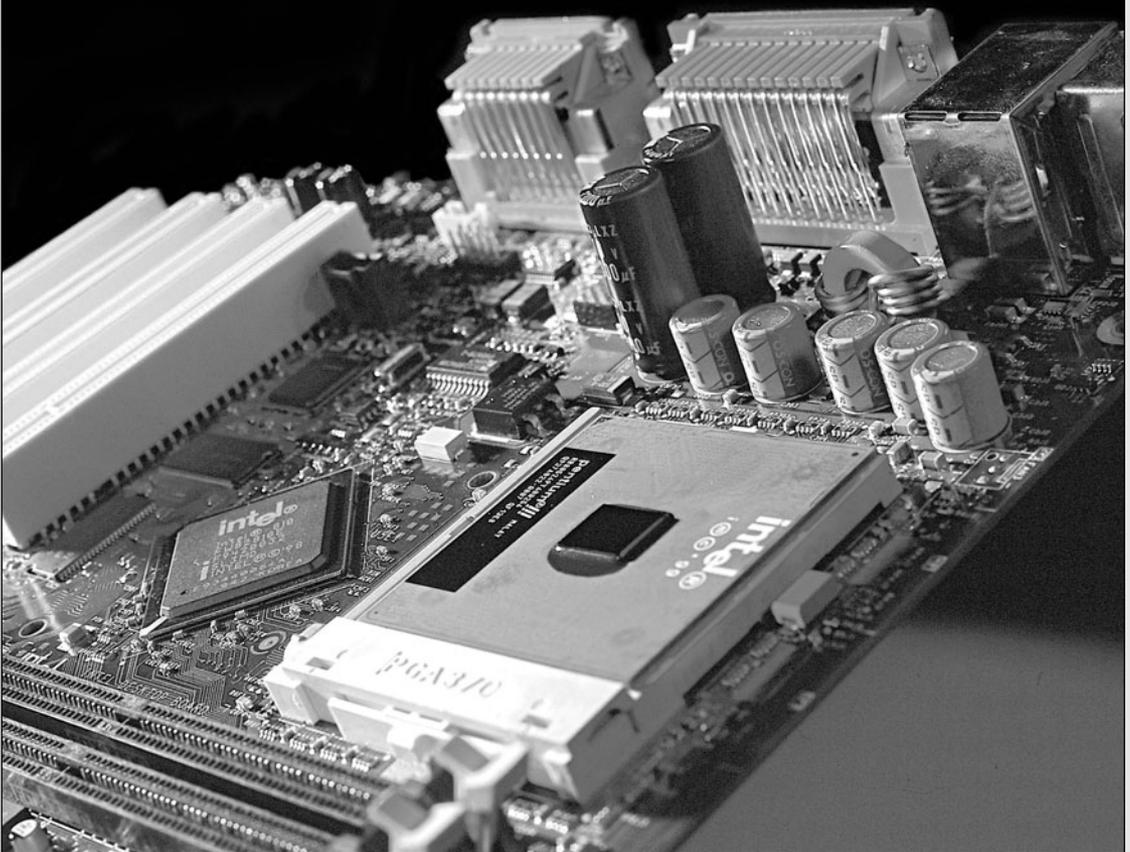


CHAPTER 3

Microprocessor Types and Specifications



Pre-PC Microprocessor History

The brain or engine of the PC is the *processor* (sometimes called *microprocessor*), or *central processing unit (CPU)*. The CPU performs the system's calculating and processing. The processor is often the most expensive single component in the system (although graphics card pricing now surpasses it in some cases); in higher-end systems it can cost up to four or more times more than the motherboard it plugs into. Intel is generally credited with creating the first microprocessor in 1971 with the introduction of a chip called the 4004. Today Intel still has control over the processor market, at least for PC systems, although over the years AMD has garnered a respectable market share. This means that all PC-compatible systems use either Intel processors or Intel-compatible processors from a handful of competitors (such as AMD or VIA/Cyrix).

Intel's dominance in the processor market hadn't always been assured. Although Intel is generally credited with inventing the processor and introducing the first one on the market, by the late 1970s the two most popular processors for personal computers were *not* from Intel (although one was a clone of an Intel processor). Personal computers of that time primarily used the Z-80 by Zilog and the 6502 by MOS Technologies. The Z-80 was noted for being an improved and less expensive clone of the Intel 8080 processor, similar to the way companies such as AMD, VIA/Cyrix, IDT, and Rise Technologies have cloned Intel's Pentium processors. In the Z-80 case, though, the clone had become far more popular than the original. Some might argue that AMD has achieved that type of status over the past year or so, but even though they have made significant gains, Intel still controls the PC processor market.

Back then I had a system containing both of those processors, consisting of a 1MHz (yes, that's 1, as in one megahertz!) 6502-based Apple II system with a Microsoft Softcard (Z-80 card) plugged into one of the slots. The Softcard contained a 2MHz Z-80 processor. This enabled me to run software for both processors on the one system. The Z-80 was used in systems of the late 1970s and early 1980s that ran the CP/M operating system, whereas the 6502 was best known for its use in the early Apple I and II computers (before the Mac).

The fate of both Intel and Microsoft was dramatically changed in 1981 when IBM introduced the IBM PC, which was based on a 4.77MHz Intel 8088 processor running the Microsoft Disk Operating System (MS-DOS) 1.0. Since that fateful decision was made to use an Intel processor in the first PC, subsequent PC-compatible systems have used a series of Intel or Intel-compatible processors, with each new one capable of running the software of the processor before it—from the 8088 to the current Pentium D/4/Celeron and Athlon XP/Athlon 64. The following sections cover the various types of processor chips that have been used in personal computers since the first PC was introduced almost two decades ago. These sections provide a great deal of technical detail about these chips and explain why one type of CPU chip can do more work than another in a given period of time.

Microprocessors from 1971 to the Present

It is interesting to note that the microprocessor had existed for only 10 years prior to the creation of the PC! Intel invented the microprocessor in 1971; the PC was created by IBM in 1981. Now more than 20 years later, we are still using systems based more or less on the design of that first PC. The processors powering our PCs today are still backward compatible in many ways with the 8088 that IBM selected for the first PC in 1981.

November 15, 2001 marked the 30th anniversary of the microprocessor, and in those 30 years processor speed has increased more than 18,500 times (from 0.108MHz to 2GHz). The story of the development of the first microprocessor, the Intel 4004, can be read in Chapter 1, "Development of the PC." The 4004 was introduced on November 15, 1971 and originally ran at a clock speed of 108KHz (108,000 cycles per second, or just over one-tenth a megahertz). The 4004 contained 2,300 transistors and was built on a 10-micron process. This means that each line, trace, or transistor could be spaced about 10 microns (millionths of a meter) apart. Data was transferred 4 bits at a time, and the maximum addressable memory was only 640 bytes. The 4004 was designed for use in a calculator but proved to be useful

for many other functions because of its inherent programmability. For example, the 4004 was used in traffic light controllers, blood analyzers, and even in the NASA Pioneer 10 deep space probe!

In April 1972, Intel released the 8008 processor, which originally ran at a clock speed of 200KHz (0.2MHz). The 8008 processor contained 3,500 transistors and was built on the same 10-micron process as the previous processor. The big change in the 8008 was that it had an 8-bit data bus, which meant it could move data 8 bits at a time—twice as much as the previous chip. It could also address more memory, up to 16KB. This chip was primarily used in dumb terminals and general-purpose calculators.

The next chip in the lineup was the 8080, introduced in April 1974, running at a clock rate of 2MHz. Due mostly to the faster clock rate, the 8080 processor had 10 times the performance of the 8008. The 8080 chip contained 6,000 transistors and was built on a 6-micron process. Similar to the previous chip, the 8080 had an 8-bit data bus, so it could transfer 8 bits of data at a time. The 8080 could address up to 64KB of memory, significantly more than the previous chip.

It was the 8080 that helped start the PC revolution because this was the processor chip used in what is generally regarded as the first personal computer, the Altair 8800. The CP/M operating system was written for the 8080 chip, and Microsoft was founded and delivered its first product: Microsoft BASIC for the Altair. These initial tools provided the foundation for a revolution in software because thousands of programs were written to run on this platform.

In fact, the 8080 became so popular that it was cloned. A company called Zilog formed in late 1975, joined by several ex-Intel 8080 engineers. In July 1976, it released the Z-80 processor, which was a vastly improved version of the 8080. It was not pin compatible but instead combined functions such as the memory interface and RAM refresh circuitry, which enabled cheaper and simpler systems to be designed. The Z-80 also incorporated a superset of 8080 instructions, meaning it could run all 8080 programs. It also included new instructions and new internal registers, so software designed for the Z-80 would not necessarily run on the older 8080. The Z-80 ran initially at 2.5MHz (later versions ran up to 10MHz) and contained 8,500 transistors. The Z-80 could access 64KB of memory.

RadioShack selected the Z-80 for the TRS-80 Model 1, its first PC. The chip also was the first to be used by many pioneering systems, including the Osborne and Kaypro machines. Other companies followed, and soon the Z-80 was the standard processor for systems running the CP/M operating system and the popular software of the day.

Intel released the 8085, its follow-up to the 8080, in March 1976. Even though it predated the Z-80 by several months, it never achieved the popularity of the Z-80 in personal computer systems. It was popular as an embedded controller, finding use in scales and other computerized equipment. The 8085 ran at 5MHz and contained 6,500 transistors. It was built on a 3-micron process and incorporated an 8-bit data bus.

Along different architectural lines, MOS Technologies introduced the 6502 in 1976. This chip was designed by several ex-Motorola engineers who had worked on Motorola's first processor, the 6800. The 6502 was an 8-bit processor like the 8080, but it sold for around \$25, whereas the 8080 cost about \$300 when it was introduced. The price appealed to Steve Wozniak, who placed the chip in his Apple I and Apple II designs. The chip was also used in systems by Commodore and other system manufacturers. The 6502 and its successors were also used in game consoles, including the original Nintendo Entertainment System (NES) among others. Motorola went on to create the 68000 series, which became the basis for the Apple Macintosh line of computers. Today those systems use the PowerPC chip, also by Motorola and a successor to the 68000 series.

All these previous chips set the stage for the first PC processors. Intel introduced the 8086 in June 1978. The 8086 chip brought with it the original x86 instruction set that is still present in current x86-compatible chips such as the Pentium 4 and AMD Athlon. A dramatic improvement over the previous chips, the 8086 was a full 16-bit design with 16-bit internal registers and a 16-bit data bus. This meant that it could work on 16-bit numbers and data internally and also transfer 16 bits at a time in and out of the chip. The 8086 contained 29,000 transistors and initially ran at up to 5MHz.

The chip also used 20-bit addressing, so it could directly address up to 1MB of memory. Although not directly backward compatible with the 8080, the 8086 instructions and language were very similar and enabled older programs to quickly be ported over to run. This later proved important to help jumpstart the PC software revolution with recycled CP/M (8080) software.

Although the 8086 was a great chip, it was expensive at the time and more importantly required expensive 16-bit board designs and infrastructure to support it. To help bring costs down, in 1979 Intel released what some called a *crippled* version of the 8086 called the 8088. The 8088 processor used the same internal core as the 8086, had the same 16-bit registers, and could address the same 1MB of memory, but the external data bus was reduced to 8 bits. This enabled support chips from the older 8-bit 8085 to be used, and far less expensive boards and systems could be made. These reasons are why IBM chose the 8088 instead of the 8086 for the first PC.

This decision would affect history in several ways. The 8088 was fully software compatible with the 8086, so it could run 16-bit software. Also, because the instruction set was very similar to the previous 8085 and 8080, programs written for those older chips could be quickly and easily modified to run. This enabled a large library of programs to be quickly released for the IBM PC, thus helping it become a success. The overwhelming blockbuster success of the IBM PC left in its wake the legacy of requiring backward compatibility with it. To maintain the momentum, Intel has pretty much been forced to maintain backward compatibility with the 8088/8086 in most of the processors it has released since then.

To date, backward compatibility has been maintained, but innovating and adding new features has still been possible. One major change in processors was the move from the 16-bit internal architecture of the 286 and earlier processors to the 32-bit internal architecture of the 386 and later chips, which Intel calls IA-32 (Intel Architecture, 32-bit). Intel's 32-bit architecture dates to 1985, and it took a full 10 years for both a partial 32-bit mainstream OS (Windows 95) as well as a full 32-bit OS requiring 32-bit drivers (Windows NT) to surface, and another 6 years for the mainstream to shift to a fully 32-bit environment for the OS and drivers (Windows XP). That's a total of 16 years from the release of 32-bit computing hardware to the full adoption of 32-bit computing in the mainstream with supporting software. I'm sure you can appreciate that 16 years is a lifetime in technology.

Now we are in the midst of another major architectural jump, as Intel and AMD are in the process of moving from 32-bit to 64-bit computing for servers, desktop PCs, and even portable PCs. Intel had introduced the IA-64 (Intel Architecture, 64-bit) in the form of the Itanium and Itanium 2 processors several years earlier, but this standard was something completely new and not an extension of the existing 32-bit technology. IA-64 was first announced in 1994 as a CPU development project with Intel and HP (codenamed Merced), and the first technical details were made available in October 1997. The result was the IA-64 architecture and Itanium chip, which was officially released in 2001.

The fact that the IA-64 architecture is not an extension of IA-32 but is instead a whole new and completely different architecture is fine for non-PC environments such as servers (for which IA-64 was designed), but the PC market has always hinged on backward compatibility. Even though emulating IA-32 within IA-64 is possible, such emulation and support is slow.

With the door now open, AMD seized this opportunity to develop 64-bit extensions to IA-32, which it calls AMD64 (originally known as x86-64). Intel eventually released its own set of 64-bit extensions, which it calls EM64T or IA-32e mode. As it turns out, the Intel extensions are almost identical to the AMD extensions, meaning they are software compatible. It seems for the first time that Intel has unarguably followed AMD's lead in the development of PC architecture.

To make 64-bit computing a reality, 64-bit operating systems and 64-bit drivers are also needed. Microsoft began providing trial versions of Windows XP Professional x64 Edition (which supports AMD64 and EM64T) in April 2005, and major computer vendors now offer systems with Windows XP Professional x64 already installed. Major hardware vendors have also developed 64-bit drivers for current and recent hardware. Linux is also available in 64-bit-compatible versions, making the move to 64-bit computing possible.

The latest development is the introduction of dual-core processors from both Intel and AMD. Dual-core processors have two full CPU cores operating off of one CPU package—in essence enabling a single processor to perform the work of two processors. Although dual-core processors don't make games (which use single execution threads and are usually not run with other applications) play faster, dual-core processors, like multiple single-core processors, split up the workload caused by running multiple applications at the same time. If you've ever tried to scan for viruses while checking email or running another application, you've probably seen how running multiple applications can bring even the fastest processor to its knees. With dual-core processors available from both Intel and AMD, your ability to get more work done in less time by multitasking is greatly enhanced. Current dual-core processors also support AMD64 or EM64T 64-bit extensions, enabling you to enjoy both dual-core and 64-bit computing's advantages.

PCs have certainly come a long way. The original 8088 processor used in the first PC contained 29,000 transistors and ran at 4.77MHz. The AMD Athlon 64FX has more than 105 million transistors, while the Pentium 4 670 (Prescott core) runs at 3.8GHz and has 169 million transistors thanks to its 2MB L2 cache. Dual-core processors, which include two processor cores and cache memory in a single physical chip, have even higher transistor counts: The Intel Pentium D processor has 230 million transistors, and the AMD Athlon 64 X2 includes over 233 million transistors. As dual-core processors and large L2 caches continue to be used in more and more designs, look for transistor counts and real-world performance to continue to increase. And the progress doesn't stop there because, according to Moore's Law, processing speed and transistor counts are doubling every 1.5–2 years.

Processor Specifications

Many confusing specifications often are quoted in discussions of processors. The following sections discuss some of these specifications, including the data bus, address bus, and speed. The next section includes a table that lists the specifications of virtually all PC processors.

Processors can be identified by two main parameters: how wide they are and how fast they are. The speed of a processor is a fairly simple concept. Speed is counted in megahertz (MHz) and gigahertz (GHz), which means millions and billions, respectively, of cycles per second—and faster is better! The width of a processor is a little more complicated to discuss because three main specifications in a processor are expressed in width. They are

- Data (I/O) bus
- Address bus
- Internal registers

Note that the processor data bus is also called the front side bus (FSB), processor side bus (PSB), or just CPU bus. All these terms refer to the bus that is between the CPU and the main chipset component (North Bridge or Memory Controller Hub). Intel uses the FSB or PSB terminology, whereas AMD uses only FSB. Personally I usually just like to say “CPU bus” in conversation or when speaking during my training seminars because that is the least confusing of the terms while also being completely accurate.

The number of bits a processor is designated can be confusing. All modern processors have 64-bit data buses; however, that does not mean they are classified as 64-bit processors. Processors such as the Pentium 4 and Athlon XP are 32-bit processors because their internal registers are 32 bits wide, although their data I/O buses are 64 bits wide and their address buses are 36 bits wide (both wider than their predecessors, the Pentium and K6 processors). The Itanium series and the AMD Opteron and Athlon 64 are 64-bit processors because their internal registers are 64 bits wide.

First, I'll present some tables describing the differences in specifications between all the PC processors; then the following sections will explain the width and other specifications in more detail. Refer to these tables as you read about the various processor specifications, and the information in the tables will become clearer.

Tables 3.1–3.4 list the Intel processors, AMD processors, and alternative processors from other manufacturers.

Table 3.1 Intel Processor Specifications

Processor	Process (Micron)	Clock Multiplier	Voltage	Internal Register Size	Data Bus Width	Max. Memory
8088	3.0	1x	5V	16-bit	8-bit	1MB
8086	3.0	1x	5V	16-bit	16-bit	1MB
286	1.5	1x	5V	16-bit	16-bit	16MB
386SX	1.5, 1.0	1x	5V	32-bit	16-bit	16MB
386SL	1.0	1x	3.3V	32-bit	16-bit	16MB
386DX	1.5, 1.0	1x	5V	32-bit	32-bit	4GB
486SX	1.0, 0.8	1x	5V	32-bit	32-bit	4GB
486SX2	0.8	2x	5V	32-bit	32-bit	4GB
487SX	1.0	1x	5V	32-bit	32-bit	4GB
486DX	1.0, 0.8	1x	5V	32-bit	32-bit	4GB
486SL ²	0.8	1x	3.3V	32-bit	32-bit	4GB
486DX2	0.8	2x	5V	32-bit	32-bit	4GB
486DX4	0.6	2x+	3.3V	32-bit	32-bit	4GB
486 Pentium OD	0.6	2.5x	5V	32-bit	32-bit	4GB
Pentium 60/66	0.8	1x	5V	32-bit	64-bit	4GB
Pentium 75-200	0.6, 0.35	1.5x+	3.3V–3.5V	32-bit	64-bit	4GB
Pentium MMX	0.35, 0.25	1.5x+	1.8V–2.8V	32-bit	64-bit	4GB
Pentium Pro	0.35	2x+	3.3V	32-bit	64-bit	64GB
Pentium II (Klamath)	0.35	3.5x+	2.8V	32-bit	64-bit	64GB
Pentium II (Deschutes)	0.35	3.5x+	2.0V	32-bit	64-bit	64GB
Pentium II PE (Dixon)	0.25	3.5x+	1.6V	32-bit	64-bit	64GB
Celeron (Covington)	0.25	3.5x+	1.8V–2.8V	32-bit	64-bit	64GB
Celeron A (Mendocino)	0.25	3.5x+	1.5V–2V	32-bit	64-bit	64GB
Celeron III (Coppermine)	0.18	4.5x+	1.5–1.75V	32-bit	64-bit	64GB
Celeron III (Tualatin)	0.13	9x+	1.5V	32-bit	64-bit	64GB
Pentium III (Katmai)	0.25	4x+	2.0–2.05V	32-bit	64-bit	64GB
Pentium III (Coppermine)	0.18	4x+	1.6–1.75V	32-bit	64-bit	64GB
Pentium III (Tualatin)	0.13	8.5x+	1.45V	32-bit	64-bit	64GB
Celeron 4 (Willamette)	0.18	4.25x+	1.6V	32-bit	64-bit	64GB
Pentium 4 (Willamette)	0.18	3x+	1.7V	32-bit	64-bit	64GB
Pentium 4A (Northwood)	0.13	4x+	1.3V	32-bit	64-bit	64GB
Pentium 4EE (Prestonia)	0.13	8x+	1.5V	32-bit	64-bit	64GB
Pentium 4E (Prescott)	0.09	8x+	1.3V	32-bit	64-bit	64GB
Celeron D	0.09	4x+	1.25V–1.4V	32-bit, 64-bit	64-bit	64GB
Pentium D (Smithfield)	0.09	3.5x+	1.25V–1.4V	32-bit, 64-bit	64-bit	64GB
Pentium EE (Glenwood)	0.09	4x	1.25V–1.4V	32-bit, 64-bit	64-bit	64GB
Pentium M (Banias)	0.13	2.25x+	0.8–1.5V	32-bit	64-bit	64GB
Pentium M (Dothan)	0.09	4.25x+	1–1.3V	32-bit	64-bit	64GB

L1 Cache	L2 Cache	L3 Cache	L2/L3 Cache Speed	Multimedia Instructions	No. of Transistors	Date Introduced
—	—	—	—	—	29,000	June '79
—	—	—	—	—	29,000	June '78
—	—	—	—	—	134,000	Feb. '82
—	—	—	Bus	—	275,000	June '88
0KB ¹	—	—	Bus	—	855,000	Oct. '90
—	—	—	Bus	—	275,000	Oct. '85
8KB	—	—	Bus	—	1.185M	Apr. '91
8KB	—	—	Bus	—	1.185M	Apr. '94
8KB	—	—	Bus	—	1.2M	Apr. '91
8KB	—	—	Bus	—	1.2M	Apr. '89
8KB	—	—	Bus	—	1.4M	Nov. '92
8KB	—	—	Bus	—	1.2M	Mar. '92
16KB	—	—	Bus	—	1.6M	Feb. '94
2x16KB	—	—	Bus	—	3.1M	Jan. '95
2x8KB	—	—	Bus	—	3.1M	Mar. '93
2x8KB	—	—	Bus	—	3.3M	Mar. '94
2x16KB	—	—	Bus	MMX	4.5M	Jan. '97
2x8KB	256KB, 512KB, 1MB	—	Core ³	—	5.5M	Nov. '95
2x16KB	512KB	—	1/2 core	MMX	7.5M	May '97
2x16KB	512KB	—	1/2 core	MMX	7.5M	May '97
2x16KB	256KB	—	Core	MMX	27.4M	Jan. '99
2x16KB	0KB	—	—	MMX	7.5M	Apr. '98
2x16KB	128KB	—	Core	MMX	19M	Aug. '98
2x16KB	128KB	—	Core	SSE	28.1M ⁴	Feb. '00
2x16KB	256KB	—	Core	SSE	44M ⁵	Oct. '01
2x16KB	512KB	—	1/2 core	SSE	9.5M	Feb. '99
2x16KB	256KB	—	Core	SSE	28.1M	Oct. '99
2x16KB	512KB	—	Core	SSE	44M	June '01
2x16KB	128KB	—	Core	SSE2	42M ⁶	May '02
12+8KB	256KB	—	Core	SSE2	42M	Nov. '00
12+8KB	512KB	—	Core	SSE2	55M	Jan. '02
12+8KB	512KB	2MB	Core	SSE2	178M	Nov. '03
12+16KB	1MB	—	Core	SSE3	125M	Feb. '04
12+16KB	256KB	—	Core	SSE3	125M	June '04
12+16KB (x2)	1MB (x2)	—	Core	SSE3	250M	Apr. '05
12+16KB (x2)	1MB (x2)	—	Core	SSE3	250M	Apr. '05
2x32KB	1MB	—	Core	SSE2	77M	Mar. '03
2x32KB	2MB	—	Core	SSE2	144M	May '04

Table 3.2 AMD Processor Specifications

Processor	Process (Microns)	CPU Clock	Voltage	Internal Register Size	Data Bus Width	Max. Memory
AMD K5	0.35	1.5x+	3.5V	32-bit	64-bit	4GB
AMD K6	0.35	2.5x+	2.2–3.2V	32-bit	64-bit	4GB
AMD K6-2	0.25	2.5x+	1.9–2.4V	32-bit	64-bit	4GB
AMD K6-3	0.25	3.5x+	1.8–2.4V	32-bit	64-bit	4GB
AMD Athlon	0.25	5x+	1.6–1.8V	32-bit	64-bit	4GB
AMD Duron	0.18	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon (Thunderbird)	0.18	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon XP (Palomino)	0.18	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon XP (Thoroughbred)	0.13	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon XP (Barton)	0.13	5.5x+	1.65V	32-bit	64-bit	4GB
Athlon 64 (ClawHammer/Winchester)	0.13, 0.09	5.5x+	1.5V	64-bit	64-bit	1TB
Athlon 64 FX (SledgeHammer)	0.13	5.5x+	1.5V	64-bit	128-bit	1TB
Athlon 64 X2 (Manchester)	0.09	5x+	1.35V–1.4V	64-bit	128-bit	1TB
Athlon 64 X2 (Toledo)	0.09	5x+	1.35V–1.4V	64-bit	128-bit	1TB

1. The 386SL contains an integral-cache controller, but the cache memory must be provided outside the chip.

2. Intel later marketed SL Enhanced versions of the SX, DX, and DX2 processors. These processors were available in both 5V and 3.3V versions and included power management capabilities.

Table 3.3 Intel/AMD Server/Workstation Processor Specifications

Processor	Process (Micron)	Clock Multiplier	Voltage	Internal Register Size	Data Bus Width	Max. Memory
Pentium II Xeon (Deschutes)	0.25	4x+	2.8V	32-bit	64-bit	64GB
Pentium III Xeon (Tanner)	0.25	5x+	2.0V	32-bit	64-bit	64GB
Pentium III E Xeon (Cascades)	0.18	4.5x+	1.65V	32-bit	64-bit	64GB
Xeon (Foster)	0.18	3.5x+	1.75V	32-bit	64-bit	64GB
Xeon (Prestonia)	0.13	4.5x+	1.5V	32-bit	64-bit	64GB
Itanium (Merced)	0.18	3x+	1.6V	64-bit	64-bit	16TB
Itanium 2 (McKinley)	0.18	3x+	1.6V	64-bit	128-bit	16TB
Itanium 2 (Madison)	0.13	3x+	1.6V	64-bit	128-bit	16TB
AMD Athlon MP (Palomino)	0.18	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon MP (Thoroughbred)	0.13	5x+	1.5–1.8V	32-bit	64-bit	4GB
AMD Athlon MP (Barton)	0.13	5.5x+	1.65V	32-bit	64-bit	4GB
AMD Opteron (SledgeHammer)	0.13	3.5x+	1.55V	64-bit	128-bit	1TB
AMD Opteron dual-core	0.09	3.5x+	1.3V	64-bit	128-bit	1TB

L1 Cache	L2 Cache	L3 Cache	L2/L3 Cache Speed	Multimedia Instructions	No. of Transistors	Date Introduced
16+8KB	—	—	Bus	—	4.3M	March '96
2x32KB	—	—	Bus	MMX	8.8M	April '97
2x32KB	—	—	Bus	3DNow!	9.3M	May '98
2x32KB	256KB	—	Core	3DNow!	21.3M	Feb. '99
2x64KB	512KB	—	1/2–1/3 core	Enh. 3DNow!	22M	June '99
2x64KB	64KB	—	Core ³	Enh. 3DNow!	25M	June '00
2x64KB	256KB	—	Core	Enh. 3DNow!	37M	June '00
2x64KB	256KB	—	Core	3DNow! Pro	37.5M	Oct. '01
2x64KB	256KB	—	Core	3DNow! Pro	37.2M	June '02
2x64KB	512KB	—	Core	3DNow! Pro	54.3M	Feb. '03
2x64KB	1MB	—	Core	3DNow! Pro (SSE3 for 0.09 process)	105.9M	Sept. '03
2x64KB	1MB	—	Core	3DNow! Pro	105.9M	Sept. '03
2x64KB (x2)	1MB	—	Core	SSE3	233.2M	May '05
2x64KB (x2)	2MB	—	Core	SSE3	233.2M	May '05

3. L2 cache runs at full-core speed but is contained in a separate chip die.

4. 128KB functional L2 cache (256KB total, 128KB disabled) uses the same die as the Pentium IIIE.

5. 256KB functional L2 cache (512KB total, 256KB disabled) uses the same die as the Pentium IIIB.

6. 128KB functional L2 cache (256KB total, 128KB disabled) uses the same die as the Pentium 4.

L1 Cache	L2 Cache	L3 Cache	L2/L3 Cache Speed	Multimedia Instructions	No. of Transistors	Date Introduced
2x16KB	512KB, 1MB, 2MB	—	Core ¹	MMX	7.5M	June '98
2x16KB	512KB, 1MB, 2MB	—	Core ¹	SSE	9.5M	Mar. '99
2x16KB	256KB, 1MB, 2MB	—	Core	SSE	28.1M, 84M, 140M	Oct. '99, May '00
12+8KB	256KB	—	Core	SSE2	42M	May '01
12+8KB	512KB	0MB, 1MB, 2MB	Core	SSE2	169M	Jan. '02
2x16KB	96KB ²	2MB, 4MB	Core	MMX	25M	May '01
2x16KB	256KB	1.5MB, 3MB	Core	MMX	221M	July '02
2x16KB	256KB	1.5MB, 6MB	Core	MMX	410M	June '03
2x64KB	256KB	—	Core	3DNow! Pro	37.5M	June '01
2x64KB	256KB	—	Core	3DNow! Pro	37.2M	Aug. '02
2x64KB	512KB	—	Core	3DNow! Pro	54.3M	May. '03
2x64KB	1MB	—	Core	3DNow! Pro	105.9M	Apr. '03
2x64KB	2MB	—	Core	SSE3	233.2M	Apr. '05

Table 3.4 Cyrix, NexGen, IDT, Rise, and VIA Processor Specifications

Processor	CPU Clock	Voltage	Internal Register Size	Data Bus Width	Max. Memory	L1 Cache
Cyrix 6x86	2x+	2.5–3.5V	32-bit	64-bit	4GB	16KB
Cyrix 6x86MX/MII	2x+	2.2–2.9V	32-bit	64-bit	4GB	64KB
Cyrix III	2.5x+	2.2V	32-bit	64-bit	4GB	64KB
NexGen Nx586	2x	4V	32-bit	64-bit	4GB	2x16KB
IDT Winchip	3x+	3.3–3.5V	32-bit	64-bit	4GB	2x32KB
IDT Winchip2/2A	2.33x+	3.3–3.5V	32-bit	64-bit	4GB	2x32KB
Rise mP6	2x+	2.8V	32-bit	64-bit	4GB	2x8KB
VIA C3 ³	6x+	1.6V	32-bit	64-bit	4GB	64KB
VIA C3 ⁴	6x+	1.35V	32-bit	64-bit	4GB	64KB
VIA C3 ⁵	5.5x+	1.35V	32-bit	64-bit	4GB	64KB
VIA C3 ⁶	7.5x+	1.4V	32-bit	64-bit	4GB	64KB

1. L2 cache runs at full-core speed but is contained in a separate chip die.

2. The Itanium also includes an additional 2MB (150M transistors) or 4MB (300M transistors) of integrated on-cartridge L3 cache running at full-core speed.

Data I/O Bus

Perhaps the most important features of a processor are the speed and width of its external data bus. This defines the rate at which data can be moved into or out of the processor.

The processor bus discussed most often is the external data bus—the bundle of wires (or pins) used to send and receive data. The more signals that can be sent at the same time, the more data can be transmitted in a specified interval and, therefore, the faster (and wider) the bus. A wider data bus is like having a highway with more lanes, which enables greater throughput.

Data in a computer is sent as digital information in which certain voltages or voltage transitions occurring within specific time intervals are used to represent data as 1s and 0s. The more wires you have, the more individual bits you can send in the same time interval. All modern processors from the original Pentium through the latest Pentium 4, Athlon XP, Athlon 64, and even the Itanium and Itanium 2 have a 64-bit (8-byte) wide data bus. Therefore, they can transfer 64 bits of data at a time to and from the motherboard chipset or system memory.

A good way to understand this flow of information is to consider a highway and the traffic it carries. If a highway has only one lane for each direction of travel, only one car at a time can move in a certain direction. If you want to increase traffic flow, you can add another lane so that twice as many cars pass in a specified time. You can think of an 8-bit chip as being a single-lane highway because 1 byte flows through at a time. (1 byte equals 8 individual bits.) The 16-bit chip, with 2 bytes flowing at a time, resembles a two-lane highway. You might have four lanes in each direction to move a large number of automobiles; this structure corresponds to a 32-bit data bus, which has the capability to move 4 bytes of information at a time. Taking this further, a 64-bit data bus is like having an 8-lane highway moving data in and out of the chip.

Another ramification of the data bus in a chip is that the width of the data bus also defines the size of a bank of memory. So, a processor with a 32-bit data bus (such as the 486) reads and writes memory 32 bits at a time, whereas processors with a 64-bit data bus (most current processors) read and write memory 64 bits at a time.

In 486 class systems, because standard 72-pin single inline memory modules (SIMMs) are only 32 bits wide, they must be installed one at a time in most 486 class systems. When used in 64-bit Pentium class systems,

L2 Cache	L3 Cache	L2/L3 Cache Speed	Multimedia Instructions	No. of Transistors	Date Introduced
—	—	Bus	—	3M	Feb. '96
—	—	Bus	MMX	6.5M	May '97
256KB	—	Core ¹	3DNow!	22M	Feb. '00
—	—	Bus	—	3.5M	Mar. '94
—	—	Bus	MMX	5.4M	Oct. '97
—	—	Bus	3DNow!	5.9M	Sep. '98
—	—	Bus	MMX	3.6M	Oct. '98
128KB	—	Bus	MMX, 3DNow!	15.2M	Mar. '01
128KB	—	Bus	MMX, 3DNow!	15.4M	Mar. '01
128KB	—	Bus	MMX, 3DNow!	15.5M	Sep. '01
128KB	—	Bus	MMX, 3DNow!	20.5M	Jan. '02

3. *Samuel 2 core (improved version of Cyrix III core).*

4. *Ezra core.*

5. *Ezra-T core.*

6. *Nehemiah core.*

they must be installed two at a time. The current module standard, dual inline memory modules (DIMMs), are 64 bits wide. So, they are normally installed one at a time, unless the system is designed or configured for dual-channel memory. Dual-channel memory reads and writes two banks simultaneously, as a way to improve system performance, which means two DIMMs must be installed at a time. To improve memory performance, most future chipsets will support and eventually require that DIMM memory modules be installed in identical pairs.

The Rambus inline memory modules (RIMMs) used in some older Pentium III and 4 systems are somewhat of an anomaly because they play by a different set of rules. They are typically only 16 or 32 bits wide. Depending on the module type and chipset, they are either used individually or in pairs.

►► See “Memory Banks,” p. 512.

Address Bus

The address bus is the set of wires that carries the addressing information used to describe the memory location to which the data is being sent or from which the data is being retrieved. As with the data bus, each wire in an address bus carries a single bit of information. This single bit is a single digit in the address. The more wires (digits) used in calculating these addresses, the greater the total number of address locations. The size (or width) of the address bus indicates the maximum amount of RAM a chip can address.

The highway analogy in the “Data I/O Bus” section can be used to show how the address bus fits in. If the data bus is the highway and the size of the data bus is equivalent to the number of lanes, the address bus relates to the house number or street address. The size of the address bus is equivalent to the number of digits in the house address number. For example, if you live on a street in which the address is limited to a two-digit (base 10) number, no more than 100 distinct addresses (00–99) can exist for that street (10^2). Add another digit, and the number of available addresses increases to 1,000 (000–999), or 10^3 .

Computers use the binary (base 2) numbering system, so a two-digit number provides only four unique addresses (00, 01, 10, and 11), calculated as 2^2 . A three-digit number provides only eight addresses (000–111), which is 2^3 . For example, the 8086 and 8088 processors use a 20-bit address bus that calculates as a maximum of 2^{20} or 1,048,576 bytes (1MB) of address locations. Table 3.5 describes the memory-addressing capabilities of processors.

