and connected directly to the motherboard
- The **CPU (central processing unit)** does the vast majority of processing for a computer
  - Called a microprocessor when talking about personal computers
  - Typically designed for a specific type of computer
    - Desktops, servers, and some notebook PCs use Intel or Advanced Micro Devices (AMD) processors
    - Portable computers and mobile devices often use Intel or AMD mobile processors(CISC) or an ARM processor (RISC) instead

Processing Speed

- Processing speed can be measured by the CPU's clock speed
  - Rated in megahertz (MHz) or gigahertz (GHz)
  - Higher CPU clock speed => more instructions processed per second
  - Alternate measure of processing speed is the number of instructions a CPU can process per second
    - Floating Point Operation -> flop
    - Megaflops (millions), gigaflops (billions), teraflops (trillions)
  - Benchmark tests can be used to evaluate overall processing speed

CPU Cores

- **Multi-core CPUs** contain the processing components (cores) of multiple independent processors in a single CPU
- **Dual-core CPUs** contain two cores
- **Quad-core CPUs** contains four cores
- Multi-core processors allow computers to work on more than one task at a time
  - They also typically use slower cores than single-core CPUS so have fewer heat problems

The GPU

- The **GPU (graphics processing unit)** takes care of the processing needed to display images (including still images, animations) on the screen
  - Can be located on the motherboard, on a video graphics board, or in the CPU package
  - Mobile processors often integrate other capabilities into the processor package (system-on-a-chip (SoC))

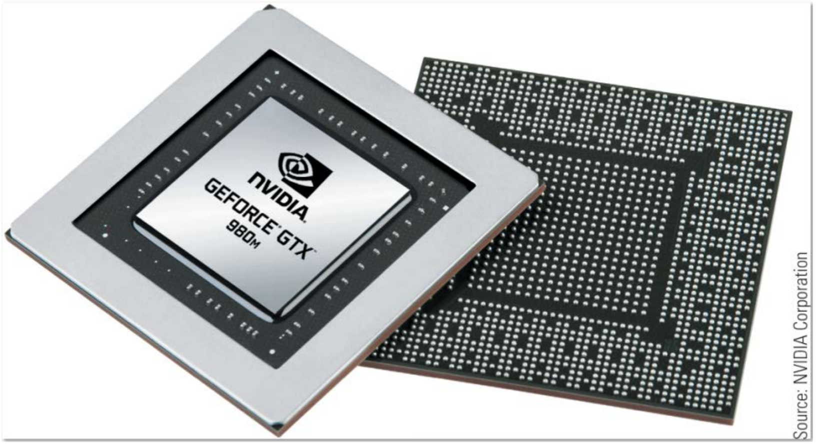


FIGURE 2-10
A GPU.

Word Size

- A computer word is the amount of data that a CPU can manipulate at one time
  - In the past, CPUs used 32-bit words (referred to as 32-bit processors); today, most CPUs are 64-bit processors
  - Word size corresponds to memory to CPU bus width

Bus Width, Bus Speed, and Bandwidth

- A **bus** is an electronic path over which data can travel
  - Found inside the CPU and on the motherboard
  - Bus width is the number of wires in the bus over which data can travel
    - A wider bus allows more data to be transferred at one time
- Bus width and bus speed together determine the bus's **bandwidth** (the amount of data that can be transferred via the bus in a given time period)
- The amount of data actual transferred under real-life conditions is called **throughput**

