Chapter 1

Motherboards, Processors, and Memory

THE FOLLOWING COMPTIA A+ 220-1101 EXAM OBJECTIVES ARE COVERED IN THIS CHAPTER:

✓ **3.2 Given a scenario, install the appropriate RAM.**

- RAM types
  - Virtual RAM
  - Small outline dual inline memory module (SODIMM)
  - Double Data Rate 3 (DDR3)
  - Double Data Rate 4 (DDR4)
  - Double Data Rate 5 (DDR5)
  - Error correction code (ECC) RAM
- Single-channel
- Dual-channel
- Triple-channel
- Quad-channel

✓ **3.4 Given a scenario, install and configure motherboards, central processing units (CPUs), and add-on cards.**

- Motherboard form factor
  - Advanced Technology eXtended (ATX)
  - Information Technology eXtended (ITX)
- Motherboard connector types
  - Peripheral Component Interconnect (PCI)
  - PCI Express (PCIe)

  - **Power connectors**
  - **SATA**
  - **eSATA**
  - **SAN**
  - **Headers**
  - **M.2**
- **Motherboard compatibility**
  - **CPU sockets**
- **Advanced Micro Devices, Inc. (AMD)**
- **Intel**
  - **Server**
  - **Multisocket**
  - **Desktop**
  - **Mobile**
- **Basic Input/Output System (BIOS)/Unified Extensible Firmware Interface (UEFI) settings**
  - **Boot options**
  - **USB permissions**
  - **Trusted Platform Module (TPM) security features**
  - **Fan considerations**
  - **Secure Boot**
  - **Boot password**
- **Encryption**
  - **TPM**
  - **Hardware security module (HSM)**
- **CPU architecture**
  - **x64/x86**
  - **Advanced RISC Machine (ARM)**
  - **Single-core**

- Multicore
- Multithreading
- Virtualization support

- Cooling
  - Fans
  - Heat sink
  - Thermal paste/pads
  - Liquid

The computers we use daily, from the largest servers to the smallest smart watches and everything in between, are collections of different electronic components and software working together in a system. Digital computers have been around since the late 1930s, so they aren't new news. As you would imagine though, their looks and functionality have evolved considerably since then.

As technology improved over the years, computers got smaller and faster, and inventors added features that required new hardware devices. Inspired by typewriters, keyboards were added for ease of input. Visual displays required a monitor and a video card and a standard interface between them. Sound was provided by a different expansion card and speakers. Because of the newly added features, PCs were modular by necessity. That is, if a new feature or functionality were needed, a new component could be added. Or if a part failed, it could be replaced by a new one. Along the way in the late 1960s, the term *personal computer (PC)* was coined to differentiate between computers designed to be used by one person versus other options, such as a mainframe or one where multiple users share a processor. This book, and the CompTIA A+ exams, focus on PC hardware and software.

NOTE Unless specifically mentioned otherwise, the terms *PC* and *computer* are used interchangeably throughout this book.

Much of the computing industry today is focused on smaller devices, such as laptops, tablets, and smartphones. Laptops have outsold desktop computers since 2005, and it seems that everyone has a smartphone glued to their hand today. Smaller devices require the same components as do their bigger desktop-sized cousins. Of course, the components are smaller and many times integrated into the same circuit board. The functionality of the individual parts is still critical, though, so what you learn here will serve you well regardless of the type of device you're working on.

Even though all parts inside a computer case are important, some are more important than others. You'll sometimes hear people refer to the "big three" computer parts, which are the motherboard, processor, and memory. Without these, a computer won't work, whereas if a sound card fails, it's probably not going to make the entire system inoperable. In this chapter, you will learn how to identify, install, and configure the big three, which are critical to computer functionality. You'll also learn about cooling systems, because too much heat will cause components to melt, which could make for a very bad day for the computer user.

NOTE With small computing devices being far more prevalent, you might wonder why we are starting the book talking about components that are primarily in desktop and larger systems. We intentionally cover this material first because it's important to understand what each component does individually. It helps you isolate the functionality of each piece of the system, which is critical for troubleshooting and fixing problems. Besides, if you are working (or want to be working) as a computer technician, you will likely see plenty of desktop-sized PCs in your travels!

Understanding Motherboards

The spine of the computer is the *motherboard*, otherwise known as the *system board* or *mainboard*. This is the *printed circuit board (PCB)*, which is a conductive series of pathways laminated to a nonconductive substrate that lines the bottom of the computer and is often of a uniform color, such as green, brown, blue, black, or red. It is the most important component in the computer because it connects all the other components together. Figure 1.1 shows a typical PC system board, as seen from above. All other components are attached to this circuit board. On the system board, you will find the central processing unit (CPU) slot or integrated CPU, underlying circuitry, expansion slots, video components, random access memory (RAM) slots, and a variety of other chips. We will be discussing each of these components throughout this book.

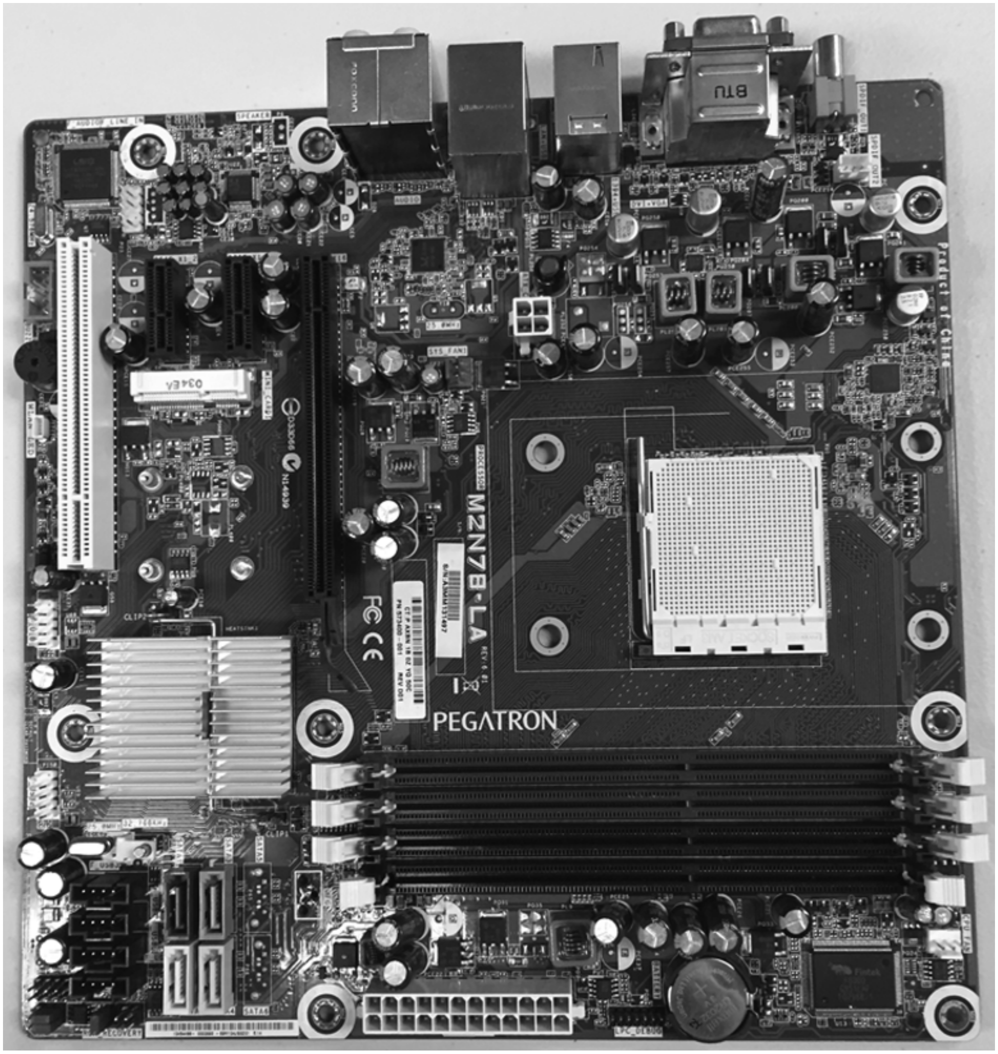
Motherboard Form Factors

There are hundreds, if not thousands, of different motherboards in the market today. It can be overwhelming trying to figure out which one is needed. To help, think of motherboards in terms of which type they are, based on a standard classification. System boards are classified by their *form factor* (design), such as ATX and ITX. Exercise care and vigilance when acquiring a motherboard and components separately. Motherboards will have different expansion slots, support certain processors and memory, and fit into some cases but not others. Be sure that the other parts are physically compatible with the motherboard you choose.

Advanced Technology Extended

Intel developed the *Advanced Technology eXtended (ATX)* motherboard in the mid-1990s to improve upon the classic AT-style motherboard architecture that had ruled the PC world for many years. The ATX motherboard has the processor and memory slots at right angles to the expansion cards, like the one in Figure 1.1. This arrangement puts the processor and memory in line with the fan output of the power supply, allowing the processor to run cooler. And because those components are not in line with the expansion cards, you can install full-length expansion cards—adapters that extend the full length of the inside of a standard computer case—in an ATX motherboard machine. ATX (and its derivatives, such as micro-ATX) is the primary PC motherboard form factor in use today. Standard ATX motherboards measure 12″ × 9.6″ (305 mm × 244 mm).

FIGURE 1.1 A typical motherboard



NOTE We will discuss expansion cards in more depth (and cover that part of Exam Objective 3.4) in Chapter 2, "Expansion Cards, Storage Devices, and Power Supplies."

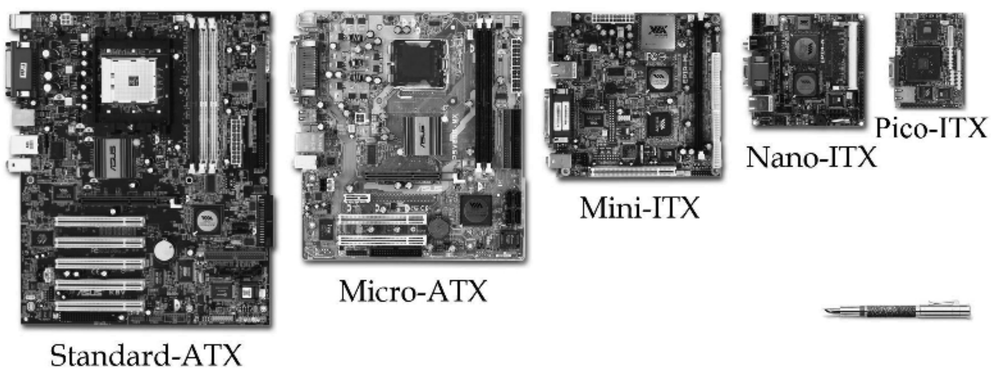
Information Technology eXtended

The *Information Technology eXtended (ITX)* line of motherboard form factors was developed by VIA Technologies in the early 2000s as a low-power, small form factor (SFF) board for specialty uses, including home-theater systems, compact desktop systems, gaming systems, and embedded components. ITX itself is not an actual form factor but a family of form factors. The family consists of the following form factors:

- **Mini-ITX—**6.7″ × 6.7″ (170 mm × 170 mm)
- **Nano-ITX—**4.7″ × 4.7″ (120 mm × 120 mm)
- **Pico-ITX—**3.9″ × 2.8″ (100 mm × 72 mm)
- **Mobile-ITX—**2.4″ × 2.4″ (60 mm × 60 mm)

The mini-ITX motherboard has four mounting holes that line up with three or four of the holes in the ATX and micro-ATX form factors. In mini-ITX boards, the rear interfaces are placed in the same location as those on the ATX motherboards. These features make mini-ITX boards compatible with ATX cases. This is where the mounting compatibility ends, because despite the PC compatibility of the other ITX form factors, they are used in embedded systems, such as set-top boxes, home entertainment systems, and smartphones, and lack the requisite mounting and interface specifications. Figure 1.2 shows the three larger forms of ITX motherboards, next to two ATX motherboards for comparison.

FIGURE 1.2 ITX motherboards



VIA Mini-ITX Form Factor Comparison by VIA Gallery from Hsintien, Taiwan; VIA Mainboards Form Factor Comparison uploaded by Kozuch, licensed under CC BY 2.0 via Commons

System Board Components

Now that you understand the basic types of motherboards and their form factors, it's time to look at the key characteristics and components of the motherboard and, where applicable, their locations relative to each other. The following list summarizes key concepts you need to know about motherboards:

- Bus architecture
- Chipsets
- Expansion slots
- Memory slots and cache
- Central processing units and processor sockets
- Power connectors
- Onboard non-volatile storage connectors (such as for hard drives)
- Motherboard headers
- BIOS/UEFI/firmware
- CMOS and CMOS battery
- Front-panel connectors and headers

In the following sections, you will learn about some of the most common components of a motherboard, what they do, and where they are located on the motherboard. We'll show what each component looks like so that you can identify it on most any motherboard that you run across. In the case of some components, this chapter provides only a brief introduction, with more detail to come in later chapters.

Bus Architecture

In a PC, data is sent from one component to another via a *bus*, which is a common collection of signal pathways. In the very early days, PCs used serial buses, which sent one bit at a time and were painfully slow. Brilliant engineers realized that they could redesign the bus and send 8 bits at a time (over synchronized separate lines), which resulted in a big speed increase. This was known as a *parallel bus*.

The downside of parallel communications is the loss of circuit length (how long the circuit could be) and throughput (how much data could move at one time). The signal could travel only a short distance, and the amount of data was limited due to the careful synchronization needed between separate lines, the speed of which must be controlled to limit skewing the arrival of the individual signals at the receiving end.

What was once old is new again, as engineers have discovered methods to make serial transmissions work at data rates that are many times faster than parallel signals. Therefore, nearly everything you see today uses a serial bus. The only limitation of serial circuits is in the capability of the transceivers, which tends to grow over time at a refreshing rate due to technical advancements. Examples of specifications that have heralded the dominance of serial communications are Serial Advanced Technology Attachment (Serial ATA, or SATA), Universal Serial Bus (USB), IEEE 1394/FireWire, and Peripheral Component Interconnect Express (PCIe).

NOTE

The term *bus* is also used in any parallel or bit-serial wiring implementation where multiple devices can be attached at the same time in parallel or in series (daisy-chained). Examples include Small Computer System Interface (SCSI), USB, and Ethernet.

On a motherboard, several different buses are used. Expansion slots of various architectures, such as PCIe, are included to allow for the insertion of external devices or adapters. Other types of buses exist within the system to allow communication between the CPU, RAM, and other components with which data must be exchanged. Except for CPU slots and sockets and memory slots, there are no insertion points for devices in many closed bus systems because no adapters exist for such an environment.

The various buses throughout a given computer system can be rated by their bus speeds. The higher the bus speed, the higher its performance. In some cases, various buses must be synchronized for proper performance, such as the system bus and any expansion buses that run at the front-side bus speed. Other times, one bus will reference another for its own speed. The internal bus speed of a CPU is derived from the front-side bus clock, for instance. The buses presented throughout this chapter are accompanied by their speeds, where appropriate.

Chipsets

A *chipset* is a collection of chips or circuits that perform interface and peripheral functions for the processor. This collection of chips is usually the circuitry that provides interfaces for memory, expansion cards, and onboard peripherals, and it generally dictates how a motherboard will communicate with the installed peripherals.

Chipsets are usually given a name and model number by the original manufacturer. For example, B550 and X570 are chipsets that support Advanced Micro Devices, Inc. (AMD) processors, and Z490 and H410 are Intel motherboard chipsets. Typically, the manufacturer and model tell you that your particular chipset has a certain set of features (for example, type of CPU and RAM supported, type and brand of onboard video, and so on). Don't worry about memorizing any chipset names—you can look them up online to understand their features.

Chipsets can be made up of one or several integrated circuit chips. Intel-based motherboards, for example, typically use two chips. To know for sure, you must check the manufacturer's documentation, especially because cooling mechanisms frequently obscure today's chipset chips, sometimes hindering visual brand and model identification.

Chipsets can be divided into two major functional groups, called Northbridge and Southbridge. Let's take a brief look at these groups and the functions they perform.

NOTE AMD and Intel have integrated the features of Northbridge and Southbridge into most of their CPUs. Therefore, the CPU provides Northbridge and Southbridge functionality as opposed to separate chipsets.

Northbridge

The *Northbridge* subset of a motherboard's chipset is the set of circuitry or chips that performs one very important function: management of high-speed peripheral communications. The Northbridge is responsible primarily for communications with integrated video using PCIe, for instance, and processor-to-memory communications. Therefore, it can be said that much of the true performance of a PC relies on the specifications of the Northbridge component and its communications capability with the peripherals it controls.

NOTE When we use the term *Northbridge*, we are referring to a functional subset of a motherboard's chipset. There isn't actually a Northbridge brand of chipset.

The communications between the CPU and memory occur over what is known as the *front-side bus (FSB)*, which is just a set of signal pathways connecting the CPU and main memory, for instance. The clock signal that drives the FSB is used to drive communications by certain other devices, such as PCIe slots, making them local-bus technologies. The *back-side bus (BSB)*, if present, is a set of signal pathways between the CPU and external cache memory. The BSB uses the same clock signal that drives the FSB. If no back-side bus exists, cache is placed on the front-side bus with the CPU and main memory.

The Northbridge is directly connected to the Southbridge (discussed next). It controls the Southbridge and helps to manage the communications between the Southbridge and the rest of the computer.

Southbridge

The *Southbridge* subset of the chipset is responsible for providing support to the slower onboard peripherals (USB, Serial and Parallel ATA, parallel ports, serial ports, and so on), managing their communications with the rest of the computer and the resources given to them. These components do not need to keep up with the external clock of the CPU and do not represent a bottleneck in the overall performance of the system. Any component that would impose such a restriction on the system should eventually be developed for FSB attachment.

In other words, if you're considering any component other than the CPU, memory and cache, or PCIe slots, the Southbridge is in charge. Most motherboards today have integrated USB, network, and analog and digital audio ports for the Southbridge to manage, for example, all of which are discussed in more detail later in this chapter or in Chapter 3, "Peripherals, Cables, and Connectors." The Southbridge is also responsible for managing communications with the slower expansion buses, such as PCI, and legacy buses.

Figure 1.3 is a photo of the chipset of a motherboard, with the heat sink of the Northbridge at the top left, connected to the heat-spreading cover of the Southbridge at the bottom right.

FIGURE 1.3 A modern computer chipset

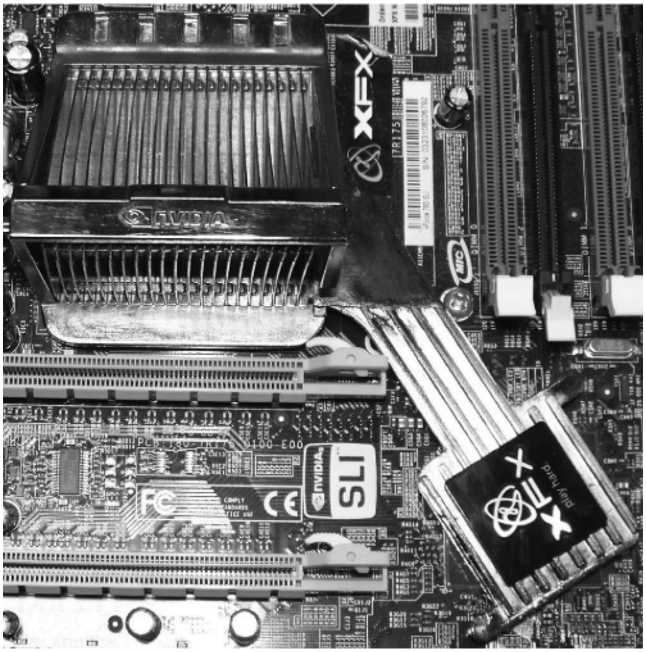
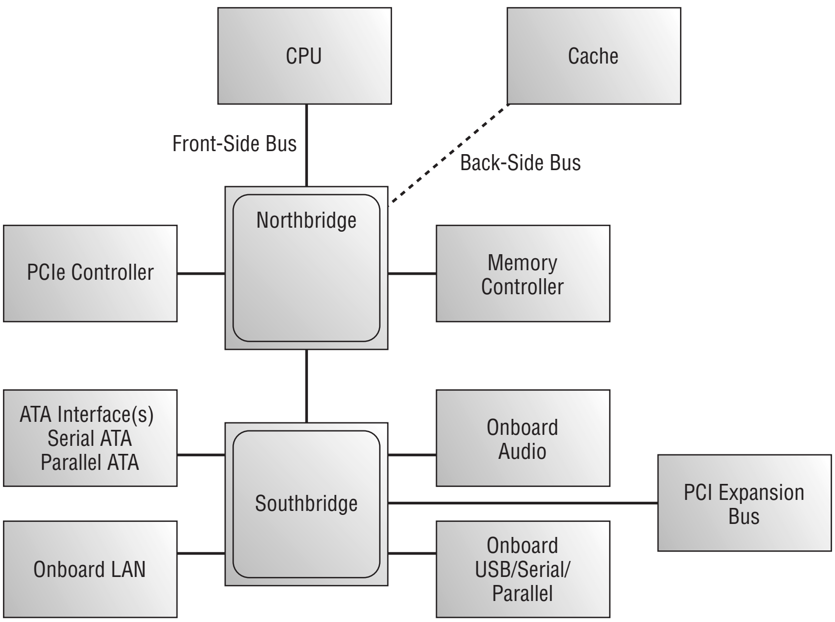


Figure 1.4 shows a schematic of a typical motherboard chipset (both Northbridge and Southbridge) and the components with which they interface. Notice which components interface with which parts of the chipset.