Table 3.5 Processor Physical Memory-Addressing Capabilities

Processor Family	Address Bus	Bytes	Kilobytes (KB)	Megabytes (MB)	Gigabytes (GB)	Terabytes (TB)
8088; 8086	20-bit	1,048,576	1,024	1	—	—
286; 386SX	24-bit	16,777,216	16,384	16	—	—
386DX; 486; Pentium	32-bit	4,294,967,296	4,194,304	4,096	4	—
K6; Duron; Athlon; Athlon XP	32-bit	4,294,967,296	4,194,304	4,096	4	—
Celeron; Pentium Pro; Pentium II; Pentium III; Pentium 4	36-bit	68,719,476,736	67,108,864	65,536	64	—
Athlon 64; Athlon 64 FX; Opteron	40-bit	1,099,511,627,776	1,073,741,824	1,048,576	1024	1
Itanium; Itanium 2	44-bit	17,592,186,044,416	17,179,869,184	16,777,216	16,384	16

Note: The terms kilobytes (KB), megabytes (MB), and terabytes (TB) are used here and throughout this chapter by convention; however, technically they should in the future be represented as kibibytes (KiB), gibibytes (GiB), and tebibytes (TiB). See www.iec.ch/zone/si/si_bytes.htm for more information.

The data bus and address bus are independent, and chip designers can use whatever size they want for each. Usually, however, chips with larger data buses have larger address buses. The sizes of the buses can provide important information about a chip's relative power, measured in two important ways. The size of the data bus is an indication of the chip's information-moving capability, and the size of the address bus tells you how much memory the chip can handle.

Internal Registers (Internal Data Bus)

The size of the internal registers indicates how much information the processor can operate on at one time and how it moves data around internally within the chip. This is sometimes also referred to as the *internal data bus*. A *register* is a holding cell within the processor; for example, the processor can add numbers in two different registers, storing the result in a third register. The register size determines the size of data on which the processor can operate. The register size also describes the type of software or commands and instructions a chip can run. That is, processors with 32-bit internal registers can run 32-bit instructions that are processing 32-bit chunks of data, but processors with 16-bit registers can't. Most advanced processors today—chips from the 386 to the Pentium 4—use 32-bit internal registers and can therefore run the same 32-bit operating systems and software. The Itanium and Athlon 64 processors have 64-bit internal registers, which require new operating systems and software to fully be utilized.

Some very old processors have an internal data bus (made up of data paths and storage units called registers) that is larger than the external data bus. The 8088 and 386SX are examples of this structure. Each chip has an internal data bus twice the width of the external bus. These designs, which sometimes are called *hybrid designs*, usually are low-cost versions of a “pure” chip. The 386SX, for example, can pass data around internally with a full 32-bit register size; for communications with the outside world, however, the chip is restricted to a 16-bit-wide data path. This design enabled a systems designer to build a lower-cost motherboard with a 16-bit bus design and still maintain software and instruction set compatibility with the full 32-bit 386. However, both the 8088 and the 386SX had lower performance than the 8086 and 386DX processors at the same speeds.

Internal registers often are larger than the data bus, which means the chip requires two cycles to fill a register before the register can be operated on. For example, both the 386SX and 386DX have internal 32-bit registers, but the 386SX must “inhale” twice (figuratively) to fill them, whereas the 386DX can do the job in one “breath.” The same thing would happen when the data is passed from the registers back out to the system bus.

The Pentium is an example of this type of design. All Pentiums have a 64-bit data bus and 32-bit registers—a structure that might seem to be a problem until you understand that the Pentium has two internal 32-bit pipelines for processing information. In many ways, the Pentium is like two 32-bit chips in one. The 64-bit data bus provides for very efficient filling of these multiple registers. Multiple pipelines are called *superscalar* architecture, which was introduced with the Pentium processor.

►► See “Pentium Processors,” p. 123.

More advanced sixth- and seventh-generation processors from Intel and AMD have as many as six internal pipelines for executing instructions. Although some of these internal pipes are dedicated to special functions, these processors can execute multiple operations in one clock cycle.

Processor Modes

All Intel and Intel-compatible 32-bit processors (from the 386 on up) can run in several modes. Processor modes refer to the various operating environments and affect the instructions and capabilities of the chip. The processor mode controls how the processor sees and manages the system memory and the tasks that use it.

The three main modes of operation with several submodes are as follows:

- Real mode (16-bit software)
- IA-32 mode:
 - Protected mode (32-bit software)
 - Virtual real mode (16-bit programs within a 32-bit environment)
- IA-32e 64-bit extension mode (also called AMD64, x86-64, or EM64T):
 - 64-bit mode (64-bit software)
 - Compatibility mode (32-bit software)

Table 3.6 summarizes the processor modes.

Table 3.6 Processor Modes

Mode	Submode	OS Required	Software	Memory Address Size	Default Operand Size	Register Width
Real	—	16-bit	16-bit	24-bit	16-bit	16-bit
IA-32	Protected	32-bit	32-bit	32-bit	32-bit	32/16-bit
	Virtual real	32-bit	16-bit	24-bit	16-bit	16-bit
IA-32e	64-bit	64-bit	64-bit	64-bit	32-bit	64-bit
	Compatibility	64-bit	32-bit	32-bit	32-bit	32/16-bit

Real Mode

Real mode is sometimes called 8086 mode because it is based on the 8086 and 8088 processors. The original IBM PC included an 8088 processor that could execute 16-bit instructions using 16-bit internal registers and could address only 1MB of memory using 20 address lines. All original PC software was created to work with this chip and was designed around the 16-bit instruction set and 1MB

memory model. For example, DOS and all DOS software, Windows 1.x through 3.x, and all Windows 1.x through 3.x applications are written using 16-bit instructions. These 16-bit operating systems and applications are designed to run on an original 8088 processor.

◀◀ See “Internal Registers (Internal Data Bus),” p. 46.

◀◀ See “Address Bus,” p. 45.

Later processors such as the 286 could also run the same 16-bit instructions as the original 8088, but much faster. In other words, the 286 was fully compatible with the original 8088 and could run all 16-bit software just the same as an 8088, but, of course, that software would run faster. The 16-bit instruction mode of the 8088 and 286 processors has become known as *real mode*. All software running in real mode must use only 16-bit instructions and live within the 20-bit (1MB) memory architecture it supports. Software of this type is usually single-tasking—only one program can run at a time. No built-in protection exists to keep one program from overwriting another program or even the operating system in memory, so if more than one program is running, one of them could bring the entire system to a crashing halt.

IA-32 Mode (32-Bit)

Then came the 386, which was the PC industry’s first 32-bit processor. This chip could run an entirely new 32-bit instruction set. To take full advantage of the 32-bit instruction set, a 32-bit operating system and a 32-bit application were required. This new 32-bit mode was referred to as *protected mode*, which alludes to the fact that software programs running in that mode are protected from overwriting one another in memory. Such protection helps make the system much more crash-proof because an errant program can’t very easily damage other programs or the operating system. In addition, a crashed program can be terminated while the rest of the system continues to run unaffected.

Knowing that new operating systems and applications—which take advantage of the 32-bit protected mode—would take some time to develop, Intel wisely built a backward compatible real mode into the 386. That enabled it to run unmodified 16-bit operating systems and applications. It ran them quite well—much more quickly than any previous chip. For most people, that was enough. They did not necessarily want any new 32-bit software; they just wanted their existing 16-bit software to run more quickly. Unfortunately, that meant the chip was never running in the 32-bit protected mode, and all the features of that capability were being ignored.

When a high-powered processor such as a Pentium 4 is running DOS (real mode), it acts like a “Turbo 8088.” Turbo 8088 means the processor has the advantage of speed in running any 16-bit programs; it otherwise can use only the 16-bit instructions and access memory within the same 1MB memory map of the original 8088. Therefore, if you have a 256MB Pentium 4 or Athlon system running Windows 3.x or DOS, you are effectively using only the first megabyte of memory, leaving the other 255MB largely unused!

New operating systems and applications that ran in the 32-bit protected mode of the modern processors were needed. Being stubborn, we resisted all the initial attempts at getting switched over to a 32-bit environment. It seems that as a user community, we are very resistant to change and would be content with our older software running faster rather than adopting new software with new features. I’ll be the first one to admit that I was one of those stubborn users myself!

Because of this resistance, true 32-bit operating systems such as Unix or variants (such as Linux), OS/2, and even Windows NT/2000 or XP have taken a long time in getting a mainstream share in the PC marketplace. Windows XP is the first full 32-bit OS that has become a true mainstream product, and that is primarily because Microsoft has coerced us in that direction with Windows 95, 98, and Me (which are mixed 16-/32-bit systems). Windows 3.x was the last full 16-bit operating system. In fact, it was not really considered a complete operating system because it ran on top of DOS.

The Itanium processor family, the AMD Opteron, and the Intel EM64T-compatible Xeon processors add 64-bit native capability to the table for servers, whereas the AMD Athlon 64 family, the Intel EM64T-compatible Pentium 4, and all Intel Pentium D processors provide this capability for desktop computers. Both processors run all the existing 32-bit software, but to fully take advantage of the processor, a

64-bit OS, drivers, and applications are required. Microsoft has released 64-bit versions of Windows XP, and several companies have released 64-bit applications for networking and workstation use.

Note

The Intel Itanium family uses the Intel-designed EPIC processor architecture. However, the AMD Athlon 64, the Opteron, and some Semprons use an AMD-developed extension of the x86 architecture Intel originally developed for 386 and higher processors. This architecture is sometimes referred to as IA32e, but AMD refers to this design as AMD64 and Intel uses the term EM64T to refer to its virtually identical 64-bit technology for the Pentium D, Pentium Extreme Edition, and 64-bit versions of the Pentium 4 desktop and Xeon workstation/server processors. Because EPIC and AMD64/EM64T architectures are different, 64-bit software written for one architecture will not work on the other without being recompiled by the software vendor. This means that software written specifically for the Intel EPIC 64-bit architecture will not run on AMD64/EM64T 64-bit processors, and vice versa.

IA-32 Virtual Real Mode

The key to the backward compatibility of the Windows 32-bit environment is the third mode in the processor: virtual real mode. *Virtual real* is essentially a virtual real mode 16-bit environment that runs inside 32-bit protected mode. When you run a DOS prompt window inside Windows, you have created a virtual real mode session. Because protected mode enables true multitasking, you can actually have several real mode sessions running, each with its own software running on a virtual PC. These can all run simultaneously, even while other 32-bit applications are running.

Note that any program running in a virtual real mode window can access up to only 1MB of memory, which that program will believe is the first and only megabyte of memory in the system. In other words, if you run a DOS application in a virtual real window, it will have a 640KB limitation on memory usage. That is because there is only 1MB of total RAM in a 16-bit environment and the upper 384KB is reserved for system use. The virtual real window fully emulates an 8088 environment, so that aside from speed, the software runs as if it were on an original real mode-only PC. Each virtual machine gets its own 1MB address space, an image of the real hardware BIOS routines, and emulation of all other registers and features found in real mode.

Virtual real mode is used when you use a DOS window to run a DOS or Windows 3.x 16-bit program. When you start a DOS application, Windows creates a virtual DOS machine under which it can run.

One interesting thing to note is that all Intel and Intel-compatible (such as AMD and Cyrix) processors power up in real mode. If you load a 32-bit operating system, it automatically switches the processor into 32-bit mode and takes control from there.

It's also important to note that some 16-bit (DOS and Windows 3.x) applications misbehave in a 32-bit environment, which means they do things that even virtual real mode does not support. Diagnostics software is a perfect example of this. Such software does not run properly in a real-mode (virtual real) window under Windows. In that case, you can still run your Pentium 4 in the original no-frills real mode by either booting to a DOS floppy or, in the case of Windows 9x (excluding Me), interrupting the boot process and commanding the system to boot plain DOS. This is accomplished on Windows 9x systems by pressing the F8 key when you see the prompt Starting Windows... on the screen or immediately after the beep when the power on self test (POST) is completed. In the latter case, it helps to press the F8 key multiple times because getting the timing just right is difficult and Windows 9x looks for the key only during a short two-second time window.

If successful, you will then see the Startup menu. You can select one of the command-prompt choices that tell the system to boot plain 16-bit real mode DOS. The choice of Safe Mode Command Prompt is best if you are going to run true hardware diagnostics, which do not normally run in protected mode and should be run with a minimum of drivers and other software loaded.

Even though Windows Me is based on Windows 98, Microsoft removed the DOS Startup menu option in an attempt to further wean us from any 16-bit operation. Windows NT, 2000, and XP also lack the capability to start up DOS in this manner. For these operating systems, you need a startup disk (CD or floppy),

which you can use to boot the system in real mode. Generally, you would do this to perform certain maintenance procedures, such as running hardware diagnostics or doing direct disk sector editing.

Although real mode is used by 16-bit DOS and “standard” DOS applications, special programs are available that “extend” DOS and allow access to extended memory (over 1MB). These are sometimes called *DOS extenders* and usually are included as part of any DOS or Windows 3.x software that uses them. The protocol that describes how to make DOS work in protected mode is called DOS protected mode interface (DPMI).

DPMI was used by Windows 3.x to access extended memory for use with Windows 3.x applications. It allowed these programs to use more memory even though they were still 16-bit programs. DOS extenders are especially popular in DOS games because they enable them to access much more of the system memory than the standard 1MB most real mode programs can address. These DOS extenders work by switching the processor in and out of real mode. In the case of those that run under Windows, they use the DPMI interface built into Windows, enabling them to share a portion of the system’s extended memory.

Another exception in real mode is that the first 64KB of extended memory is actually accessible to the PC in real mode, despite the fact that it’s not supposed to be possible. This is the result of a bug in the original IBM AT with respect to the 21st memory address line, known as A20 (A0 is the first address line). By manipulating the A20 line, real-mode software can gain access to the first 64KB of extended memory—the first 64KB of memory past the first megabyte. This area of memory is called the *high memory area (HMA)*.

IA-32e 64-Bit Extension Mode (AMD64, x86-64, EM64T)

64-bit extension mode is an enhancement to the IA-32 architecture originally designed by AMD and later adopted by Intel. Processors with 64-bit extension technology can run in real (8086) mode, IA-32 mode, or IA-32e mode. IA-32 mode enables the processor to run in protected mode and virtual real mode. IA-32e mode allows the processor to run in 64-bit mode and compatibility mode, which means you can run both 64-bit and 32-bit applications simultaneously. IA-32e mode includes two submodes:

- *64-bit mode*. Enables a 64-bit operating system to run 64-bit applications
- *Compatibility mode*. Enables a 64-bit operating system to run most existing 32-bit software

IA-32e 64-bit mode is enabled by loading a 64-bit operating system and is used by 64-bit applications. In the 64-bit submode, the following new features are available:

- 64-bit linear memory addressing
- Physical memory support beyond 4GB (limited by the specific processor)
- Eight new general-purpose registers (GPRs)
- Eight new registers for streaming SIMD extensions (MMX, SSE, SSE2, and SSE3)
- 64-bit-wide GPRs and instruction pointers

IE-32e compatibility mode enables 32-bit and 16-bit applications to run under a 64-bit operating system. Unfortunately, legacy 16-bit programs that run in virtual real mode (that is, DOS programs) are not supported and will not run, which is likely to be the biggest problem for many users. Similar to 64-bit mode, compatibility mode is enabled by the operating system on an individual code basis, which means 64-bit applications running in 64-bit mode can operate simultaneously with 32-bit applications running in compatibility mode.

What we need to make all this work is a 64-bit operating system and, more importantly, 64-bit drivers for all our hardware to work under that OS. A 64-bit OS already exists in two versions:

- Windows XP 64-bit Edition for Itanium
- Windows XP Professional x64 Edition

Of those, the first is for IA-64 processors, such as Itanium and Itanium 2, and has been available in a released production version since 2001. The latter is for IA-32 processors with 64-bit extensions, such as the Athlon 64, Opteron, some Semprons, the Pentium D, the Pentium Extreme Edition, and some Xeon and Pentium 4 processors supporting 64-bit extensions, and is now available on shipping systems, as an upgrade from Windows XP Professional, or as a 360-day trial version. Note that Microsoft uses the term *x64* to refer to processors that support either AMD64 or EM64T because AMD and Intel's extensions to the standard IA32 architecture are practically identical and can be supported with a single version of Windows.

Note

Early versions of EM64T-equipped processors from Intel lacked support for the LAHF and SAHF instructions used in the AMD64 instruction set. However, Pentium 4 and Xeon DP processors using core steppings G1 and higher completely support these instructions; a BIOS update is also needed.

The differences between Windows XP 32-bit and 64-bit versions are shown in Table 3.7.

Table 3.7 Windows XP 32-Bit Versus 64-Bit

Address Space	Windows XP 32-Bit	Windows XP 64-Bit
Physical memory	4GB	128GB
Virtual memory	4GB	16TB
Paging file	16TB	512TB
Paged pool	470MB	128GB
Non-paged pool	256MB	128GB
System cache	1GB	1TB

The major difference between 32-bit and 64-bit Windows XP is memory support, specifically breaking the 4GB barrier found in 32-bit Windows systems. Windows XP 32-bit supports up to 4GB of physical or virtual memory, with up to 2GB of dedicated memory per process. Windows XP 64-bit Edition supports up to 128GB of physical memory and up to 16TB of virtual memory. Support for more memory means applications can preload more data into either physical or virtual memory, which the processor can access much more quickly. If you need more than 4GB of RAM, 64-bit systems and 64-bit Windows are required.

Windows XP 64-bit runs 32-bit Windows applications with no problems, but it does not run DOS applications or other programs that run in virtual real mode. Also, drivers are another big problem. 32-bit processes cannot load 64-bit dynamic link libraries (DLLs), and 64-bit processes cannot load 32-bit DLLs. This essentially means that, for all the devices you have connected to your system, you need both 32-bit and 64-bit drivers for them to work. Acquiring 64-bit drivers for older devices or devices that are no longer supported can be difficult or impossible. Even for new devices, it can be a couple of years before manufacturers provide 64-bit drivers as a standard feature. Before installing a 64-bit version of Windows, be sure to check with the vendors of your internal and add-on hardware for 64-bit drivers. Keep in mind that drivers made for Itanium do not work with x64-compatible processors.

You should keep all the memory size, software, and driver issues in mind when considering the transition from 32-bit to 64-bit technology. The transition from 32-bit hardware to mainstream 32-bit computing took 16 years. As I've already stated, it might not take 16 years for 64-bit computing to become mainstream, but it will most likely take at least a few years.

Processor Speed Ratings

A common misunderstanding about processors is their different speed ratings. This section covers processor speed in general and then provides more specific information about Intel, AMD, and VIA/Cyrix processors.

A computer system's clock speed is measured as a frequency, usually expressed as a number of cycles per second. A crystal oscillator controls clock speeds using a sliver of quartz sometimes contained in what looks like a small tin container. Newer systems include the oscillator circuitry in the motherboard chipset, so it might not be a visible separate component on newer boards. As voltage is applied to the quartz, it begins to vibrate (oscillate) at a harmonic rate dictated by the shape and size of the crystal (sliver). The oscillations emanate from the crystal in the form of a current that alternates at the harmonic rate of the crystal. This alternating current is the clock signal that forms the time base on which the computer operates. A typical computer system runs millions or billions of these cycles per second, so speed is measured in megahertz or gigahertz. (One hertz is equal to one cycle per second.) An alternating current signal is like a sine wave, with the time between the peaks of each wave defining the frequency (see Figure 3.1).

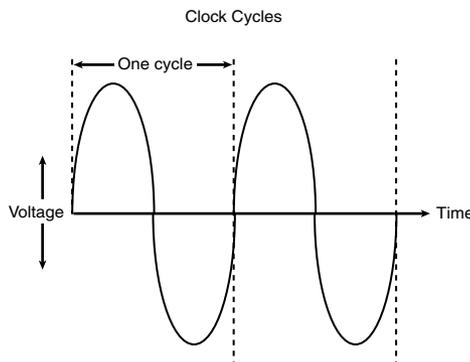


Figure 3.1 Alternating current signal showing clock cycle timing.

Note

The hertz was named for the German physicist Heinrich Rudolf Hertz. In 1885, Hertz confirmed the electromagnetic theory, which states that light is a form of electromagnetic radiation and is propagated as waves.

A single cycle is the smallest element of time for the processor. Every action requires at least one cycle and usually multiple cycles. To transfer data to and from memory, for example, a modern processor such as the Pentium 4 needs a minimum of three cycles to set up the first memory transfer and then only a single cycle per transfer for the next three to six consecutive transfers. The extra cycles on the first transfer typically are called *wait states*. A wait state is a clock tick in which nothing happens. This ensures that the processor isn't getting ahead of the rest of the computer.

►► See "SIMMs, DIMMs, and RIMMs," p. 492.

The time required to execute instructions also varies:

- **8086 and 8088.** The original 8086 and 8088 processors take an average of 12 cycles to execute a single instruction.
- **286 and 386.** The 286 and 386 processors improve this rate to about 4.5 cycles per instruction.
- **486.** The 486 and most other fourth-generation Intel-compatible processors, such as the AMD 5x86, drop the rate further, to about 2 cycles per instruction.
- **Pentium, K6 series.** The Pentium architecture and other fifth-generation Intel-compatible processors, such as those from AMD and Cyrix, include twin instruction pipelines and other improvements that provide for operation at one or two instructions per cycle.
- **Pentium Pro, Pentium II/III/4/D/Extreme Edition/Celeron, and Athlon/Athlon XP/Athlon 64/Athlon 64FX/Duron/Sempron.** These P6 and P7 (sixth- and seventh-generation) processors can execute as many as three or more instructions per cycle.

Different instruction execution times (in cycles) make comparing systems based purely on clock speed or number of cycles per second difficult. How can two processors that run at the same clock rate perform differently with one running “faster” than the other? The answer is simple: efficiency.

The main reason the 486 was considered fast relative to a 386 is that it executes twice as many instructions in the same number of cycles. The same thing is true for a Pentium; it executes about twice as many instructions in a given number of cycles as a 486. Therefore, given the same clock speed, a Pentium is twice as fast as a 486, and consequently a 133MHz 486 class processor (such as the AMD 5x86-133) is not even as fast as a 75MHz Pentium! That is because Pentium megahertz are “worth” about double what 486 megahertz are worth in terms of instructions completed per cycle. The Pentium II and III are about 50% faster than an equivalent Pentium at a given clock speed because they can execute about that many more instructions in the same number of cycles.

Unfortunately, after the Pentium III, it becomes much more difficult to compare processors on clock speed alone. This is because the different internal architectures make some processors more efficient than others, but these same efficiency differences result in circuitry that is capable of running at different maximum speeds. The less efficient the circuit, the higher the clock speed it can attain, and vice versa.

Comparing relative processor performance, you can see that a 1GHz Pentium III is about equal to a (theoretical) 1.5GHz Pentium, which is about equal to a 3GHz 486, which is about equal to a 6GHz 386 or 286, which is about equal to a 12GHz 8088. The original PC's 8088 ran at only 4.77MHz; today, we have systems that are comparatively at least 2,500 times faster! As you can see, you must be careful in comparing systems based on pure MHz alone because many other factors affect system performance.

Evaluating CPU performance can be tricky. CPUs with different internal architectures do things differently and can be relatively faster at certain things and slower at others. To fairly compare various CPUs at different clock speeds, Intel has devised a specific series of benchmarks called the *iCOMP (Intel Comparative Microprocessor Performance) index* that can be run against processors to produce a relative gauge of performance. The iCOMP index benchmark has been updated twice and released in original iCOMP, iCOMP 2.0, and now iCOMP 3.0 versions.

Table 3.8 shows the relative power, or iCOMP 2.0 index, for several processors.

Note

Note that this reflects the most recent iCOMP index. Intel uses other benchmarks for the Pentium 4 and subsequent processors.

Table 3.8 Intel iCOMP 2.0 Index Ratings

Processor	iCOMP 2.0 Index	Processor	iCOMP 2.0 Index
Pentium 75	67	Pentium Pro 200	220
Pentium 100	90	Celeron 300	226
Pentium 120	100	Pentium II 233	267
Pentium 133	111	Celeron 300A	296
Pentium 150	114	Pentium II 266	303
Pentium 166	127	Celeron 333	318
Pentium 200	142	Pentium II 300	332
Pentium-MMX 166	160	Pentium II OverDrive 300	351
Pentium Pro 150	168	Pentium II 333	366
Pentium-MMX 200	182	Pentium II 350	386
Pentium Pro 180	197	Pentium II OverDrive 333	387
Pentium-MMX 233	203	Pentium II 400	440
Celeron 266	213	Pentium II 450	483

The iCOMP 2.0 index is derived from several independent benchmarks and is a stable indication of relative processor performance. The benchmarks balance integer with floating-point and multimedia performance.

When Intel developed the Pentium III, it discontinued the iCOMP 2.0 index and released the iCOMP 3.0 index. iCOMP 3.0 is an updated benchmark that incorporates an increasing use of 3D, multimedia, and Internet technology and software, as well as the increasing use of rich data streams and computer-intensive applications, including 3D, multimedia, and Internet technology. iCOMP 3.0 combines six benchmarks: WinTune 98 Advanced CPU Integer test, CPUmark 99, 3D WinBench 99-3D Lighting and Transformation Test, MultimediaMark 99, Jmark 2.0 Processor Test, and WinBench 99-FPU WinMark. These newer benchmarks take advantage of the SSE (Streaming SIMD Extensions), additional graphics and sound instructions built into the PIII. Without taking advantage of these new instructions, the PIII would benchmark at about the same speed as a PII at the same clock rate.

Table 3.9 shows the iCOMP Index 3.0 ratings for Pentium II and III processors.

Table 3.9 Intel iCOMP 3.0 Ratings

Processor	iCOMP 3.0 Index	Processor	iCOMP 3.0 Index
Pentium II 350	1000	Pentium III 650	2270
Pentium II 450	1240	Pentium III 700	2420
Pentium III 450	1500	Pentium III 750	2540
Pentium III 500	1650	Pentium III 800	2690
Pentium III 550	1780	Pentium III 866	2890
Pentium III 600	1930	Pentium III 1000	3280
Pentium III 600E	2110		

Intel and AMD currently rate their latest processors using the commercially available BAPCo SYSmark 2002 and 2004 benchmark suites. The ratings for the various processors under the 2002 and 2004 benchmark suites are shown in Tables 3.10 and 3.11.

Table 3.10 SYSmark 2002 Scores for Various Processors

CPU	Clock	SYSmark 2002 Rating	CPU	Clock	SYSmark 2002 Rating
Pentium 4 Extreme Edition	3.2GHz	362	AMD Athlon XP	1.72GHz	195
Pentium 4	3.2GHz	344	Pentium 4	1.9GHz	192
Pentium 4	3.0GHz	328	Pentium 4	1.8GHz	187
Pentium 4	3.06GHz	324	Pentium 4	1.7GHz	178
Pentium 4	2.8GHz	312	Pentium 4	1.6GHz	171
Pentium 4	2.6GHz	295	AMD Athlon XP	1.67GHz	171
Pentium 4	2.67GHz	285	Pentium 4	1.5GHz	162
Pentium 4	2.53GHz	273	AMD Athlon XP	1.53GHz	149
Pentium 4	2.4GHz	264	Pentium III	1.2GHz	108
Pentium 4	2.26GHz	252	Pentium III	1.3GHz	104
Pentium 4	2.2GHz	238	Pentium III	1.13GHz	100
Pentium 4	2.0GHz	222	Pentium III	1.0GHz	92

Table 3.11 SYSmrk 2004 Scores for Various Processors

CPU	Clock	SYSmrk 2004 Rating	CPU	Clock	SYSmrk 2004 Rating
Intel Pentium 4EE	3.4GHz	225	AMD Athlon XP 3200+	2.2GHz	163
Intel Pentium 4E	3.4GHz	218	Intel Pentium 4C	2.4GHz	153
Intel Pentium 4EE	3.2GHz	215	AMD Athlon XP 2800+	2.25GHz	151
AMD Athlon FX-53	2.4GHz	213	AMD Athlon XP 2700+	2.18GHz	148
Intel Pentium 4C	3.4GHz	212	Intel Pentium 4B	2.8GHz	144
Intel Pentium 4E	3.2GHz	204	AMD Athlon XP 2600+	2.08GHz	144
AMD Athlon FX-51	2.2GHz	200	AMD Athlon XP 2400+	2.0GHz	133
AMD Athlon 64 3400+	2.2GHz	195	Intel Pentium 4B	2.4GHz	130
AMD Athlon 64 3200+	2.2GHz	194	Intel Celeron	2.8GHz	117
Intel Pentium 4C	3.0GHz	193	Intel Celeron	2.7GHz	115
Intel Pentium 4E	2.8GHz	182	AMD Athlon XP 1800+	1.53GHz	111
AMD Athlon 64 3200+	2.0GHz	180	Intel Celeron	2.5GHz	110
AMD Athlon 64 3000+	2.0GHz	178	Intel Celeron	2.4GHz	104
Intel Pentium 4C	2.8GHz	174	Intel Pentium 4A	2.0GHz	104
AMD Athlon 64 2800+	1.8GHz	164	Intel Pentium III	1.0GHz	64

SYSmrk 2002 and 2004 are commercially available application-based benchmarks that reflect the normal usage of business users employing modern Internet content creation and Microsoft Office applications. However, it is important to note that the scores listed here are produced by complete systems and are affected by things such as the specific version of the processor, the motherboard and chipset used, the amount and type of memory installed, the speed of the hard disk, and other factors. For complete disclosure of the other factors resulting in the given scores, see the BAPCo website.

SYSmrk 2002 incorporates the following applications, which it uses for testing:

- *Internet Content Creation.* Includes Adobe Photoshop 6.01, Premiere 6.0, Microsoft Windows Media Encoder 7.1, Macromedia Dreamweaver 4, and Flash 5
- *Office Productivity.* Includes Microsoft Word 2002, Excel 2002, PowerPoint 2002, Outlook 2002, Access 2002, Netscape Communicator 6.0, Dragon NaturallySpeaking Preferred v.5, WinZip 8.0, and McAfee VirusScan 5.13

SYSmrk 2004 incorporates the following applications, which it uses for testing:

- *Internet Content Creation.* Includes Adobe After Effects 5.5, Adobe Photoshop 7.01, Adobe Premiere 6.5, Discreet 3ds max 5.1, Macromedia Dreamweaver MX, Macromedia Flash MX, Microsoft Windows Media Encoder 9 Series, Network Associates McAfee VirusScan 7.0, and WinZip Computing WinZip 8.1
- *Office Productivity.* Includes Adobe Acrobat 5.0.5, Microsoft Access 2002, Microsoft Excel 2002, Microsoft Internet Explorer 6, Microsoft Outlook 2002, Microsoft PowerPoint 2002, Microsoft Word 2002, Network Associates McAfee VirusScan 7.0, ScanSoft Dragon NaturallySpeaking 6 Preferred, and WinZip Computing WinZip 8.1

The latest version of SYSmrk—SYSmrk 2004 SE—was introduced in June 2005 and now supports the Windows XP Professional x64 Edition. SYSmrk 2004 SE uses the same applications used by SYSmrk 2004, but it incorporates changes to its operation that are designed to more closely mirror how a typical user works with the application mix featured in each module.

SYSMark runs various scripts to do actual work using these applications and is used by many companies for testing and comparing PC systems and components. It is a much more modern and real-world benchmark than the iCOMP benchmark Intel previously used, and because it is available to anybody, the results can be independently verified. SYSMark 2002, 2004, and 2004SE can be purchased from BAPCO at www.bapco.com or from FutureMark at www.futuremark.com.

Processor Speeds and Markings Versus Motherboard Speed

Another confusing factor when comparing processor performance is that virtually all modern processors since the 486DX2 run at some multiple of the motherboard speed. For example, a Pentium 4 2.53GHz chip runs at a multiple of 19/4 (4.75x) times the motherboard speed of 533MHz, whereas an AMD Athlon XP 2800+ using the latest Barton core (2.083GHz) runs at 75/12 (6.25x) times the motherboard speed of 333MHz. Up until early 1998, most motherboards ran at 66MHz or less. Starting in April 1998, Intel released both processors and motherboard chipsets designed to run at 100MHz.

By late 1999, chipsets and motherboards running at 133MHz became available to support the newer Pentium III processors. At that time, AMD Athlon motherboards and chipsets were introduced running a 100MHz clock but using a double transfer technique for an effective 200MHz data rate between the Athlon processor and the main chipset North Bridge chip.

In 2000 and 2001, processor bus speeds advanced further to 266MHz for the AMD Athlon and Intel Itanium and 400MHz to 533MHz for the Pentium 4. In 2002, the AMD Athlon XP processors began to support a processor bus speed of 333MHz. In 2003, Intel introduced the first Pentium 4 processors that supported a processor bus speed of 800MHz; later that year, Intel introduced the Pentium 4 Extreme Edition, which supports a processor bus speed of 1066MHz. Typically, the speed of the CPU bus is selected to match whatever memory types Intel and AMD want to support. Most of the modern CPU bus speeds are based on the speeds of the CPU as well as the available SDRAM, DDR SDRAM, and RDRAM memory. Note that the processor bus speed of Pentium 4, Pentium D, and Pentium Extreme Edition processors is not directly equivalent to a particular memory speed.

Note

See Chapter 4, "Motherboards and Buses," for more information on chipsets and bus speeds.

You can set the motherboard speed and multiplier setting via jumpers or another configuration mechanism (such as BIOS setup) on the motherboard. Modern systems use a variable-frequency synthesizer circuit usually found in the main motherboard chipset to control the motherboard and CPU speed. Most Pentium motherboards have three or four speed settings. The processors used today are available in a variety of versions that run at different frequencies based on a given motherboard speed. For example, most of the Pentium chips run at a speed that is some multiple of the true motherboard speed. For example, Pentium-class processors and motherboards run at the speeds shown in Table 3.12.

Note

For information on specific AMD, Cyrix, or VIA processors, see their respective sections later in this chapter.

Table 3.12 Intel Processor and Motherboard Speeds

CPU Type	CPU Speed (MHz)	CPU Clock Multiplier	Motherboard Speed (MHz)
Pentium	75	1.5x	50
Pentium	60	1x	60
Pentium	90	1.5x	60
Pentium	120	2x	60
Pentium	150	2.5x	60
Pentium/Pentium Pro	180	3x	60
Pentium	66	1x	66
Pentium	100	1.5x	66
Pentium	133	2x	66

Table 3.12 Continued

CPU Type	CPU Speed (MHz)	CPU Clock Multiplier	Motherboard Speed (MHz)
Pentium/Pentium Pro	166	2.5x	66
Pentium/Pentium Pro	200	3x	66
Pentium/Pentium II	233	3.5x	66
Pentium(Mobile)/Pentium II/ Celeron	266	4x	66
Pentium II/Celeron	300	4.5x	66
Pentium II/Celeron	333	5x	66
Pentium II/Celeron	366	5.5x	66
Celeron	400	6x	66
Celeron	433	6.5x	66
Celeron	466	7x	66
Celeron	500	7.5x	66
Celeron	533	8x	66
Celeron	566	8.5x	66
Celeron	600	9x	66
Celeron	633	9.5x	66
Celeron	667	10x	66
Celeron	700	10.5x	66
Celeron	733	11x	66
Celeron	766	11.5x	66
Pentium II	350	3.5x	100
Pentium II	400	4x	100
Pentium II/III	450	4.5x	100
Pentium III	500	5x	100
Pentium III	550	5.5x	100
Pentium III	600	6x	100
Pentium III	650	6.5x	100
Pentium III	700	7x	100
Pentium III	750	7.5x	100
Pentium III/Celeron	800	8x	100
Pentium III/Celeron	850	8.5x	100
Pentium III/Celeron	900	9x	100
Pentium III/Celeron	950	9.5x	100
Pentium III/Celeron	1000	10x	100
Pentium III/Celeron	1100	11x	100
Pentium III/Celeron	1200	12x	100
Pentium III/Celeron	1300	13x	100
Pentium III/Celeron	1400	14x	100
Pentium III	533	4x	133
Pentium III	600	4.5x	133
Pentium III	667	5x	133
Pentium III	733	5.5x	133
Pentium III	800	6x	133
Pentium III	866	6.5x	133
Pentium III	933	7x	133
Pentium III	1000	7.5x	133
Pentium III	1066	8x	133
Pentium III	1133	8.5x	133
Pentium III	1200	9x	133

Table 3.12 Continued

CPU Type	CPU Speed (MHz)	CPU Clock Multiplier	Motherboard Speed (MHz)
Pentium III	1266	9.5x	133
Pentium III	1333	10x	133
Pentium III	1400	10.5x	133
Pentium 4	1300	3.25x	400
Pentium 4	1400	3.5x	400
Pentium 4	1500	3.75x	400
Pentium 4	1600	4x	400
Pentium 4/Celeron	1700	4.25x	400
Pentium 4	1800	4.5x	400
Pentium 4	1900	4.75x	400
Pentium 4	2000	5x	400
Pentium 4	2200	5.5x	400
Pentium 4	2400	6x	400
Pentium 4	2260	4.25x	533
Pentium 4	2400	4.5x	533
Pentium 4	2400	3x	800
Pentium 4	2500	6.25x	400
Pentium 4	2533	4.75x	533
Pentium 4	2600	6.5x	400
Pentium 4	2600	3.25x	800
Pentium 4	2660	5x	533
Pentium 4	2800	5.25x	533
Pentium 4, Pentium D	2800	3.5x	800
Pentium 4, Pentium D	3000	3.75x	800
Pentium 4	3060	5.75x	533
Pentium 4, Pentium D, Pentium EE	3200	4x	800
Pentium 4	3400	4.25x	800
Pentium 4 EE	3460	3.25x	1066
Pentium 4	3600	4.5x	800
Pentium 4 EE	3730	3.5x	1066
Pentium 4	3800	4.75x	800
Itanium	733	2.75x	266
Itanium	800	3x	266
Itanium 2	1000	2.5x	400
Itanium 2	1300	3.25x	400
Itanium 2	1300	2.43x	533
Itanium 2	1600	4x	400
Itanium 2	1600	3x	533
Itanium 2	1666	4.17x	400
Itanium 2	1666	2.5x	667

If all other variables are equal—including the type of processor, the number of wait states (empty cycles) added to different types of memory accesses, and the width of the data bus—you can compare two systems by their respective clock rates. However, the construction and design of the memory controller (contained in the motherboard chipset) as well as the type and amount of memory installed can have an enormous effect on a system's final execution speed.