Example of Bus Width

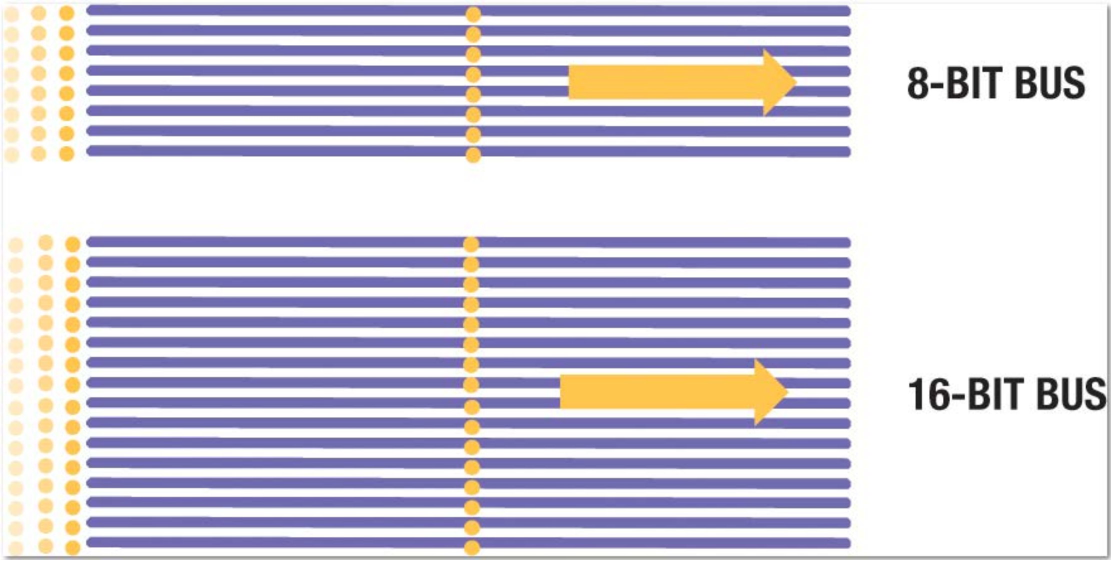


FIGURE 2-11

Bus width. A wider bus can transfer more data at one time than a narrower bus.

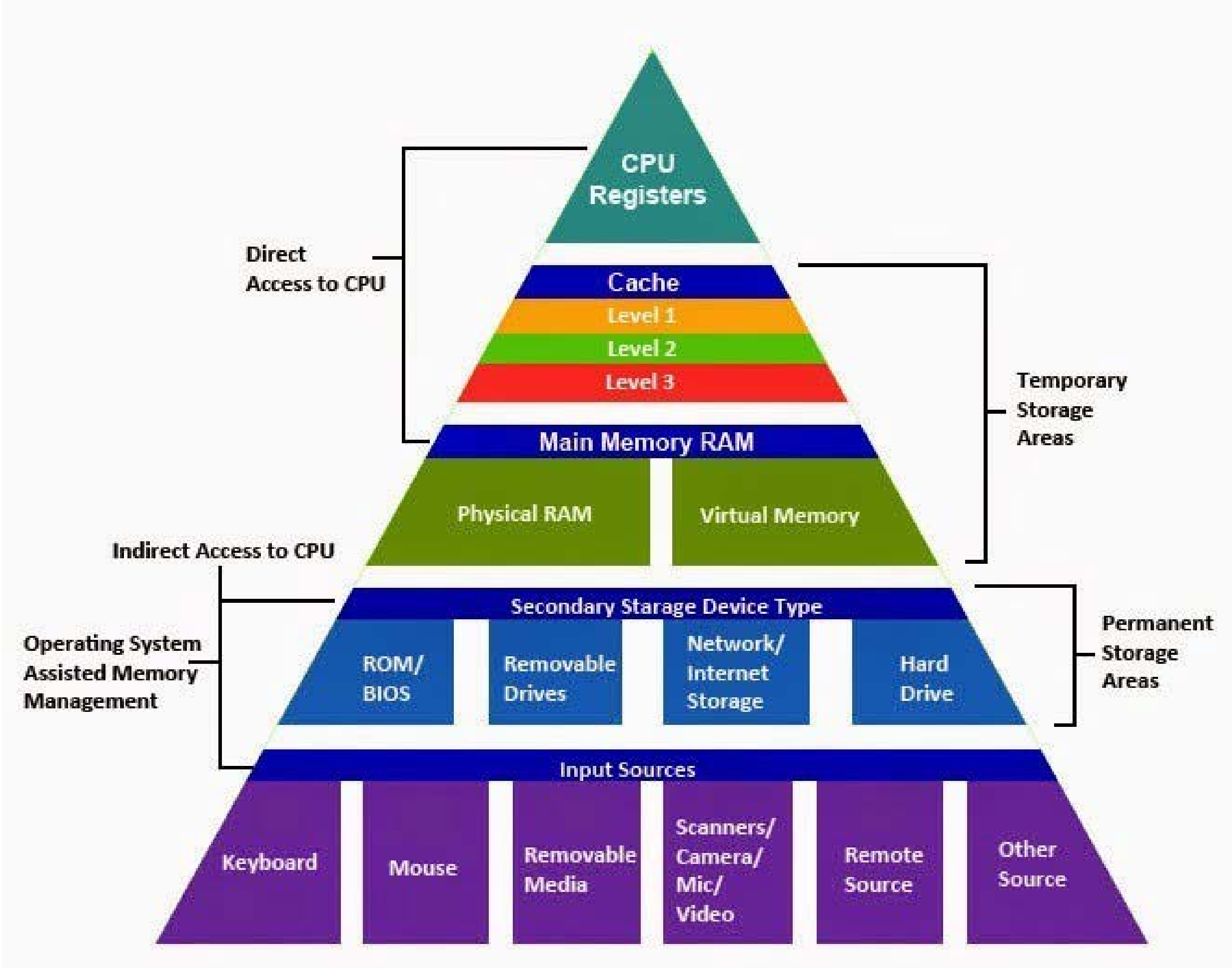
Buses Continued

- An **expansion bus** connects the CPU to peripheral (typically input and output) devices
- The **memory bus** connects the CPU directly to RAM
- The **frontside bus (FSB)** connects the CPU to the chipset that connects the CPU to the rest of the bus architecture