FIGURE 1.4 A schematic of a typical motherboard chipset



Expansion Slots

The most visible parts of any motherboard are the *expansion slots*. These are small plastic slots, usually from 1 to 6 inches long and approximately ½ inch wide. As their name suggests, these slots are used to install various devices in the computer to expand its capabilities. Some expansion devices that might be installed in these slots include video, network, sound, and disk interface cards.

If you look at the motherboard in your computer, you will more than likely see one of the main types of expansion slots used in computers today, which are PCI and PCIe. In the following sections, we will cover how to visually identify the different expansion slots on the motherboard.

PCI Expansion Slots

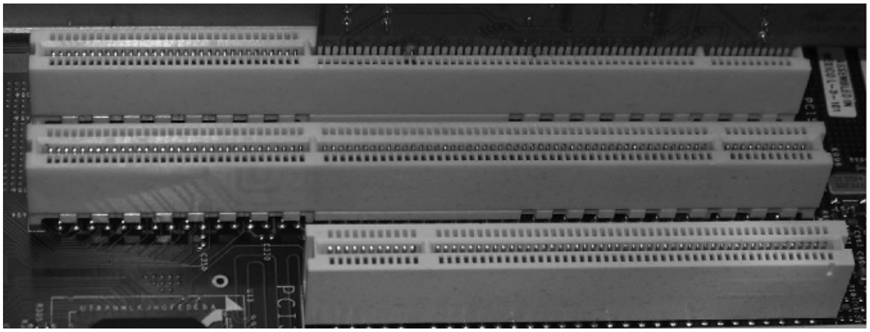
It's now considered an old technology, but many motherboards in use today still contain 32-bit *Peripheral Component Interconnect (PCI)* slots. They are easily recognizable because they are only around 3 inches long and classically white, although modern boards take liberties with the color. PCI slots became extremely popular with the advent of Pentium-class processors in the mid-1990s. Although popularity has shifted from PCI to PCIe, the PCI slot's service to the industry cannot be ignored; it has been an incredibly prolific architecture for many years.

PCI expansion buses operate at 33 MHz or 66 MHz (version 2.1) over a 32-bit (4-byte) channel, resulting in data rates of 133 MBps and 266 MBps, respectively, with 133 MBps being the most common, server architectures excluded. PCI is a shared-bus topology, however, so mixing 33 MHz and 66 MHz adapters in a 66 MHz system will slow all adapters to 33 MHz. Older servers might have featured 64-bit PCI slots as well, since version 1.0, which double the 32-bit data rates. See the sidebar "Arriving at the Exact Answer" for help with understanding the math involved in frequencies and bit rates.

> **Arriving at the Exact Answer**
>
> To get the math exactly right when dealing with frequencies and data rates ending in 33 and 66, you have to realize that every 33 has an associated one-third (1/3), and every 66 has an associated two-thirds (2/3). The extra quantities are left off of the final result but must be added back on to get the math exactly right. The good news is that omitting these small values from the equation still gets you close, and a bit of experience with the numbers leads to being able to make the connection on the fly.

PCI slots and adapters are manufactured in 3.3V and 5V versions. Universal adapters are keyed to fit in slots based on either of the two voltages. The notch in the card edge of the common 5V slots and adapters is oriented toward the front of the motherboard, and the notch in the 3.3V adapters toward the rear. Figure 1.5 shows several PCI expansion slots. Note the 5V 32-bit slot in the foreground and the 3.3V 64-bit slots. Also notice that a universal 32-bit card, which has notches in both positions, is inserted into and operates fine in the 64-bit 3.3V slot in the background.

FIGURE 1.5 PCI expansion slots



PCIe Expansion Slots

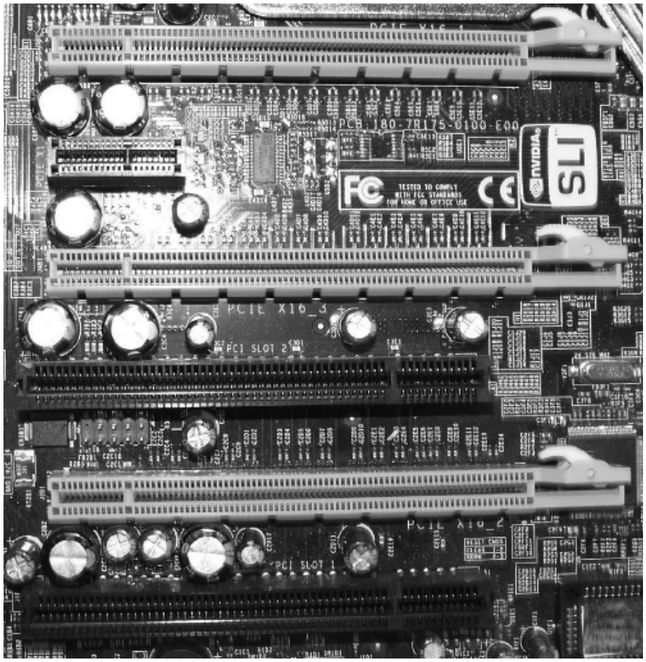
The most common expansion slot architecture that is being used by motherboards is *PCI Express (PCIe)*. It was designed to be a replacement for PCI, as well as an older video card standard called accelerated graphics port (AGP). PCIe has the advantage of being faster than AGP while maintaining the flexibility of PCI. PCIe has no plug compatibility with either AGP or PCI. Some modern PCIe motherboards can be found with regular PCI slots for backward compatibility, but AGP slots have not been included for many years.

PCIe is casually referred to as a bus architecture to simplify its comparison with other bus technologies. True expansion buses share total bandwidth among all slots, each of which taps into different points along the common bus lines. In contrast, PCIe uses a switching component with point-to-point connections to slots, giving each component full use of the corresponding bandwidth and producing more of a star topology versus a bus. Furthermore, unlike other expansion buses, which have parallel architectures, PCIe is a serial technology, striping data packets across multiple serial paths to achieve higher data rates.

PCIe uses the concept of *lanes*, which are the switched point-to-point signal paths between any two PCIe components. Each lane that the switch interconnects between any two intercommunicating devices comprises a separate pair of wires for both directions of traffic. Each PCIe pairing between cards requires a negotiation for the highest mutually supported number of lanes. The single lane or combined collection of lanes that the switch interconnects between devices is referred to as a *link*.

There are seven different link widths supported by PCIe, designated x1 (pronounced "by 1"), x2, x4, x8, x12, x16, and x32, with x1, x4, and x16 being the most common. The x8 link width is less common than these but more common than the others. A slot that supports a particular link width is of a physical size related to that width because the width is based on the number of lanes supported, requiring a related number of wires. As a result, an x8 slot is longer than an x1 slot but shorter than an x16 slot. Every PCIe slot has a 22-pin portion in common toward the rear of the motherboard, which you can see in Figure 1.7, in which the rear of the motherboard is to the left. These 22 pins comprise mostly voltage and ground leads. (The PCIe slots are the longer and lighter ones in Figure 1.6.)

FIGURE 1.6 PCIe expansion slots



Four major versions of PCIe are currently available in the market: 1.x, 2.x, 3.0, and 4.0. For the four versions, a single lane, and therefore an x1 slot, operates in each direction (or transmits and receives from either communicating device's perspective), at a data rate of 250 MBps (almost twice the rate of the most common PCI slot), 500 MBps, approximately 1 GBps, and roughly 2 GBps, respectively.

NOTE

PCIe 5.0 was formally ratified by the PCI Special Interest Group (PCI-SIG) in 2019, and motherboards supporting the architecture started hitting the market in late 2021. Much like its predecessors, it doubles the speed of the previous version. Therefore, a PCIe 5.0 x1 adapter operates at about 4 MBps in each direction. The slots are forward and backward compatible. For example, you can put a PCIe 4.0 video card into a motherboard with a PCIe 5.0 slot, but you won't get the full performance that PCIe 5.0 is capable of. The inverse is true as well—a PCIe 5.0 card will work in a 4.0 slot but at the 4.0 standard's speed. PCIe 6.0 is expected around 2023.

An associated bidirectional link has a nominal throughput of double these rates. Use the doubled rate when comparing PCIe to other expansion buses because those other rates are for bidirectional communication. This means that the 500 MBps bidirectional link of an x1 slot in the first version of PCIe was comparable to PCI's best, a 64-bit slot running at 66 MHz and producing a throughput of 533 MBps.

NOTE *Bidirectional* means that data flows in both directions, often simultaneously. *Unidirectional* means data flows in only one direction.

Combining lanes simply results in a linear multiplication of these rates. For example, a PCIe 1.1 x16 slot is capable of 4 GBps of throughput in each direction, 16 times the 250 MBps x1 rate. Bidirectionally, this fairly common slot produces a throughput of 8 GBps. Each subsequent PCIe specification doubles this throughput. The aforementioned PCIe 5.0 will produce bidirectional throughput of approximately 128 GBps, which is faster than some DDR4 standards (which is to say, it's really, really fast).

Using Shorter Cards in Longer Slots

Up-plugging is defined in the PCIe specification as the ability to use a higher-capability slot for a lesser adapter. In other words, you can use a shorter (fewer-lane) card in a longer slot. For example, you can insert an x8 card into an x16 slot. The x8 card won't completely fill the slot, but it will work at x8 speeds if up-plugging is supported by the motherboard. Otherwise, the specification requires up-plugged devices to operate at only the x1 rate. This is something you should be aware of and investigate in advance. Down-plugging is possible only on open-ended slots, although not specifically allowed in the official specification. Even if you find or make (by cutting a groove in the end) an open-ended slot that accepts a longer card edge, the inserted adapter cannot operate faster than the slot's maximum rated capability because the required physical wiring to the PCIe switch in the Northbridge is not present.

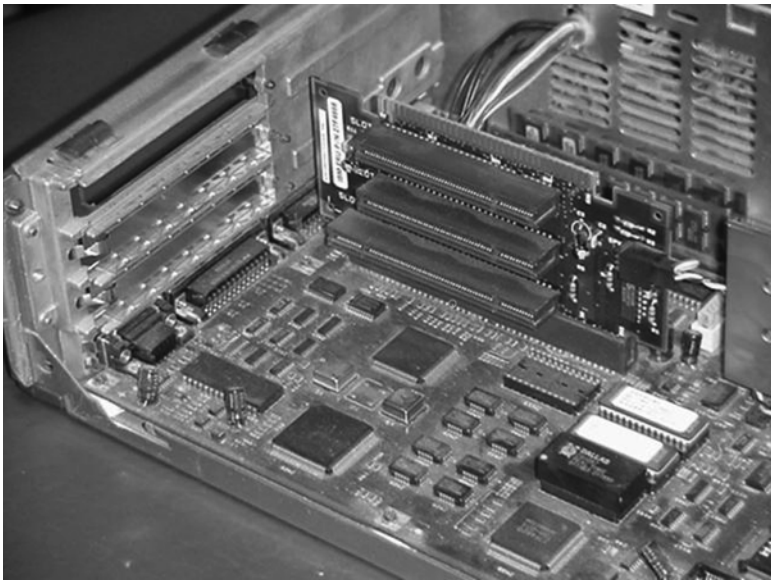
Because of its high data rate, PCIe is the current choice of gaming aficionados. Additionally, technologies similar to NVIDIA's Scalable Link Interface (SLI) allow such users to combine preferably identical graphics adapters in appropriately spaced PCIe x16 slots with a hardware bridge to form a single virtual graphics adapter. The job of the bridge is to provide non-chipset communication among the adapters. The bridge is not a requirement for SLI to work, but performance suffers without it. SLI-ready motherboards allow two, three, or four PCIe graphics adapters to pool their graphics processing units (GPUs) and memory to feed graphics output to a single monitor attached to the adapter acting as the primary SLI device. SLI implementation results in increased graphics performance over single-PCIe and non-PCIe implementations.

Refer back to Figure 1.6, which is a photo of an SLI-ready motherboard with three PCIe x16 slots (every other slot, starting with the top one), one PCIe x1 slot (second slot from the top), and two PCI slots (first and third slots from the bottom). Notice the latch and tab that secures the x16 adapters in place by their hooks. Any movement of these high-performance devices can result in temporary failure or poor performance.

Using Riser Cards

Most PC expansion cards plug directly into the motherboard. A special type of expansion card, called a *riser card*, provides additional slots that other expansion cards plug into. The other expansion cards are then parallel to the motherboard, as opposed to perpendicular. Figure 1.7 shows an example of a riser card. It's an older example, but it illustrates the concept well.

FIGURE 1.7 Riser card in a motherboard



Riser cards aren't often found in desktop PCs today but do still find some use in rack-mounted servers with low-profile cases. The motherboard must be designed to accept a riser card, and the case needs to be built for it as well, for the external portion of the expansion card to be accessible.

Memory Slots and Cache

Memory, or *random access memory (RAM)*, slots are the next most notable slots on a motherboard. These slots are designed for the modules that hold memory chips that make up primary memory, which is used to store currently used data and instructions for the CPU. Many types of memory are available for PCs today. In this chapter, you will become familiar with the appearance and specifications of the slots on the motherboard so that you can identify them and appropriately install or replace RAM.

For the most part, PCs today use memory chips arranged on a small circuit board. A *dual in-line memory module (DIMM)* is one type of circuit board. Today's DIMMs differ in the number of conductors, or pins, that each particular physical form factor uses. Some common examples include 168-, 184-, 240-, and 288-pin configurations. In addition, laptop memory comes in smaller form factors known as small outline DIMMs (SODIMMs) and Micro-DIMMs. More detail on memory packaging and the technologies that use them can be found later in this chapter in the section "Understanding Memory."

Memory slots are easy to identify on a motherboard. Classic DIMM slots were usually black and, like all memory slots, were placed very close together. DIMM slots with color-coding are more common these days, however. The color-coding of the slots acts as a guide to the installer of the memory. See the section "Single-, Dual-, Triple-, and Quad-Channel Memory" later in this chapter for more on the purpose of this color-coding. Consult the motherboard's documentation to determine the specific modules allowed as well as their required orientation. The number of memory slots varies from motherboard to motherboard, but the structure of the different slots is similar. Metal pins in the bottom make contact with the metallic pins on each memory module. Small metal or plastic tabs on each side of the slot keep the memory module securely in its slot. Figure 1.8 shows four memory slots, with the CPU socket included for reference.

FIGURE 1.8 Double Data Rate (DDR) memory slots

Sometimes, the amount of primary memory installed is inadequate to service additional requests for memory resources from newly launched applications. When this condition occurs, the user may receive an "out of memory" error message and an application may fail to launch. One solution for this is to use the hard drive as additional RAM. This space on the hard drive is known as a *swap file* or a *paging file*. The technology in general is known as *virtual memory* or *virtual RAM*. The paging file is called PAGEFILE.SYS in modern Microsoft operating systems. It is an optimized space that can deliver information to RAM at the request of the memory controller faster than if it came from the general storage pool of the drive. It's located at `c:\pagefile.sys` by default. Note that virtual memory cannot be used directly from the hard drive; it must be paged into RAM as the oldest contents of RAM are paged out to the hard drive to make room. The memory controller, by the way, is the chip that manages access to RAM as well as adapters that have had a few hardware memory addresses reserved for their communication with the processor.

Nevertheless, relying too much on virtual memory (check your page fault statistics in the Reliability and Performance Monitor) results in the entire system slowing down noticeably. An inexpensive and highly effective solution is to add physical memory to the system, thus reducing its reliance on virtual memory. More information on virtual memory and its configuration can be found in Chapter 13, "Operating System Basics."

Another type of memory common in PCs is *cache memory*, which is small and fast and logically sits between the CPU and RAM. Cache is a very fast form of memory forged from static RAM, which is discussed in detail in the section "Understanding Memory" later in this chapter. Cache improves system performance by predicting what the CPU will ask for next and prefetching this information before being asked. This paradigm allows the cache to be smaller in size than the RAM itself. Only the most recently used data and code or that which is expected to be used next is stored in cache.

You'll see three different cache designations:

Level 1 Cache *L1 cache* is the smallest and fastest, and it's on the processor die itself. In other words, it's an integrated part of the manufacturing pattern that's used to stamp the processor pathways into the silicon chip. You can't get any closer to the processor than that.

Though the definition of L1 cache has not changed much over the years, the same is not true for other cache levels. L2 and L3 cache used to be on the motherboard but now have moved on-die in most processors as well. The biggest differences are the speed and whether they are shared.

Level 2 Cache *L2 cache* is larger but a little slower than L1 cache. For processors with multiple cores, each core will generally have its own dedicated L1 and L2 caches. A few processors share a common L2 cache among the cores.

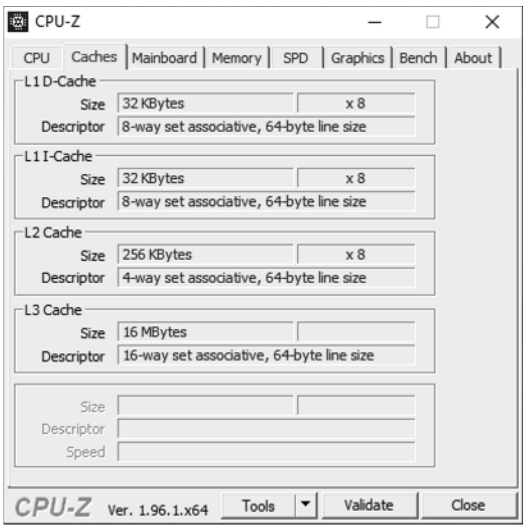
Level 3 Cache *L3 cache* is larger and slower than L1 or L2, and is usually shared among all processor cores.

The typical increasing order of capacity and distance from the processor die is L1 cache, L2 cache, L3 cache, RAM, and HDD/SSD (hard disk drive and solid-state drive—more on these in Chapter 2). This is also the typical decreasing order of speed. The following list includes representative capacities of these memory types. The cache capacities are for each core of the 10th generation Intel Core i7 processor. The other capacities are simply modern examples.

- L1 cache—80 KB (32 KB for instructions and 48 KB for data)
- L2 cache—512 KB
- L3 cache—8–16 MB
- RAM—16–256 GB
- HDD/SSD—100s of GB to several TB

One way to find out how much cache your system has is to use a utility such as CPU-Z, as shown in Figure 1.9. CPU-Z is freeware that can show you the amount of cache, processor name and number, motherboard and chipset, and memory specifications. It can be found at `www.cpuid.com`.

FIGURE 1.9 Cache in a system



NOTE

Figure 1.9 shows L1D and L1I caches. Many CPUs will split L1 cache into cache for data (L1D) or instructions (L1I). It's highly unlikely you'll be tested on this, but it's interesting to know.

Central Processing Unit and Processor Socket

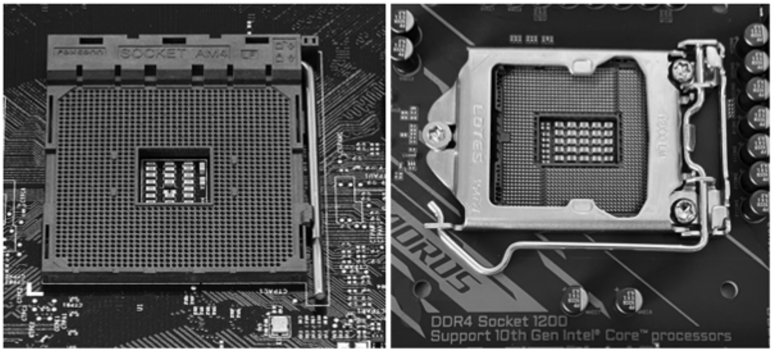
The "brain" of any computer is the *central processing unit (CPU)*. There's no computer without a CPU. There are many different types of processors for computers—so many, in fact, that you will learn about them later in this chapter in the section "Understanding Processors."

Typically, in today's computers, the processor is the easiest component to identify on the motherboard. It is usually the component that has either a fan or a heat sink (usually both) attached to it (as shown in Figure 1.10). These devices are used to draw away and disperse the heat that a processor generates. This is done because heat is the enemy of microelectronics. Today's processors generate enough heat that, without heat dispersal, they would permanently damage themselves and the motherboard in a matter of minutes, if not seconds.

