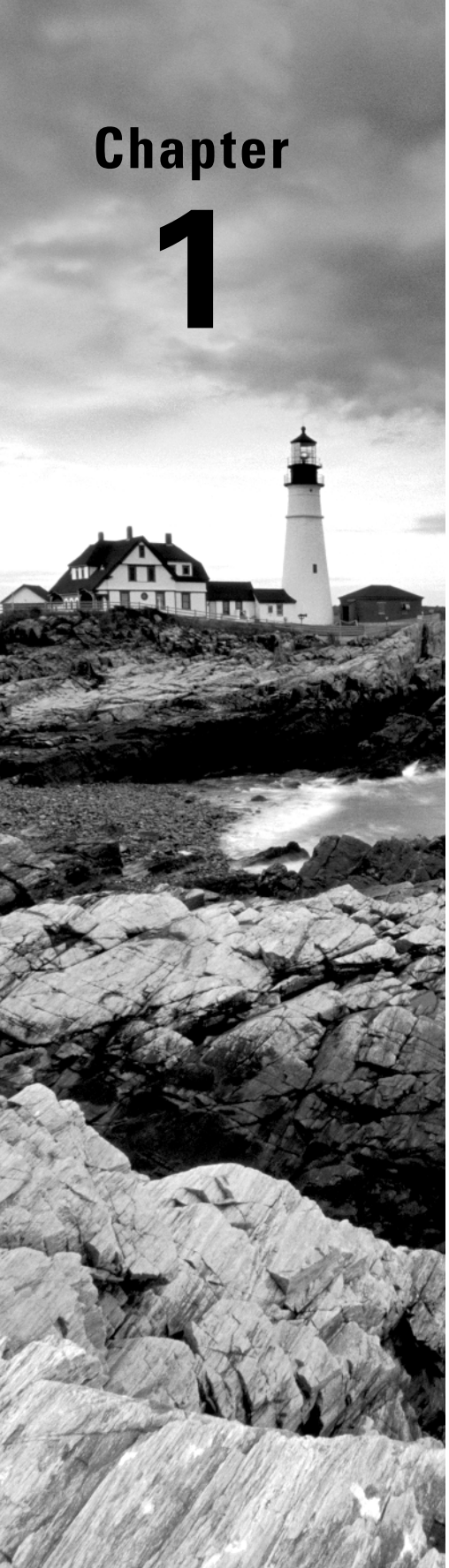
Part I: 220-901

PART

I



Chapter 1

Motherboards, Processors, and Memory

THE FOLLOWING COMPTIA A+ 220-901 OBJECTIVES ARE COVERED IN THIS CHAPTER:

✓ **1.1 Given a scenario, configure settings and use BIOS/UEFI tools on a PC.**

- Install firmware upgrades – flash BIOS
- BIOS component information: RAM, Hard drive, Optical drive, CPU, Boot sequence, Enabling and disabling devices, Date/time, Clock speeds, Virtualization support
- BIOS security (passwords, drive encryption: TPM, lo-jack, secure boot)
- Use built-in diagnostics
- Monitoring: Temperature monitoring, Fan speeds, Intrusion detection/notification, Voltage, Clock, Bus speed

✓ **1.2 Explain the importance of motherboard components, their purposes, and properties.**

- Sizes: ATX, Micro-ATX, Mini-ITX, ITX)
- Expansion slots: PCI, PCI-X, PCIe, miniPCI
- RAM slots
- CPU sockets
- Chipsets: Northbridge/Southbridge, CMOS battery
- Power connections and types
- Fan connectors
- Front/top panel connectors: USB, Audio, Power button, Power light, Drive activity lights, Reset button
- Bus speeds

✓ **1.3 Compare and contrast RAM types and features.**

- Types: DDR, DDR2, DDR3, SODIMM, DIMM, Parity vs. non-parity, ECC vs. non-ECC, RAM configurations (Single channel vs. dual channel vs. triple channel), Single sided vs. double sided, Buffered vs. unbuffered
- RAM compatibility and speed

✓ **1.6 Differentiate among various CPU types and features, and select the appropriate cooling method.**

- Socket types: Intel (775, 1155, 1156, 1366, 1150, 2011), AMD (AM3, AM3+, FM1, FM2, FM2+)
- Characteristics (Speeds, Cores, Cache size/type, Hyper-threading, Virtualization support, Architecture [32-bit vs. 64-bit], Integrated GPU, Disable execute bit)
- Cooling (Heat sink, Fans, Thermal paste, Liquid-based, Fanless/passive)

A personal computer (PC) is a computing device made up of many distinct electronic components that all function together in order to accomplish some useful task, such as adding up the numbers in a spreadsheet or helping you write a letter. Note that this definition describes a computer as having many distinct parts that work together. Most computers today are modular. That is, they have components that can be removed and replaced with another component of the same function but with different specifications in order to improve performance. Each component has a specific function. In this chapter, you will learn about the core components that make up a typical PC, what their functions are, and how they work together inside the PC.

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

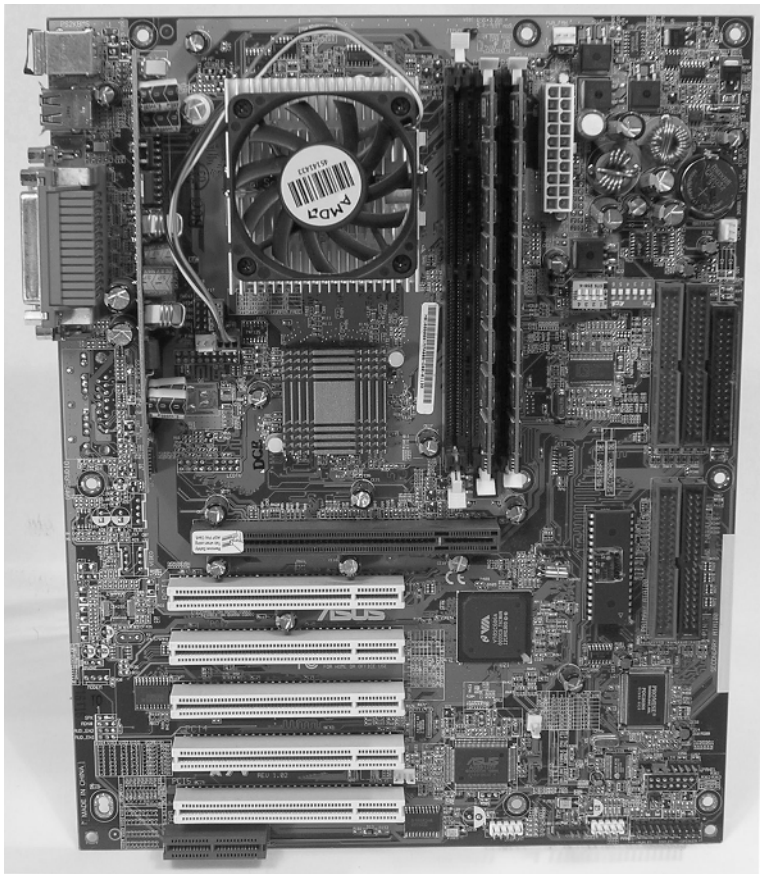
In this chapter, you will learn how to identify system components common to most personal computers, including the following:

- Motherboards
- Processors
- Memory
- Cooling systems

Identifying Components of 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 olive, brown, or blue. It is the most important component in the computer because it connects all of 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), 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.

FIGURE 1.1 A typical system board



System Board Form Factors

System boards are classified by their form factor (design), such as ATX, micro ATX, and ITX. Exercise care and vigilance when acquiring a motherboard and case separately. Some cases are less accommodating than others, and they might not be 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. 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) is the primary motherboard in use today. Standard ATX motherboards measure 12″ × 9.6″ (305mm × 244mm).

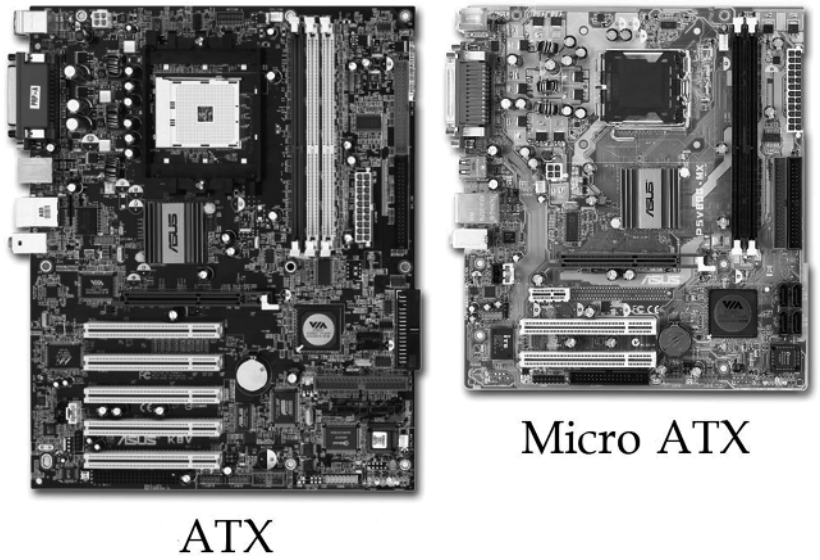
Micro ATX

A form factor that is designed to work in standard ATX cases, as well as its own smaller cases, is known as *micro ATX* (also referred to as µATX). Micro ATX follows the ATX principle of component placement for enhanced cooling over pre-ATX designs but with a smaller footprint. Some trade-offs come with this smaller form. For the compact use of space, you must give up quantity; that is, quantity of memory slots, motherboard headers, expansion slots, and integrated components. You also have fewer micro ATX chassis bays, although the same small-scale motherboard can fit into much larger cases if your original peripherals are still a requirement.

Be aware that micro ATX systems tend to be designed with power supplies of lower wattage in order to help keep power consumption and heat production down. This is generally acceptable with the standard, reduced micro ATX suite of components. As more off-board USB ports are added and larger cases are used with additional in-case peripherals, a larger power supply might be required.

Micro ATX motherboards share their width, mounting hole pattern, and rear interface pattern with ATX motherboards but are shallower and square, measuring 9.6″ × 9.6″ (244mm × 244mm). They were designed to be able to fit into full-size ATX cases. Figure 1.2 shows a full-size ATX motherboard next to a micro ATX motherboard.

FIGURE 1.2 ATX and micro ATX motherboards



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

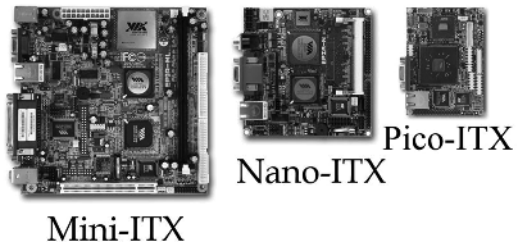
ITX

The *ITX* line of motherboard form factors was developed by VIA as a low-power, small form factor (SFF) board for specialty uses, such as home-theater 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″ (170mm × 170mm)
- Nano-ITX—4.7″ × 4.7″ (120mm × 120mm)
- Pico-ITX—3.9″ × 2.8″ (100mm × 72mm)
- Mobile-ITX—2.4″ × 2.4″ (60mm × 60mm)

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 chassis. 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, and lack the requisite mounting and interface specifications. Figure 1.3 shows the three larger forms of ITX motherboard.