PCI, PCIe, and USB

- The PCI bus used to be the most common type of expansion bus
- Today, **PCI Express (PCIe)** buses are more common
  - PCIe x16 is a 16-bit bus and is used to connect monitors to a computer
  - PCIe x1 is a 1-bit bus and is used to connect other peripherals
  - PCIe buses are extremely fast
- A **universal serial bus (USB)** connects USB devices to a computer
  - 127 different devices can connect via a single USB port
  - Extremely versatile

Memory Hierachy



We will return to this chart below

Registers and ROM

- **Registers** are high-speed memory locations built into the CPU
  - Used to store data and intermediary results during processing
  - Fastest type of memory
- **ROM (read-only memory)** consists of non-volatile chips located on the motherboard into which data or programs permanently stored (boot process & UEFI Boot)
  - Retrieved by the computer when needed
  - Being replaced with flash memory

Cache Memory

- **Cache memory** is a special group of very fast circuitry usually built into the CPU (internal cache memory)
  - More cache memory typically means faster processing
  - Cache memory level numbers indicate the order in which the various levels of cache are accessed by the CPU
    - Level 1 is fastest, then Level 2, then Level 3

Memory

- **Memory** refers to chip-based storage, or locations that a computer uses to store data on a temporary basis
  - **Volatile** memory (content is erased when the device is shut off)
  - **Non-volatile** memory (content is retained when the device is shut off)
- **Random access memory (RAM)** is the computer's main memory or system memory
  - Stores essential parts of operating system, programs, and data the computer is currently using
  - Consists of electronic circuits etched onto chips
    - Mobile devices typically use embedded memory chips
    - Servers and personal computers use circuit boards called memory modules plugged into the motherboard

Memory Addressing

- Each location in memory has an address
  - Usually stored in one or more consecutive addresses, depending on its size
- Computer system sets up and maintains directory tables to facilitate retrieval of the data
- What's wrong with this image => zero based addressing

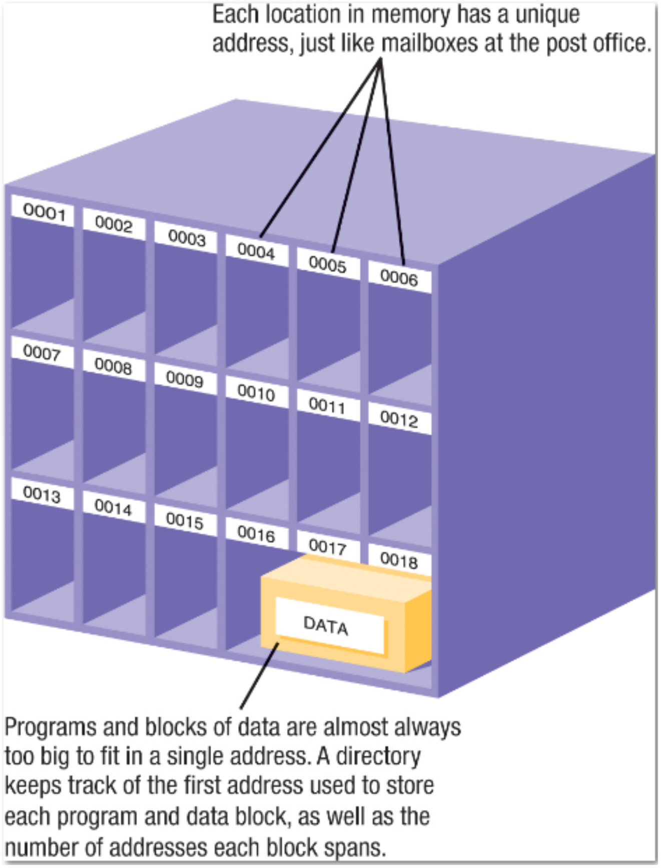


FIGURE 2-13
Memory addressing.

Characteristics of RAM

- Volatile
- Measured in bytes (amount dependent on CPU, operating system => address space)
- Most personal computers use SDRAM (synchronous dynamic)
- Double-Data Rate (DDR) RAM sends data twice as often as ordinary SDRAM or prior versions of RAM
  - DDR2, DDR3, DDR4
- Dual-channel memory architecture has two paths that go to and from memory; tri-channel (three paths) and quad-channel (four paths) memory architecture used for higher performance