In building a processor, a manufacturer tests it at various speeds, temperatures, and pressures. After the processor is tested, it receives a stamp indicating the maximum safe speed at which the unit will

operate under the wide variation of temperatures and pressures encountered in normal operation. These ratings are clearly marked on the processor package.

Cyrix Processor Speeds

Cyrix/IBM/VIA 6x86 processors—which were designed to compete with the Intel Pentium, early Pentium II, and AMD K5 and K6 series of processors—used a PR (performance rating) scale that was not equal to the true clock speed in megahertz. For example, the Cyrix 6x86MX/MII-PR366 actually runs at only 250MHz (2.5×100MHz). This is a little misleading—you must set up the motherboard as if a 250MHz processor were being installed, instead of the 366MHz you might suspect. Unfortunately, this led people to believe these systems were faster than they really were.

Table 3.13 shows the relationship between the Cyrix 6x86, 6x86MX, and M-II P-Ratings versus the actual chip speeds in MHz.

Table 3.13 Cyrix P-Ratings Versus Actual Chip Speeds in MHz

CPU Type	P-Rating	Actual CPU Speed (MHz)	Clock Multiplier	Motherboard Speed (MHz)
6x86	PR90	80	2x	40
6x86	PR120	100	2x	50
6x86	PR133	110	2x	55
6x86	PR150	120	2x	60
6x86	PR166	133	2x	66
6x86	PR200	150	2x	75
6x86MX	PR133	100	2x	50
6x86MX	PR133	110	2x	55
6x86MX	PR150	120	2x	60
6x86MX	PR150	125	2.5x	50
6x86MX	PR166	133	2x	66
6x86MX	PR166	137.5	2.5x	55
6x86MX	PR166	150	3x	50
6x86MX	PR166	150	2.5x	60
6x86MX	PR200	150	2x	75
6x86MX	PR200	165	3x	55
6x86MX	PR200	166	2.5x	66
6x86MX	PR200	180	3x	60
6x86MX	PR233	166	2x	83
6x86MX	PR233	187.5	2.5x	75
6x86MX	PR233	200	3x	66
6x86MX	PR266	207.5	2.5x	83
6x86MX	PR266	225	3x	75
6x86MX	PR266	233	3.5x	66
M-II	PR300	225	3x	75
M-II	PR300	233	3.5x	66
M-II	PR333	250	3x	83
M-II	PR366	250	2.5x	100
M-II	PR400	285	3x	95
M-II	PR433	300	3x	100
Cyrix III	PR433	350	3.5x	100
Cyrix III	PR466	366	3x	122
Cyrix III	PR500	400	3x	133
Cyrix III	PR533	433	3.5x	124
Cyrix III	PR533	450	4.5x	100

Note that a given P-Rating can mean several different actual CPU speeds—for example, a Cyrix 6x86MX-PR200 might actually be running at 150MHz, 165MHz, 166MHz, or 180MHz, but *not* at 200MHz.

This P-Rating was supposed to indicate speed in relation to an Intel Pentium processor, but the processor being compared to in this case is the original non-MMX, small L1 cache version running on an older motherboard platform with an older chipset and slower technology memory. The P-Rating did not compare well against the Celeron, Pentium II, or Pentium III processors. In other words, the MII-PR366 really ran at only 250MHz and compared well against Intel processors running at closer to that speed, making the ratings somewhat misleading.

AMD Processor Speeds

AMD's Athlon XP processors were excellent performers and included several notable features, but they also brought with them a resurrection of the infamous Cyrix/AMD performance rating. This is a simulated MHz number that does not indicate the actual speed of the chip but instead indicates an estimate of the relative MHz of a first-generation Intel Pentium 4 that would be approximately equal in performance. If this sounds confusing, that's because it is!

As time marched on and CPU architecture evolved, this method of rating chips had to be revised and eventually abandoned. Although AMD uses model numbers to identify the newer Sempron and Athlon 64 product families, the model numbers for these chips are not specifically intended to compare the processors to Intel processors. As is increasingly the case with both Intel and AMD processors, to gauge processor performance, there's no substitute for knowing the particular features (CPU speed, motherboard speed, L2 cache size, and so on) of a given processor.

Table 3.14 shows the P-Rating (model number) and actual speeds of the AMD K5, K6, Athlon, Athlon XP, Duron, and Sempron (Socket A) processors.

Table 3.14 AMD P-Ratings Versus Actual Chip Speeds in MHz

CPU Type	P-Rating	Actual CPU Speed (MHz)	Clock Multiplier	Motherboard Speed (MHz)
K5	75	75	1.5x	50
K5	90	90	1.5x	60
K5	100	100	1.5x	66
K5	120	90	1.5x	60
K5	133	100	1.5x	66
K5	166	116	1.75x	66
K6	166	166	2.5x	66
K6	200	200	3x	66
K6	233	233	3.5x	66
K6	266	266	4x	66
K6	300	300	4.5x	66
K6-2	233	233	3.5x	66
K6-2	266	266	4x	66
K6-2	300	300	4.5x	66
K6-2	300	300	3x	100
K6-2	333	333	5x	66
K6-2	333	333	3.5x	95
K6-2	350	350	3.5x	100
K6-2	366	366	5.5x	66
K6-2	380	380	4x	95
K6-2	400	400	6x	66
K6-2	400	400	4x	100

Table 3.14 Continued

CPU Type	P-Rating	Actual CPU Speed (MHz)	Clock Multiplier	Motherboard Speed (MHz)
K6-2	450	450	4.5x	100
K6-2	475	475	5x	95
K6-2	500	500	5x	100
K6-2	533	533	5.5x	97
K6-2	550	550	5.5x	100
K6-3	400	400	4x	100
K6-3	450	450	4.5x	100
Athlon	500	500	2.5x	200
Athlon	550	550	2.75x	200
Athlon/Duron	600	600	3x	200
Athlon/Duron	650	650	3.25x	200
Athlon/Duron	700	700	3.5x	200
Athlon/Duron	750	750	3.75x	200
Athlon/Duron	800	800	4x	200
Athlon/Duron	850	850	4.25x	200
Athlon/Duron	900	900	4.5x	200
Athlon/Duron	950	950	4.75x	200
Athlon/Duron	1000	1000	5x	200
Athlon/Duron	1100	1100	5.5x	200
Athlon/Duron	1200	1200	6x	200
Athlon/Duron	1300	1300	6.5x	200
Athlon/Duron	1400	1400	7x	200
Athlon	1000	1000	3.75x	266
Athlon	1133	1133	4.25x	266
Athlon	1200	1200	4.5x	266
Athlon	1333	1333	5x	266
Athlon	1400	1400	5.25x	266
Athlon XP	1500+	1333	5x	266
Athlon XP	1600+	1400	5.25x	266
Athlon XP	1700+	1466	5.5x	266
Athlon XP	1800+	1533	5.75x	266
Athlon XP	1900+	1600	6x	266
Sempron	2000+ ²	1500	4.5x	333
Athlon XP	2000+	1666	6.25x	266
Athlon XP	2100+	1733	6.5x	266
Sempron	2200+ ⁴	1500	4.5x	333
Athlon XP	2200+	1800	6.75x	266
Sempron	2300+ ²	1583	4.75x	333
Sempron	2400+ ²	1667	5x	333
Athlon XP	2400+	2000	7.5x	266
Sempron	2500+ ²	1750	5.25x	333
Athlon XP	2500+ ¹	1833	5.5x	333
Sempron	2600+ ²	1833	5.5x	333
Athlon XP	2600+ ¹	1917	5.75x	333
Athlon XP	2600+	2083	6.25x	333
Athlon XP	2600+	2133	8x	266
Athlon XP	2700+	2167	6.5x	333

Table 3.14 Continued

CPU Type	P-Rating	Actual CPU Speed (MHz)	Clock Multiplier	Motherboard Speed (MHz)
Sempron	2800+ ^{2,4}	2000	6x	333
Athlon XP	2800+ ¹	2083	6.25x	333
Athlon XP	2800+	2250	6.75x	333
Sempron	3000+ ³	2000	6x	333
Athlon XP	3000+ ¹	2167	6.5x	333
Athlon XP	3200+ ¹	2200	5.5x	400
Athlon XP	3200+ ¹	2333	7x	333

Note that the 200MHz and 266MHz bus speeds in the Athlon/Duron use 100MHz and 133MHz clock signals and two transfers per cycle for double the effective rate. Some motherboards refer to the CPU bus speed by the half-speed 100MHz or 133MHz clock and therefore use twice the clock multiplier settings shown here.

1. These processors use the Barton core, which uses a larger 512KB cache instead of other Athlon XP models' 256KB cache to improve performance.
2. Sempron Model 8 (256KB L2 cache).
3. Sempron Model 10 (512KB L2 cache).
4. Sempron Model 10 with 256KB L2 cache.

The marketing problem that led to the need for performance ratings and model numbers for processors is real: How do you market a chip that performs faster than its predecessors or its rivals when both are running at the same clock speed or lower? For example, an AMD Athlon XP with an actual clock speed of 2GHz is significantly faster than a 2GHz Pentium 4 (Northwood) and in fact performs about equal to a 2.4GHz Pentium 4 (hence, AMD called its model the Athlon XP 2400+). This apparent disparity in performance is because the P4 uses a different architecture that utilizes a deeper instruction pipeline with more stages. The original version of the Pentium 4 had a 20-stage pipeline, which compared to a 10-stage pipeline in the Athlon and a 10-stage pipeline in the Pentium III/Celeron (see Table 3.15).

Table 3.15 Number of Pipelines per CPU

Processor	Pipeline Depth	Processor	Pipeline Depth
Pentium III	10-stage	Athlon 64/64 FX	12-stage
Pentium M	10-stage	Pentium 4	20-stage
Athlon/XP	10-stage	Pentium 4 Prescott	31-stage

A deeper pipeline effectively breaks instructions down into smaller microsteps, which allows overall higher clock rates to be achieved using the same silicon technology. However, it also means that overall fewer instructions can be executed in a single cycle as compared to processors with shorter pipelines. This is because, if a branch prediction or speculative execution step fails (which happens fairly frequently inside the processor as it attempts to line up instructions in advance), the entire pipeline has to be flushed and refilled. Thus, if you compared an Athlon to a Pentium III to a Pentium 4 all running at the same clock speed, the Athlon and Pentium III would both beat the Pentium 4 running typical benchmarks because they would execute more instructions in the same number of cycles.

Although this would sound bad for the Pentium 4, it really isn't. Intel's reasoning was sound: Even though the deeper pipeline might be 30% less efficient overall, it more than makes up for this by allowing at least 50% greater clock speeds than the Athlon XP or Pentium III could muster. The deeper 20- or 31-stage pipeline in the P4 architecture enables significantly higher clock speeds to be achieved using the same silicon die process as other chips. As an example, the Athlon XP and Pentium 4 were originally made using the same 0.18-micron process (which describes the line width

of components etched on the chips). The P4's 20-stage pipeline enabled the 0.18-micron die process to result in chips running up to 2.0GHz, whereas the same process achieves only 1.73GHz in the 10-stage Athlon XP and only 1.13GHz in the 10-stage Pentium III/Celeron. Using the newer 0.13-micron process, the Pentium 4 runs up to 3.4GHz and the Athlon XP tops out at 2.2GHz (3200+ model) in the same introduction timeframe. The latest Pentium 4 models (and the dual-core Pentium D and Pentium Extreme Edition) use the 0.09-micron process to reach clock speeds up to 3.8GHz. Even though the Pentium 4 executes fewer instructions in each cycle, the overall higher cycling speeds make up for the loss of efficiency. So, in the end, for the initial crop of Athlon XP and Pentium 4 processors, higher clock speed versus more efficient processing effectively cancelled each other out.

Note

If you want to determine the designed clock speed for any type of AMD processor (the actual clock speed could vary according to motherboard overclocking, underclocking, or power management clock speed adjustments), go to AMD's website and download the Data Sheet for the processor model you are interested in. You will find a table in each data sheet that lists the actual MHz (divide by 1,000 for the GHz) for each model.

One thing is clear in all of this confusion: Raw MHz (or GHz) is not always a good way to compare chips, and generating pseudo-MHz numbers can only make things more confusing for the uninitiated. Even Intel moved away from using clock speed as its primary marketing designation. It still notes the speeds of its chips, but the processors are labeled and marketed primarily by model number. This is necessary because the relative difference between each model number is based not just on the CPU's speed, but also on architectural and other differences that affect overall performance.

►► See "Intel Processor Model Numbers," p. 207, for more information.

Overclocking

As is discussed in detail in Chapter 21, "PC Mods: Overclocking and Cooling," in some systems, the processor speed can be set higher than the rating on the chip; this is called *overclocking* the chip. In many cases, you can get away with a certain amount of overclocking because Intel, AMD, and others often build safety margins into their ratings. So, a chip rated for, say, 800MHz might in fact run at 900MHz or more but instead be down-rated to allow for a greater margin of reliability. By overclocking, you are using this margin and running the chip closer to its true maximum speed. I don't normally recommend overclocking for a novice, but if you are comfortable playing with your system settings, and you can afford and are capable of dealing with any potential consequences, overclocking might enable you to get 10%–20% or more performance from your system.

Overclocking Pitfalls

If you are intent on overclocking, there are several issues to consider. One is that most Intel processors since the Pentium II are multiplier-locked before they are shipped out. Therefore, the chip ignores any changes to the multiplier setting on the motherboard. Actually, both Intel and AMD lock the multipliers on most of their newer processors, but the AMD processors use solder bridges on top of the chip that can be manipulated if you are careful and somewhat mechanically inclined. Although originally done to prevent re-marketers from fraudulently relabeling processors (creating "counterfeit" chips), this has impacted the computing performance enthusiast, leaving tweaking the motherboard bus speed as the only easy way (or in some cases, the only way possible) to achieve a clock speed higher than standard.

You can run into problems increasing motherboard bus speed, as well. Most older Intel motherboards, for example, simply don't support clock speeds other than the standard 66MHz, 100MHz, 133MHz, 400MHz, 533MHz or 800MHz settings. Newer Intel boards have a "burn-in" feature that allows you to increase the default processor bus speed (and also the speed of the processor core) by up to 4%. That is relatively mild, but achievable with most chips. Most other brands of motherboards allow changing the bus speeds by even greater amounts, as well as in small increments, sometimes as small as

1MHz changes. Small incremental changes in clock multiplier speeds, rather than large jumps, are the best way to coax a bit more performance out of a particular processor. This is because a given chip is generally overclockable by a certain percentage. The smaller the steps you can take when increasing speed, the more likely that you'll be able to come close to the actual maximum speed of the chip without going over that amount and causing system instability.

For example, the Asus P5LD2 motherboard for Socket 775 Pentium 4 processors supports the standard motherboard bus speeds of 533MHz, 800MHz, and 1066MHz. However, it also permits 1MHz adjustments of the front-side bus (which is multiplied by 4 to obtain the motherboard bus) to enable you to fine-tune your processor speed. Assume you have a 2.8GHz processor with an 800MHz motherboard (FSB) bus. The CPU frequency is 200MHz and is multiplied by 4 to obtain the motherboard bus (FSB) speed: $800\text{MHz} \times 3.5 = 2800\text{MHz}$, or 2.8GHz. Here are the actual speeds you could achieve by adjusting the CPU frequency from 200MHz (the standard setting for Intel CPUs with an 800MHz motherboard bus or FSB) to 220MHz:

CPU Frequency Set in BIOS	CPU Frequency Multiplier to Calculate FSB	Front-Side Bus (FSB)	Multiplier to Calculate Processor Speed	Processor Speed
200MHz	4x	800MHz	3.5x	2.80GHz
201MHz	4x	804MHz	3.5x	2.814GHz
202MHz	4x	808MHz	3.5x	2.828GHz
203MHz	4x	812MHz	3.5x	2.842GHz
204MHz	4x	816MHz	3.5x	2.856GHz
205MHz	4x	820MHz	3.5x	2.870GHz
206MHz	4x	824MHz	3.5x	2.884GHz
207MHz	4x	828MHz	3.5x	2.898GHz
208MHz	4x	832MHz	3.5x	2.912GHz
209MHz	4x	836MHz	3.5x	2.926GHz
210MHz	4x	840MHz	3.5x	2.940GHz
211MHz	4x	844MHz	3.5x	2.954GHz
212MHz	4x	848MHz	3.5x	2.968GHz
213MHz	4x	852MHz	3.5x	2.982GHz
214MHz	4x	856MHz	3.5x	2.996GHz
215MHz	4x	860MHz	3.5x	3.010GHz
216MHz	4x	864MHz	3.5x	3.024GHz
217MHz	4x	868MHz	3.5x	3.038GHz
218MHz	4x	872MHz	3.5x	3.052GHz
219MHz	4x	876MHz	3.5x	3.066GHz
220MHz	4x	880MHz	3.5x	3.080GHz

Typically, a 10%–20% increase is successful, especially if your system offers excellent cooling and you can also adjust CPU voltage and other settings. So with this motherboard, you are likely to get your processor running 200MHz or faster than it was originally designed for.

An issue when it comes to increasing CPU bus speeds is that the other buses in the system will typically be similarly affected. Thus, if you increase the CPU bus speed by 10%, you might also be increasing the PCI or AGP bus by the same amount, and your video, network, or other cards might not be able to keep up. This is something that varies from board to board, so you have to consider each example as a potentially unique case. If possible, configure the AGP and PCI buses to run at their normal speeds through the appropriate BIOS settings.

Overclocking Socket A Processors

The AMD Athlon and Duron processors in the FC-PGA (flip-chip pin grid array) format, which plugs into Socket A, have special solder bridges on the top face of the chip that can be modified to change or remove the lock from the internal multiplier on the chip. This can increase the speed of the chip without changing the motherboard bus speed, thus affecting other buses or cards.

The selected multiplier is set or locked by very small solder connections between solder dots (contacts) on the surface of the chip. You can completely unlock the chip by bridging or disconnecting the appropriate dots. Unfortunately, it is somewhat difficult to add or remove these bridges; you usually have to mask off the particular bridge you want to create and, rather than dripping solder onto it, literally paint the bridge with silver or copper paint. For example, you can use the special copper paint sold in small vials at any auto parts store for repairing the window defogger grids. The real problem is that the contacts are very small, and if you bridge to adjacent rather than opposite contacts, you can render the chip nonfunctional. An Xacto knife or razor blade can be used to remove the bridges if desired. If you are not careful, you can easily damage a processor worth several hundred dollars. If you are leery of making such changes, you should try bus overclocking instead because this is done in the BIOS Setup and can easily be changed or undone without any mechanical changes to the chip.

CPU Voltage Settings

Another trick used by overclockers is playing with the voltage settings for the CPU. All modern CPU sockets and slots, including Slot 1, Slot A, Socket 8, Socket 370, Socket 423, Socket 462 (Socket A), Socket 478, Socket 754, Socket 775, Socket 939, and Socket 940, have automatic voltage detection. With this detection, the system detects and sets the correct voltage by reading certain pins on the processor. Some motherboards, such as those made by Intel, do not allow any manual changes to these settings. Other motherboards, such as the Asus P5LD2 mentioned earlier, allow you to tweak the voltage settings from the automatic setting up or down by fractions of a volt. Some experimenters have found that by either increasing or decreasing voltage slightly from the standard, a higher speed of overclock can be achieved with the system remaining stable.

My recommendation is to be careful when playing with voltages because you can damage the chip in this manner. Even without changing voltage, overclocking with an adjustable bus speed motherboard is very easy and fairly rewarding. I do recommend you make sure you are using a high-quality board, good memory, and especially a good system chassis with additional cooling fans and a heavy-duty power supply. See Chapter 19, "Power Supplies," for more information on upgrading power supplies and chassis. Especially when overclocking, it is essential that the system components and the CPU remain properly cooled. Going a little bit overboard on the processor heatsink and adding extra cooling fans to the case never hurts and in many cases helps a great deal when hotrodding a system in this manner.

Cache Memory

As processor core speeds increased, memory speeds could not keep up. How could you run a processor faster than the memory from which you feed it without having performance suffer terribly? The answer was cache. In its simplest terms, *cache memory* is a high-speed memory buffer that temporarily stores data the processor needs, allowing the processor to retrieve that data faster than if it came from main memory. But there is one additional feature of a cache over a simple buffer, and that is intelligence. A cache is a buffer with a brain.

A buffer holds random data, usually on a first in, first out, or first in, last out basis. A cache, on the other hand, holds the data the processor is most likely to need in advance of it actually being needed. This enables the processor to continue working at either full speed or close to it without having to wait for the data to be retrieved from slower main memory. Cache memory is usually made up of static RAM (SRAM) memory integrated into the processor die, although older systems with cache also used chips installed on the motherboard.

►► See "Cache Memory: SRAM," p. 476.

For the vast majority of desktop systems, there are two levels of processor/memory cache used in a modern PC: Level 1 (L1) and Level 2 (L2). Some processors—most of them designed for use in servers such as the Itanium series from Intel—also have Level 3 cache. The most notable desktop processor to use an L3 cache is the Pentium 4 Extreme Edition. These caches and how they function are described in the following sections.

Internal Level 1 Cache

All modern processors starting with the 486 family include an integrated L1 cache and controller. The integrated L1 cache size varies from processor to processor, starting at 8KB for the original 486DX and now up to 32KB, 64KB, or more in the latest processors.

To understand the importance of cache, you need to know the relative speeds of processors and memory. The problem with this is that processor speed usually is expressed in MHz or GHz (millions or billions of cycles per second), whereas memory speeds are often expressed in nanoseconds (billionths of a second per cycle). Most newer types of memory express the speed in either MHz or in megabyte per second (MBps) bandwidth (throughput).

Both are really time- or frequency-based measurements, and a chart comparing them can be found in Table 6.3 in Chapter 6, “Memory.” In this table, you will note that a 233MHz processor equates to 4.3-nanosecond cycling, which means you would need 4ns memory to keep pace with a 200MHz CPU. Also note that the motherboard of a 233MHz system typically runs at 66MHz, which corresponds to a speed of 15ns per cycle and requires 15ns memory to keep pace. Finally, note that 60ns main memory (common on many Pentium-class systems) equates to a clock speed of approximately 16MHz. So, a typical Pentium 233 system has a processor running at 233MHz (4.3ns per cycle), a motherboard running at 66MHz (15ns per cycle), and main memory running at 16MHz (60ns per cycle). This might seem like a rather dated example, but in a moment, you will see that the figures listed here make it easy for me to explain how cache memory works.

Because L1 cache is always built into the processor die, it runs at the full-core speed of the processor internally. By full-core speed, I mean this cache runs at the higher clock multiplied internal processor speed rather than the external motherboard speed. This cache basically is an area of very fast memory built into the processor and is used to hold some of the current working set of code and data. Cache memory can be accessed with no wait states because it is running at the same speed as the processor core.

Using cache memory reduces a traditional system bottleneck because system RAM is almost always much slower than the CPU; the performance difference between memory and CPU speed has become especially large in recent systems. Using cache memory prevents the processor from having to wait for code and data from much slower main memory, therefore improving performance. Without the L1 cache, a processor would frequently be forced to wait until system memory caught up.

Cache is even more important in modern processors because it is often the only memory in the entire system that can truly keep up with the chip. Most modern processors are clock multiplied, which means they are running at a speed that is really a multiple of the motherboard into which they are plugged. The Pentium 4 2.8GHz, for example, runs at a multiple of 5.25 times the true motherboard speed of 533MHz. The main memory is one half this speed (266MHz) because the Pentium 4 uses a quad-pumped memory bus. Because the main memory is plugged into the motherboard, it can run only at 266MHz maximum. The only 2.8GHz memory in such a system is the L1 and L2 caches built into the processor core. In this example, the Pentium 4 2.8GHz processor has 20KB of integrated L1 cache (8KB data cache and 12KB execution trace cache) and 512KB of L2, all running at the full speed of the processor core.

►► See “Memory Module Speed,” p. 513.

If the data the processor wants is already in the internal cache, the CPU does not have to wait. If the data is not in the cache, the CPU must fetch it from the Level 2 cache or (in less sophisticated system designs) from the system bus, meaning main memory directly.

How Cache Works

To learn how the L1 cache works, consider the following analogy.

This story involves a person (in this case you) eating food to act as the processor requesting and operating on data from memory. The kitchen where the food is prepared is the main system memory (typically DDR or DDR2 DIMMs). The cache controller is the waiter, and the L1 cache is the table at which you are seated.

Okay, here's the story. Say you start to eat at a particular restaurant every day at the same time. You come in, sit down, and order a hot dog. To keep this story proportionately accurate, let's say you normally eat at the rate of one bite (byte? <g>) every four seconds (233MHz = about 4ns cycling). It also takes 60 seconds for the kitchen to produce any given item that you order (60ns main memory).

So, when you first arrive, you sit down, order a hot dog, and you have to wait for 60 seconds for the food to be produced before you can begin eating. After the waiter brings the food, you start eating at your normal rate. Pretty quickly you finish the hot dog, so you call the waiter over and order a hamburger. Again you wait 60 seconds while the hamburger is being produced. When it arrives, you again begin eating at full speed. After you finish the hamburger, you order a plate of fries. Again you wait, and after it is delivered 60 seconds later, you eat it at full speed. Finally, you decide to finish the meal and order cheesecake for dessert. After another 60-second wait, you can eat cheesecake at full speed. Your overall eating experience consists of mostly a lot of waiting, followed by short bursts of actual eating at full speed.

After coming into the restaurant for two consecutive nights at exactly 6 p.m. and ordering the same items in the same order each time, on the third night the waiter begins to think, "I know this guy is going to be here at 6 p.m., order a hot dog, a hamburger, fries, and then cheesecake. Why don't I have these items prepared in advance and surprise him? Maybe I'll get a big tip." So you enter the restaurant and order a hot dog, and the waiter immediately puts it on your plate, with no waiting! You then proceed to finish the hot dog and right as you are about to request the hamburger, the waiter deposits one on your plate. The rest of the meal continues in the same fashion, and you eat the entire meal, taking a bite every four seconds, and never have to wait for the kitchen to prepare the food. Your overall eating experience this time consists of all eating, with no waiting for the food to be prepared, due primarily to the intelligence and thoughtfulness of your waiter.

This analogy exactly describes the function of the L1 cache in the processor. The L1 cache itself is the table that can contain one or more plates of food. Without a waiter, the space on the table is a simple food buffer. When stocked, you can eat until the buffer is empty, but nobody seems to be intelligently refilling it. The waiter is the cache controller who takes action and adds the intelligence to decide which dishes are to be placed on the table in advance of your needing them. Like the real cache controller, he uses his skills to literally guess which food you will require next, and if and when he guesses right, you never have to wait.

Let's now say on the fourth night you arrive exactly on time and start off with the usual hot dog. The waiter, by now really feeling confident, has the hot dog already prepared when you arrive, so there is no waiting.

Just as you finish the hot dog, and right as he is placing a hamburger on your plate, you say "Gee, I'd really like a bratwurst now; I didn't actually order this hamburger." The waiter guessed wrong, and the consequence is that this time you have to wait the full 60 seconds as the kitchen prepares your brat. This is known as a *cache miss*, in which the cache controller did not correctly fill the cache with the data the processor actually needed next. The result is waiting, or in the case of a sample 233MHz Pentium system, the system essentially throttles back to 16MHz (RAM speed) whenever a cache miss occurs.

According to Intel, the L1 cache in most of its processors has approximately a 90% hit ratio (some processors, such as the Pentium 4, are slightly higher). This means that the cache has the correct data 90% of the time, and consequently the processor runs at full speed—233MHz in this example—90% of the time. However, 10% of the time the cache controller guesses wrong and the data has to be retrieved out of the significantly slower main memory, meaning the processor has to wait. This essentially throttles the system back to RAM speed, which in this example was 60ns or 16MHz.

In this analogy, the processor was 14 times faster than the main memory. Memory speeds have increased from 16MHz (60ns) to 333MHz (3.0ns) or faster in the latest systems, but processor speeds have also risen to 3GHz and beyond, so even in the latest systems, memory is still 7.5 or more times *slower* than the processor. Cache is what makes up the difference.

The main feature of L1 cache is that it has always been integrated into the processor core, where it runs at the same speed as the core. This, combined with the hit ratio of 90% or greater, makes L1 cache very important for system performance.

Level 2 Cache

To mitigate the dramatic slowdown every time an L1 cache miss occurs, a secondary (L2) cache is employed.

Using the restaurant analogy I used to explain L1 cache in the previous section, I'll equate the L2 cache to a cart of additional food items placed strategically in the restaurant such that the waiter can retrieve food from the cart in only 15 seconds (versus 60 seconds from the kitchen). In an actual Pentium class (Socket 7) system, the L2 cache is mounted on the motherboard, which means it runs at motherboard speed—66MHz, or 15ns in this example. Now, if you ask for an item the waiter did not bring in advance to your table, instead of making the long trek back to the kitchen to retrieve the food and bring it back to you 60 seconds later, he can first check the cart where he has placed additional items. If the requested item is there, he will return with it in only 15 seconds. The net effect in the real system is that instead of slowing down from 233MHz to 16MHz waiting for the data to come from the 60ns main memory, the data can instead be retrieved from the 15ns (66MHz) L2 cache. The effect is that the system slows down from 233MHz to 66MHz.

Newer processors have integrated L2 cache that runs at the same speed as the processor core, which is also the same speed as the L1 cache. For the analogy to describe these newer chips, the waiter would simply place the cart right next to the table you were seated at in the restaurant. Then, if the food you desired wasn't on the table (L1 cache miss), it would merely take a longer reach over to the adjacent L2 cache (the cart, in this analogy) rather than a 15-second walk to the cart as with the older designs.

Level 3 Cache

A few processors, primarily those designed for very high-performance desktop operation or enterprise-level servers, contain a third level of cache known as L3 cache. Relatively few processors have L3 cache, but those that do access it at the same speed as L1 and L2 cache.

Extending the restaurant analogy I used to explain L1 and L2 caches, I'll equate L3 cache to another cart of additional food items placed in the restaurant next to the cart used to symbolize L2 cache. If the food item needed was not on the table (L1 cache miss) or on the first food cart (L2 cache miss), the waiter could then reach over to the second food cart to retrieve a necessary item.

Although Intel has used L3 caches with the first version of the Pentium 4 Extreme Edition processor and with the Itanium 2 and Xeon MP server processors, more recent desktop processors—including the dual-core Pentium D and Pentium Extreme Edition processors—use large L2 caches instead of a separate L3 cache.

Cache Performance and Design

Just as with the L1 cache, most L2 caches have a hit ratio also in the 90% range; therefore, if you look at the system as a whole, 90% of the time it will be running at full speed (233MHz in this example) by retrieving data out of the L1 cache. Ten percent of the time it will slow down to retrieve the data from the L2 cache. Ninety percent of the time the processor goes to the L2 cache, the data will be in the L2, and 10% of that time it will have to go to the slow main memory to get the data because of an L2 cache miss. So, by combining both caches, our sample system runs at full processor speed 90% of the time (233MHz in this case), at motherboard speed 9% (90% of 10%) of the time (66MHz in this case),

and at RAM speed about 1% (10% of 10%) of the time (16MHz in this case). You can clearly see the importance of both the L1 and L2 caches; without them the system uses main memory more often, which is significantly slower than the processor.

This brings up other interesting points. If you could spend money doubling the performance of either the main memory (RAM) or the L2 cache, which would you improve? Considering that main memory is used directly only about 1% of the time, if you doubled performance there, you would double the speed of your system only 1% of the time! That doesn't sound like enough of an improvement to justify much expense. On the other hand, if you doubled L2 cache performance, you would be doubling system performance 9% of the time, a much greater improvement overall. I'd much rather improve L2 than RAM performance.

The processor and system designers at Intel and AMD know this and have devised methods of improving the performance of L2 cache. In Pentium (P5) class systems, the L2 cache usually was found on the motherboard and had to therefore run at motherboard speed. Intel made the first dramatic improvement by migrating the L2 cache from the motherboard directly into the processor and initially running it at the same speed as the main processor. The cache chips were made by Intel and mounted next to the main processor die in a single chip housing. This proved too expensive, so with the Pentium II, Intel began using cache chips from third-party suppliers such as Sony, Toshiba, NEC, Samsung, and others. Because these were supplied as complete packaged chips and not raw die, Intel mounted them on a circuit board alongside the processor. This is why the Pentium II was designed as a cartridge rather than what looked like a chip.

One problem was the speed of the available third-party cache chips. The fastest ones on the market were 3ns or higher, meaning 333MHz or less in speed. Because the processor was being driven in speed above that, in the Pentium II and initial Pentium III processors Intel had to run the L2 cache at half the processor speed because that is all the commercially available cache memory could handle. AMD followed suit with the Athlon processor, which had to drop L2 cache speed even further in some models to two-fifths or one-third the main CPU speed to keep the cache memory speed less than the 333MHz commercially available chips.

Then a breakthrough occurred, which first appeared in Celeron processors 300A and above. These had 128KB of L2 cache, but no external chips were used. Instead, the L2 cache had been integrated directly into the processor core just like the L1. Consequently, both the L1 and L2 caches now would run at full processor speed, and more importantly scale up in speed as the processor speeds increased in the future. In the newer Pentium III, as well as all the Xeon and Celeron processors, the L2 cache runs at full processor core speed, which means there is no waiting or slowing down after an L1 cache miss. AMD also achieved full-core speed on-die cache in its later Athlon and Duron chips. Using on-die cache improves performance dramatically because 9% of the time the system would be using the L2, it would now remain at full speed instead of slowing down to one-half or less the processor speed or, even worse, slow down to motherboard speed as in Socket 7 designs. Another benefit of on-die L2 cache is cost, which is less because now fewer parts are involved.

Let's revisit the restaurant analogy using a modern Pentium 4 3.6GHz processor. You would now be taking a bite every half second (3.6GHz = 0.28ns cycling). The L1 cache would also be running at that speed, so you could eat anything on your table at that same rate (the table = L1 cache). The real jump in speed comes when you want something that isn't already on the table (L1 cache miss), in which case the waiter reaches over to the cart (which is now directly adjacent to the table) and nine out of ten times is able to find the food you want in just over one-quarter second (L2 speed = 3.6GHz or 0.28ns cycling). In this more modern system, you would run at 3.6GHz 99% of the time (L1 and L2 hit ratios combined) and slow down to RAM speed (wait for the kitchen) only 1% of the time as before. With faster memory running at 400MHz (2.5ns), you would have to wait only 2.5 seconds for the food to come from the kitchen. If only restaurant performance would increase at the same rate processor performance has!

Cache Organization

You know that cache stores copies of data from various main memory addresses. Because the cache cannot hold copies of the data from all the addresses in main memory simultaneously, there has to be a way to know which addresses are currently copied into the cache so that, if we need data from those addresses, it can be read from the cache rather than from the main memory. This function is performed by Tag RAM, which is additional memory in the cache that holds an index of the addresses that are copied into the cache. Each line of cache memory has a corresponding address tag that stores the main memory address of the data currently copied into that particular cache line. If data from a particular main memory address is needed, the cache controller can quickly search the address tags to see whether the requested address is currently being stored in the cache (a *hit*) or not (a *miss*). If the data is there, it can be read from the faster cache; if it isn't, it has to be read from the much slower main memory.

Various ways of organizing or mapping the tags affect how cache works. A cache can be mapped as fully associative, direct-mapped, or set associative.

