Chapter 1

The Computer's Brain: Processors and Memory

In This Chapter

- Processor performance
- Processor types
- History and evolution of Intel processors
- Intel's competition—AMD, multiprocessor computers
- Physical memory
- Bus architecture and bus types

pro•cess *v*: to complete a series of actions

Every computer consists of a microprocessor and memory. Without the two, the computer would not function. The microprocessor, commonly referred to as the central processing unit (CPU), is the brain of the computer. Like the human brain, the CPU is responsible for managing the timing of each operation and carrying out the instructions or commands from an application or the operating system.

The CPU uses memory as a place to store or retrieve information. Memory comes in several forms, such as random access memory (RAM) and read-only memory (ROM). Memory provides a temporary location for storing information and contains more permanent system configuration information.

Introduction to Processors

The most central component to the computer is the processor. It is responsible for executing the instructions that are given to the computer. The processor determines the operating systems you can use, the software applications you can run on the computer, and the computer's ability and performance. It is also typically one of the major factors in computer cost. Computers that contain newer and more powerful processors are more expensive than computers with less complex processors. This has led processor manufacturers to offer several different lines of processors for the home user, business workstation, and server markets.

Processor Performance

The goal of processor performance is to make applications run faster. Performance is commonly defined by how long it takes for a specific task to be executed. Traditionally, processor performance has been defined as how many instructions can be completed in each clock cycle, or instructions per clock (IPC), times the number of clock cycles. Thus, performance is measured as

IPC × Frequency

Processor Types: A First Look

So many types of computer processors, also referred to as microprocessors, are on the market today that it can be quite confusing to wade through them all. All processors are not created equal, and each processor has its own characteristics that make it unique. For instance, a processor that is built around an architecture common to other processors of the same time period may actually operate at double or triple the speed. Fierce competition among the various chip makers lays the groundwork for new technological innovations and constant improvements.

The most obvious difference among processors is the physical appearance of the chips, meaning that many processors differ noticeably from one another in size and shape. The first processor that Intel released was packaged in a small chip that contained two rows of 20 pins each. As processor technology improved, the shape and packaging scheme of the processor also changed. Modern processors, such as the Intel Core i7 class processors, use the same socket as the Xeon processors and can only be placed on the *motherboard*, which has the appropriate socket. This design also reduces the cost involved in producing the *CPU*.

Another noticeable difference among processors is the type of instruction set they use. The instruction sets that are most common to processors are either *Complex Instruction Set Computing (CISC)* or *Reduced Instruction Set Computing (RISC)*.

CISC has been a common method of processing operations, especially in Intel CPUs. CISC uses a set of commands, which include subcommands that require additional CPU memory and time to process. Each command must go through

motherboard
The main board in a computer that manages communication between devices internally and externally.

central processing unit (CPU)
The microprocessor, or brain, of the computer. It uses logic to perform mathematical operations that are used to manipulate data.

Complex Instruction Set Computing (CISC)
A full complement of instructions used by a processor to complete tasks such as mathematical calculations. Used in the most common type of processor produced; Intel processors are currently based on this standard.

Reduced Instruction Set Computing (RISC)
A reduced set of instructions used by a processor. PowerPC and Alpha processors are manufactured using this standard. The reduced instruction set enables a microprocessor to operate at higher speeds.

a decode unit, located inside the CPU, to be broken down into *microcode*. The microcode is then processed one command at a time, which slows computing.

microcode
The smallest form of an instruction in a CPU.

RISC, on the other hand, uses smaller commands that enable it to operate at higher speeds. The smaller commands work directly with microcode, so there is no need for a decode unit. This factor—along with a RISC chip's capability to execute multiple commands simultaneously—dramatically increases the processing power.

Finally, different manufacturers design processors to varying specifications. You should be sure that the processor type and model you choose are compatible with the operating system that you want to use. If the processor is not 100-percent compatible with the operating system, the computer will not operate at its best or might not work at all.

NOTE

The terms processor, microprocessor, chip, and CPU are used interchangeably.

Deciphering Processor Terminology

For most computer novices, terms such as *microcode efficiency* and *internal cache RAM* can sound like part of a foreign language. To help you keep things straight, here are some common terms and their definitions:

Clock cycles The internal speed of a computer or processor expressed in *megahertz (MHz)* or *gigahertz (GHz)*. The faster the clock speed, the faster the computer performs a specific operation.

megahertz (MHz)
One million cycles per second. The internal clock speed of a microprocessor is expressed in MHz.

gigahertz (GHz)
One billion cycles per second. The internal clock speed of a microprocessor is expressed in GHz.

CPU speed The number of operations that are processed in one second.

Data path The number of *bits* that can be transported into the processor chip during one operation.

bit
A binary digit. The digit is the smallest unit of information and represents either an off state (zero) or an on state (one).

Floating-point unit (FPU), or math coprocessor A secondary processor that speeds operations by taking over math calculations of decimal numbers. Also called a numeric processor.

Level 1 (L1), or internal, cache Memory in the CPU that is used to temporarily store instructions and data while they are waiting to be processed. One of the distinguishing features of different processors is the amount of internal cache that is supported.

Level 2 (L2), or backside, cache Memory that is used by the CPU to temporarily store data that is waiting to be processed. Originally located on the motherboard, CPU architectures such as the Pentium II, III, and 4 have incorporated L2 cache directly on the same board as the CPU. This holds true in today's i5 and i7 processors with L2 and L3 cache on board. The CPU can access the on-board L2 cache two to four times faster than it can access the L2 cache on the motherboard.

Level 3 (L3), or backside, cache Memory that is used by the CPU to temporarily store data that is waiting to be processed and is used in conjunction with the L2 cache. The L3 cache is used to hold memory feeds to the L2 cache, and its memory is typically slower than the L2 memory but faster than main memory.

Microcode efficiency The capability of a CPU to process microcode in a manner that uses the least amount of time and completes the greatest number of operations.

Word size The largest number in bits that can be processed during one operation.

NOTE **All the computer's components, including the processor, are installed on the motherboard. This fiberglass sheet is designed for a specific type of CPU. When purchasing a motherboard, you should check with the motherboard manufacturer to determine which types of CPUs are supported.**

The Intel Processor Lineup

Over time, Intel has introduced several generations of microprocessors. Each processor type is referred to as a generation; each is based on the new technological enhancements of the day. With each product release come new software and hardware products to take advantage of the new technology.

Several generations of Intel processors are available today. Since the arrival of the first Intel chip in the IBM PC, Intel has dominated the market. It seems that every time you turn around, a new chip promises greater performance and processing capabilities than the previous one.

