


Computer Organization and Application

Lecture 5
Introduction to Memory unit, Memory Organization & classification

Dr Mohammed Fadhil
Email: mohammed.fadhil1@uomus.edu.iq

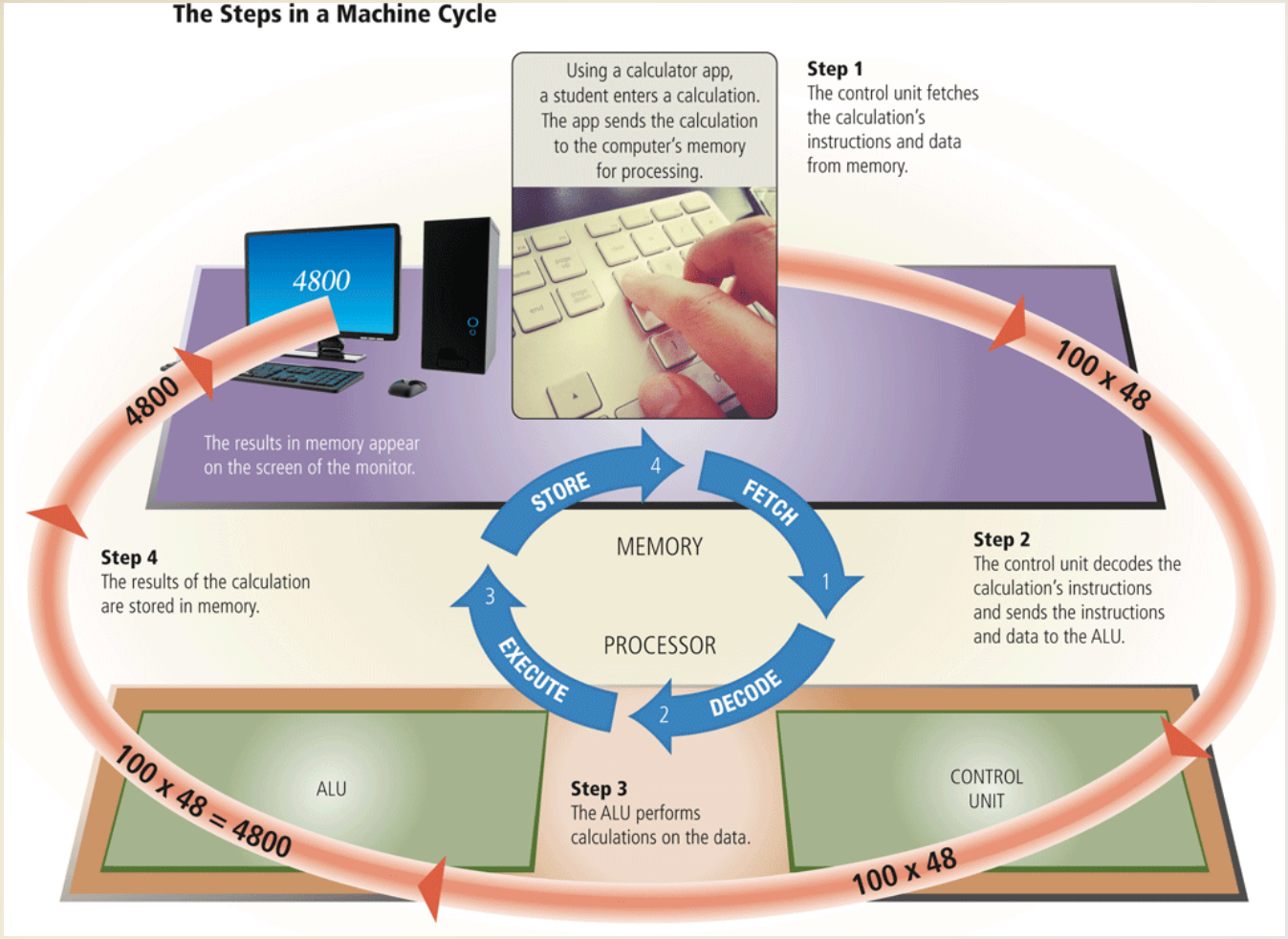
Objectives Overview

- Define a bit, and describe how a series of bits represents data
- Explain how program and application instructions transfer in and out of memory
- Understand the memory unit
- Differentiate among the various types of memory
- Overview of memory access times

