THE ESSENTIALS OF

Computer Organization and Architecture

THIRD EDITION

Linda Null
Julia Lobur

Chapter 1
Introduction

Chapter 1 Objectives

- Know the difference between computer organization and computer architecture.
- Understand units of measure common to computer systems.
- Appreciate the evolution of computers.
- Understand the computer as a layered system.
- Be able to explain the von Neumann architecture and the function of basic computer components.

1.1 Overview

Why study computer organization and architecture?

- Design better programs, including system software such as compilers, operating systems, and device drivers.
- Optimize program behavior.
- Evaluate (benchmark) computer system performance.
- Understand time, space, and price tradeoffs.

1.1 Overview

- Computer organization
  - Encompasses all physical aspects of computer systems.
  - E.g., circuit design, control signals, memory types.
  - *How does a computer work?*
- Computer architecture
  - Logical aspects of system implementation as seen by the programmer.
  - E.g., instruction sets, instruction formats, data types, addressing modes.
  - *How do I design a computer?*

1.2 Computer Components

- There is no clear distinction between matters related to computer organization and matters relevant to computer architecture.
- Principle of Equivalence of Hardware and Software:
  - ***Any task done by software can also be done using hardware, and any operation performed directly by hardware can be done using software.****

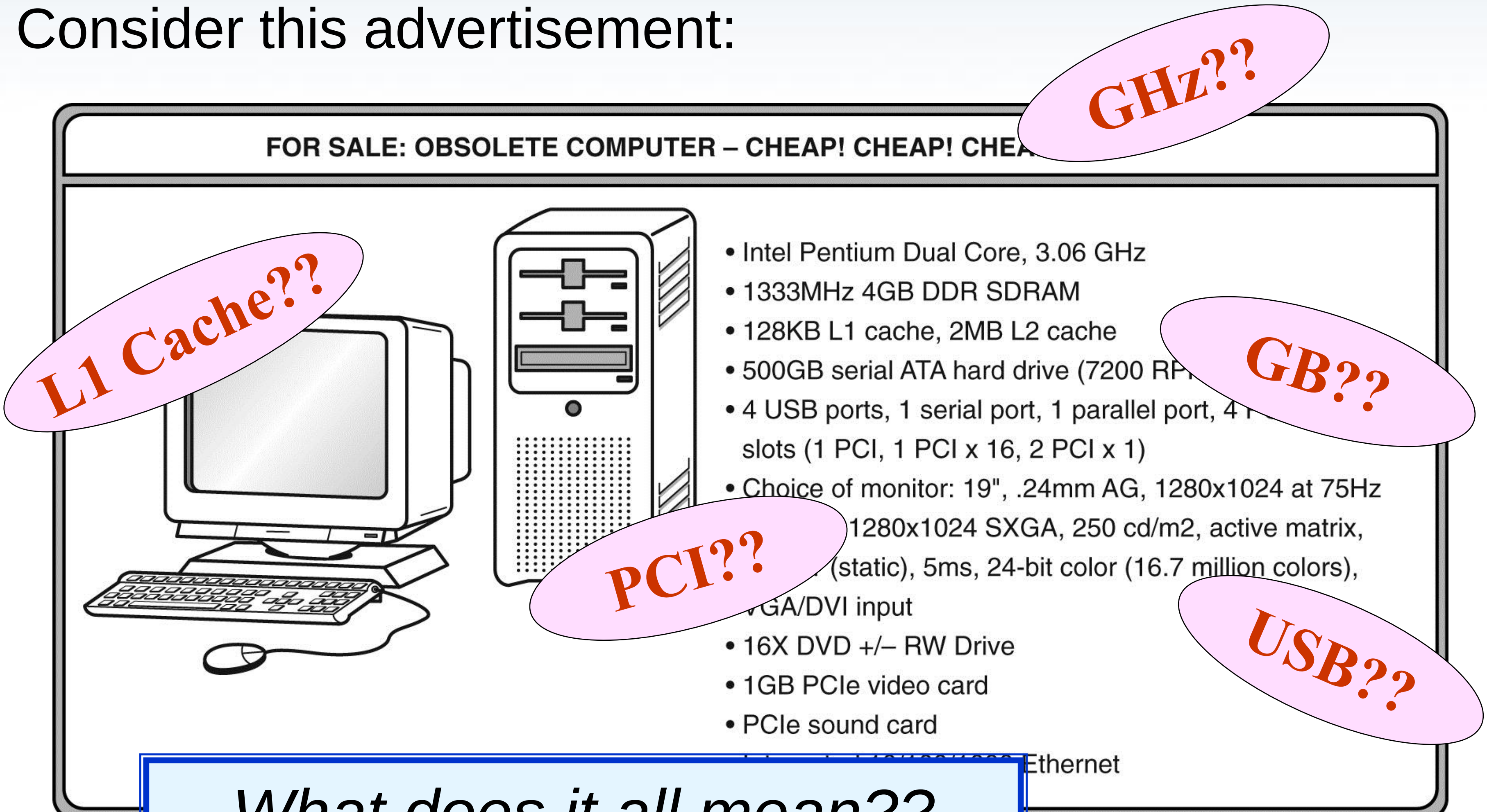
* Assuming speed is not a concern.

1.2 Computer Components

- At the most basic level, a computer is a device consisting of three pieces:
  - A processor to interpret and execute programs
  - A memory to store both data and programs
  - A mechanism for transferring data to and from the outside world.

1.3 An Example System

Consider this advertisement:



What does it all mean??

1.3 An Example System

Measures of capacity and speed:

- Kilo- (K) = 1 thousand = 10^3 and 2^{10}
- Mega- (M) = 1 million = 10^6 and 2^{20}
- Giga- (G) = 1 billion = 10^9 and 2^{30}
- Tera- (T) = 1 trillion = 10^{12} and 2^{40}
- Peta- (P) = 1 quadrillion = 10^{15} and 2^{50}
- Exa- (E) = 1 quintillion = 10^{18} and 2^{60}
- Zetta- (Z) = 1 sextillion = 10^{21} and 2^{70}
- Yotta- (Y) = 1 septillion = 10^{24} and 2^{80}

Whether a metric refers to a power of ten or a power of two <u>typically</u> depends upon what is being measured.

1.3 An Example System

- Hertz = clock cycles per second (frequency)
  – 1MHz = 1,000,000Hz
  – Processor speeds are measured in MHz or GHz.
- Byte = a unit of storage
  – 1KB = 2^{10} = 1024 Bytes
  – 1MB = 2^{20} = 1,048,576 Bytes
  – 1GB = 2^{30} = 1,099,511,627,776 Bytes
  – Main memory (RAM) is measured in GB
  – Disk storage is measured in GB for small systems, TB (2^{40}) for large systems.

1.3 An Example System

Measures of time and space:

- Milli- (m) = 1 thousandth = 10^{-3}
- Micro- (μ) = 1 millionth = 10^{-6}
- Nano- (n) = 1 billionth = 10^{-9}
- Pico- (p) = 1 trillionth = 10^{-12}
- Femto- (f) = 1 quadrillionth = 10^{-15}
- Atto- (a) = 1 quintillionth = 10^{-18}
- Zepto- (z) = 1 sextillionth = 10^{-21}
- Yocto- (y) = 1 septillionth = 10^{-24}

1.3 An Example System

- Millisecond = 1 thousandth of a second
  - Hard disk drive access times are often 10 to 20 milliseconds.
- Nanosecond = 1 billionth of a second
  - Main memory access times are often 50 to 70 nanoseconds.
- Micron (micrometer) = 1 millionth of a meter
  - Circuits on computer chips are measured in microns.