What makes Intel the market leader is its ability to bring the newest innovations in chip technology to the public, usually before its competitors, who are not far behind. Competition is fierce, and each manufacturer attempts to improve on the designs of the others, releasing similar chips that promise better performance.

The following table shows the specifications for some of the newer Intel processors issued to date. You should read the specifications and reviews for each processor to understand its capabilities and reliability.

Model	Clock Speed	Number of Cores	Cache Size
Core i3-530	2.93 GHz	2	4 MB
Core i3-550	3.20 GHz	2	4 MB
Core i3-330M*	2.13 GHz	2	3 MB
Core i3-370M	2.40 GHz	2	3 MB
Core i3-37M	2.40 GHz	2	3 MB
Core i5-750	2.66GHz	4	8 MB

Model	Clock Speed	Number of Cores	Cache Size
Core i5-760	2.8 GHz	4	8 MB
Core i5-430M	2.26 GHz	2	3 MB
Core i5-540M	2.53GHz	2	3 MB
Core i7-975	3.33 GHz	4	8 MB
Core i7-980X	3.33 GHz	6	12 MB
Core i7-920XM	2.00 GHz	4	8 MB
Core i7-940XM	2.13 GHz	4	8 MB

* Mobile Processor

Factors Affecting Performance

Many factors come together to determine the performance of any computer. All other factors being equal, faster components will give better performance, but any computer will be limited by its "weakest links." As an analogy, consider that putting a larger engine in a standard automobile will make it faster, but only if the automobile is going in a straight line. As soon as you try to make the car follow a twisting road, other components such as the drivetrain and the tires can limit the performance of the larger engine.

Within a processor family, faster processors will outperform slower processors. But when we're comparing processors from different families, that rule does not apply. For example, the rating of 400 MHz for a processor from one family does not indicate that it will run significantly faster than a 333 MHz processor from a more advanced processor family.

As you learned earlier, clock cycles and data path are two factors that can influence the performance of your computer. Other factors are

Cache memory Very fast memory that sits between the CPU and the main RAM. Cache memory can be as fast as 5 to 10 nanoseconds, whereas main RAM is usually not faster than 60 to 70 nanoseconds. (Yes, a lower number is better here because it indicates that the memory takes less time to move data.)

Bus speed The rate at which data can be transferred between the CPU and the rest of the motherboard. Typical bus speeds are 1 GHz and higher with the current standard for motherboards entering the market.

NOTE

The type of peripherals on your computer can affect system performance. If your application spends a lot of time accessing your hard disk, selecting a better-performing disk system would improve CPU efficiency. Storage systems are covered in detail in Chapter 2, "Storing Your Files: Data Storage."

History of Intel Chips

Intel released the world's first microprocessor, the Intel 4004, in 1971. It was a 4-bit microprocessor containing a programmable controller chip that could process 45 instructions. The 4 bits meant that the chip had four lines for data to travel on, much like a four-lane highway. Because of its limitations, it was implemented only in a few early video games and some other devices. The following year, Intel released the 8008, an 8-bit microprocessor with enhanced memory storage and the capability to process 48 instructions.

Intel then began to research and develop faster, more capable processors. From that research emerged the 8080, which could process instructions 10 times faster than its predecessors. Although the speed had dramatically improved, it was still limited by the number of instructions it could process. Finally, in 1978, Intel broke many barriers by releasing the first of many computer-ready microprocessors, the 8086. The 8086 was a breakthrough technology with a bus speed of 16 bits and the capability to support and use 1 MB of RAM. Unfortunately, the cost of manufacturing such a chip and compatible 16-bit components made the chip unaffordable. Intel responded the following year with the production of an 8-bit chip, the 8088.

Intel continued to break new ground as the release of each new generation of processor offered improved functions and processing capabilities. The most dramatic improvement was the number of instructions, based on a scale of millions, that the processor could process in one second. This rate, referred to as *millions of instructions per second (MIPS)*, ranges from 0.75 MIPS for the 8088 to over 159,000 MIPS for the Core i7 990X.

The second most dramatic improvement was the speed of the internal clock, measured in gigahertz. All processors are driven by an internal clock mechanism that keeps the rhythm of the chip, much like the rhythm of a heartbeat. The faster the speed of the internal clock, the faster the processor can process instructions. Intel continued to increase the speed of the internal clock from 4.77 MHz for the original 8088 to more than 3.6 GHz for the newest generation of Core i7 Intel microprocessors.

The Pentium Family

Intel released the Pentium chip to take advantage of *Peripheral Component Interconnect (PCI) bus architecture*. This processor consisted of 3.1 million *transistors* and a new 64-bit data path. The chip was originally designed to operate at 66 MHz but was scaled down to 60 MHz to support the new transistor design, which was experiencing heat and power problems. The first chips deployed also suffered from a bug in the microcode that hampered the processor's capability to calculate complex mathematical equations with precision. This problem was immediately fixed through a new batch of chips.

millions of instructions per second (MIPS)
A measurement of the number of microcode instructions that a CPU or microprocessor can complete in one second, or cycle.

bus architecture
Any linear pathway on which electrical signals travel and carry data from a source to a destination.

transistor
A microscopic electronic device that uses positive electrons to create the binary value of one, or "on," and negative electrons to create the binary value of zero, or "off." Modern CPUs have millions of transistors.

Peripheral Component Interconnect (PCI)
A bus standard for the transfer of data between the CPU, expansion cards, and RAM. PCI communicates at 33 MHz.

The most significant development in the Pentium was the use of two parallel 32-bit *pipeline*s that enabled it to execute twice the number of instructions as previous Intel processors—a technological advancement that Intel named superscalar technology. Almost all processors today use this technology.

Released with the Pentium family of processors was *Multimedia Extension (MMX)* technology. MMX technology is often referred to as multimedia-enhanced technology, but this is not completely accurate. MMX-equipped processors contained additional instruction code sets that increased the processing speed for audio, video, and graphical data by up to 60 percent as compared to traditional Pentium processors. The MMX chips dramatically improved the response time of games and multimedia-based applications.

The types of Pentium processors include

- Pentium
- Pentium MMX
- Pentium Pro
- Pentium II (PII)
- Celeron
- Pentium II Xeon
- Pentium III (PIII)
- Pentium III Xeon
- Itanium/Itanium 2
- Pentium 4
- Tualatin
- Core i3
- Core i5
- Core i7

pipeline
A place in the processor where operations occur in a series of stages. The operation is not complete until it has passed through all stages.