Recall: Machine Cycle

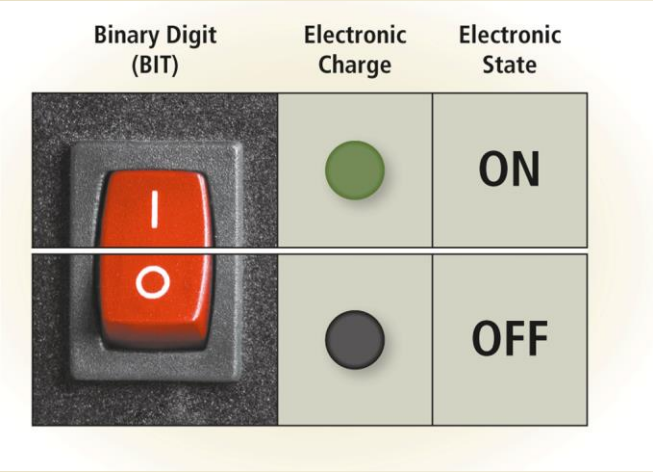
- A processor follows a machine cycle for each instruction, involving four basic steps:
  1. **Fetching:** Gets the instruction or data from memory.
  2. **Decoding:** Translates the instruction into signals the computer understands.
  3. **Executing:** Carries out the instruction's commands.
  4. **Storing** (if needed): Saves the result back to memory (not to an external storage device).

Recall: Machine Cycle

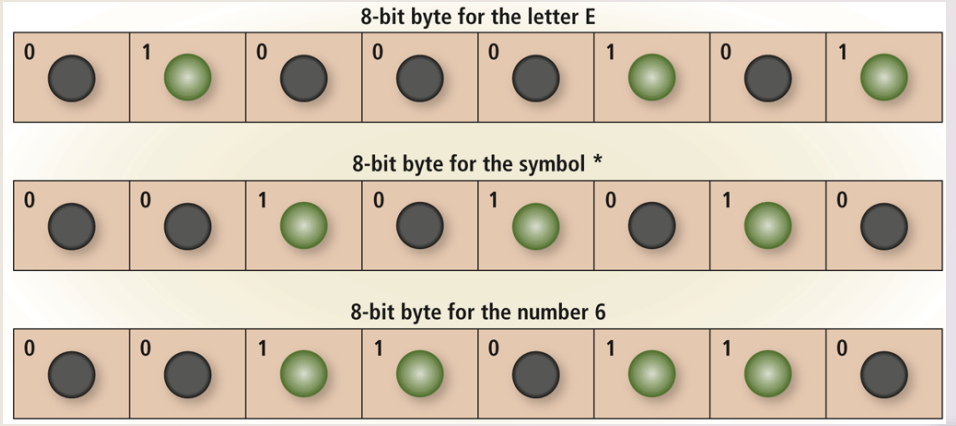


Recall: Data Representation

- The circuitry in a computer or mobile device represents the **on** **or the** **off** **states** electronically by the presence or absence of an electronic charge
  - **On** is represented by the digit 1 (presence of an electronic charge).
  - **Off** is represented by the digit 0 (absence of an electronic charge).



- A single binary digit, or **bit**, is the smallest unit of data a computer can process.
- **Eight bits** grouped together as a **unit** are called a **byte**. **A byte** represents 256 individual characters (letters, numbers, symbols) in the computer or mobile device



Data Representation

- **Coding schemes** define how combinations of 0s and 1s represent different characters:
  - Example: **00110100** represents the number **4**.
  - The widely-used coding scheme **ASCII** ***(American Standard Code for Information Interchange)*** assigns codes to characters, making human-computer interaction possible.
- **Data Conversion Process:**
  - When a key is pressed on a keyboard, it is converted into a binary code by a chip.
  - The system unit then processes this binary code in memory as a series of bytes.
  - After processing, software converts the bytes back into human-readable characters for display on the screen.

ASCII	SYMBOL	ASCII	SYMBOL
00110000	0	01001110	N
00110001	1	01001111	O
00110010	2	01010000	P
00110011	3	01010001	Q
00110100	4	01010010	R
00110101	5	01010011	S
00110110	6	01010100	T
00110111	7	01010101	U
00111000	8	01010110	V
00111001	9	01010111	W
01000001	A	01011000	X
01000010	B	01011001	Y
01000011	C	01011010	Z
01000100	D	00100001	!
01000101	E	00100010	
01000110	F	00100011	#
01000111	G	00100100	$
01001000	H	00100101	%
01001001	I	00100110	&
01001010	J	00101000	(
01001011	K	00101001	)
01001100	L	00101010	*
01001101	M	00101011	+

How a Letter Is Converted to Binary Form and Back

Step 1

A user presses the capital letter **T** (SHIFT+T keys) on the keyboard, which in turn creates a special code, called a scan code, for the capital letter **T**.

Step 2

The scan code for the capital letter **T** is sent to the system unit.

Step 3

The system unit converts the scan code for the capital letter **T** to its ASCII binary code (01010100) and stores it in memory for processing.