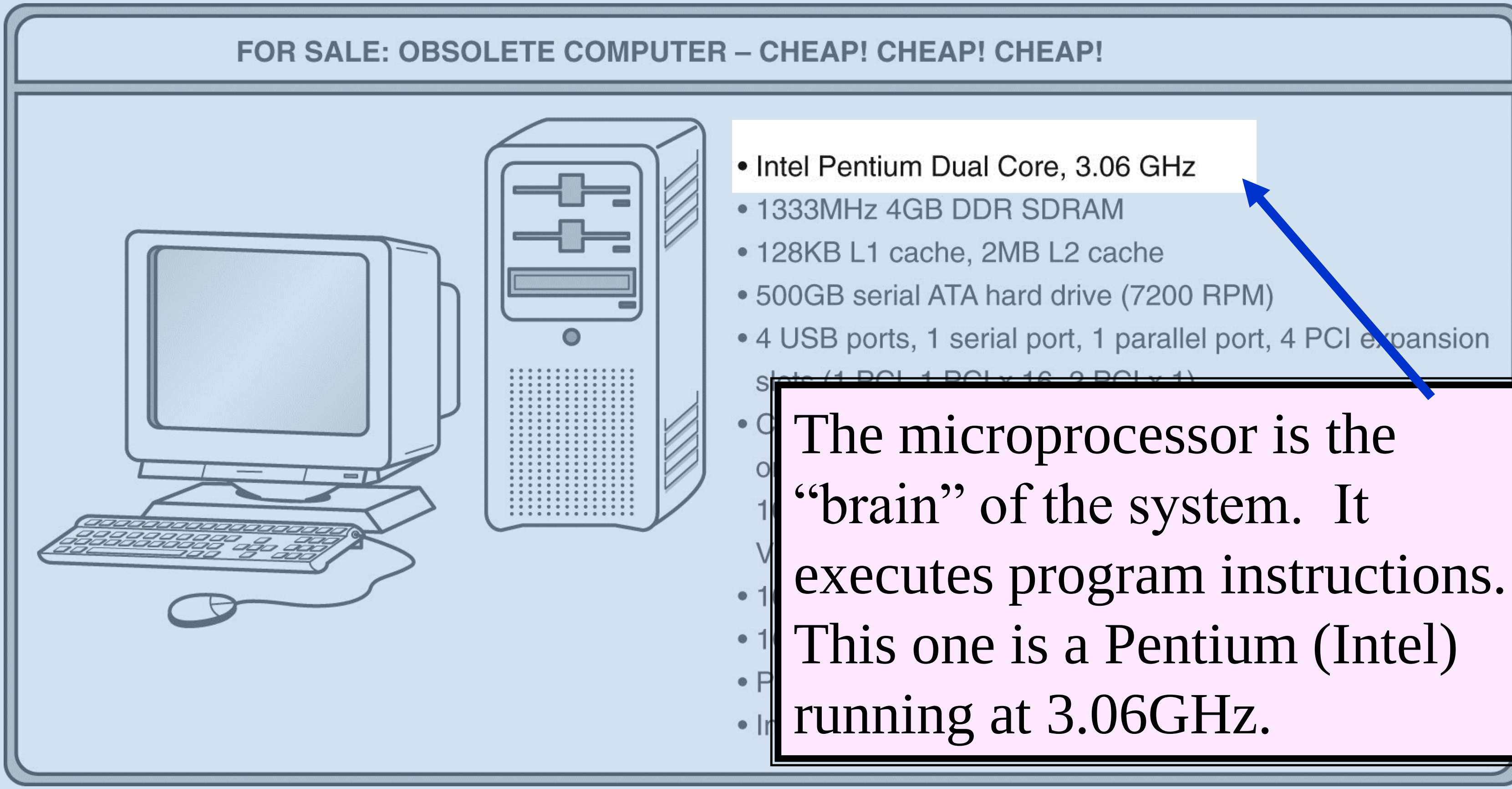
1.3 An Example System

- We note that cycle time is the reciprocal of clock frequency.
- A bus operating at 133MHz has a cycle time of 7.52 nanoseconds:

$$133{,}000{,}000 \text{ cycles/second} = 7.52\text{ns/cycle}$$

Now back to the advertisement ...

1.3 An Example System



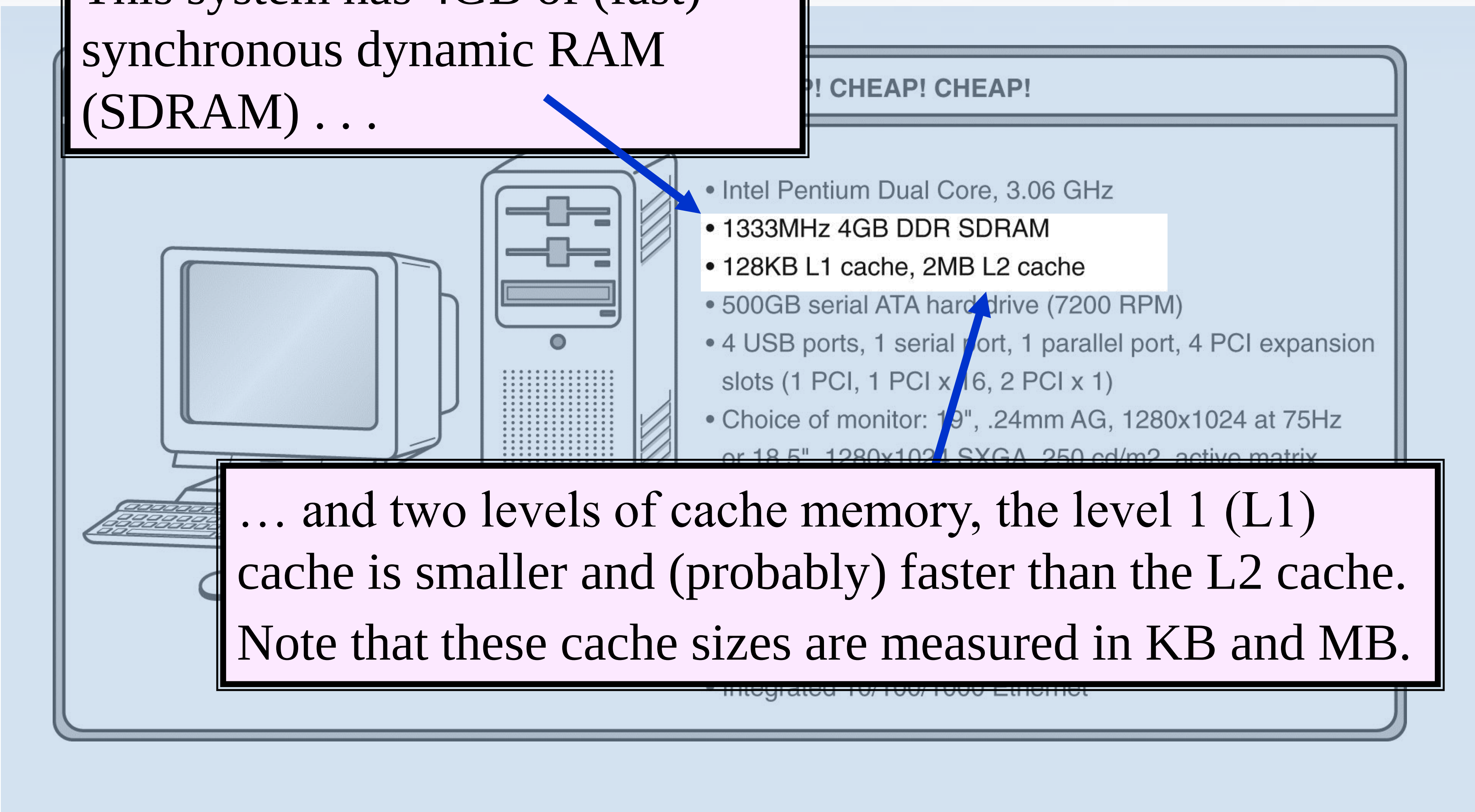
The microprocessor is the "brain" of the system. It executes program instructions. This one is a Pentium (Intel) running at 3.06GHz.

1.3 An Example System

- Computers with large main memory capacity can run larger programs with greater speed than computers having small memories.
- RAM is an acronym for *random access memory*. Random access means that memory contents can be accessed directly if you know its location.
- Cache is a type of temporary memory that can be accessed faster than RAM.