In a fully associative mapped cache, when a request is made for data from a specific main memory address, the address is compared against all the address tag entries in the cache tag RAM. If the requested main memory address is found in the tag (a *hit*), the corresponding location in the cache is returned. If the requested address is not found in the address tag entries, a *miss* occurs and the data must be retrieved from the main memory address instead of the cache.

In a direct-mapped cache, specific main memory addresses are preassigned to specific line locations in the cache where they will be stored. Therefore, the tag RAM can use fewer bits because when you know which main memory address you want, only one address tag needs to be checked and each tag needs to store only the possible addresses a given line can contain. This also results in faster operation because only one tag address needs to be checked for a given memory address.

A set associative cache is a modified direct-mapped cache. A direct-mapped cache has only one set of memory associations, meaning a given memory address can be mapped into (or associated with) only a specific given cache line location. A two-way set associative cache has two sets, so that a given memory location can be in one of two locations. A four-way set associative cache can store a given memory address into four different cache line locations (or sets). By increasing the set associativity, the chance of finding a value increases; however, it takes a little longer because more tag addresses must be checked when searching for a specific location in the cache. In essence, each set in an *n*-way set associative cache is a subcache that has associations with each main memory address. As the number of subcaches or sets increases, eventually the cache becomes fully associative—a situation in which any memory address can be stored in any cache line location. In that case, an *n*-way set associative cache is a compromise between a fully associative cache and a direct-mapped cache.

In general, a direct-mapped cache is the fastest at locating and retrieving data from the cache because it has to look at only one specific tag address for a given memory address. However, it also results in more misses overall than the other designs. A fully associative cache offers the highest hit ratio but is the slowest at locating and retrieving the data because it has many more address tags to check through. An *n*-way set associative cache is a compromise between optimizing cache speed and hit ratio, but the more associativity there is, the more hardware (tag bits, comparator circuits, and so on) is required, making the cache more expensive. Obviously, cache design is a series of tradeoffs, and what works best in one instance might not work best in another. Multitasking environments such as Windows are good examples of environments in which the processor needs to operate on different areas of memory simultaneously and in which an *n*-way cache can improve performance.

The organization of the cache memory in the 486 and MMX Pentium family is called a *four-way set associative cache*, which means that the cache memory is split into four blocks. Each block also is

organized as 128 or 256 lines of 16 bytes each. The following table shows the associativity of various processor L1 and L2 caches.

Processor	L1 Cache Associativity	L2 Cache Associativity
486	Four-way	Not in CPU
Pentium (non-MMX)	Two-way	Not in CPU
Pentium MMX	Four-way	Not in CPU
Pentium Pro/II/III	Four-way	Four-way (off-die)
Pentium III/4	Four-way	Eight-way (on-die)

The contents of the cache must always be in sync with the contents of main memory to ensure that the processor is working with current data. For this reason, the internal cache in the 486 family is a *write-through* cache. Write-through means that when the processor writes information out to the cache, that information is automatically written through to main memory as well.

By comparison, the Pentium and later chips have an internal write-back cache, which means that both reads and writes are cached, further improving performance. Even though the internal 486 cache is write-through, the system can employ an external write-back cache for increased performance. In addition, the 486 can buffer up to 4 bytes before actually storing the data in RAM, improving efficiency in case the memory bus is busy.

Another feature of improved cache designs is that they are nonblocking. This is a technique for reducing or hiding memory delays by exploiting the overlap of processor operations with data accesses. A *nonblocking* cache enables program execution to proceed concurrently with cache misses as long as certain dependency constraints are observed. In other words, the cache can handle a cache miss much better and enable the processor to continue doing something nondependent on the missing data.

The cache controller built into the processor also is responsible for watching the memory bus when alternative processors, known as *bus masters*, are in control of the system. This process of watching the bus is referred to as *bus snooping*. If a bus master device writes to an area of memory that also is stored in the processor cache currently, the cache contents and memory no longer agree. The cache controller then marks this data as invalid and reloads the cache during the next memory access, preserving the integrity of the system.

All PC processor designs that support cache memory include a feature known as a translation look-aside buffer (TLB) to improve recovery from cache misses. The TLB is a table inside the processor that stores information about the location of recently accessed memory addresses. The TLB speeds up the translation of virtual addresses to physical memory addresses. To improve TLB performance, several recent processors have increased the number of entries in the TLB, as AMD did when it moved from the Athlon Thunderbird core to the Palomino core. Pentium 4 processors that support HT Technology have a separate instruction TLB (iTLB) for each virtual processor thread.

►► See "Hyper-Threading Technology," p. 77.

As clock speeds increase, cycle time decreases. Newer systems don't use cache on the motherboard any longer because the faster system memory used in modern systems can keep up with the motherboard speed. Modern processors all integrate the L2 cache into the processor die just like the L1 cache. This enables the L2 to run at full-core speed because it is now a part of the core. Cache speed is always more important than size. The rule is that a smaller but faster cache is always better than a slower but bigger cache. Table 3.16 illustrates the need for and function of L1 (internal) and L2 (external) caches in modern systems.

Table 3.16 CPU Speeds Relative to Cache, RAM, and Motherboard

CPU Type	Pentium	Pentium Pro	Pentium II
CPU speed	233MHz	200MHz	450MHz
L1 cache speed	4.3ns (233MHz)	5.0ns (200MHz)	2.2ns (450MHz)
L1 cache size	16K	32K	32K
L2 cache type	onboard	on-chip	on-chip
L2 speed ratio	2/7	1/1	1/2
L2 cache speed	1.5ns (66MHz)	5ns (200MHz)	4.4ns (225MHz)
L2 cache size	varies ¹	256KB ²	512KB
CPU bus speed	66MHz	66MHz	100MHz
Memory bus speed	60ns (16MHz)	60ns (16MHz)	10ns (100MHz)
CPU Type	AMD K6-2	AMD K6-3	Pentium III
CPU speed	550MHz	450MHz	1.4GHz
L1 cache speed	1.8ns (550MHz)	2.2ns (450MHz)	0.71ns (1.4GHz)
L1 cache size	64K	64K	32K
L2 cache type	onboard	on-die	on-die
L2 speed ratio	2/11	1/1	1/1
L2 cache speed	10ns (100MHz)	2.2ns (450MHz)	0.71ns (1.4GHz)
L2 cache size	varies ¹	256KB	512KB
CPU bus speed	100MHz	100MHz	133MHz
Memory bus speed	10ns (100MHz)	10ns (100MHz)	7.5ns (133MHz)
CPU speed	2.2GHz	2.4GHz	4.0GHz
L1 cache speed	0.45ns (2.2GHz)	0.42ns (2.4GHz)	0.28ns (3.6GHz)
L1 cache size	128K	128K	28K
L2 cache type	on-die	on-die	on-die
L2 speed ratio	1/1	1/1	1/1
L2 cache speed	0.45ns (2.2GHz)	0.42ns (2.4GHz)	0.28ns (3.6GHz)
L2 cache size	512KB	1024KB	1024KB
CPU bus speed	400MHz	400MHz	800MHz
Memory bus speed	2.5ns (400MHz)	2.5ns (400MHz)	2.5ns (400MHz) ³

1. The L2 cache is on the motherboard, and the amount depends on which board is chosen and how much is installed.

2. The Pentium Pro was also available with 512KB and 1024KB L2 cache.

3. Dual-channel memory uses two banks simultaneously, doubling the throughput.

As you can see, having two levels of cache between the very fast CPU and the much slower main memory helps minimize any wait states the processor might have to endure, especially those with the on-die L2. This enables the processor to keep working closer to its true speed.

Processor Features

As new processors are introduced, new features are continually added to their architectures to help improve everything from performance in specific types of applications to the reliability of the CPU as a whole. The next few sections take a look at some of these technologies, including System Management Mode (SMM), Superscalar Execution, MMX, SSE, 3DNow!, HT Technology, and dual-core processing.

SMM (Power Management)

Spurred on primarily by the goal of putting faster and more powerful processors in laptop computers, Intel has created power-management circuitry. This circuitry enables processors to conserve energy use and lengthen battery life. This was introduced initially in the Intel 486SL processor, which is an enhanced version of the 486DX processor. Subsequently, the power-management features were universalized and incorporated into all 75MHz and faster Pentium and later processors. This feature set is called SMM, which stands for *system management mode*.

SMM circuitry is integrated into the physical chip but operates independently to control the processor's power use based on its activity level. It enables the user to specify time intervals after which the CPU will be partially or fully powered down. It also supports the Suspend/Resume feature that allows for instant power on and power off, used mostly with laptop PCs. These settings are typically controlled via system BIOS settings.

Superscalar Execution

The fifth-generation Pentium and newer processors feature multiple internal instruction execution pipelines, which enable them to execute multiple instructions at the same time. The 486 and all preceding chips can perform only a single instruction at a time. Intel calls the capability to execute more than one instruction at a time *superscalar* technology. This technology provides additional performance compared with the 486.

►► See "Pentium Processors," p. 123.

Superscalar architecture usually is associated with high-output Reduced Instruction Set Computer (RISC) chips. A RISC chip has a less complicated instruction set with fewer and simpler instructions. Although each instruction accomplishes less, overall the clock speed can be higher, which can usually increase performance. The Pentium is one of the first Complex Instruction Set Computer (CISC) chips to be considered superscalar. A CISC chip uses a richer, fuller-featured instruction set, which has more complicated instructions. As an example, say you wanted to instruct a robot to screw in a light bulb. Using CISC instructions, you would say

1. Pick up the bulb.
2. Insert it into the socket.
3. Rotate clockwise until tight.

Using RISC instructions, you would say something more along the lines of

1. Lower hand.
2. Grasp bulb.
3. Raise hand.
4. Insert bulb into socket.
5. Rotate clockwise one turn.
6. Is bulb tight? If not, repeat step 5.
7. End.

Overall, many more RISC instructions are required to do the job because each instruction is simpler (reduced) and does less. The advantage is that there are fewer overall commands the robot (or processor) has to deal with and it can execute the individual commands more quickly, and thus in many cases execute the complete task (or program) more quickly as well. The debate goes on whether RISC or CISC is really better, but in reality there is no such thing as a pure RISC or CISC chip—it is all just a matter of definition, and the lines are somewhat arbitrary.

Intel and compatible processors have generally been regarded as CISC chips, although the fifth- and sixth-generation versions have many RISC attributes and internally break CISC instructions down into RISC versions.

MMX Technology

MMX technology was originally named for multimedia extensions, or matrix math extensions, depending on whom you ask. Intel officially states that it is actually not an abbreviation and stands for nothing other than the letters MMX (not being an abbreviation was apparently required so that the letters could be trademarked); however, the internal origins are probably one of the preceding. MMX technology was introduced in the later fifth-generation Pentium processors as a kind of add-on that improves video compression/decompression, image manipulation, encryption, and I/O processing—all of which are used in a variety of today's software.

MMX consists of two main processor architectural improvements. The first is very basic; all MMX chips have a larger internal L1 cache than their non-MMX counterparts. This improves the performance of any and all software running on the chip, regardless of whether it actually uses the MMX-specific instructions.

The other part of MMX is that it extends the processor instruction set with 57 new commands or instructions, as well as a new instruction capability called single instruction, multiple data (SIMD).

Modern multimedia and communication applications often use repetitive loops that, while occupying 10% or less of the overall application code, can account for up to 90% of the execution time. SIMD enables one instruction to perform the same function on multiple pieces of data, similar to a teacher telling an entire class to “sit down,” rather than addressing each student one at a time. SIMD enables the chip to reduce processor-intensive loops common with video, audio, graphics, and animation.

Intel also added 57 new instructions specifically designed to manipulate and process video, audio, and graphical data more efficiently. These instructions are oriented to the *highly parallel* and often repetitive sequences frequently found in multimedia operations. *Highly parallel* refers to the fact that the same processing is done on many data points, such as when modifying a graphic image. The main drawbacks to MMX were that it worked only on integer values and used the floating-point unit for processing, so time was lost when a shift to floating-point operations was necessary. These drawbacks were corrected in the additions to MMX from Intel and AMD.

Intel licensed the MMX capabilities to competitors such as AMD and Cyrix, who were then able to upgrade their own Intel-compatible processors with MMX technology.

SSE, SSE2, and SSE3

In February 1999, Intel introduced the Pentium III processor and included in that processor an update to MMX called Streaming SIMD Extensions (SSE). These were also called Katmai New Instructions (KNI) up until their debut because they were originally included on the Katmai processor, which was the codename for the Pentium III. The Celeron 533A and faster Celeron processors based on the Pentium III core also support SSE instructions. The earlier Pentium II and Celeron 533 and lower (based on the Pentium II core) do not support SSE.

SSE includes 70 new instructions for graphics and sound processing over what MMX provided. SSE is similar to MMX; in fact, besides being called KNI, SSE was also called MMX-2 by some before it was released. In addition to adding more MMX style instructions, the SSE instructions allow for floating-point calculations and now use a separate unit within the processor instead of sharing the standard floating-point unit as MMX did.

SSE2 was introduced in November 2000, along with the Pentium 4 processor, and adds 144 additional SIMD instructions. SSE2 also includes all the previous MMX and SSE instructions.

SSE3 was introduced in February 2004, along with the Pentium 4 Prescott processor, and adds 13 new SIMD instructions to improve complex math, graphics, video encoding, and thread synchronization. SSE3 also includes all the previous MMX, SSE, and SSE2 instructions.

The Streaming SIMD Extensions consist of new instructions, including SIMD floating-point, additional SIMD integer, and cacheability control instructions. Some of the technologies that benefit from the Streaming SIMD Extensions include advanced imaging, 3D video, streaming audio and video (DVD playback), and speech-recognition applications. The benefits of SSE include the following:

- Higher resolution and higher quality image viewing and manipulation for graphics software
- High-quality audio, MPEG2 video, and simultaneous MPEG2 encoding and decoding for multimedia applications
- Reduced CPU utilization for speech recognition, as well as higher accuracy and faster response times when running speech-recognition software

The SSEx instructions are particularly useful with MPEG2 decoding, which is the standard scheme used on DVD video discs. SSE-equipped processors should therefore be more capable of performing MPEG2 decoding in software at full speed without requiring an additional hardware MPEG2 decoder card. SSE-equipped processors are much better and faster than previous processors when it comes to speech recognition, as well.

One of the main benefits of SSE over plain MMX is that it supports single-precision floating-point SIMD operations, which have posed a bottleneck in the 3D graphics processing. Just as with plain MMX, SIMD enables multiple operations to be performed per processor instruction. Specifically, SSE supports up to four floating-point operations per cycle; that is, a single instruction can operate on four pieces of data simultaneously. SSE floating-point instructions can be mixed with MMX instructions with no performance penalties. SSE also supports data *prefetching*, which is a mechanism for reading data into the cache before it is actually called for.

Note that for any of the SSE instructions to be beneficial, they must be encoded in the software you are using, so SSE-aware applications must be used to see the benefits. Most software companies writing graphics- and sound-related software today have updated those applications to be SSE aware and use the features of SSE. For example, high-powered graphics applications such as Adobe Photoshop support SSE instructions for higher performance on processors equipped with SSE. Microsoft includes support for SSE in its DirectX 6.1 and later video and sound drivers, which are included with Windows 98 Second Edition, Windows Me, Windows NT 4.0 (with service pack 5 or later), Windows 2000, and Windows XP.

SSE is an extension to MMX; SSE2 is an extension to SSE; and SSE3 is an extension to SSE2. Therefore, processors that support SSE3 also support the SSE2 instructions, processors that support SSE2 also support SSE, and processors that support SSE also support the original MMX instructions. This means that standard MMX-enabled applications run as they did on MMX-only processors.

The first AMD processors to support SSE3 are the 0.09-micron versions of the Athlon 64 and all versions of the dual-core Athlon 64 X2.

3DNow!, Enhanced 3DNow!, and Professional 3DNow!

3DNow! technology was originally introduced as AMD's alternative to the SSE instructions in the Intel processors. Actually, 3DNow! was first introduced in the K6 series before Intel released SSE in the Pentium III, and then AMD added Enhanced 3DNow! to the Athlon and Duron processors. The latest version, Professional 3DNow!, was introduced in the first Athlon XP processors. AMD licensed MMX from Intel, and all its K6 series, Athlon, Duron, and later processors include full MMX instruction support. Not wanting to additionally license the SSE instructions being developed by Intel, AMD first came up with a

different set of extensions beyond MMX called 3DNow!. Introduced in May 1998 in the K6-2 processor and enhanced when the Athlon was introduced in June 1999, 3DNow!, and Enhanced 3DNow! are sets of instructions that extend the multimedia capabilities of the AMD chips beyond MMX. This enables greater performance for 3D graphics, multimedia, and other floating-point-intensive PC applications.

3DNow! technology is a set of 21 instructions that uses SIMD techniques to operate on arrays of data rather than single elements. Enhanced 3DNow! adds 24 more instructions (19 SSE and 5 DSP/communications instructions) to the original 21 for a total of 45 new instructions. Positioned as an extension to MMX technology, 3DNow! is similar to the SSE found in the Pentium III and Celeron processors from Intel. According to AMD, 3DNow! provides approximately the same level of improvement to MMX as did SSE, but in fewer instructions with less complexity. Although similar in capability, they are not compatible at the instruction level, so software specifically written to support SSE does not support 3DNow!, and vice versa. The latest version of 3DNow!, 3DNow! Professional, adds 51 SSE commands to 3DNow! Enhanced, meaning that 3DNow! Professional now supports all SSE commands, meaning that AMD chips now essentially have SSE capability. Unfortunately, AMD includes SSE2 only on the Athlon 64, Athlon 64FX, and Opteron 64-bit processors.

Just as with SSE, 3DNow! also supports single precision floating-point SIMD operations and enables up to four floating-point operations per cycle. 3DNow! floating-point instructions can be mixed with MMX instructions with no performance penalties. 3DNow! also supports data prefetching.

Also like SSE, 3DNow! is well supported by software, including Windows 9x, Windows NT 4.0, and all newer Microsoft operating systems. 3DNow!-specific support is no longer a big issue if you are using an Athlon XP or Athlon 64 processor because they now fully support SSE through their support of 3DNow! Professional.

Dynamic Execution

First used in the P6 or sixth-generation processors, dynamic execution enables the processor to execute more instructions on parallel, so tasks are completed more quickly. This technology innovation is comprised of three main elements:

- *Multiple branch prediction.* Predicts the flow of the program through several branches
- *Dataflow analysis.* Schedules instructions to be executed when ready, independent of their order in the original program
- *Speculative execution.* Increases the rate of execution by looking ahead of the program counter and executing instructions that are likely to be necessary

Branch Prediction

Branch prediction is a feature formerly found only in high-end mainframe processors. It enables the processor to keep the instruction pipeline full while running at a high rate of speed. A special fetch/decode unit in the processor uses a highly optimized branch prediction algorithm to predict the direction and outcome of the instructions being executed through multiple levels of branches, calls, and returns. It is similar to a chess player working out multiple strategies in advance of game play by predicting the opponent's strategy several moves into the future. By predicting the instruction outcome in advance, the instructions can be executed with no waiting.

Dataflow Analysis

Dataflow analysis studies the flow of data through the processor to detect any opportunities for out-of-order instruction execution. A special dispatch/execute unit in the processor monitors many instructions and can execute these instructions in an order that optimizes the use of the multiple superscalar execution units. The resulting out-of-order execution of instructions can keep the execution units busy even when cache misses and other data-dependent instructions might otherwise hold things up.

Speculative Execution

Speculative execution is the processor's capability to execute instructions in advance of the actual program counter. The processor's dispatch/execute unit uses dataflow analysis to execute all available instructions in the instruction pool and store the results in temporary registers. A retirement unit then searches the instruction pool for completed instructions that are no longer data dependent on other instructions to run or which have unresolved branch predictions. If any such completed instructions are found, the results are committed to memory by the retirement unit or the appropriate standard Intel architecture in the order they were originally issued. They are then retired from the pool.

Dynamic execution essentially removes the constraint and dependency on linear instruction sequencing. By promoting out-of-order instruction execution, it can keep the instruction units working rather than waiting for data from memory. Even though instructions can be predicted and executed out of order, the results are committed in the original order so as not to disrupt or change program flow. This enables the P6 to run existing Intel architecture software exactly as the P5 (Pentium) and previous processors did—just a whole lot more quickly!

Dual Independent Bus Architecture

The Dual Independent Bus (DIB) architecture was first implemented in the sixth-generation processors from Intel and AMD. DIB was created to improve processor bus bandwidth and performance. Having two (dual) independent data I/O buses enables the processor to access data from either of its buses simultaneously and in parallel, rather than in a singular sequential manner (as in a single-bus system). The main (often called front-side) processor bus is the interface between the processor and the motherboard or chipset. The second (back-side) bus in a processor with DIB is used for the L2 cache, enabling it to run at much greater speeds than if it were to share the main processor bus.

Two buses make up the DIB architecture: the L2 cache bus and the main CPU bus, often called FSB (front-side bus). The P6 class processors from the Pentium Pro to the Celeron, Pentium II/III/4, and Athlon/Duron processors can use both buses simultaneously, eliminating a bottleneck there. The dual bus architecture enables the L2 cache of the newer processors to run at full speed inside the processor core on an independent bus, leaving the main CPU bus (FSB) to handle normal data flowing in and out of the chip. The two buses run at different speeds. The front-side bus or main CPU bus is coupled to the speed of the motherboard, whereas the back-side or L2 cache bus is coupled to the speed of the processor core. As the frequency of processors increases, so does the speed of the L2 cache.

The key to implementing DIB was to move the L2 cache memory off the motherboard and into the processor package. L1 cache always has been a direct part of the processor die, but L2 was larger and originally had to be external. By moving the L2 cache into the processor, the L2 cache could run at speeds more like the L1 cache, much faster than the motherboard or processor bus.

DIB also enables the system bus to perform multiple simultaneous transactions (instead of singular sequential transactions), accelerating the flow of information within the system and boosting performance. Overall, DIB architecture offers up to three times the bandwidth performance over a single-bus architecture processor.

Hyper-Threading Technology

Computers with two or more physical processors have long had a performance advantage over single-processor computers when the operating system supported multiple processors, as is the case with Windows NT 4.0, 2000, XP Professional, and Linux. However, dual-processor motherboards and systems have always been more expensive than otherwise-comparable single processor systems, and upgrading a dual-processor-capable system to dual-processor status can be difficult with only one processor because of the need to match processor speeds and specifications. However, Intel's Hyper-Threading (HT) Technology allows a single processor to handle two independent sets of instructions at the same time. In essence, HT Technology converts a single physical processor into two virtual processors.

Intel originally introduced HT Technology in its line of Xeon processors for servers in March 2002. HT Technology enables multiprocessor servers to act as if they had twice as many processors installed. HT Technology was introduced on Xeon workstation-class processors with a 533MHz system bus and later found its way into PC processors, with the Pentium 4 3.06GHz processor in November 2002. HT Technology is also present in all Pentium 4 processors with 800MHz CPU bus speed (2.4GHz up through 3.8GHz) as well as the Pentium 4 Extreme Edition and the dual-core Pentium Extreme Edition. However, the dual-core Pentium D does not include HT Technology.

How Hyper-Threading Works

Internally, an HT-enabled processor has two sets of general-purpose registers, control registers, and other architecture components, but both logical processors share the same cache, execution units, and buses. During operations, each logical processor handles a single thread (see Figure 3.2).

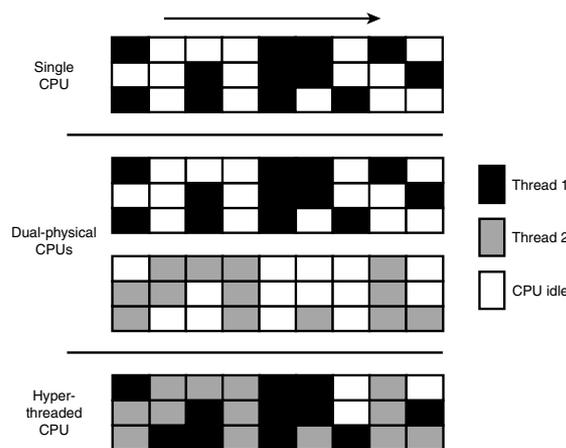


Figure 3.2 A processor with HT Technology enabled can fill otherwise-idle time with a second process, improving multitasking and performance of multithreading single applications.

Although the sharing of some processor components means that the overall speed of an HT-enabled system isn't as high as a true dual-processor system would be, speed increases of 25% or more are possible when multiple applications or a single multithreaded application is being run.

Hyper-Threading Requirements

The first HT-enabled processor was the Intel Pentium 4 3.06GHz. All 3.06GHz and faster Pentium 4 models support HT Technology, as do all processors 2.4GHz and faster that use the 800MHz bus. However, an HT-enabled P4 processor by itself can't bring the benefits of HT Technology to your system. You also need the following:

- *A compatible motherboard (chipset).* It might need a BIOS upgrade.
- *BIOS support to enable/disable HT Technology.* If your operating system doesn't support HT Technology, you should disable this feature. Application performance varies (some faster, some slower) when HT Technology is enabled. If this is a matter of concern, you should perform application-based benchmarks with HT Technology enabled and disabled to determine whether your application mix will benefit from using HT Technology.
- *A compatible operating system such as Windows XP.* When hyper-threading is enabled, the Device Manager shows two processors.

Intel's newer chipsets for the Pentium 4 support HT Technology; see the listing in Chapter 4 for details. However, if your motherboard or computer was released before HT Technology was introduced, you will need a BIOS upgrade from the motherboard or system vendor to be able to use HT Technology. Although Windows NT 4.0 and Windows 2000 are designed to use multiple physical processors, HT Technology requires specific operating system optimizations to work correctly. Linux distributions based on kernel 2.4.18 and higher also support HT Technology.

Dual-core Technology

HT Technology is designed to simulate two processors in a single physical unit. With properly written software, HT Technology can improve application performance. Unfortunately, many applications do not support HT Technology and slow down when HT Technology is enabled. However, applications do not need to be rewritten to take advantage of multiple processors or dual-core processors. A dual-core processor, as the name implies, contains two processor cores in a single processor package. A dual-core processor provides virtually all the advantages of a multiple-processor computer at a cost lower than two matched processors.

Both AMD and Intel introduced dual-core x86-compatible desktop processors in 2005. AMD's entry—the Athlon 64 X2—can be installed in most Socket 939 motherboards designed for the original single-core Athlon 64 or Athlon 64 FX processors. A BIOS upgrade might be necessary in some situations. AMD also introduced dual-core versions of the Opteron workstation and server processor in 2005. Intel's first dual-core processors—the Pentium Extreme Edition and the Pentium D—use the same Socket 775 as the most recent Pentium 4 models. However, they require new motherboards using the Intel 945 and 955 series chipsets or third-party chipsets that support dual-core operation.

▶▶ For more information, see "Dual-Core Processors," p. 203.

Processor Manufacturing

Processors are manufactured primarily from silicon, the second most common element on the planet (only the element oxygen is more common). Silicon is the primary ingredient in beach sand; however, in that form it isn't pure enough to be used in chips.

The manner in which silicon is formed into chips is a lengthy process that starts by growing pure silicon crystals via what is called the Czochralski method (named after the inventor of the process). In this method, electric arc furnaces transform the raw materials (primarily quartz rock that is mined) into metallurgical-grade silicon. Then to further weed out impurities, the silicon is converted to a liquid, distilled, and then redeposited in the form of semiconductor-grade rods, which are 99.999999% pure. These rods are then mechanically broken up into chunks and packed into quartz crucibles, which are loaded into electric crystal pulling ovens. There the silicon chunks are melted at more than 2,500° Fahrenheit. To prevent impurities, the ovens usually are mounted on very thick concrete cubes—often on a suspension to prevent any vibration, which would damage the crystal as it forms.

After the silicon is melted, a small seed crystal is inserted into the molten silicon and slowly rotated (see Figure 3.3). As the seed is pulled out of the molten silicon, some of the silicon sticks to the seed and hardens in the same crystal structure as the seed. By carefully controlling the pulling speed (10–40 millimeters per hour) and temperature (approximately 2,500°F), the crystal grows with a narrow neck that then widens into the full desired diameter. Depending on the chips being made, each ingot is 200mm (approximately 8") or 300mm (12") in diameter and more than 5 feet long, weighing hundreds of pounds.

The ingot is then ground into a perfect 200mm- (8") or 300mm-diameter (12") cylinder, with a small, flat cut on one side for positioning accuracy and handling. Each ingot is then cut with a high-precision diamond saw into more than a thousand circular wafers, each less than a millimeter thick (see Figure 3.4). Each wafer is then polished to a mirror-smooth surface.

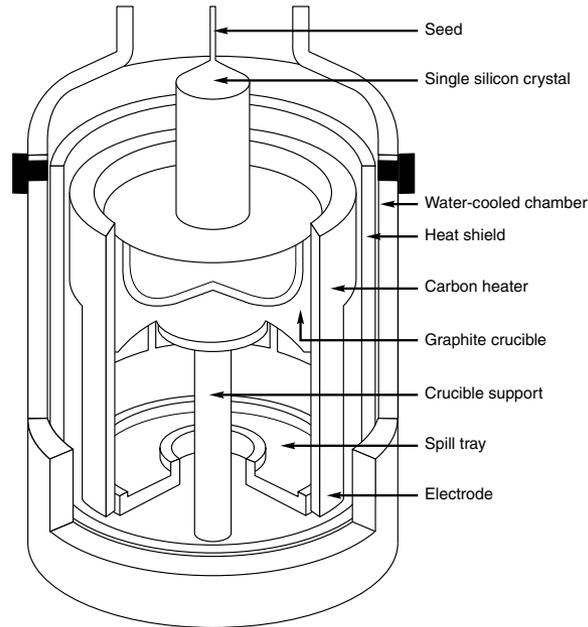


Figure 3.3 Growing a pure silicon ingot in a high-pressure, high-temperature oven.

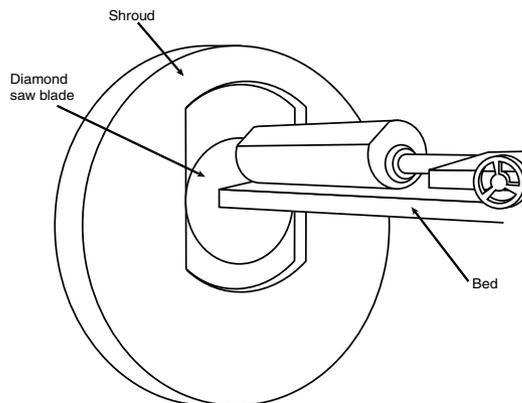


Figure 3.4 Slicing a silicon ingot into wafers with a diamond saw.

Chips are manufactured from the wafers using a process called *photolithography*. Through this photographic process, transistors and circuit and signal pathways are created in semiconductors by depositing different layers of various materials on the chip, one after the other. Where two specific circuits intersect, a transistor or switch can be formed.

The photolithographic process starts when an insulating layer of silicon dioxide is grown on the wafer through a vapor deposition process. Then a coating of photoresist material is applied, and an image of that layer of the chip is projected through a mask onto the now light-sensitive surface.

Doping is the term used to describe chemical impurities added to silicon (which is naturally a nonconductor), creating a material with semiconductor properties. The projector uses a specially created

mask, which is essentially a negative of that layer of the chip etched in chrome on a quartz plate. Modern processors have 20 or more layers of material deposited and partially etched away (each requiring a mask) and up to six or more layers of metal interconnects.

As the light passes through a mask, the light is focused on the wafer surface, exposing the photoresist with the image of that layer of the chip. Each individual chip image is called a *die*. A device called a *stepper* then moves the wafer over a little bit, and the same mask is used to imprint another chip die immediately next to the previous one. After the entire wafer is imprinted with a layer of material and photoresist, a caustic solution washes away the areas where the light struck the photoresist, leaving the mask imprints of the individual chip circuit elements and pathways. Then, another layer of semiconductor material is deposited on the wafer with more photoresist on top, and the next mask is used to expose and then etch the next layer of circuitry. Using this method, the layers and components of each chip are built one on top of the other until the chips are completed.

Some of the masks are used to add the *metallization* layers, which are the metal interconnects used to tie all the individual transistors and other components together. Most older chips use aluminum interconnects, although during 2002 many moved to copper. The first commercial PC processor chip to use copper was the 0.18-micron Athlon made in AMD's Dresden fab, and Intel shifted the Pentium 4 to copper with the 0.13-micron Northwood version (see Figure 3.5). Copper is a better conductor than aluminum and allows smaller interconnects with less resistance, meaning smaller and faster chips can be made. The reason copper hadn't been used until recently is that there were difficult corrosion problems to overcome during the manufacturing process that were not as much of a problem with aluminum. Now that these problems have been solved, more and more chips are fabricated with copper interconnects.

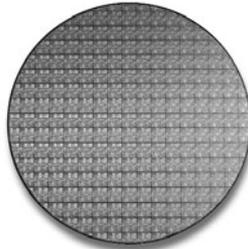


Figure 3.5 200mm wafer of the 0.13-micron Pentium 4 processors.

Note

The Pentium III and Celeron chips with the "coppermine" (codename for the 0.18-micron die used in those chips) die used aluminum and not copper metal interconnects as many people assume. In fact, the chip name had nothing to do with metal; the codename instead came from the Coppermine River in the Northwest Territory of Canada. Intel has long had a fondness for using codenames based on rivers (and sometimes, other geological features), especially those in the northwest region of the North American continent. For example, an older version of the Pentium III (0.25-micron die) was code-named Katmai, after an Alaskan river. Intel codenames read like the travel itinerary of a whitewater rafting enthusiast: Deerfield, Foster, Northwood, Tualatin, Gallatin, McKinley, and Madison are all rivers in Oregon, California, Alaska, Montana, and—in the case of Deerfield—Massachusetts and Vermont.

Another technology that is becoming common is the use of silicon on insulator (SOI) instead of CMOS technology. AMD uses SOI for its 90-nanometer (0.09-micron) processors, and it's expected that SOI, which provides better insulation than CMOS for transistors, will continue to grow in popularity.

A completed circular wafer has as many chips imprinted on it as can possibly fit. Because each chip usually is square or rectangular, there are some unused portions at the edges of the wafer, but every attempt is made to use every square millimeter of surface.

The industry is going through several transitions in chip manufacturing. The trend in the industry is to use both larger wafers and a smaller chip die process. *Process* refers to the size and spacing of the individual circuits and transistors on the chip. In late 2001 and into 2002, chip manufacturing processes began moving from the 0.18-micron to the 0.13-micron process, the metal interconnects on the die began moving from aluminum to copper, and wafers began moving from 200mm (8") to 300mm (12") in diameter. The larger 300mm wafers alone enable more than double the number of chips to be made, compared to the 200mm used previously. The smaller 0.13-micron and 0.09-micron (90-nanometer) processes enables more transistors to be incorporated into the die while maintaining a reasonable die size and allowing for a sufficient yield. This means the trend for incorporating more and more cache within the die will continue, and transistor counts will rise to 1 billion per chip or more by 2010.

As an example of how this can affect a particular chip, let's look at the original Pentium 4. The standard wafer size used in the industry for many years was 200mm in diameter, or just under 8". This results in a wafer of about 31,416 square millimeters in area. The first version of the Pentium 4 with the Willamette core used a 0.18-micron process with aluminum interconnects on a die that was 217 square millimeters in area, had 42 million transistors, and was made on 200mm wafers. Therefore, up to 145 of these chips could fit on a 200mm (8") wafer.

The Pentium 4 processors with the Northwood core that followed it use a smaller 0.13-micron process with copper interconnects on a die that is 131 square millimeters in area with 55 million transistors. Northwood has double the on-die L2 cache (512KB) as compared to Willamette, which is why the transistor count is significantly higher. Even with the higher transistor count, the smaller 0.13-micron process results in a die that is more than 60% smaller, allowing up to 240 chips to fit on the same 200mm (8") wafer that could hold only 145 Willamette die.

Starting in early 2002, Intel began producing Northwood on the larger 300mm wafers, which have a surface area of 70,686 square millimeters. These wafers have 2.25 times the surface area of the smaller 200mm wafers, enabling more than double the number of chips to be produced per wafer. In the case of the Pentium 4 Northwood, up to 540 chip dies fit on a 300mm wafer. By combining the smaller die with the larger wafer, Pentium 4 production has increased by more than 3.7 times since the chip was first introduced. This is one reason newer chips are often more plentiful and less expensive than older ones.

In 2004, the industry began moving to the 90-nanometer (0.09-micron) process, allowing even smaller and faster chips to be made. Most new chips in 2005 were based on the 0.09-micron process, and it's expected that this will continue throughout 2006.