FIGURE 1.10 Two heat sinks, one with a fan

CPU sockets are almost as varied as the processors that they hold. Sockets are basically flat and have several columns and rows of holes or pins arranged in a square, as shown in Figure 1.11. The left socket is known as Socket AM4, made for AMD processors such as the Ryzen, and has holes to receive the pins on the CPU. This is known as a *pin grid array (PGA)* arrangement for a CPU socket. The holes and pins are in a row/column orientation, an array of pins. The right socket is known as LGA 1200, and there are spring-loaded pins in the socket and a grid of lands on the CPU. The *land grid array (LGA)* is a newer technology that places the delicate pins (yet more sturdy than those on chips) on the cheaper motherboard instead of on the more expensive CPU, opposite to the way that the aging PGA does. The device with the pins has to be replaced if the pins become too damaged to function. The PGA and LGA are mentioned again later in this chapter in the section "Understanding Processors."

FIGURE 1.11 CPU socket examples



Modern CPU sockets have a mechanism in place that reduces the need to apply considerable force to the CPU to install a processor, which was necessary in the early days of personal computing. Given the extra surface area on today's processors, excessive pressure applied in the wrong manner could damage the CPU packaging, its pins, or the motherboard itself. For CPUs based on the PGA concept, *zero insertion force (ZIF)* sockets are exceedingly popular. ZIF sockets use a plastic or metal lever on one of the two lateral edges to lock or release the mechanism that secures the CPU's pins in the socket. The CPU rides on the mobile top portion of the socket, and the socket's contacts that mate with the CPU's pins are in the fixed bottom portion of the socket. The image of Socket AM4 shown on the left in Figure 1.11 illustrates the ZIF locking mechanism at the right edge of the socket.

For processors based on the LGA concept, a socket with a different locking mechanism is used. Because there are no receptacles in either the motherboard or the CPU, there is no opportunity for a locking mechanism that holds the component with the pins in place. LGA-compatible sockets, as they're called despite the misnomer, have a lid that closes over the CPU and is locked in place by an L-shaped arm that borders two of the socket's edges. The nonlocking leg of the arm has a bend in the middle that latches the lid closed when the other leg of the arm is secured. The right image in Figure 1.11 shows an LGA socket with no CPU installed and the locking arm secured over the lid's tab, along the bottom edge.

Listing out all the desktop PC socket types you might encounter would take a long time. Instead, we'll give you a sampling of some that you might see. The first thing you might notice is that sockets are made for Intel or AMD processors, but not both. Keep that compatibility in mind when replacing a motherboard or a processor. Make sure that the processor and motherboard were designed for each other (even within the Intel or AMD families); otherwise, they won't fit each other and won't work. Table 1.1 lists some common desktop socket/CPU relationships. Servers and laptops/tablets generally have different sockets altogether, although some CPU sockets will support processors designed for desktops or servers.

TABLE 1.1 Desktop PC socket types and the processors they support

Socket	Released	Type	Processors
LGA 1200	2020	LGA	Intel Comet Lake and Rocket Lake
Socket AM4	2017	PGA	AMD Ryzen 3, Ryzen 5, Ryzen 7, Ryzen 9, Athlon 200GE
Socket TR4	2017	LGA	AMD Ryzen Threadripper
LGA 2066 (Socket R4)	2017	LGA	Intel Skylake-X and Kaby Lake-X
LGA 1151 (Socket H4)	2015	LGA	Intel Skylake, Kaby Lake, and Coffee Lake
Socket FM2+	2014	PGA	AMD Kaveri and Godavari
Socket AM1	2014	PGA	AMD Athlon and Sempron
LGA 1150 (Socket H3)	2013	LGA	Intel Haswell, Haswell Refresh, and Broadwell
Socket FM2	2012	PGA	AMD Trinity

NOTE

Some legacy sockets, such as Socket 5 and Socket 7, supported both AMD and Intel platforms, but it's unlikely that you will see one in the wild, as they are over 25 years old now.

Real World Scenario

Playing the Name Game

Back in simpler times, Intel marketed only one or two processors at a time. For example, when the 80486DX was in the market, they didn't have much else. Then the Pentium came along, followed by the Celeron (which everyone knew was a stripped-down Pentium), and the Pentium M, which was the mobile version of the Pentium. The active roster of processors started to expand.

Now, there are a lot of active chips, including the Xeon and the Core series (i9, i7, i5, i3, and m3), and even the Pentium and Celeron are still kicking around. And with new innovation coming nearly every year, it's difficult to keep track of what's current.

In about 2013, Intel began to publicize its project code names a bit more to make it easier to keep track of the new technology bundles. No longer were techs able to remember that an i7 required a specific type of socket, but they needed to know *which* i7 they were dealing with. For example, a Broadwell-based (2014) i7 requires an LGA 1150 socket, whereas most Kaby Lake (2016) i7s use an LGA 1151. One pin makes a huge difference!

Odds are you aren't going to need to memorize all possible Intel chip and motherboard socket combinations for the exam or the real world. In the real world, the Internet's always there to remind you of which socket you need for a specific processor. To give you a flavor, though, Table 1.2 covers a few of Intel's recent CPU architecture codenames and the associated processor model numbers.

To make things even more interesting, you might hear references to the *n*th generation of the i5, i7, or i9 processor. For example, Rocket Lake is considered 11th generation. Intel codenames over the past several years have had some variation of a "Lake" name, such as Coffee Lake (2017), Comet Lake (2019), and Rocket Lake (2020). Perhaps the names keep it interesting for the engineers.

When you go to the parts store, though, you're probably not going to ask for a Rocket Lake i7. You'll ask for an i7, see what they have in stock—usually they will have the current generation and possibly the previous one, and then figure out what motherboard you need. (Or alternatively, based on your motherboard, you'll figure out which i7 you can use.)

What does all of this mean you need to do? First, get comfortable knowing that a new architecture may require a different motherboard. Second, feel free to memorize some of the more common combinations of motherboards and processors if you find it helpful, although you won't be tested on it. It may be handy in the real world to know that if you have an 11th gen Intel chip, you need a board with an LGA 1200 socket. (Or you can always Google it!)

TABLE 1.2 Select Intel desktop processors

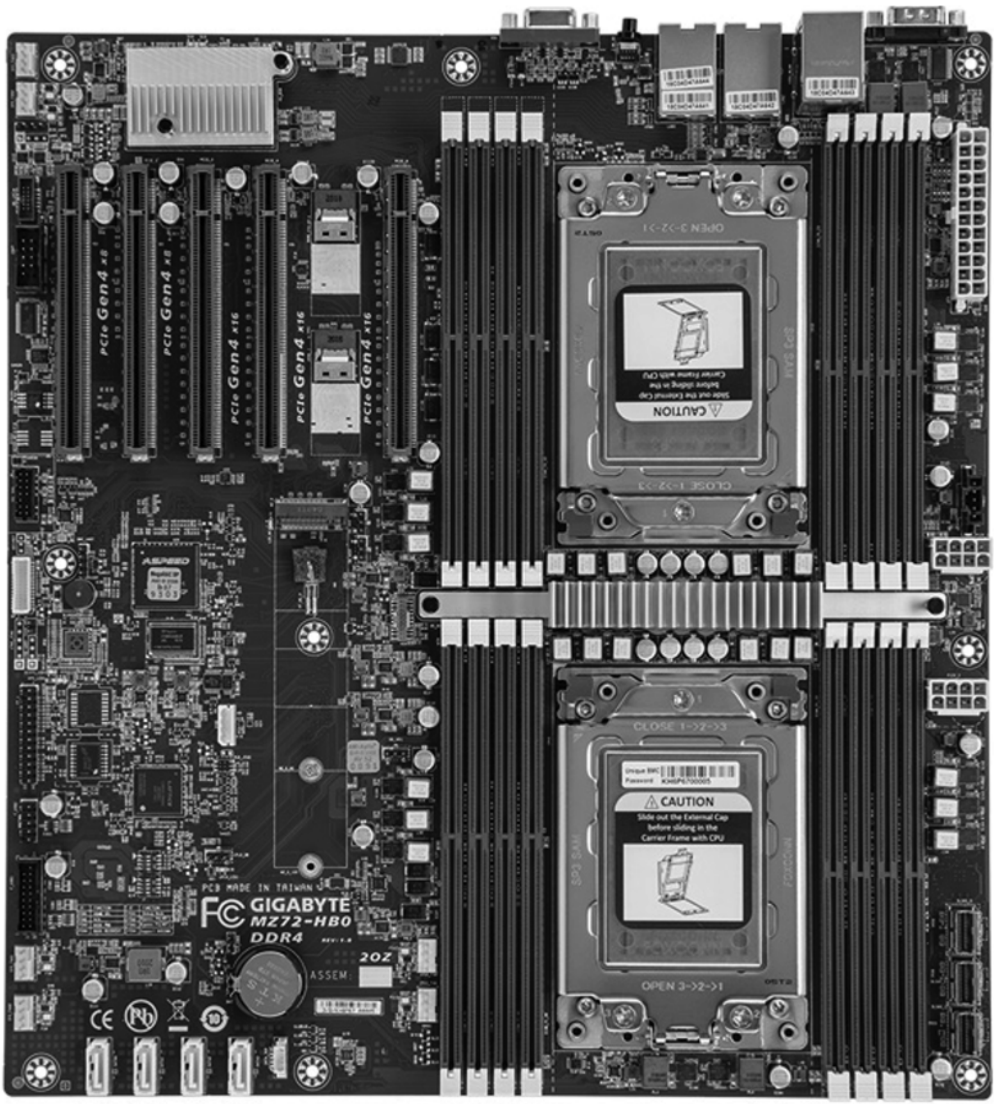
Name (Year)	Gen	Socket	Core i9	Core i7	Core i5	Core i3
Alder Lake (2021)	12th	LGA 1700	129xx	127xx	126xx	n/a
Rocket Lake (2020)	11th	LGA 1200	119xx	117xx	116xx	n/a
					115xx	
					114xx	
Comet Lake (2019)	10th	LGA 1200	109xx	107xx	106xx	103xx
					105xx	101xx
					104xx	

Multisocket and Server Motherboards

When it comes to motherboard compatibility, the two biggest things to keep in mind are the processor type and the case. If either of those are misaligned with what the motherboard supports, you're going to have problems.

Thus far, as we've talked about desktop motherboards and their CPU sockets, we have shown examples of boards that have just one socket. There are motherboards that have more than one CPU socket and conveniently, they are called *multisocket* (typically written as two words) motherboards. Figure 1.12 shows a two-socket motherboard made by GIGABYTE. The two CPU sockets are easily identifiable and note that each CPU socket has eight dedicated memory slots.

FIGURE 1.12 GIGABYTE multisocket motherboard

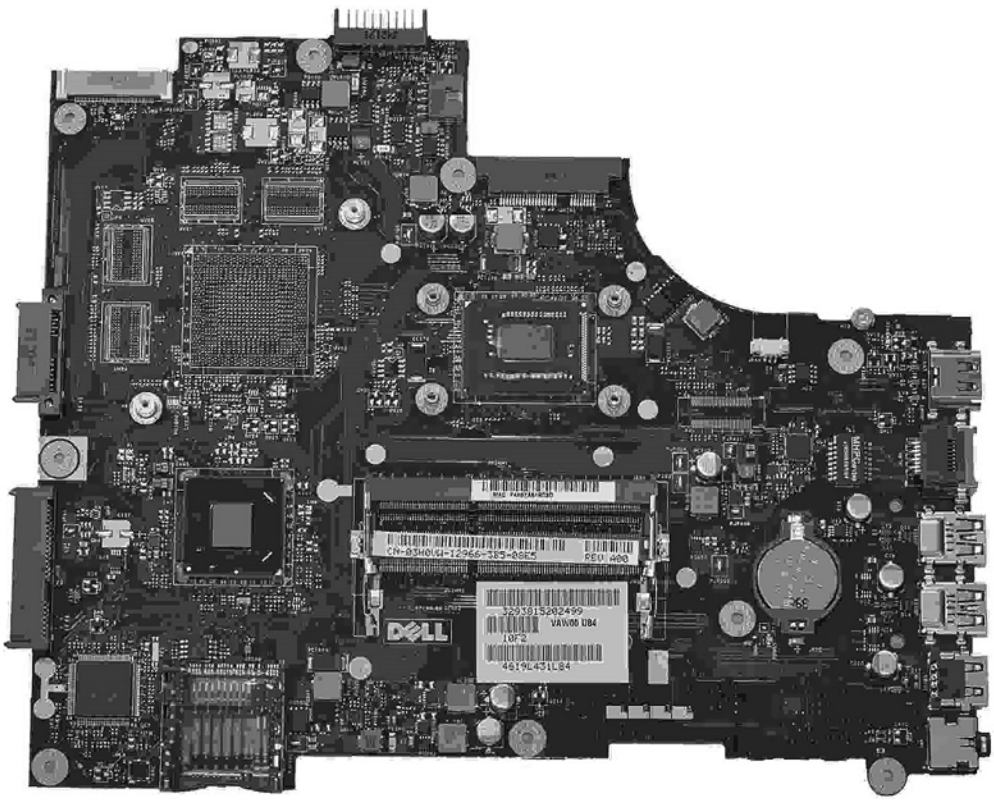


Trying to categorize server motherboards can be a bit challenging. Servers are expected to do a lot more work than the average PC, so it makes sense that servers need more powerful hardware. Servers can, and quite often do, make do with a single processor on a "normal" PC motherboard. At the same time, there are motherboards designed specifically for servers that support multiple processors (two and four sockets are common) and have expanded memory and networking capabilities as well. Further, while server motherboards are often ATX-sized, many server manufacturers create custom boards to fit inside their chassis. Regardless, multisocket and server motherboards will generally use the same CPU sockets that other motherboards use.

Mobile Motherboards

In small mobile devices, space is at a premium. Some manufacturers will use standard small-factor motherboards, but most create their own boards to fit inside specific cases. An example of an oddly shaped Dell laptop motherboard is shown in Figure 1.13. When replacing a laptop motherboard, you almost always need to use one from the exact same model, otherwise it won't fit inside the case.

FIGURE 1.13 Dell laptop motherboard

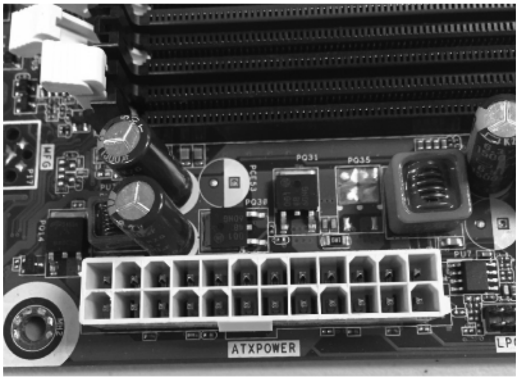


Nearly all laptop processors are soldered onto the motherboard, so you don't have to worry about CPU socket compatibility. If the CPU dies, you replace the entire motherboard. We will cover laptop components more extensively in Chapter 9, "Laptop and Mobile Device Hardware."

Power Connectors

In addition to these sockets and slots on the motherboard, a special connector (the 24-pin white block connector shown in Figure 1.14) allows the motherboard to be connected to the power supply to receive power. This connector is where the ATX power connector (mentioned in Chapter 2 in the section "Understanding Power Supplies") plugs in.

FIGURE 1.14 An ATX power connector on a motherboard



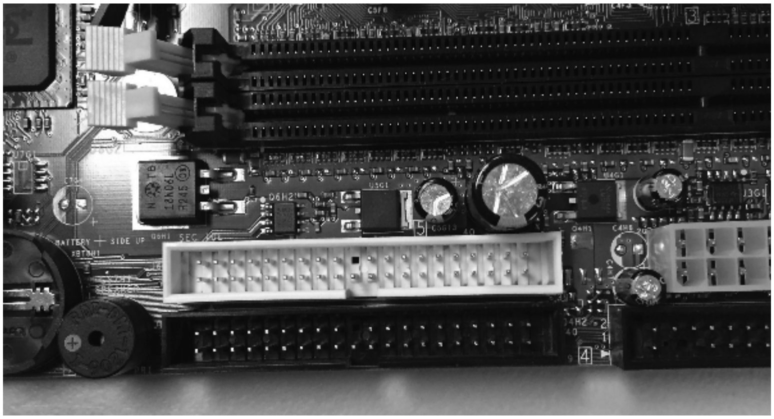
Onboard Nonvolatile Storage Connectors

Nearly all users store data, and the most widely used data storage device is a hard drive. Hard drives are great because they store data even when the device is powered off, which explains why they are sometimes referred to as nonvolatile storage. There are multiple types of hard drives, and we'll get into them in more detail in Chapter 2. Of course, those drives need to connect to the motherboard, and that's what we'll cover here.

Integrated Drive Electronics/Parallel ATA

At one time, *integrated drive electronics (IDE)* drives were the most common type of hard drive found in computers. Though often thought of in relation to hard drives, IDE was much more than a hard drive interface; it was also a popular interface for many other drive types, including optical drives and tape drives. Today, we call it IDE Parallel ATA (PATA) and consider it to be a legacy technology. Figure 1.15 shows two PATA interfaces; you can see that one pin in the center is missing (as a key) to ensure that the cable gets attached properly. The industry now favors Serial ATA instead.

FIGURE 1.15 Two PATA hard drive connectors



Serial ATA

Serial ATA (SATA) began as an enhancement to the original ATA specifications, also known as *IDE* and, today, *PATA*. Technology is proving that the orderly progression of data in a single-file path is superior to placing multiple bits of data in parallel and trying to synchronize their transmission to the point that all of the bits arrive simultaneously. In other words, if you can build faster transceivers, serial transmissions are simpler to adapt to the faster rates than are parallel transmissions.

The first version of SATA, known as SATA 1.5 Gbps (and also by the less-preferred terms SATA I and SATA 150), used an 8b/10b-encoding scheme that requires 2 non-data overhead bits for every 8 data bits. The result is a loss of 20 percent of the rated bandwidth. The silver lining, however, is that the math becomes quite easy. Normally, you have to divide by 8 to convert bits to bytes. With 8b/10b encoding, you divide by 10. Therefore, the 150 MBps throughput for which this version of SATA was nicknamed is easily derived as 1/10 of the 1.5 Gbps transfer rate. The original SATA specification also provided for hot swapping at the discretion of the motherboard and drive manufacturers.

Similar math works for SATA 3 Gbps, tagged as SATA II and SATA 300, and SATA 6 Gbps, which you might hear called SATA III or SATA 600. Note that each subsequent version doubles the throughput of the previous version. Figure 1.16 shows four SATA headers on a motherboard that will receive the data cable. Note that identifiers silkscreened onto motherboards often enumerate such headers. The resulting numbers are not related to the SATA version that the header supports. Instead, such numbers serve to differentiate headers from one another and to map to firmware identifiers, often visible within the BIOS configuration utility.

Another version of SATA that you will see is *external SATA (eSATA)*. As you might expect based upon the name, this technology was developed for devices that reside outside of the case, not inside it. Many motherboards have an eSATA connector built in. If not, you can buy an expansion card that has eSATA ports and plugs into internal SATA connectors. Figure 1.17 shows an example of how the two ports are different. Finally, SATA and eSATA standards are compatible. In other words, SATA 6 Gbps equals eSATA 6 Gbps.

FIGURE 1.16 Four SATA headers

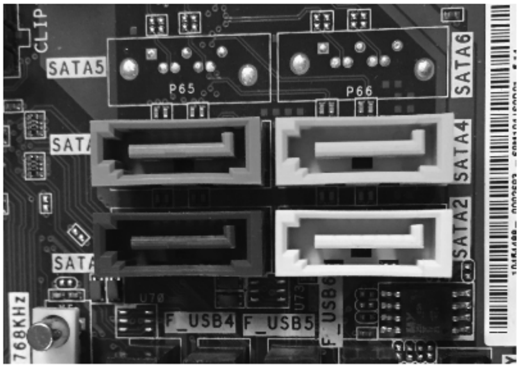
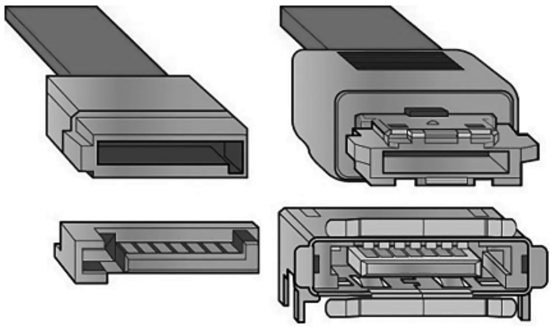


FIGURE 1.17 SATA (left) and eSATA (right) cables and ports



> **NOTE** SATA and eSATA ports do not provide power, like USB does. Therefore, when connecting SATA and eSATA drives, a separate power connection is required in addition to the data cable.

M.2

The most recent development in expansion connections is *M.2* (pronounced “M dot 2”). So far it’s primarily used for hard drives, but other types of devices, such as Wi-Fi, Bluetooth, Global Positioning System (GPS), and near-field communication (NFC) adapters are built for M.2 as well. We will cover M.2 in more depth in Chapter 2, thanks to how important it is to storage solutions.

It’s important to call out that M.2 is a form factor, not a bus standard. The form factor supports existing SATA, USB, and PCIe buses. This means that if you hook up a SATA device to an M.2 slot (with the appropriate connector), the device speed will be regulated by SATA standards. Figure 1.18 shows two M.2 connectors on a motherboard.