Multimedia Extension (MMX)
A processor technology that dramatically improves the response time of games and multimedia-based applications. The technology was introduced through the MMX-equipped line of Intel Pentium chips.

Pentium

The Pentium chip introduced the world to the first parallel 32-bit data path, which enabled the Pentium to process 64 bits—twice as much data as before. The Pentium was the first microprocessor chip designed to work with the PCI bus specification and had internal clock speeds ranging from 60 MHz to 200 MHz.

Pentium MMX

The Pentium with MMX technology included an expanded instruction code set with 57 new MMX microcode instructions. MMX enabled the microprocessor to increase the processing speed of audio, video, and graphics by up to 60 percent.

Pentium Pro

The Pentium Pro was the successor to Intel's Pentium processor. One of the unique features of this microprocessor was its internal RISC architecture with CISC-RISC translator service. The translator service was able to use the CISC set of instructions, common to all Intel chips, convert them to the RISC set, the faster of the two, and then complete the tasks as necessary using RISC.

The architectural enhancement that really distinguished the Pro from the original Pentium would influence how most microprocessors would later be developed. The Pro was two chips in one: On the bottom of the Pentium was the actual processor. Connected directly overhead of the processor was an L2 cache. By placing the L2 cache close to the processor, Intel was able to greatly increase the performance of the Pentium Pro.

Pentium II (PII)

The Intel Pentium II, or PII, processor was essentially an enhanced Pentium Pro processor with MMX extensions, cache memory, and a new interface design. The PII was designed to fit into a *single-edge cartridge (SEC)* that plugs into a 242-pin slot.

single-edge cartridge (SEC)
An advanced packaging scheme that the Intel Pentium II and later models used. The processor was encased in a cartridge module with a single edge that plugged into a 242-pin slot on the system board, much as an expansion card plugged into the system board.

Celeron

The only noticeable difference between the Celeron and regular Pentium II processors is the lack of cache memory within its cartridge. Later models of the Celeron include cache memory on the same chip as the processor.

Pentium II Xeon

One of the major enhancements in the Pentium II Xeon was a larger on-board cache. This processor was available with either 1 MB or 2 MB of L2 cache and a clock rate of 450 MHz.

Pentium III (PIII)

With its faster clock rates (up to 733 MHz), the Pentium III supported demanding applications such as full-screen, full-motion video and realistic graphics. Seventy new instructions had been added to make technologies such as 3D graphics, video, speech, and imaging faster and more affordable for mainstream users. Each Pentium III also contained a unique processor serial number. Intel's intent behind this feature was to enhance system security and asset tracking. However, many individuals object to the serial number as infringement on their privacy because it could be used to identify computers on the Internet.

Pentium III Xeon

The Pentium III Xeon processor challenged RISC-based servers in both price/performance and raw performance. It's available in speeds of up to 550 MHz and supports configurations that have more than one processor in the same box.

Itanium/Itanium 2

The Itanium processor employs a 64-bit architecture and enhanced instruction handling to greatly increase the performance of computational and multimedia operations and supports clock speeds of up to 800 MHz. The Itanium 2 processor uses a 128-bit architecture and supports speeds of 900 MHz and 1 GHz.

Pentium 4

The Pentium 4 introduced architectural changes that allowed the processor to increase performance by processing more instructions per clock cycle. This technology is referred to as Hyper Pipeline and allows for 20 pipeline stages as opposed to the 10 pipeline stages used in the Pentium III family. Other enhancements were added through NetBurst technology, which includes such features as improved L1 and L2 caches and the Rapid Execution Engine. Current Pentium 4 processors can support speeds of up to 2.53 GHz.

Tualatin

The Tualatin processor was originally designed to be a logical next step in the Pentium III family. However, as the schedule for this processor slipped, Intel shifted focus to the Pentium 4 processor family. As a result, the Tualatin processor was not released until mid-2001. The Tualatin processors support speeds of up to 1.2 GHz.

Core i3 Series

The Core i3 is one of Intel's newer baseline processors. There are two categories for the Core i3: the 300 series, which contains all laptop products, and the 500 series, which contains all desktop products. All Core i3 processors have dual cores and all have Intel Hyper-Threading, which allows each CPU core to execute two processing threads. One thing that the Core i3s do not have is Intel's dynamic clock speed technology called Turbo Boost, which can result in more significant performance increases. The Core i3 has an integrated graphics processor.

Core i5 Series

The Core i5 is a mid-range line processor from Intel and has either dual or quad cores. There are four categories for the Core i5. The 400 and 500 series processors are designed for mobile computing and have dual cores. The 600 and 700 are desktop processors: The 600 has dual cores and the 700 has quad cores. All Core i5 processors have Turbo Boost, which is one of the primary differences separating the Core i3 series from the Core i5. The Core i5 700 series is the only one in the Core i5 family that does not have an integrated graphics processor.

Core i7 Series

The Core i7 is the high-end line from Intel and it has dual, quad, and hex core processors. All Core i7 processors have Hyper-Threading and Turbo Boost. The i7 (as well as the i3 and i5) processor supports 32-bit or 64-bit versions of an operating system. The drawback to a 32-bit system is that it supports only up to 3.3 GB of RAM, so you do not get the benefit of any additional RAM installed on the system.

NOTE

Although most of us would like to get our hands on the new high-speed processors, the reality is that it will be a while before they are affordable. Also, to really reap the benefits of those high-speed CPUs, computers need to have equivalently high-powered hardware. That is why you will see the first high-speed high CPUs only in expensive servers.

Stacking Up the Competition

Many manufacturers have atempted to compete with Intel to produce microprocessor chips. For many years, Intel's competitors produced clone copies of its chips, often slightly altering the original design to allow for faster processing speeds. A good example of this was the release by Advanced Micro Devices (AMD) of a 40 MHz version of the 386 processor to rival the 33 MHz version that Intel was producing. Non-Intel, or clone, chips became popular because of their cheaper price and improved features.

In addition to clone chips, other manufacturers produced powerful processors that were not based on Intel architecture. Digital Equipment Corporation (DEC), Sun Microsystems, IBM, and Motorola all produced powerful CPUs. Most of these chips were RISC-based CPUs designed to meet two needs: First, RISC-type chips can meet the powerful speed demands of Unix workstations; second, companies want to differentiate themselves from Intel to increase sales.

When Intel released the Pentium generation of processors, the clone manufacturers adopted their own unique naming conventions that diverged from the path that Intel laid with a new release. At the same time, Intel was experiencing a problem with the early release of its Pentium line—a high-level mathematical division problem. Intel's competitors took advantage of the opportunity by releasing their chips to compete with the Pentium processor.