In 2007, we'll likely see a move toward a 65-nanometer process, and we'll see a 45-nanometer process in 2010. These advancements in process will allow 1 billion transistors per chip in 2010! All these will still be made on 300mm wafers because the next wafer transition isn't expected until 2013, when a transition to 450mm wafers is being considered. Table 3.17 lists the CPU process transitions.

Table 3.17 Past, Current, and Future CPU Process Transitions

Date:	1989	1991	1993	1995	1997	1999	2001	2004	2007	2010	2013	2016
Process (micron):	1.0	0.8	0.5	0.35	0.25	0.18	0.13	0.09	0.065	0.045	0.032	0.022
Process (nm):	1000	800	500	350	250	180	130	90	65	45	32	22

Note that not all the chips on each wafer will be good, especially as a new production line starts. As the manufacturing process for a given chip or production line is perfected, more and more of the chips will be good. The ratio of good to bad chips on a wafer is called the *yield*. Yields well under 50% are common when a new chip starts production; however, by the end of a given chip's life, the yields are normally in the 90% range. Most chip manufacturers guard their yield figures and are very secretive

about them because knowledge of yield problems can give their competitors an edge. A low yield causes problems both in the cost per chip and in delivery delays to their customers. If a company has specific knowledge of competitors' improving yields, it can set prices or schedule production to get higher market share at a critical point.

After a wafer is complete, a special fixture tests each of the chips on the wafer and marks the bad ones to be separated out later. The chips are then cut from the wafer using either a high-powered laser or diamond saw.

After being cut from the wafers, the individual dies are then retested, packaged, and retested again. The packaging process is also referred to as *bonding* because the die is placed into a chip housing in which a special machine bonds fine gold wires between the die and the pins on the chip. The package is the container for the chip die, and it essentially seals it from the environment.

After the chips are bonded and packaged, final testing is done to determine both proper function and rated speed. Different chips in the same batch often run at different speeds. Special test fixtures run each chip at different pressures, temperatures, and speeds, looking for the point at which the chip stops working. At this point, the maximum successful speed is noted and the final chips are sorted into bins with those that tested at a similar speed. For example, the Pentium 4 2.0A, 2.2, 2.26, 2.4, and 2.53GHz are all exactly the same chip made using the same die. They were sorted at the end of the manufacturing cycle by speed.

One interesting thing about this is that as a manufacturer gains more experience and perfects a particular chip assembly line, the yield of the higher-speed versions goes way up. So, of all the chips produced from a single wafer, perhaps more than 75% of them check out at the highest speed and only 25% or less run at the lower speeds. The paradox is that Intel often sells a lot more of the lower-priced, lower-speed chips, so it just dips into the bin of faster ones, labels them as slower chips, and sells them that way. People began discovering that many of the lower-rated chips actually ran at speeds much higher than they were rated, and the business of overclocking was born.

Processor Re-marking

As people learned more about how processors are manufactured and graded, an interesting problem arose: Unscrupulous vendors began re-marking slower chips and reselling them as if they were faster. Often the price between the same chip at different speed grades can be substantial—in the hundreds of dollars—so by changing a few numbers on the chip, the potential profits can be huge. Because most of the Intel and AMD processors are produced with a generous safety margin—that is, they typically run well past their rated speeds—the re-marked chips would seem to work fine in most cases. Of course, in many cases they wouldn't work fine, and the system would end up crashing or locking up periodically.

At first, the re-marked chips were just a case of rubbing off the original numbers and restamping with new official-looking numbers. These were easy to detect, though. Re-markers then resorted to manufacturing completely new processor housings, especially for the plastic-encased Slot 1 and Slot A processors from Intel and AMD that were popular in the late '90s and still quite common just a few years ago. Although it might seem to be a huge bother to make a custom plastic case and swap it with the existing case, because the profits can be huge, criminals find it very lucrative. This type of re-marking is a form of organized crime and isn't just some kid in his basement with sandpaper and a rubber stamp.

Intel and AMD have seen fit to put a stop to some of the re-marking by building overclock protection in the form of a multiplier lock into most of their chips dating back nearly 10 years. This is usually done in the bonding or cartridge manufacturing process, where the chips are intentionally altered so they won't run at any speeds higher than they are rated. Usually this involves changing the bus frequency (BF) pins or traces on the chip, which control the internal multipliers the chip uses. At one point, many feared that fixing the clock multiplier would put an end to hobbyist overclocking, but that proved not to be the case. Enterprising individuals found ways to run their motherboards at bus speeds higher than normal, so even though the CPU generally won't allow a higher multiplier, you can still run it at a speed higher than it was designed for by ramping up the speed of the processor bus.

The real problem with the overclock protection as implemented by Intel and AMD is that the professional counterfeiter has often been able to figure out a way around it by modifying the chip physically. Today's socketed processors are much more immune to these re-marking attempts, but it is still possible, particularly because the evidence can be hidden under a heatsink. To protect yourself from purchasing a fraudulent chip, verify the specification numbers and serial numbers with Intel and AMD before you purchase. Also beware where you buy your hardware. Purchasing over online auction sites can be extremely dangerous because defrauding the purchaser is so easy. Also, traveling computer show/flea market arenas can be a hotbed of this type of activity. Finally, I recommend purchasing only "boxed" or retail-packaged versions of the Intel and AMD processors, rather than the raw OEM versions. The boxed versions are shrink-wrapped and contain a high-quality heatsink, documentation, and a 3-year warranty with the manufacturer.

Fraudulent computer components are not limited to processors. I have seen fake memory, fake mice, fake video cards, fake cache memory, counterfeit operating systems and applications, and even fake motherboards. The hardware that is faked usually works but is of inferior quality to the type it is purporting to be. For example, one of the most highly counterfeited pieces of hardware at one time was the Microsoft mouse. They originally sold for \$35 wholesale, yet I could purchase cheap mice from overseas manufacturers for as little as \$2.32 each. It didn't take somebody long to realize that if they made the \$2 mouse look like a \$35 Microsoft mouse, they could sell it for \$20 and people would think they were getting a genuine article for a bargain, while the thieves ran off with a substantial profit.

PGA Chip Packaging

Variations on the pin grid array (PGA) chip packaging have been the most commonly used chip packages over the years. They were used starting with the 286 processor in the 1980s and are still used today, although not in all CPU designs. PGA takes its name from the fact that the chip has a grid-like array of pins on the bottom of the package. PGA chips are inserted into sockets, which are often of a zero insertion force (ZIF) design. A ZIF socket has a lever to allow for easy installation and removal of the chip.

Most Pentium processors use a variation on the regular PGA called staggered pin grid array (SPGA), in which the pins are staggered on the underside of the chip rather than in standard rows and columns. This was done to move the pins closer together and decrease the overall size of the chip when a large number of pins is required. Figure 3.6 shows a Pentium Pro that uses the dual-pattern SPGA (on the right) next to an older Pentium 66 that uses the regular PGA. Note that the right half of the Pentium Pro shown here has additional pins staggered among the other rows and columns.

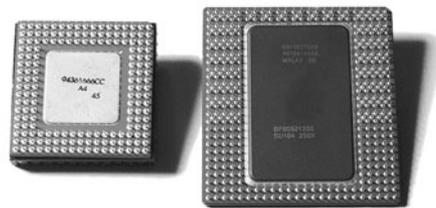


Figure 3.6 PGA on Pentium 66 (left) and dual-pattern SPGA on Pentium Pro (right).

Older PGA variations had the processor die mounted in a cavity underneath the substrate, with the top surface facing up if you turned the chip upside down. The die was then wire-bonded to the chip package with hundreds of tiny gold wires connecting the connections at the edge of the chip with the internal connections in the package. After the wire bonding, the cavity was sealed with a metal cover. This was an expensive and time-consuming method of producing chips, so cheaper and more efficient packaging methods were designed.

Most modern processors are built on a form of flip-chip pin grid array (FC-PGA) packaging. This type still plugs into a PGA socket, but the package itself is dramatically simplified. With FC-PGA, the raw silicon die is mounted face down on the top of the chip substrate, and instead of wire bonding, the connections are made with tiny solder bumps around the perimeter of the die. The edge is then sealed with a fillet of epoxy. With the original versions of FC-PGA, you could see the backside of the raw die sitting on the chip.

Unfortunately, there were some problems with attaching the heatsink to an FC-PGA chip. The heatsink sat on the top of the die, which acted as a pedestal. If you pressed down on one side of the heatsink excessively during the installation process (such as when you were attaching the clip), you risked cracking the silicon die and destroying the chip. This was especially a problem as heatsinks became larger and heavier and the force applied by the clip became greater.

AMD decreased the risk of damage by adding rubber spacers to each corner of the chip substrate for the Athlon XP, thus preventing the heatsink from tilting excessively during installation. Still, these bumpers could compress, and it was all too easy to crack the die.

Intel revised its packaging with a newer FC-PGA2 version used in later Pentium III and all Pentium 4 processors. This incorporates a protective metal cap, dubbed a heat spreader, to protect the CPU from damage when the heatsink is attached. Ironically, the first processor for PCs to use a heat spreader was actually made by AMD for its K6 family of processors.

The Athlon 64 processor family uses a heatsink design different from the Athlon XP. On the Athlon 64 family, the heatsink is attached to a clip. The clip is then screwed to the motherboard, which helps prevent damage to the processor. The Athlon 64, Opteron, and Socket 754 versions of the Sempron also use a heat spreader on top of the processor die, enabling larger and heavier heatsinks to be installed without any potential damage to the processor core.

Future packaging directions are headed toward what is called BBUL (bumpless build-up layer) packaging. This will embed the die completely in the package; in fact, the package layers will be built up around and on top of the die, fully encapsulating it within the package. This will embed the chip die and allow for a full flat surface for attaching the heatsink, as well as shorter internal interconnections within the package.

Single Edge Contact and Single Edge Processor Packaging

Intel and AMD used cartridge- or board-based packaging for some of their processors from 1997 through 2000. This packaging was called single edge contact cartridge (SECC) or single edge processor package (SEPP) and consisted of the CPU and optional separate L2 cache chips mounted on a circuit board that looked similar to an oversized memory module and that plugged into a slot. In some cases, the boards were covered with a plastic cartridge cover.

The SEC cartridge is an innovative—if a bit unwieldy—package design that incorporates the back-side bus and L2 cache internally. It was used as a cost-effective method for integrating L2 cache into the processor before it was feasible to include the cache directly inside the processor die.

A less expensive version of the SEC is called the single edge processor (SEP) package. The SEP package is basically the same circuit board containing processor and (optional) cache, but without the fancy plastic cover. This was used mainly by the lower-cost early Celeron processors. The SEP package plugs directly into the same Slot 1 connector used by the standard Pentium II or III. Four holes on the board enable the heatsink to be installed.

Slot 1, as shown in Figure 3.7, is the connection to the motherboard and has 242 pins. AMD used the same physical slot but rotated it 180° and called it Slot A. The SEC cartridge or SEP processor is plugged into the slot and secured with a processor-retention mechanism, which is a bracket that holds it in place. There also might be a retention mechanism or support for the processor heatsink. Figure 3.8 shows the parts of the cover that make up the SEC package. Note the large thermal plate used to aid in dissipating the heat from this processor. The SEP package is shown in Figure 3.9.

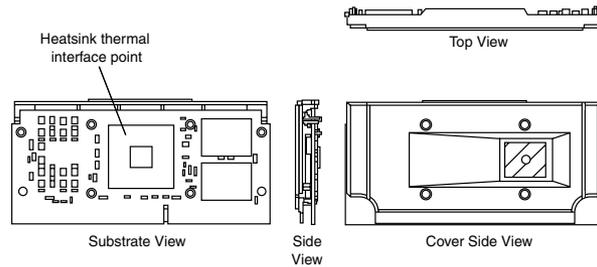


Figure 3.10 SECC2 packaging used in newer Pentium II and III processors.

Processor Socket and Slot Types

Intel and AMD have created a set of socket and slot designs for their processors. Each socket or slot is designed to support a different range of original and upgrade processors. Table 3.18 shows the designations for the various 486 and newer processor sockets/slots and lists the chips designed to plug into them.

Table 3.18 CPU Socket and Slot Types and Specifications

Chip Class	Socket	Pins	Layout	Voltage	Supported Processors	Introduced
Intel/AMD 486 class	Socket 1	169	17×17 PGA	5V	486 SX/SX2, DX/DX2, DX4 OD	Apr. '89
	Socket 2	238	19×19 PGA	5V	486 SX/SX2, DX/DX2, DX4 OD, 486 Pentium OD	Mar. '92
	Socket 3	237	19×19 PGA	5V/3.3V	486 SX/SX2, DX/DX2, DX4, 486 Pentium OD, AMD 5x86	Feb. '94
	Socket 6 ¹	235	19×19 PGA	3.3V	486 DX4, 486 Pentium OD	Feb. '94
Intel/AMD 586 (Pentium) class	Socket 4	273	21×21 PGA	5V	Pentium 60/66, OD	Mar. '93
	Socket 5	320	37×37 SPGA	3.3V/3.5V	Pentium 75-133, OD	Mar. '94
	Socket 7	321	37×37 SPGA	VRM	Pentium 75-233+, MMX, OD, AMD K5/K6, Cyrix M1/II	June '95
Intel 686 (Pentium II/III) class	Socket 8	387	Dual-pattern SPGA	Auto VRM	Pentium Pro, OD	Nov. '95
	Slot 1 (SC242)	242	Slot	Auto VRM	Pentium II/III, Celeron SECC	May '97
	Socket 370	370	37×37 SPGA	Auto VRM	Celeron/Pentium III PPGA/FC-PGA	Nov. '98
Intel Pentium 4 class	Socket 423	423	39×39 SPGA	Auto VRM	Pentium 4 FC-PGA	Nov. '00
	Socket 478	478	26×26 mPGA	Auto VRM	Pentium 4/Celeron FC-PGA2	Oct. '01
	Socket T (LGA775)	775	30×33 LGA	Auto VRM	Pentium 4/Celeron/Pentium D/Pentium Extreme Edition/LGA775	June '04
AMD K7 class	Slot A	242	Slot	Auto VRM	AMD Athlon SECC	June '99
	Socket A (462)	462	37×37 SPGA	Auto VRM	AMD Athlon/Athlon XP/Duron PGA/FC-PGA	June '00

Table 3.18 Continued

Chip Class	Socket	Pins	Layout	Voltage	Supported Processors	Introduced
AMD K8 class	Socket 754	754	29×29 mPGA	Auto VRM	AMD Athlon 64	Sep. '03
	Socket 939	939	31×31 mPGA	Auto VRM	AMD Athlon 64 v.2	June '04
	Socket 940	940	31×31 mPGA	Auto VRM	AMD Athlon 64FX, Opteron	Apr. '03
Intel/AMD server and workstation class	Slot 2 (SC330)	330	Slot	Auto VRM	Pentium II/III Xeon	Apr. '98
	Socket 603	603	31×25 mPGA	Auto VRM	Xeon (P4)	May '01
	Socket 604	604	31×25 mPGA	Auto VRM	Xeon (P4)	Oct. '03
	Socket PAC418	418	38×22	Auto VRM split SPGA	Itanium	May '01
	Socket PAC611	611	25×28	Auto VRM mPGA	Itanium 2	July '02
	Socket 940	940	31×31 mPGA	Auto VRM	AMD Athlon 64FX, Opteron	Apr. '03

1. Socket 6 was never actually implemented in any systems.

FC-PGA = Flip-chip pin grid array

FC-PGA2 = FC-PGA with an Integrated Heat Spreader (IHS)

OD = OverDrive (retail upgrade processors)

PAC = Pin array cartridge

PGA = Pin grid array

PPGA = Plastic pin grid array

SC242 = Slot connector, 242 pins

SC330 = Slot connector, 330 pins

SECC = Single edge contact cartridge

SPGA = Staggered pin grid array

mPGA = Micro pin grid array

VRM = Voltage regulator module with variable voltage output determined by module type or manual jumpers

Auto VRM = Voltage regulator module with automatic voltage selection determined by processor Voltage ID (VID) pins

Sockets 1, 2, 3, and 6 are 486 processor sockets and are shown together in Figure 3.11 so you can see the overall size comparisons and pin arrangements between these sockets. Sockets 4, 5, 7, and 8 are Pentium and Pentium Pro processor sockets and are shown together in Figure 3.12 so you can see the overall size comparisons and pin arrangements between these sockets. More detailed drawings of each socket are included throughout the remainder of this section with thorough descriptions of the sockets.

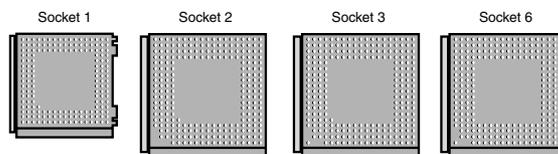


Figure 3.11 486 processor sockets.

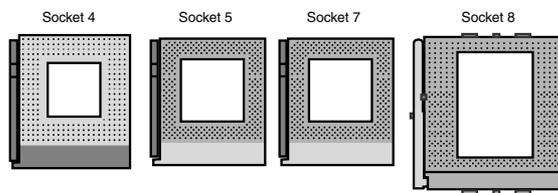


Figure 3.12 Pentium and Pentium Pro processor sockets.

Zero Insertion Force

When the Socket 1 specification was created, manufacturers realized that if users were going to upgrade processors, they had to make the process easier. The socket manufacturers found that 100 lbs. of insertion force is required to install a chip in a standard 169-pin screw Socket 1 motherboard. With this much force involved, you easily could damage either the chip or the socket during removal or reinstallation. Because of this, some motherboard manufacturers began using low insertion force (LIF) sockets, which required only 60 lbs. of insertion force for a 169-pin chip. With the LIF or standard socket, I usually advise removing the motherboard—that way you can support the board from behind when you insert the chip. Pressing down on the motherboard with 60–100 lbs. of force can crack the board if it is not supported properly. A special tool is also required to remove a chip from one of these sockets. As you can imagine, even the low insertion force was relative, and a better solution was needed if the average person was ever going to replace his CPU.

Manufacturers began using ZIF sockets in Socket 1 designs, and all processor sockets from Socket 2 and higher have been of the ZIF design. ZIF is required for all the higher-density sockets because the insertion force would simply be too great otherwise. ZIF sockets almost eliminate the risk involved in installing or removing a processor because no insertion force is necessary to install the chip and no tool is needed to extract one. Most ZIF sockets are handle-actuated: You lift the handle, drop the chip into the socket, and then close the handle. This design makes installing or removing a processor an easy task.

Socket 1

The original OverDrive socket, now officially called Socket 1, is a 169-pin PGA socket. Motherboards that have this socket can support any of the 486SX, DX, and DX2 processors and the DX2/OverDrive versions. This type of socket is found on most 486 systems that originally were designed for OverDrive upgrades. Figure 3.13 shows the pinout of Socket 1.

The original DX processor draws a maximum 0.9 amps of 5V power in 33MHz form (4.5 watts) and a maximum 1 amp in 50MHz form (5 watts). The DX2 processor, or OverDrive processor, draws a maximum 1.2 amps at 66MHz (6 watts). This minor increase in power requires only a passive heatsink consisting of aluminum fins that are glued to the processor with thermal transfer epoxy. Passive heatsinks don't have any mechanical components like fans. Heatsinks with fans or other devices that use power are called *active* heatsinks. OverDrive processors rated at 40MHz or less do not have heatsinks.

Socket 2

When the DX2 processor was released, Intel was already working on the new Pentium processor. The company wanted to offer a 32-bit, scaled-down version of the Pentium as an upgrade for systems that originally came with a DX2 processor. Rather than just increasing the clock rate, Intel created an all-new chip with enhanced capabilities derived from the Pentium.

The chip, called the Pentium OverDrive processor, plugs into a processor socket with the Socket 2 or Socket 3 design. These sockets hold any 486 SX, DX, or DX2 processor, as well as the Pentium OverDrive. Because this chip is essentially a 32-bit version of the (normally 64-bit) Pentium chip, many have taken to calling it a Pentium-SX. It was available in 25/63MHz and 33/83MHz versions. The first number indicates the base motherboard speed; the second number indicates the actual operating speed of the Pentium OverDrive chip. As you can see, it is a clock-multiplied chip that runs at 2.5 times the motherboard speed. Figure 3.14 shows the pinout configuration of the official Socket 2 design.

Notice that although the chip for Socket 2 is called Pentium OverDrive, it is not a full-scale (64-bit) Pentium. Intel released the design of Socket 2 a little prematurely and found that the chip ran too hot for many systems. The company solved this problem by adding a special active heatsink to the Pentium OverDrive processor. This active heatsink is a combination of a standard heatsink and a built-in electric fan. Unlike the aftermarket glue-on or clip-on fans for processors that you might have seen, this one actually draws 5V power directly from the socket to drive the fan. No external connection to disk drive cables or the power supply is required. The fan/heatsink assembly clips and plugs directly into the processor and provides for easy replacement if the fan fails.

Caution

Because this jumper must be manually set, a user could install a 3.3V processor in this socket when it is configured for 5V operation. This installation instantly destroys the chip when the system is powered on. So, it is up to the end user to ensure that this socket is properly configured for voltage, depending on which type of processor is installed. If the jumper is set in 3.3V configuration and a 5V processor is installed, no harm will occur, but the system will not operate properly unless the jumper is reset for 5V.

Socket 4

Socket 4 is a 273-pin socket designed for the original Pentium processors. The original Pentium 60MHz and 66MHz version processors had 273 pins and plugged into Socket 4. It is a 5V-only socket because all the original Pentium processors run on 5V. This socket accepts the original Pentium 60MHz or 66MHz processor and the OverDrive processor. Figure 3.16 shows the pinout specification of Socket 4.

Somewhat amazingly, the original Pentium 66MHz processor consumes up to 3.2 amps of 5V power (16 watts), not including power for a standard active heatsink (fan). The 66MHz OverDrive processor that replaced it consumes a maximum 2.7 amps (13.5 watts), including about 1 watt to drive the fan. Even the original 60MHz Pentium processor consumes up to 2.91 amps at 5V (14.55 watts). It might seem strange that the replacement processor, which is twice as fast, consumes less power than the original, but this has to do with the manufacturing processes used for the original and OverDrive processors.

Although both processors run on 5V, the original Pentium processor was created with a circuit size of 0.8 micron, making that processor much more power-hungry than the 0.6-micron circuits used in the OverDrive and the other Pentium processors. Shrinking the circuit size is one of the best ways to decrease power consumption. Although the OverDrive processor for Pentium-based systems draws less power than the original processor, additional clearance might have to be allowed for the active heatsink assembly that is mounted on top. As in other OverDrive processors with built-in fans, the power to run the fan is drawn directly from the chip socket, so no separate power-supply connection is required. Also, the fan is easy to replace should it ever fail.

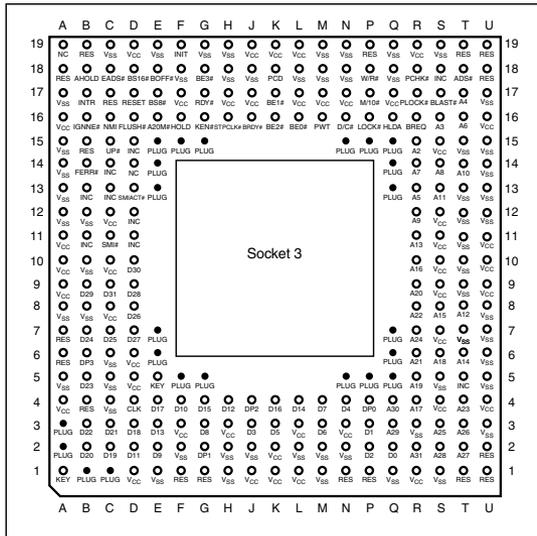


Figure 3.15 273-pin Intel Socket 3 configuration.

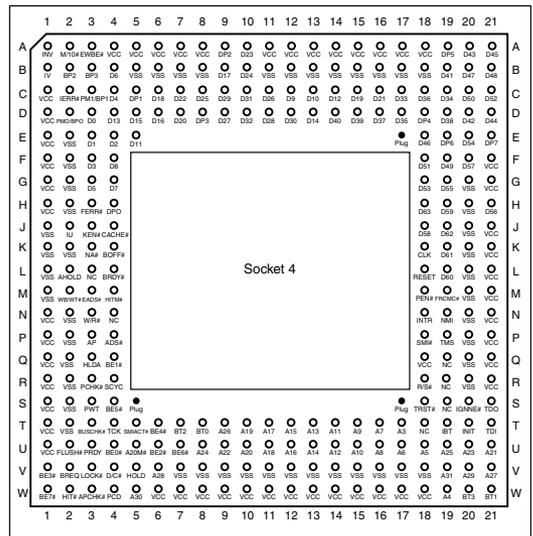


Figure 3.16 273-pin Intel Socket 4 configuration.

Socket 5

When Intel redesigned the Pentium processor to run at 75MHz, 90MHz, and 100MHz, the company went to a 0.6-micron manufacturing process and 3.3V operation. This change resulted in lower power consumption: only 3.25 amps at 3.3V (10.725 watts). Therefore, the 100MHz Pentium processor used far less power than even the original 60MHz version. This resulted in lower power consumption and enabled the extremely high clock rates without overheating.

The Pentium 75 and higher processors actually have 296 pins, although they plug into the official Intel Socket 5 design, which calls for a total of 320 pins. The additional pins are used by the Pentium OverDrive for Pentium processors. This socket has the 320 pins configured in a staggered PGA, in which the individual pins are staggered for tighter clearance.

Several OverDrive processors for existing Pentiums were available. These usually were later design chips with integral voltage regulators to enable operating on the higher voltages the older chips originally required. Intel no longer sells these; however, companies such as PowerLeap do still sell upgrade chips for older systems. Figure 3.17 shows the standard pinout for Socket 5.

The Pentium OverDrive for Pentium processors has an active heatsink (fan) assembly that draws power directly from the chip socket. The chip requires a maximum 4.33 amps of 3.3V to run the chip (14.289 watts) and 0.2 amp of 5V power to run the fan (one watt), which results in a total power consumption of 15.289 watts. This is less power than the original 66MHz Pentium processor requires, yet it runs a chip that is as much as four times faster!

Socket 6

The last 486 socket was designed for the 486 DX4 and the 486 Pentium OverDrive processor. Socket 6 was intended as a slightly redesigned version of Socket 3 and had an additional 2 pins plugged for proper chip keying. Socket 6 has 235 pins and accepts only 3.3V 486 or OverDrive processors. Although Intel went to the trouble of designing this socket, it never was built or implemented in any systems. Motherboard manufacturers instead stuck with Socket 3.

Socket 7 (and Super7)

Socket 7 is essentially the same as Socket 5 with one additional key pin in the opposite inside corner of the existing key pin. Socket 7, therefore, has 321 pins total in a 37×37 SPGA arrangement. The real difference with Socket 7 is not with the socket itself, but with the companion voltage regulator module (VRM) circuitry on the motherboard that must accompany it.

The VRM is either a small circuit board or a group of circuitry embedded in the motherboard that supplies the proper voltage level and regulation of power to the processor.

The main reason for the VRM is that Intel and AMD wanted to drop the voltages the processors would use from the 3.3V or 5V supplied to the motherboard by the power supply. Rather than require custom power supplies for different processors, the VRM converts the 3.3V or 5V to the proper voltage for the particular CPU you are using. Intel released different versions of the Pentium and Pentium-MMX processors that ran on 3.3V (called VR), 3.465V (called VRE), or 2.8V. Equivalent processors from AMD, Cyrix, and others used voltages from 3.3V to 1.8V. Because of the variety of voltages that might be required to support different processors, most motherboard manufacturers started including VRM sockets or building adaptable VRMs into their Pentium motherboards.

Figure 3.18 shows the Socket 7 pinout.

AMD, along with Cyrix and several chipset manufacturers, pioneered an improvement or extension to the Intel Socket 7 design called Super Socket 7 (or Super7), taking it from 66MHz to 95MHz and 100MHz. This enabled faster Socket 7–type systems to be made, supporting processors up to 500MHz, which are nearly as fast as some of the newer Slot 1– and Socket 370–type systems using Intel processors. Super7 systems also have support for the AGP video bus, as well as Ultra DMA hard disk controllers and advanced power management.

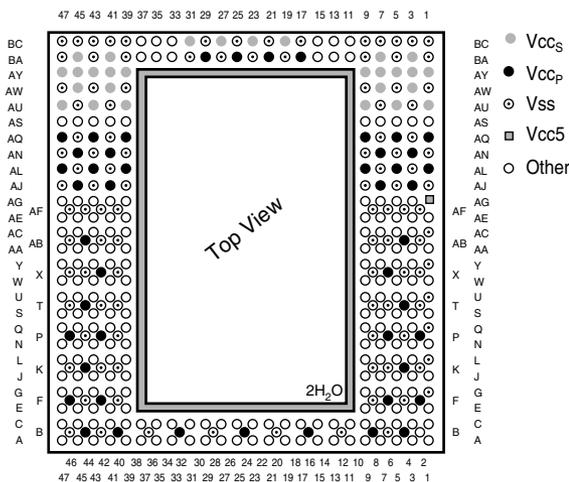


Figure 3.19 Socket 8 (Pentium Pro) pinout showing power pin locations.

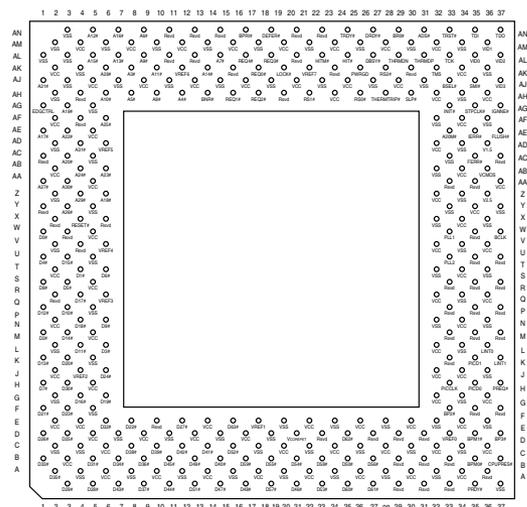


Figure 3.20 Socket 370 (PGA-370) Pentium III/Celeron pinout (top view).

The Celeron was gradually shifted over to PGA-370, although for a time both were available. All Celeron processors at 333MHz and lower were available only in the Slot 1 version. Celeron processors from 366MHz to 433MHz were available in both Slot 1 and Socket 370 versions; all Celeron processors from 466MHz and up through 1.4GHz are available only in the Socket 370 version.

Starting in October 1999, Intel also introduced Pentium III processors with integrated cache that plug into Socket 370. These use a packaging called flip chip pin grid array (FC-PGA), in which the raw die is mounted on the substrate upside down. The slot version of the Pentium III was more expensive and no longer necessary because of the on-die L2 cache.

Note that because of some voltage changes and one pin change, many original Socket 370 motherboards do not accept the later FC-PGA Socket 370 versions of the Pentium III and Celeron. Pentium III processors in the FC-PGA form have two RESET pins and require VRM 8.4 specifications. Prior motherboards designed only for the older versions of the Celeron are referred to as *legacy motherboards*, and the newer motherboards supporting the second RESET pin and VRM 8.4 specification are referred to as *flexible motherboards*. Contact your motherboard or system manufacturer for information to see whether your socket is the flexible version. Some motherboards, such as the Intel CA810, do support the VRM 8.4 specifications and supply proper voltage, but without Vtt support the Pentium III processor in the FC-PGA package will be held in RESET#. The last versions of the Pentium III and Celeron III use the Tualatin core design, which also requires a revised socket to operate. Motherboards that can handle Tualatin-core processors are known as *Tualatin-ready* and use different chipsets from those not designed to work with the Tualatin-core processor. Companies that sell upgrade processors offer products that enable you to install a Tualatin-core Pentium III or Celeron III processor into a motherboard that lacks built-in Tualatin support.

Installing a Pentium III processor in the FC-PGA package into an older motherboard is unlikely to damage the motherboard. However, the processor itself could be damaged. Pentium III processors in the 0.18-micron process operate at either 1.60V or 1.65V, whereas the Intel Celeron processors operate at 2.00V. The motherboard could be damaged if the motherboard BIOS fails to recognize the voltage identification of the processor. Contact your PC or motherboard manufacturer before installation to ensure compatibility.

A motherboard with a Slot 1 can be designed to accept almost any Celeron, Pentium II, or Pentium III processor. To use the socketed Celerons and Pentium III processors, several manufacturers have made available a low-cost slot-to-socket adapter sometimes called a *slot-ket*. This is essentially a Slot 1 board

containing only a Socket 370, which enables you to use a PGA processor in any Slot 1 board. A typical slot-let adapter is shown in the “Celeron” section later in this chapter.

►► See “Celeron,” p. 148.

Socket 423

Socket 423 is a ZIF-type socket introduced in November 2000 for the original Pentium 4. Figure 3.21 shows Socket 423.

Socket 423 supports a 400MHz processor bus, which connects the processor to the Memory Controller Hub (MCH), which is the main part of the motherboard chipset and similar to the North Bridge in earlier chipsets. Pentium 4 processors up to 2GHz were available for Socket 423; all faster versions require Socket 478 instead.

Socket 423 uses a unique heatsink mounting method that requires standoffs attached either to the chassis or to a special plate that mounts underneath the motherboard. This was designed to support the weight of the larger heatsinks required for the Pentium 4. Because of this, many Socket 423 motherboards require a special chassis that has the necessary additional standoffs installed. Fortunately, the need for these standoffs was eliminated with the newer Socket 478 for Pentium 4 processors.

The processor uses five voltage ID (VID) pins to signal the VRM built into the motherboard to deliver the correct voltage for the particular CPU you install. This makes the voltage selection completely automatic and foolproof. Most Pentium 4 processors for Socket 423 require 1.7V. A small triangular mark indicates the pin-1 corner for proper orientation of the chip.

Socket 478

Socket 478 is a ZIF-type socket for the Pentium 4 and Celeron 4 (Celerons based on the Pentium 4 core) introduced in October 2001. It was specially designed to support additional pins for future Pentium 4 processors and speeds over 2GHz. The heatsink mounting is different from the previous Socket 423, allowing larger heatsinks to be attached to the CPU. Figure 3.22 shows Socket 478.

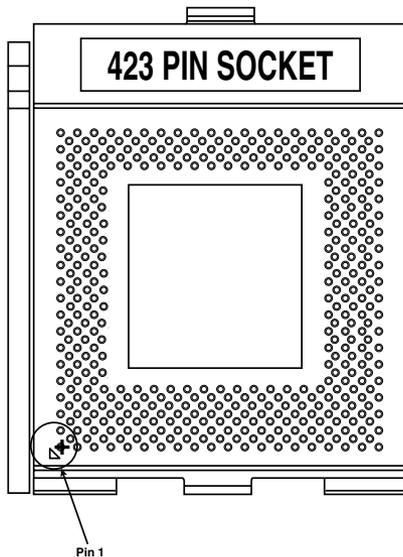


Figure 3.21 Socket 423 (Pentium 4) showing pin 1 location.

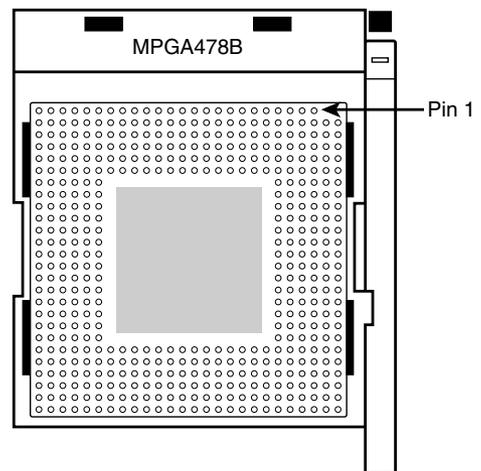


Figure 3.22 Socket 478 (Pentium 4) showing pin 1 location.

Socket 478 supports a 400MHz, 533MHz, or 800MHz processor bus that connects the processor to the memory controller hub (MCH), which is the main part of the motherboard chipset.

Socket 478 uses a heatsink attachment method that clips the heatsink directly to the motherboard, and not the CPU socket or chassis (as with Socket 423). Therefore, any standard chassis can be used, and the special standoffs used by Socket 423 boards are not required. This heatsink attachment allows for a much greater clamping load between the heatsink and processor, which aids cooling.

Socket 478 processors use five VID pins to signal the VRM built into the motherboard to deliver the correct voltage for the particular CPU you install. This makes the voltage selection completely automatic and foolproof. A small triangular mark indicates the pin-1 corner for proper orientation of the chip.