Step 4

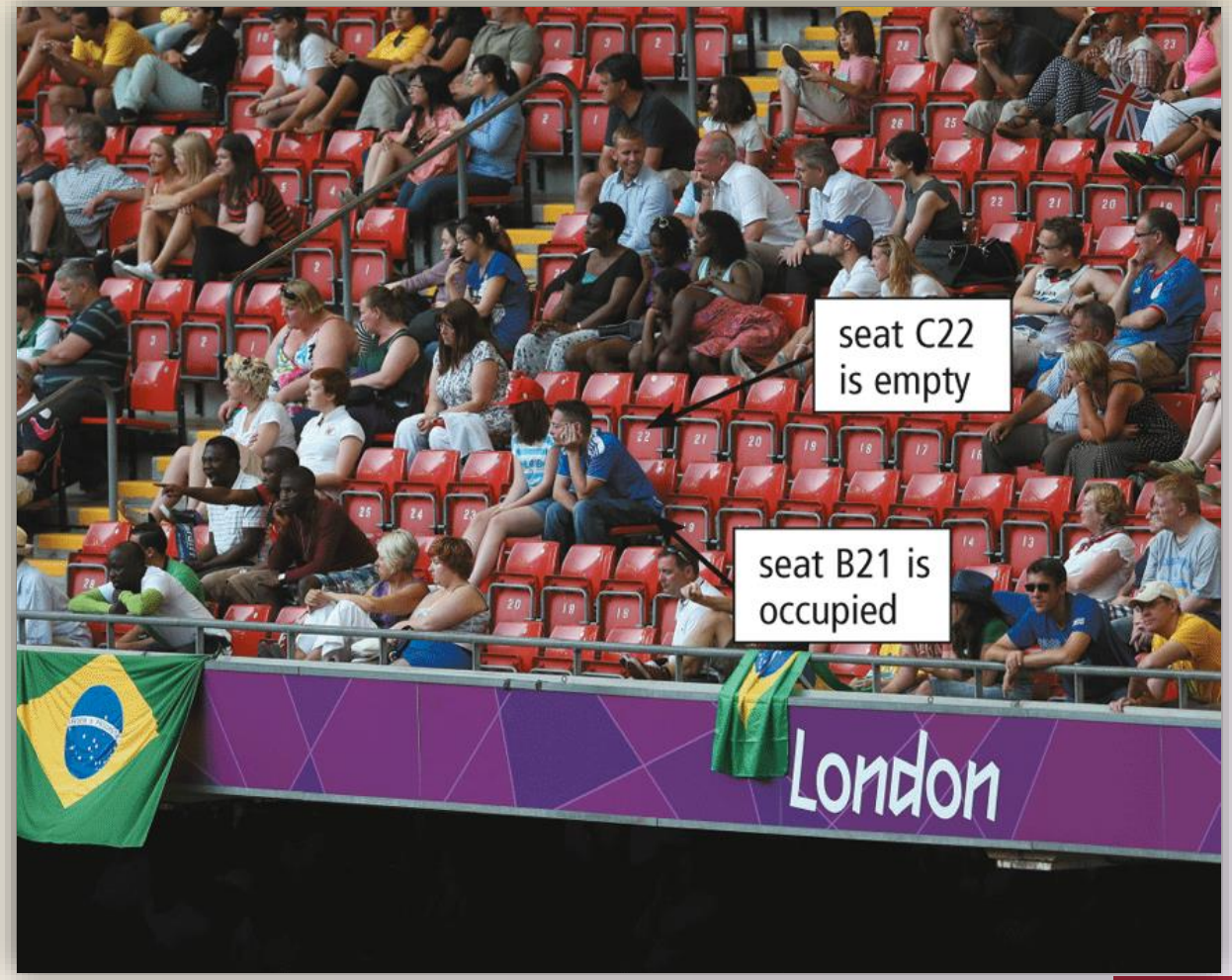
After processing, the binary code for the capital letter **T** is converted to an image and displayed on the output device.

Memory

- **Memory** consists of electronic components that store instructions waiting to be executed by the processor, data needed by those instructions, and the results of processing the data (information).
- **Memory usually consists of one or more chips** on the motherboard or some other circuit board in the computer.
- Memory stores **three basic categories of items**:
  1. **the operating system and other system software** that control or maintain the computer and its devices;
  2. **application programs** that carry out a specific task such as word processing;
  3. **the data** being processed by the application programs and **resulting information**.
- This role of memory to store both data and programs is known as the **stored program concept.**

Bytes and Addressable Memory

- **A byte (character)** is the basic unit of storage in memory.
- When data and instructions are loaded from storage devices, they are stored in memory as bytes.
- **Each byte is stored temporarily** in a **memory location with a unique address.**
- An **address** is a unique number **identifying the location of each byte in memory.**
- Seats in a football stadium are similar to addresses in memory:
  - *Each seat, identified by a unique seat number, holds one person at a time.*
  - *Similarly, each memory location, identified by an address, holds a single byte.*
  - *Both a seat and a memory location can be empty.*
- To access data or instructions in memory, the computer references the specific addresses holding these bytes.