FIGURE 1.18 Two M.2 slots

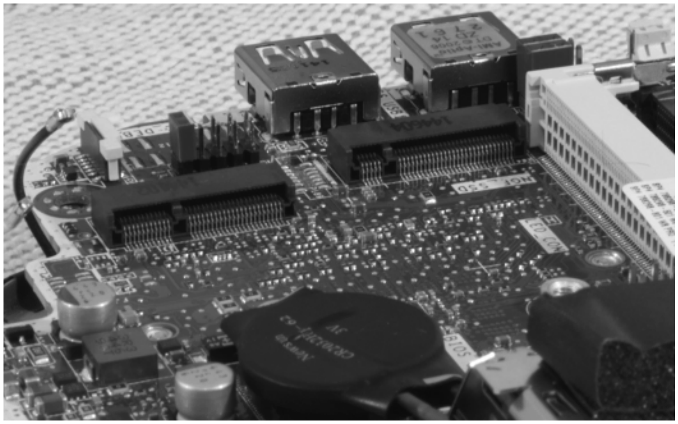


Photo credit: Andrew Cunningham/Ars Technica

NOTE: M.2 is closely associated with hard drives, and the M.2 form factor is also in CompTIA A+ exam objective 3.3, "Given a scenario, select and install storage devices." Because of that, we will go into more depth about M.2 and hard drive data transfer speeds in Chapter 2 when we discuss solid-state drives (SSDs).

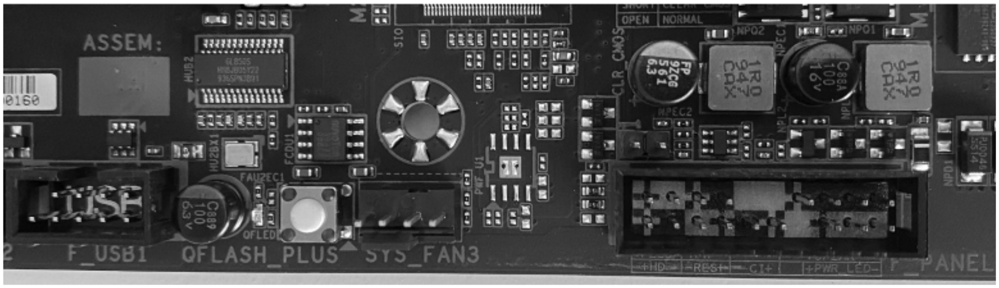
Motherboard Headers

From the time of the very first personal computer, there has been a minimum expectation as to the buttons and LEDs that should be easily accessible to the user. At first, they generally appeared on the front of the case. In today's cases, buttons and LEDs have been added and placed on the top of the case or on a beveled edge between the top and the front. They have also been left on the front or have been used in a combination of these locations. These buttons and lights, as well as other external connectors, plug into the motherboard through a series of pins known as *headers*. Examples of items that are connected using a header include:

- Power button
- Power light
- Reset button
- Drive activity lights
- Audio jacks
- USB ports

Headers for different connections are often spread throughout different locations on the motherboard—finding the right one can sometimes be a frustrating treasure hunt. Other headers are grouped together. For example, most of the headers for the items on the front or top panel of the case are often co-located. The purpose for the header will be printed on the motherboard, and while that may tell you what should connect there, it often lacks detail in how it should be connected. The motherboard manufacturer's website is a good place to go if you need a detailed diagram or instructions. Figure 1.19 shows several headers on a motherboard. On the left is a USB header, then a system fan header in the center, and a block of front panel headers on the right, including the hard drive light, reset button, chassis intrusion detector, and power light.

FIGURE 1.19 Motherboard headers



Power Button and Light

Users expect a *power button* to use to turn the computer on. (These were on the side or back of very early PCs.) The soft power feature available through the front power button, which is no more than a relay, allows access to multiple effects through the contact on the motherboard, based on how long the button is pressed. These effects can be changed through the BIOS or operating system. Users also expect a *power light*, often a green LED, to assure them that the button did its job.

Reset Button

The *reset button* appeared as a way to reboot the computer from a cold startup point without removing power from the components. Keeping the machine powered tends to prolong the life of the electronics affected by power cycling. Pressing the reset button also gets around software lockups because the connection to the motherboard allows the system to restart from the hardware level. One disadvantage to power cycling is that certain circuits, such as memory chips, might need time to drain their charge for the reboot to be completely successful. This is why there is always a way to turn the computer off as well.

Drive Activity Light

In the early days of personal computing, the hard disk drive's LED had to be driven by the drive itself. Before long, the motherboard was equipped with drive headers, so adding pins to drive the *drive activity light* was no issue. These days, all motherboards supply this connectivity. The benefit of having one LED for all internal drives is that all the drives are represented on the front panel when only one LED is provided. The disadvantage might be that you cannot tell which drive is currently active. This tends to be a minor concern because you often know which drive you've accessed. If you haven't intentionally accessed any drive, it's likely the drive that holds the operating system or virtual-memory swap file is being accessed by the system itself. In contrast, external drives with removable media, such as optical drives, supply their own activity light on their faceplate.

Audio Jacks

Early generations of optical drives had to have a special cable attached to the rear of the drive, which was then attached to the sound card if audio CDs were to be heard through the speakers attached to the sound card. Sound emanating from a CD-ROM running an application, such as a game, did not have to take the same route and could travel through the same path from the drive as general data. The first enhancement to this arrangement came in the form of a front 3.5 mm jack on the drive's faceplate that was intended for headphones but could also have speakers connected to it. The audio that normally ran across the special cable was rerouted to the front jack when something was plugged into it.

Many of today's motherboards have 10-position pin headers designed to connect to standardized front-panel audio modules. Some of these modules have legacy AC'97 analog ports on them, whereas others have high-definition (HD) audio connections. Motherboards that accommodate both have a BIOS setting that enables you to choose which header you want to activate, with the HD setting most often being the default.

USB Ports

So many temporarily attached devices feature USB connectivity, such as USB keys (flash drives) and cameras, that front-panel connectivity is a must. Finding your way to the back of the system unit for a brief connection is hardly worth the effort in some cases. For many years, motherboards have supplied one or more 10-position headers for internal connectivity of front-panel USB ports. Because this header size is popular for many connectors, only 9 positions tend to have pins protruding, while the 10th position acts as a key, showing up in different spots for each connector type to discourage the connection of the wrong cable. Figure 1.20 shows USB headers on a motherboard. The labels "USB56" and "USB78" indicate that one block serves ports 5 and 6, while the other serves ports 7 and 8, all of which are arbitrary, based on the manufacturer's numbering convention. In each, the upper left pin is "missing," which is the key.

FIGURE 1.20 Two motherboard USB headers

BIOS/UEFI and the POST Routine

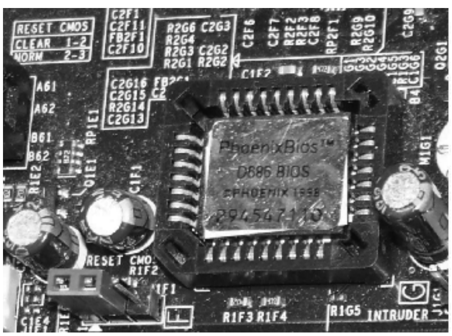
Firmware is the name given to any software that is encoded in hardware, usually a read-only memory (ROM) chip, and it can be run without extra instructions from the operating system. Most computers, large printers, and devices with no operating system use firmware in some sense. The best example of firmware is a computer's *Basic Input/Output System (BIOS)*, which is burned into a chip. Also, some expansion cards, such as SCSI cards and graphics adapters, use their own firmware utilities for setting up peripherals.

The BIOS chip, also referred to as the ROM BIOS chip, is one of the most important chips on the motherboard. This special memory chip contains the BIOS system software that boots the system and allows the operating system to interact with certain hardware in the computer in lieu of requiring a more complex device driver to do so. The BIOS chip is easily identified: If you have a brand-name computer, this chip might have on it the name of the manufacturer and usually the word *BIOS*. For clones, the chip usually has a sticker or printing on it from one of the major BIOS manufacturers (AMI, Phoenix, Award, Winbond, and others). On later motherboards, the BIOS might be difficult to identify or it might even be integrated into the Southbridge, but the functionality remains regardless of how it's implemented.

The successor to the BIOS is the *Unified Extensible Firmware Interface (UEFI)*. The extensible features of the UEFI allow for the support of a vast array of systems and platforms by allowing the UEFI access to system resources for storage of additional modules that can be added at any time. In the following section, you'll see how a security feature known as Secure Boot would not be possible with the classic BIOS. It is the extensibility of the UEFI that makes such technology feasible.

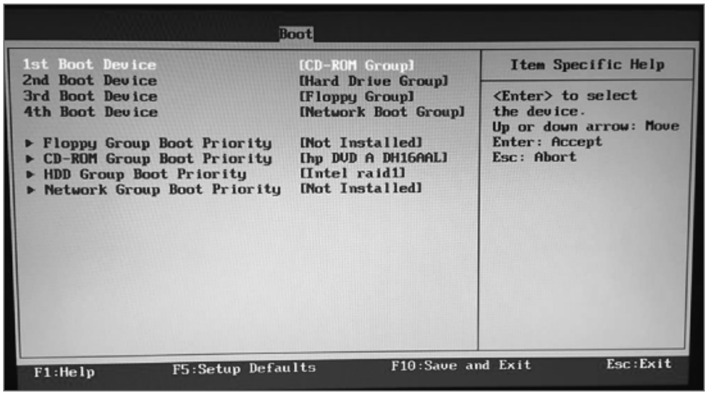
Figure 1.21 gives you an idea of what a modern BIOS/UEFI chip might look like on a motherboard. Despite the 1998 copyright on the label, which refers only to the oldest code present on the chip, this particular chip can be found on motherboards produced as late as 2009. Notice also the Reset CMOS jumper at the lower left and its configuration silkscreen at the upper left. You might use this jumper to clear the CMOS memory, discussed shortly, when an unknown password, for example, is keeping you out of the BIOS/UEFI configuration utility. The jumper in the photo is in the clear position, not the normal operating position. System bootup is typically not possible in this state.

FIGURE 1.21 A BIOS chip on a motherboard



At a basic level, the BIOS/UEFI controls system boot options such as the sequence of drives from which it will look for operating system boot files. The boot sequence menu from a BIOS/UEFI is shown in Figure 1.22. Other interface configuration options will be available too, such as enabling or disabling integrated ports or an integrated video card. A popular option on corporate computers is to disable the USB ports, which can increase security and decrease the risk of contracting a virus.

FIGURE 1.22 BIOS boot sequence



Most BIOS/UEFI setup utilities have more to offer than a simple interface for making selections and saving the results. For example, these utilities often offer diagnostic routines that you can use to have the BIOS/UEFI analyze the state and quality of the same components that it inspects during bootup, but at a much deeper level.

Consider the scenario where a computer is making noise and overheating. You can use the BIOS/UEFI configuration utility to access built-in diagnostics to check the rotational speed of the motherboard fans. If the fans are running slower than expected, the noise could be related to the bearings of one or more fans, causing them to lose speed and, thus, cooling capacity.

There is often also a page within the utility that gives you access to such bits of information as current live readings of the temperature of the CPU and the ambient temperature of the interior of the system unit. On such a page, you can set the temperature at which the BIOS/UEFI sounds a warning tone and the temperature at which the BIOS/UEFI shuts down the system to protect it. You can also monitor the instantaneous fan speeds, bus speeds, and voltage levels of the CPU and other vital landmarks to make sure that they are all within acceptable ranges. You might also be able to set a lower fan speed threshold at which the system warns you. In many cases, some of these levels can be altered to achieve such phenomena as overclocking, which is using the BIOS/UEFI to set the system clock higher than what the CPU is rated for, or undervolting, which is lowering the voltage of the CPU and RAM, which reduces power consumption and heat production.

BIOS/UEFI Security and Encryption

The BIOS/UEFI has always played a role in system security. Since the early days of the personal computer, the BIOS allowed the setting of two passwords—the user (or boot) password and the supervisor/administrator, or access, password. The boot password is required to leave the initial power-on screens and begin the process of booting an operating system. The administrator password is required before entering the BIOS/UEFI configuration utility. It is always a good idea to set the administrator password, but the boot password should not be set on public systems that need to boot on their own, in case of an unforeseen power cycle.

In more recent years, the role of the BIOS/UEFI in system security has grown substantially. BIOS/UEFI security has been extended to a point where the operating system is ready to take it over. The BIOS/UEFI was a perfect candidate to supervise security and integrity in a platform-independent way. Coupled with the *Trusted Platform Module (TPM)*, a dedicated security coprocessor, or cryptoprocessor, the BIOS can be configured to boot the system only after authenticating the boot device. This authentication confirms that the hardware being booted to has been tied to the system containing the BIOS/UEFI and TPM, a process known as *sealing*. Sealing the devices to the system also prohibits the devices from being used after removing them from the system. For further security, the keys created can be combined with a PIN or password that unlocks their use or with a USB flash drive that must be inserted before booting.

Microsoft's BitLocker uses the TPM to encrypt the entire drive. Normally, only user data can be encrypted, but BitLocker encrypts operating-system files, the Registry, the hibernation file, and so on, in addition to those files and folders that file-level encryption secures. If any changes have occurred to the Windows installation, the TPM does not release the keys required to decrypt and boot to the secured volume. TPM is configured in Windows under Start ➢ Settings ➢ Update & Security ➢ Windows Security ➢ Device security, as shown in Figure 1.23.

Most motherboards come with a TPM chip installed, but if they don't, it's not possible to add one. In those situations, you can enable the same functionality by using a *hardware security module (HSM)*. An HSM is a security device that can manage, create, and securely store encryption keys—it enables users to safely encrypt and decrypt data. An HSM can take a few different forms. The simplest is a USB or PCIe device that plugs into a system. It could be set up for file encryption and decryption, required for the computer to boot, or both. For large-scale solutions, HSM-enabled servers can provide crypto services to an entire network.

When a certain level of UEFI is used, the system firmware can also check digital signatures for each boot file it uses to confirm that it is the approved version and has not been tampered with. This technology is known as *Secure Boot*. An example of a BIOS/UEFI's boot security screen is shown in Figure 1.24. The boot files checked include option ROMs (defined in the following section), the boot loader, and other operating-system boot files. Only if the signatures are valid will the firmware load and execute the associated software.

FIGURE 1.23 Windows TPM configuration screen

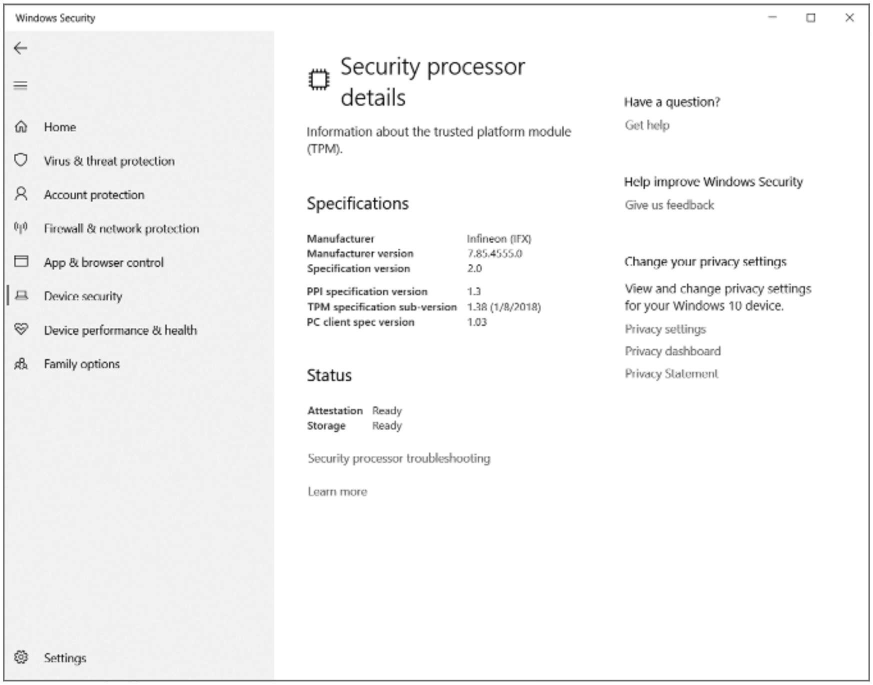
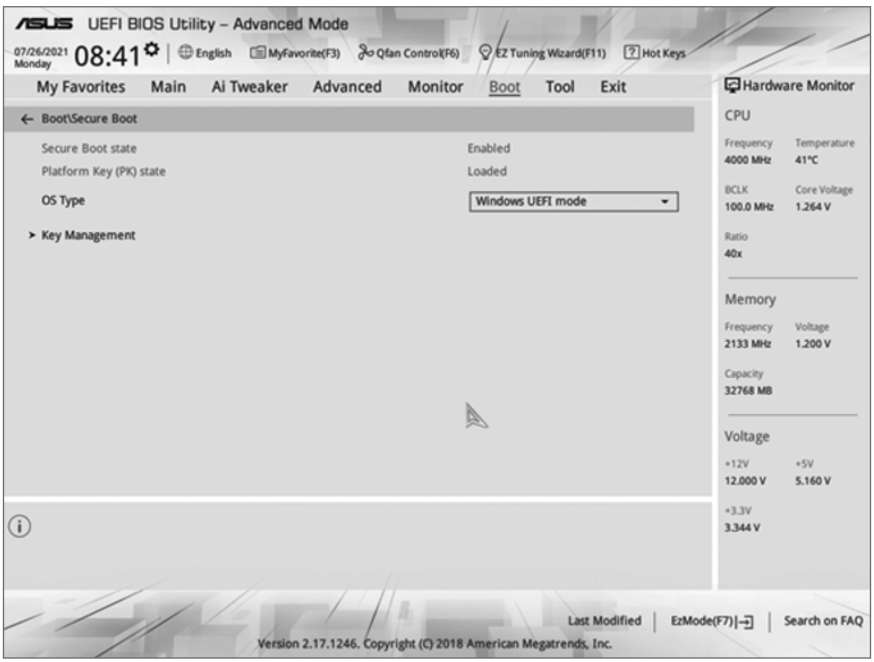


FIGURE 1.24 Secure boot in UEFI



The problem can now arise that a particular operating system might not be supported by the database of known-good signatures stored in the firmware. In such a situation, the system manufacturer can supply an extension that the UEFI can use to support that operating system—a task not possible with traditional BIOS-based firmware.

NOTE

LoJack for Laptops is a UEFI-enabled security system developed by Absolute Software. It performs similarly to other UEFI security methods already discussed, such as drive encryption, but adds a few features. For example, LoJack can remotely track stolen laptops as well as lock and delete certain files.

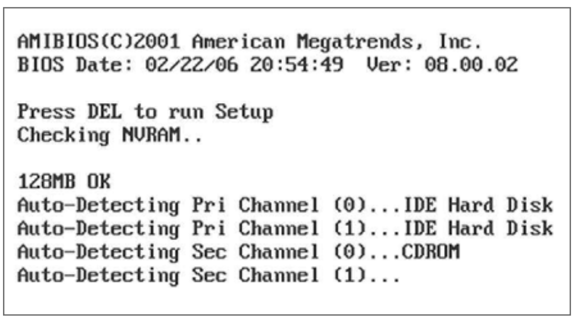
Some BIOS firmware can monitor the status of a contact on the motherboard for intrusion detection. If the feature in the BIOS is enabled and the sensor on the chassis is connected to the contact on the motherboard, the removal of the cover will be detected and logged by the BIOS. This can occur even if the system is off, thanks to the CMOS battery. At the next bootup, the BIOS will notify you of the intrusion. No notification occurs over subsequent boots unless additional intrusion is detected.

POST

A major function of the BIOS/UEFI is to perform a process known as a *power-on self-test (POST)*. POST is a series of system checks performed by the system BIOS/UEFI and other high-end components, such as the SCSI BIOS and the video BIOS, known collectively as *option ROMs*. Among other things, the POST routine verifies the integrity of the BIOS/UEFI itself. It also verifies and confirms the size of primary memory. During POST, the BIOS also analyzes and catalogs other forms of hardware, such as buses and boot devices, as well as manages the passing of control to the specialized BIOS/UEFI routines mentioned earlier. The BIOS/UEFI is responsible for offering the user a key sequence to enter the configuration routine as POST is beginning. Finally, once POST has completed successfully, the BIOS/UEFI selects the boot device highest in the configured boot order and executes the master boot record (MBR) or similar construct on that device so that the MBR can call its associated operating system's boot loader and continue booting up.

The POST process can end with a beep code or displayed code that indicates the issue discovered. Each BIOS/UEFI publisher has its own series of codes that can be generated. Figure 1.25 shows a simplified POST display during the initial boot sequence of a computer.

FIGURE 1.25 An example of a system POST screen



Flashing the System BIOS/UEFI

If ever you find that a hardware upgrade to your system is not recognized, even after the latest and correct drivers have been installed, perhaps a BIOS/UEFI upgrade, also known as *flashing the BIOS*, is in order. Only certain hardware benefits from a BIOS/UEFI upgrade, such as drives and a change of CPU or RAM types. Very often, this hardware is recognized immediately by the BIOS/UEFI and has no associated driver that you must install. So, if your system doesn't recognize the new device, and there's no driver to install, the BIOS/UEFI is a logical target.