Socket A (Socket 462)

AMD introduced Socket A, also called Socket 462, in June 2000 to support the PGA versions of the Athlon and Duron processors. It is designed as a replacement for Slot A used by the original Athlon processor. Because the Athlon has now moved to incorporate L2 cache on-die, and the low-cost Duron was manufactured only in an on-die cache version, there was no longer a need for the expensive cartridge packaging the original Athlon processors used.

Socket A has 462 pins and 11 plugs oriented in an SPGA form (see Figure 3.23). Socket A has the same physical dimensions and layout as Socket 370; however, the location and placement of the plugs prevent Socket 370 processors from being inserted. Socket A supports 31 voltage levels from 1.100V to 1.850V in 0.025V increments, controlled by the VID0-VID4 pins on the processor. The automatic voltage regulator module circuitry typically is embedded on the motherboard.

There are 11 total plugged holes, including 2 of the outside pin holes at A1 and AN1. These are used to allow for keying to force the proper orientation of the processor in the socket. The pinout of Socket A is shown in Figure 3.24.

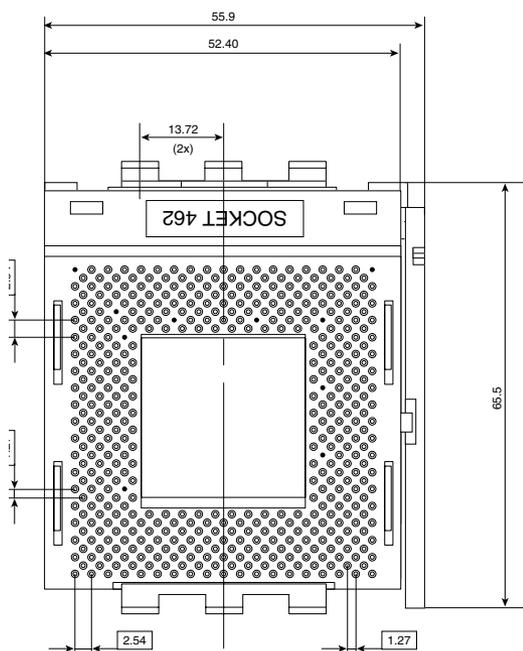


Figure 3.23 Socket A (Socket 462) Athlon/Duron layout.

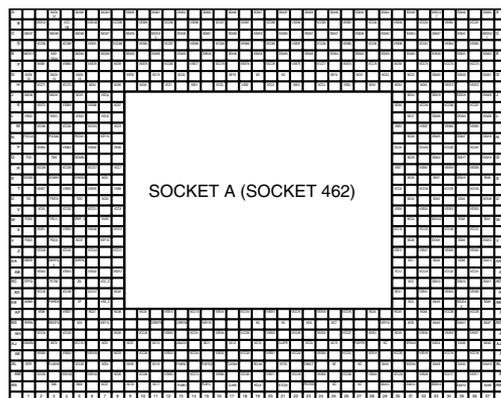


Figure 3.24 Socket A (Socket 462) Athlon/Duron pinout (top view).

After the introduction of Socket A, AMD moved all Athlon (including all Athlon XP) processors to this form factor, phasing out Slot A. In addition, for a time AMD also sold a reduced L2 cache version of the Athlon called the Duron in this form factor. In 2005, AMD discontinued the Athlon XP and introduced the AMD Sempron in both Socket A and Socket 754 form factors. The first Athlon 64 processors also used Socket 754, but most current Athlon 64 processors now use Socket 939.

Caution

Just because a chip can plug into a socket doesn't mean it will work. The newer Athlon XP and Socket A Sempron processors require different voltages, BIOS, and chipset support than earlier Socket A Athlon and Duron processors. As always, make sure your motherboard supports the processor you intend to install.

Socket 603

Socket 603 is used with the Intel Xeon processor in DP (dual processor) and MP (multiple processor) configurations. These are typically used in motherboards designed for use in network file servers. Figure 3.25 shows Socket 603.

Socket 754

Socket 754 is used with the initial releases of the AMD Athlon 64 processors. Socket 754 is also used by some versions of the AMD Sempron, AMD's economy processor line. This socket supports single-channel unbuffered DDR SDRAM. Figure 3.26 shows an overhead view of this socket.

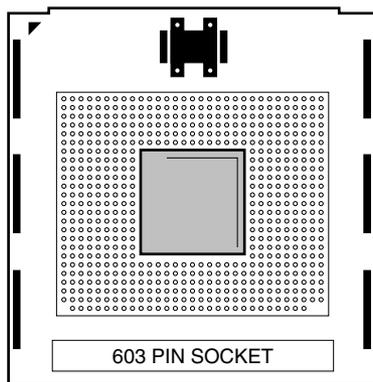


Figure 3.25 Socket 603 is used by the Intel Xeon processor.

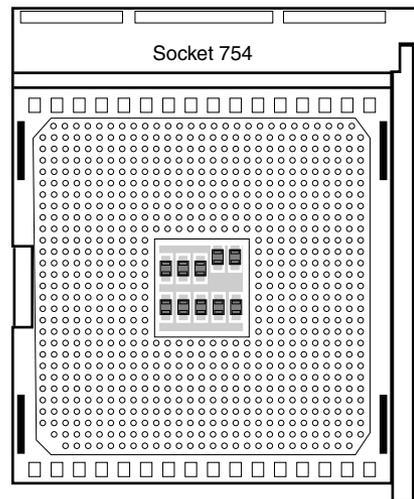


Figure 3.26 Socket 754. The large cutout corner at the lower left indicates pin 1.

Socket 939 and 940

Socket 939 is used with the Socket 939 versions of the AMD Athlon 64, 64 FX, and 64 X2 (see Figure 3.27). It's also used by some recent versions of the AMD Opteron processor for workstations and servers. Motherboards using this socket support conventional unbuffered DDR SDRAM modules in either single- or dual-channel mode, rather than the server-oriented (more expensive) registered modules required by Socket 940 motherboards. Sockets 939 and 940 have different pin arrangements and processors for each and are not interchangeable.

Socket 940 is used with the Socket 940 version of the AMD Athlon 64 FX, as well as most AMD Opteron processors (see Figure 3.28). Motherboards using this socket support only registered DDR SDRAM modules in dual-channel mode. Because the pin arrangement is different, Socket 939 processors do not work in Socket 940, and vice versa.

Socket T

Socket T (LGA775) is used by the latest versions of the Intel Pentium 4 Prescott processor and the Pentium D and Pentium Extreme Edition processors, as well as some versions of the Celeron D. The first-generation Prescott processors used Socket 478. Socket T is unique in that it uses a land grid array format, so the pins are on the socket, rather than the processor. The first LGA processors were the Pentium II and Celeron processors in 1997; in those processors LGA packaging was used for the chip mounted on the Slot-1 cartridge.

LGA uses gold pads (called *lands*) on the bottom of the substrate to replace the pins used in PGA packages. In socketed form, it allows for much greater clamping forces and therefore greater stability and improved thermal transfer (better cooling). LGA is really just a recycled version of what was previously called LCC (leadless chip carrier) packaging. This was used way back on the 286 processor in '84, which had gold lands around the edge only (there were far fewer pins back then). In other ways LGA is simply a modified version of ball grid array (BGA), with gold lands replacing the solder balls, making it more suitable for socketed (rather than soldered) applications. The early LCC packages were ceramic, whereas the first Pentium II LGA packages were plastic, with the package soldered to a cartridge substrate. These days (and for the future) the LGA package is organic and directly socketed instead. On a technical level, the Pentium 4 LGA chips combine several packaging technologies that have all been used in the past, including organic land grid array (OLGA) for the substrate and controlled collapse chip connection (C4) flip-chip for the actual processor die (see Figure 3.29).

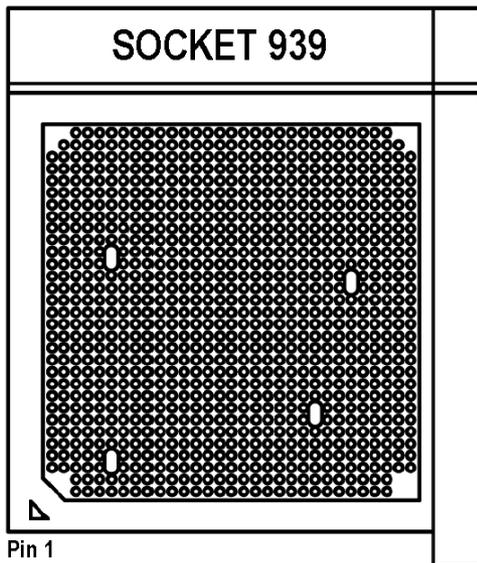


Figure 3.27 Socket 939. The cutout corner and triangle at the lower left indicate pin 1.

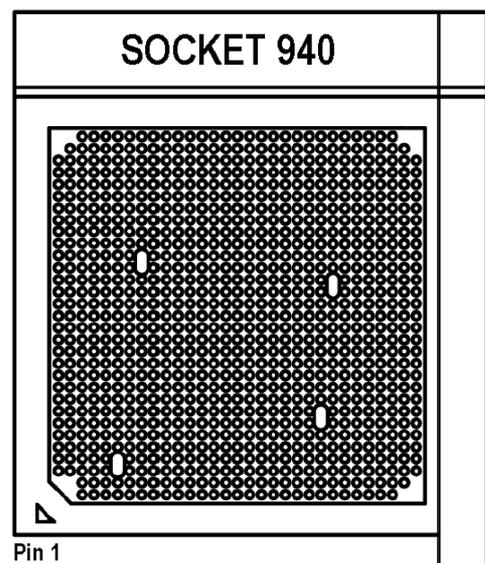


Figure 3.28 Socket 940. The cutout corner and triangle at the lower left indicate pin 1.

Socket M2

In the second quarter of 2006, AMD introduced processors that use a new socket, called Socket M2 (see Figure 3.30). AMD intends for M2 to be the eventual replacement for the confusing array of Socket 754, Socket 939, and Socket 940 form factors it uses for the Athlon 64, Athlon 64 FX, Athlon 64 X2, Opteron, and Socket 754 AMD Sempron processors.

Although Socket M2 contains 940 pins—the same number as used by Socket 940—Socket M2 is designed to support the integrated dual-channel DDR2 memory controllers that were added to the Athlon 64 and Opteron processor families in 2006. Processors designed for Sockets 754, 939, and 940 include DDR memory controllers and are not pin compatible with Socket M2.

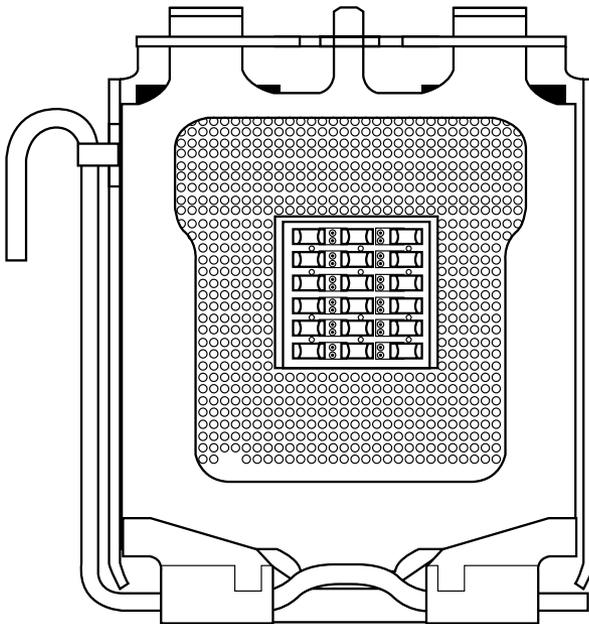


Figure 3.29 Socket T. The release lever on the left is used to raise the clamp out of the way to permit the processor to be placed over the contacts.

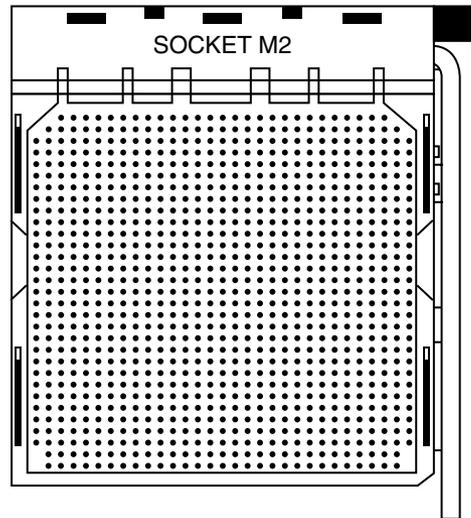


Figure 3.30 Socket M2. The cutout corner at the lower left indicates pin 1.

Processor Slots

After introducing the Pentium Pro with its integrated L2 cache, Intel discovered that the physical package it chose was very costly to produce. Intel was looking for a way to easily integrate cache and possibly other components into a processor package, and it came up with a cartridge or board design as the best way to do this. To accept its new cartridges, Intel designed two types of slots that could be used on motherboards.

Slot 1 is a 242-pin slot designed to accept Pentium II, Pentium III, and most Celeron processors. Slot 2, on the other hand, is a more sophisticated 330-pin slot designed for the Pentium II Xeon and Pentium III Xeon processors, which are primarily for workstations and servers. Besides the extra pins,

the biggest difference between Slot 1 and Slot 2 is the fact that Slot 2 was designed to host up to four-way or more processing in a single board. Slot 1 allows only single or dual processing functionality.

Note that Slot 2 is also called SC330, which stands for slot connector with 330 pins. Intel later discovered less-expensive ways to integrate L2 cache into the processor core and no longer produces Slot 1 or Slot 2 processors. Both Slot 1 and Slot 2 processors are now obsolete, and many systems using these processors have been retired or upgraded with socket-based motherboards.

Slot 1 (SC242)

Slot 1, also called SC242 (slot connector 242 pins), is used by the SEC design that is used with the cartridge-type Pentium II/III and Celeron processors (see Figure 3.31).

◀◀ See "Single Edge Contact and Single Edge Processor Packaging," p. 85.

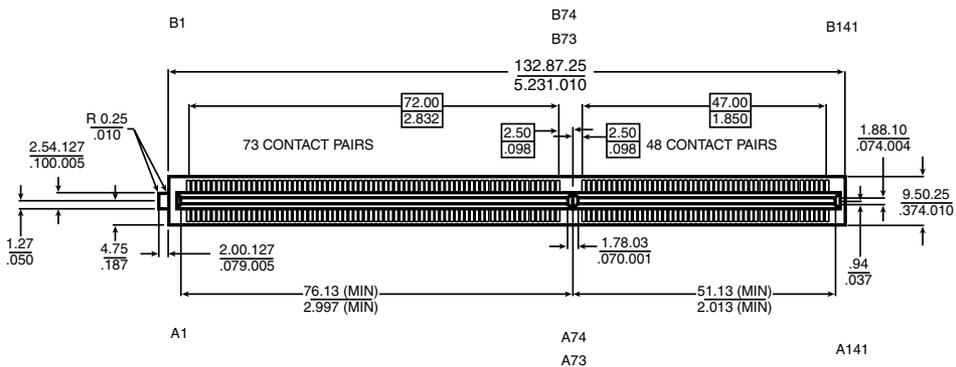


Figure 3.31 Slot 1 connector dimensions and pin layout.

Slot 2 (SC330)

Slot 2, otherwise called SC330 (slot connector 330 pins), is used on high-end motherboards that support the Pentium II Xeon and Pentium III Xeon processors. Figure 3.32 shows the Slot 2 connector.

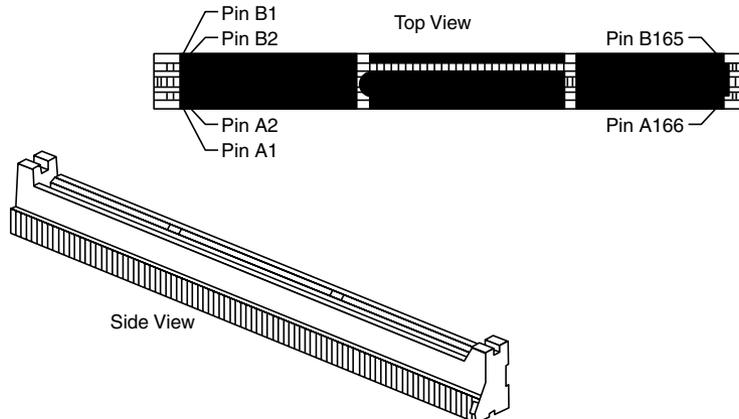


Figure 3.32 Slot 2 (SC330) connector dimensions and pin layout.

The Pentium II Xeon and Pentium III Xeon processors are designed in a cartridge similar to, but larger than, that used for the standard Pentium II/III. Figure 3.33 shows the Xeon cartridge.

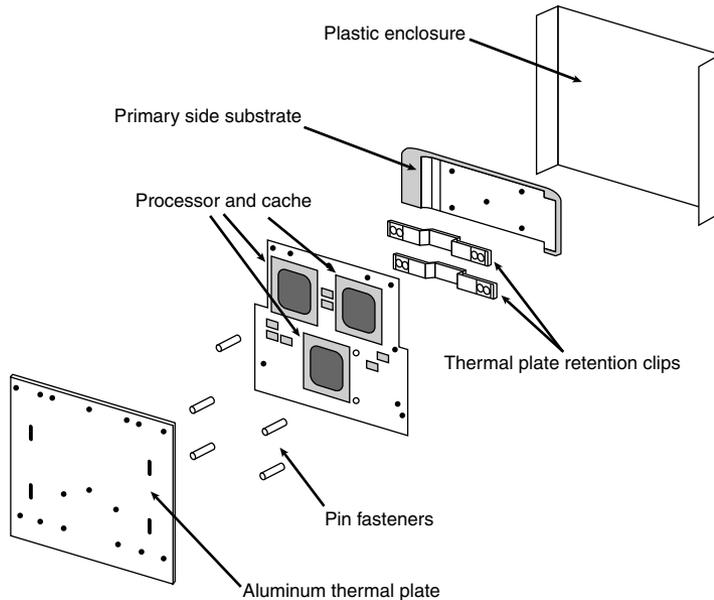


Figure 3.33 Pentium II/III Xeon cartridge.

Slot 2 motherboards were used in higher-end systems such as workstations or servers based on the Pentium II Xeon or Pentium III Xeon. These versions of the Xeon differ from the standard Pentium II and slot-based Pentium III mainly by virtue of having full-core speed L2 cache, and in some versions more of it. The additional pins allow for additional signals needed by multiple processors.

CPU Operating Voltages

One trend that is clear to anybody who has been following processor design is that the operating voltages have gotten lower and lower. The benefits of lower voltage are threefold. The most obvious is that with lower voltage comes lower overall power consumption. By consuming less power, the system is less expensive to run, but more importantly for portable or mobile systems, it runs much longer on existing battery technology. The emphasis on battery operation has driven many of the advances in lowering processor voltage because this has a great effect on battery life.

The second major benefit is that with less voltage and therefore less power consumption, less heat is produced. Processors that run cooler can be packed into systems more tightly and last longer.

The third major benefit is that a processor running cooler on less power can be made to run faster. Lowering the voltage has been one of the key factors in enabling the clock rates of processors to go higher and higher. This is because the lower the voltage, the shorter the time needed to change a signal from low to high.

Until the release of the mobile Pentium and both desktop and mobile Pentium MMX, most processors used a single voltage level to power both the core as well as run the input/output circuits. Originally, most processors ran both the core and I/O circuits at 5V, which was later reduced to 3.5V or 3.3V to lower power consumption. When a single voltage is used for both the internal processor core power as well as the external processor bus and I/O signals, the processor is said to have a single or unified power plane design.

When originally designing a version of the Pentium processor for mobile or portable computers, Intel came up with a scheme to dramatically reduce the power consumption while still remaining compatible with the existing 3.3V chipsets, bus logic, memory, and other components. The result was a dual-plane or split-plane power design in which the processor core ran off a lower voltage while the I/O circuits remained at 3.3V. This originally was called voltage reduction technology (VRT) and first debuted in the Mobile Pentium processors released in 1996. Later, this dual-plane power design also

appeared in desktop processors such as the Pentium MMX, which used 2.8V to power the core and 3.3V for the I/O circuits. Now most recent processors, whether for mobile or desktop use, feature a dual-plane power design. Some of the more recent Mobile Pentium II processors run on as little as 1.6V for the core while still maintaining compatibility with 3.3V components for I/O.

Knowing the processor voltage requirements is not a big issue with Socket 8, Socket 370, Socket 478, Socket A, Socket 604, Socket 754, Socket 940, Pentium Pro (Socket 8), or Pentium II (Slot 1 or Slot 2) processors because these sockets and slots have special VID pins the processor uses to signal to the motherboard the exact voltage requirements. This enables the voltage regulators built into the motherboard to be automatically set to the correct voltage levels by merely installing the processor.

Unfortunately, this automatic voltage setting feature was not available on Super7, Socket 7, and earlier motherboard and processor designs. Therefore, you usually must set jumpers or otherwise configure the motherboard according to the voltage requirements of the processor you are installing. Pentium (Socket 4, 5, or 7) processors have run on a number of voltages, but the most recent MMX versions all use 2.8V—except for mobile Pentium processors, which are as low as 1.8V. Table 3.19 lists the voltage settings used by Intel Pentium (non-MMX) Socket 7 processors that use a single power plane and a dual power plane. A single power plane means that both the CPU core and the I/O pins run at the same voltage, whereas a dual power plane means that the core and I/O voltage values are different.

Table 3.19 Socket 7 Single- and Dual-Plane Processor Voltages

Voltage Setting	Processor	Core Voltage	I/O Voltage	Voltage Planes
VRE (3.5V)	Intel Pentium	3.5V	3.5V	Single
STD (3.3V)	Intel Pentium	3.3V	3.3V	Single
MMX (2.8V)	Intel MMX Pentium	2.8V	3.3V	Dual
VRE (3.5V)	AMD K5	3.5V	3.5V	Single
3.2V	AMD-K6	3.2V	3.3V	Dual
2.9V	AMD-K6	2.9V	3.3V	Dual
2.4V	AMD-K6-2/K6-3	2.4V	3.3V	Dual
2.2V	AMD-K6/K6-2	2.2V	3.3V	Dual
VRE (3.5V)	Cyrix 6x86	3.5V	3.5V	Single
2.9V	Cyrix 6x86MX/M-II	2.9V	3.3V	Dual
MMX (2.8V)	Cyrix 6x86L	2.8V	3.3V	Dual
2.45V	Cyrix 6x86LV	2.45V	3.3V	Dual

Generally, the acceptable range is plus or minus 5% from the nominal intended setting.

Most Socket 7 and later Pentium motherboards supply several voltages (such as 2.5V, 2.7V, 2.8V, and 2.9V) for compatibility with future devices. A voltage regulator built into the motherboard converts the power supply voltage into the various levels the processor core requires. Check the documentation for your motherboard and processor to find the appropriate settings.

The Pentium Pro and Pentium II processors were the first to automatically determine their voltage settings by controlling the motherboard-based voltage regulator through built-in VID pins. Those are explained in more detail later in this chapter.

- ▶▶ See "Pentium Pro Processors," p. 134.
- ▶▶ See "Pentium II Processors," p. 138.

Note that on the STD or VRE settings, the core and I/O voltages are the same; these are single-plane voltage settings. Any time a voltage other than STD or VRE is set, the motherboard defaults to a dual-plane voltage setting where the core voltage can be specifically set, while the I/O voltage remains constant at 3.3V no matter what.

Socket 5 was designed to supply only STD or VRE settings, so any processor that can work at those settings can work in Socket 5 as well as Socket 7. Older Socket 4 designs can supply only 5V, and they have a completely different pinout (fewer pins overall), so using a processor designed for Socket 7 or Socket 5 in Socket 4 is not possible.

Most Socket 7 and later Pentium motherboards supply several voltages (such as 2.2V, 2.4V, 2.5V, 2.7V, 2.8V, and 2.9V as well as the older STD or VRE settings) for compatibility with many processors. A voltage regulator built into the motherboard converts the power supply voltage into the various levels required by the processor core. Check the documentation for your motherboard and processor to find the appropriate settings.

Starting with the Pentium Pro, all newer processors automatically determine their voltage settings by controlling the motherboard-based voltage regulator. That's done through built-in VID pins.

For hotrodding purposes, many newer motherboards for these processors have override settings that allow for manual voltage adjustment if desired. Many people have found that when attempting to overclock a processor, increasing the voltage by a tenth of a volt or so often helps. Of course, this increases the heat output of the processor and must be accounted for with adequate heatsinking and case cooling.

Note

Although modern processors use VID pins to enable the processor to select the correct voltage, newer processor that use the same processor socket as older processors might use a voltage setting not supported by the motherboard. Before upgrading an existing motherboard with a new processor, make sure the motherboard will support the processor's voltage and other features. You might need to install a BIOS upgrade before upgrading the processor to ensure that the processor is properly recognized by the motherboard.

Heat and Cooling Problems

Heat can be a problem in any high-performance system. The higher-speed processors consume more power and therefore generate more heat. The processor is usually the single most power-hungry chip in a system, and in most situations, the fan inside your computer case is incapable of handling the load without some help.

To ensure a constant flow of air and more consistent performance, most processors include some form of heatsink, which is designed to draw heat away from the processor. Additionally, most heatsinks incorporate fans so they don't have to rely on the airflow within the system. Heatsinks with fans are referred to as *active* heatsinks (see Figure 3.34). Active heatsinks have a power connection. Older ones often used a spare disk drive power connector, but most recent heatsinks plug in to dedicated heatsink power connections found on the newer motherboards. Heatsink power connections also provide a connection used to monitor fan performance through the BIOS Hardware Monitor or PC Health screen. Fan performance can also be displayed within the operating system by using a monitoring program.

Processor cooling, including heatsinks, is covered in detail in Chapter 21.

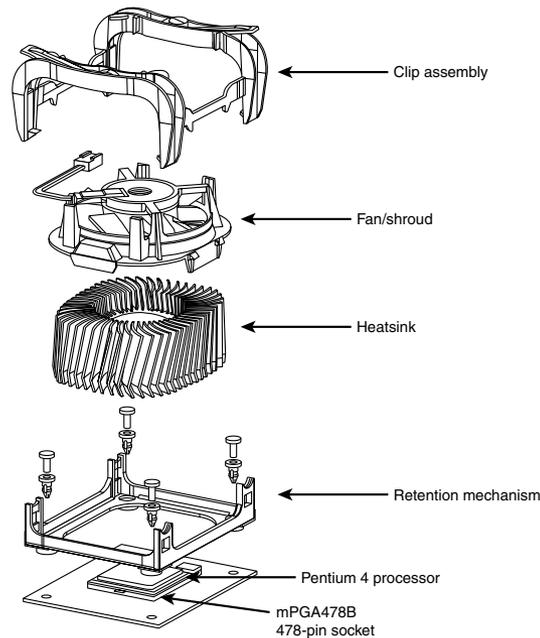


Figure 3.34 Active heatsink suitable for a Pentium 4 processor using Socket 478.

Math Coprocessors (Floating-Point Units)

This section covers the floating-point unit (FPU) contained in the processor, which was formerly a separate external math coprocessor in the 386 and older chips. Older central processing units designed by Intel (and cloned by other companies) used an external math coprocessor chip. However, when Intel introduced the 486DX, it included a built-in math coprocessor, and every processor built by Intel (and AMD and Cyrix, for that matter) since then includes a math coprocessor. Coprocessors provide hardware for floating-point math, which otherwise would create an excessive drain on the main CPU. Math chips speed your computer's operation only when you are running software designed to take advantage of the coprocessor. All the subsequent fifth- and sixth-generation Intel and compatible processors (such as those from AMD and Cyrix) have featured an integrated floating-point unit.

Math chips (as coprocessors sometimes are called) can perform high-level mathematical operations—long division, trigonometric functions, roots, and logarithms, for example—at 10–100 times the speed of the corresponding main processor. The operations performed by the math chip are all operations that make use of noninteger numbers (numbers that contain digits after the decimal point). The need to process numbers in which the decimal is not always the last character leads to the term *floating point* because the decimal (point) can move (float), depending on the operation. The integer units in the primary CPU work with integer numbers, so they perform addition, subtraction, and multiplication operations. The primary CPU is designed to handle such computations; these operations are not offloaded to the math chip.

The instruction set of the math chip is different from that of the primary CPU. A program must detect the existence of the coprocessor and then execute instructions written explicitly for that coprocessor; otherwise, the math coprocessor draws power and does nothing else. Fortunately, most modern programs that can benefit from the use of the coprocessor correctly detect and use the coprocessor. These programs usually are math intensive: spreadsheet programs, database applications, statistical programs, and graphics programs, such as computer-aided design (CAD) software. Word processing programs do not benefit from a math chip and therefore are not designed to use one.

Table 3.20 summarizes the coprocessors available for the Intel family of processors.

Table 3.20 Math Coprocessor Summary

Processor	Coprocessor	Processor	Coprocessor
8086	8087	486DX4/5x86	Built-in FPU
8088	8087	Intel Pentium/Pentium MMX	Built-in FPU
286	287	Cyrix 6x86/MI/MII	Built-in FPU
386SX	387SX	AMD K5/K6/Athlon/Duron	Built-in FPU
386DX	387DX	Pentium II/III/Celeron/Xeon	Built-in FPU
486SX	487SX, DX2/OverDrive ¹	Pentium 4/Celeron	Built-in FPU
487SX ¹	Built-in FPU	Pentium D/EE/Celeron D	Built-in FPU
486SX2	DX2/OverDrive ²	Athlon 64/FX/X2/Opteron	Built-in FPU
486DX	Built-in FPU	Itanium/Itanium II	Built-in FPU
486DX2	Built-in FPU		

FPU = Floating-point unit

1. The 487SX chip is a modified pinout 486DX chip with the math coprocessor enabled. When you plug in a 487SX chip, it disables the 486SX main processor and takes over all processing.
2. The DX2/OverDrive is equivalent to the SX2 with the addition of a functional FPU.

Although virtually all processors since the 486 series have built-in floating-point units, they vary in performance. Historically, the Intel processor FPUs have dramatically outperformed those from AMD and Cyrix, although AMD and Cyrix are achieving performance parity in their newer offerings.

Within each of the original 8087 group, the maximum speed of the math chips varies. A suffix digit after the main number, as shown in Table 3.21, indicates the maximum speed at which a system can run a math chip.

Table 3.21 Maximum Math Chip Speeds

Part	Speed	Part	Speed
8087	5MHz	287	6MHz
8087-3	5MHz	287-6	6MHz
8087-2	8MHz	287-8	8MHz
8087-1	10MHz	287-10	10MHz

The 387 math coprocessors and the 486 or 487 and Pentium processors always indicate their maximum speed ratings in MHz in the part number suffix. A 486DX2-66, for example, is rated to run at 66MHz. Some processors incorporate clock multiplication, which means they can run at different speeds compared with the rest of the system.

Most systems that use the 386 or earlier processors are socketed for a math coprocessor as an option, but they do not include a coprocessor as standard equipment. A few systems on the market at that time didn't even have a socket for the coprocessor because of cost and size considerations. These systems were usually low-cost or portable systems, such as older laptops, the IBM PS/1, and the PCjr. For more specific information about math coprocessors, see the discussions of the specific chips—8087, 287, 387, and 487SX—in the later sections. Table 3.22 shows the specifications of the various math coprocessors.

Table 3.22 Older Intel Math Coprocessor Specifications

Name	Power Consumption	Case Minimum Temperature	Case Maximum Temperature	No. of Transistors	Date Introduced
8087	3 watts	0°C, 32°F	85°C, 185°F	45,000	1980
287	3 watts	0°C, 32°F	85°C, 185°F	45,000	1982
287XL	1.5 watts	0°C, 32°F	85°C, 185°F	40,000	1990
387SX	1.5 watts	0°C, 32°F	85°C, 185°F	120,000	1988
387DX	1.5 watts	0°C, 32°F	85°C, 185°F	120,000	1987

Most often, you can learn which CPU and math coprocessor are installed in a particular system by checking the markings on the chip.

Note

Most applications that formerly used floating-point math now use SSE, SSE2, or SSE3 instructions instead. These instructions are faster and more accurate than x87 floating-point math.

Processor Bugs

Processor manufacturers use specialized equipment to test their own processors, but you have to settle for a little less. The best processor-testing device to which you have access is a system that you know is functional; you then can use the diagnostics available from various utility software companies or your system manufacturer to test the motherboard and processor functions.

Perhaps the most infamous of these bugs is the floating-point division math bug in the early Pentium processors. This and a few other bugs are discussed in detail later in this chapter.

Because the processor is the brain of a system, most systems don't function with a defective processor. If a system seems to have a dead motherboard, try replacing the processor with one from a functioning motherboard that uses the same CPU chip. You might find that the processor in the original board is the culprit. If the system continues to play dead, however, the problem is elsewhere, most likely in the motherboard, memory, or power supply. See the chapters that cover those parts of the system for more information on troubleshooting those components. I must say that in all my years of troubleshooting and repairing PCs, I have rarely encountered defective processors.

A few system problems are built in at the factory, although these bugs or design defects are rare. By learning to recognize these problems, you can avoid unnecessary repairs or replacements. Each processor section describes several known defects in that generation of processors, such as the infamous floating-point error in the Pentium. For more information on these bugs and defects, see the following sections, and check with the processor manufacturer for updates.

Microcode and the Processor Update Feature

All processors can contain design defects or errors. Many times, the effects of any given bug can be avoided by implementing hardware or software workarounds. Intel documents these bugs and workarounds well for its processors in its processor Specification Update manual; this manual is available from Intel's website. Most of the other processor manufacturers also have bulletins or tips on their websites listing any problems or special fixes or patches for their chips.

Previously, the only way to fix a processor bug was to work around it or replace the chip with one that had the bug fixed. Starting with the Intel P6 and P7 processors, including the Pentium Pro through Pentium D, many bugs in a processor's design can be fixed by altering the *microcode* in the processor. Microcode is essentially a set of instructions and tables in the processor that control how the processor operates. These processors incorporate a new feature called *reprogrammable microcode*,

which enables certain types of bugs to be worked around via microcode updates. The microcode updates reside in the motherboard ROM BIOS and are loaded into the processor by the motherboard BIOS during the POST. Each time the system is rebooted, the fix code is reloaded, ensuring that it will have the bug fix installed anytime the processor is operating.

The updated microcode for a given processor is provided by Intel to the motherboard manufacturer so it can incorporate the microcode into the flash ROM BIOS for the board. This is one reason it is important to install the most recent motherboard BIOS anytime you install a new processor. If your processor is newer than your motherboard ROM BIOS code, it probably doesn't include updated microcode to support your processor. In that case, you should visit the website of your motherboard manufacturer so you can download and install the latest BIOS update for your motherboard.

Processor Codenames

Intel, AMD, and Cyrix have always used codenames when talking about future processors. The codenames usually are not supposed to become public, but they typically do. They can often be found in online and print news and magazine articles talking about future-generation processors. Sometimes, they even appear in motherboard manuals because the manuals are written before the processors are officially introduced. Table 3.23 lists processor codenames for reference purposes.