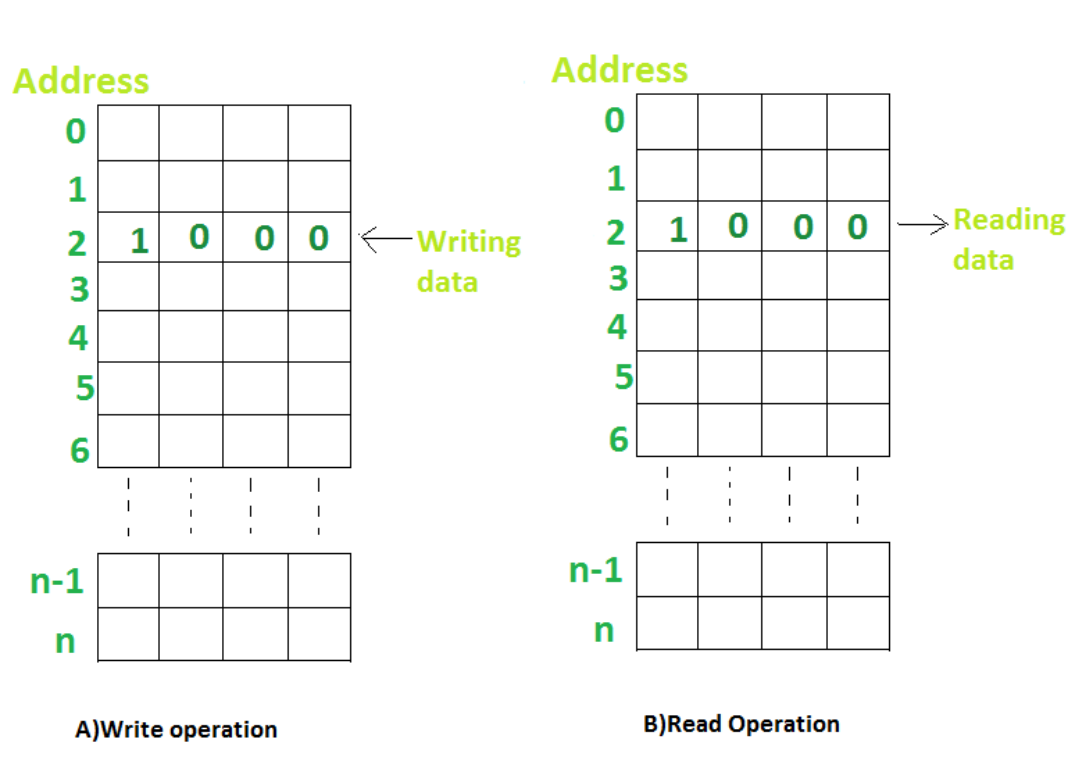


Memory Organization

- **Memories** are made up of **registers**. Each register in the memory is one storage location.
- **Storage location** is also called as **memory location**. Memory locations are identified using **Address**.
- The total number of bit a memory can store is its **capacity**.

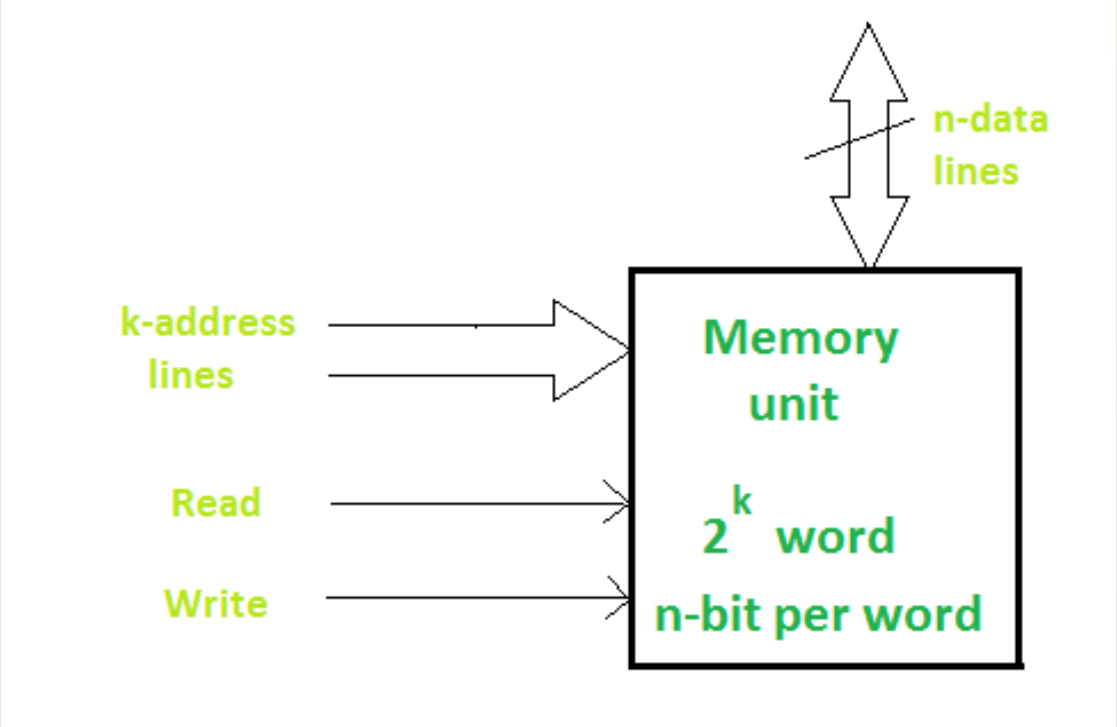
- A storage element is called a **Cell**. Each register is made up of storage element in which one bit of data is stored.
- The data in a memory are **stored** and **retrieved** by the process called **writing** and **reading** respectively.