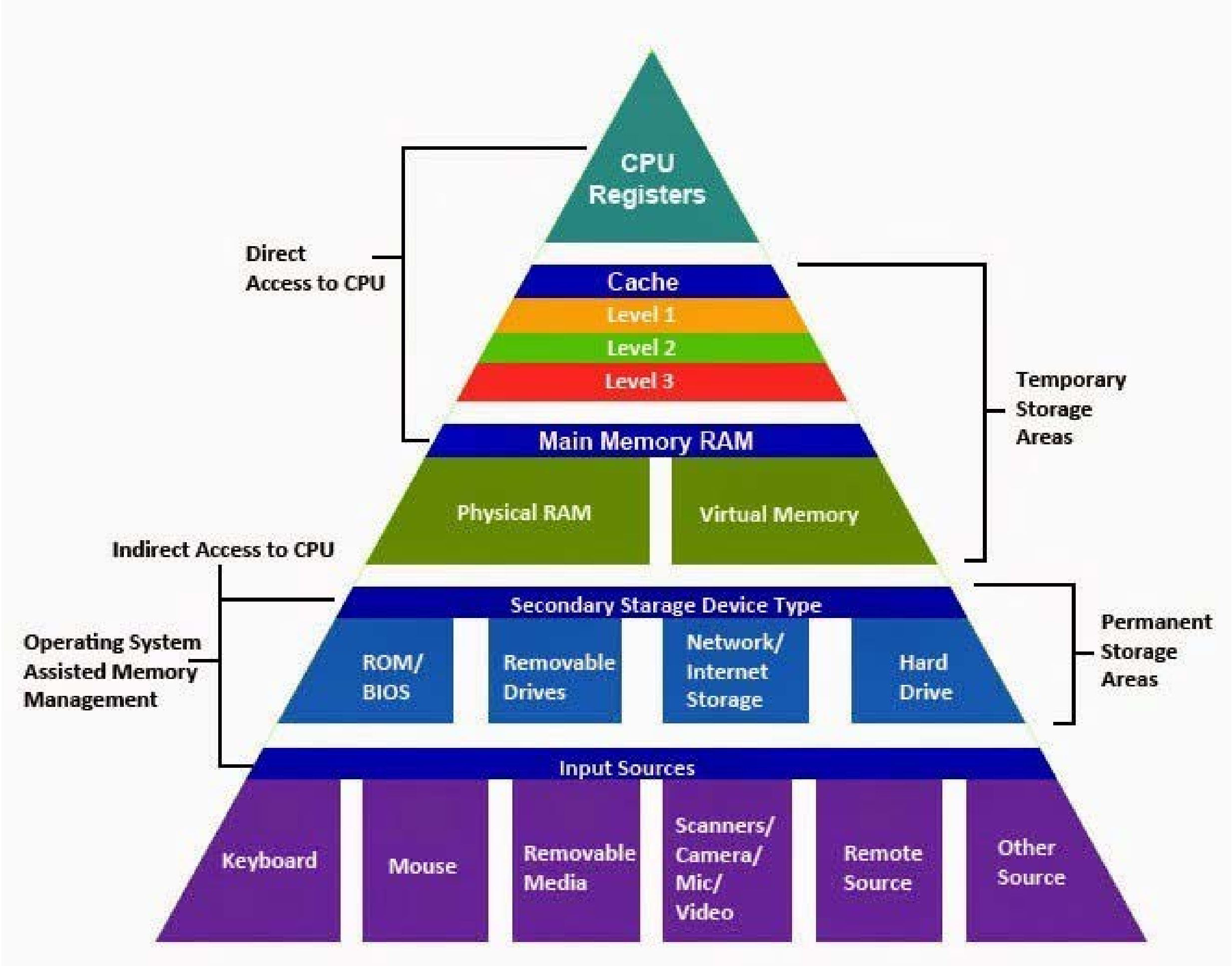
Flash Memory

- **Flash memory** consists of nonvolatile memory chips that can be used for storage
  - Have begun to replace ROM for storing system information (BIOS)
  - Stores firmware for personal computers and other devices
  - Built into many types of devices (tablets, smartphones, and digital cameras) for user storage
  - Built into some storage devices (solid-state hard drives, USB flash drives, etc.)

Solid State Storage (not in text)

- NAND (Not And) (ex. USB Drives)
  - Inexpensive, High Capacity, High Write Speed, Medium Read Speed, Difficult Code Execution, Low Power Requirements
- NOR (Not Or)
  - Expensive, Lower Capacity, Low Write Speed, High Read Speed, Easy Code Execution, High Power Requirements

Memory Hierachy



See ciss100.com LM2 for Memory Hierachy presentation

Fans, Heat Sinks, and Other Cooling Components

- Fans are used on most personal computers to help cool the CPU and system unit
  - Heat is an ongoing problem for CPU and computer manufacturers
    - Can damage components
    - Cooler chips run faster
- Heat sinks are small components typically made out of aluminum with fins that help to dissipate heat
- Some portable computers and virtually all mobile devices don’t include a fan; instead thermal transfer materials are used to spread out the heat generated

Other Types of Cooling Systems

- Liquid cooling systems
  - Cool the computer with liquid-filled tubes
- Immersion cooling
  - Hardware is actually submerged into units filled with a liquid cooling solution
- Notebook cooling stand
  - Cools the underside of a notebook computer
- Other cooling methods, such as ion pump cooling systems, are under development