FIGURE 1.3 ITX motherboards



VIA Mainboards 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 components found on the motherboard and their locations relative to each other. Many of the following components can be found on a typical motherboard:

- Chipsets
- Expansion slots and buses
- Memory slots and external cache
- CPUs and their sockets
- Power connectors
- Onboard disk drive connectors

- Keyboard connectors
- Integrated peripheral ports and headers
- BIOS/firmware
- CMOS battery
- Front-panel connectors

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.

Before we can talk about specific components, however, you need to understand the concepts underlying serial and parallel connectivity, the two main categories of bus architecture.

Bus Architecture

There has been a quiet revolution taking place in the computer industry for quite some time now. Unlike in the early days of personal computing, when parallel communication pathways (made up of multiple synchronized wires or traces) dominated single-file serial connections, this revolution has brought a shift toward serial communications. Once engineers created transmitters and receivers capable of sustaining data rates many times those of parallel connections, they found no need to tie these pathways together in a parallel circuit. The downside of parallel communications is the loss of circuit length and throughput—how far the signal can travel and the amount of data moved per unit of time, respectively—due to the careful synchronization of the separate lines, the speed of which must be controlled to limit skewing the arrival of the individual signals at the receiving end.

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 migration toward the dominance of serial communications are Serial ATA (SATA), Universal Serial Bus (USB), IEEE 1394/FireWire, and Peripheral Component Interconnect Express (PCIe).

Parallel computer-system components work on the basis of a bus. A *bus*, in this sense, is a common collection of signal pathways over which related devices communicate within the computer system. Slots are incorporated at certain points in expansion buses of various architectures, such as PCI, to allow for the insertion of external devices, or adapters, usually with no regard as to which adapters are inserted into which slots; insertion is generally arbitrary. Other types of buses exist within the system to allow communication between the CPU and components with which data must be exchanged. Except for CPU slots and sockets and memory slots, there are no insertion points in such closed buses because no adapters exist for such an environment.

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.

The various buses throughout a given computer system can be rated by their bus speeds. The higher the bus speed, the higher the performance of which the bus is capable. In some cases, various buses must be synchronized for proper performance, such as the system bus and any expansion buses that run at the frontside-bus speed. Other times, one bus will reference another for its own speed. The internal bus speed of a CPU is derived from the frontside-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. Typically, the manufacturer and model also tell you that your particular chipset has a certain set of features (for example, type of RAM supported, type and brand of onboard video, and so on).

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.

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.

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 *frontside 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 *backside bus (BSB)*, if present, is a set of signal pathways between the CPU and Level 2 or Level 3 (external) cache memory. The BSB uses the same clock signal that drives the FSB. If no backside bus exists, cache is placed on the frontside 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 (PS/2, parallel ports, serial ports, Serial and Parallel ATA, 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 PS/2, USB, LAN, analog and digital audio, and FireWire 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 and Expansion." The Southbridge is also responsible for managing communications with the slower expansion buses, such as PCI, and legacy buses.

Figure 1.4 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 1.4 A modern computer chipset

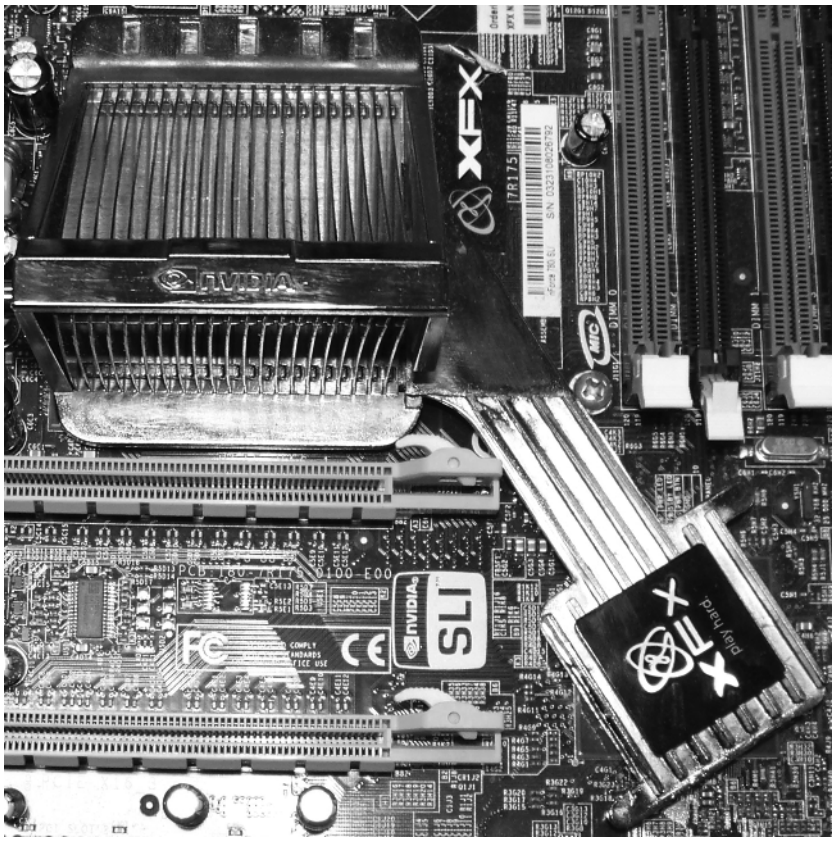
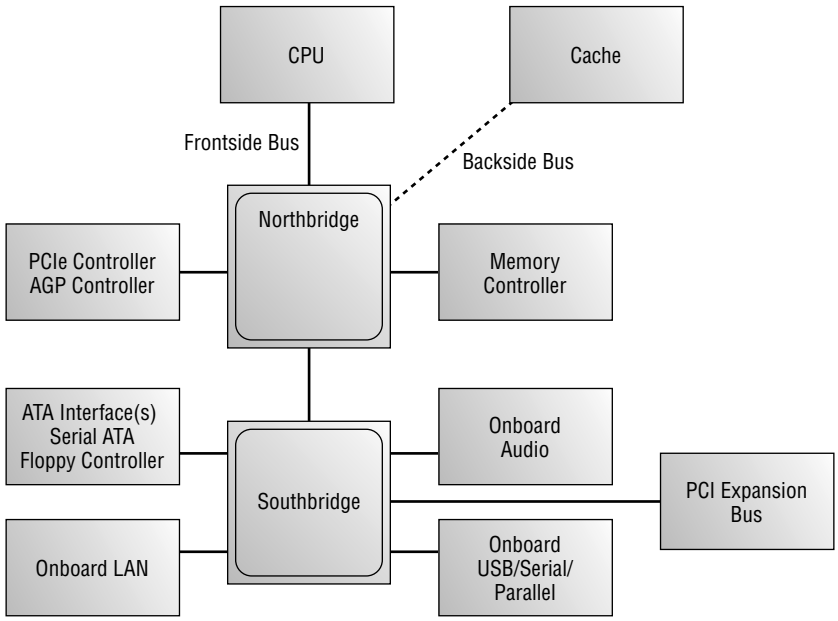


Figure 1.5 shows a schematic of a typical motherboard chipset (both Northbridge and Southbridge) and the components with which they interface. Notice which components interface with which parts of the chipset.

FIGURE 1.5 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 1/2-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:

- PCI
- PCIe
- PCI-X

Each type differs in appearance and function. In the following sections, we will cover how to visually identify the different expansion slots on the motherboard. Personal Computer Memory Card International Association (PCMCIA) buses, such as PC Card, CardBus, Mini PCI, ExpressCard, and PCIe Mini, are more related to laptops than to desktop computers, and they are covered in Chapter 9, "Understanding Laptops."

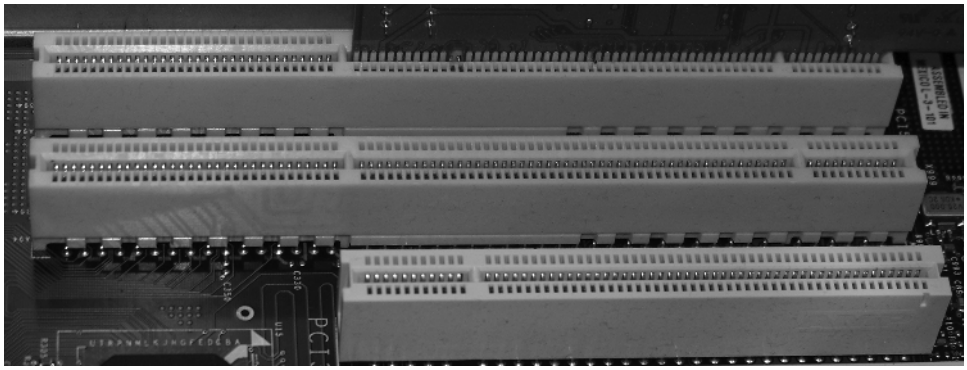
PCI Expansion Slots

The motherboards of many computers in use today 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. 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 33MHz or 66MHz (version 2.1) over a 32-bit (4-byte) channel, resulting in data rates of 133MBps and 266MBps, respectively, with 133MBps being the most common, server architectures excluded. PCI is a shared-bus topology, however, so mixing 33 MHz and 66MHz adapters in a 66MHz system will slow all adapters to 33MHz. 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 in this chapter titled "Arriving at the Exact Answer" for help with understanding the math involved in frequencies and bit rates.

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 1.6 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 1.6 PCI expansion slots



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-X Expansion Slots

Visually indistinguishable from 64-bit PCI, because it uses the same slots, *PCI-Extended (PCI-X)* takes the 66MHz maximum frequency of PCI to new heights. Version 1.0 of the specification provided a 66MHz (533MBps) implementation as well as the most commonly deployed PCI-X offering, 133MHz (1066MBps). Version 2.0 introduced the current—and likely final—maximum, 533MHz. With an 8-byte (64-bit) bus, this translates to maximum throughput of 4266MBps, roughly 4.3GBps. Additionally, PCI-X version 2.0 supports a 266MHz (2133MBps) bus. Because PCI-X slots are physically compatible with PCI adapters, and because all PCI-X slots support the 66MHz minimum clock rate, PCI-X slots are compatible with 66MHz PCI adapters.

PCI-X is targeted at server platforms with its speed and support for hot-plugging, but it is still no match for the speeds available with PCIe, which all but obviates PCI-X today and made PCI-X version 2.0 obsolete not long after its release. PCI-X also suffers from the same shared-bus topology as PCI, resulting in all adapters falling back to the frequency of the slowest inserted adapter.