- A **word** is a **group of bits** where a memory unit stores **binary information**. **A word with group of 8 bits is called a byte.**



Memory Locations and Addresses

- A **memory unit** consists of **data lines**, **address selection lines**, and **control lines** that specify the direction of transfer.
- **Data lines** provide the information to be stored in memory. **The control inputs** specify the direct transfer. The **k-address** lines specify the word chosen.
- When there are **k-address lines**, ***2^k memory words can be accessed.***
- **For example**, a memory with **64k words** and a **word size of 1 byte**, then this memory unit has 64 * 1024 = 65536 memory locations. **The address of these locations varies from 0 to 65535.**



Memory Sizes

- Manufacturers state the size of memory and storage devices in terms of the number of bytes the chip or device has available for storage.
- **Storage devices** hold data, instructions, and information for future use, while **most memory** holds these items ***temporarily***.
- Units of measurement:
  - **Kilobyte (KB)**: 1 KB = 1,024 bytes (often rounded to 1,000 bytes for simplicity).
    - Example: 100 KB ≈ 100,000 bytes (characters).
  - **Megabyte (MB)**: ≈ 1 million bytes.
  - **Gigabyte (GB)**: ≈ 1 billion bytes.
  - **Terabyte (TB)**: ≈ 1 trillion bytes.

Terms commonly used to define memory sizes

Memory Sizes

Term	Abbreviation	Approximate Number of Bytes	Exact Number of Bytes	Approximate Number of Pages of Text
Kilobyte	KB or K	1 thousand	1,024	1/2
Megabyte	MB	1 million	1,048,576	500
Gigabyte	GB	1 billion	1,073,741,824	500,000
Terabyte	TB	1 trillion	1,099,511,627,776	500,000,000

Question?

- **Q/ A computer has 32 MB of memory. How many bits are needed to address any single byte in memory?**
  - Convert 32 MB to bytes:
    - **1 Megabyte (MB) = 2^{20} bytes (or 1,048,576 bytes).**
  - Therefore, 32 MB of memory is:
    - **$32 \times 2^{20} = 2^5 \times 2^{20} = 2^{25}$ bytes**
    - So, the memory contains **2^{25} individual bytes.**
  - Determine the Number of Addressable Units:
    - Since each byte in memory has a unique address, we need to determine how many bits are required to represent 2^{25} unique addresses.
  - Calculate the Number of Bits Needed:
    - In general, to address *N* unique locations, we need $\log_2(N) = bits$.
    - Here, ***N*** **= 2^{25}** , so:
      - $\log_2(2^{25}) = 25$ **bits**
  - This means we need **25 bits** to uniquely address each byte in **32 MB** of memory.

Main Memory Organization

- The number of bits stored in a **register** is called a **memory word**.
- **Memory devices (chips)** are available in various word sizes.
- Each word in memory has a **specific binary number** known as its **memory address.**
- For the CPU to interact with main memory, it must be able to:
  - **Select the chip** (determine which chip to access),
  - **Identify the memory location** (locate the specific address within the chip),
  - **Identify the memory operation** (specify whether to **read** from or **write** to the memory location).
- To perform these functions, a memory chip needs the following terminals:

I. **Address Lines**:
   - Used to identify the number of memory locations within the chip.
   - The number of address lines n determines the number of locations in the chip by the relation 2^n.
   - Example: **If n = 9, then the chip contains 2^9 =512 locations.**

II. **Data Lines**:
   - Represent the data input/output lines in a memory chip.
   - Example: **A 2M x 8 memory has 8 data lines and 21 address lines since:**

$$2M = 2 \times 2^{20} = 2^{21}$$

III. **Control Lines**:
   - Memory chips typically have two control lines:
     1. **R/W line**: Specifies the operation type, whether to **read** or **write**.
     2. **Chip Select (CS) line**: Activates the desired chip in a multi-chip memory system.

Types of Memory

- The system unit contains **two types of memory**:

Volatile memory

- Loses its contents when the computer's power is turned off
- Volatile memory is temporary
- Example includes RAM

Nonvolatile memory

- Does not lose contents when power is removed from the computer
- Nonvolatile memory is permanent
- Examples include ROM, flash memory, and CMOS

RAM

- When people mention **"computer memory,"** they usually mean **RAM (Random Access Memory),** also known as **main memory**.
- RAM is made up of memory chips that the processor and other devices can read from and write to.
- When a computer is powered on, some operating system files (like those for desktop appearance) load into RAM from a storage device, such as a ***hard drive***.
- These files **stay** in RAM as long as the computer has continuous power.
- When more programs and data are needed, they load into RAM from storage as well.