The subsections that follow provide an overview of these processors:

- AMD
- Cyrix
- PowerPC
- Alpha

WARNING

Be sure to check the Microsoft *Hardware Compatibility List (HCL)* before you attempt to buy non-Intel processors. Many of these processors may be obsolete, or no longer widely used. We are including discussion here for historical purposes.

Advanced Micro Devices (AMD)

In 1996, AMD introduced the K5 to compete with the already-released Intel Pentium processor. The K5 was released in a 64-bit version as a follow-up to the earlier K5x86, which resembled a higher-performance 486-based processor. The performance of the K5 equaled that of the Pentium at a reduced cost to the consumer.

Hardware Compatibility List (HCL) Provided by Microsoft, the HCL lists all hardware that has been tested by Microsoft and has proved to work with a particular operating system. Hardware not on the HCL might work, but it is not certain to.

AMD soon followed the K5 generation with the release of the K6 processor. The K6 offered a boost by accelerating the audio, video, and 3D capabilities of the chip in processing software, and adding MMX technology to compete with the Celeron, Pentium II, and Pentium III.

In addition to competing with the latest PC processors equipped with MMX technology, the AMD K6 offered a bigger "bang for the buck." The AMD K6 processors still plugged into standard motherboards by using current technology with chipset and *Basic Input/Output System (BIOS)* support, without needing special motherboards required by processors such as the Intel Pentium Pro and Pentium II models.

Basic Input/Output System (BIOS) Software located in a ROM chip that is responsible for communicating directly between the computer hardware and the operating system.

The next-generation processors were the Athlon (formerly the K7) and the Duron. Unlike earlier AMD processors, these processors could not be used with standard motherboards. Instead, they required a special Athlon- or Duron-compatible motherboard. One factor that set these chips apart from their Intel and Cyrix counterparts is that they used RISC technology. By using the reduced instruction set, they were able to process instructions at a more rapid rate. This capability and other improved implementations in design enabled these chips to often outperform their Intel counterparts.

NOTE

The Athlon family of processors is geared toward workstations and servers, while the Duron family of processors is geared toward lower-end business and home users.

AMD's latest processors are the Phenom and Phenom II. The original Phenom was released in 2007 and is considered the first true quad core processor. The Phenom II was released in 2008 and supports dual through hex core processors. The Phenom II offers an L3 cache size of 6 MB, which leads to increased performance.

Cyrix

Cyrix introduced a rival to the Intel Pentium processor in 1995. The first generation of its non-clone processor was named the MI 6x86 series. Although early releases of the MI encountered heat-related issues, Cyrix resolved the issues and produced a model that did not suffer from the initial design problem. The improved chip offered lower power consumption requirements that enabled the chip to operate at cooler temperatures.

Although the chip was originally designed to rival the Pentium, it included additional features found in the Intel Pentium Pro. One of the important features of the MI processor was that it could predict the next instruction to process before encountering it, thereby considerably boosting processor performance.

A follow-up to the MI series of processors was the MII series, a direct competitor with the Celeron and Pentium II CPUs. The improved design included additional optimization, enabling instructions to be processed faster than by other processors. The MII processors' improved capabilities were overshadowed by software incompatibilities that made them unable to take full advantage of the improved timing. Cyrix later released software utilities and patches to address the timing issues. The MII featured a set of 57 new instructions that were fully compatible with what was the industry-standard MMX software.

PowerPC

Apple, IBM, and Motorola developed the PowerPC as a new microprocessor technology. The PowerPC microprocessor used RISC technology to produce a high processing rate. The innovative design of the PowerPC chip enabled it to deliver high-performance computing power with lower power consumption than its counterparts.

The IBM RS/6000 Unix-based workstation and Apple Macintosh computers used the PowerPC chip. Its design was much different from the traditional design of the Intel microprocessors. The term *PowerPC* referred to more than just a type of processor; it was also an architecture standard that outlined specifications by which manufacturers could design processors. The resulting designs that followed the specifications offered performance advantages and innovative manufacturing techniques such as those IBM created.

IBM developed one of the most significant changes in processor manufacturing, which it called *Silicon-on-Insulator (SOI)* technology. The PowerPC 750, marketed by Apple Computer as the G3, was the first chip released that used this new manufacturing method. SOI technology provided increased processor performance while it offered low power consumption. Low power consumption is the key to producing products such as handheld devices, which operate for long periods of time powered by a battery. Since the PowerPC 750, Motorola has released the PowerPC MPC 7400, which is more commonly known in Apple circles as the G4.

Silicon-on-Insulator (SOI)
The microchip manufacturing innovation that IBM invented. It is based on the capability to enhance silicon technology by reducing the time it takes to move electricity through a conductor.

Alpha

The Alpha was a high-speed microprocessor that DEC developed. The Alpha processor was typically found in workstations and servers that needed more processing power than that found in Intel-based servers.

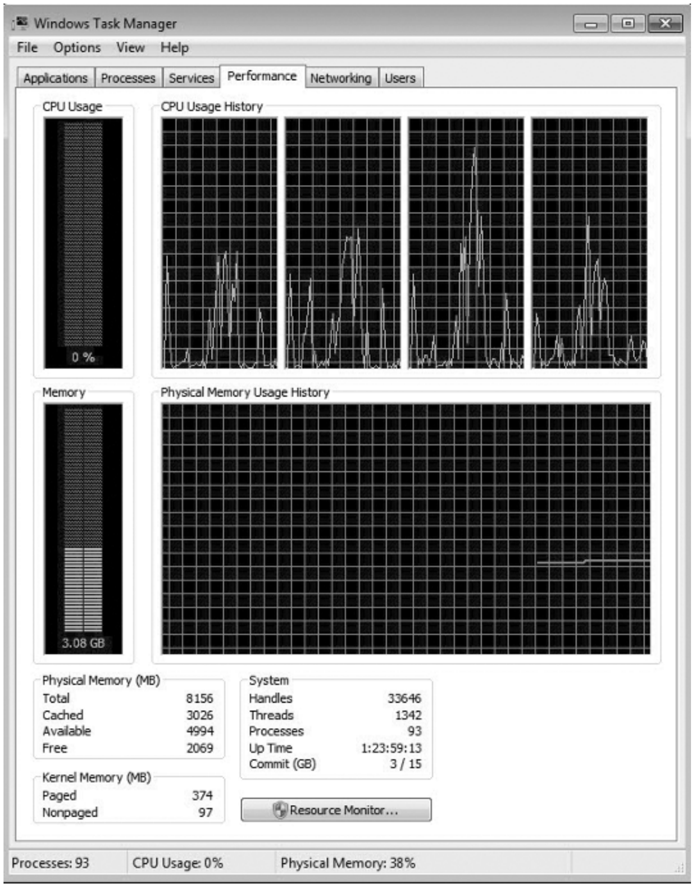
One of the Alpha chip's selling points was that it was the only other chip besides the Intel x86 generations of processors that could run the Microsoft Windows NT operating system. Alpha-equipped workstations were often characterized as the fastest NT workstations on the planet. NT took advantage of the Alpha's capability to produce or generate graphics up to eight times faster than Intel Pentium–based systems. Alpha chips were also commonly found in Unix workstations.