PCIe Expansion Slots

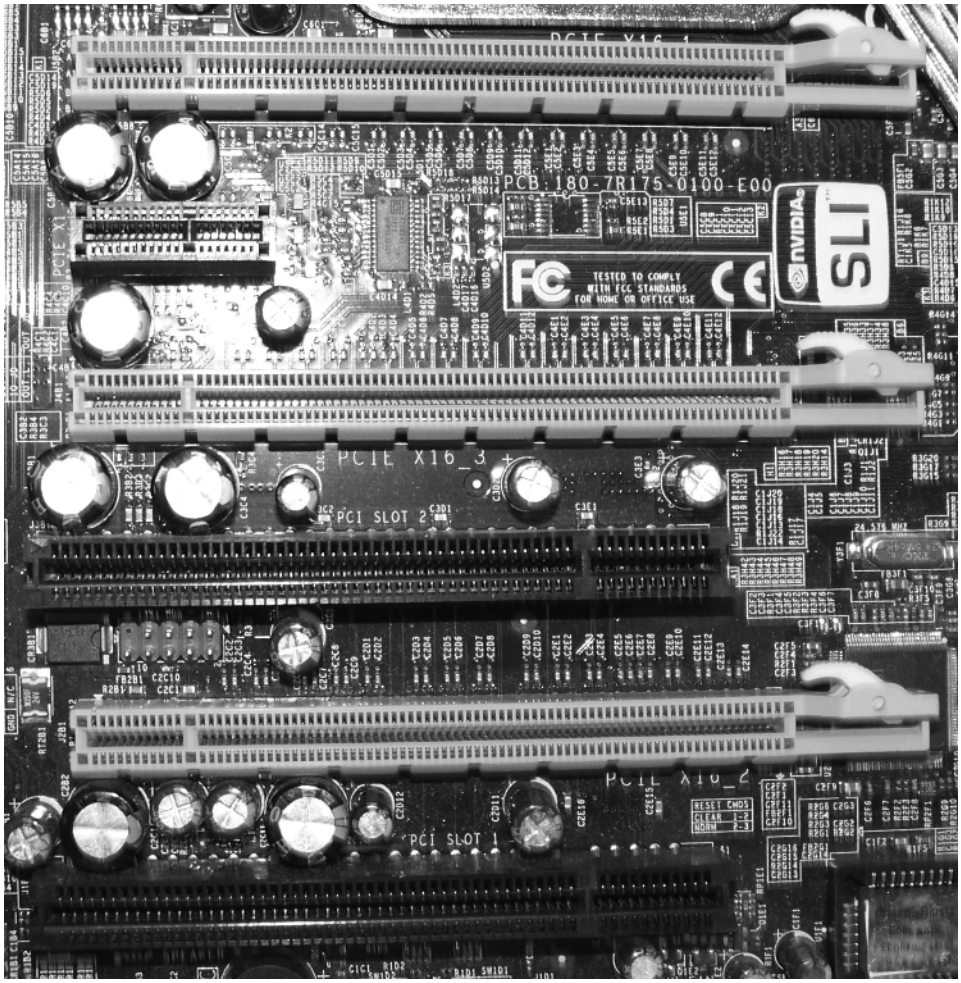
The latest expansion slot architecture that is being used by motherboards is *PCI Express (PCIe)*. It was designed to be a replacement for AGP, or accelerated graphics port, and PCI. PCIe has the advantage of being faster than AGP while maintaining the flexibility of PCI. PCIe has no plug compatibility with either AGP or PCI. As a result, 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, a x8 slot is longer than a x1 slot but shorter than a 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.

FIGURE 1.7 PCIe expansion slots



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

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 500MBps bidirectional link of a x1 slot in the first version of PCIe was comparable to PCI's best, a 64-bit slot running at 66MHz and producing a throughput of 533MBps.

Combining lanes results in a linear multiplication of these rates. For example, a PCIe 1.1 x16 slot is capable of 4GBps of throughput in each direction, 16 times the 250MBps x1 rate. Bidirectionally, this fairly common slot produces a throughput of 8GBps. Later PCIe specifications increase this throughput even more.

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 a x8 card into a 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 SLI master. SLI implementation results in increased graphics performance over single-PCIe and non-PCIe implementations.

Refer back to Figure 1.7, 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.

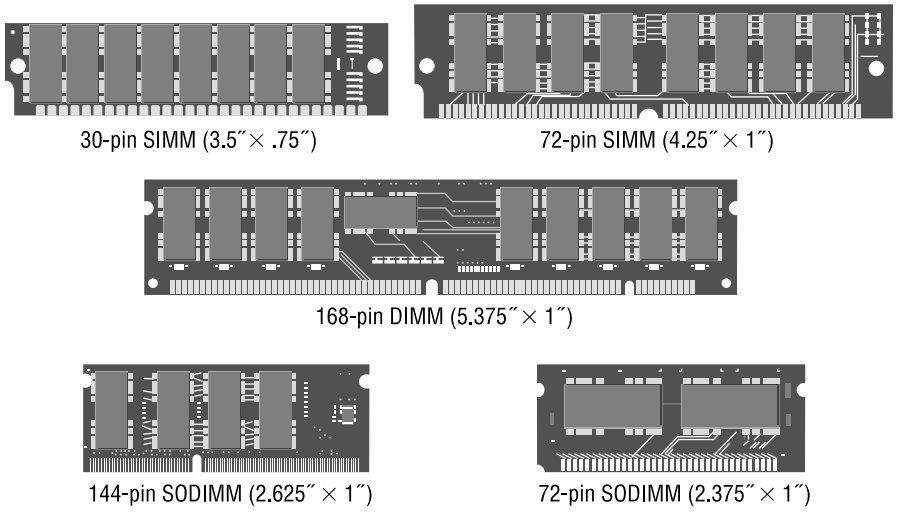
Memory Slots and Cache

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 and varied 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.

For the most part, PCs today use memory chips arranged on a small circuit board. A *dual inline 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-, and 240-pin configurations. In addition, laptop memory comes in smaller form factors known as *small outline DIMMs (SODIMMs)* and MicroDIMMs. The single inline memory module (SIMM) is an older memory form factor that began the trend of placing memory chips on modules. More detail on memory packaging and the technologies that use them can be found later in this chapter in the section "Identifying Purposes and Characteristics of Memory." Figure 1.8 shows the form factors for some once-popular memory modules. Notice how they basically look the same, but that the module sizes and keying notches are different.

FIGURE 1.8 Different memory module form factors



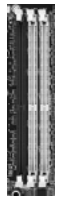
Memory slots are easy to identify on a motherboard. Classic DIMM slots were usually black and, like all memory slots, were placed very close together. DIMM slots with color-coding are more common these days, however. The color-coding of the slots acts as a guide to the installer of the memory. See the section "Single-, Dual-, and Triple-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.

Table 1.1 identifies the types of memory slots that can be seen in various images throughout this chapter. All slots listed have the characteristic release-tabs at each end in common. The ATX motherboard in Figure 1.2 and the motherboard in Figure 1.4 exhibit dual-channel indicators, by way of dual-colored slots. Figure 1.9 shows only the memory slots from the ATX motherboard in Figure 1.2 as a reference.

TABLE 1.1 List of memory slots in Chapter 1 images

Image number	Type of memory slot
Figure 1.1	168-pin SDR SDRAM
Figure 1.2	184-pin DDR SDRAM
Figure 1.3 (mini-ITX)	184-pin DDR SDRAM
Figure 1.4 (partial visibility)	240-pin DDR2 SDRAM
Figure 1.12 (partial visibility)	168-pin SDR SDRAM
Figure 1.22	168-pin SDR SDRAM

FIGURE 1.9 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*. The swap file, PAGEFILE.SYS in modern Microsoft operating systems, 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. 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."

When it's not the amount of RAM in a system that you need to enhance but its speed, engineers can add *cache memory* 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 "Identifying Purposes and Characteristics of 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. Cache on the motherboard is known as external cache because it is external to the processor; it's also referred to as Level 2 cache (*L2 cache*). Level 1 cache (*L1 cache*), by comparison, is internal cache because it is built into the processor's silicon wafer, or *die*. The word *core* is often used interchangeably with the word *die*.

It is now common for chipmakers to use extra space in the processor's packaging to bring the L2 cache from the motherboard closer to the CPU. When L2 cache is present in the processor's packaging, but not on-die, the cache on the motherboard is referred to as Level 3 cache (*L3 cache*). Unfortunately, due to the de facto naming of cache levels, the term *L2 cache* alone is not a definitive description of where the cache is located. The terms *L1 cache* and *L3 cache* do not vary in their meaning, however.

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, "Storage Devices and Power Supplies"). 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 original Intel Core i7 processor. The other capacities are simply modern examples.

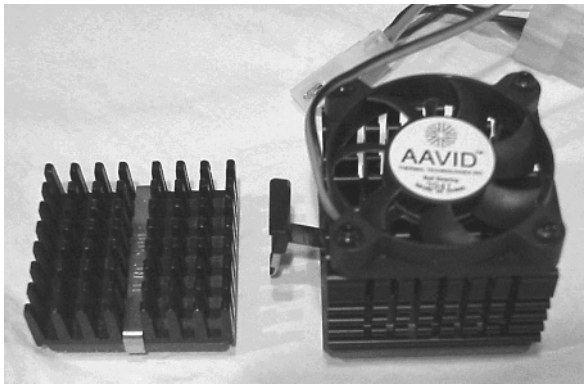
- L1 cache—64KB (32KB each for data and instructions)
- L2 cache—256KB
- L3 cache—4MB–12MB
- RAM—4–16GB
- HDD/SSD—100s–1000s of GB

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 "Identifying Purposes and Characteristics of 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. Theoretically, a Pentium (or higher) processor generates enough heat that, without the heat sink, it would permanently damage itself and the motherboard in a matter of hours or even minutes.

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 top socket is known as Socket A or Socket 462, made for earlier AMD processors such as the Athlon, 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 bottom socket is known as Socket T or Socket *LGA 775*, 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 "Identifying Purposes and Characteristics of Processors."

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 462 shown in Figure 1.11 illustrates the ZIF locking mechanism at the edge of the socket along the bottom of the photo.

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 bottom image in Figure 1.11 shows an LGA socket with no CPU installed and the locking arm secured over the lid's tab (right-edge in the photo).

FIGURE 1.11 CPU socket examples

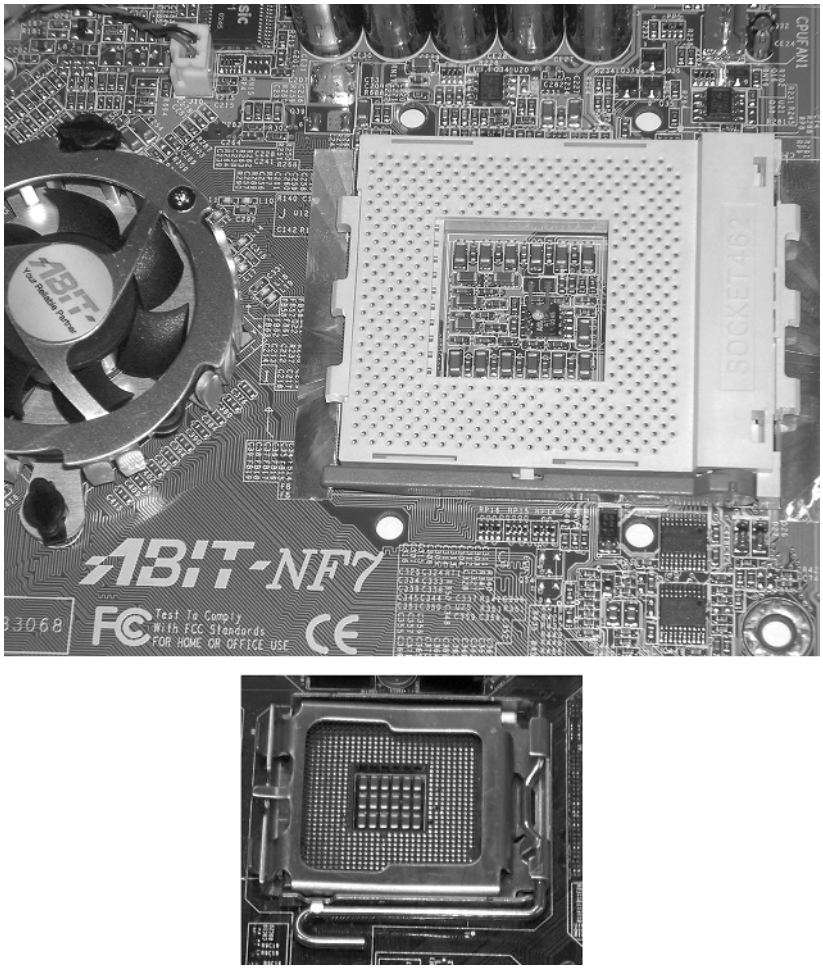


Table 1.2 lists some common socket/CPU relationships.