Let's be clear about the fact that we are not talking about entering the BIOS/UEFI setup utility and making changes to settings and subsequently saving your changes before exiting and rebooting. What we are referring to here is a replacement of the burned-in code within the BIOS itself. You might even notice after the upgrade that the BIOS setup utility looks different or has different pages and entries than before.

On older systems and certain newer ones, a loss of power during the upgrade results in catastrophe. The system becomes inoperable until you replace the BIOS/UEFI chip, if possible, or the motherboard itself. Most new systems, however, have a fail-safe or two. This could be a portion of the BIOS/UEFI that does not get flashed and has just enough code to boot the system and access the upgrade image. It could be a passive section to which the upgrade is installed and switched to only if the upgrade is successful. Sometimes this is controlled on screen. At other times, there may be a mechanism, such as a jumper, involved in the recovery of the BIOS/UEFI after a power event occurs. The safest bet is to make sure that your laptop has plenty of battery power and is connected to AC power or your desktop is connected to an uninterruptible power supply (UPS).

In all cases, if you think you need a BIOS/UEFI upgrade, do not consult the BIOS/UEFI manufacturer. Instead, go to the motherboard or system manufacturer and check its website. The motherboard or system manufacturer vendors have personalized their BIOS/UEFI code after licensing it from the BIOS/UEFI publisher. The vendor will give you access to the latest code as well as the appropriate flashing utility for its implementation.

CMOS and CMOS Battery

Your PC has to keep certain settings when it's turned off and its power cord is unplugged:

- Date
- Time
- Hard drive/optical drive configuration
- Memory
- CPU settings, such as overclocking
- Integrated ports (settings as well as enable/disable)

- Boot sequence
- Power management
- Virtualization support
- Security (passwords, Trusted Platform Module settings, LoJack)

Consider a situation where you added a new graphics adapter to your desktop computer, but the built-in display port continues to remain active, prohibiting the new interface from working. The solution might be to alter your BIOS/UEFI configuration to disable the internal graphics adapter, so that the new one will take over. Similar reconfiguration of your BIOS/UEFI settings might be necessary when overclocking—or changing the system clock speed—is desired, or when you want to set BIOS/UEFI-based passwords or establish TPM-based whole-drive encryption, as with Microsoft's BitLocker. While not so much utilized today, the system date and time can be altered in the BIOS/UEFI configuration utility of your system; once, in the early days of personal computing, the date and time actually might have needed to be changed this way.

Your PC keeps these settings in a special memory chip called the *complementary metal oxide semiconductor (CMOS)* memory chip. Actually, CMOS (usually pronounced *see-moss*) is a manufacturing technology for integrated circuits. The first commonly used chip made from CMOS technology was a type of memory chip, the memory for the BIOS/UEFI. As a result, the term *CMOS* stuck and is the accepted name for this memory chip.

The BIOS/UEFI starts with its own default information and then reads information from the CMOS, such as which hard drive types are configured for this computer to use, which drive(s) it should search for boot sectors, and so on. Any overlapping information read from the CMOS overrides the default information from the BIOS/UEFI. A lack of corresponding information in the CMOS does not delete information that the BIOS knows natively. This process is a merge, not a write-over. CMOS memory is usually *not* upgradable in terms of its capacity and might be integrated into the BIOS/UEFI chip or the Southbridge.

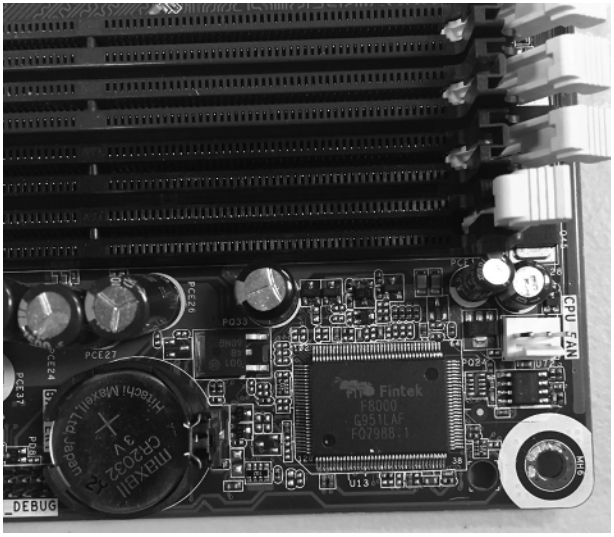
NOTE

Although there are technical differences, in the real world the terms BIOS/UEFI and CMOS (and BIOS/UEFI chip and CMOS chip) are used interchangeably. You will hear someone say, "Did you check the BIOS/UEFI?" or, "Did you check CMOS settings?" and they're referring to the same thing. Even though it's an oversimplification, most people feel comfortable with thinking about the BIOS/UEFI as the firmware that controls the boot process, whereas CMOS is the chip itself. It's incredibly unlikely that the A+ exam or any person is going to ask you to differentiate between the two, unless you get a job focusing specifically on those technologies.

To keep its settings, integrated circuit-based memory must have power constantly. When you shut off a computer, anything that is left in this type of memory is lost forever. The CMOS manufacturing technology produces chips with very low power requirements. As a result, today's electronic circuitry is more susceptible to damage from electrostatic discharge (ESD). Another ramification is that it doesn't take much of a power source to keep CMOS chips from losing their contents.

To prevent CMOS from losing its rather important information, motherboard manufacturers include a small battery called the *CMOS battery* to power the CMOS memory, shown in the bottom-left corner of Figure 1.26. The batteries come in different shapes and sizes, but they all perform the same function. Most CMOS batteries look like large watch batteries or small cylindrical batteries. Today's CMOS batteries are most often of a long-life, non-rechargeable lithium chemistry.

FIGURE 1.26 CMOS battery



NOTE

There's a lot to know about motherboards—they are complex devices that connect everything together inside the computer! For the A+ exam, remember that you will need to be familiar with the following concepts:

- Form factors, such as ATX and ITX
- Connector types including PCI, PCIe, power, SATA, eSATA, M.2, and headers
- Compatibility concerns such as AMD and Intel CPU sockets and different board types such as server, multisocket, desktop, and mobile
- BIOS/UEFI boot options, USB permissions, TPM, fan configurations, secure boot, boot passwords
- Security and encryption using TPM and HSM

Understanding Processors

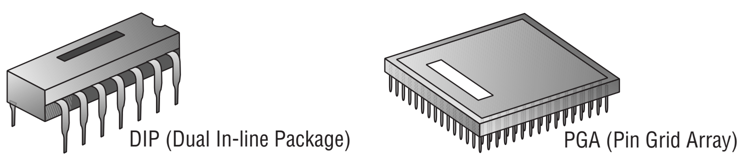
Now that you've learned the basics of the motherboard, you need to learn about the most important component on the motherboard: the CPU. The role of the CPU, or central processing unit, is to control and direct all the activities of the computer using both external and internal buses. From a technical perspective, the job of the CPU is to process, or do math, on large strings of binary numbers—0s and 1s. It is a processor chip consisting of an array of *millions* of transistors. Intel and Advanced Micro Devices, Inc. (AMD) are the two largest PC-compatible CPU manufacturers. Their chips were featured in Table 1.1 during the discussion of the sockets into which they fit.

Today's AMD and Intel CPUs should be compatible with every PC-based operating system and application in the market. It's possible that you could run into an app that doesn't work quite right on an AMD chip, but those cases are exceedingly rare. From a compatibility standpoint, the most important thing to remember is that the motherboard and processor need to be made for each other. The rest of the hardware plugs into the motherboard and will be CPU brand agnostic.

NOTE: The term *chip* has grown to describe the entire package that a technician might install in a socket. However, the word originally denoted the silicon wafer that is generally hidden within the carrier that you actually see. The external pins that you see are structures that can withstand insertion into a socket and are carefully threaded from the wafer's minuscule contacts. Just imagine how fragile the structures must be that you *don't* see.

Older CPUs are generally square, with contacts arranged in a pin grid array (PGA). Prior to 1981, chips were found in a rectangle with two rows of 20 pins known as a *dual in-line package (DIP)*—see Figure 1.27. There are still integrated circuits that use the DIP form factor; however, the DIP form factor is no longer used for PC CPUs. Most modern CPUs use the LGA form factor. Figure 1.11, earlier in this chapter, shows an LGA socket next to a PGA socket. Additionally, the ATX motherboard in Figure 1.2 has a PGA socket, whereas the micro ATX motherboard has an LGA.

FIGURE 1.27 DIP and PGA



Intel and AMD both make extensive use of an inverted socket/processor combination of sorts. As mentioned earlier, the LGA packaging calls for the pins to be placed on the motherboard, while the mates for these pins are on the processor packaging. As with PGA, LGA is named for the landmarks on the processor, not the ones on the motherboard. As a result, the grid of metallic contact points, called *lands*, on the bottom of the CPU gives this format its name.

You can easily identify which component inside the computer is the CPU because it is a large square lying flat on the motherboard with a very large heat sink and fan (refer to Figure 1.10). The CPU is almost always located very close to the RAM to improve system speed, as shown in Figure 1.1, Figure 1.2, and Figure 1.8.

CPU Architecture

As noted in the previous section, the functional job of the processor is to do math on very large strings of 0s and 1s. How the CPU goes about doing that depends upon its architecture. For commonly used processors, there are two major categories—those based on Complex Instruction Set Computing (CISC) and those based on Reduced Instruction Set Computer (RISC).

x64/x86

CISC (pronounced like *disk*, but with a "c") and RISC (pronounced *risk*) are examples of an instruction set architecture (ISA). Essentially, it's the set of commands that the processor can execute. Both types of chips, when combined with software, can ultimately perform all the same tasks. They just go about it differently. When programmers develop code, they develop it for a CISC or a RISC platform.

As the CISC name implies, instructions sent to the computer are relatively complex (as compared to RISC), and as such they can do multiple mathematical tasks with one instruction, and each instruction can take several clock cycles to complete. We'll talk more about CPU speeds in the "CPU Characteristics" section later, but for now, know that if a CPU is advertised as having 3.8 GHz speed, that means it can complete roughly 3.8 billion cycles in one second. The core of a processor can only do one thing at a time—it just does them very, very quickly so it looks like it's multitasking.

CISC was the original ISA for microprocessors, and the most well-known example of CISC technology is the x64/x86 platform popularized by Intel. AMD processors are CISC chips as well. So where did the terms x64 and x86 come from? First, just a bit more theory.

There is a set of data lines between the CPU and the primary memory of the system—remember the *bus*? The most common bus today is 64 bits wide, although there are still some 32-bit buses kicking around out there. In older days, buses could theoretically be as narrow as 2 bits, and 8-bit and 16-bit buses were popular for CPUs for several years. The wider the bus, the more data that can be processed per unit of time, and hence, more work can be performed. Internal registers in the CPU might be only 32 bits wide, but with a 64-bit system bus, two separate pipelines can receive information simultaneously. For true 64-bit CPUs, which have 64-bit internal registers and can run x64 versions of Microsoft operating systems, the external system data bus will always be 64 bits wide or some larger multiple thereof.

In the last paragraph we snuck in the term *x64*, and by doing that we also defined it. It refers to processors that are designed to work with 64 bits of data at a time. To go along with it, the operating system must also be designed to work with x64 chips.

Contrast that with processors that can handle only 32 bits of information at once. Those are referred to as *x86* processors. You might look at that last sentence and be certain that we made a typo, but we didn't. For a long time when 32-bit processors were the fastest on the PC market, Intel was the dominant player. Their CPUs had names like 80386 (aka i386) and 80486 (i486) and were based on the older 16-bit 80286 and 8086. Since the i386 and i486 were the most popular standards, the term x86 sprung up to mean a 32-bit architecture. So even though it may seem counterintuitive due to the numbers, x64 is newer and faster than x86.

Advanced RISC Machine

Moving into the RISC architecture, the primary type of processor used today is known as an *Advanced RISC Machine (ARM)* CPU. Depending on who you talk to and which sources you prefer, there are conflicting stories on if that's actually the right acronym, as ARM is also known as an Acorn RISC Machine. Regardless of what it stands for, ARM is a competing technology to Intel and AMD x64-based CPUs.

Based on the RISC acronym, one might think that the reduced set of instructions the processor can perform makes it inferior somehow, but that's not the case. Tasks just need to get executed in different ways. To use a human example, let's say that we tell you to add the number 7 to itself seven times. One way to do that is to use one step of multiplication: 7 × 7 equals 49. But what if we said you can't use multiplication? You can still get to the answer by using addition. It will just take you seven steps instead of one. Same answer, different process. That's kind of how RISC compares to CISC.

RISC processors have some advantages over their CISC counterparts. They can be made smaller than CISC chips and they produce less heat, making them ideal for mobile devices. In fact, nearly all smartphones use RISC-based chips, such as Apple's A15 and Samsung's Exynos series processors. On the downside, RISC processors use more memory than CISC ones do because it takes more code to complete a task with a RISC chip.

Like x64/x86, ARM processors have evolved over time—64-bit implementations are the most current, and they are designated ARM64; 32-bit versions are known simply as ARM.

CPU Cores

Older processors were *single-core*, meaning that there was one set of instruction pathways through the processor. Therefore, they could process one set of tasks at a time. Designers then figured out how to speed up the computer by creating multiple cores within one processor package. Each core effectively operates as its own independent processor, provided that the operating system and applications are able to support *multicore* technology. (Nearly all do.)

Today, almost all desktop CPUs in the market are multicore. The number of cores you want may determine the processor to get. For example, the 10th-generation Intel Core i7 has eight cores whereas the i5 has six.

NOTE

As an interesting aside, throughout the 1980s and 1990s, processor clock speeds grew at very fast rates. But they haven't gotten much faster over the last 10–15 years as designers ran into technical limitations for increasing speeds. Instead, computers have gotten faster because of the introduction of multicore processors, which is basically like having four (or six or eight) separate processors in one computer.

CPU Characteristics

When looking for a processor, you might have several decisions to make. Do you want an Intel or AMD CPU? Which model? How fast should it be? What features does it need to support? In this section, we will take a look at some characteristics of processor performance.

Speed

The speed of the processor is generally described in clock frequency. Older chips were rated in megahertz (MHz) and new chips in gigahertz (GHz). Since the dawn of the personal computer industry, motherboards have included oscillators, quartz crystals shaved down to a specific geometry so that engineers know exactly how they will react when a current is run through them. The phenomenon of a quartz crystal vibrating when exposed to a current is known as the *piezoelectric effect*. The crystal (XTL) known as the *system clock* keeps the time for the flow of data on the motherboard. How the front-side bus uses the clock leads to an *effective* clock rate known as the *FSB speed*. As discussed in the section "Types of Memory" later in this chapter, the FSB speed is computed differently for different types of RAM (DDR3, DDR4, DDR5, and so forth). From here, the CPU multiplies the FSB speed to produce its own internal clock rate, producing the third *speed* mentioned thus far.

As a result of the foregoing tricks of physics and mathematics, there can be a discrepancy between the front-side bus frequency and the internal frequency that the CPU uses to latch data and instructions through its pipelines. This disagreement between the numbers comes from the fact that the CPU is capable of splitting the clock signal it receives from the external oscillator that drives the front-side bus into multiple regular signals for its own internal use. In fact, you might be able to purchase a number of processors rated for different (internal) speeds that are all compatible with a single motherboard that has a front-side bus rated, for instance, at 1,333 MHz. Furthermore, you might be able to adjust the internal clock rate of the CPU that you purchased through settings in the BIOS.

The speed of a processor can also be tweaked by *overclocking*, or running the processor at a higher speed than the one at which the manufacturer rated it. Running at a higher speed requires more voltage and also generates more heat, which can shorten the life of the CPU. Manufacturers often discourage the practice (of course, they want you to just buy a faster and more expensive CPU), and it usually voids any warranty. However, some chips are sold today that specifically give you the ability to overclock them. Our official recommendation is to not do it unless the manufacturer says it's okay. If you're curious, plenty of information on how to overclock is available online.

Multithreading and Hyper-Threading

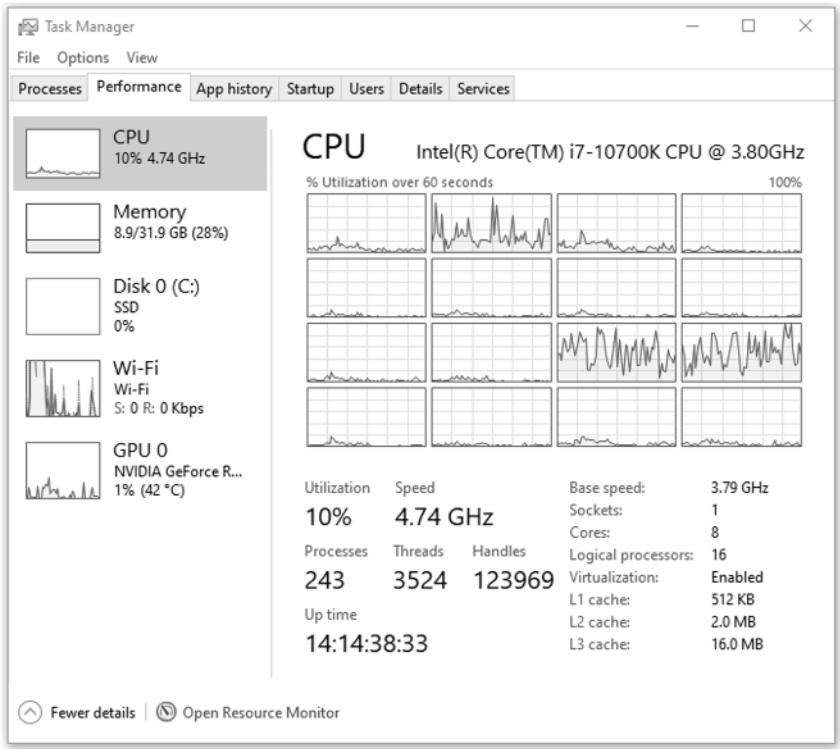
The string of instructions that a CPU runs is known as a *thread*. Old processors were capable of running only one thread at a time, whereas newer ones can run multiple threads at once. This is called *multithreading*.

Intel markets their multithreading technology as *Hyper-Threading Technology (HTT)*. HTT is a form of simultaneous multithreading (SMT). SMT takes advantage of a modern CPU's superscalar architecture. Superscalar processors can have multiple instructions operating on separate data in parallel.

HTT-capable processors appear to the operating system to be two processors. As a result, the operating system can schedule two processes at the same time, as in the case of symmetric multiprocessing (SMP), where two or more processors use the same system resources. In fact, the operating system must support SMP in order to take advantage of HTT. If the current process stalls because of missing data caused by, say, cache or branch prediction issues, the execution resources of the processor can be reallocated for a different process that is ready to go, reducing processor downtime.

HTT manifests itself in the Windows 10 Task Manager by, for example, showing graphs for twice as many CPUs as the system has cores. These virtual CPUs are listed as logical processors (see Figure 1.28).

FIGURE 1.28 Logical processors in Windows



For an in-market example, compare the Intel i5 with the Intel i7. Similar models will have the same number of cores (say, four), but the i7 supports HTT, whereas the i5 does not. This gives the i7 a performance edge over its cousin. The i9 will be even one further step up from the i7. For everyday email and Internet use, the differences won't amount to much. But for someone who is using resource-intensive apps such as online gaming or virtual reality, the differences, especially between i5 and i9 processors, can be important.

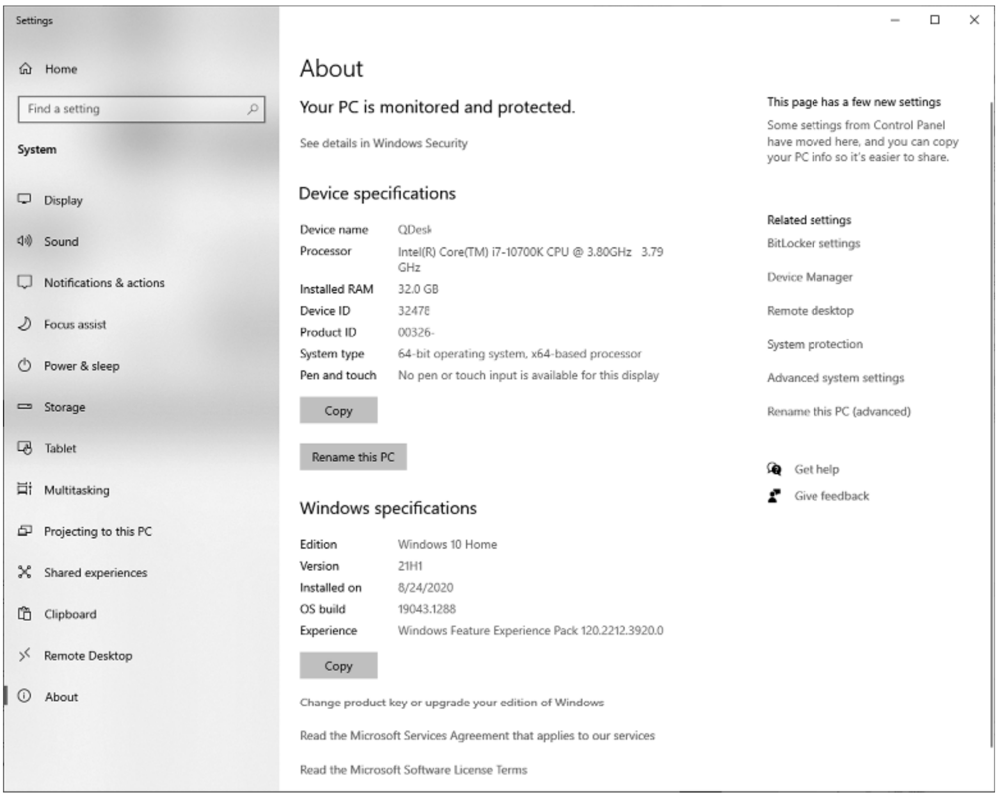
Real World Scenario

Which CPU Do You Have?