NOTE

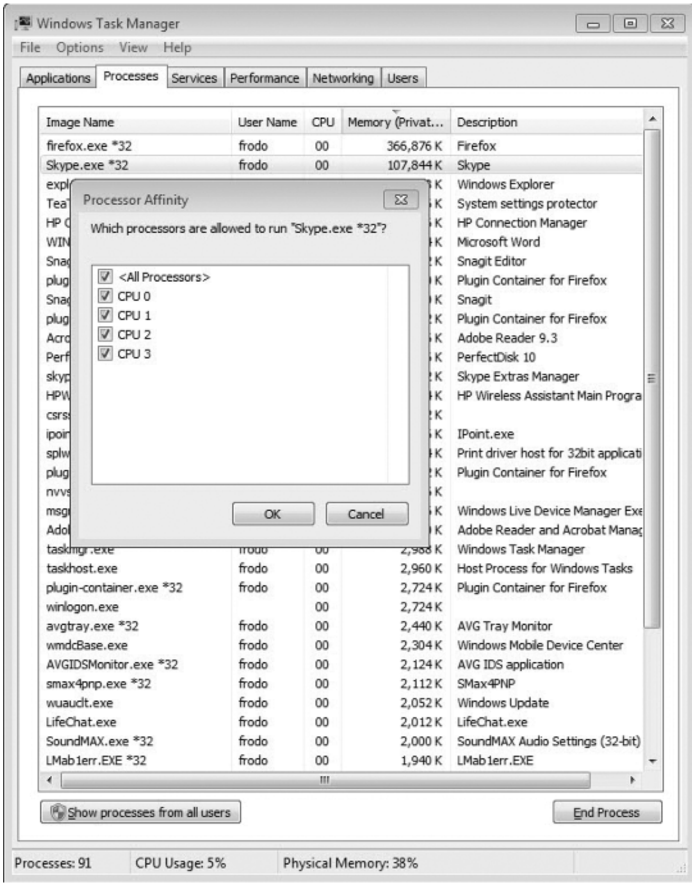
Although Microsoft and DEC did not support Windows 2000, Windows XP, or Windows Server 2003 on the Alpha platform, other operating systems, including HP's version of Unix and Linux, were supported. Other manufacturers, such as Samsung, produced servers and workstations that ran Alpha processors.

Using Multiprocessor and Multicore Computers

Desktop and laptop computers can have multiple processors and multiple cores within a processor. Computers that contain more than one processor and/or core can scale to meet the needs of more demanding application programs. Microsoft server and desktop operating systems can all take advantage of the increased computing power of multiple processors and cores. The following graphics show the difference between Symmetrical Multiprocessing and Asymmetrical Multiprocessing.

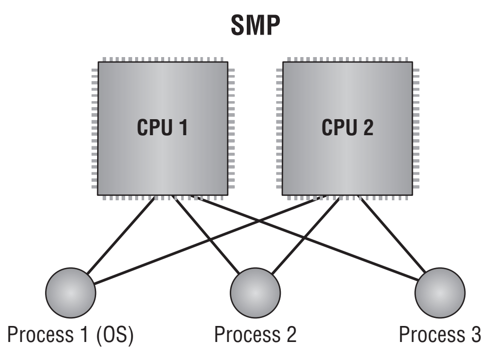


If you have a multi-CPU computer you can implement processor affinity. With processor affinity, you can configure applications to establish a relationship between worker process and one or more CPUs to more efficiently use CPU caches. Processor affinity is used in conjunction with the processor affinity mask setting to specify CPUs.



CPUs in a multiprocessing system may all be treated as equals, or they may be reserved for special purposes. *Symmetric multiprocessing (SMP)* occurs when all processors are treated as equals. In SMP, all tasks are shared equally. The tasks are split among the processors.

symmetric multiprocessing (SMP)
A computer architecture that uses multiple CPUs to improve a computer's performance. As performance demands increase on an SMP-capable computer, additional CPUs can be added to boost performance. During operation, if one CPU is idle, it can be given any task to perform. In the following graphic, you can see how tasks are assigned.



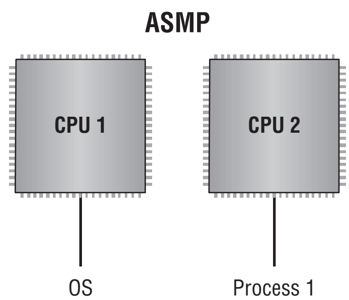
asymmetric multiprocessing (ASMP)
A computer architecture that uses multiple CPUs to improve the performance of the computer. In the ASMP model, one CPU is dedicated to managing tasks (which usually involves managing the other CPUs) for the computer system. The remaining CPUs process user tasks.

non-uniform memory access (NUMA)
A multiprocessing system used by some larger computers where different memory allocations are given to different processors.

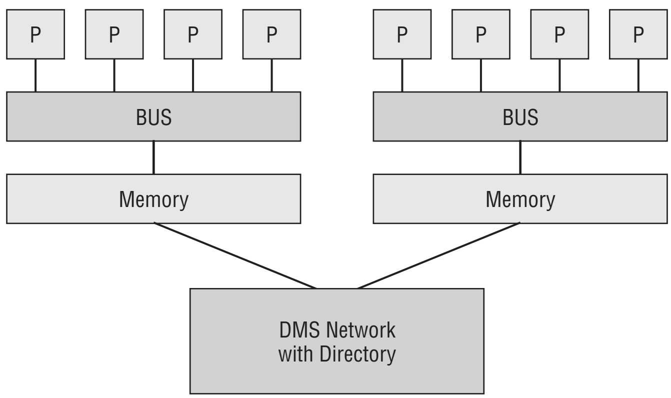
input/output (I/O) devices
Devices that communicate with a computer—such as monitors, printers, keyboards, and mice.