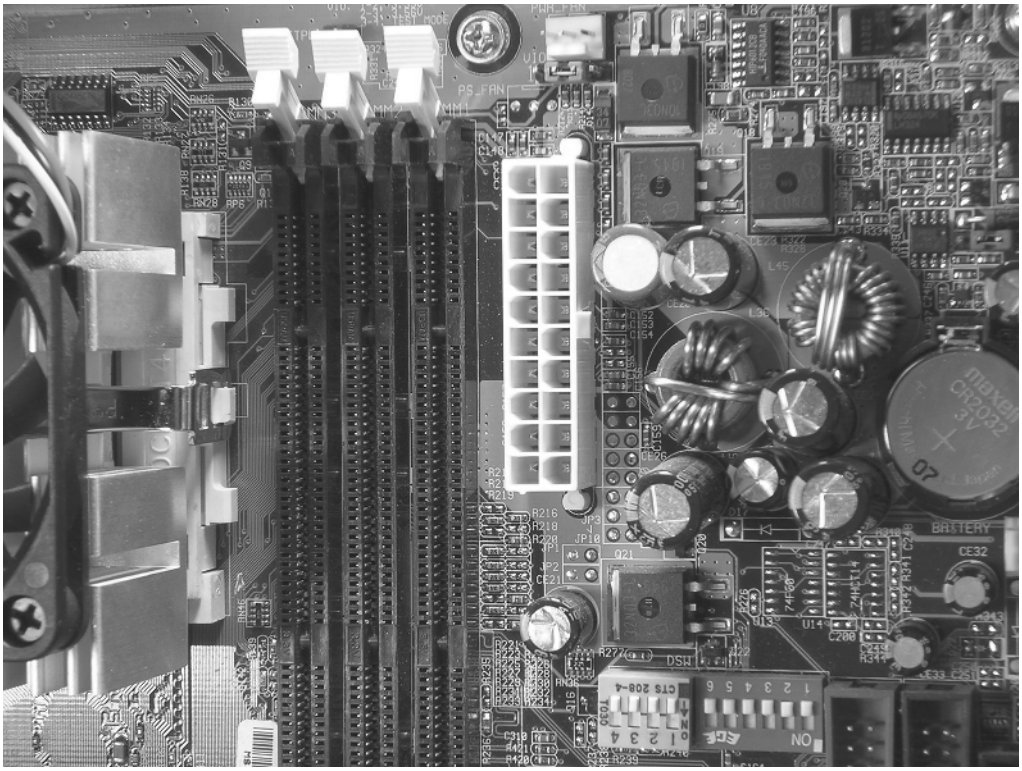
TABLE 1.2 Socket types and the processors they support

Socket	Processors
LGA 775 (Socket T)	Intel only: Pentium 4, Pentium 4 Extreme Edition (single core), Pentium D, Celeron D, Pentium Extreme Edition (dual core), Core 2 Duo, Core 2 Extreme, Core 2 Quad, Xeon, Celeron (4xx, Exxxx series).
LGA 1156 (Socket H)	Intel only: Celeron (G1xxx series), Core i3, Core i5, Core i7 (8xx series), Pentium (G6xxx series), Xeon (34xx series).
LGA 1155 (Socket H2)	Intel only: Replacement for LGA 1156 to support CPUs based on the Sandy Bridge (such as Celeron G4xx and G5xx) and eventual Ivy Bridge architectures.
LGA 1150 (Socket H3)	Replacement for LGA 1155 to support the Haswell and Broadwell processors, which rely on x8x and x9x desktop chipsets and C22x single-Xeon server chipsets.
LGA 1366 (Socket B)	Intel only: Core i7 (9xx series), Xeon (35xx, 36xx, 55xx, 56xx series), Intel Celeron P1053.
LGA 2011 (Socket R)	Replacement for LGA 1366. ▪ Original LGA 2011-0 socket is used for Sandy Bridge-E (desktop Core i7-38xx, -39xx) and -EP (Xeon E5) as well as Ivy Bridge-E (desktop Core i7-48xx and -49xx) and -EP processors (Xeon E5 v2). ▪ LGA 2011-1 is used for Ivy Bridge-EX (Xeon E7 v2) CPUs. ▪ LGA 2011-v3 socket is used for Haswell-E (desktop, X99 chipset) and Haswell-EP (Xeon E5 v3) CPUs, supporting DDR4 memory. ▪ The three sockets are not electrically compatible.
Socket AM3	AMD only: DDR3 capable CPUs only (thus not compatible with AM2+ CPUs), such as Phenom II, Athlon II, Sempron, Opteron 138x, and has the potential to accept AM3+ CPUs.
Socket AM3+	AMD only: Specified for CPUs based on the Bulldozer microarchitecture and designed to accept AM3 CPUs.
Socket FM1	AMD only: Designed to accept AMD Fusion Accelerated Processing Units (APUs), for desktops, which incorporate CPUs and GPUs, such as early A Series APUs.
Socket FM2	AMD only: A 904-pin desktop socket for Trinity and Richland APUs.
Socket FM2+	AMD only: A 906-pin desktop socket for Steamroller-based APUs Kaveri and Godavari. Accepts chips that fit socket FM2 as well, but the converse is not also true.

Power Connectors

In addition to these sockets and slots on the motherboard, a special connector (the 20-pin white block connector shown in Figure 1.12) 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 "Identifying Purposes and Characteristics of Power Supplies") plugs in.

FIGURE 1.12 An ATX power connector on a motherboard



Firmware

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) routine, 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.

BIOS and POST

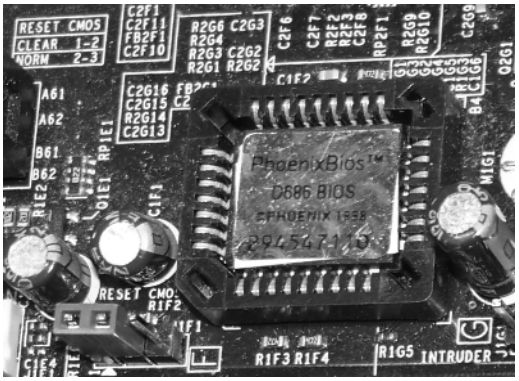
One of the most important chips on the motherboard is the *basic input/output system (BIOS)* chip, also referred to as the ROM BIOS chip. This special memory chip contains the BIOS system software that boots the system and allows the operating system to interact with certain hardware in the computer 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 so on). 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.

BIOS

Figure 1.13 gives you an idea of what a modern BIOS might look like. 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 lower left and its configuration silkscreen at 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 configuration utility. The jumper in the photo is in the clear position, not the normal operating position. System boot-up is typically not possible in this state.

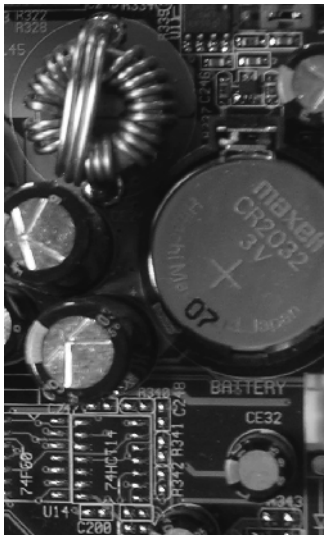
FIGURE 1.13 A BIOS chip on a motherboard



Most BIOS setup utilities have more to offer than a simple interface for making selections and saving the results. As always, you can enter the utility to check to see if the clock appears

to be losing time, possibly due to a dying battery. (See Figure 1.14.) Today, these utilities also offer diagnostic routines that you can use to have the BIOS analyze the state and quality of the same components that it inspects during boot-up, but at a much deeper level. Consider the scenario where a computer is making noise and overheating. You can use the BIOS 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.

FIGURE 1.14 A CR2032 CMOS battery



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 sounds a warning tone and the temperature at which the BIOS shuts the system down 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 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.

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 boot-up, the BIOS will notify you of the intrusion. No notification occurs over subsequent boots unless additional intrusion is detected.

The BIOS 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 password and the supervisor, or access, password. The *user password* is required to leave the initial power-on screens and begin the process of booting an operating system. The *supervisor password* is required before entering the BIOS configuration utility. It is always a good idea to set the supervisor password, but the user password should not be set on public systems that need to boot on their own in case of an unforeseen power-cycle.

In later years, the role of the BIOS in system security grew substantially. Somehow, security needed to be extended to a point before the operating system was ready to take it over. The BIOS 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 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.

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*. 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.

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, not a task possible with traditional BIOS-based firmware.

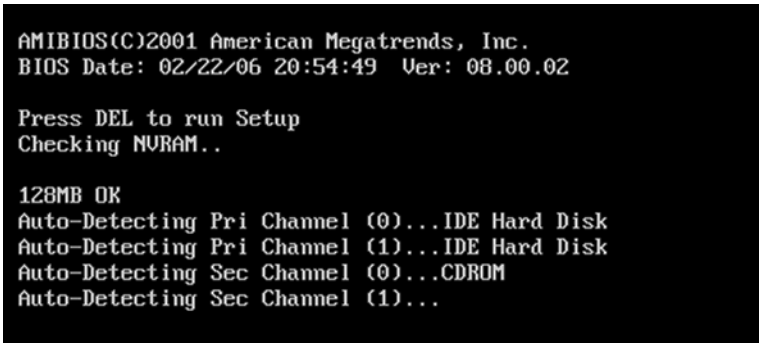
POST

A major function of the BIOS 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 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 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 managing the passing of control to the specialized BIOS routines mentioned earlier. The BIOS 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 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 publisher has its own series of codes that can be generated. Figure 1.15 shows a simplified POST display during the initial boot sequence of a computer.

FIGURE 1.15 An example of a BIOS boot screen



Flashing the System BIOS

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 upgrade, also known as *flashing the BIOS*, is in order. Only certain hardware benefits from a BIOS upgrade, such as drives and a change of CPU or RAM types. Very often, this hardware is recognized immediately by the BIOS 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 is a logical target.

Let's be clear about the fact that we are not talking about entering the BIOS 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 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 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

continues

continued

onscreen. At other times, there may be a mechanism, such as a jumper, involved in the recovery of the BIOS 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, regardless of the BIOS maker, you should not consult BIOS companies—AMI, Award, Phoenix, and so forth. Instead, go back to the motherboard or system manufacturer; check its website, for example. The motherboard or system manufacturer vendors have personalized their BIOS code after licensing it from the BIOS 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)

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 here might be to alter your BIOS configuration to disable the internal graphics adapter, so that the new one will take over. Similar reconfiguration of your BIOS settings might be necessary when overclocking—or changing the system clock speed—is desired, or when you want to set BIOS-based passwords or establish TPM-based whole-drive encryption, as with Microsoft's BitLocker (see Chapter 19, "Security"). While not so much utilized today, the system date and time can be altered in the BIOS 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. As a result, the term *CMOS* stuck and is the accepted name for this memory chip.

The BIOS 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. 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 chip or the Southbridge.

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. 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, nonrechargeable lithium chemistry.

When Absolute Software licensed the name LoJack, which was originally used as the name of a locating service for motor vehicles, the company replaced it as the name of its CompuTrace product, which allowed computing devices to be electronically tracked and controlled through a technology referred to as persistence. It's the ability to control devices that concerns many industry professionals. Because many laptop vendors incorporate the LoJack code and persistence into their BIOS firmware, there is a concern that attackers can redirect the service to rogue servers that can then gain control of legions of corporate systems. Furthermore, not all vendors incorporate the ability to disable this feature.