1.3 An Example System

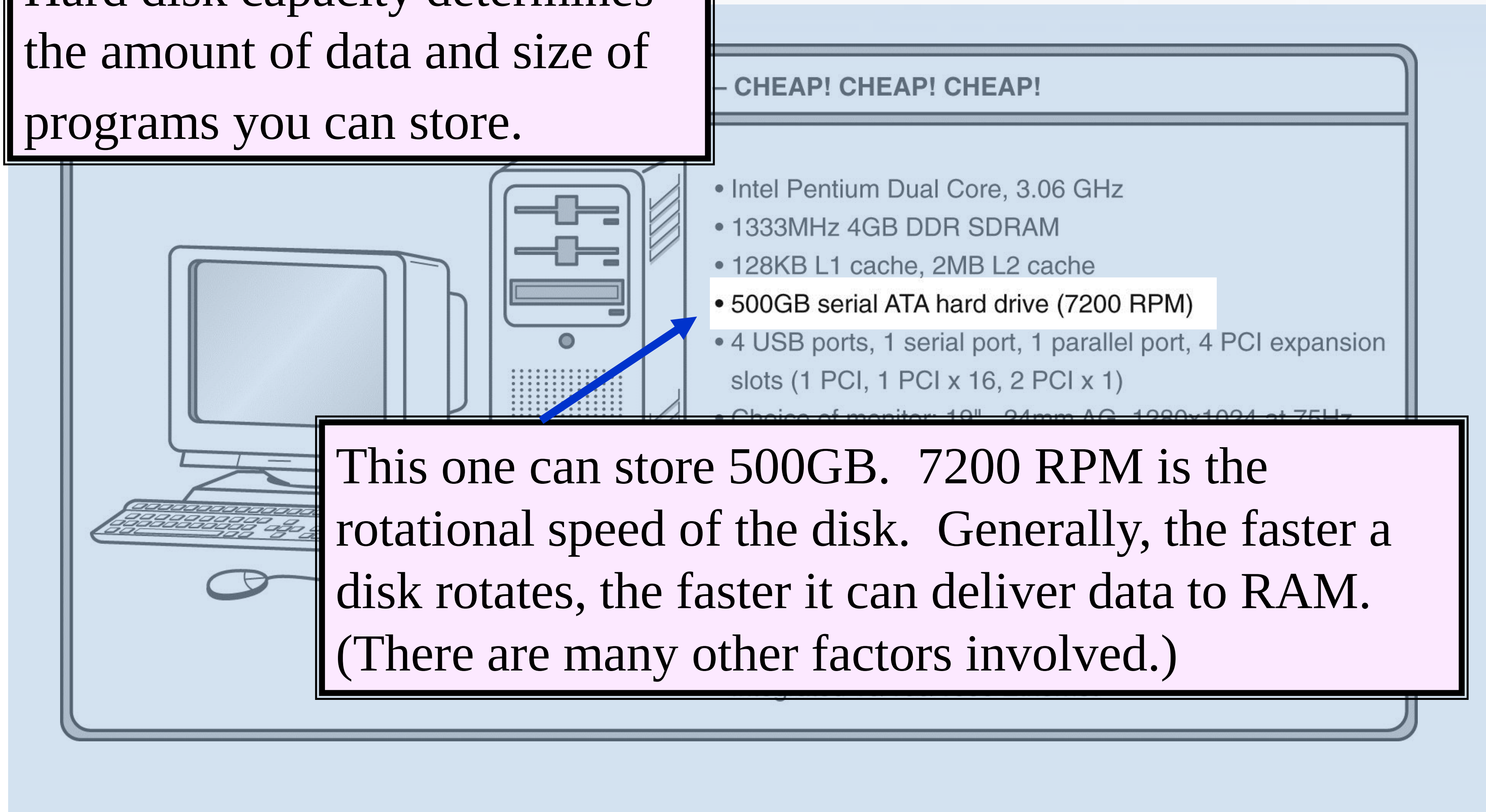
This system has 4GB of (fast) synchronous dynamic RAM (SDRAM) . . .



… and two levels of cache memory, the level 1 (L1) cache is smaller and (probably) faster than the L2 cache. Note that these cache sizes are measured in KB and MB.

1.3 An Example System

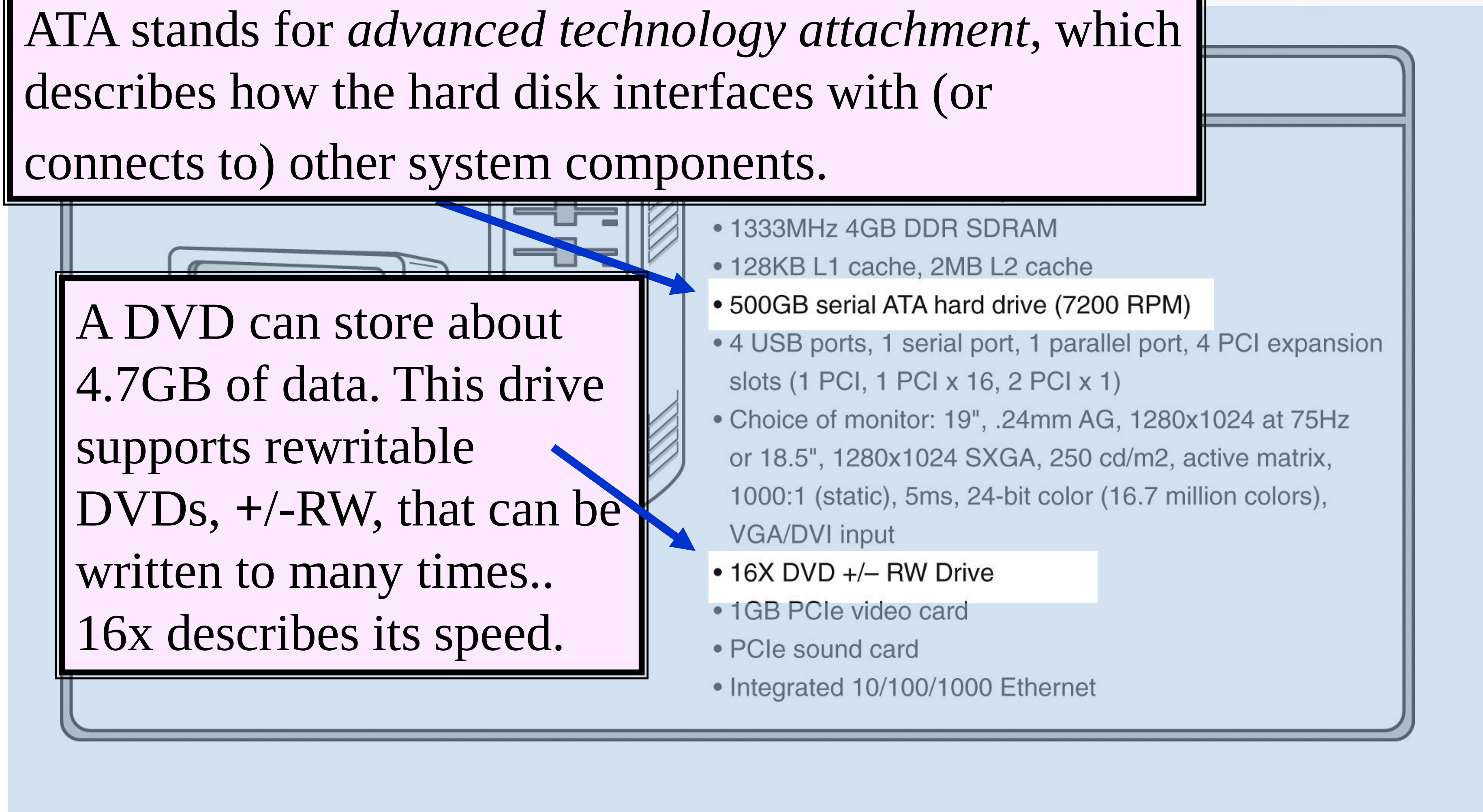
Hard disk capacity determines the amount of data and size of programs you can store.



This one can store 500GB. 7200 RPM is the rotational speed of the disk. Generally, the faster a disk rotates, the faster it can deliver data to RAM. (There are many other factors involved.)