In computers where all CPUs are not equal, system resources may be divided in several ways, including *asymmetric multiprocessing (ASMP)* and *non-uniform memory access (NUMA) multiprocessing.*

In ASMP, one processor is reserved to run the operating system and the *input/output (I/O) devices*. The second ASMP processor runs the application, including the other miscellaneous tasks that the first processor does not handle. This method is often inefficient, because one processor can become busier than the other.



When using NUMA, access time to system memory depends on where the memory is located relative to a specific processor. Also, it is faster to access the system's local memory than it is to access non-local memory. Remote memory (local to another processor) or memory that is shared between processors could present performance issues.



Cooling a System

As with all computers, cooling is paramount. Rarely is a computer too cool in normal circumstances. By ensuring your computer has adequate airflow with intake/exhaust fans, you will assist tremendously. Also, regularly clean the internal case. How dirty your system gets will depend on its location (office, home, or factory) as well as its physical environment (dust, dirt, rodents, snakes, and insects). In some extreme cases, you can use a liquid-cooled system to chill your computer, but if you do, you must always be concerned with leakage and resultant damage to the electronics in the case.

Physical Memory

Memory is an important part of any computer's system. Memory is used in every function of a computer, and it can have a major effect on computer performance. If you are going to get the most out of your computer, you must understand the types of physical memory and how to select the type that is best suited to your computing needs.

NOTE

Memory has always been a consideration with computers and that still holds true with today's computers. In Windows 7 and Windows Server 2008 R2, memory is more important than ever. The minimum requirement for Windows 7 (32-bit) is 1 GB, and 2 GB is recommended. Windows Server 2008 R2 has a minimum requirement of 512 MB.

Memory is basically a series of cells with an address. Each memory cell stores a small piece of information, and each memory cell is identified by a unique address so the processor knows where the cell resides and can easily access it. Computers use several types of memory, each serving a different purpose.

Random Access Memory (RAM)

Random access memory (RAM), often referred to as main memory, is a temporary type of memory that the computer uses as a work area. This type of memory is dynamic (sometimes it is also referred to as volatile memory), meaning that it is constantly changing because of the activity of the processor. When you shut off the power to the computer, RAM loses everything stored in it. RAM stores program instructions and related data for the CPU to quickly access without having to extract data from a slower type of storage device, such as the hard disk or a USB storage device.

random access memory (RAM)
A temporary memory location that stores the operating system, applications, and files that are currently in use. The content of this type of memory is constantly changing. When you shut down the computer, all information in this type of memory is lost.

Hard disks and USB storage devices are more permanent forms of data storage. Programs and their output data are stored on disks or chips for future use. When you shut off the power to the computer, the data on the storage media is intact. However, accessing data and program instructions from storage media can take over a 100 times longer than from RAM.

RAM Types

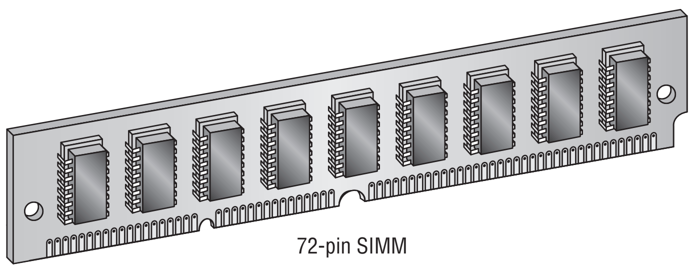
Every computer needs RAM, but which type? Not all types of RAM will work on a computer. Some physically won't fit in the RAM socket, and others will fit but won't work, preventing the computer from passing the *power-on self-test (POST)*.

power-on self-test (POST)
A set of diagnostic tests that are used to determine the state of hardware installed in the computer. Some components that fail the POST, such as bad RAM or a disconnected keyboard, will prevent the computer from booting up properly.

To select the right type of RAM, you need to know your CPU type and motherboard. Some CPUs, such as the Intel i5 and i7, work only with motherboards designed for their specific chips' make. The motherboard is typically designed to meet the highest performance levels of a particular CPU and, therefore, it determines which types of physical RAM can be used. RAM comes in one of two types: Single Inline Memory Modules (SIMMs), an older specification, and Dual Inline Memory Modules (DIMMs), the current standard.

SIMMs

SIMMs are physically different from DIMMs. Older SIMMs were designed with 30 pins that connected to a slot in the motherboard. These modules were slow and typically had to be added in groups of two or four identical SIMMs to be recognized by the BIOS. The current model is a 72-pin SIMM. On motherboards designed for the Pentium processor, the SIMMs must be added in pairs.

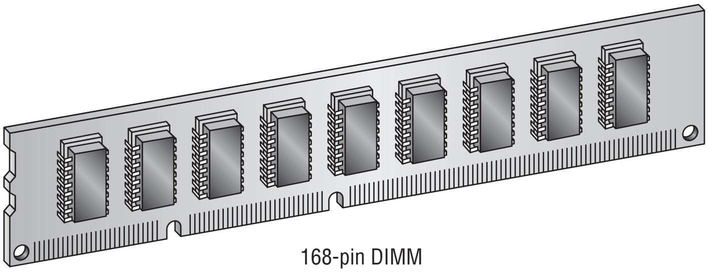


72-pin SIMM

DIMMs

parity
An extra bit found on some memory modules. Non-parity memory has 8 bits. Parity adds an extra bit that is used to keep track of the other 8 bits. This can help prevent memory errors and is recommended for use in servers.

DIMMs have 240 pins with a data path of either 64 bits for non-parity memory or 72 bits for *parity* memory. DIMMs have the largest data path of any memory module. The wider data path makes the chip as fast as the data path on the CPU. This means that the DIMMs can be added one at a time and in varying sizes. It is because of this improved performance and flexibility that DIMMs have become popular in today's personal computers. A variant of the DIMM module is the SO-DIMM (Small Outline DIMM), which is used in laptops.



168-pin DIMM

RAM Speed

Identifying the type of RAM that will physically fit into your computer is only one part of the selection process. Also consider the performance of the RAM you select. Two types of RAM to choose from include Extended Data Out (EDO) and Synchronous Dynamic RAM (SDRAM). Each offers improved performance over older models. Check with your computer manufacturer to see which type of RAM is supported on your computer.

EDO RAM

EDO RAM uses dual-pipeline architecture that enables the unit to store data (write) at the same time it sends it out (reads). EDO RAM is limited to a bus speed of 66 MHz due to its non-parity design. EDO RAM can be purchased in 72-pin SIMMs or 168-pin DIMMs.