RAM

- The **processor** interprets and executes program instructions while the program is in RAM.
- While running, the contents of RAM may change.
- RAM can hold multiple programs at the same time, as long as there's enough space.
- Most RAM is **volatile**, meaning it loses its contents when power is turned off.
- To keep data for future use, **it must be saved from RAM to a storage device (e.g., a hard disk).**

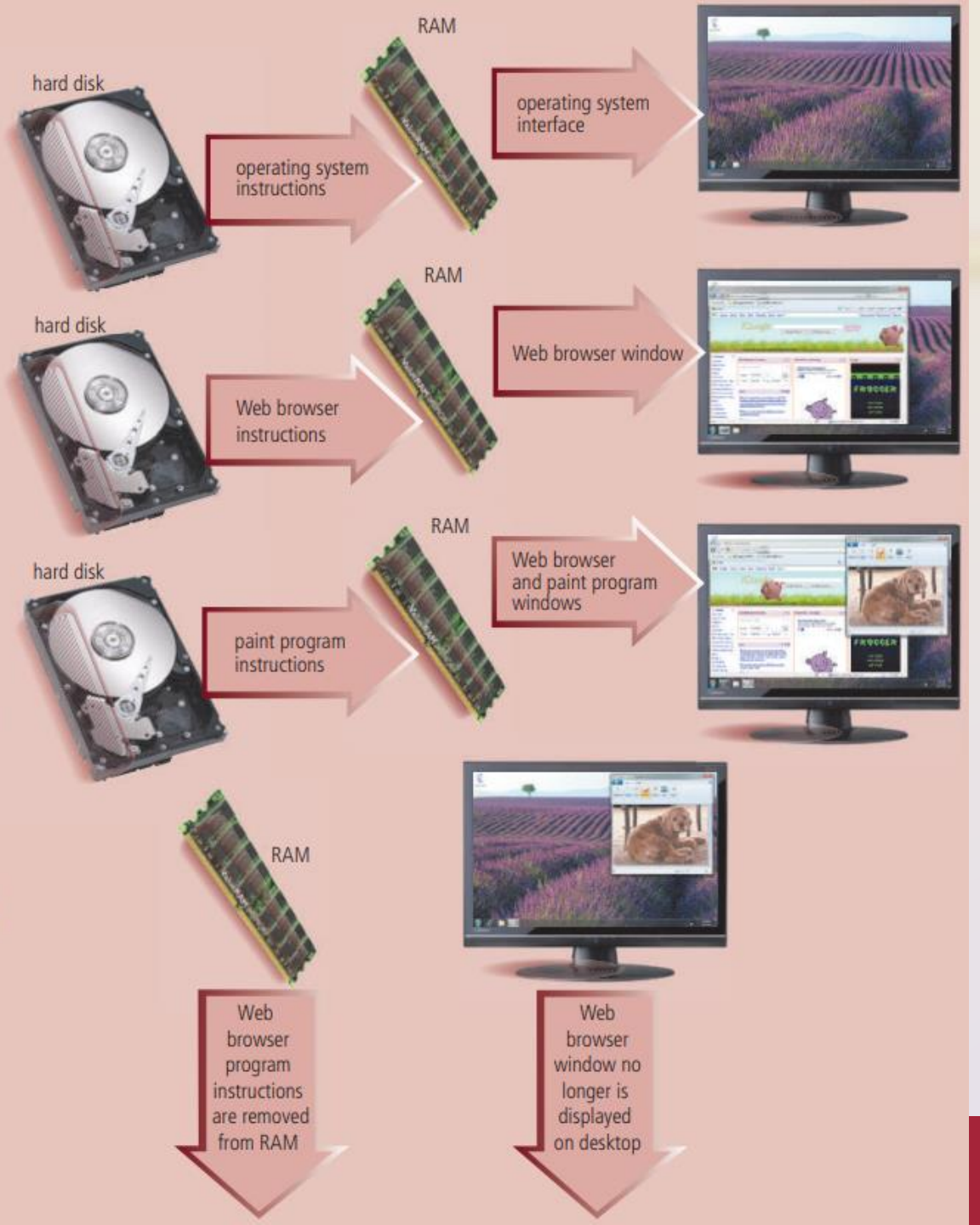
How Program Instructions Transfer in and out of RAM

Step 1
When you start the computer, certain operating system files are loaded into RAM from the hard disk. The operating system displays the user interface on the screen.

Step 2
When you start a Web browser, the program's instructions are loaded into RAM from the hard disk. The Web browser and certain operating system instructions are in RAM. The Web browser window appears on the screen.