Expansion Slots and Expansion Cards

- An **expansion slot** is a location on the motherboard into which expansion cards are inserted
- An **expansion card** is a circuit board inserted into an expansion slot
  - Used to add additional functionality or to attach a peripheral device
- Smaller devices may integrate capabilities directly into the device
- USB adapters can be used with portable computers and some mobile devices

Ports and Connectors

- A **port** is a connector on the exterior of a computer's system unit to which a device may be attached
  - Typical desktop computer ports HDMI to connect a monitor (VGA and Digital Video Interface (DVI) are older standards)
  - Network ports connect a device to a wired network
  - USB ports connect USB devices; can be USB-C
  - Others include IrDA and Bluetooth ports, flash memory card slots, audio ports, eSATA ports, and Thunderbolt ports (Apple)
  - Most computers support the Plug and Play standard
  - USB and Thunderbolt devices are hot-swappable

How the CPU Works

- The CPU (central processing unit) consists of a variety of circuitry and components packaged together
  - The transistor is the key element of the microprocessor
    - Made of semi-conductor material that controls the flow of electrons inside a chip
    - Today's CPUs contain hundreds of millions of transistors; the number doubles about every 18 months (Moore's Law)
  - Electronic impulses move from one part of the CPU to another to process data
  - The architecture and components included in a CPU (referred to as microarchitecture) vary from processor to processor

CPU Core Components

- The **arithmetic/logic unit (ALU)** performs arithmetic involving integers and logical operations
- The **floating point unit (FPU)** performs decimal arithmetic
- The **control unit** coordinates and controls activities within a CPU core
- The **prefetch unit** attempts to retrieve data and instructions before they are needed for processing in order to avoid delays

CPU Core Components (cont'd)

- The **decode unit** translates instructions from the prefetch unit so that they are understood by the control unit, ALU, and FPU
- The registers and internal cache memory store data and instructions needed by the CPU
- The **bus interface unit** allows the core to communicate with other CPU components