SDRAM

SDRAM is similar in design to EDO RAM in that it writes at the same time that it reads, vastly accelerating data along. SDRAM is a popular choice over EDO RAM due to its high bus speed of 100 MHz and its low cost.

DDR SDRAM

This was the first Double Data Rate (DDR) design. The benefit of this design was that the reading and writing of information was done on both cycles, meaning a greater effective data rate (twice the actual speed of the clock and address lines). DDR RAM speeds ranged from 200 to 400 MHz. DDR RAM was used in older computers from the Athlon 64 and Pentium 4 era. DDR RAM capacities typically ranged from 128 MB to 1 GB.

DDR2 SDRAM

DDR2 (DDR type 2) is an improvement on the interface specification with a higher bandwidth interface than DDR, making for improved performance. It is not compatible with DDR because of voltage and timing differences. DDR2 has a lower voltage requirement of 1.8 V compared to 2.5 V in DDR. DDR2 has data rates of 400–1066 MHz. DDR2 RAM capacities typically range from 512 MB to 4 GB.

DDR3 SDRAM

The DDR3 specification has a higher bandwidth interface compared with that of the DDR2, which allows for improved performance. As with DDR, it is not compatible with earlier versions of DDR due to voltage and timing differences. Two main benefits of DDR3 over DDR2 are that it can transfer data at twice the rate (800–2133 MHz) and it has a lower voltage rate of 1.5 V (nearly 30 percent lower than DDR2). DDR3 RAM supports chip sizes of up to 8 GB.

Read-Only Memory (ROM)

read-only memory (ROM)
A type of memory that has data precopied onto it. The data can only be read from and cannot be overwritten. ROM is used to store the BIOS software.

Read-only memory (ROM) is a special type of memory in which data is written onto a chip during manufacturing. Information stored in ROM is permanent and can only be changed in special circumstances. ROM stores the BIOS, the set of instructions a computer uses during the first stages of initialization. Without the BIOS, the computer would not have a mechanism to verify that the main hardware components are installed and functioning properly.

ROM can also hold other, non-volatile instruction sets to include updates for firmware (for computers and computer components), updates for networking devices (Cisco routers and switches), and updates for voice over Internet protocol (VOIP) components (e.g., audio gateways). In many newer systems and devices, ROM chips have been replaced with Flash memory cards. These devices can include various audio/video devices, household appliances, and industrial equipment.

Other ROM Types

Besides the basic ROM chip discussed earlier, other ROM chips are used in computers and small computing devices. The ROM chips described next are programmable, meaning that information can be recorded onto them. These types of chips are important because they enable software that is critical to the computer start-up process to be updated.

PROM

Programmable read-only memory (PROM) is a special type of chip that is manufactured without any configuration. Manufacturers can then *burn in*, or program, the chip to contain whatever configuration is needed.

EPROM

Erasable programmable read-only memory (EPROM) maintains its contents without using electrical power. The stored contents of an EPROM chip are erased and reprogrammed by removing the protective cover and using special equipment to reprogram the chip.

EEPROM

Electrically erasable programmable ROM (EEPROM) typically maintains the BIOS code, which can be updated either with a downloadable file or with a disk that the BIOS manufacturer supplies.

Bus Architecture

When configuring the hardware for a new computer, you have to consider the CPU and motherboard, as discussed earlier in this chapter. In addition, you need to decide which *expansion cards* to install. Expansion cards include sound cards, video adapters, and *network interface cards (NICs)*.

expansion card
An add-on device, such as a sound or video card, that is installed directly into an expansion slot built into a motherboard. The card must be of the same bus architecture as the slot on the motherboard.

network interface card (NIC)
A device that connects a computer to the physical cable media and produces signals for transferring data.

The expansion cards fit into expansion slots that are built into the motherboard. The most common exceptions are special types of computers such as laptop computers. Expansion cards and the slots they fit into can have several different connector types. The connector types are physically different from one another and have varying performance characteristics.

Some reasons why expansion slots are useful are

- The earliest motherboards didn't have room for all the necessary components.
- The expansion slots add flexibility in the event that you need to replace a failed expansion card without having to buy a complete new motherboard.
- Most motherboards have several types of expansion slots. The older type of expansion slots, described in the next section, are available to support older expansion cards, protecting consumers' original investment in their hardware.

In most modern computers, you insert a new add-in, or expansion, card as follows:

WARNING

Before opening any electronic device, make sure that the power is turned off and that you are grounded. To ground yourself, use a special tool called an electrostatic discharge wrist strap. One end of the wrist strap attaches to you, and the

other end clips to the metal case of the computer. Using the strap prevents you from shocking the computer and possibly causing irreversible damage.

1. Turn the power off and disconnect the power cord from the case.
2. Open the case.
3. Make sure the card that you are trying to install is the proper type for an open expansion slot.
4. Insert the card and fasten it down.
5. Close the case.
6. Turn the power back on.

Bus Types

Many types of buses have been introduced since the personal computer was created. Some, such as Industry Standard Architecture (ISA), have had long histories. Others, such as IBM's Micro Channel Architecture (MCA), were never widely adopted for one reason or another. PCI was one of the most widely used (described earlier in the section, "The Pentium Family"), though it is rapidly being phased out in favor of PCIe. Not long ago, PCI was seen as adding better performance to emerging high-speed computers. Now, even the once-sought-after PCI is considered sluggish.

Accelerated Graphics Port (AGP)

megabytes per second (MBps)
A measurement of the transfer speed of a device in terms of millions of bytes per second.

AGP was developed as a replacement for PCI. AGP uses the Intel two-chip 440LX AGP set. This set of chips sits directly on the motherboard and provides similar functionality to PCI. The new chips are responsible for handling the transfer of data between memory, the processor, and the ISA cards all at the same time. Transfer of data to and from PCI cards still occurs at 132 *megabytes per second (MBps)* at 33 MHz. The significant change from PCI is in the speed of transfers to RAM and to the accelerated graphics port. Both have transfer speeds of 528 MBps. This fourfold performance increase provides a significant boost, speeding data along to high-speed CPUs and RAM.

Enhanced Industry Standard Architecture (EISA)

In response to IBM's proprietary MCA bus, the other major hardware vendors (led primarily by Compaq) developed this enhanced bus design.