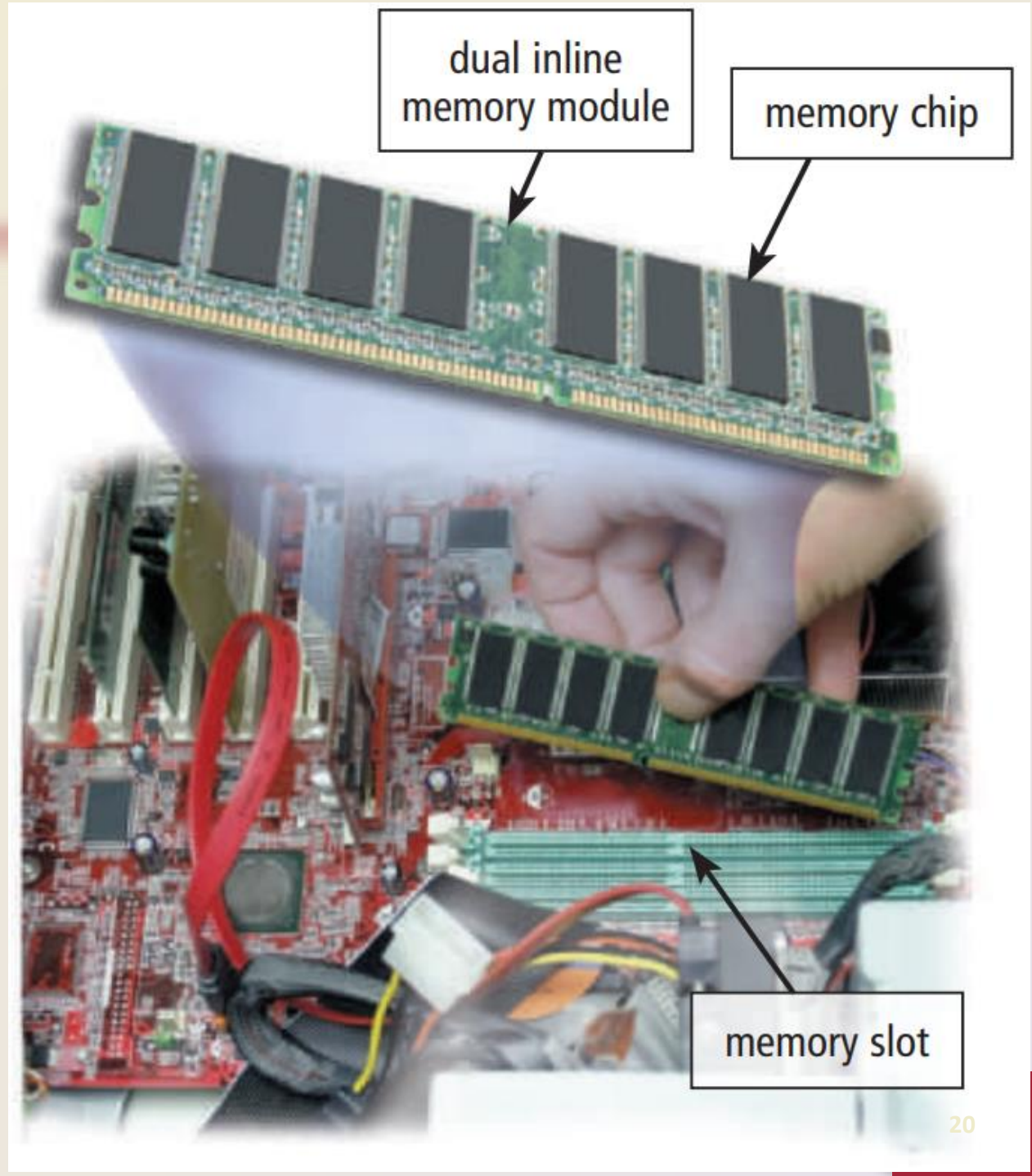
Step 3
When you start a paint program, the program's instructions are loaded into RAM from the hard disk. The paint program, along with the Web browser and certain operating system instructions, are in RAM. The paint program window appears on the screen.

Step 4
When you quit a program, such as the Web browser, its program instructions are removed from RAM. The Web browser no longer is displayed on the screen.



RAM

- **Three basic types of RAM chips exist:**
  - **Dynamic RAM (DRAM):**
    - Must be constantly re-energized or it loses data.
    - Many versions exist, most of which are faster than the basic DRAM.
  - **Static RAM (SRAM):**
    - Faster and more reliable than DRAM.
    - Needs less re-energizing, hence "static."
  - **Magnetoresistive RAM (MRAM):**
    - Uses magnetic charges to store data instead of electrical.
    - Offers more storage, lower power consumption, and faster access than electronic RAM.
- **RAM chips** usually reside on a **memory module**, which is a small circuit board. **Memory slots** on the motherboard hold memory modules



RAM Configurations

- **The amount of RAM** needed depends on the software you plan to use.
- A computer runs programs that are in RAM, **so more RAM means faster performance.**
- Software usually lists a minimum RAM requirement, but more RAM is often needed for optimal performance.
- General recommendations:
  - **Home users** (basic tasks like word processing): at least **1 GB** of RAM.
  - **Business users** (tasks like accounting, spreadsheets, multimedia): **2 to 8 GB** of RAM.
  - **Advanced users** (graphics-intensive tasks or professional web design): **8 GB** or more of RAM.
- Current computers offer RAM options ranging from **1 GB to 128 GB**.