Table 3.23 Processor Codenames

AMD Codename	Description
X5	5x86-133 [Socket 3]
SSA5	K5 (original PR75-PR100) [Socket 5, 7]
5k86	K5 (newer PR120-PR200) [Socket 7]
K6	Original AMD K6 core; canceled
NX686	NexGen K6 core; became the K6 [Socket 7]
Little Foot	0.25 μ m K6 [Socket 7]
Chompers	K6-2 [Socket 7, Super7]
Sharptooth	K6-3 [Super7]
Argon	Formerly K7
K7	Athlon [Slot A]
K75	0.18 μ m Athlon [Slot A]
K76	0.18 μ m Athlon (copper interconnects) [Slot A]
K8	Athlon 64
Thunderbird	Athlon [Slot A, Socket A]
Mustang	Athlon w/large L2; canceled
Corvette	Former mobile Athlon (now Palomino)
Palomino	0.18 μ m Athlon XP/MP, Mobile Athlon 4 [Socket A]
Thoroughbred-A	0.13 μ m Athlon XP/MP 1700–2100+ [Socket A]
Thoroughbred-B	0.13 μ m Athlon XP 1700–2400+, 2600–2800+ [Socket A]; Sempron 2200–2800+ [Socket A]
Barton	0.13 μ m Athlon XP/MP w/512K L2 [Socket A]
Thorton	Athlon XP (256KB L2 cache) [Socket A]
Spitfire	Duron [Socket A]
Camaro	Former Morgan
Morgan	Mobile Duron and Model 7 Duron 900MHz–1.3GHz [Socket A]

Table 3.23 Continued

AMD Codename	Description
Applebred	Duron 1.4GHz–1.8GHz
Appaloosa	0.13 μ m Morgan [Socket A]
ClawHammer	Athlon 64 (64-bit CPU) [Socket 754 and Socket 939]
ClawHammer DP	Early name for Opteron DP [Socket 940]
Newcastle	Athlon 64 [Socket 754 and Socket 939]
Winchester	0.09 μ m Athlon 64 [Socket 939]
San Diego	0.09 μ m Athlon 64 and Athlon 64 FX w/SSE3 extensions [Socket 939]
Venice	0.09 μ m Athlon 64 w/SSE3 extensions [Socket 939]
Odessa	0.09 μ m mobile Athlon 64
Manchester	Athlon 64 X2 w/512KB L2 cache and SSE3 extensions [Socket 939]
Toledo	Athlon 64 X2 w/1024KB L2 cache and SSE3 extensions [Socket 939]
SledgeHammer	Opteron w/large L2 [Socket 940]
Palermo	0.09 μ m Sempron [Socket 754]
Paris	Sempron [Socket 754]
Oakville	Mobile Athlon 64 and Sempron [Socket 754]
Windsor	Athlon 64 X2 and Athlon 64 FX-62 [Socket M2]
Orleans	Athlon 64 [Socket M2]
Manila	Sempron [Socket M2]
Intel Codename	Description
P23	486SX [Socket 1, 2, 3]
P23S	486SX SL-enhanced [Socket 1, 2, 3]
P23N	487SX (coprocessor) [Socket 1]
P4	486DX [Socket 1, 2, 3]
P4S	486DX SL-enhanced [Socket 1, 2, 3]
P24	486DX2 [Socket 1, 2, 3]
P24S	486DX2 SL-enhanced [Socket 1, 2, 3]
P24D	486DX2 (write-back cache) [Socket 3]
P24C	486DX4 [Socket 3]
P23T	486DXODP (486 OverDrive) [Socket 3]
P4T	486DXODPR (486 OverDrive) [Socket 1, 2, 3]
P24T	PODP5V (Pentium OverDrive) [Socket 2, 3]
P24CT	Pentium OverDrive 3.3V [Socket 2, 3]
P5	Pentium 60/66MHz [Socket 4]
P5T	Pentium OverDrive 120/133MHz [Socket 4]
P54C	Pentium 75MHz–120MHz [Socket 5, 7]
P54CQS	Pentium 120MHz–133MHz [Socket 5, 7]
P54CS	Pentium 120MHz–200MHz [Socket 7]
P54CT(A)	Pentium OverDrive [Socket 5, 7]
P55C	Pentium MMX [Socket 7]
P54CTB	Pentium OverDrive MMX [Socket 5, 7]
Tillamook	Mobile Pentium MMX [Mobile Module]

Table 3.23 Continued

Intel Codename	Description
P6	Pentium Pro [Socket 8]
P6T	Pentium II OverDrive [Socket 8]
Klamath	0.35 μ m Pentium II [Slot 1]
Deschutes	0.25 μ m Pentium II [Slot 1]
Drake	0.25 μ m Pentium II Xeon [Slot 2]
Tonga	Mobile Pentium II
Covington	Celeron (cacheless Pentium II) [Slot 1]
Mendocino	0.25 μ m Celeron w/128KB on-die L2 [Slot 1, Socket 370]
Dixon	Mobile Pentium II w/256KB on-die L2
Katmai	0.25 μ m Pentium III w/SSE [Slot 1]
Tanner	0.25 μ m Pentium III Xeon w/SSE [Slot 2]
Coppermine	0.18 μ m Pentium III w/on-die L2 [Slot 1, Socket 370]
Tualatin	0.13 μ m Pentium III [Socket 370]
Coppermine-T	0.18 μ m Pentium III w/Tualatin voltage [Socket 370]
Cascades	0.18 μ m Pentium III Xeon [Slot 2]
Coppermine-128	0.18 μ m Celeron w/128KB L2 [Socket 370]
Timna	Mobile Celeron w/DRAM controller; canceled
P68	Willamette
Willamette	0.18 μ m Pentium 4 [Socket 423, 478]
Northwood	0.13 μ m Pentium 4 [Socket 478]
Prescott	0.09 μ m Pentium 4 w/HT; Celeron D [Socket 478]; Celeron D [Socket 775]
Smithfield	Pentium D, Pentium Extreme Edition [Socket 775]
Presler	0.065 μ m Pentium D
Conroe	0.065 μ m Pentium D with reduced power consumption
Banias	130nm Pentium M w/1MB L2
Yonah	Pentium M (dual-core) and Celeron M (single-core)
Merom	64-bit version of Yonah
Foster	Xeon DP [Socket 603]
Foster MP	Xeon MP [Socket 603]
Prestonia	0.13 μ m Xeon DP [Socket 603]
Gallatin	0.13 μ m Xeon MP [Socket 603]
Nocona	0.09 μ m Xeon [Socket 603]; Pentium 4 Extreme Edition [Socket 478 and Socket 775]
Dothan	90nm Pentium M w/2MB L2
P7	Former Merced (Itanium)
Merced	Itanium [PAC 418]
McKinley	Itanium 2 w/3MB on-die L3 [PAC 418]
Madison	0.13 μ m Itanium 2
Deerfield	Low-cost Madison
Montecito	0.09 μ m Madison
Shavano	Future Itanium family chip

Note that the codenames and information listed in these tables are used before the processor is officially introduced. After a chip is introduced, the codename is dropped and the chip is thereafter referred to by the marketing name used at the time of the introduction. Because many of these names refer to chips that are not yet officially released, the names or specifications might change. For chipset codenames, see Chapter 4.

P1 (086) First-Generation Processors

The first generation of processors represents the series of chips from Intel that were found in the first PCs. IBM, as the architect of the PC at the time, chose Intel processors and support chips to build the PC motherboard, setting a standard that would hold for many subsequent processor generations to come.

8088 and 8086 Processors

Intel introduced the 8086 back in June 1978. The 8086 was one of the first 16-bit processor chips on the market; at the time, virtually all other processors were 8-bit designs. The 8086 had 16-bit internal registers and could run a new class of software using 16-bit instructions. It also had a 16-bit external data path, so it could transfer data to memory 16 bits at a time.

The address bus was 20 bits wide, which enabled the 8086 to address a full 1MB (2^{20}) of memory. This was in stark contrast to most other chips of that time that had 8-bit internal registers, an 8-bit external data bus, and a 16-bit address bus allowing a maximum of only 64KB of RAM (2^{16}).

Unfortunately, most of the personal computer world at the time was using 8-bit processors, which ran 8-bit CP/M (Control Program for Microprocessors) operating systems and software. The board and circuit designs at the time were largely 8-bit, as well. Building a full 16-bit motherboard and memory system was costly, pricing such a computer out of the market.

The cost was high because the 8086 needed a 16-bit data bus rather than a less expensive 8-bit bus. Systems available at that time were 8-bit, and slow sales of the 8086 indicated to Intel that people weren't willing to pay for the extra performance of the full 16-bit design. In response, Intel introduced a kind of crippled version of the 8086, called the 8088. The 8088 essentially deleted 8 of the 16 bits on the data bus, making the 8088 an 8-bit chip as far as data input and output were concerned. However, because it retained the full 16-bit internal registers and the 20-bit address bus, the 8088 ran 16-bit software and was capable of addressing a full 1MB of RAM.

For these reasons, IBM selected the 8-bit 8088 chip for the original IBM PC. Years later, IBM was criticized for using the 8-bit 8088 instead of the 16-bit 8086. In retrospect, it was a very wise decision. IBM even covered up the physical design in its ads, which at the time indicated its new PC had a "high-speed 16-bit microprocessor." IBM could say that because the 8088 still ran the same powerful 16-bit software the 8086 ran, just a little more slowly. In fact, programmers universally thought of the 8088 as a 16-bit chip because there was virtually no way a program could distinguish an 8088 from an 8086. This enabled IBM to deliver a PC capable of running a new generation of 16-bit software, while retaining a much less expensive 8-bit design for the hardware. Because of this, the IBM PC was actually priced less at its introduction than the most popular PC of the time, the Apple II. For the trivia buffs out there, the IBM PC listed for \$1,265 and included only 16KB of RAM, whereas a similarly configured Apple II cost \$1,355.

Even though the 8088 was introduced in June 1979, the original IBM PC that used the processor did not appear until August 1981. Back then, a significant lag time often occurred between the introduction of a new processor and systems that incorporated it. That is unlike today, when new processors and systems using them often are released on the same day.

The 8088 in the IBM PC ran at 4.77MHz; the average instruction on the 8088 took 12 cycles to complete.

Computer users sometimes wonder why a 640KB conventional-memory barrier exists if the 8088 chip can address 1MB of memory. The conventional-memory barrier exists because IBM reserved 384KB of the upper portion of the 1024KB (1MB) address space of the 8088 for use by adapter cards and system BIOS. The lower 640KB is the conventional memory in which DOS and software applications execute.

80186 and 80188 Processors

After Intel produced the 8086 and 8088 chips, it created versions of these chips with some of the required support components integrated within the processor.

The relationship between the 80186 and 80188 is the same as that of the 8086 and 8088; the 80188 is essentially an 8-bit interface version of the 80186. The advantage of the 80186 and 80188 is that they combine on a single chip 15–20 of the 8086–8088 series system components—a fact that can greatly reduce the number of components in a computer design. The 80186 and 80188 chips were used for highly intelligent peripheral adapter cards of that age, such as network adapters.

8087 Coprocessor

The math coprocessor or floating-point unit that was paired with the 8086 chip was called the 8087 numeric data processor (NDP), the math coprocessor, or simply the math chip. The 8087 is designed to perform high-level math operations at many times the speed of the main processor. The primary advantage of using this chip is the increased execution speed in number-crunching programs, such as spreadsheet applications.

P2 (286) Second-Generation Processors

The second generation of PC processors allowed for a great leap in system speed and processing efficiency. With these chips we went from moving 8 bits of data around to moving 16 bits at a time. The following section details the second-generation PC processor, the 286.

286 Processors

The Intel 80286 (normally abbreviated as 286) processor did not suffer from the compatibility problems that damned the 80186 and 80188. The 286 chip, first introduced in 1982, is the CPU behind the original IBM PC AT (Advanced Technology). Other computer makers manufactured what came to be known as IBM clones, with many of these manufacturers calling their systems AT-compatible or AT-class computers.

When IBM developed the AT, it selected the 286 as the basis for the new system because the chip provided compatibility with the 8088 used in the PC and the XT. Therefore, software written for those chips should run on the 286. The 286 chip is many times faster than the 8088 used in the XT, and at the time it offered a major performance boost to PCs used in businesses. The processing speed, or throughput, of the original AT (which ran at 6MHz) is five times greater than that of the PC running at 4.77MHz. The die for the 286 is shown in Figure 3.35.

286 systems are faster than their predecessors for several reasons. The main reason is that 286 processors are much more efficient in executing instructions. An average instruction takes 12 clock cycles on the 8086 or 8088, but takes an average of only 4.5 cycles on the 286 processor. Additionally, the 286 chip can handle up to 16 bits of data at a time through an external data bus twice the size of the 8088.

The 286 chip has two modes of operation: real mode and protected mode. The two modes are distinct enough to make the 286 resemble two chips in one. In real mode, a 286 acts essentially the same as an 8086 chip and is fully *object-code compatible* with the 8086 and 8088. (A processor with object-code compatibility can run programs written for another processor without modification and execute every system instruction in the same manner.)

In the protected mode of operation, the 286 was truly something new. In this mode, a program designed to take advantage of the chip's capabilities believes that it has access to 1GB of memory (including virtual memory). The 286 chip, however, can address only 16MB of hardware memory. A significant failing of the 286 chip is that it cannot switch from protected mode to real mode without a hardware reset (a warm reboot) of the system. (It can, however, switch from real mode to protected mode without a reset.) A major improvement of the 386 over the 286 is that software can switch the 386 from real mode to protected mode, and vice versa. See the section "Processor Modes," earlier in this chapter for more information.

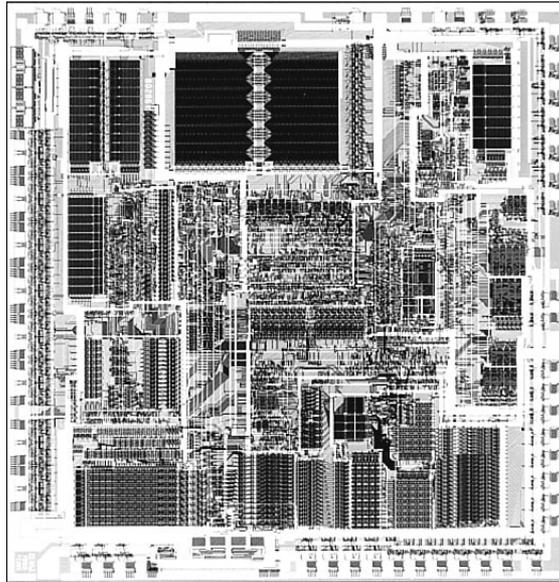


Figure 3.35 286 Processor die. *Photograph used by permission of Intel Corporation.*

Only a small amount of software that took advantage of the 286 chip was sold until Windows 3.0 offered standard mode for 286 compatibility; by that time, the hottest-selling chip was the 386. Still, the 286 was Intel's first attempt to produce a CPU chip that supported multitasking, in which multiple programs run at the same time.

80287 Coprocessor

The 80287, internally, is the same math chip as the 8087, although the pins used to plug them into the motherboard are different. Both the 80287 and the 8087 operate as though they are identical.

In most systems, the 80286 internally divides the system clock by 2 to derive the processor clock. The 80287 internally divides the system-clock frequency by 3. For this reason, most AT-type computers run the 80287 at one-third the system clock rate, which also is two-thirds the clock speed of the 80286. Because the 286 and 287 chips are asynchronous, the interface between the 286 and 287 chips is not as efficient as with the 8088 and 8087.

P3 (386) Third-Generation Processors

The third generation represents perhaps the most significant change in processors since the first PC. The big deal was the migration from processors that handled 16-bit operations to true 32-bit chips. The third-generation processors were so far ahead of their time, it took fully 10 years before 32-bit operating systems and software became mainstream, and by that time the third-generation chips had become a memory. The following section details the third-generation processors.

386 Processors

The Intel 80386 (usually abbreviated as 386) caused quite a stir in the PC industry because of the vastly improved performance it brought to the personal computer. Compared with 8088 and 286 systems, the 386 chip offered greater performance in almost all areas of operation.

The 386 is a full 32-bit processor optimized for high-speed operation and multitasking operating systems. Intel introduced the chip in 1985, but the 386 appeared in the first systems in late 1986 and

early 1987. The Compaq Deskpro 386 and systems made by several other manufacturers introduced the chip; somewhat later, IBM used the chip in its PS/2 Model 80.

The 386 can execute the real-mode instructions of an 8086 or 8088, but in fewer clock cycles. The 386 was as efficient as the 286 in executing instructions—the average instruction took about 4.5 clock cycles. In raw performance, therefore, the 286 and 386 actually seemed to be at almost equal clock rates. The 386 offered greater performance in other ways, mainly because of additional software capability (modes) and a greatly enhanced memory management unit (MMU). The die for the 386 is shown in Figure 3.36.

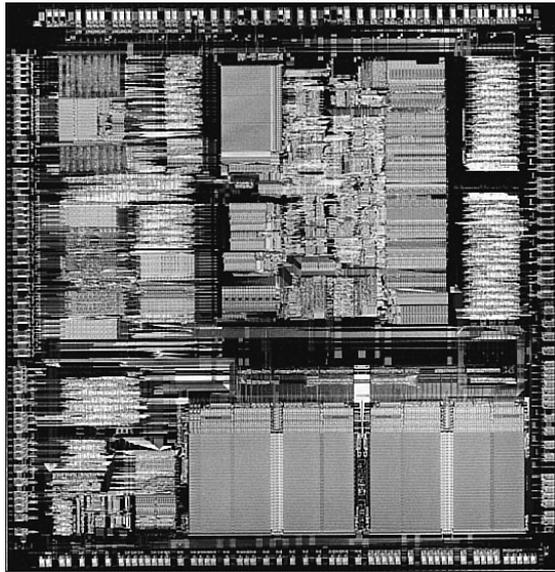


Figure 3.36 386 processor die. *Photograph used by permission of Intel Corporation.*

The 386 can switch to and from protected mode under software control without a system reset—a capability that makes using protected mode more practical. In addition, the 386 includes a new mode, called virtual real mode, which enables several real-mode sessions to run simultaneously under protected mode.

The protected mode of the 386 is fully compatible with the protected mode of the 286. The protected mode for both chips often is called their native mode of operation because these chips are designed for advanced operating systems such as Windows NT/2000/XP, which run only in protected mode. Intel extended the memory-addressing capabilities of 386 protected mode with a new MMU that provided advanced memory paging and program switching. These features were extensions of the 286 type of MMU, so the 386 remained fully compatible with the 286 at the system-code level.

The 386 chip's virtual real mode was also new. In virtual real mode, the processor could run with hardware memory protection while simulating an 8086's real-mode operation. Multiple copies of DOS and other operating systems, therefore, could run simultaneously on this processor, each in a protected area of memory. If the programs in one segment crashed, the rest of the system was protected.

Numerous variations of the 386 chip were manufactured, some of which are less powerful and some of which are less power hungry. The following sections cover the members of the 386-chip family and their differences.

386DX Processors

The 386DX chip was the first of the 386 family members that Intel introduced. The 386 is a full 32-bit processor with 32-bit internal registers, a 32-bit internal data bus, and a 32-bit external data bus. The 386 contains 275,000 transistors in a very large scale integration (VLSI) circuit. The chip comes in a 132-pin package and draws approximately 400 milliamperes (ma), which is less power than even the 8086 requires. The 386 has a smaller power requirement because it is made of Complementary Metal-Oxide Semiconductor (CMOS) materials. The CMOS design enables devices to consume extremely low levels of power.

The Intel 386 chip was available in clock speeds ranging from 16MHz–33MHz; other manufacturers, primarily AMD and Cyrix, offered comparable versions with speeds up to 40MHz.

The 386DX can address 4GB of physical memory. Its built-in virtual memory manager enables software designed to take advantage of enormous amounts of memory to act as though a system has 64TB of memory. (A terabyte, or TB, is 1,099,511,627,776 bytes of memory, or about 1000GB.)

386SX Processors

The 386SX was designed for systems designers looking for 386 capabilities at 286 system prices. Similar to the 286, the 386SX is restricted to only 16 bits when communicating with other system components, such as memory. Internally, however, the 386SX is identical to the DX chip; the 386SX has 32-bit internal registers and can therefore run 32-bit software. The 386SX uses a 24-bit memory-addressing scheme like that of the 286, rather than the full 32-bit memory address bus of the standard 386. The 386SX, therefore, can address a maximum 16MB of physical memory rather than the 4GB of physical memory the 386DX can address. Before it was discontinued, the 386SX was available in clock speeds ranging from 16MHz to 33MHz.

The 386SX signaled the end of the 286 because of the 386SX chip's superior MMU and the addition of the virtual real mode. Under a software manager such as Windows or OS/2, the 386SX can run numerous DOS programs at the same time. The capability to run 386-specific software is another important advantage of the 386SX over any 286 or older design. For example, Windows 3.1 runs nearly as well on a 386SX as it does on a 386DX.

386SL Processors

The 386SL is another variation on the 386 chip. This low-power CPU had the same capabilities as the 386SX, but it was designed for laptop systems in which low power consumption was necessary. The SL chips offered special power-management features that were important to systems that ran on batteries. The SL chip also offered several sleep modes to conserve power.

The chip included an extended architecture that contained a System Management Interrupt (SMI), which provided access to the power-management features. Also included in the SL chip was special support for LIM (Lotus Intel Microsoft) expanded memory functions and a cache controller. The cache controller was designed to control a 16KB–64KB external processor cache.

These extra functions account for the higher transistor count in the SL chips (855,000) compared with even the 386DX processor (275,000). The 386SL was available in 25MHz clock speed.

Intel offered a companion to the 386SL chip for laptops called the 82360SL I/O subsystem. The 82360SL provided many common peripheral functions, such as serial and parallel ports, a direct memory access (DMA) controller, an interrupt controller, and power-management logic for the 386SL processor. This chip subsystem worked with the processor to provide an ideal solution for the small size and low power-consumption requirements of portable and laptop systems.

80387 Coprocessor

Although the 80387 chips ran asynchronously, 386 systems were designed so that the math chip ran at the same clock speed as the main CPU. Unlike the 80287 coprocessor, which was merely an 8087 with different pins to plug into the AT motherboard, the 80387 coprocessor was a high-performance math chip specifically designed to work with the 386.

All 387 chips used a low power-consumption CMOS design. The 387 coprocessor had two basic designs: the 387DX coprocessor, which was designed to work with the 386DX processor, and the 387SX coprocessor, which was designed to work with the 386SX, SL, or SLC processor.

Intel originally offered several speeds for the 387DX coprocessor. But when the company designed the 33MHz version, a smaller mask was required to reduce the lengths of the signal pathways in the chip. This increased the performance of the chip by roughly 20%.

Note

Because Intel lagged in developing the 387 coprocessor, some early 386 systems were designed with a socket for a 287 coprocessor. Performance levels associated with that combination, however, left much to be desired.

Installing a 387DX is easy, but you must be careful to orient the chip in its socket properly; otherwise, the chip will be destroyed. The most common cause of burned pins on the 387DX is incorrect installation. In many systems, the 387DX was oriented differently from other large chips. Follow the manufacturer's installation instructions carefully to avoid damaging the 387DX; Intel's warranty does not cover chips that are installed incorrectly.

Several manufacturers developed their own versions of the Intel 387 coprocessors, some of which were touted as being faster than the original Intel chips. The general compatibility record of these chips was very good.

P4 (486) Fourth-Generation Processors

The third generation had been a large change from the previous generations of processors. With the fourth generation, more refinement than complete redesign was accomplished. Even so, Intel, AMD, and others managed to literally double processor performance with their fourth-generation processors. The following section defines the fourth-generation processors from Intel, AMD, and others.

486 Processors

In the race for more speed, the Intel 80486 (normally abbreviated as 486) was another major leap forward. The additional power available in the 486 fueled tremendous growth in the software industry. Tens of millions of copies of Windows, and millions of copies of OS/2, have been sold largely because the 486 finally made the GUI of Windows and OS/2 a realistic option for people who work on their computers every day.

Four main features make a given 486 processor roughly twice as fast as an equivalent MHz 386 chip. These features are

- *Reduced instruction-execution time.* A single instruction in the 486 takes an average of only two clock cycles to complete, compared with an average of more than four cycles on the 386. Clock-multiplied versions, such as the DX2 and DX4, further reduced this to about two cycles per instruction.
- *Internal (Level 1) cache.* The built-in cache has a hit ratio of 90%–95%, which describes how often zero-wait-state read operations occur. External caches can improve this ratio further.
- *Burst-mode memory cycles.* A standard 32-bit (4-byte) memory transfer takes two clock cycles. After a standard 32-bit transfer, more data up to the next 12 bytes (or three transfers) can be transferred with only one cycle used for each 32-bit (4-byte) transfer. Thus, up to 16 bytes of contiguous, sequential memory data can be transferred in as little as five cycles instead of eight cycles or more. This effect can be even greater when the transfers are only 8 bits or 16 bits each.
- *Built-in (synchronous) enhanced math coprocessor (some versions).* The math coprocessor runs synchronously with the main processor and executes math instructions in fewer cycles than previous designs did. On average, the math coprocessor built into the DX-series chips provides two to three times greater math performance than an external 387 chip.

The 486 chip is about twice as fast as the 386, so a 386DX-40 is about as fast as a 486SX-20. This made the 486 a much more desirable option, primarily because it could more easily be upgraded to a DX2 or DX4 processor at a later time. You can see why the arrival of the 486 rapidly killed off the 386 in the marketplace.

Most of the 486 chips were offered in a variety of maximum speed ratings, varying from 16MHz up to 133MHz. Additionally, 486 processors have slight differences in overall pin configurations. The DX, DX2, and SX processors have a virtually identical 168-pin configuration, whereas the OverDrive chips have either the standard 168-pin configuration or a specially modified 169-pin OverDrive (sometimes also called 487SX) configuration. If your motherboard has two sockets, the primary one likely supports the standard 168-pin configuration, and the secondary (OverDrive) socket supports the 169-pin OverDrive configuration. Most of the later 486-based motherboards with a single ZIF socket support any of the 486 processors except the DX4. The DX4 is different because it requires 3.3V to operate instead of 5V, like most other chips up to that time.

A processor rated for a given speed always functions at any of the lower speeds. A 100MHz-rated 486DX4 chip, for example, runs at 75MHz if it is plugged into a 25MHz motherboard. Note that the DX2/OverDrive processors operate internally at two times the motherboard clock rate, whereas the DX4 processors operate at two, two-and-one-half, or three times the motherboard clock rate. Table 3.24 shows the various speed combinations that can result from using the DX2 or DX4 processors with different motherboard clock speeds.

Table 3.24 Intel DX2 and DX4 Operating Speeds Versus CPU Bus (Motherboard) Clock Speeds

CPU Bus Speed	DX2/DX4 Speed (2× Mode)	DX4 Speed (2.5× Mode)	DX4 Speed (3× Mode)
16MHz	32MHz	40MHz	48MHz
20MHz	40MHz	50MHz	60MHz
25MHz	50MHz	63MHz	75MHz
33MHz	66MHz	83MHz	100MHz
40MHz	80MHz	100MHz	120MHz
50MHz	100MHz	n/a	n/a

The internal multiplier of the DX4 processor is controlled by the CLKMUL (clock multiplier) signal at pin R-17 (Socket 1) or S-18 (Socket 2, 3, or 6). In most cases, one or two jumpers will be on the board near the processor socket to control the settings for these pins. The motherboard documentation should cover these settings if they can be changed.

One interesting capability here is to run the DX4-100 chip in a doubled mode with a 50MHz motherboard speed. This gives you a very fast memory bus, along with the same 100MHz processor speed, as if you were running the chip in a 33/100MHz tripled mode.

Many VL-Bus motherboards can run the VL-Bus slots in a buffered mode, add wait states, or even selectively change the clock only for the VL-Bus slots to keep them compatible. In most cases, they don't run properly at 50MHz. Consult your motherboard—or even better, your chipset documentation—to see how your board is set up.

Caution

When upgrading an existing system, you should be sure that your socket supports the chip you are installing. This was especially true when putting a DX4 processor in an older system. In that scenario, you needed some type of adapter to regulate the voltage down to 3.3V. Putting the DX4 in a 5V socket destroys the chip! See the earlier section on processor sockets for more information.

486DX Processors

The original Intel 486DX processor was introduced on April 10, 1989, and systems using this chip first appeared during 1990. The first chips had a maximum speed rating of 25MHz; later versions of the 486DX were available in 33MHz- and 50MHz-rated versions. The 486DX originally was available only in a 5V, 168-pin PGA version, but later became available in 5V, 196-pin plastic quad flat pack (PQFP) and 3.3V, 208-pin small quad flat pack (SQFP). These latter form factors were available in SL enhanced versions, which were intended primarily for portable or laptop applications in which saving power is important.

Two main features separate the 486 processor from its predecessors:

- The 486DX integrates functions such as the math coprocessor, cache controller, and cache memory into the chip.
- The 486 also was designed with easy installation and upgradeability in mind; double-speed OverDrive upgrades were available for most systems.

The 486DX processor was fabricated with low-power CMOS technology. The chip has a 32-bit internal register size, a 32-bit external data bus, and a 32-bit address bus. These dimensions are equal to those of the 386DX processor. The internal register size is where the “32-bit” designation used in advertisements comes from. The 486DX chip contains 1.2 million transistors on a piece of silicon no larger than your thumbnail. This figure is more than four times the number of components on 386 processors and should give you a good indication of the 486 chip’s relative power. The die for the 486 is shown in Figure 3.37.

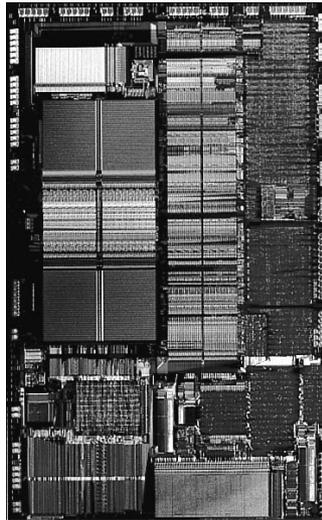


Figure 3.37 486 processor die. Photograph used by permission of Intel Corporation.

The standard 486DX contains a processing unit, floating-point unit (math coprocessor), memory-management unit, and cache controller with 8KB of internal-cache RAM. Due to the internal cache and a more efficient internal processing unit, the 486 family of processors can execute individual instructions in an average of only 2 processor cycles. Compare this figure with the 286 and 386 families, both of which execute an average 4.5 cycles per instruction. Compare it also with the original 8086 and 8088 processors, which execute an average 12 cycles per instruction. At a given clock rate (MHz), therefore, a 486 processor is roughly twice as efficient as a 386 processor; a 16MHz 486SX is roughly equal to a 33MHz 386DX system; and a 20MHz 486SX is equal to a 40MHz 386DX system. Any of the faster 486s are way beyond the 386 in performance.

The 486 is fully instruction-set-compatible with previous Intel processors, such as the 386, but offers several additional instructions (most of which have to do with controlling the internal cache).

Similar to the 386DX, the 486 can address 4GB of physical memory and manage as much as 64TB of virtual memory. The 486 fully supports the three operating modes introduced in the 386: real mode, protected mode, and virtual real mode:

- *Real mode.* In this mode, the 486 (similar to the 386) runs unmodified 8086-type software.
- *Protected mode.* In this mode, the 486 (similar to the 386) offers sophisticated memory paging and program switching.
- *Virtual real mode.* In this mode, the 486 (similar to the 386) can run multiple copies of DOS or other operating systems while simulating an 8086's real-mode operation. Under an operating system such as Windows or OS/2, therefore, both 16-bit and 32-bit programs can run simultaneously on this processor with hardware memory protection. If one program crashes, the rest of the system is protected, and you can reboot the blown portion through various means, depending on the operating software.

The 486DX series has a built-in math coprocessor that sometimes is called an MCP (math coprocessor) or FPU. This series is unlike previous Intel CPU chips, which required you to add a math coprocessor if you needed faster calculations for complex mathematics. The FPU in the 486DX series is 100% software-compatible with the external 387 math coprocessor used with the 386, but it delivers more than twice the performance. It runs in synchronization with the main processor and executes most instructions in half as many cycles as the 386.

486SL

The 486SL was a short-lived, standalone chip. The SL enhancements and features became available in virtually all the 486 processors (SX, DX, and DX2) in what are called SL enhanced versions. SL enhancement refers to a special design that incorporates special power-saving features.

The SL enhanced chips originally were designed to be installed in laptop or notebook systems that run on batteries, but they found their way into desktop systems, as well. The SL-enhanced chips featured special power-management techniques, such as sleep mode and clock throttling, to reduce power consumption when necessary. These chips were available in 3.3V versions, as well.

Intel designed a power-management architecture called system management mode (SMM). This mode of operation is totally isolated and independent from other CPU hardware and software. SMM provides hardware resources such as timers, registers, and other I/O logic that can control and power down mobile-computer components without interfering with any of the other system resources. SMM executes in a dedicated memory space called system management memory, which is not visible and does not interfere with operating system and application software. SMM has an interrupt called system management interrupt (SMI), which services power-management events and is independent from—and a higher priority than—any of the other interrupts.

SMM provides power management with flexibility and security that were not available previously. For example, an SMI occurs when an application program tries to access a peripheral device that is powered down for battery savings, which powers up the peripheral device and re-executes the I/O instruction automatically.

Intel also designed a feature called Suspend/Resume in the SL processor. The system manufacturer can use this feature to provide the portable computer user with instant on-and-off capability. An SL system typically can resume (instant on) in 1 second from the suspend state (instant off) to exactly where it left off. You do not need to reboot, load the operating system, or load the applications and their data. Instead, simply push the Suspend/Resume button and the system is ready to go.

The SL CPU was designed to consume almost no power in the suspend state. This feature means that the system can stay in the suspend state possibly for weeks and yet start up instantly right where it left off. An SL system can keep working data in normal RAM memory safe for a long time while it is in the suspend state, but saving to a disk still is prudent.

486SX

The 486SX, introduced in April 1991, was designed to be sold as a lower-cost version of the 486. The 486SX is virtually identical to the full DX processor, but the chip does not incorporate the FPU or math coprocessor portion.

As you read earlier in this chapter, the 386SX was a scaled-down (some people would say crippled) 16-bit version of the full-blown 32-bit 386DX. The 386SX even had a completely different pinout and was not interchangeable with the more powerful DX version. The 486SX, however, is a different story. The 486SX is, in fact, a full-blown 32-bit 486 processor that is basically pin compatible with the DX. A few pin functions are different or rearranged, but each pin fits into the same socket.

The 486SX chip was more a marketing quirk than new technology. Early versions of the 486SX chip actually were DX chips that showed defects in the math-coprocessor section. Instead of being scrapped, the chips were packaged with the FPU section disabled and sold as SX chips. This arrangement lasted for only a short time; thereafter, SX chips got their own mask, which is different from the DX mask. (A *mask* is the photographic blueprint of the processor and is used to etch the intricate signal pathways into a silicon chip.) The transistor count dropped to 1.185 million (from 1.2 million) to reflect this new mask.

The 486SX was available in 16MHz-, 20MHz-, 25MHz-, and 33MHz-rated speeds, and a 486 SX/2 was also available that ran at up to 50MHz or 66MHz. The 486SX typically was made in a 168-pin version, although other surface-mount versions were manufactured in SL-enhanced models.

Despite what Intel's marketing and sales information implies, no technical provision exists for adding a separate math coprocessor to a 486SX system; neither was a separate math coprocessor chip ever available to plug in. Instead, Intel wanted you to add a new 486 processor with a built-in math unit and disable the SX CPU that already was on the motherboard.

487SX

The 487SX math coprocessor, as Intel called it, really is a complete 25MHz 486DX CPU with an extra pin added and some other pins rearranged. When the 487SX is installed in the extra socket provided in a 486SX CPU-based system, the 487SX turns off the existing 486SX via a new signal on one of the pins. The extra key pin actually carries no signal itself and exists only to prevent improper orientation when the chip is installed in a socket.

The 487SX takes over all CPU functions from the 486SX and also provides math coprocessor functionality in the system. At first glance, this setup seems rather strange and wasteful, so perhaps further explanation is in order. Fortunately, the 487SX turned out to be a stopgap measure while Intel prepared its real surprise: the OverDrive processor. The DX2/OverDrive speed-doubling chips, which are designed for the 487SX 169-pin socket, have the same pinout as the 487SX. These upgrade chips are installed in exactly the same way as the 487SX; therefore, any system that supports the 487SX also supports the DX2/OverDrive chips.

Originally, Intel discouraged users from removing the existing chip from the socket and replacing it with a 487SX (or even a DX or DX2/OverDrive). Instead, Intel recommended that PC manufacturers include a dedicated upgrade (OverDrive) socket in their systems because several risks were involved in removing the original CPU from a standard socket. (The following section elaborates on those risks.) Later Intel recommended—or even insisted on—the use of a single processor socket of a ZIF design, which makes upgrading an easy task physically.

DX2/OverDrive and DX4 Processors

On March 3, 1992, Intel introduced the DX2 speed-doubling processors. On May 26, 1992, Intel announced that the DX2 processors also would be available in a retail version called OverDrive. Originally, the OverDrive versions of the DX2 were available only in 169-pin versions, which meant that they could be used only with 486SX systems that had sockets configured to support the rearranged pin configuration.

On September 14, 1992, Intel introduced 168-pin OverDrive versions for upgrading 486DX systems. These processors could be added to existing 486 (SX or DX) systems as an upgrade, even if those systems did not support the 169-pin configuration. When you use this processor as an upgrade, you install the new chip in your system, which subsequently runs twice as fast.

The DX2/OverDrive processors run internally at twice the clock rate of the host system. If the motherboard clock is 25MHz, for example, the DX2/OverDrive chip runs internally at 50MHz; likewise, if the motherboard is a 33MHz design, the DX2/OverDrive runs at 66MHz. The DX2/OverDrive speed doubling has no effect on the rest of the system; all components on the motherboard run the same as they do with a standard 486 processor. Therefore, you do not have to change other components (such as memory) to accommodate the double-speed chip. The DX2/OverDrive chips have been available in several speeds. Three speed-rated versions have been offered:

- 40MHz DX2/OverDrive for 16MHz or 20MHz systems
- 50MHz DX2/OverDrive for 25MHz systems
- 66MHz DX2/OverDrive for 33MHz systems

Notice that these ratings indicate the maximum speed at which the chip is capable of running. You could use a 66MHz-rated chip in place of the 50MHz- or 40MHz-rated parts with no problem, although the chip will run only at the slower speeds. The actual speed of the chip is double the motherboard clock frequency. When the 40MHz DX2/OverDrive chip is installed in a 16MHz 486SX system, for example, the chip functions only at 32MHz—exactly double the motherboard speed. Intel originally stated that no 100MHz DX2/OverDrive chip would be available for 50MHz systems—which technically has not been true because the DX4 could be set to run in a clock-doubled mode and used in a 50MHz motherboard (see the discussion of the DX4 processor in this section).