Front- and Top-Panel Connectors

From the time of the very first personal computer, there has been a minimum expectation as to the buttons and LEDs that should appear 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.

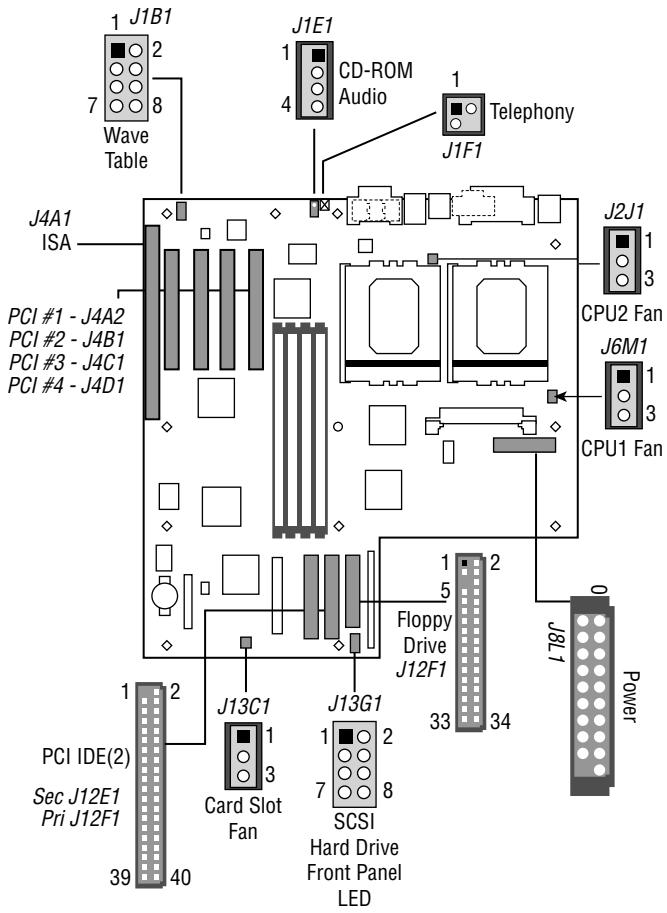
Users expect a *power button* to use to turn the computer on (these were on the side or back of very early PCs). The soft power feature available through the front power button, which is no more than a relay, allows access to multiple effects through the contact on the motherboard, based on how long 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. As time progressed, users were introduced to new things on the front panel of their computers. Each of these components depends on connectivity to the motherboard for its functionality. As a result, most motherboards have these standardized connections in common. The following list includes the

majority of these landmarks (including the power button and power light, which were just discussed):

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

So common are the various interfaces and indicators found on the front panel of today's computer chassis that the industry has standardized on a small number of connectors, making attachment to motherboards much simpler. Figure 1.16 shows a typical motherboard header. Consult the motherboard's documentation for a particular model's pin assignments.

FIGURE 1.16 The front-panel motherboard header



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. The cable 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 3.5mm 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 while others have high-definition (HD) audio connections. Motherboards that accommodate both have a BIOS setting that allows 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 applications, only 9 positions tend to have pins protruding, while the 10th position acts as a key, showing up in different spots for each application to discourage the connection of the wrong cable.

Figure 1.17 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.

FIGURE 1.17 Two motherboard USB headers

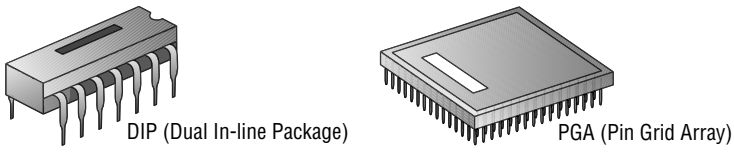
Identifying Purposes and Characteristics of Processors

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. It is a processor chip consisting of an array of *millions* of transistors. Intel and Advanced Micro Devices (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.

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 1.18. 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" below a PGA socket. Additionally, the ATX motherboard in Figure 1.2 has a PGA socket, while the micro ATX motherboard has an LGA.

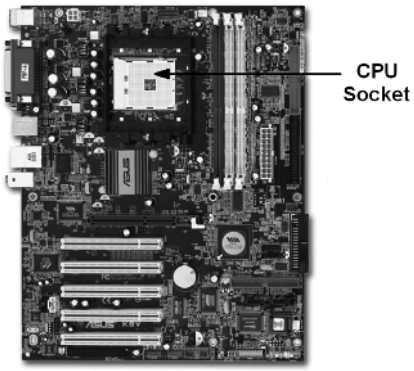
FIGURE 1.18 DIP and PGA



Intel and AMD both make extensive use of an inverted socket/processor combination of sorts. As mentioned earlier, the land grid array (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 (as shown earlier in Figure 1.10). Figure 1.19 points out the location of the CPU in relation to the other components on a typical ATX motherboard. Notice how prominent the CPU socket is.

FIGURE 1.19 The location of a CPU on a typical motherboard



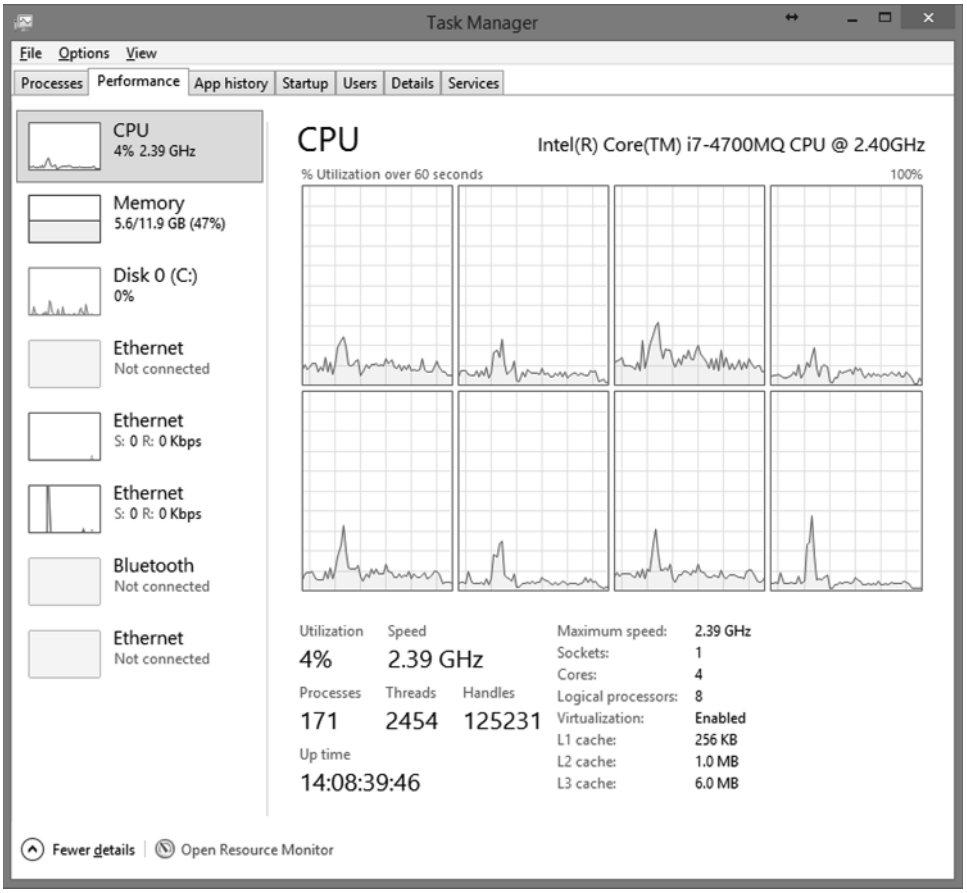
Modern processors may feature the following characteristics:

Hyperthreading This term refers to Intel's *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 8.x 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 1.20).

FIGURE 1.20 Logical processors in Windows



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.

If you have a no-name clone, look in the System Properties pages, found by right-clicking My Computer (Computer in Vista and Windows 7) and selecting Properties. The General tab, which is the default, contains this information. Even more detailed information can be found by running the System Information utility from Start ➢ Accessories ➢ System Tools or by entering **`msinfo32.exe`** in the Start ➢ Run dialog box.

Another 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 that tells you what chip you are using.

Multicore A processor that exhibits a *multicore architecture* has multiple completely separate processor dies in the same package. The operating system and applications see multiple processors in the same way that they see multiple processors in separate sockets. As with HTT, the operating system must support SMP to benefit from the separate processors. In addition, SMP is not a benefit if the applications that are run on the SMP system are not written for parallel processing. Dual-core and quad-core processors are common specific examples of the multicore technology.

Don't be confused by Intel's Core 2 labeling. The numeric component does not imply that there are two cores. There was a Core series of 32-bit mobile processors that featured one (Solo) or two (Duo) processing cores on a single die (silicon wafer). The same dual-core die was used for both classes of Core CPU. The second core was disabled for Core Solo processors.

The 64-bit Core 2 product line can be thought of as a second generation of the Core series. Core 2, by the way, reunited Intel mobile and desktop computing—the Pentium 4 family had a separate Pentium M for mobile computing. Intel describes and markets the microcode of certain processors as "Core microarchitecture." As confusing as it may sound, the Core 2 processors are based on the Core microarchitecture; the Core processors are not. Core 2 processors come in Solo (mobile only), Duo, and four-core (Quad) implementations. Solo and Duo processors have a single die; Quad processors have two Duo dies. A more capable Extreme version exists for the Duo and Quad models.

Processors, such as certain models of AMD's Phenom series, can contain an odd number of multiple cores as well. The triple-core processor, which obviously contains three cores, is the most common implementation of multiple odd cores.

Throttling CPU throttling allows for reducing the operating frequency of the CPU during times of less demand or during battery operation. CPU throttling is very common in processors for mobile devices, where heat generation and system-battery drain are key issues of full power usage. You might discover throttling in action when you use a utility that reports a lower CPU clock frequency than expected. If the load on the system does not require full-throttle operation, there is no need to push such a limit.

Speed The speed of the processor is generally described in clock frequency (MHz or 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 frontside bus uses the clock leads to an *effective* clock rate known as the FSB speed. As shown in the section "Types of Memory" later in this chapter, the FSB speed is computed differently for different types of RAM (DDR, DDR2, 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 frontside 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 frontside 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 frontside bus rated, for instance, at 1333MHz. Furthermore, you might be able to adjust the internal clock rate of the CPU that you purchased through settings in the BIOS. The successful technician needs to be familiar with more basic information than this, however. The sidebar titled "Matching System Components" explains these basics.

Matching System Components

In a world of clock doubling, tripling, quadrupling, and so forth, it becomes increasingly important to pay attention to what you are buying when you purchase CPUs, memory, and motherboards a la carte. The only well-known relationship that exists in the marketplace among these components is the speed of the FSB (in MHz) and the throughput of the memory (in MBps). Because 8 bytes are transferred in parallel by a processor with a 64-bit (64 bits = 8 bytes) system data bus, you have to know the FSB rating before you choose the RAM for any particular modern motherboard. For example, an FSB of 800MHz requires memory rated at a throughput of 6400MBps (800 million cycles per second × 8 bytes per cycle).

Matching CPUs with motherboards or CPUs with memory requires consulting the documentation or packaging of the components. Generally, the CPU gets selected first. Once you know the CPU you want, the motherboard tends to come next. You must choose a motherboard that features a socket compatible with your chosen CPU. The FSB or QuickPath Interconnect (QPI) used on the selected motherboard/CPU dictates the RAM that you should purchase.