A CPU Core

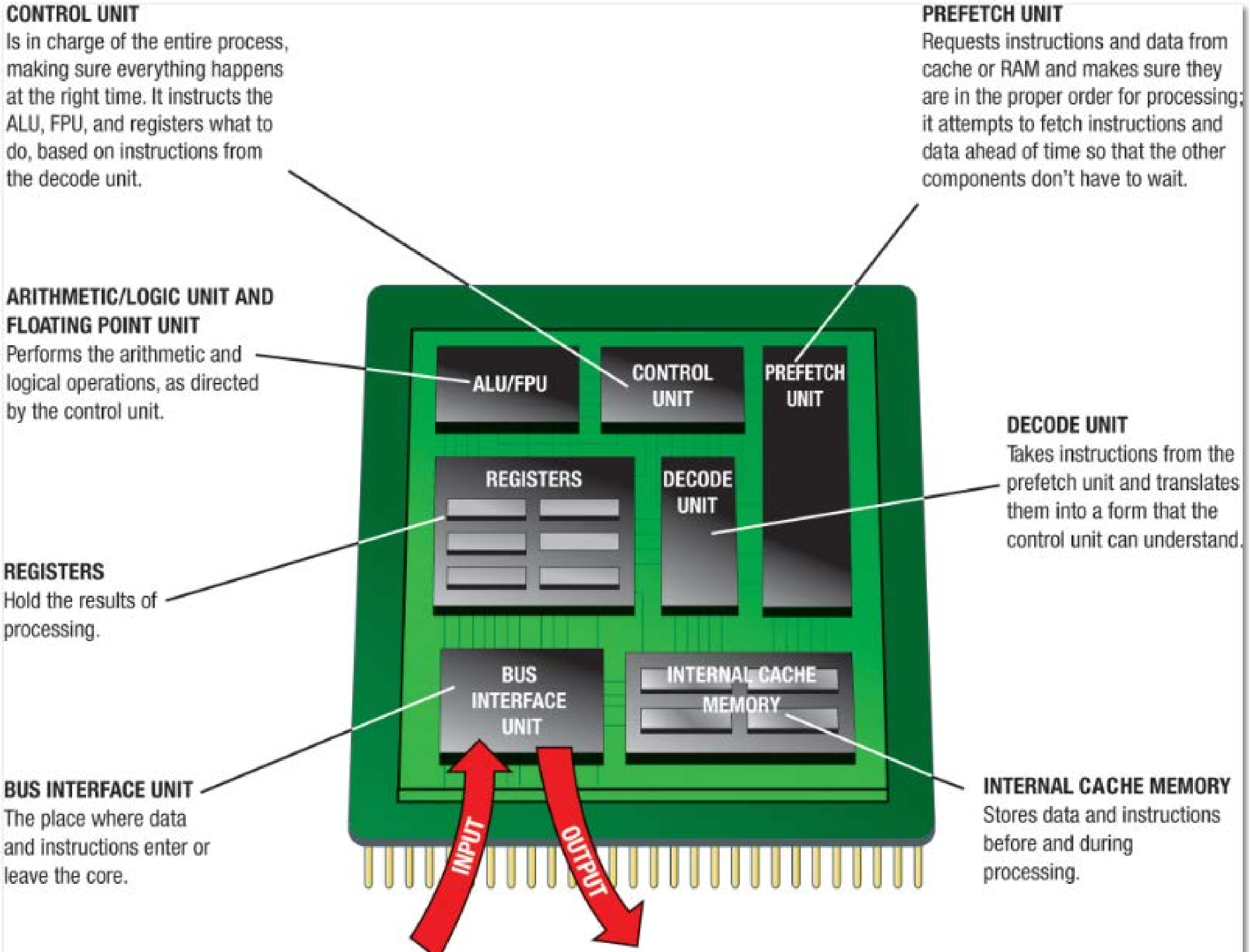


FIGURE 2-21
Inside a CPU core.

The Machine (Fetch-execute) Cycle

- A **machine cycle** occurs whenever the CPU processes a single piece of microcode
  - It consists of four operations:
    - Fetch
    - Decode
    - Execute
    - Store

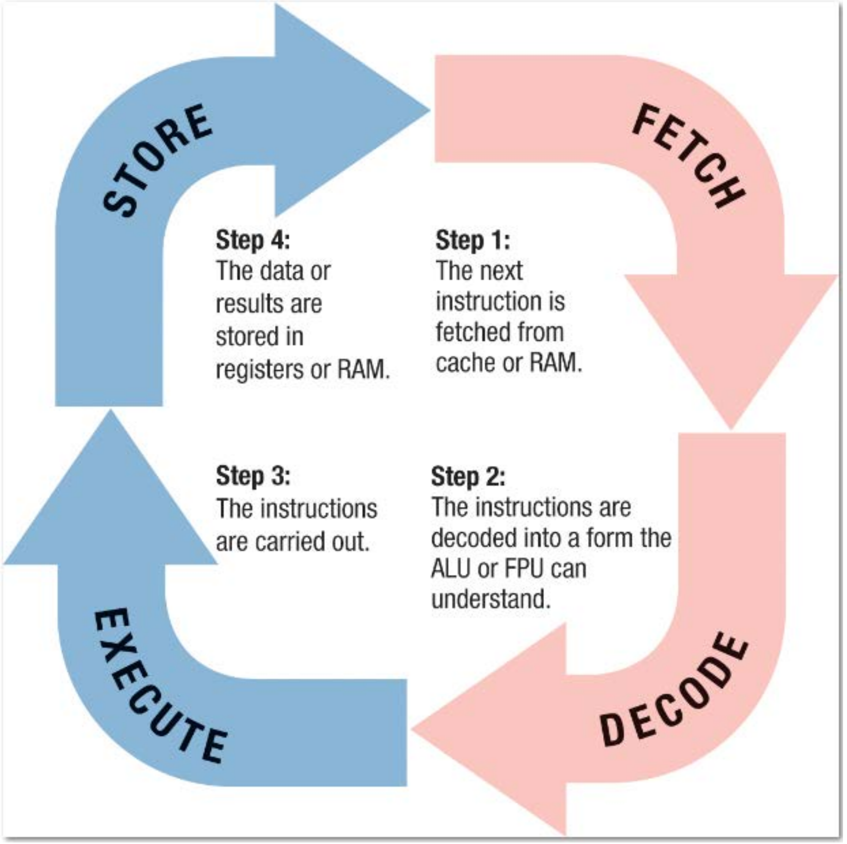


FIGURE 2-22

A machine cycle.
A machine cycle is typically accomplished in four steps.

The System Clock

- The **system clock** is a timing mechanism within the computer system that synchronizes the computer's operations
  - Located on the motherboard
  - Sends out a signal on a regular basis to all computer components
  - Each signal is a cycle
  - Number of cycles per second is measured in hertz (Hz)
  - One megahertz = one million ticks of the system clock

Clock Speed

- Computers can run at a multiple or fraction of the system clock speed
  - Many PC system clocks run at 200 MHz; all devices run at a fraction or multiplier of the clock speed
  - A CPU clock speed of 2 GHz means the CPU clock “ticks” 10 times during each system clock tick
  - During each CPU clock tick, one or more pieces of microcode are processed
  - A CPU with a higher clock speed processes more instructions per second than the same CPU with a lower CPU clock speed

Strategies for Making Faster and Better Computers: Pipelining

- **Pipelining** allows multiple instructions to be processed at one time
  - A new instruction begins as soon as the previous instruction completes a stage of the machine cycle