The surest way to determine which CPU your computer is using is to open the case and view the numbers stamped on the CPU, a process that today requires removal of the active heat sink. However, you may be able to get an idea without opening the case and removing the heat sink and fan because many manufacturers place a very obvious sticker somewhere on the case indicating the processor type. Failing this, you can always go to the manufacturer's website and look up the information on the model of computer you have.

An easier way may be to look in Control Panel ➢ System, to get the About screen as shown in Figure 1.29. Even more detailed information can be found by running the System Information by clicking Start and entering **msinfo32.exe**. Click System Information when it appears as the Best Match. Of course, third-party utilities such as CPU-Z that we showed you earlier will work too. A final way to determine a computer's CPU is to save your work, exit any open programs, and restart the computer. Watch closely as the computer boots back up. You should see a notation during the POST routine that tells you what chip you are using.

FIGURE 1.29 System information in the About screen



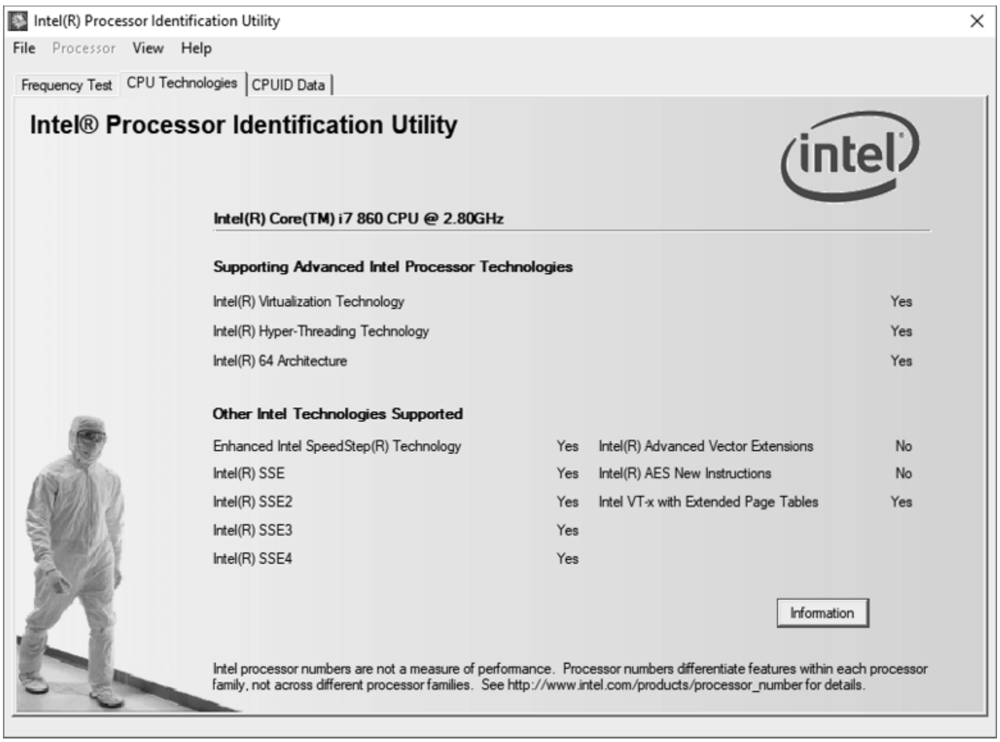
Virtualization Support

Many of today's CPUs support virtualization in hardware, which eases the burden on the system that software-based virtualization imposes. For more information on virtualization, see Chapter 8, "Virtualization and Cloud Computing." AMD calls its virtualization technology *AMD-V* (V for virtualization), whereas Intel calls theirs *Virtualization Technology (VT)*. Most processors made today support virtual technology, but not all. Keep in mind that the BIOS/UEFI and operating system must support it as well for virtualization to work. You may need to manually enable the virtualization support in the BIOS/UEFI before it can be used. If you have an Intel processor and would like to check its support of VT, visit the following site to download the Intel Processor Identification Utility:

`https://downloadcenter.intel.com/download/7838`

As shown in Figure 1.30, the CPU Technologies tab of this utility tells you if your CPU supports Intel VT.

FIGURE 1.30 Intel Processor Identification Utility



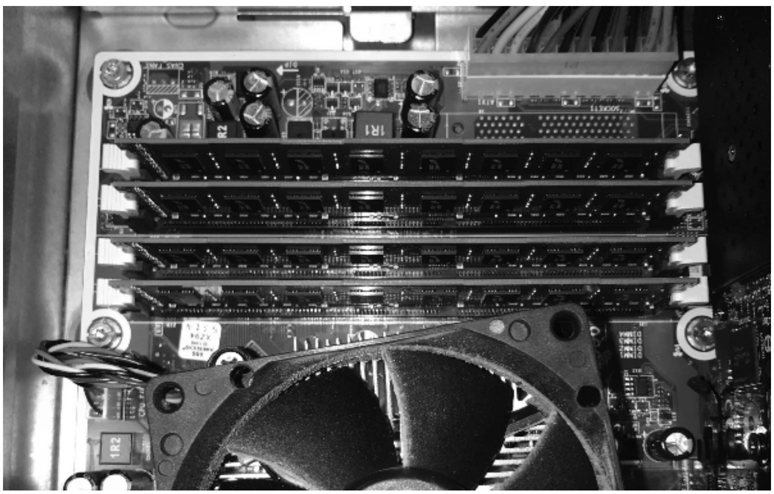
The CompTIA A+ objectives for CPUs aren't quite as long as they were for motherboards. For CPUs, be sure to understand the differences between x64/x86/ARM, single-core and multicore, multithreading, and virtualization support.

Understanding Memory

"More memory, more memory, I don't have enough memory!" Adding memory is one of the most popular, easy, and inexpensive ways to upgrade a computer. As the computer's CPU works, it stores data and instructions in the computer's memory. Contrary to what you might expect from an inexpensive solution, memory upgrades tend to afford the greatest performance increase as well, up to a point. Motherboards have memory limits; operating systems have memory limits; CPUs have memory limits.

To identify memory visually within a computer, look for several thin rows of small circuit boards sitting vertically, potentially packed tightly together near the processor. In situations where only one memory stick is installed, it will be that stick and a few empty slots that are tightly packed together. Figure 1.31 shows where memory is located in a system—in this case, all four banks are full.

FIGURE 1.31 Location of memory within a system



Important Memory Terms

There are a few technical terms and phrases that you need to understand with regard to memory and its function:

- Parity checking
- Error-correction code (ECC)
- Single- and double-sided memory
- Single-, dual-, triple-, and quad-channel memory

NOTE: ECC will sometimes be referred to as error correcting code as well. As with other acronyms in computing that might have multiple ways to spell it out, focus less on what it stands for and more on what it does.

These terms are discussed in detail in the following sections.

Parity Checking and Memory Banks

Parity checking is a rudimentary error-checking scheme that offers no error correction. Parity checking works most often on a byte, or 8 bits, of data. A ninth bit is added at the transmitting end and removed at the receiving end so that it does not affect the actual data transmitted. If the receiving end does not agree with the parity that is set in a particular byte, a parity error results. The four most common parity schemes affecting this extra bit are known as *even*, *odd*, *mark*, and *space*. Even and odd parity are used in systems that actually compute parity. Mark (a term for a digital pulse, or 1 bit) and space (a term for the lack of a pulse, or a 0 bit) parity are used in systems that do not compute parity but expect to see a fixed bit value stored in the parity location. Systems that do not support or reserve the location required for the parity bit are said to implement *non-parity memory*.

The most basic model for implementing memory in a computer system uses eight memory chips to form a set. Each memory chip holds millions or billions of bits of information, each in its own *cell*. For every byte in memory, one bit is stored in each of the eight chips. A ninth chip is added to the set to support the parity bit in systems that require it. One or more of these sets, implemented as individual chips or as chips mounted on a memory module, form a *memory bank*.

A bank of memory is required for the computer system to recognize electrically that the minimum number of memory components or the proper number of additional memory components has been installed. The width of the system data bus, the external bus of the processor, dictates how many memory chips or modules are required to satisfy a bank. For example, one 32-bit, 72-pin SIMM (single in-line memory module) satisfies a bank for an old 32-bit CPU, such as a i386 or i486 processor. Two such modules are required to satisfy a bank for a 64-bit processor—a Pentium, for instance. However, only a single 64-bit, 168-pin DIMM is required to satisfy the same Pentium processor. For those modules that have fewer than eight or nine chips mounted on them, more than 1 bit for every byte is being handled by some of the chips. For example, if you see three chips mounted, the two larger chips customarily handle 4 bits, a nibble, from each byte stored, and the third, smaller chip handles the single parity bit for each byte.

Even and odd parity schemes operate on each byte in the set of memory chips. In each case, the number of bits set to a value of 1 is counted up. If there is an even number of 1 bits in the byte (0, 2, 4, 6, or 8), even parity stores a 0 in the ninth bit, the parity bit; otherwise, it stores a 1 to even up the count. Odd parity does just the opposite, storing a 1 in the parity bit to make an even number of 1s odd and a 0 to keep an odd number of 1s odd. You can see that this is effective only for determining if there was a blatant error in the set of bits received, but there is no indication as to where the error is and how to fix it. Furthermore, the total 1-bit count is not important, only whether it's even or odd. Therefore, in either the even or odd scheme, if an even number of bits is altered in the same byte during transmission, the error goes undetected because flipping 2, 4, 6, or all 8 bits results in an even number of 1s remaining even and an odd number of 1s remaining odd.

Mark and space parity are used in systems that want to see 9 bits for every byte transmitted but don't compute the parity bit's value based on the bits in the byte. Mark parity always uses a 1 in the parity bit, and space parity always uses a 0. These schemes offer less error detection capability than the even and odd schemes because only changes in the parity bit can be detected. Again, parity checking is *not* error correction; it's error detection only, and not the best form of error detection at that. Nevertheless, an error can lock up the entire system and display a memory parity error. Enough of these errors and you need to replace the memory. Therefore, parity checking remains from the early days of computing as an effective indicator of large-scale memory and data-transmission failure, such as with serial interfaces attached to analog modems or networking console interfaces, but not so much for detecting random errors.

In the early days of personal computing, almost all memory was parity-based. As quality has increased over the years, parity checking in the RAM subsystem has become more uncommon. As noted earlier, if parity checking is not supported, there will generally be fewer chips per module, usually one less per column of RAM.

Error Checking and Correction

The next step in the evolution of memory error detection is known as *error-correction code (ECC)*. If memory supports ECC, check bits are generated and stored with the data. An algorithm is performed on the data and its check bits whenever the memory is accessed. If the result of the algorithm is all zeros, then the data is deemed valid and processing continues. ECC can detect single- and double-bit errors and actually correct single-bit errors. In other words, if a particular byte—group of 8 bits—contains errors in 2 of the 8 bits, ECC can recognize the error. If only 1 of the 8 bits is in error, ECC can correct the error.

Single- and Double-Sided Memory

Commonly speaking, the terms *single-sided memory* and *double-sided memory* refer to how some memory modules have chips on one side and others have chips on both sides. Double-sided memory is essentially treated by the system as two separate memory modules. Motherboards that support such memory have memory controllers that must switch between the two “sides” of the modules and, at any particular moment, can access only the side to which they have switched. Double-sided memory allows more memory to be inserted into a computer, using half the physical space of single-sided memory, which requires no switching by the memory controller.

Single-, Dual-, Triple-, and Quad-Channel Memory

Standard memory controllers manage access to memory in chunks of the same size as the system bus’s data width. This is considered communicating over a single channel. Most modern processors have a 64-bit system data bus. This means that a standard memory controller can transfer exactly 64 bits of information at a time. Communicating over a single channel is a bottleneck in an environment where the CPU and memory can both operate faster than the conduit between them. Up to a point, every channel added in parallel between the CPU and RAM serves to ease this constriction.

Memory controllers that support dual-channel and greater memory implementation were developed in an effort to alleviate the bottleneck between the CPU and RAM. *Dual-channel memory* is the memory controller’s coordination of two memory banks to work as a synchronized set during communication with the CPU, doubling the specified system bus width from the memory’s perspective. *Triple-channel memory*, then, demands the coordination of three memory modules at a time. *Quad-channel memory* is the coordination of four memory modules at once. Collectively, they are known as *multichannel memory* implementations.

Because today’s processors largely have 64-bit external data buses, and because one stick of memory satisfies this bus width, there is a 1:1 ratio between banks and modules. This means that implementing multichannel memory in today’s most popular computer systems requires that two, three, or four memory modules be installed at a time. Note, however, that it’s the motherboard, not the memory, that implements multichannel memory (more on this in a moment). *Single-channel memory*, in contrast, is the classic memory model that dictates only that a complete bank be satisfied whenever memory is initially installed or added. One bank supplies only half the width of the effective bus created by dual-channel support, for instance, which by definition pairs two banks at a time.

In almost all cases, multichannel implementations support single-channel installation, but poorer performance should be expected. Multichannel motherboards often include slots of different colors, usually one of each color per set of slots. To use only a single channel, you populate slots of the same color, skipping neighboring slots to do so. Filling neighboring slots in a dual-channel motherboard takes advantage of its dual-channel capability.

Because of the special tricks that are played with memory subsystems to improve overall system performance, care must be taken during the installation of disparate memory modules. In the worst case, the computer will cease to function when modules of different speeds, different capacities, or different numbers of sides are placed together in slots of the same channel. If all of these parameters are identical, there should be no problem with pairing modules. Nevertheless, problems could still occur when modules from two different manufacturers or certain unsupported manufacturers are installed, all other parameters being the same. Technical support or documentation from the manufacturer of your motherboard should be able to help with such issues.

Although it's not the make-up of the memory that leads to multichannel support but instead the technology on which the motherboard is based, some memory manufacturers still package and sell pairs and triplets of memory modules in an effort to give you peace of mind when you're buying memory for a system that implements multichannel memory architecture. Keep in mind, the motherboard memory slots have the distinctive color-coding, not the memory modules.

I Can't Fill All My Memory Slots

As a reminder, most motherboard manufacturers document the quantity and types of modules that their equipment supports. Consult your documentation when you have questions about supported memory. Most manufacturers require that slower memory be inserted in lower-numbered memory slots. This is because such a system adapts to the first module it sees, looking at the lower-numbered slots first. Counterintuitively, however, it might be required that you install modules of larger capacity rather than smaller modules in lower-numbered slots.

Additionally, memory technology continues to advance after each generation of motherboard chipsets is announced. Don't be surprised when you attempt to install a single module of the highest available capacity in your motherboard and the system doesn't recognize the module, either by itself or with others. That capacity of module might not have been in existence when the motherboard's chipset was released. Sometimes, flashing the BIOS is all that is required. Other times it just won't work. Consult the motherboard's documentation.

One common point of confusion, not related to capacity, when memory is installed is the lack of recognition of four modules when two or three modules work fine, for example. In such a case, let's say your motherboard's memory controller supports a total of four modules. Recall that a double-sided module acts like two separate modules. If you are using double-sided memory, your motherboard might limit you to two such modules comprising four sides (essentially four virtual modules), even though you have four slots on the board. If instead you start with three single-sided modules, when you attempt to install a double-sided module in the fourth slot, you are essentially asking the motherboard to accept five modules, which it cannot.

Finally, know that dual- and quad-channel memory support is common today, but triple-channel is less so. As with anything else, check the motherboard's documentation to see what it supports.

Types of Memory

Memory comes in many formats. Each one has a particular set of features and characteristics, making it best suited for a particular application. Some decisions about the application of the memory type are based on suitability; others are based on affordability to consumers or marketability to computer manufacturers. The following list gives you an idea of the vast array of memory types and subtypes:

- DRAM (dynamic random access memory)

    ADRAM (asynchronous DRAM)

    FPM DRAM (fast page mode DRAM)

    EDO DRAM (extended data out DRAM)

    BEDO DRAM (burst EDO DRAM)

    SDRAM (synchronous DRAM)

    SDR SDRAM (single data rate SDRAM)

    DDR SDRAM (double data rate SDRAM)

    DDR2 SDRAM (double data rate, version two, SDRAM)

    DDR3 SDRAM (double data rate, version three, SDRAM)

    DDR4 SDRAM (double data rate, version four, SDRAM)

    DDR5 SDRAM (double data rate, version five, SDRAM)

- SRAM (static random access memory)
- ROM (read-only memory)

Pay particular attention to all synchronous DRAM types as that's the most common type in use. Note that the type of memory does not dictate the packaging of the memory. Conversely, however, you might notice one particular memory packaging holding the same type of memory every time you come across it. Nevertheless, there is no requirement to this end. Let's detail the intricacies of some of these memory types.

DRAM

DRAM is dynamic random access memory. This is what most people are talking about when they mention RAM. When you expand the memory in a computer, you are adding DRAM chips. You use DRAM to expand the memory in the computer because it's a cheaper type of memory. Dynamic RAM chips are cheaper to manufacture than most other types because they are less complex. *Dynamic* refers to the memory chips' need for a constant update signal (also called a *refresh signal*) in order to keep the information that is written there. If this signal is not received every so often, the information will bleed off and cease to exist. Currently, the most popular implementations of DRAM are based on synchronous DRAM and include DDR3 and DDR4. Occasionally you will see some DDR2, and DDR5 is new so it hasn't been widely adopted yet. Before discussing these technologies, let's take a quick look at the legacy asynchronous memory types, none of which should appear on modern exams.

Asynchronous DRAM

Asynchronous DRAM (ADRAM) is characterized by its independence from the CPU's external clock. Asynchronous DRAM chips have codes on them that end in a numerical value that is related to (often 1/10 of the actual value of) the access time of the memory. Access time is essentially the difference between the time when the information is requested from memory and the time when the data is returned. Common access times attributed to asynchronous DRAM were in the 40- to 120-nanosecond (ns) vicinity. A lower access time is obviously better for overall performance.

Because ADRAM is not synchronized to the front-side bus, you would often have to insert wait states through the BIOS setup for a faster CPU to be able to use the same memory as a slower CPU. These wait states represented intervals in which the CPU had to mark time and do nothing while waiting for the memory subsystem to become ready again for subsequent access.

Common asynchronous DRAM technologies included fast page mode (FPM), extended data out (EDO), and burst EDO (BEDO). Feel free to investigate the details of these particular technologies, but a thorough discussion of these memory types is not necessary here. The A+ technician should be concerned with synchronous forms of RAM, which are the only types of memory being installed in mainstream computer systems today.

Synchronous DRAM

Synchronous DRAM (SDRAM) shares a common clock signal with the computer's system-bus clock, which provides the common signal that all local-bus components use for each step that they perform. This characteristic ties SDRAM to the speed of the FSB and hence the processor, eliminating the need to configure the CPU to wait for the memory to catch up.

Originally, *SDRAM* was the term used to refer to the only form of synchronous DRAM on the market. As the technology progressed, and more was being done with each clock signal on the FSB, various forms of SDRAM were developed. What was once called simply SDRAM needed a new name retroactively. Today, we use the term *single data rate SDRAM (SDR SDRAM)* to refer to this original type of SDRAM.

SDR SDRAM *SDR SDRAM* is a legacy RAM technology, and it is presented here only to provide a basis for the upcoming discussion of DDR and other more advanced RAM. With SDR SDRAM, every time the system clock ticks, 1 bit of data can be transmitted per data pin, limiting the bit rate per pin of SDRAM to the corresponding numerical value of the clock's frequency. With today's processors interfacing with memory using a parallel data-bus width of 8 bytes (hence the term *64-bit processor*), a 100 MHz clock signal produces 800 MBps. That's mega *bytes* per second, not mega *bits*. Such memory modules are referred to as PC100, named for the true FSB clock rate upon which they rely. PC100 was preceded by PC66 and succeeded by PC133, which used a 133 MHz clock to produce 1,066 MBps of throughput.