1.3 An Example System

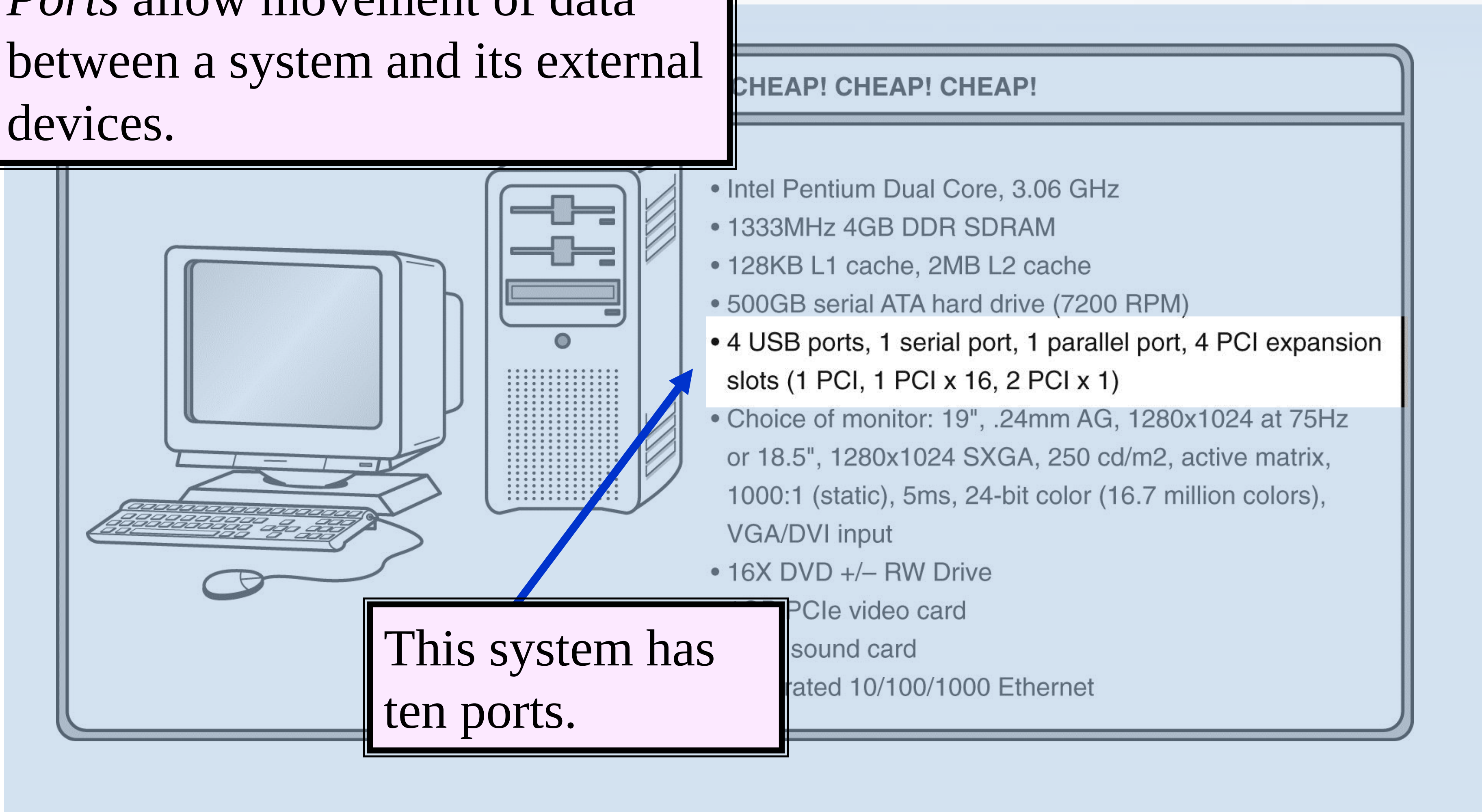
ATA stands for *advanced technology attachment*, which describes how the hard disk interfaces with (or connects to) other system components.



A DVD can store about 4.7GB of data. This drive supports rewritable DVDs, +/-RW, that can be written to many times.. 16x describes its speed.

1.3 An Example System

Ports allow movement of data between a system and its external devices.



This system has ten ports.

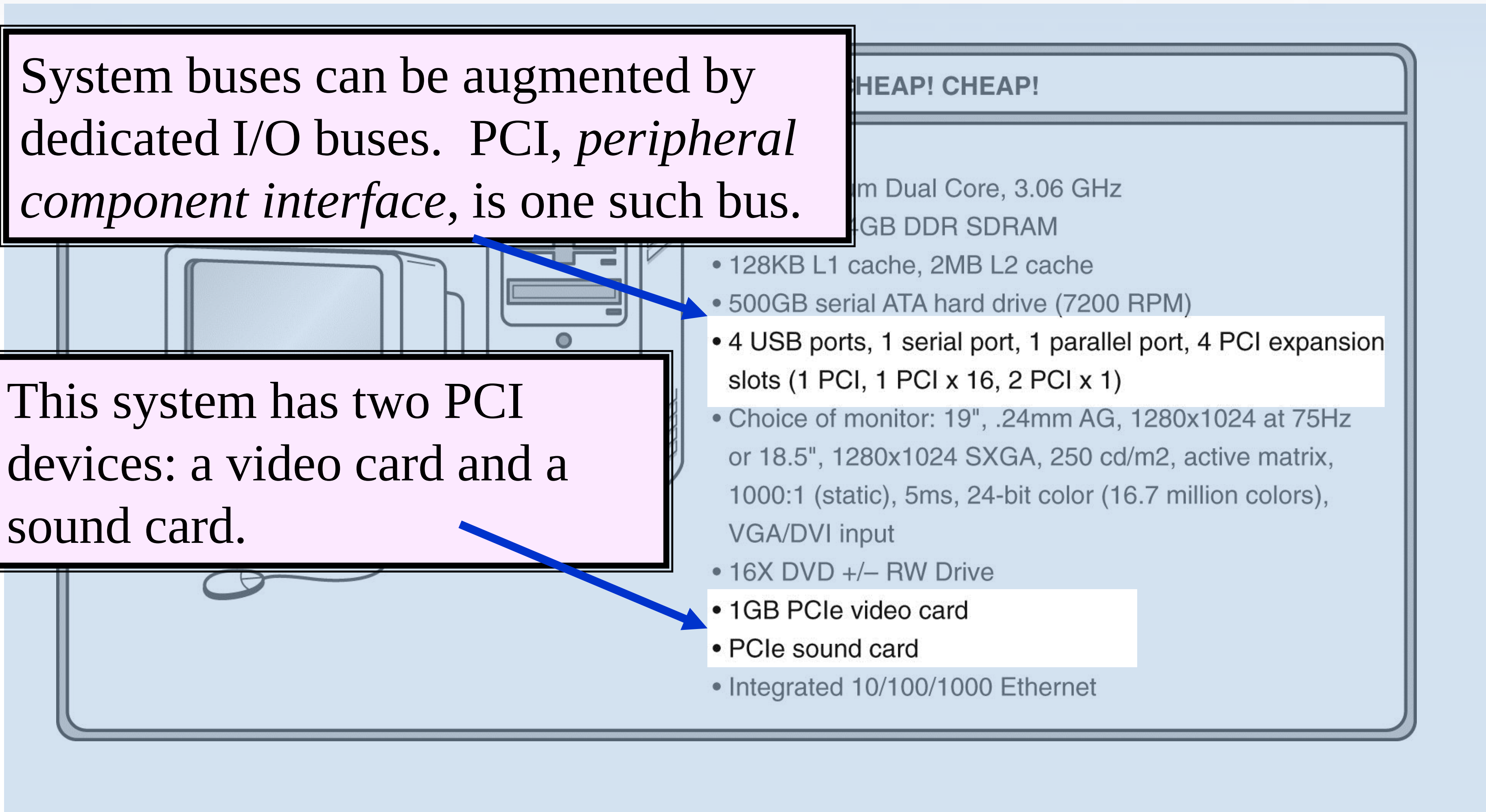
1.3 An Example System

- Serial ports send data as a series of pulses along one or two data lines.
- Parallel ports send data as a single pulse along at least eight data lines.
- USB, Universal Serial Bus, is an intelligent serial interface that is self-configuring. (It supports "plug and play.")