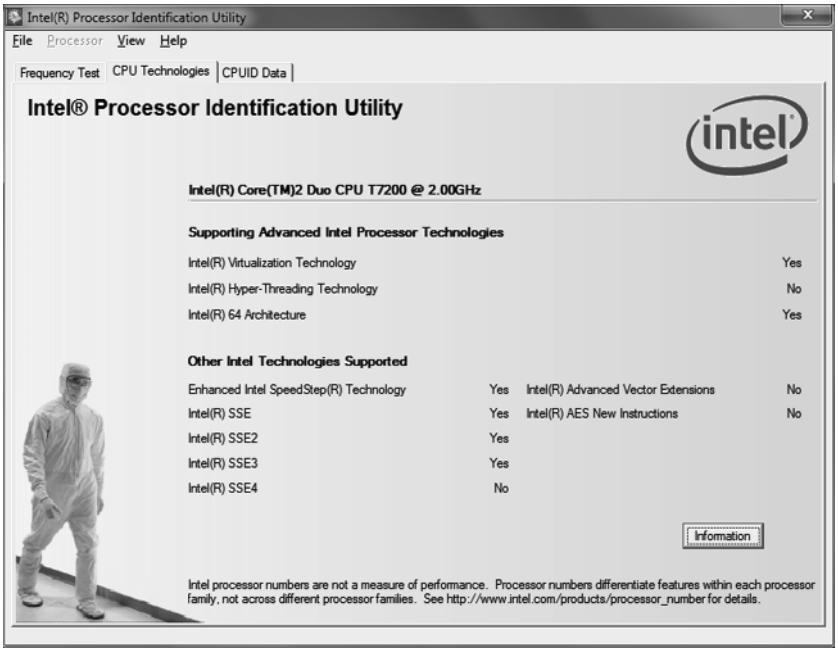
32- and 64-bit processors The set of data lines between the CPU and the primary memory of the system can be 32 or 64 bits wide, among other widths. The wider the bus, the more data that can be processed per unit of time, and hence, more work can be performed. Internal registers in the CPU might be only 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.

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 20, "Network Services, Cloud Computing, and Virtualization." Unlike AMD's AMD-V (*V* for virtualization) technology, which is widely inclusive of AMD's CPUs, Intel's Virtualization Technology (VT) is used by Intel to segment its market for CPUs made concurrently. For example, you can find Intel VT on the Core 2 Duo processor in the E6000 series and most of the E8000 series but not in the E7000 series. In some cases, you must also first enable the virtualization support in the BIOS 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:

`downloadcenter.intel.com/Detail_Desc.aspx?ProductID=1881&DwnldID=7838`

As shown in Figure 1.21, the CPU Technologies tab of this utility tells you if your CPU supports Intel VT.

FIGURE 1.21 Intel Processor Identification utility



Integrated GPU Intel and AMD both have a line of low-power CPUs, originally aimed at the netbook and embedded markets, that have built-in graphics processing units (GPUs). Building in specialized functionality to CPUs is nothing new, but before now, math coprocessors were some of the most complex features added on to the die of CPUs. A GPU, then, which is normally a large chip on your graphics adapter, is quite a bit more complex than anything heretofore integrated into the CPU. Integrated GPUs take much of the burden off of the CPU itself in addition to minimizing the amount of off-package communication that must occur, which improves overall system performance. As if that were not enough, the CPUs in this class are quite a bit smaller than standard CPUs. The Intel Atom and AMD Fusion (now simply APU for Accelerated Processing Unit) lines of CPUs have built-in GPUs and open the door for other complex systems to be built into future processors.

Disable execute bit Modern CPUs respond to the operating system's setting of the *disable execute bit*, more accurately known as the *no-execute (NX) bit*, for an area of memory by refusing to execute any code placed into that memory location. The result is that malicious buffer overrun attacks are less likely to succeed. A similar, but non NX-based, support feature has been in existence since the Intel 80286 processor. Use of the NX bit provides more granular linear addressing. In contrast, the 286 applied protection to entire segments at a time. Windows began operating-system NX support with Windows XP, calling it Data Execution Prevention (DEP). Intel refers to the NX bit as the eXecute Disable (XD) bit.

Identifying Purposes and Characteristics of Memory

"More memory, more memory, I don't have enough memory!" Today, 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 1.22 shows where memory is located in a system.

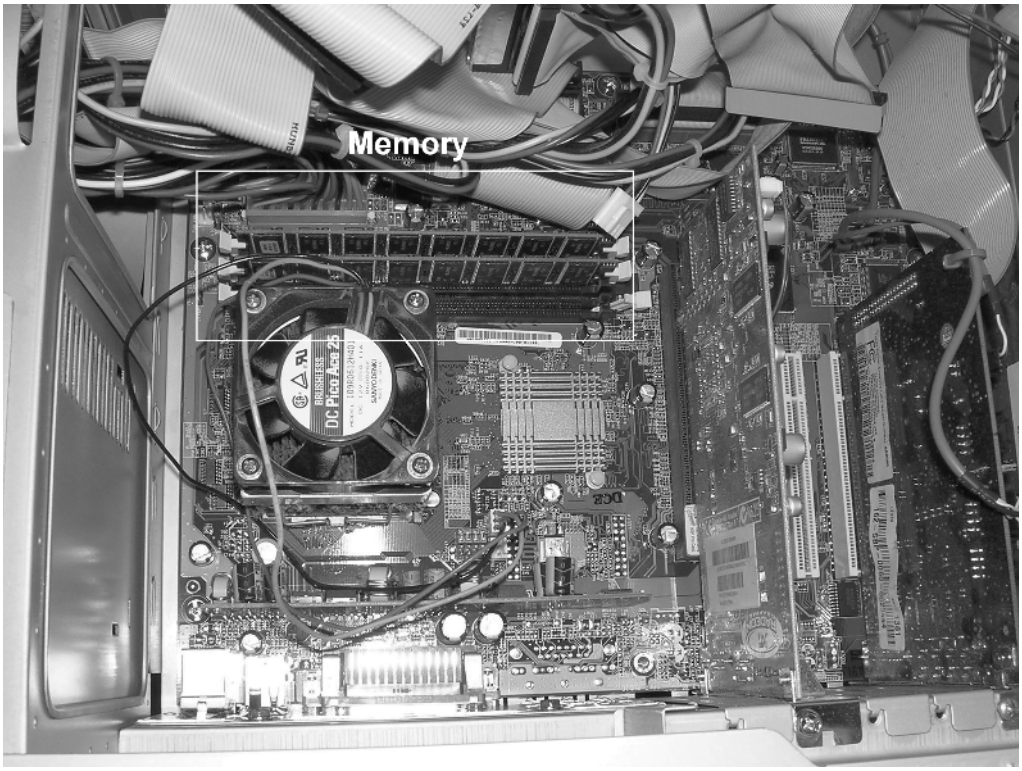
Important Memory Terms

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

- Parity checking
- Error-correcting code (ECC)

- Single- and double-sided memory
- Single-, dual-, and triple-channel memory
- Buffered and unbuffered memory

FIGURE 1.22 Location of memory within a system



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 inline memory module) satisfies a bank for an old 32-bit CPU, such as a 386 or 486 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 nybble, 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 are 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 rare. 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-correcting 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 while 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-, and Triple-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- and triple-channel 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.

The major difference between dual- and triple-channel architectures is that triple-channel memory employs a form of interleaving that reduces the amount of information transferred by each module. Nevertheless, there is an overall performance increase over that of dual-channel memory because of the ability to access more information per unit of time with triple-channel memory.

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 dual- and triple-channel memory in today's most popular computer systems requires that pairs or triads of memory modules be installed at a time.

Note, however, that it's the motherboard, not the memory, that implements dual- and triple-channel 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. The same loss of performance occurs when only two modules are installed in a triple-channel motherboard. Multichannel motherboards 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 dual-channel 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 dual- or triple-channel memory architecture. Keep in mind, the motherboard memory slots have the distinctive color-coding, not the memory modules.

Buffered and Unbuffered Memory

In technical terms, a *buffer* is a temporary storage area that takes some of the load off of the primary circuit. For instance, a network-interface buffer can store inbound packets when the CPU is currently unable to give undivided attention to the packets, or it can store outbound packets when available network bandwidth is low or the receiver has throttled its flow control. Buffers used in this sense are a form of hardware register. Registers are characterized by multiple cells, each of which stores a bit (binary digit) of information, accessed in parallel, at the same time.

In the high-end workstation and server market, there are two types of memory that are considered appropriate for the task at hand. The two types differ in the presence of a buffer or the lack thereof between the chips and the system's memory controller.

When the ECC memory mentioned earlier is referred to only as ECC, it is a form of *unbuffered* DIMM (see the section "Memory Packaging" later in this chapter for an explanation of DIMMs). Because *buffer* and *register* are interchangeable terms, in this context, this type of memory is also referred to as unregistered, and the modules are referred to as UDIMMs.

A common misconception could very well be that the term *unregistered* implies that there is an authority that provides certification for memory modules. Instead, the term refers to the hardware registers, or buffers, present in this type of memory module.

Buffered, or registered memory modules (RDIMMs), include specialized chips that act as buffers for all signals from the memory controller, except, in some cases, the data signals. By buffering these signals, the electrical load placed on the controller is reduced because the memory controller communicates in series with the register, instead of in parallel with the memory chips. The register performs the parallel communication with the chips.

Load-reduced DIMMs (LRDIMMs) are a form of registered DIMMs that increase performance by maintaining parallel communication, thus avoiding the performance hit of the conversion from serial to parallel.

The reduced electrical load on the memory controller leads to an increase in system stability when more modules are placed in a computer system. As a result, such systems can support more memory than those containing UDIMMs.

Before you jump to the conclusion that RDIMMs are, by definition, non-ECC, consider these points. The ECC function can be present with all forms of RDIMMs as well. UDIMMs may also be non-ECC. Nevertheless, the accepted industry naming convention is that the term *ECC*, alone, refers to UDIMMs. It would be correct to say, however, that UDIMMs and RDIMMs alike can be ECC or non-ECC based.

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

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, whether in print or online, 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. Consult the motherboard's documentation.

One common point of confusion, not related to capacity, when memory is installed is 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.

Pay particular attention to all synchronous DRAM types. 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 DDR, DDR2, and DDR3. 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 frontside 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 now considered 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 100MHz clock signal produces 800MBps. 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 133MHz clock to produce 1066MBps 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 100MHz clock gives a DDR SDRAM system the impression of a 200MHz 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 1.23 explains how this relationship works.

FIGURE 1.23 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 100MHz clock and two operations per cycle, motherboard makers will market their boards as having an FSB of 200MHz. Multiplying this effective rate by 8 bytes transferred per cycle, the data rate is 1600MBps. 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 1600MBps. The chips that go into making PC1600 modules are named DDR200 for the effective FSB frequency of 200MHz. 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 400MHz, meaning an actual clock signal of 200MHz, 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 333MHz FSB (actual clock is 166MHz)? Well, just using the 8:1 rule mentioned earlier, you might be on the lookout for a PC2667 module. However, note 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 333MHz. 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 2667MBps 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 100MHz actual clock, transfers data in four operations per cycle (effective 400MHz FSB) and still 8 bytes per operation, for a total of 3200MBps. 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 667MHz 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 5333MBps 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 that was 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 133MHz at the low end to less than 300MHz. 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 1066MHz to more than 2000MHz in DDR3 systems. These memory devices are named following the conventions established earlier. Therefore, if you buy a motherboard with a 1600MHz 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 12800MBps throughput.