Note that throughput in megabytes per second is easily computed as eight times the rating in the name. This trick works for the more advanced forms of SDRAM as well. The common thread is the 8-byte system data bus. Incidentally, you can double throughput results when implementing dual-channel memory.

DDR SDRAM *Double data rate (DDR) SDRAM* earns its name by doubling the transfer rate of ordinary SDRAM; it does so by double-pumping the data, which means transferring a bit per pin on both the rising and falling edges of the clock signal. This obtains twice the transfer rate at the same FSB clock frequency. It's the increasing clock frequency that generates heating issues with newer components, so keeping the clock the same is an advantage. The same 100 MHz clock gives a DDR SDRAM system the impression of a 200 MHz clock compared to an SDR SDRAM system. For marketing purposes, and to aid in the comparison of disparate products (DDR vs. SDR, for example), the industry has settled on the practice of using this effective clock rate as the speed of the FSB.

Module Throughput Related to FSB Speed

There is always an 8:1 module-to-chip (or module-to-FSB speed) numbering ratio because of the 8 bytes that are transferred at a time with 64-bit processors (*not* because of the ratio of 8 bits per byte). The formula in Figure 1.32 explains how this relationship works.

FIGURE 1.32 The 64-bit memory throughput formula

FSB in MHz	(~~cycles~~/second)
X 8 bytes	(bytes/~~cycle~~)
throughput	(bytes/second)

Because the actual system clock speed is rarely mentioned in marketing literature, on packaging, or on store shelves for DDR and higher, you can use this advertised FSB frequency in your computations for DDR throughput. For example, with a 100 MHz clock and two operations per cycle, motherboard makers will market their boards as having an FSB of 200 MHz. Multiplying this effective rate by 8 bytes transferred per cycle, the data rate is 1,600 MBps. Because DDR made throughput a bit trickier to compute, the industry began using this final throughput figure to name the memory modules instead of the actual frequency, which was used when naming SDR modules. This makes the result seem many times better (and much more marketable), while it's really only twice (or so) as good, or close to it.

In this example, the module is referred to as PC1600, based on a throughput of 1,600 MBps. The chips that go into making PC1600 modules are named DDR200 for the effective FSB frequency of 200 MHz. Stated differently, the industry uses DDR200 memory chips to manufacture PC1600 memory modules.

Let's make sure that you grasp the relationship between the speed of the FSB and the name for the related chips as well as the relationship between the name of the chips (or the speed of the FSB) and the name of the modules. Consider an FSB of 400 MHz, meaning an actual clock signal of 200 MHz, by the way—the FSB is double the actual clock for DDR, remember. It should be clear that this motherboard requires modules populated with DDR400 chips and that you'll find such modules marketed and sold as PC3200.

Let's try another. What do you need for a motherboard that features a 333 MHz FSB (actual clock is 166 MHz)? Well, just using the 8:1 rule mentioned earlier, you might be on the lookout for a PC2667 module. Note, however, that sometimes the numbers have to be played with a bit to come up with the industry's marketing terms. You'll have an easier time finding PC2700 modules that are designed specifically for a motherboard like yours, with an FSB of 333 MHz. The label isn't always technically accurate, but round numbers sell better, perhaps. The important concept here is that if you find PC2700 modules and PC2667 modules, there's absolutely no difference; they both have a 2667 MBps throughput rate. Go for the best deal; just make sure that the memory manufacturer is reputable.

DDR2 SDRAM Think of the *2* in *DDR2* as yet another multiplier of 2 in the SDRAM technology, using a lower peak voltage to keep power consumption down (1.8V vs. the 2.5V of DDR). Still double-pumping, DDR2, like DDR, uses both sweeps of the clock signal for data transfer. Internally, DDR2 further splits each clock pulse in two, doubling the number of operations it can perform per FSB clock cycle. Through enhancements in the electrical interface and buffers, as well as through adding off-chip drivers, DDR2 nominally produces four times the throughput that SDR is capable of producing.

Continuing the DDR example, DDR2, using a 100 MHz actual clock, transfers data in four operations per cycle (effective 400 MHz FSB) and still 8 bytes per operation, for a total of 3,200 MBps. Just as with DDR, chips for DDR2 are named based on the perceived frequency. In this case, you would be using DDR2-400 chips. DDR2 carries on the effective FSB frequency method for naming modules but cannot simply call them PC3200 modules because those already exist in the DDR world. DDR2 calls these modules PC2-3200. (Note the dash to keep the numeric components separate.)

As another example, it should make sense that PC2-5300 modules are populated with DDR2-667 chips. Recall that you might have to play with the numbers a bit. If you multiply the well-known FSB speed of 667 MHz by 8 to figure out what modules you need, you might go searching for PC2-5333 modules. You might find someone advertising such modules, but most compatible modules will be labeled PC2-5300 for the same marketability mentioned earlier. They both support 5,333 MBps of throughput.

DDR3 SDRAM The next generation of memory devices was designed to roughly double the performance of DDR2 products. Based on the functionality and characteristics of DDR2's proposed successor, most informed consumers and some members of the industry surely assumed the forthcoming name would be DDR4. This was not to be, however, and DDR3 was born. This naming convention proved that the 2 in DDR2 was not meant to be a multiplier but instead a revision mark of sorts. Well, if DDR2 was the second version of DDR, then DDR3 is the third. *DDR3* is a memory type, designed to be twice as fast as the DDR2 memory, that operates with the same system clock speed. Just as DDR2 was required to lower power consumption to make up for higher frequencies, DDR3 must do the same. In fact, the peak voltage for DDR3 is only 1.5V.

The most commonly found range of actual clock speeds for DDR3 tends to be from 133 MHz at the low end to less than 300 MHz. Because double-pumping continues with DDR3, and because four operations occur at each wave crest (eight operations per cycle), this frequency range translates to common FSB implementations from 1,066 MHz to more than 2,000 MHz in DDR3 systems. These memory devices are named following the conventions established earlier. Therefore, if you buy a motherboard with a 1,600 MHz FSB, you know immediately that you need a memory module populated with DDR3-1600 chips, because the chips are always named for the FSB speed. Using the 8:1 module-to-chip/FSB naming rule, the modules that you need would be called PC3-12800, supporting a 12,800 MBps throughput.

The earliest DDR3 chips, however, were based on a 100 MHz actual clock signal, so we can build on our earlier example, which was also based on an actual clock rate of 100 MHz. With eight operations per cycle, the FSB on DDR3 motherboards is rated at 800 MHz, quite a lot of efficiency while still not needing to change the original clock with which our examples began. Applying the 8:1 rule again, the resulting RAM modules for this motherboard are called PC3-6400 and support a throughput of 6,400 MBps, carrying chips called DDR3-800, again named for the FSB speed.

DDR4 SDRAM Continuing the inevitable march of technology, *DDR4* is the next iteration of SDRAM on the market. As you would expect, the speed of DDR4 is roughly double that of DDR3. DDR3 provided data rates of approximately 800 Mbps to 2,133 Mbps, whereas DDR4 ranges between 1,600 Mbps and 3,200 Mbps. DDR also runs at a lower voltage—1.2 volts. Finally, DDR4 can support more memory per module, up to 512 GB per chip. Realistically, though, no one produces motherboards (or RAM) that support that quantity. The largest you will see are 64 GB sticks.

DDR5 SDRAM After a long wait, *DDR5* finally hit the market at the end of 2021. Intel's Alder Lake platform was the first to support it; AMD chips could support DDR5 in early 2022 with the Zen 4 release.

DDR5 doubles the speed of DDR4 to 6.4 Gbps, as is expected for a new memory standard. Improved power efficiency means it runs at 1.1 volts. DDR5 is also the first memory module to be available in up to 128 GB modules.

SRAM

Static random access memory (SRAM) doesn't require a refresh signal like DRAM does. The chips are more complex and are thus more expensive. However, they are considerably faster. DRAM access times come in at 40 nanoseconds (ns) or more; SRAM has access times faster than 10ns. SRAM is classically used for cache memory.

ROM

ROM stands for read-only memory. It is called read-only because you could not write to the original form of this memory. Once information had been etched on a silicon chip and manufactured into the ROM package, the information couldn't be changed. Some form of ROM is normally used to store the computer's BIOS because this information normally does not change often.

The system ROM in the original IBM PC contained the power-on self-test (POST), BIOS, and cassette BASIC. Later, IBM computers and compatibles included everything but the cassette BASIC. The system ROM enables the computer to "pull itself up by its bootstraps," or *boot* (find and start the operating system).

Through the years, different forms of ROM were developed that could be altered, later ones more easily than earlier ones. The first generation was the *programmable ROM (PROM)*, which could be written to for the first time in the field using a special programming device, but then no more. You may liken this to the burning of a DVD-R.

The *erasable PROM (EPROM)* followed the PROM, and it could be erased using ultraviolet light and subsequently reprogrammed using the original programming device. These days, flash memory is a form of *electronically erasable PROM (EEPROM)*. Of course, it does not require UV light to erase its contents, but rather a slightly higher than normal electrical pulse.

NOTE

Although the names of these memory devices are different, they all contain ROM. Therefore, regardless which of these technologies is used to manufacture a BIOS chip, it's never incorrect to say that the result is a ROM chip.

Memory Packaging

The memory slots on a motherboard are designed for particular module form factors or styles. RAM historically evolved from form factors no longer seen for such applications, such as dual in-line package (DIP), single in-line memory module (SIMM), and single in-line pin package (SIPP). The most popular form factors for primary memory modules today are as follows:

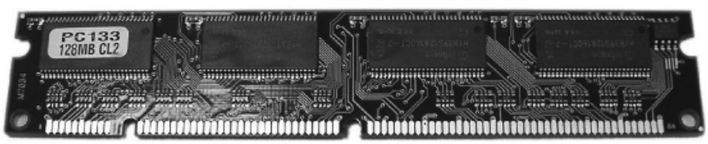
- DIMM (dual in-line memory module)
- SODIMM (small outline dual in-line memory module)

Desktop computers will use DIMMs. Laptops and smaller devices require SODIMMs or smaller memory packaging. So, in addition to coordinating the speed of the components, their form factor is an issue that must be addressed.

DIMM

One type of memory package is known as a DIMM, which stands for dual in-line memory module. DIMMs are 64-bit memory modules that are used as a package for the SDRAM family: SDR, DDR, DDR2, DDR3, DDR4, and DDR5. The term *dual* refers to the fact that, unlike their SIMM predecessors, DIMMs differentiate the functionality of the pins on one side of the module from the corresponding pins on the other side. With 84 pins per side, this makes 168 independent pins on each standard SDR module, as shown with its two keying notches as well as the last pin labeled 84 on the right side in Figure 1.33. SDR SDRAM modules are no longer part of the CompTIA A+ objectives, and they are mentioned here as a foundation only.

FIGURE 1.33 An SDR dual in-line memory module (DIMM)



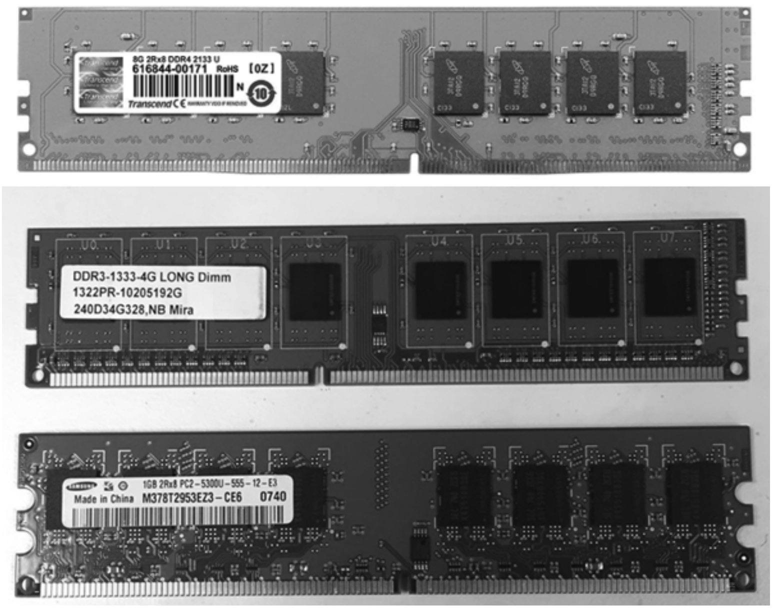
The DIMM used for DDR memory has a total of 184 pins and a single keying notch, whereas the DIMM used for DDR2 has a total of 240 pins, one keying notch, and possibly an aluminum cover for both sides, called a *heat spreader* and designed like a heat sink to dissipate heat away from the memory chips and prevent overheating. The DDR3 DIMM is similar to that of DDR2. It has 240 pins and a single keying notch, but the notch is in a different location to avoid cross-insertion. Not only is the DDR3 DIMM physically incompatible with DDR2 DIMM slots, it's also electrically incompatible. A DDR4 DIMM is the same length as a DDR3 DIMM, but is about 0.9mm taller and has 288 pins. The key is in a different spot, so you can't put DDR4 memory into a DDR2 or DDR3 slot. Finally, DDR5 has 288 pins as DDR4 does but is keyed differently so that DDR4 modules won't fit into DDR5 slots, and vice versa. Table 1.3 summarizes some key differences between the types of DDR we've introduced in this chapter.

TABLE 1.3 DDR characteristics

Characteristic	DDR	DDR2	DDR3	DDR4	DDR5
Pins	184	240	240	288	288
Max memory	1 GB	8 GB	32 GB	64 GB	128 GB
Channels	1	1	1	1	2
Voltage	2.5 v	1.8 v	1.5 v	1.2 v	1.1 v

Figure 1.34 shows, from top to bottom, DDR4, DDR3, and DDR2 DIMMs.

FIGURE 1.34 DDR4, DDR3, and DDR2 DIMMs



Inserting and Removing Memory Modules

The original single in-line memory modules had to be inserted into their slots at a 45° angle. The installer then had to apply slight pressure as the module was maneuvered upright at a 90° angle to the motherboard, where a locking mechanism would grip the module and prevent it from returning to its 45° position. This procedure created a pressure that reinforced the contact of the module with its slot. Releasing the clips on either end of the module unlocked it and allowed it to return to 45°, where it could be removed.

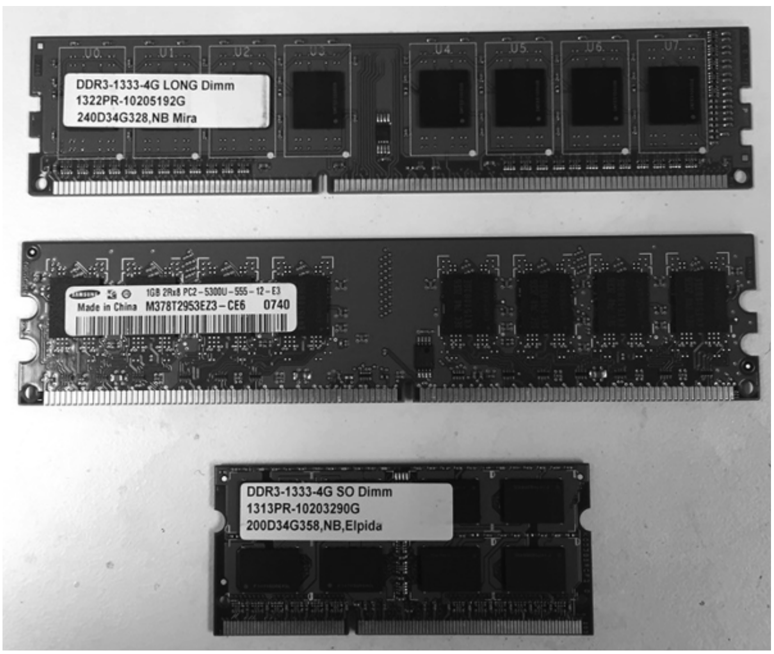
DIMM slots, by comparison, have no spring action. DIMMs are inserted straight into the slot with the locking tabs pulled away from the module. The locking tabs are at either end of the module, and they automatically snap into place, securing the module. Pulling the tabs away from the module releases the module from the slot, allowing it to be effortlessly removed.

SODIMM

Laptop computers and other computers that require much smaller components don't use standard RAM packages, such as DIMMs. Instead, they call for a much smaller memory form factor, such as a *small outline DIMM (SODIMM)*. SODIMMs are available in many physical implementations, including the older 32-bit (72- and 100-pin) configuration and newer 64-bit (144-pin SDR SDRAM, 200-pin DDR/DDR2, 204-pin DDR3, 260-pin DDR4, and 262-pin DDR5) configurations.

All 64-bit modules have a single keying notch. The 144-pin module's notch is slightly off center. Note that although the 200-pin SODIMMs for DDR and DDR2 have slightly different keying, it's not so different that you don't need to pay close attention to differentiate the two. They are not, however, interchangeable. DDR3, DDR4, and DDR5 are keyed differently from the others as well. Figure 1.34 shows a DDR3 SODIMM compared to DDR3 and DDR2 DIMMs.

FIGURE 1.35 DDR3 SODIMM vs. DDR3 and DDR2 DIMMs



NOTE: For the A+ exam, be sure to know the differences between SODIMMs and DIMMs, DDR3, DDR4, and DDR5, ECC RAM, and single-, dual-, triple-, and quad-channel RAM. Also understand what virtual RAM is.

Understanding Cooling Systems

It's a basic concept of physics: electronic components turn electricity into work and heat. The excess heat must be dissipated or it will shorten the life of the components. In some cases (like with the CPU), the component will produce so much heat that it can destroy itself in a matter of seconds if there is not some way to remove this extra heat.

Air-cooling methods are used to cool the internal components of most PCs. With air cooling, the movement of air removes the heat from the component. Sometimes, large blocks of metal called *heat sinks* are attached to a heat-producing component in order to dissipate the heat more rapidly.

Fans

When you turn on a computer, you will often hear lots of whirring. Contrary to popular opinion, the majority of the noise isn't coming from the hard disk (unless it's about to go bad). Most of this noise is coming from the various fans inside the computer. Fans provide airflow within the computer.

Most PCs have a combination of the following seven fans:

Front Intake Fan This fan is used to bring fresh, cool air into the computer for cooling purposes.

Rear Exhaust Fan This fan is used to take hot air out of the case.

Power Supply Exhaust Fan This fan is usually found at the back of the power supply, and it is used to cool the power supply. In addition, this fan draws air from inside the case into vents in the power supply. This pulls hot air through the power supply so that it can be blown out of the case. The front intake fan assists with this airflow. The rear exhaust fan supplements the power supply fan to achieve the same result outside of the power supply.

CPU Fan This fan is used to cool the processor. Typically, this fan is attached to a large heat sink, which is in turn attached directly to the processor.

Chipset Fan Some motherboard manufacturers replaced the heat sink on their onboard chipset with a heat sink and fan combination as the chipset became more advanced. This fan aids in the cooling of the onboard chipset (especially useful when overclocking—setting the system clock frequency higher than the default).

Video Card Chipset Fan As video cards get more complex and have higher performance, more video cards have cooling fans directly attached. Despite their name, these fans don't attach to a chipset in the same sense as to a chipset on a motherboard. The chipset here is the set of chips mounted on the adapter, including the GPU and graphics memory. On many late-model graphics adapters, the equivalent of a second slot is dedicated to cooling the adapter. The cooling half of the adapter has vents in the backplane bracket to exhaust the heated air.

Memory Module Fan The more capable memory becomes of keeping up with the CPU, the hotter the memory runs. As an extra measure of safety, regardless of the presence of heat spreaders on the modules, an optional fan setup for your memory might be in order. See the upcoming section "Memory Cooling" for more information.

Motherboard Fan Power Connectors

It's important to be aware of the two main types of fan connections found on today's motherboards. One of these connectors has only three connections, while the other has four. The fan connectors and motherboard headers are interchangeable between the two pinouts, but if a chassis fan has four conductors, it's a sign that it's calling for connectivity to an extra +5VDC (volts direct current) connection that the most common 3-pin header doesn't offer. A more rare 3-pin chassis-fan connector features a +12VDC power connection for heavier-duty fans and a rotation pin used as an input to the motherboard for sensing the speed of the fan.

4-pin CPU connections place the ground and power connections in pins 1 and 2, respectively, so that 2-pin connectors can be used to power older fans. The 4-pin header also offers a tachometer input signal from the fan on pin 3 so that the speed of the fan can be monitored by the BIOS and other utilities. Look for markings such as *CPU FAN IN* to identify this function. Pin 4 might be labeled *CPU FAN PWM* to denote the pulse-width modulation that can be used to send a signal to the fan to control its speed. This is the function lost when a 3-pin connector is placed in the correct position on a 4-pin header. Four-pin chassis-fan connectors can share the tachometer function but replace the speed control function with the extra 5V mentioned earlier.

Other power connections and types will be covered in Chapter 2, including the Molex connector, which can be used to power chassis and CPU fans using an adapter or the built-in connector on mostly older fans manufactured before the motherboard connectors were standardized. Figure 1.36 shows two 3-pin chassis-fan headers on a motherboard.