1.3 An Example System

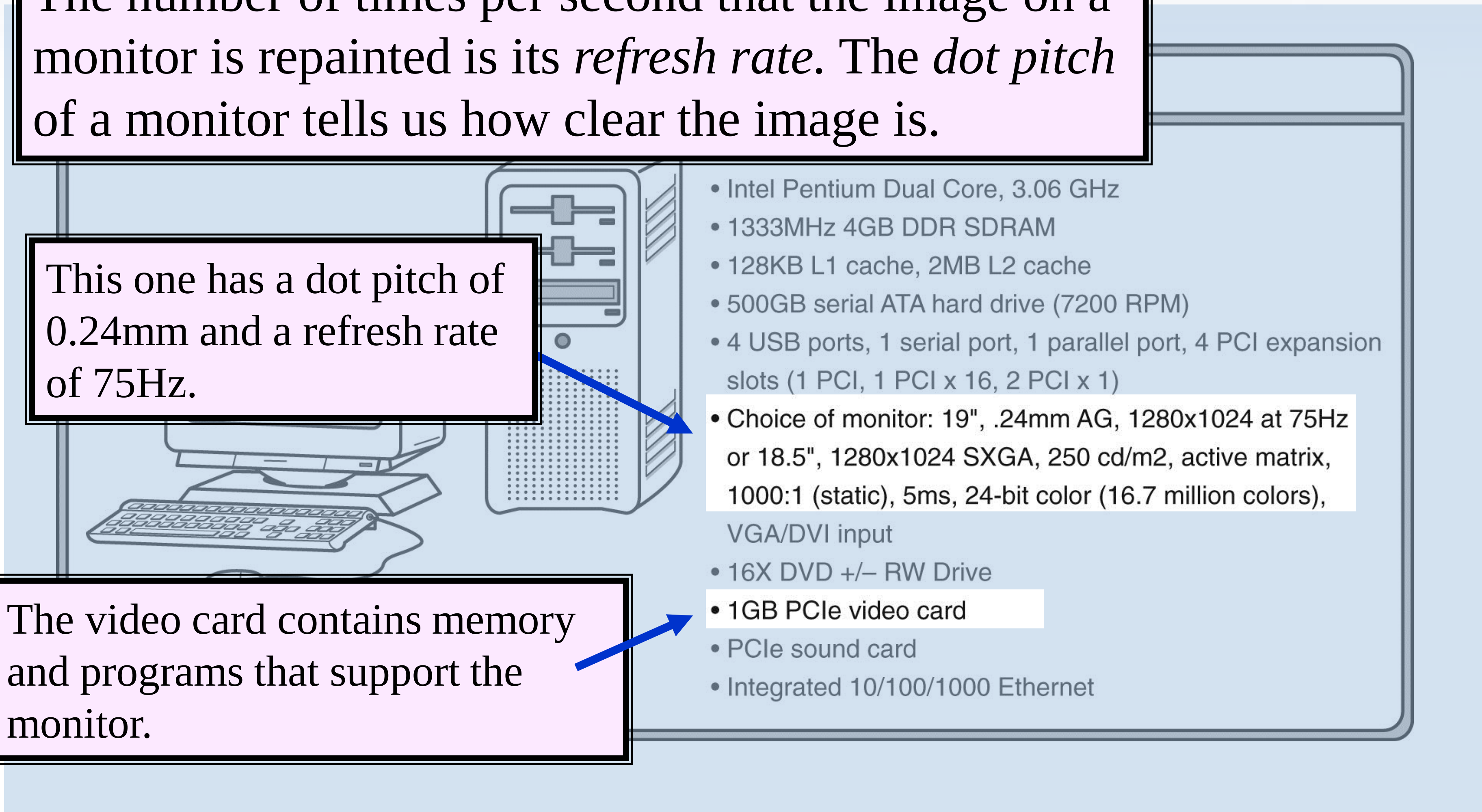
System buses can be augmented by dedicated I/O buses. PCI, *peripheral component interface*, is one such bus.

This system has two PCI devices: a video card and a sound card.



1.3 An Example System

The number of times per second that the image on a monitor is repainted is its *refresh rate*. The *dot pitch* of a monitor tells us how clear the image is.



This one has a dot pitch of 0.24mm and a refresh rate of 75Hz.

The video card contains memory and programs that support the monitor.

1.3 An Example System

Throughout the remainder of the book you will see how these components work and how they interact with software to make complete computer systems.

This statement raises two important questions:

What assurance do we have that computer components will operate as we expect?

And what assurance do we have that computer components will operate together?

1.4 Standards Organizations

- There are many organizations that set computer hardware standards-- to include the interoperability of computer components.
- Throughout this book, and in your career, you will encounter many of them.
- Some of the most important standards-setting groups are . . .

1.4 Standards Organizations

- The Institute of Electrical and Electronic Engineers (IEEE)
  - Promotes the interests of the worldwide electrical engineering community.
  - Establishes standards for computer components, data representation, and signaling protocols, among many other things.

1.4 Standards Organizations

- The International Telecommunications Union (ITU)
  - Concerns itself with the interoperability of telecommunications systems, including data communications and telephony.
- National groups establish standards within their respective countries:
  - The American National Standards Institute (ANSI)
  - The British Standards Institution (BSI)

1.4 Standards Organizations

- The International Organization for Standardization (ISO)
  - Establishes worldwide standards for everything from screw threads to photographic film.
  - Is influential in formulating standards for computer hardware and software, including their methods of manufacture.

Note: ISO is **not** an acronym. ISO comes from the Greek, *isos*, meaning “equal.”

1.5 Historical Development

- To fully appreciate the computers of today, it is helpful to understand how things got the way they are.
- The evolution of computing machinery has taken place over several centuries.
- In modern times computer evolution is usually classified into four generations according to the salient technology of the era.

We note that many of the following dates are approximate.

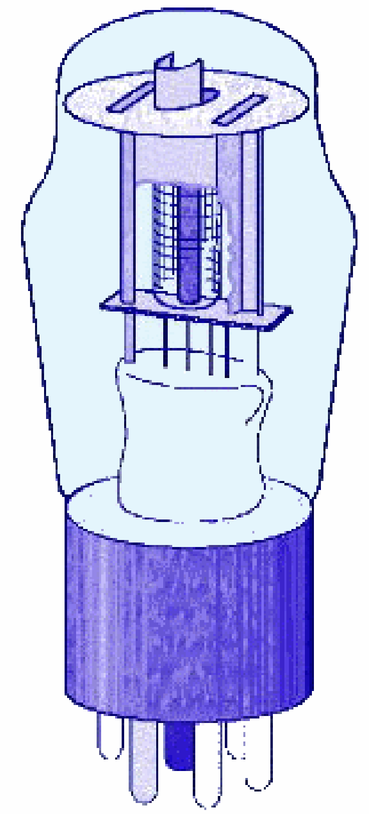
1.5 Historical Development

- Generation Zero: Mechanical Calculating Machines (1642 - 1945)
  - Calculating Clock - Wilhelm Schickard (1592 - 1635).
  - Pascaline - Blaise Pascal (1623 - 1662).
  - Difference Engine - Charles Babbage (1791 - 1871), also designed but never built the Analytical Engine.
  - Punched card tabulating machines - Herman Hollerith (1860 - 1929).

Hollerith cards were commonly used for computer input well into the 1970s.

1.5 Historical Development

- The First Generation: Vacuum Tube Computers (1945 - 1953)



  - Atanasoff Berry Computer (1937 - 1938) solved systems of linear equations.
  - John Atanasoff and Clifford Berry of Iowa State University.

1.5 Historical Development

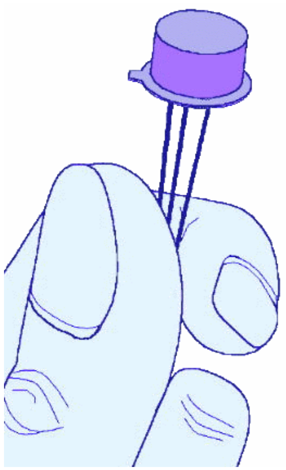
- The First Generation: Vacuum Tube Computers (1945 - 1953)
  - Electronic Numerical Integrator and Computer (ENIAC)
  - John Mauchly and J. Presper Eckert
  - University of Pennsylvania, 1946
- The ENIAC was the first general-purpose computer.