The earliest DDR3 chips, however, were based on a 100MHz actual clock signal, so we can build on our earlier example, which was also based on an actual clock rate of 100MHz. With eight operations per cycle, the FSB on DDR3 motherboards is rated at 800MHz, 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 6400MBps, carrying chips called DDR3-800, again named for the FSB speed.

Real World Scenario

Choosing the Right Memory for Your CPU

Picking out the right memory for your CPU and motherboard is all about understanding the minimum performance required for the CPU you choose. Sometimes, the motherboard you choose to mate with the CPU makes your decision for you. If you go with the cheaper motherboard, you might find that just a single channel of DDR2 is all you need to worry about. Otherwise, the more expensive boards might support dual- or triple-channel memory and require DDR3 modules. It's usually safe to assume that the higher price of admission gets you better performance. This is generally true on paper, but you might find that the higher-end setup doesn't knock your socks off the way you expected.

Let's say that you head down to your local computer store where motherboards, CPUs, memory, and other computer components are sold à la carte. You're interested in putting together your own system from scratch. Usually, you will have a CPU in mind that you would like to use in your new system. Assume you choose, for example, an Intel Core 2 Quad Q9650 processor. It's fast enough at 3GHz, but it calls for an older LGA 775 socket, meaning that you'll save a bit of money on that performance but you won't be approaching the state of the art. Nevertheless, the FSB with which this CPU is outfitted runs at a healthy 1333MHz, and its associated chipsets call for DDR3 memory. As a result, you

will need to purchase one or more modules that contain DDR3-1333 chips, especially if you buy a motherboard that supports dual-channel memory. Therefore, you'll be buying PC3-10600 modules (multiplying 1333 by 8 and adjusting for marketing). Recall the 8:1 module-to-chip/FSB naming convention.

Perhaps you'd prefer the pricier Intel Core i7-990X Extreme Edition six-core processor. With a little research, you discover that Intel did away with the FSB by moving the memory controller out of the Northbridge and into the CPU. What remains is what Intel calls the QPI, or QuickPath Interconnect. QPI is a PCIe-like path that uses 20 bidirectional lanes to move data exceedingly fast between the CPU and RAM, requiring less capable RAM than FSB CPUs to do the same amount of work. The Core i7 requires a motherboard that has an LGA 1366 socket and supports a minimum of DDR3-1066 memory chips (chips, not modules). Therefore, you'll be buying at least one stick of your chosen capacity of PC3-8500 (multiplying 1066 by 8), two or three sticks if you decide to take advantage of the chip's ability to access as many as three channels of memory. These days, you'll have better luck starting with a minimum of PC3-10600 modules because PC3—8500s are becoming harder to find. The Core i7 is designed for desktop use; server-level processors that use the Intel QPI include the Xeon and Itanium.

To put each of the SDRAM types into perspective, consult Table 1.3, which summarizes how each technology in the SDRAM arena would achieve a transfer rate of 3200MBps, even if only theoretically.

TABLE 1.3 How some memory types transfer 3200MBps per channel

Memory Type	Actual/Effective (FSB) Clock Frequency (MHz)	Bytes per Transfer
DDR SDRAM PC3200	200/400	8
DDR2 SDRAM PC2-3200	100/400	8
DDR3 SDRAM PC3-3200*	50/400	8

*PC3-3200 does not exist and is too slow for DDR3.

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 CD-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.

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

First of all, it should be noted that each motherboard supports memory based on the speed of the frontside bus (or the CPU's QPI) and the memory's form factor. For example, if the motherboard's FSB is rated at a maximum speed of 1333MHz and you install memory that is rated at 1066MHz, the memory will operate at only 1066MHz, if it works at all, thus making the computer operate slower than it could. In their documentation, most motherboard manufacturers list which type(s) of memory they support as well as its maximum speeds and required pairings.

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 inline package (DIP), single inline memory module (SIMM), and single inline pin package (SIPP). The most popular form factors for primary memory modules today are as follows:

- DIMM (dual inline memory module)
- SODIMM (small outline dual inline memory module)

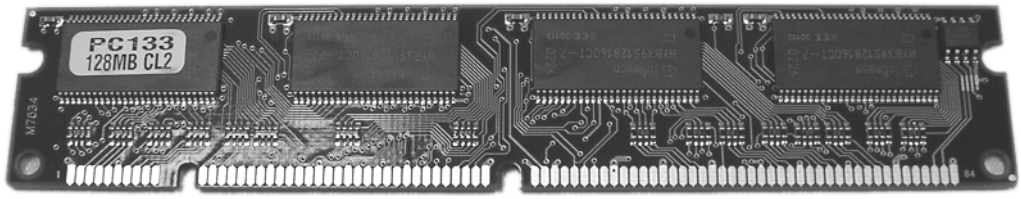
Note also that the various CPUs on the market tend to support only one form of physical memory packaging due to the memory controller in the Northbridge or CPU itself (QPI). For example, the Intel Pentium 4 class of processors was always paired with DIMMs. Laptops and

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 inline memory module. DIMMs are 64-bit memory modules that are used as a package for the SDRAM family: SDR, DDR, DDR2, and DDR3. 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 1.24. SDR SDRAM modules are no longer part of the CompTIA A+ objectives, and they are mentioned here as a foundation only.

FIGURE 1.24 An SDR dual inline memory module (DIMM)



The DIMM used for DDR memory has a total of 184 pins and a single keying notch, while 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.

Figure 1.25 shows a DDR2 module. A matched pair of DDR3 modules with heat spreaders, suitable for dual-channel use in a high-end graphics adapter or motherboard, is shown in Figure 1.26.

FIGURE 1.25 A DDR2 SDRAM module

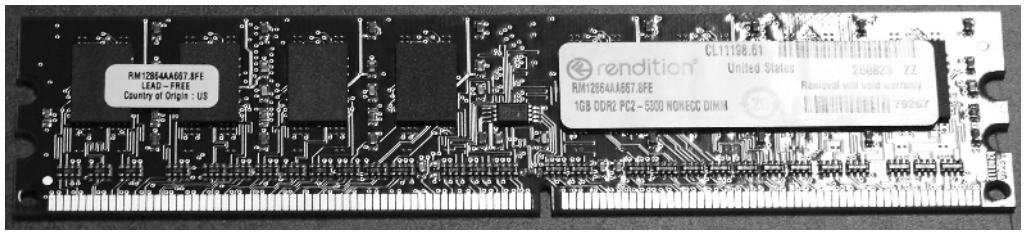
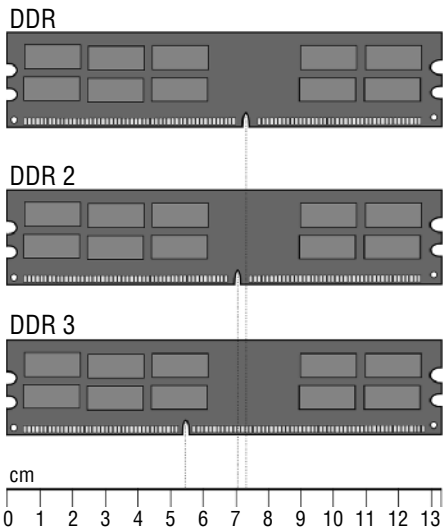


Figure 1.27 shows the subtle differences among various DIMM form factors.

FIGURE 1.26 A pair of DDR3 SDRAM modules

FIGURE 1.27 Aligned DIMM modules



"Desktop DDR Memory Comparison" by Martini - Own work. Licensed under Public Domain via Wikimedia Commons

Inserting and Removing Memory Modules

The original single inline 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

Notebook 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. 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, and 204-pin DDR3) 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. Figure 1.28 shows an example of a 144-pin, 64-bit SDR module. Figure 1.29 is a photo of a 200-pin DDR2 SODIMM.

FIGURE 1.28 144-pin SODIMM

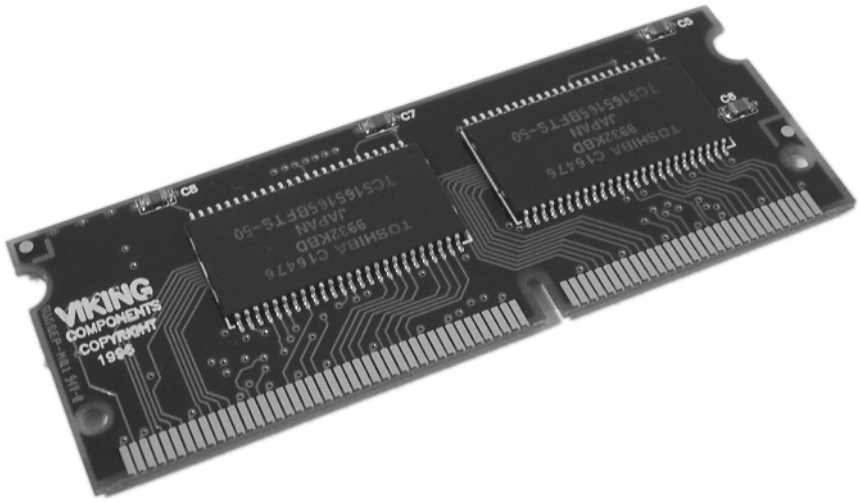


FIGURE 1.29 200-pin DDR2 SODIMM

Identifying Purposes and Characteristics of 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 these 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 three-pin header doesn't offer. A more rare three-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.

Four-pin CPU connections place the ground and power connections in pins 1 and 2, respectively, so that two-pin connectors can be used to power older fans. The four-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 three-pin connector is placed in the correct position on a four-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, "Storage Devices and Power Supplies," 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

continues

continued

before the motherboard connectors were standardized. Figure 1.30 shows two three-pin chassis-fan headers on a motherboard.

FIGURE 1.30 Three-pin chassis-fan headers

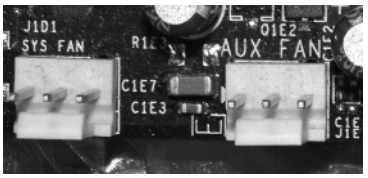
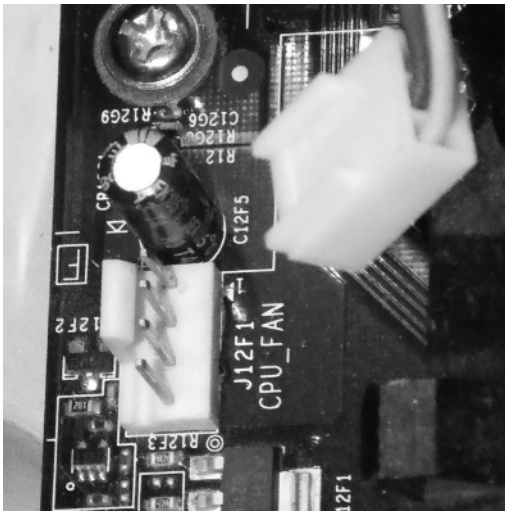


Figure 1.31 shows a four-pin CPU fan header with an approaching three-pin connector from the fan. Note that the keying tab is lined up with the same three pins it's lined up with in the three-pin connectors.