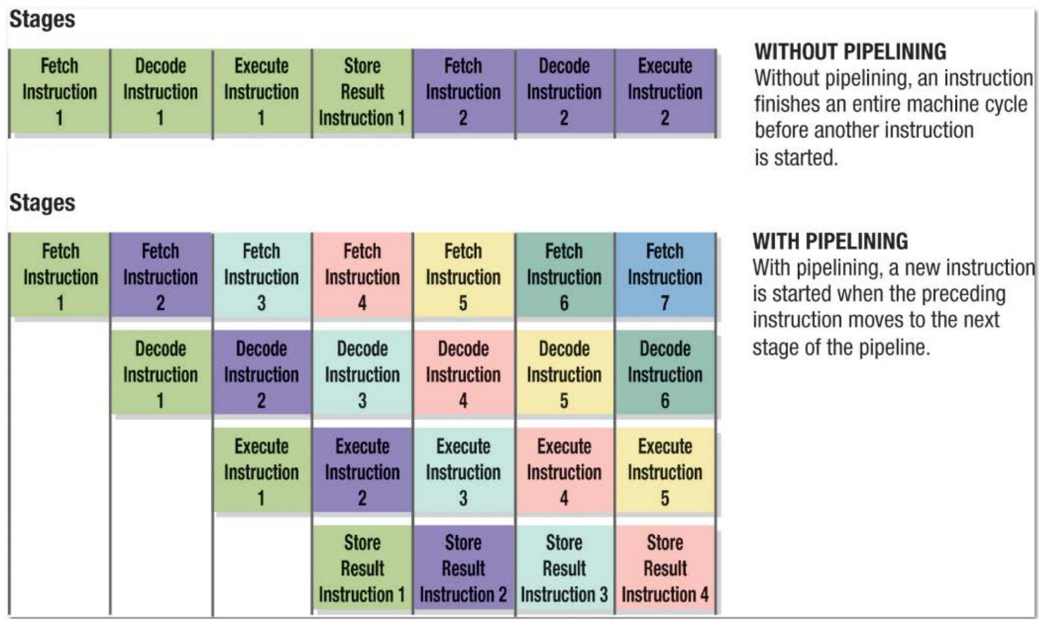


FIGURE 2-25

Pipelining. Pipelining streamlines the machine cycle by executing different stages of multiple instructions at the same time so that the different parts of the CPU are idle less often.

Multiprocessing and Parallel Processing

- Using more than one processor or processor core is common today
  - **Multiprocessing:** Each processor or core typically works on a different job to process these jobs faster than with a single processor
    - 1 processor for executing Word Processor
    - 1 processor to play mp3
  - **Parallel processing**: Multiple processors or cores work together to process a single job as fast as possible
  - Multithreading: the ability of a CPU (or software) to execute multiple streams of instructions (called threads) within a single program at the same time

Sequential vs. Simultaneous Processing (from LM5)