1.5 Historical Development

- The First Generation: Vacuum Tube Computers (1945 - 1953)
  - The IBM 650 first mass-produced computer. (1955)
    - It was phased out in 1969.
  - Other major computer manufacturers of this period include UNIVAC, Engineering Research Associates (ERA), and Computer Research Corporation (CRC).
    - UNIVAC and ERA were bought by Remington Rand, the ancestor of the Unisys Corporation.
    - CRC was bought by the Underwood (typewriter) Corporation, which left the computer business.

1.5 Historical Development

- The Second Generation: Transistorized Computers (1954 - 1965)



  - IBM 7094 (scientific) and 1401 (business)
  - Digital Equipment Corporation (DEC) PDP-1
  - Univac 1100
  - Control Data Corporation 1604.
  - . . . and many others.

These systems had few architectural similarities.

1.5 Historical Development

- The Third Generation: Integrated Circuit Computers (1965 - 1980)
  – IBM 360
  – DEC PDP-8 and PDP-11
  – Cray-1 supercomputer
  – . . . and many others.
- By this time, IBM had gained overwhelming dominance in the industry.
  – Computer manufacturers of this era were characterized as IBM and the BUNCH (Burroughs, Unisys, NCR, Control Data, and Honeywell).

1.5 Historical Development

- The Fourth Generation: VLSI Computers (1980 - ????)
  - Very large scale integrated circuits (VLSI) have more than 10,000 components per chip.
  - Enabled the creation of microprocessors.
  - The first was the 4-bit Intel 4004.
  - Later versions, such as the 8080, 8086, and 8088 spawned the idea of "personal computing."

1.5 Historical Development

- Moore's Law (1965)
  - Gordon Moore, Intel founder
  - "The density of transistors in an integrated circuit will double every year."
- Contemporary version:
  - "The density of silicon chips doubles every 18 months."

But this "law" cannot hold forever ...

1.5 Historical Development

- Rock's Law
  - Arthur Rock, Intel financier
  - "The cost of capital equipment to build semiconductors will double every four years."
  - In 1968, a new chip plant cost about $12,000.

At the time, $12,000 would buy a nice home in the suburbs.

An executive earning $12,000 per year was "making a very comfortable living."

1.5 Historical Development

- Rock's Law
  - In 2010, a chip plants under construction cost well over $4 billion.

$4 billion is more than the gross domestic product of some small countries, including Barbados, Mauritania, and Rwanda.

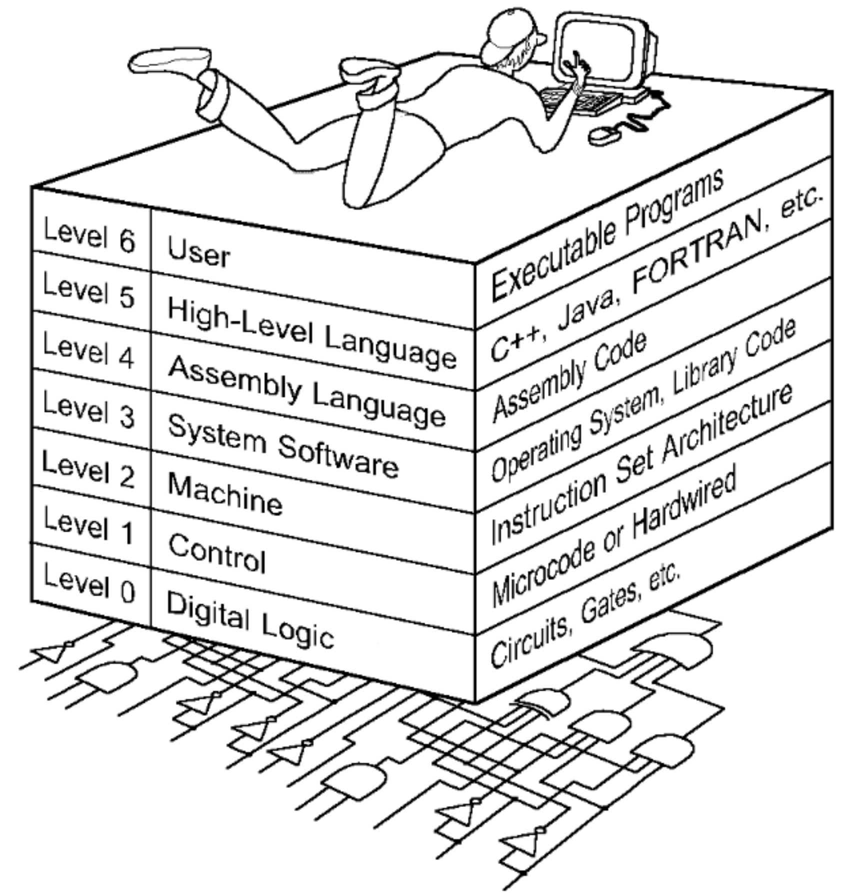
  - For Moore's Law to hold, Rock's Law must fall, or vice versa. But no one can say which will give out first.

1.6 The Computer Level Hierarchy

- Computers consist of many things besides chips.
- Before a computer can do anything worthwhile, it must also use software.
- Writing complex programs requires a “divide and conquer” approach, where each program module solves a smaller problem.
- Complex computer systems employ a similar technique through a series of virtual machine layers.

1.6 The Computer Level Hierarchy

- **Each virtual machine layer is an abstraction of the level below it.**
- **The machines at each level execute their own particular instructions, calling upon machines at lower levels to perform tasks as required.**
- **Computer circuits ultimately carry out the work.**



1.6 The Computer Level Hierarchy

- Level 6: The User Level
  - Program execution and user interface level.
  - The level with which we are most familiar.
- Level 5: High-Level Language Level
  - The level with which we interact when we write programs in languages such as C, Pascal, Lisp, and Java.

1.6 The Computer Level Hierarchy

- Level 4: Assembly Language Level
  - Acts upon assembly language produced from Level 5, as well as instructions programmed directly at this level.
- Level 3: System Software Level
  - Controls executing processes on the system.
  - Protects system resources.
  - Assembly language instructions often pass through Level 3 without modification.

1.6 The Computer Level Hierarchy

- Level 2: Machine Level
  - Also known as the Instruction Set Architecture (ISA) Level.
  - Consists of instructions that are particular to the architecture of the machine.
  - Programs written in machine language need no compilers, interpreters, or assemblers.

1.6 The Computer Level Hierarchy

- Level 1: Control Level
  - A *control unit* decodes and executes instructions and moves data through the system.
  - Control units can be *microprogrammed* or *hardwired*.
  - A microprogram is a program written in a low-level language that is implemented by the hardware.
  - Hardwired control units consist of hardware that directly executes machine instructions.

1.6 The Computer Level Hierarchy

- Level 0: Digital Logic Level
  - This level is where we find digital circuits (the chips).
  - Digital circuits consist of gates and wires.
  - These components implement the mathematical logic of all other levels.

1.7 The von Neumann Model

- On the ENIAC, all programming was done at the digital logic level.
- Programming the computer involved moving plugs and wires.
- A different hardware configuration was needed to solve every unique problem type.

Configuring the ENIAC to solve a "simple" problem required many days labor by skilled technicians.

1.7 The von Neumann Model

- Inventors of the ENIAC, John Mauchley and J. Presper Eckert, conceived of a computer that could store instructions in memory.
- The invention of this idea has since been ascribed to a mathematician, John von Neumann, who was a contemporary of Mauchley and Eckert.
- Stored-program computers have become known as von Neumann Architecture systems.