FIGURE 1.31 A four-pin CPU fan header

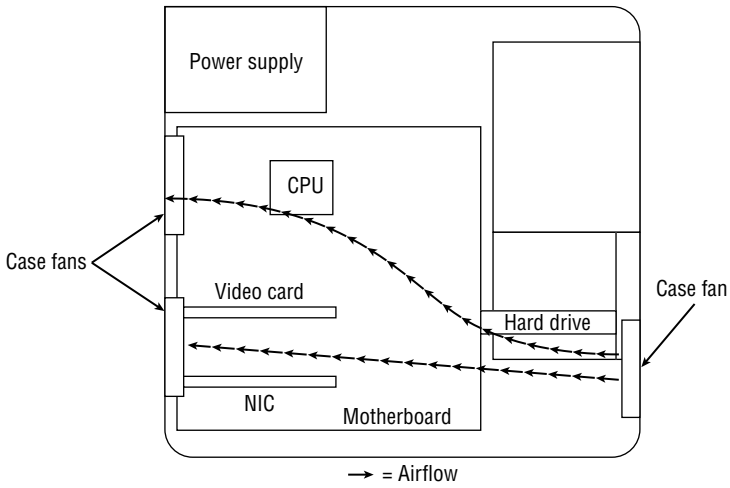


This physical aspect and the similar pin functions are what make these connectors interchangeable, provided the header's function matches the role of the fan being connected. Figure 1.32 shows the resulting unused pin on the four-pin header. Again, controlling the fan's speed is not supported in this configuration.

FIGURE 1.32 Position of a three-pin connector on a four-pin header

Ideally, the airflow inside a computer should resemble what is shown in Figure 1.33, where the back of the chassis is shown on the left in the image.

FIGURE 1.33 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 5'QF" expansion bay. This device contains fans that draw in cool air over the hard drive, thus cooling it. Figure 1.34 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 1.34 An active hard disk cooler

Chipset Cooling

Every motherboard has a chip or chipset that controls how the computer operates. As with 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 an instant. 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 a few different types of CPU cooling methods, but the most important can be grouped into three broad categories: air cooling, liquid cooling, and fanless and passive cooling methods.

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.

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 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.

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 preapplied, 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. There is only one fan needed: the fan on the radiator to cool the water. So a liquid-cooled system can run extremely quietly.

Liquid cooling, while more efficient than air cooling and much quieter, has its drawbacks. Most liquid-cooling systems are more expensive than supplemental fan sets and require less familiar components, such as a reservoir, pump, water block(s), hose, and radiator.

The relative complexity of installing liquid cooling systems, coupled with the perceived danger of liquids in close proximity to electronics, leads most computer owners to consider liquid cooling a novelty or a liability. The primary market for liquid cooling is the high-performance niche that engages in overclocking to some degree. However, developments in active air cooling, including extensive piping of heat away from the body of the heat sink, have kept advanced cooling methods out of the forefront. Nevertheless, advances in fluids with safer electrolytic properties and even viscosities keep liquid cooling viable.

In the next sections, you will notice a spate of seeming liquid-cooling methods. While these methods use liquid in the execution of their cooling duties, liquid to them is analogous to the heat used in cartridges to effect printing in inkjet printers. In other words, the following cooling systems are no more liquid-cooling methods than inkjet printers are thermal printers.

Fanless and Passive CPU Cooling Methods

Advancements in air cooling led to products like the Scythe Ninja series, which is a stack of thin aluminum fins with copper tubing running up through them. Some of the hottest-running CPUs can be passively cooled with a device like this, using only the existing air-movement scheme from your computer's case. Adding a fan to the side, however, adds to the cooling efficiency but also to the noise level, even though Scythe calls this line Ninja because of how quiet it is.

In addition to standard and advanced air-cooling methods, there are other methods of cooling a CPU (and other chips as well). Many of these are *fanless*, in that they do not include a fan but are still considered active because they require power to operate. Others require neither a fan nor power, making them *passive* methods. Some of these methods might appear somewhat unorthodox, but they often deliver extreme results.

Experimental methods can also result in permanent damage to your computer, so try them at your own risk.

Heat Pipes

Heat pipes are closed systems that employ some form of tubing filled with a liquid suitable for the applicable temperature range. Pure physics is used with this technology to achieve cooling to ambient temperatures; no outside mechanism is used. One end of the heat pipe is heated by the component being cooled. This causes the liquid at the heated end to evaporate and increase the relative pressure at that end of the heat pipe with respect to the cooler end. This pressure imbalance causes the heated vapor to equalize the pressure by migrating to the cooler end, where the vapor condenses and releases its heat, warming the nonheated end of the pipe. The cooler environment surrounding this end transfers the heat away from the pipe by convection. The condensed liquid drifts to the pipe's walls and is drawn back to the heated end of the heat pipe by gravity or by a wicking material or texture that lines the inside of the pipe. Once the liquid returns, the process repeats.

Heat pipes are found throughout the computing industry but are particularly beneficial in smaller devices, even as large as laptops. This is because heat pipes work in the absence of cavernous spaces that support airflow. A simple radiator of sorts at the cool end of the pipes, coupled with a simple fan, is enough to keep such devices running cool indefinitely.

Peltier Cooling Devices

Water- and air-cooling devices are extremely effective by themselves, but they are more effective when used with a device known as a *Peltier cooling element*. These devices, also known as thermoelectric coolers (TECs), facilitate the transfer of heat from one side of the element, made of one material, to the other side, made of a different material. Thus they have a hot side and a cold side. The cold side should always be against the CPU surface, and optimally, the hot side should be mated with a heat sink or water block for heat dissipation. Consequently, TECs are not meant to replace air-cooling mechanisms but to complement them.

One of the downsides to TECs is the likelihood of condensation because of the sub-ambient temperatures these devices produce. Closed-cell foams can be used to guard against damage from condensation.

Phase-Change Cooling

With phase-change cooling, the cooling effect from the change of a liquid to a gas is used to cool the inside of a PC. It is a very expensive method of cooling, but it does work. Most often, external air-conditioner-like pumps, coils, and evaporators cool the coolant, which is sent, ice cold, to the heat sink blocks on the processor and chipset. Think of it as a water-cooling system that chills the water below room temperature. Unfortunately, this is easily the noisiest cooling method in this discussion. Its results cannot be ignored, however; it is possible to get CPU temps in the range of –4° F (–20° C). Normal CPU temperatures hover between 104° F and 122° F (40° C and 50° C).

The major drawback to this method is that in higher-humidity conditions, condensation can be a problem. The moisture from the air condenses on the heat sink and can run off onto and under the processor, thus shorting out the electronics. Designers of phase-change cooling systems offer solutions to help ensure that this isn't a problem. Products in the form of foam; silicone adhesive; and greaseless, noncuring adhesives are available to seal the

surface and perimeter of the processor. Additionally, manufacturers sell gaskets and shims that correspond to specific processors, all designed to protect your delicate and expensive components from damage.

Liquid Nitrogen and Helium Cooling

In the interest of completeness, there is a novel approach to super-cooling processors that is ill advised under all but the most extreme circumstances. By filling a vessel placed over the component to be cooled with a liquid form of nitrogen or, for an even more intense effect, helium, temperatures from –100° C to –240° C can be achieved. The results are short lived and only useful in overclocking with a view to setting records. The processor is not likely to survive the incident, due to the internal stress from the extreme temperature changes as well as the stress placed on the microscopic internal joints by the passage of excessive electrons.

Undervolting

Not an attachment, undervolting takes advantage of the property of physics whereby reduction in voltage has an exponential effect on the reduction of power consumption and associated heat production. Undervolting requires a BIOS (where the setting is made) and CPU combination that supports it.

You should monitor the system for unpredictable adverse effects. One of your troubleshooting steps might include returning the CPU voltage to normal and observing the results.

Summary

In this chapter, we took a tour of the system components of a PC. You learned about some of the elements that make up a PC, such as the motherboard, CPU, memory, BIOS/UEFI, and firmware. You'll learn about other PC components in the following chapters. You learned about the various methods used for cooling a PC. You also saw what many of these items look like and how they function.

Exam Essentials

Know the types of system boards. Know the characteristics of and differences between ATX, micro ATX, and ITX motherboards.

Know the components of a motherboard. Be able to describe motherboard components, such as chipsets, expansion slots, memory slots, external cache, CPUs, processor sockets, power connectors, BIOS/UEFI (firmware), and CMOS batteries.

Understand the purposes and characteristics of processors. Be able to discuss the different processor packaging, old and new, and know the meaning of the terms *hyperthreading*, *cores*, *cache*, *speed*, *virtualization support*, and *integrated GPU*.

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.

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 of the circuitry necessary for other components or devices to communicate with one another?
   - **A.** Motherboard
   - **B.** Adapter card
   - **C.** Hard drive
   - **D.** Expansion bus
2. Which packaging is used for DDR SDRAM memory?
   - **A.** 168-pin DIMM
   - **B.** 72-pin SIMM
   - **C.** 184-pin DIMM
   - **D.** 240-pin DIMM
3. What memory chips would you find on a stick of PC3-16000?
   - **A.** DDR-2000
   - **B.** DDR3-2000
   - **C.** DDR3-16000
   - **D.** PC3-2000
4. Which motherboard design style features smaller size and lower power consumption?
   - **A.** ATX
   - **B.** AT
   - **C.** Micro ATX
   - **D.** ITX
5. Which of the following socket types is required for the Intel Core i7-9xx desktop series?
   - **A.** LGA 1366
   - **B.** LGA 1156
   - **C.** LGA 1155
   - **D.** LGA 775
6. Which of the following is a socket technology that is designed to ease insertion of modern CPUs?
   - **A.** Socket 1366
   - **B.** ZIF
   - **C.** LPGA
   - **D.** SPGA

7. Which of the following is *not* controlled by the Northbridge?
   - **A.** PCIe
   - **B.** SATA
   - **C.** AGP
   - **D.** Cache memory

8. Which of the following is used to store data and programs for repeated use? Information can be added and deleted at will, and it does *not* lose its data when power is removed.
   - **A.** Hard drive
   - **B.** RAM
   - **C.** Internal cache memory
   - **D.** ROM

9. Which socket type is required for an AMD Phenom II that uses DDR3 RAM?
   - **A.** AM2
   - **B.** AM2+
   - **C.** AM3
   - **D.** Socket 940

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. 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. To avoid software-based virtualization, which two components need to support hardware-based virtualization?
    - **A.** Memory
    - **B.** Hard drive
    - **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 supply connector
    - **C.** Expansion slot
    - **D.** Jumper

14. Your Core i5 fan has a four-pin connector, but your motherboard only has a single three-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 four-pin connector into the three-pin header.
    - **B.** Buy an adapter.
    - **C.** Leave the plug disconnected and just use the heat sink.
    - **D.** Add an extra chassis fan.

15. What is the combined total speed of a PCIe 2.0 x16 slot?
    - **A.** 500MBps
    - **B.** 16Gbps
    - **C.** 8GBps
    - **D.** 16GBps

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. Which of the following is most helpful when flashing the BIOS on a desktop computer system?
    - **A.** Floppy diskette drive
    - **B.** Uninterruptible power supply
    - **C.** An Internet connection
    - **D.** The Windows administrator password

18. Intel and AMD have integrated which of the following into their Atom and APU processor lines that had not been integrated before?
    - **A.** A GPU
    - **B.** A math coprocessor
    - **C.** The frontside bus
    - **D.** The memory controller

19. You have just purchased a motherboard that has an LGA775 socket for an Intel Pentium 4 processor. What type of memory modules will you most likely need for this motherboard?

A. DIP

B. SIMM

C. RIMM

D. DIMM

20. What type of expansion slot is preferred today for high-performance graphics adapters?

A. AGP

B. PCIe

C. PCI

D. ISA

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 inline 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.