FIGURE 1.36 3-pin chassis-fan headers

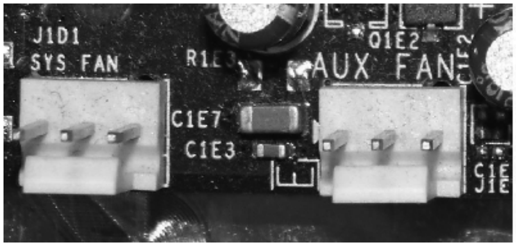
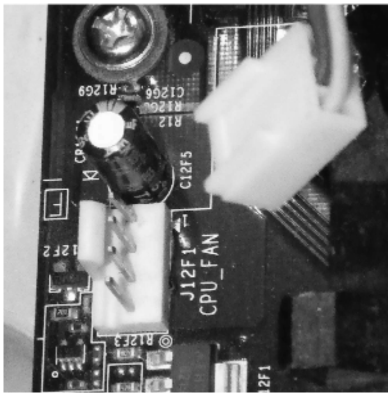


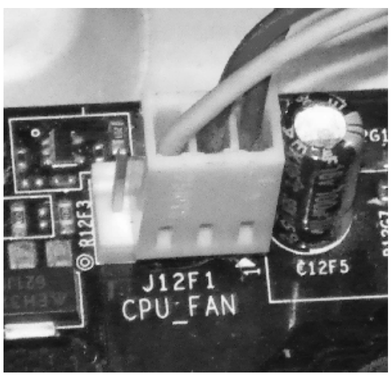
Figure 1.37 shows a 4-pin CPU fan header with an approaching 3-pin connector from the fan. Note that the keying tab is lined up with the same three pins it's lined up with in the 3-pin connectors.

FIGURE 1.37 A 4-pin CPU fan header



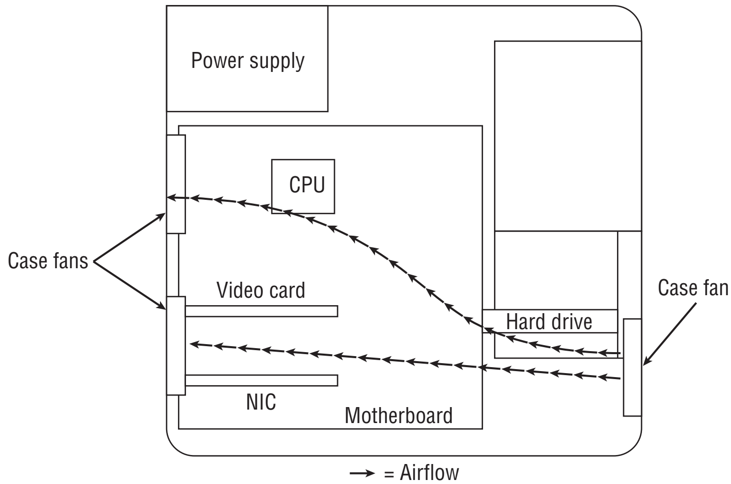
This physical aspect and the similar pin functions are what make these connectors interchangeable, provided that the header's function matches the role of the fan being connected. Figure 1.38 shows the resulting unused pin on the 4-pin header. Again, controlling the fan's speed is not supported in this configuration.

FIGURE 1.38 Position of a 3-pin connector on a 4-pin header



Ideally, the airflow inside a computer should resemble what is shown in Figure 1.39, where the back of the chassis is shown on the left in the image.

FIGURE 1.39 System unit airflow



Note that you must pay attention to the orientation of the power supply's airflow. If the power supply fan is an exhaust fan, as assumed in this discussion, the front and rear fans will match their earlier descriptions: front, intake; rear, exhaust. If you run across a power supply that has an intake fan, the orientation of the supplemental chassis fans should be reversed as well. The rear chassis fan(s) should always be installed in the same orientation the power supply fan runs to avoid creating a small airflow circuit that circumvents the cross-flow of air throughout the case. The front chassis fan and the rear fans should always be installed in reverse orientation to avoid having them fight against each other and thereby reduce the internal airflow. Reversing supplemental chassis fans is usually no harder than removing four screws and flipping the fan. Sometimes, the fan might just snap out, flip, and then snap back in, depending on the way it is rigged up.

Memory Cooling

If you are going to start overclocking your computer, you will want to do everything in your power to cool all of its components, and that includes the memory.

There are two methods of cooling memory: passive and active. The passive memory cooling method just uses the ambient case airflow to cool the memory through the use of enhanced heat dissipation. For this, you can buy either heat sinks or, as mentioned earlier, special "for memory chips only" devices known as *heat spreaders*. Recall that these are special aluminum or copper housings that wrap around memory chips and conduct the heat away from them.

Active cooling, on the other hand, usually involves forcing some kind of cooling medium (air or water) around the RAM chips themselves or around their heat sinks. Most often, active cooling methods are just high-speed fans directing air right over a set of heat spreaders.

Hard Drive Cooling

You might be thinking, "Hey, my hard drive is doing work all the time. Is there anything I can do to cool it off?" There are both active and passive cooling devices for hard drives. Most common, however, is the active cooling bay. You install a hard drive in a special device that fits into a 5¼″ expansion bay. This device contains fans that draw in cool air over the hard drive, thus cooling it. Figure 1.40 shows an example of one of these active hard drive coolers. As you might suspect, you can also get heat sinks for hard drives.

FIGURE 1.40 An active hard disk cooler

Chipset Cooling

Every motherboard has a chip or chipset that controls how the computer operates. Like other chips in the computer, the chipset is normally cooled by the ambient air movement in the case. However, when you overclock a computer, the chipset may need to be cooled more because it is working harder than it normally would be. Therefore, it is often desirable to replace the onboard chipset cooler with a more efficient one. Refer back to Figure 1.4 for a look at a modern chipset cooling solution.

CPU Cooling

Probably the greatest challenge in cooling is the computer's CPU. It is the component that generates the most heat in a computer (aside from some pretty insane GPUs out there). As a matter of fact, if a modern processor isn't actively cooled all of the time, it will generate enough heat to burn itself up in seconds. That's why most motherboards have an internal CPU heat sensor and a CPU_FAN sensor. If no cooling fan is active, these devices will shut down the computer before damage occurs.

There are multiple CPU cooling methods, but the two most common are air cooling and liquid cooling.

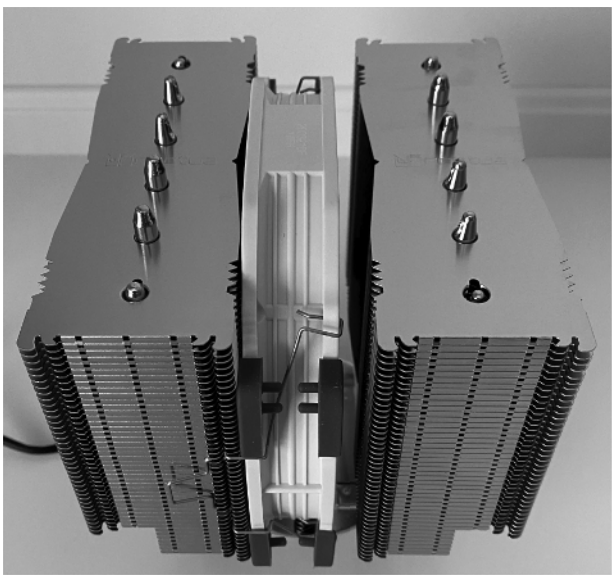
Air Cooling

The parts inside most computers are cooled by air moving through the case. The CPU is no exception. However, because of the large amount of heat produced, the CPU must have (proportionately) the largest surface area exposed to the moving air in the case. Therefore, the heat sinks on the CPU are the largest of any inside the computer.

The CPU fan often blows air down through the body of the heat sink to force the heat into the ambient internal air where it can join the airflow circuit for removal from the case. However, in some cases, you might find that the heat sink extends up farther, using radiator-type fins, and the fan is placed at a right angle and to the side of the heat sink. This design moves the heat away from the heat sink immediately instead of pushing the air down through the heat sink. CPU fans can be purchased that have an adjustable rheostat to allow you to dial in as little airflow as you need, aiding in noise reduction but potentially leading to accidental overheating.

It should be noted that the highest-performing CPU coolers use copper plates in direct contact with the CPU. They also use high-speed and high-CFM cooling fans to dissipate the heat produced by the processor. CFM is short for cubic feet per minute, an airflow measurement of the volume of air that passes by a stationary object per minute. Figure 1.41 shows a newer, large heat sink with a fan in the center. In the picture it can be tough to gauge size—this unit is about six inches across! (And to be fair, this heat sink should have a second fan on one of the sides, but with the second fan the heatsink wouldn't fit into the author's case—the RAM was in the way.)

FIGURE 1.41 Large heat sink and fan



Most new CPU heat sinks use tubing to transfer heat away from the CPU. With any cooling system, the more surface area exposed to the cooling method, the better the cooling. Plus, the heat pipes can be used to transfer heat to a location away from the heat source before cooling. This is especially useful in cases where the form factor is small and with laptops, where open space is limited.

With advanced heat sinks and CPU cooling methods like this, it is important to improve the thermal transfer efficiency as much as possible. To that end, cooling engineers came up with a glue-like compound that helps to bridge the extremely small gaps between the CPU and the heat sink, which avoids superheated pockets of air that can lead to focal damage of the CPU. This product is known as *thermal transfer compound*, or simply *thermal compound* (alternatively, *thermal grease* or *thermal paste*), and it can be bought in small tubes. Single-use tubes are also available and alleviate the guesswork involved with how much you should apply. Watch out, though; this stuff makes quite a mess and doesn't want to come off your fingers very easily. An alternative to the paste is a small *thermal pad*, which provides heat conductivity between the processor and the heat sink.

Apply the compound by placing a bead in the center of the heat sink, not on the CPU, because some heat sinks don't cover the entire CPU package. That might sound like a problem, but some CPUs don't have heat-producing components all the way out to the edges. Some CPUs even have a raised area directly over the silicon die within the packaging, resulting in a smaller contact area between the components. You should apply less than you think you need because the pressure of attaching the heat sink to the CPU will spread the compound across the entire surface in a very thin layer. It's advisable to use a clean, lint-free applicator of your choosing to spread the compound around a bit as well, just to get the spreading started. You don't need to concern yourself with spreading it too thoroughly or too neatly because the pressure applied during attachment will equalize the compound quite well. During attachment, watch for oozing compound around the edges, clean it off immediately, and use less next time.

Improving and Maintaining CPU Cooling

In addition to using thermal compound, you can enhance the cooling efficiency of a CPU heat sink by lapping the heat sink, which smoothens the mating surface using a very fine sanding element, about 1000 grit in the finishing stage. Some vendors of the more expensive heat sinks will offer this service as an add-on.

If your CPU has been in service for an extended period of time, perhaps three years or more, it is a smart idea to remove the heat sink and old thermal compound and then apply fresh thermal compound and reattach the heat sink. Be careful, though; if your thermal paste has already turned into thermal "glue," you can wrench the processor right out of the socket, even with the release mechanism locked in place. Invariably, this damages the pins on the chip. Try running the computer for a couple of minutes to warm the paste and then try removing the heat sink again.

Counterintuitively perhaps, you can remove a sealed heat sink from the processor by gently rotating the heat sink to break the paste's seal. Again, this can be made easier with heat.

If the CPU has risen in the socket already, however, rotating the heat sink would be an extremely bad idea. Sometimes, after you realize that the CPU has risen a bit and that you need to release the mechanism holding it in to reseat it, you find that the release arm is not accessible with the heat sink in place. This is an unfortunate predicament that will present plenty of opportunity to learn.

If you've ever installed a brand-new heat sink onto a CPU, you've most likely used thermal compound or the thermal compound patch that was already applied to the heat sink for you. If your new heat sink has a patch of thermal compound pre-applied, don't add more. If you ever remove the heat sink, don't try to reuse the patch or any other form of thermal compound. Clean it all off and start fresh.

Liquid Cooling

Liquid cooling is a technology whereby a special water block is used to conduct heat away from the processor (as well as from the chipset). Water is circulated through this block to a radiator, where it is cooled.

The theory is that you could achieve better cooling performance through the use of liquid cooling. For the most part, this is true. However, with traditional cooling methods (which use air and water), the lowest temperature you can achieve is room temperature. Plus, with liquid cooling, the pump is submerged in the coolant (generally speaking), so as it works, it produces heat, which adds to the overall liquid temperature.

The main benefit to liquid cooling is silence. Only one fan is needed: the fan on the radiator to cool the water. So, a liquid-cooled system can run extremely quietly.

There are two major classifications of liquid cooling systems in use with PCs today: all-in-one (AIO) coolers and custom loop systems. AIO systems are relatively easy to install—they require about as much effort as a heat sink and fan—and comparably priced to similarly effective air systems. Figure 1.42 shows an example from Corsair, with the pump in front and the fans behind it, attached to the radiator.

FIGURE 1.42 AIO liquid cooling system

AIO systems come in three common sizes: 120 mm (with one fan, and the most common), 240 mm (two fans, for overclocked components), and 360 mm (three fans, for high-end multicore overclocked components). Options with RGB lighting are readily available if that's the style you want. Custom loop systems can quickly become complex and expensive, but many hardcore gamers swear by their performance. The components are essentially the same as those in an AIO system—there's a radiator, pump, fans, some tubes, and liquid. However, each part is purchased separately, and some assembly is required.

TIP

Water cooling systems often require more room inside the case than heat sinks, as well as special headers on the motherboard (or in the case) to support the fans. As always, check your documentation to be sure that your other hardware is compatible!

NOTE

Cooling objectives you need to know for the exam include fans, heat sinks, thermal paste and pads, and liquid cooling.

Summary

In this chapter, we took a tour of the key internal system components of a PC. You learned primarily about the "big three," which are the motherboard, processor, and memory. Included in the motherboard discussion were form factors, connector types, such as PCIe, SATA, M.2, and headers, BIOS/UEFI settings, encryption, and the CMOS battery. CPU topics included features such as compatibility, architecture, multithreading, and virtualization. With RAM you learned about different types (SODIMMs, DDR2, DDR3, DDR4, and DDR5) and concepts such as single-, dual-, triple-, and quad-channel, error correction, and parity.

Finally, the chapter ended with cooling systems, which keep the components from damaging themselves with excess heat. This chapter laid the foundation for the rest of the book, including the next few chapters on additional hardware components.

Exam Essentials

Know three form factors of system boards. Know the characteristics of and differences between ATX and ITX motherboards.

Know the components of a motherboard. Be able to describe, identify, and replace (where applicable) motherboard components, such as chipsets, expansion slots, memory slots, processor sockets, BIOS/UEFI (firmware), and CMOS batteries.

Be able to identify and differentiate motherboard connector types. Understand the differences between PCI, PCIe, SATA, eSATA, and M.2 connectors, as well as power connectors and headers.

Understand core concepts of motherboard compatibility. Know that Intel and AMD chips use different sockets and therefore are incompatible with each other. Also know differences between server, multisocket, desktop, and laptop motherboards.

Understand CPU architecture. Know the differences between x64, x86, and ARM processors, implications of single-core versus multicore CPUs, and multithreading and virtualization support.

Know what the BIOS/UEFI is responsible for. The BIOS/UEFI controls boot options, fan speeds, USB permissions, and security options such as the boot password, TPM, HSM, and Secure Boot.

Understand the purposes and characteristics of memory. Know about the characteristics that set the various types of memory apart from one another. This includes the actual types of memory, such as DRAM (which includes several varieties), SRAM, ROM, and CMOS, as well as memory packaging, such as DIMMs and SODIMMs. Also have a firm understanding of the different levels of cache memory as well as its purpose in general.

Know how to replace RAM, given a scenario. RAM must match the motherboard in form factor and in speed. For example, some motherboards will only accept DDR4 or DDR5 memory. The speed should be compatible with the motherboard, as indicated in the documentation.

Understand the purposes and characteristics of cooling systems. Know the different ways internal components can be cooled and how overheating can be prevented.

Review Questions

The answers to the chapter review questions can be found in Appendix A.

1. Which computer component contains all the circuitry necessary for other components or devices to communicate with one another?
   A. Motherboard
   B. CPU
   C. RAM
   D. Expansion bus
2. You are told by a technician over the phone that you need to bring replacement DDR4 SDRAM memory. What type of packaging is used for DDR4 SDRAM memory?
   A. 224-pin DIMM
   B. 240-pin DIMM
   C. 288-pin DIMM
   D. 296-pin DIMM
3. You need to replace memory in a desktop PC and go to purchase RAM. When you are at the store, you need to find the appropriate type of memory. What memory chips would you find on a stick of PC3-16000?
   A. DDR-2000
   B. DDR3-2000
   C. DDR3-1600
   D. PC3-2000
4. A client wants you to build a new PC for her, with a smaller case and lower power requirements. When selecting a motherboard, which form factor should you choose for a smaller size and lower power consumption?
   A. ATX
   B. AT
   C. DTX
   D. ITX
5. A motherboard failed on a desktop PC with an Intel Core i5-10600K processor. When you are looking for a replacement motherboard, which CPU socket does it need to have?
   A. LGA 1366
   B. LGA 1150
   C. LGA 1200
   D. LGA 1700

6. You have just replaced a processor in a computer and now need to add a cooling mechanism. What should you use to attach the cooling system to the processor?

   A. Heat sink

   B. Thermal paste

   C. Fan

   D. Superglue

7. A technician asks you how to get a computer's processor to run faster than it currently does. What are they trying to achieve?

   A. Hyper-Threading

   B. Overclocking

   C. Virtualization

   D. Multicore support

8. You are assembling a new computer and are looking at new cooling systems. Which of the following cooling systems requires the use of a pump?

   A. Liquid

   B. Thermal paste

   C. Heat sink

   D. Heat sink plus fan

9. Which of the following types of processors will likely generally be preferred for mobile devices?

   A. x64

   B. x86

   C. ARM

   D. LGA

10. You press the front power button on a computer and the system boots. Later, you press it briefly and the system hibernates. When you press it again, the system resumes. You press and hold the button and the system shuts down. What is this feature called?

    A. Programmable power

    B. Soft power

    C. Relay power

    D. Hot power

11. You are training new technicians to install RAM and pointing out differences between packages. Which of the following are the numbers of pins that can be found on DIMM modules used in desktop motherboards? (Choose two.)

    A. 180

    B. 184

    C. 200

D. 204

E. 232

F. 240

12. You are installing a new computer, and the user wants to use virtualization. Which hardware components need to support virtual technology for this to work properly? (Choose two.)

A. RAM

B. Motherboard

C. CPU

D. BIOS

13. You find out that a disgruntled ex-employee's computer has a boot password that must be entered before the operating system is ever loaded. There is also a password preventing your access to the BIOS utility. Which of the following motherboard components can most likely be used to return the computer to a state that will allow you to boot the system without knowing the password?

A. Cable header

B. Power reset connector

C. Toggle switch

D. Jumper

14. Your Core i5 fan has a 4-pin connector, but your motherboard only has a single 3-pin header with the CPU_FAN label. Which of the following will be the easiest solution to get the necessary cooling for your CPU?

A. Plug the 4-pin connector into the 3-pin header.

B. Buy a 4-pin to 3-pin adapter.

C. Leave the plug disconnected and use only the heat sink.

D. Add an extra chassis fan.

15. You are installing a new video card into a PCIe slot. What is the combined total throughput of a PCIe 2.0 x16 slot?

A. 500 MBps

B. 1 GBps

C. 16 GBps

D. 32 GBps

16. Which of the following allows you to perform the most complete restart of the computer without removing power?

A. Start ➢ Restart

B. Start ➢ Hibernate

C. Reset button

D. Power button

17. You have just found out that a new UEFI upgrade exists for your computer, and you obtain a copy on a flash drive to install it. Which of the following is most helpful when flashing the UEFI on a desktop computer system?
    - A. DVD-ROM
    - B. Uninterruptable power supply
    - C. Internet connection
    - D. Windows administrator password

18. The 32 GB of DDR4 memory has failed in a client's desktop computer. The client requests that you upgrade the system for better performance when replacing the RAM. Which of the following is the best option?
    - A. Install two new 32 GB DDR4 RAM modules.
    - B. Install two new 128 GB DDR4 RAM modules.
    - C. Install two new 32 GB DDR5 RAM modules.
    - D. Install two new 128 GB DDR5 RAM modules.

19. You are replacing an HP laptop's motherboard. Which type of motherboard should you use?
    - A. ATX
    - B. AT
    - C. One for the specific model
    - D. mini-ITX

20. You need to install a new video card into a desktop computer. What type of expansion slot is preferred today for high-performance graphics adapters?
    - A. AGP
    - B. PCIe
    - C. PCI
    - D. SATA

Performance-Based Question 1

You will encounter performance-based questions on the A+ exams. The questions on the exam require you to perform a specific task, and you will be graded on whether or not you were able to complete the task. The following requires you to think creatively in order to measure how well you understand this chapter's topics. You may or may not see similar questions on the actual A+ exams. To see how your answer compares to the authors', refer to Appendix B.

You have been asked to remove a dual in-line memory module and insert one with a larger capacity in its place. Describe the process for doing so.

Performance-Based Question 2

Identify the component each arrow points to in the following image of an ATX motherboard.