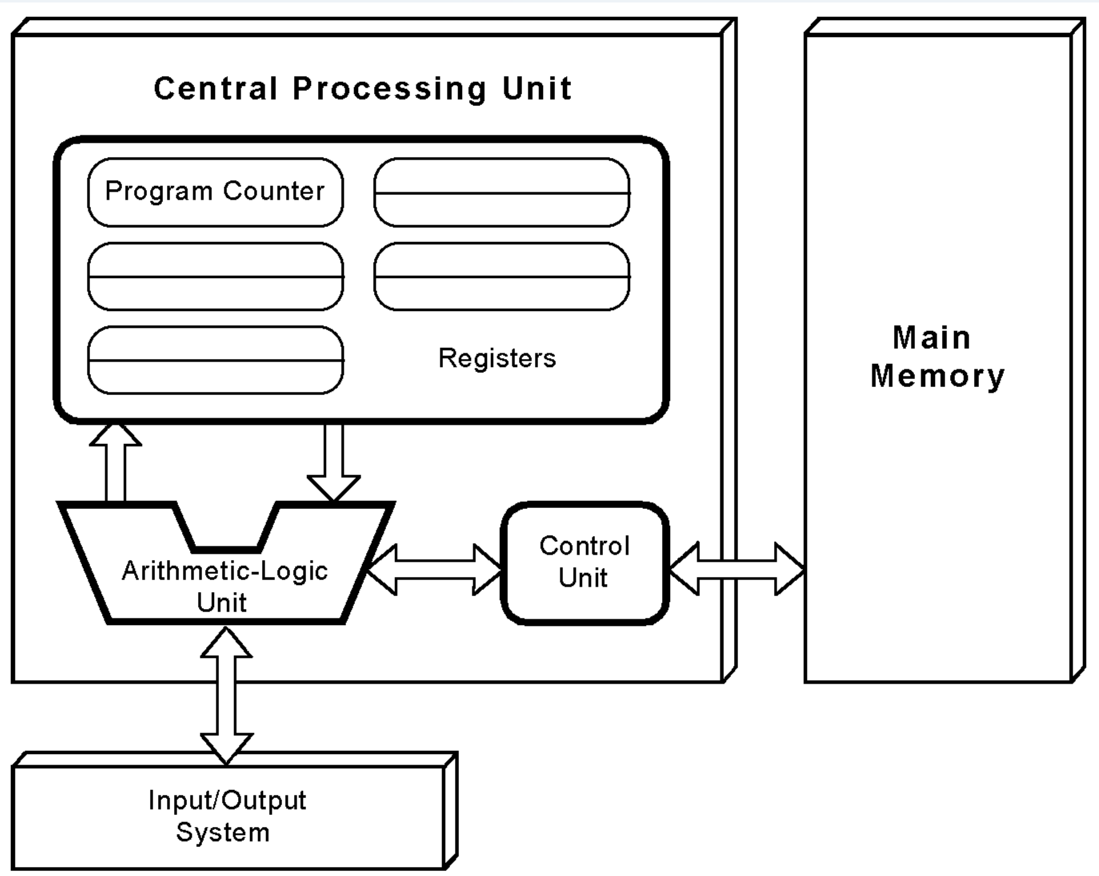
1.7 The von Neumann Model

- Today's stored-program computers have the following characteristics:
  - Three hardware systems:
    - A central processing unit (CPU)
    - A main memory system
    - An I/O system
  - The capacity to carry out sequential instruction processing.
  - A single data path between the CPU and main memory.
    - This single path is known as the *von Neumann bottleneck*.

1.7 The von Neumann Model

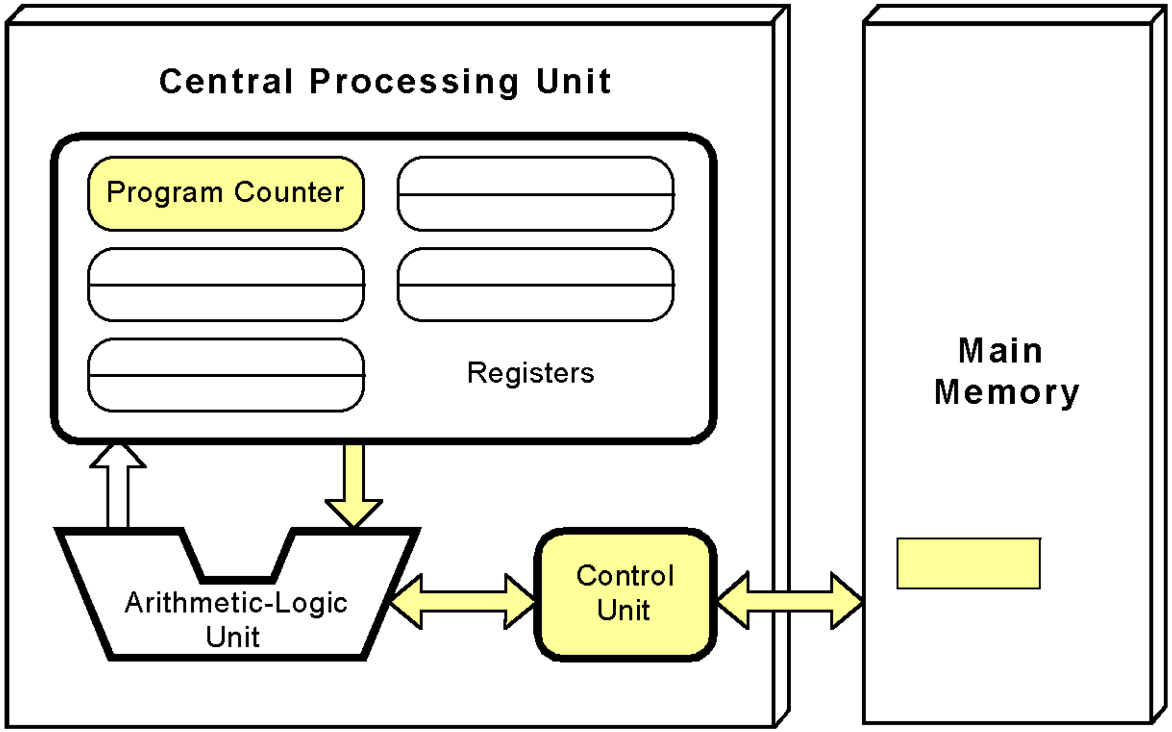
- **This is a general depiction of a von Neumann system:**

- **These computers employ a fetch-decode-execute cycle to run programs as follows . . .**



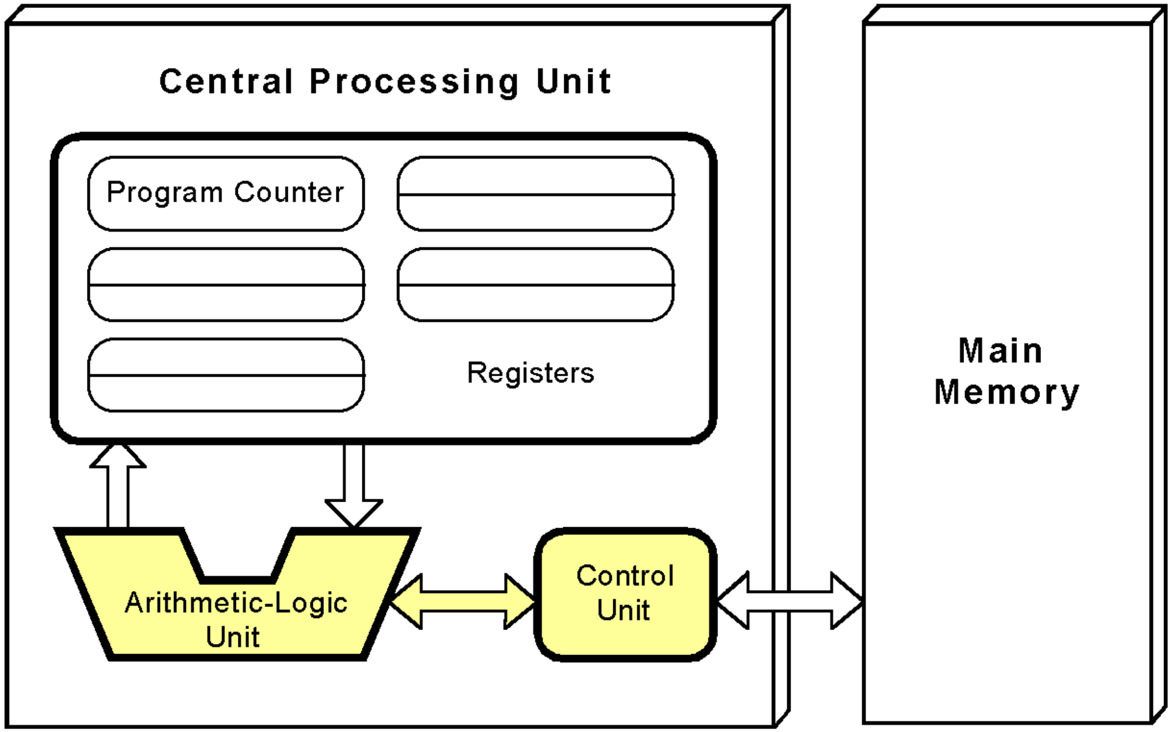
1.7 The von Neumann Model

- **The control unit fetches the next instruction from memory using the program counter to determine where the instruction is located.**