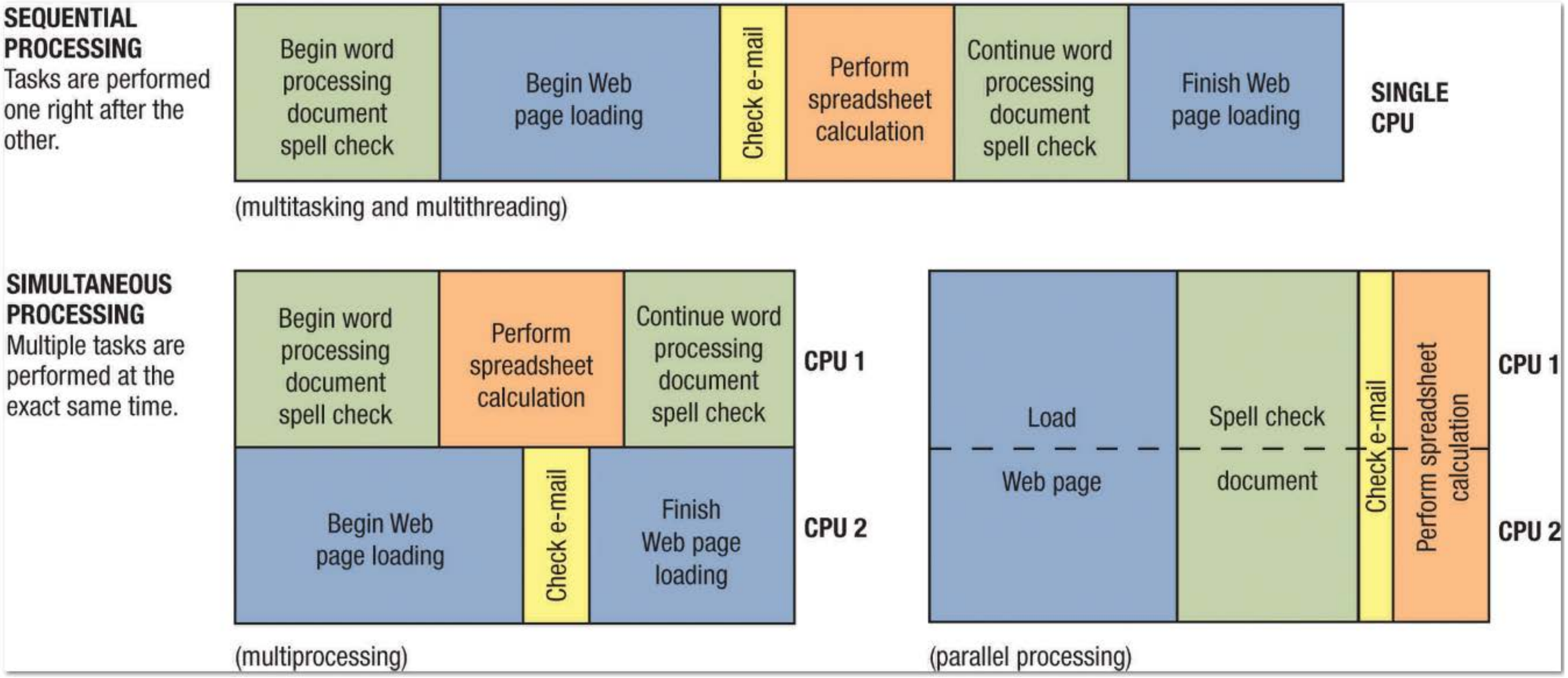


FIGURE 5-7
Sequential vs. simultaneous processing.

Improved Architecture and Materials

- Improved architecture
  - Smaller components, faster memory, faster bus speeds, increasing number of CPU cores, integrated GPUs, support for virtualization, and increased 3D graphics processing
- Improved materials
  - Traditionally, CPUs used aluminum circuitry on silicon chips
  - Alternate materials include copper chips, and high-k, germanium and other III-V materials
  - **Graphene** consists of flat sheets are carbon one atom tall
    - Lightest and strongest known material
    - Graphene chips are faster than silicon chips and require less power

Nanotechnology

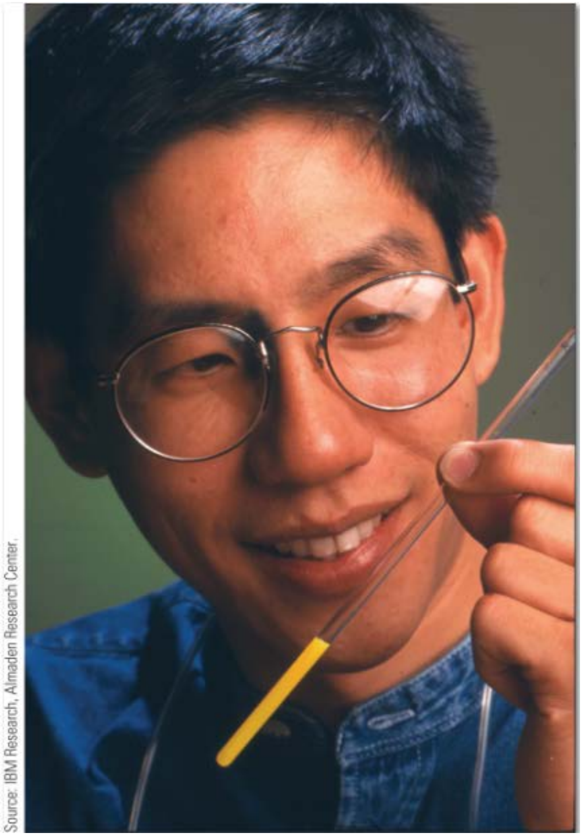
- **Nanotechnology** is the science of creating tiny computers and components less than 100 nanometers in size
  - **Carbon nanotubes (CNTs)** are tiny, hollow tubes of graphene
  - **Graphene** consists of flat sheets are carbon one atom tall
    - Lightest and strongest known material
    - Graphene chips are faster than silicon chips and require less power

Optical Computing and Silicon Photonics

- **Optical computers** use light to perform digital computations
  - Can be smaller and faster than electronic computers
  - Opto-electronic computers use both optical and electronic components
- Silicon photonics uses light for data transfers within and among silicon chips
  - Expected to be used to transfer very large quantities of data at very high speeds between chips in servers, mainframes, and supercomputers

Quantum Computing

- **Quantum computing** applies the principles of quantum physics and quantum mechanics to computers
  - Utilizes atoms or nuclei working together as quantum bits (qubits)
  - Qubits function simultaneously as the computer's processor and memory and can represent more than two states
  - Used for specialized applications, such as encryption and code breaking



Source: IBM Research, Almaden Research Center.

FIGURE 2-30

Quantum computers. This vial of liquid contains a 7-qubit computer.