RAM Configurations

- Can I add more RAM to my computer?

Check your computer documentation to see how much RAM you can add.

RAM modules are relatively inexpensive and usually include easy-to-follow installation instructions.

Be sure to purchase RAM that is compatible with your brand and model of computer.

Cache

- **Cache** improves computer processing times by **storing frequently used instructions and data**. Two types of cache are memory cache and disk cache. This lecture discusses **memory cache.**
- **Types of Memory Cache:**
  - **L1 Cache:**
    - **Built directly into the processor chip.**
    - **Small capacity (8 KB to 128 KB).**
  - **L2 Cache:**
    - **Slightly slower than L1 cache.**
    - **Larger capacity (64 KB to 16 MB).**
  - Current processors include **Advanced Transfer Cache:**
    - **A type of L2 cache built on the processor chip.**
    - **Makes processors much faster.**
    - **Commonly has 512 KB to 12 MB in personal computers.**

Cache

- How Cache Works:
  - Cache speeds up processing time because it stores frequently used instructions and data.
  - When the processor needs an instruction or data, it searches memory in this order:
    - **L1 Cache**
    - **L2 Cache**
    - **RAM** (with a greater delay in processing for each level of memory it must search). If the instruction or data is **not found in memory , then;**
    - **Storage** (like a hard disk or optical disc, with a slower speed storage medium).

ROM

- Read-only memory (ROM) refers to memory chips storing permanent data and instructions.
- The data on most ROM chips **cannot be modified** — hence, the name ***read only***.
- **ROM is nonvolatile**, which means its contents are **not lost** when power is removed from the computer.
- **Firmware**: ROM chips that contain permanently written data, instructions, or information pre-recorded by the manufacturer.

Flash Memory

- **Flash memory** is a type of **nonvolatile memory** that can be erased and rewritten electronically.
- Computers use flash memory to **store startup instructions, making it easy to update**.
- Flash memory enables automatic updates, like adjusting for daylight savings time.
- It is commonly used in mobile devices to store data and applications, such as:
  - **Smartphones, Portable media players, PDAs (Personal Digital Assistants), Printers, Digital cameras, Automotive devices, Digital voice recorders, Pagers,**
- Some **portable media players** use flash memory for **music storage**, while others use **tiny hard disks or flash memory cards.**
- **Flash memory cards** are **removable devices** that contain flash memory on a **card** rather than on a **chip**.

Flash Memory

- **How much music can I store on a portable media player?**

Portable media players that store music on flash memory chips can hold up to 16,000 songs. Portable media players with tiny hard disks have a much greater storage capacity — from 1,000 to more than 80,000 songs.

CMOS

- Some memory chips, like RAM and flash memory, use **CMOS (Complementary Metal-Oxide Semiconductor) technology.**
- CMOS is popular because **it offers high speed and low power consumption.**
- **CMOS memory** can **retain information using battery power, even when the computer is turned off.**
- Battery-backed CMOS memory keeps **the calendar, date, and time up to date when the computer is off.**
- Flash memory chips that store startup information also often use **CMOS** technology.

Memory Access Times

- **Access Time**: The time the processor takes to read data, instructions, and information from memory.
- **Impact on Performance**: Faster access time leads to faster data processing by the computer.
- **Memory vs. Hard Disk:** Memory access is over 200,000 times faster than hard disk access due to the hard disk's mechanical parts.
- **Measurement Units:**
  - **Often measured in nanoseconds (ns), where 1 ns = one billionth of a second.**
  - **Some manufacturers use MHz (e.g., 800 MHz RAM) to indicate access speeds.**
- **Focus on Memory Size**: Despite its importance, manufacturers usually highlight memory size (capacity) rather than access time when advertising computers.

Access Time Terminology

Term	Abbreviation	Speed
Millisecond	ms	One-thousandth of a second
Microsecond	µs	One-millionth of a second
Nanosecond	ns	One-billionth of a second
Picosecond	ps	One-trillionth of a second

10 million operations = 1 blink

QUIZ YOURSELF

i. **What is memory unit in CPU?**
ii. **Which is the smallest unit of measurement of data?**
iii. **What is the role of cache memory?**
iv. **What is memory bottleneck? [SEARCH]**
v. **A computer has 128 MB of memory. Each word in this computer is eight bytes. How many bits are needed to address any single word in memory?**
vi. **Find the true statement below. Then, rewrite the remaining false statements so that they are true.**

1. A computer's memory access time directly affects how fast the computer processes data.
2. A gigabyte (GB) equals approximately 1 trillion bytes.
3. Memory cache helps speed the processes of the computer because it stores seldom used instructions and data.
4. Most computers are analog, which means they recognize only two discrete states: on and off.
5. Most RAM retains its contents when the power is removed from the computer.
6. Read-only memory (ROM) refers to memory chips storing temporary data and instructions.

THANK YOU ☺