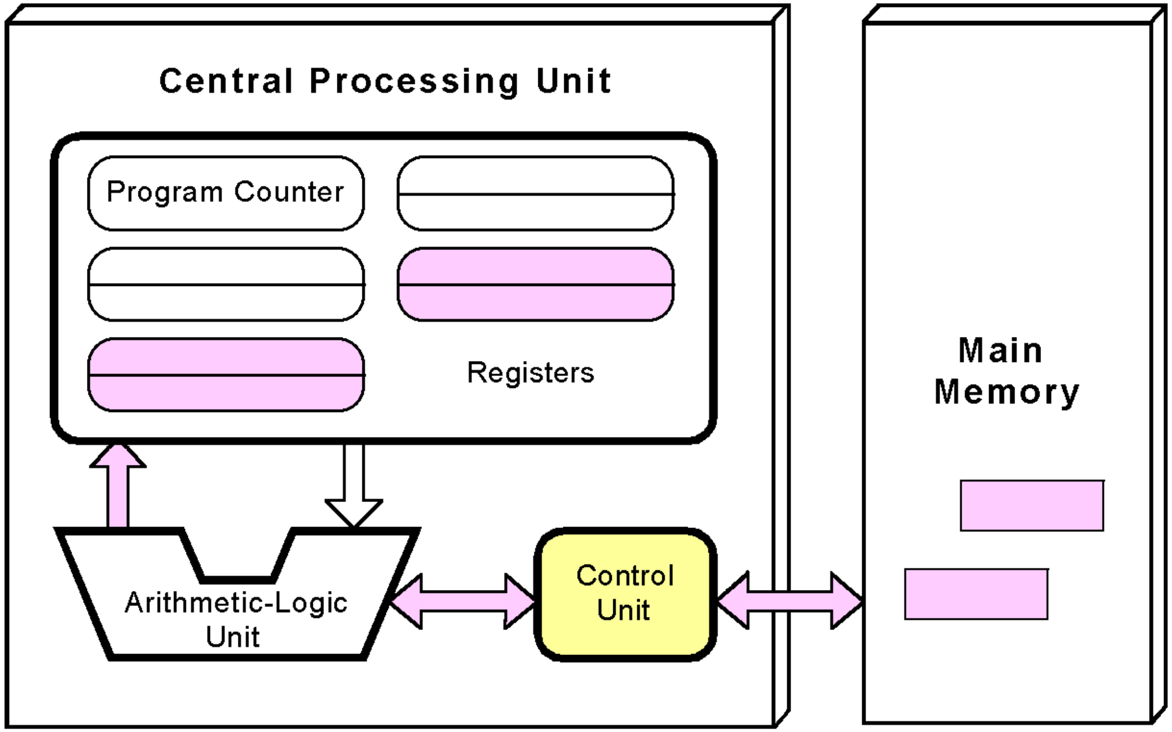
1.7 The von Neumann Model

- **The instruction is decoded into a language that the ALU can understand.**



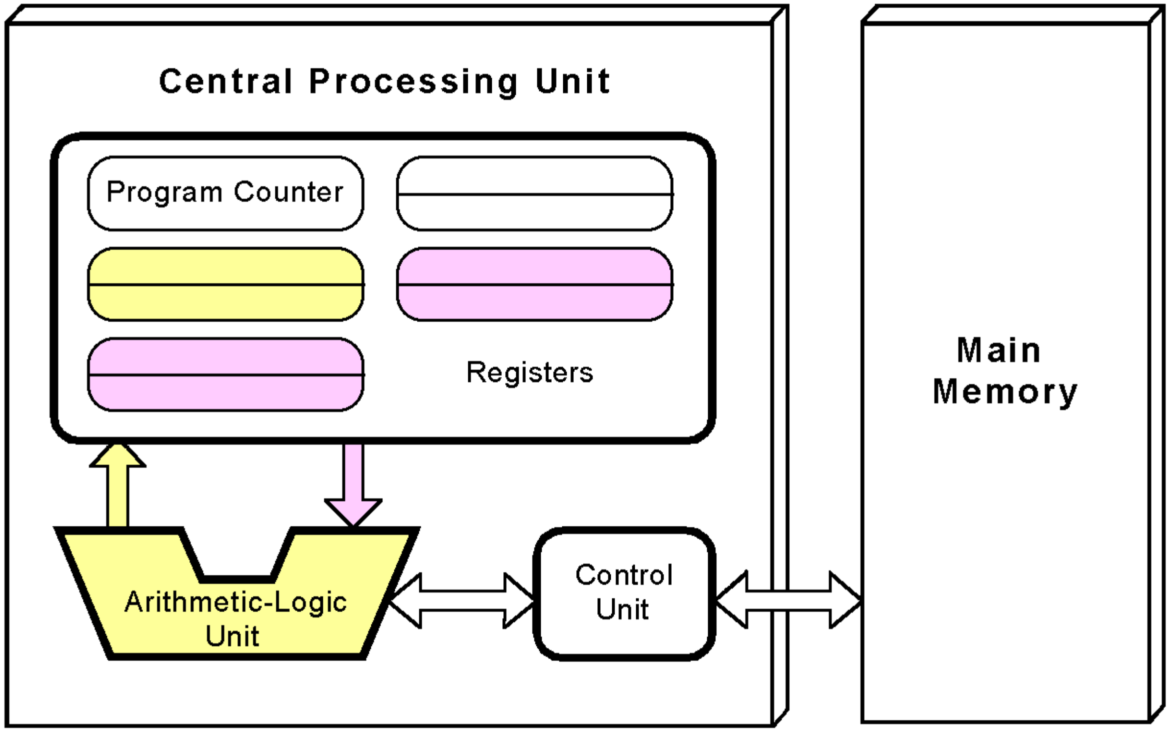
1.7 The von Neumann Model

- **Any data operands required to execute the instruction are fetched from memory and placed into registers within the CPU.**



1.7 The von Neumann Model

- **The ALU executes the instruction and places results in registers or memory.**



1.8 Non-von Neumann Models

- Conventional stored-program computers have undergone many incremental improvements over the years.
- These improvements include adding specialized buses, floating-point units, and cache memories, to name only a few.
- But enormous improvements in computational power require departure from the classic von Neumann architecture.
- Adding processors is one approach.

1.8 Non-von Neumann Models

- In the late 1960s, high-performance computer systems were equipped with dual processors to increase computational throughput.
- In the 1970s supercomputer systems were introduced with 32 processors.
- Supercomputers with 1,000 processors were built in the 1980s.
- In 1999, IBM announced its Blue Gene system containing over 1 million processors.

1.8 Non-von Neumann Models

- Multicore architectures have multiple CPUs on a single chip.
- Dual-core and quad-core chips are commonplace in desktop systems.
- Multi-core systems provide the ability to multitask
  - E.g., browse the Web while burning a CD
- Multithreaded applications spread mini-processes, *threads*, across one or more processors for increased throughput.

Conclusion

- This chapter has given you an overview of the subject of computer architecture.
- You should now be sufficiently familiar with general system structure to guide your studies throughout the remainder of this course.
- Subsequent chapters will explore many of these topics in great detail.

End of Chapter 1