FireWire (IEEE 1394)

FireWire or IEEE 1394 is a specification for a high-speed serial bus interface standard. FireWire is used for high-speed communications and isochronous real-time data transfer; it is frequently used by personal computers, as well as in digital audio, digital video, automotive, and aeronautics applications. The interface is variously known by the brand names of FireWire (Apple), i.LINK (Sony), and Lynx (Texas Instruments). The IEEE 1394 replaced parallel SCSI in many applications, because of lower implementation costs and a simplified, more adaptable cabling system. FireWire is also available in wireless, fiber-optic, and coaxial versions using the isochronous protocols.

Nearly all digital still and video cameras have a four-circuit IEEE 1394 interface. The Firewire connection is the primary transfer mechanism for high-end professional audio and video equipment. Most computers built since 2003 have built-in FireWire/i.LINK ports.

HDMI

HDMI (High-Definition Multimedia Interface) is a compact audio/video interface for transmitting uncompressed digital data. Typically this is found on devices that are connected to digital audio/video sources such as DVD players/Blu-ray Disc players, camcorders, personal computers and laptops, video game consoles such as the PlayStation 3 and Xbox 360, and AV receivers. HDMI is also used to hook up your laptop or other device to an HDTV, HD monitors, or other projectors.

IBM Micro Channel Architecture (MCA)

IBM's third version of a motherboard expansion bus increased the width of the bus (to 32 bits) and increased the speed. However, unlike with the two original bus designs, IBM didn't freely allow all the other hardware vendors to build cards that were compatible with the MCA specifications and as a result, it was not widely adopted and was eventually discontinued.

IBM PC

The original IBM PC supported 8-bit expansion cards that ran at the same speed as its Intel 8088 processor, 4.77 MHz.

IBM PC-AT, or Industry Standard Architecture (ISA)

The IBM PC-AT introduced two major enhancements: The data path was increased (by the use of a second connector) to 16 bits, and the speed of the expansion cards, usually fixed at 8.33 MHz, was made independent of the processor speed.

Peripheral Component Interconnect (PCI)

The PCI architecture is a 32-bit-wide local bus design that runs at 33 MHz. Due to their local bus design, PCI devices have direct access to the CPU local bus. The PCI local bus is connected to the CPU local bus and system memory bus via a PCI-Host bridge. This is a caching device that provides the interface between the CPU, memory, and PCI local bus. The cache enables the CPU to hand off executions to the PCI bus in order to free up valuable CPU resources. The CPU can continue to fetch information from the caching bridge while the cache controller provides an expansion device with access to system memory.

More than one communication on more than one bus can occur at the same time. This concurrent bus operation could not happen with previous architectures (such as VESA). Additionally, PCI expansion devices are fully independent of the CPU local bus; there is no CPU dependency at all. This design enables the CPU to be upgraded without requiring new designs for devices on the CPU or expansion buses.

Peripheral Component Interconnect Express (PCIe)

PCIe is a new computer expansion card standard that will eventually replace PCI and AGP bus standards. The new PCIe standard offers a faster bus throughput and a smaller physical connector footprint and supports hardware I/O virtualization. The latest version of PCIe is 3.0 with availability of devices announced in June 2011. Work on PCIe 4.0 has started. The latest version of PCIe supports a sustained transfer rate in excess of 500 MBps.

Universal Serial Bus (USB)

The original USB specification was released in 1996, followed by USB 1.1, 2.0, and 3.0 (which was adopted in 2008, with the first device released in January 2010). The USB port is an expected and standard port on computers, networking devices, and elsewhere (some airlines have USB ports available for passengers, and they can be found on some televisions, monitors, keyboards, and elsewhere).

A specific standard is applied to USB ports that is used to establish communication between USB devices and a host controller (usually a personal computer). Though the USB effectively replaced earlier ports, it is in turn being replaced with newer and faster buses, including FireWire, eSATA, PCIe and USB 3. You can use a USB port to connect a wide array of devices including mice, keyboards, digital cameras, printers, and more. In fact, USB devices can be daisy-chained (up to 127 devices), where one port may have several devices connected to it through a USB hub.

One key fact about USB is that, in almost all cases, it is a Plug and Play device. The drivers for the devices are included on the devices (for instance, some mass storage devices) or are on companion media (CD-ROM). The first widely used 1.1 USB devices had a speed of 12 MBps; USB 2.0 followed with an accepted speed of up to 480 MBps. USB 3.0 has speeds of up to 5 GBps (as well as backward compatibility with USB 2.0 and reduced power consumption).

USB ports have a distinct advantage over older port standards such as RS-232 or parallel ports; they can provide power to devices that are connected to them. As a result, USB devices do need a separate external power source (such as a small external hard drive).

Video Electronics Standards Association (VESA) Local Bus (VL-Bus)

The VL-Bus was not a replacement for the other bus types but was instead usually used as an auxiliary bus. The primary devices that supported the VL-Bus were, as you might expect because of its name, video cards. However, some high-performance disk controllers were released that used this standard. Using VL-Bus technology, especially over the long term, had limitations. Major limitations of the VL-Bus included a restriction in the number of VL-Bus devices, a maximum 32-bit data path (preventing expansion to the new Intel Pentium 64-bit systems), and a clock-speed limit of only 33 MHz.

Terms to Know

asymmetric multiprocessing (ASMP)

Basic Input/Output (BIOS)

bit

bus architecture

Complex Instruction Set Computing (CISC)

central processing unit (CPU)

expansion card

gigahertz (GHz)

Hardware Compatibility List (HCL)

input/output (I/O) devices

megabytes per second (MBps)

megahertz (MHz)

microcode

millions of instructions per second (MIPS)

Multimedia Extension (MMX)

motherboard

network interface card (NIC)

non-uniform memory access (NUMA) multiprocessing

parity

Peripheral Component Interconnect (PCI)

pipeline

power-on self-test (POST)

random access memory (RAM)

Reduced Instruction Set Computing (RISC)

read-only memory (ROM)

single-edge cartridge (SEC)

symmetric multiprocessing (SMP)

Silicon-on-Insulator (SOI)

transistor

Review Questions

1. Which processor was released in the first IBM PC?
2. How did the 8086 differ from the 8088?
3. What does *CPU speed* refer to?
4. What is a DIMM and where is it used?
5. How does real mode differ from protected mode?
6. What does *clock cycles* refer to?
7. What is EEPROM?
8. What does PROM stand for, and where is it used?
9. What is a PCIe?
10. How many transistors made up the original Pentium processor?
11. How does asymmetrical multiprocessing differ from symmetrical?
12. What is a math coprocessor?
13. What is the recommended amount of RAM for Windows 7?
14. What is the primary difference between RAM and ROM?
15. What performance gains does PCI have over the EISA bus architecture?