The only part of the DX2 chip that doesn't run at double speed is the bus interface unit, a region of the chip that handles I/O between the CPU and the outside world. By translating between the differing internal and external clock speeds, the bus interface unit makes speed doubling transparent to the rest of the system. The DX2 appears to the rest of the system to be a regular 486DX chip, but one that seems to execute instructions twice as fast.

DX2/OverDrive chips are based on the 0.8-micron circuit technology that was first used in the 50MHz 486DX. The DX2 contains 1.2 million transistors in a three-layer form. The internal 8KB cache, integer, and floating-point units all run at double speed. External communication with the PC runs at normal speed to maintain compatibility.

Besides upgrading existing systems, one of the best parts of the DX2 concept was the fact that system designers could introduce very fast systems by using cheaper motherboard designs, rather than the more costly designs that would support a straight high-speed clock. Therefore, a 50MHz 486DX2 system was much less expensive than a straight 50MHz 486DX system. The system board in a 486DX-50 system operates at a true 50MHz. The 486DX2 CPU in a 486DX2-50 system operates internally at 50MHz, but the motherboard operates at only 25MHz.

You might be thinking that a true 50MHz DX processor-based system still would be faster than a speed-doubled 25MHz system, and this generally is true. But, the differences in speed actually are very slight—a real testament to the integration of the 486 processor and especially to the cache design.

When the processor has to go to system memory for data or instructions, for example, it must do so at the slower motherboard operating frequency (such as 25MHz). Because the 8KB internal cache of the 486DX2 has a hit rate of 90%–95%, however, the CPU must access system memory only 5%–10% of the time for memory reads. Therefore, the performance of the DX2 system can come very close to that of a true 50MHz DX system and cost much less. Even though the motherboard runs at only 33.33MHz, a system with a DX2 66MHz processor ends up being faster than a true 50MHz DX system, especially if the DX2 system has a good L2 cache.

Many 486 motherboard designs also include a secondary cache that is external to the cache integrated into the 486 chip. This external cache allows for much faster access when the 486 chip calls for external-memory access. The size of this external cache can vary anywhere from 16KB to 512KB or more. When you add a DX2 processor, an external cache is even more important for achieving the greatest performance gain. This cache greatly reduces the wait states the processor must add when writing to system memory or when a read causes an internal cache miss. For this reason, some systems perform better with the DX2/OverDrive processors than others, usually depending on the size and efficiency of the external-memory cache system on the motherboard. Systems that have no external cache still enjoy a near-doubling of CPU performance, but operations that involve a great deal of memory access are slower.

Although the standard DX4 technically was not sold as a retail part, it could be purchased from several vendors, along with the 3.3V voltage adapter needed to install the chip in a 5V socket. These adapters have jumpers that enable you to select the DX4 clock multiplier and set it to 2x, 2.5x, or 3x mode. In a 50MHz DX system, you could install a DX4/voltage-regulator combination set in 2x mode for a motherboard speed of 50MHz and a processor speed of 100MHz! Although you might not be able to take advantage of certain VL-Bus adapter cards, you will have one of the fastest 486-class PCs available.

Intel also sold a special DX4 OverDrive processor that included a built-in voltage regulator and heatsink that are specifically designed for the retail market. The DX4 OverDrive chip is essentially the same as the standard 3.3V DX4 with the main exception that it runs on 5V because it includes an on-chip regulator. Also, the DX4 OverDrive chip runs only in the tripled speed mode, and not the 2x or 2.5x modes of the standard DX4 processor. Intel OverDrive products were discontinued several years ago, as were third-party equivalents.

Pentium OverDrive for 486SX2 and DX2 Systems

The Pentium OverDrive Processor became available in 1995. An OverDrive chip for 486DX4 systems had been planned, but poor marketplace performance of the SX2/DX2 chip resulted in it never seeing the light of day. One thing to keep in mind about the 486 Pentium OverDrive chip is that although it was intended primarily for SX2 and DX2 systems, it should work in any upgradeable 486SX or DX system that has a Socket 2 or Socket 3. If you want to install one in an older system, you can check Intel's online upgrade guide for compatibility, located at <http://support.intel.com/support/processors/overdrive/>.

The Pentium OverDrive processor is designed for systems that have a processor socket that follows the Intel Socket 2 specification. This processor also works in systems that have a Socket 3 design, although you should ensure that the voltage is set for 5V rather than 3.3V. The Pentium OverDrive chip includes a 32KB internal L1 cache and the same superscalar (multiple instruction path) architecture of the real Pentium chip. Besides a 32-bit Pentium core, these processors feature increased clock-speed operation due to internal clock multiplication and incorporate an internal write-back cache (standard with the Pentium). If the motherboard supports the write-back cache function, increased performance is realized. Unfortunately, most motherboards, especially older ones with the Socket 2 design, support only write-through cache.

Most tests of these OverDrive chips show them to be only slightly ahead of the DX4-100 and behind the DX4-120 and true Pentium 60, 66, or 75. Based on the relative affordability of low-end "real" Pentiums (in their day), it was hard not to justify making the step up to a Pentium system.

AMD 486 (5x86)

AMD made a line of 486-compatible chips that installed into standard 486 motherboards. In fact, AMD made the fastest 486 processor available, which it called the Am5x86(TM)-P75. The name was a little misleading because the 5x86 part made some people think that this was a fifth-generation Pentium-type processor. In reality, it was a fast clock-multiplied (4x clock) 486 that ran at four times the speed of the 33MHz 486 motherboard you plugged it into.

The 5x86 offered high-performance features such as a unified 16KB write-back cache and 133MHz core clock speed; it was approximately comparable to a Pentium 75, which is why it was denoted with a P-75 in the part number. It was the ideal choice for cost-effective 486 upgrades, where changing the motherboard is difficult or impossible.

Not all 486 motherboards support the 5x86. The best way to verify that your motherboard supports the chip is by checking with the documentation that came with the board. Look for keywords such as "Am5X86," "AMD-X5," "clock-quadrupled," "133MHz," or other similar wording. Another good way to determine whether your motherboard supports the AMD 5x86 is to look for it in the listed models on AMD's website.

There are a few things to note when installing a 5x86 processor into a 486 motherboard:

- *The operating voltage for the 5x86 is 3.45V +/- 0.15V.* Not all motherboards have this setting, but most that incorporate a Socket 3 design should. If your 486 motherboard is a Socket 1 or 2 design, you cannot use the 5x86 processor directly. The 3.45V processor does not operate in a 5V socket and can be damaged. To convert a 5V motherboard to 3.45V, processors with adapters could be purchased from several vendors, such as Kingston, PowerLeap, and Evergreen. These companies and others sold the 5x86 complete with a voltage regulator adapter attached in an easy-to-install package. These versions are ideal for older 486 motherboards that don't have a Socket 3 design and might still be available in the surplus or closeout market. If not, you can create your own upgrade kit with a processor, voltage adapter, and heatsink/fan.
- *It is generally better to purchase a new motherboard, processor, and RAM than to buy one of these adapters.* Buying a new motherboard is also better than using an adapter because the older BIOS might not understand the requirements of the processor as far as speed is concerned. BIOS updates often are required with older boards.
- *Most Socket 3 motherboards have jumpers, enabling you to set the voltage manually.* Some boards don't have jumpers, but have voltage autodetect instead. These systems check the VOLDET pin (pin S4) on the microprocessor when the system is powered on.
- *The VOLDET pin is tied to ground (Vss) internally to the microprocessor.* If you cannot find any jumpers for setting voltage, you can check the motherboard as follows: Switch the PC off, remove the microprocessor, connect pin S4 to a Vss pin on the ZIF socket, power on, and check any Vcc pin with a voltmeter. This should read 3.45 ([pm] 0.15) volts. See the previous section on CPU sockets for the pinout.
- *The 5x86 requires a 33MHz motherboard speed, so be sure the board is set to that frequency.* The 5x86 operates at an internal speed of 133MHz. Therefore, the jumpers must be set for "clock-quadrupled" or "4x clock" mode. By setting the jumpers correctly on the motherboard, the CLKMUL pin (pin R17) on the processor will be connected to ground (Vss). If there is no 4x clock setting, the standard DX2 2x clock setting should work.
- *Some motherboards have jumpers that configure the internal cache in either write-back (WB) or write-through (WT) mode.* They do this by pulling the WB/WT pin (pin B13) on the microprocessor to logic High (Vcc) for WB or to ground (Vss) for WT. For best performance, configure your system in WB mode; however, reset the cache to WT mode if problems running applications occur or the floppy drive doesn't work right (DMA conflicts).
- *The 5x86 runs hot, so a heatsink is required.* It normally must have a fan, and most upgrade kits included a fan.

In addition to the 5x86, the AMD-enhanced 486 product line included 80MHz; 100MHz; and 120MHz CPUs. These are the A80486DX2-80SV8B (40MHz×2), A80486DX4-100SV8B (33MHz×3), and A80486DX4-120SV8B (40MHz×3).

Cyrix/TI 486

The Cyrix 486DX2/DX4 processors were available in 100MHz, 80MHz, 75MHz, 66MHz, and 50MHz versions. Similar to the AMD 486 chips, the Cyrix versions are fully compatible with Intel's 486 processors and work in most 486 motherboards.

The Cx486DX2/DX4 incorporates an 8KB write-back cache, an integrated floating-point unit, advanced power management, and SMM, and was available in 3.3V versions.

Note

TI originally made all the Cyrix-designed 486 processors, and under the agreement it also sold them under the TI name. They are essentially the same as the Cyrix chips.

P5 (586) Fifth-Generation Processors

After the fourth-generation chips such as the 486, Intel and other chip manufacturers went back to the drawing board to come up with new architectures and features that they would later incorporate into what they called fifth-generation chips. This section defines the fifth-generation processors from Intel, AMD, and others.

Pentium Processors

On October 19, 1992, Intel announced that the fifth generation of its compatible microprocessor line (codenamed P5) would be named the Pentium processor rather than the 586, as everybody had assumed. Calling the new chip the 586 would have been natural, but Intel discovered that it could not trademark a number designation, and the company wanted to prevent other manufacturers from using the same name for any clone chips they might develop. The actual Pentium chip shipped on March 22, 1993. Systems that used these chips were only a few months behind.

The Pentium is fully compatible with previous Intel processors, but it differs from them in many ways. At least one of these differences is revolutionary: The Pentium features twin data pipelines, which enable it to execute two instructions at the same time. The 486 and all preceding chips can perform only a single instruction at a time. Intel calls the capability to execute two instructions at the same time superscalar technology. This technology provides additional performance compared with the 486.

With superscalar technology, the Pentium can execute many instructions at a rate of two instructions per cycle. Superscalar architecture usually is associated with high-output RISC chips. The Pentium is one of the first CISC chips to be considered superscalar. The Pentium is almost like having two 486 chips under the hood. Table 3.25 shows the Pentium processor specifications.

Table 3.25 Pentium Processor Specifications

Introduced	March 22, 1993 (first generation); March 7, 1994 (second generation)
Maximum rated speeds	60/66MHz (first generation); 75/90/100/120/133/150/166/200MHz (second generation)
CPU clock multiplier	1x (first generation); 1.5x–3x (second generation)
Register size	32-bit
External data bus	64-bit
Memory address bus	32-bit
Maximum memory	4GB
Integral-cache size	8KB code; 8KB data

Table 3.25 Continued

Integral-cache type	Two-way set associative; write-back data
Burst-mode transfers	Yes
Number of transistors	3.1 million (first generation); 3.3 million (second generation)
Circuit size	0.8 micron (60/66MHz); 0.6 micron (75MHz–100MHz); 0.35 micron (120MHz and up)
External package	273-pin PGA; 296-pin SPGA; tape carrier
Math coprocessor	Built-in FPU
Power management	SMM; enhanced in second generation
Operating voltage	5V (first generation); 3.465V, 3.3V, 3.1V, 2.9V (second generation)

PGA = Pin grid array

SPGA = Staggered pin grid array

The two instruction pipelines within the chip are called the *u*- and *v*-pipes. The *u*-pipe, which is the primary pipe, can execute all integer and floating-point instructions. The *v*-pipe is a secondary pipe that can execute only simple integer instructions and certain floating-point instructions. The process of operating on two instructions simultaneously in the different pipes is called *pairing*. Not all sequentially executing instructions can be paired, and when pairing is not possible, only the *u*-pipe is used. To optimize the Pentium's efficiency, you can recompile software to enable more instructions to be paired.

The Pentium processor has a branch target buffer (BTB), which employs a technique called branch prediction. It minimizes stalls in one or more of the pipes caused by delays in fetching instructions that branch to nonlinear memory locations. The BTB attempts to predict whether a program branch will be taken and then fetches the appropriate instructions. The use of branch prediction enables the Pentium to keep both pipelines operating at full speed. Figure 3.38 shows the internal architecture of the Pentium processor.

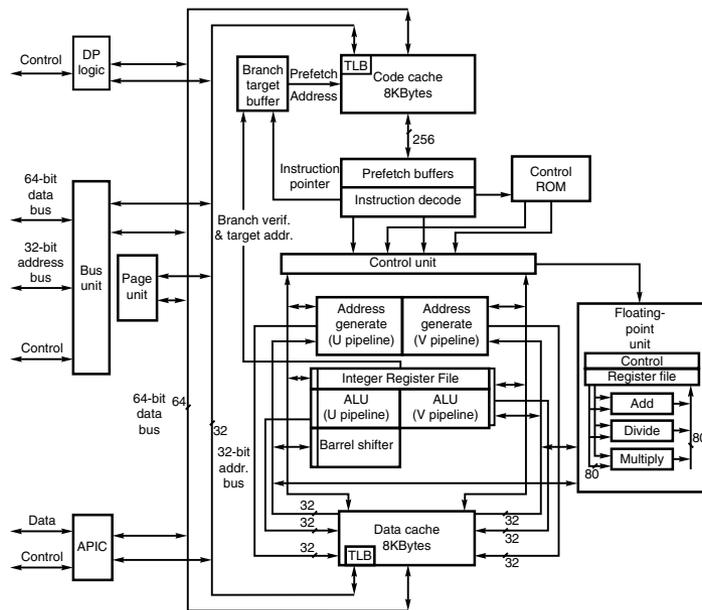


Figure 3.38 Pentium processor internal architecture.

The Pentium has a 32-bit address bus width, giving it the same 4GB memory-addressing capabilities as the 386DX and 486 processors. But the Pentium expands the data bus to 64 bits, which means it can move twice as much data into or out of the CPU, compared with a 486 of the same clock speed. The 64-bit data bus requires that system memory be accessed 64 bits wide, so each bank of memory is 64 bits.

On most Pentium-based motherboards, memory is installed via SIMMs or DIMMs. SIMMs are available in 8-bit-wide and 32-bit-wide versions, whereas DIMMs are 64 bits wide. In addition, versions are available with additional bits for parity or error correcting code (ECC) data. Most Pentium systems use the 32-bit-wide SIMMs—two of these SIMMs per bank of memory. Most Pentium motherboards have at least four of these 32-bit SIMM sockets, providing for a total of two banks of memory. Later Pentium systems and most Pentium II systems still in use today use DIMMs, which are 64 bits wide—just like the processor's external data bus, so only one DIMM is used per bank. This makes installing or upgrading memory much easier because DIMMs can go in one at a time and don't have to be matched up in pairs.

► See "SIMMs, DIMMs, and RIMMs," p. 492, and "Memory Banks," p. 512.

Even though the Pentium has a 64-bit data bus that transfers information 64 bits at a time into and out of the processor, the Pentium has only 32-bit internal registers. As instructions are being processed internally, they are broken down into 32-bit instructions and data elements and processed in much the same way as in the 486. Some people thought that Intel was misleading them by calling the Pentium a 64-bit processor, but 64-bit transfers do indeed take place. Internally, however, the Pentium has 32-bit registers that are fully compatible with the 486.

The Pentium has two separate internal 8KB caches, compared with a single 8KB or 16KB cache in the 486. The cache-controller circuitry and the cache memory are embedded in the CPU chip. The cache mirrors the information in normal RAM by keeping a copy of the data and code from different memory locations. The Pentium cache also can hold information to be written to memory when the load on the CPU and other system components is less. (The 486 makes all memory writes immediately.)

The separate code and data caches are organized in a two-way set associative fashion, with each set split into lines of 32 bytes each. Each cache has a dedicated translation lookaside buffer (TLB) that translates linear addresses to physical addresses. You can configure the data cache as write-back or write-through on a line-by-line basis. When you use the write-back capability, the cache can store write operations and reads, further improving performance over read-only write-through mode. Using write-back mode results in less activity between the CPU and system memory—an important improvement because CPU access to system memory is a bottleneck on fast systems. The code cache is an inherently write-protected cache because it contains only execution instructions and not data, which is updated. Because burst cycles are used, the cache data can be read or written very quickly.

Systems based on the Pentium can benefit greatly from secondary processor caches (L2), which usually consist of up to 512KB or more of extremely fast (15ns or less) SRAM chips. When the CPU fetches data that is not already available in its internal processor (L1) cache, wait states slow the CPU. If the data already is in the secondary processor cache, however, the CPU can go ahead with its work without pausing for wait states.

The Pentium uses a Bipolar Complementary Metal-Oxide Semiconductor (BiCMOS) process and super-scalar architecture to achieve the high level of performance expected from the chip. BiCMOS adds about 10% to the complexity of the chip design, but adds about 30%–35% better performance without a size or power penalty.

All 75MHz and faster Pentium processors are SL enhanced—they incorporate the SMM to provide full control of power-management features, which helps reduce power consumption. The second-generation Pentium processors (75MHz and faster) incorporate a more advanced form of SMM that includes processor clock control. This enables you to throttle the processor up or down to control power use. You can even stop the clock with these more advanced Pentium processors, putting the processor in a state of suspension that requires very little power. The second-generation Pentium processors run on 3.3V power (instead of 5V), reducing power requirements and heat generation even further.

Many Pentium motherboards supply either 3.465V or 3.3V. The 3.465V setting is called VRE (voltage reduced extended) by Intel and is required by some versions of the Pentium, particularly some of the 100MHz versions. The standard 3.3V setting is called STD (standard), which most of the second-generation Pentiums use. STD voltage means anything in a range from 3.135V to 3.465V with 3.3V nominal. Additionally, a special 3.3V setting called VR (voltage reduced) reduces the range from 3.300V to 3.465V with 3.38V nominal. Some of the processors require this narrower specification, which most motherboards provide. Here is a summary:

Voltage Specification	Nominal	Tolerance	Minimum	Maximum
STD (standard)	3.30V	±0.165	3.135V	3.465V
VR (voltage reduced)	3.38V	±0.083	3.300V	3.465V
VRE (VR extended)	3.50V	±0.100	3.400V	3.600V

For even lower power consumption, Intel introduced special Pentium processors with voltage reduction technology in the 75 to 266MHz family; the processors were intended for mobile computer applications. They did not use a conventional chip package and were instead mounted using a new format called tape carrier packaging (TCP). The tape carrier packaging does not encase the chip in ceramic or plastic as with a conventional chip package, but instead covers the actual processor die directly with a thin, protective plastic coating. The entire processor is less than 1mm thick, or about half the thickness of a dime, and weighs less than 1 gram. They were sold to system manufacturers in a roll that looks very much like a filmstrip.

The TCP processor is directly affixed (soldered) to the motherboard by a special machine, resulting in a smaller package, lower height, better thermal transfer, and lower power consumption. Special solder plugs on the circuit board located directly under the processor draw heat away and provide better cooling in the tight confines of a typical notebook or laptop system—no cooling fans are required. For more information on mobile processors and systems, see the chapter “Portable PCs” included on the disc accompanying this book.

The Pentium, like the 486, contains an internal math coprocessor or FPU. The FPU in the Pentium was rewritten to perform significantly better than the FPU in the 486 yet still be fully compatible with the 486 and 387 math coprocessors. The Pentium FPU is estimated to be two to as much as ten times faster than the FPU in the 486. In addition, the two standard instruction pipelines in the Pentium provide two units to handle standard integer math. (The math coprocessor handles only more complex calculations.) Other processors, such as the 486, have only a single-standard execution pipe and one integer math unit. Interestingly, the Pentium FPU contains a flaw that received widespread publicity. See the discussion in the section “Pentium Defects,” later in this chapter.

First-Generation Pentium Processors

The Pentium was offered in three basic designs, each with several versions. The first-generation design came in 60MHz and 66MHz processor speeds. This design used a 273-pin PGA form factor and ran on 5V power. In this design, the processor ran at the same speed as the motherboard—in other words, a 1x clock was used.

The first-generation Pentium was created through a 0.8-micron BiCMOS process. Unfortunately, this process, combined with the 3.1 million transistor count, resulted in a die that was overly large and complicated to manufacture. As a result, reduced yields kept the chip in short supply; Intel could not make them fast enough. The 0.8-micron process was criticized by other manufacturers, including Motorola and IBM, which had been using 0.6-micron technology for their most advanced chips. The huge die and 5V operating voltage caused the 66MHz versions to consume up to an incredible 3.2 amps or 16 watts of power, resulting in a tremendous amount of heat and problems in some systems that did not employ conservative design techniques. Fortunately, adding a fan to the processor solved most cooling problems, as long as the fan kept running.

Much of the criticism leveled at Intel for the first-generation Pentium was justified. Some people realized that the first-generation design was just that; they knew that new Pentium versions, made in a more advanced manufacturing process, were coming. Many of those people advised against purchasing any Pentium system until the second-generation version became available.

Tip

A cardinal rule of computing is never buy the first generation of any processor. Although you can wait forever because something better always will be on the horizon, a little waiting is worthwhile in many cases.

Those who purchased first-generation Pentiums still had a way out, however. As with previous 486 systems, Intel released OverDrive upgrade chips that effectively doubled the processor speed of the Pentium 60 or 66. These are a single-chip upgrade, meaning they replace the existing CPU. Because subsequent Pentiums are incompatible with the Pentium 60/66 Socket 4 arrangement, these OverDrive chips and comparable upgrades available from some third-party sources were the only way to upgrade an existing first-generation Pentium without replacing the motherboard.

Generally, it was better to consider a complete motherboard replacement, which would accept a newer design processor that would potentially be many times faster, than to upgrade using just an OverDrive processor, that might only be twice as fast.

Second-Generation Pentium Processors

Intel announced the second-generation Pentium on March 7, 1994. This processor was introduced in 90MHz and 100MHz versions, with a 75MHz version not far behind. Eventually, 120MHz, 133MHz, 150MHz, 166MHz, and 200MHz versions were also introduced. The second-generation Pentium uses 0.6-micron (75/90/100MHz) BiCMOS technology to shrink the die and reduce power consumption. The newer, faster 120MHz (and higher) second-generation versions incorporate an even smaller die built on a 0.35-micron BiCMOS process. These smaller dies are not changed from the 0.6-micron versions; they are basically a photographic reduction of the P54C die. The die for the Pentium is shown in Figure 3.39. Additionally, these new processors run on 3.3V power. The 100MHz version consumes a maximum of 3.25 amps of 3.3V power, which equals only 10.725 watts. Further up the scale, the 150MHz chip uses 3.5 amps of 3.3V power (11.6 watts); the 166MHz unit draws 4.4 amps (14.5 watts); and the 200MHz processor uses 4.7 amps (15.5 watts).

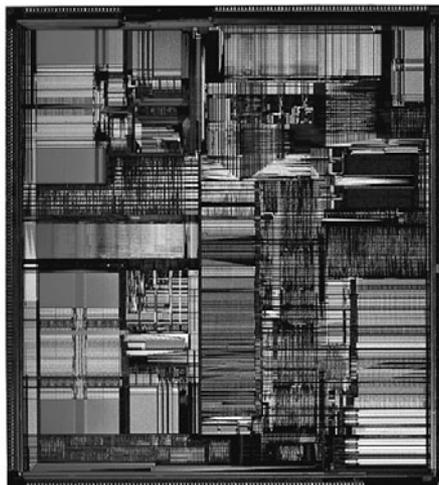


Figure 3.39 Pentium processor die. *Photograph used by permission of Intel Corporation.*

The second-generation Pentium processors use a 296-pin SPGA form factor that is physically incompatible with the first-generation versions. The only way to upgrade from the first generation to the second was to replace the motherboard. The second-generation Pentium processors also have 3.3 million transistors—more than the earlier chips. The extra transistors exist because additional clock-control SL enhancements were added, along with an on-chip advanced programmable interrupt controller (APIC) and dual-processor interface.

The APIC and dual-processor interfaces are responsible for orchestrating dual-processor configurations in which two second-generation Pentium chips can process on the same motherboard simultaneously. Many of the Pentium motherboards designed for file servers come with dual Socket 7 specification sockets, which fully support the multiprocessing capability of the new chips. Software support for what usually is called symmetric multiprocessing (SMP) was integrated into operating systems such as Windows NT and OS/2.

The second-generation Pentium processors use clock-multiplier circuitry to run the processor at speeds faster than the bus. The 150MHz Pentium processor, for example, can run at 2.5 times the bus frequency, which normally is 60MHz. The 200MHz Pentium processor can run at a 3x clock in a system using a 66MHz bus speed.

Virtually all Pentium motherboards had three speed settings: 50MHz, 60MHz, and 66MHz. Pentium chips were available with a variety of internal clock multipliers that caused the processor to operate at various multiples of these motherboard speeds. Refer to Table 3.7 for a list of the speeds of Pentium processors and motherboards.

The core-to-bus frequency ratio or clock multiplier is controlled in a Pentium processor by two pins on the chip labeled BF1 and BF2. Table 3.26 shows how the state of the BFx pins affects the clock multiplication in the Pentium processor.

Table 3.26 Pentium BFx Pins and Clock Multipliers

BF1	BF2	Clock Multiplier	Bus Speed (MHz)	Core Speed (MHz)
0	1	3x	66	200
0	1	3x	60	180
0	1	3x	50	150
0	0	2.5x	66	166
0	0	2.5x	60	150
0	0	2.5x	50	125
1	0	2x/4x	66	133/266 ¹
1	0	2x	60	120
1	0	2x	50	100
1	1	1.5x/3.5x	66	100/233 ¹
1	1	1.5x	60	90
1	1	1.5x	50	75

1. The 233MHz and 266MHz processors have modified the 1.5x and 2x multipliers to 3.5x and 4x, respectively.

Not all chips support all the bus frequency (BF) pins or combinations of settings. In other words, some of the Pentium processors operate only at specific combinations of these settings or might even be fixed at one particular setting. Many of the later Pentium motherboards included jumpers or

switches that enabled you to control the BF pins and, therefore, alter the clock-multiplier ratio within the chip. In theory, you could run a 75MHz-rated Pentium chip at 133MHz by changing jumpers on the motherboard. This is called overclocking and is discussed in the section “Overclocking,” earlier in this chapter. What Intel has done to discourage overclockers in its most recent Pentiums is discussed near the end of the “Processor Manufacturing” section of this chapter.

Intel also offered a single-chip OverDrive upgrade for second-generation Pentiums. These OverDrive chips are fixed at a 3x multiplier; they replace the existing Socket 5 or 7 CPU, increase processor speed up to 200MHz (with a 66MHz motherboard speed), and add MMX capability. Simply stated, a Pentium 100, 133, or 166 system equipped with the OverDrive chip has a processor speed of 200MHz. Perhaps the best feature of these Pentium OverDrive chips is that they incorporate MMX technology to improve multimedia application performance.

Pentium-MMX Processors

A third generation of Pentium processors (codenamed P55C) was released in January 1997, and incorporates what Intel calls MMX technology into the second-generation Pentium design (see Figure 3.40). These Pentium-MMX processors were manufactured in clock rates of 66/166MHz, 66/200MHz, and 66/233MHz and in a mobile-only version, which is 66/266MHz. The MMX processors have a lot in common with other second-generation Pentiums, including superscalar architecture, multiprocessor support, on-chip local APIC controller, and power-management features. New features include a pipelined MMX unit, 16KB code, write-back cache (versus 8KB in earlier Pentiums), and 4.5 million transistors. Pentium-MMX chips are produced on an enhanced 0.35-micron CMOS silicon process that allows for a lower 2.8V voltage level. The newer mobile 233MHz and 266MHz processors are built on a 0.25-micron process and run on only 1.8V. With this newer technology, the 266 processor actually uses less power than the non-MMX 133.

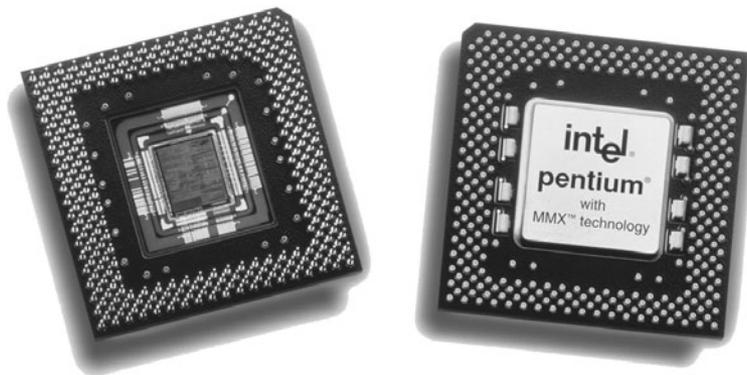


Figure 3.40 Pentium MMX. The left side shows the underside of the chip with the cover plate removed exposing the processor die. *Photograph used by permission of Intel Corporation.*

To use the Pentium-MMX, the motherboard must be capable of supplying the lower (2.8V or less) voltage these processors use. To enable a more universal motherboard solution with respect to these changing voltages, Intel developed the Socket 7 with VRM. The VRM is a socketed module that plugs in next to the processor and supplies the correct voltage. Because the module is easily replaced, reconfiguring a motherboard to support any of the voltages required by the newer Pentium processors is easy.

Of course, lower voltage is nice, but MMX is what this chip is really all about. MMX incorporates a process Intel calls single instruction multiple data (SIMD), which enables one instruction to perform

the same function on many pieces of data. Fifty-seven new instructions designed specifically to handle video, audio, and graphics data have been added to the chip.

Pentium Defects

Probably the most famous processor bug in history is the now legendary flaw in the Pentium FPU. It has often been called the FDIV bug because it affects primarily the FDIV (floating-point divide) instruction, although several other instructions that use division are also affected. Intel officially refers to this problem as Errata No. 23, titled “Slight precision loss for floating-point divides on specific operand pairs.” The bug has been fixed in the D1 or later steppings of the 60/66MHz Pentium processors, as well as the B5 and later steppings of the 75/90/100MHz processors. The 120MHz and higher processors are manufactured from later steppings, which do not include this problem. Tables listing all the variations of Pentium processors and steppings and how to identify them appear later in this chapter.

This bug caused a tremendous fervor when it first was reported on the Internet by mathematician Thomas R. Nicely of Lynchburg College in Virginia in October 1994. Within a few days, news of the defect had spread nationwide, and even people who did not have computers had heard about it. The Pentium incorrectly performed floating-point division calculations with certain number combinations, with errors anywhere from the third digit on up.

By the time the bug was publicly discovered outside of Intel, the company had already incorporated the fix into the next stepping of both the 60/66MHz and the 75/90/100MHz Pentium processor, along with the other corrections Intel had made.

After the bug was made public and Intel admitted to already knowing about it, a fury erupted. As people began checking their spreadsheets and other math calculations, many discovered they had also encountered this problem and did not know it. Others who had not encountered the problem had their faith in the core of their PCs very shaken. People had come to put so much trust in the PC that they had a hard time coming to terms with the fact that it might not even be capable of doing math correctly!

One interesting result of the fervor surrounding this defect is that people are less likely to implicitly trust their PCs and are therefore doing more testing and evaluating of important results. The bottom line is that if your information and calculations are important enough, you should implement some results tests. Several math programs were found to have problems. For example, a bug was discovered in the yield function of Excel 5.0 that some were attributing to the Pentium processor. In this case, the problem turned out to be the software (which has been corrected in versions 5.0c and later).

Intel finally decided that in the best interest of the consumer and its public image, it would begin a lifetime replacement warranty on the affected processors. Therefore, if you ever encounter one of the Pentium processors with the Errata 23 floating-point bug, Intel will replace the processor with an equivalent one without this problem.

If you are still using a Pentium-based system and wonder whether you might have a system affected by this bug, visit the Intel “FDIV Replacement Program” page at <http://support.intel.com/support/processors/pentium/fdiv/>. Here you can find information on how to determine whether your processor is affected and how to obtain a free replacement for an affected processor. As long as Intel receives the original CPU back within a specified amount of time, there will be no charges to you. Intel has indicated that these defective processors are destroyed and will not be remarketed or resold in another form.

Testing for the FPU Bug

Testing a Pentium for this bug is relatively easy. All you have to do is execute one of the test division cases cited here and see whether your answer compares to the correct result.

The division calculation can be done in a spreadsheet (such as Lotus 1-2-3, Microsoft Excel, or any other), the Microsoft Windows built-in calculator, or any other calculating program that uses the FPU. Make sure that for the purposes of this test the FPU has not been disabled. That typically requires some special command or setting specific to the application and, of course, ensures that the test comes out correct, regardless of whether the chip is flawed.

The most severe Pentium floating-point errors occur as early as the third significant digit of the result. Here is an example of one of the more severe instances of the problem:

$$962,306,957,033 / 11,010,046 = 87,402.6282027341 \text{ (correct answer)}$$

$$962,306,957,033 / 11,010,046 = 87,399.5805831329 \text{ (flawed Pentium)}$$

Note

Note that your particular calculator program might not show the answer to the number of digits shown here. Most spreadsheet programs limit displayed results to 13 or 15 significant digits.

As you can see in the previous case, the error turns up in the third most significant digit of the result. In an examination of more than 5,000 integer pairs in the 5- to 15-digit range found to produce Pentium floating-point division errors, errors beginning in the sixth significant digit were the most likely to occur.

Several workarounds are available for this bug, but they extract a performance penalty. Because Intel has agreed to replace any Pentium processor with this flaw under a lifetime warranty replacement program, the best workaround is a free replacement or an upgrade to a more modern system.

Power Management Bugs

Starting with the second-generation Pentium processors, Intel added functions that enable these CPUs to be installed in energy-efficient systems. These are usually called *Energy Star systems* because they meet the specifications imposed by the EPA Energy Star program, but they are also unofficially called *green PCs* by many users.

Unfortunately, there have been several bugs with respect to these functions, causing them to either fail or be disabled. These bugs are in some of the functions in the power-management capabilities accessed through SMM. These problems are applicable only to the second-generation 75/90/100MHz processors because the first-generation 60/66MHz processors do not have SMM or power-management capabilities, and all higher-speed (120MHz and up) processors have the bugs fixed.

Most of the problems are related to the STPCLK# pin and the HALT instruction. If this condition is invoked by the chipset, the system will hang. For most systems, the only workaround for this problem is to disable the power-saving modes, such as suspend or sleep. Unfortunately, this means that your green PC won't be so green anymore! The best way to repair the problem is to replace the processor with a later stepping version that does not have the bug. These bugs affect the B1 stepping version of the 75/90/100MHz Pentiums, and they were fixed in the B3 and later stepping versions.

Pentium Processor Models and Steppings

We know that like software, no processor is truly ever perfect. From time to time, the manufacturers gather up what problems they have found and put into production a new stepping, which consists of a new set of masks that incorporate the corrections. Each subsequent stepping is better and more refined than the previous ones. Although no microprocessor is ever perfect, they come closer to perfection with each stepping. In the life of a typical microprocessor, a manufacturer might go through half a dozen or more such steppings.

See *Upgrading and Repairing PCs, Tenth Anniversary Edition*, which is included on the disc, for tables showing the Pentium processor steppings and revisions. This information is also available online from Intel via its website.

To determine the specifications of a given processor, you must look up the S-spec number in the table of processor specifications. To find your S-spec number, you have to read it off the chip directly. It can be found printed on both the top and bottom of the chip. If your heatsink is glued on, remove the chip and heatsink from the socket as a unit and read the numbers from the bottom of the chip. Then, you can look up the S-spec number in the Specification Guide Intel publishes (via its website); it tells you the specifications of that particular processor. Intel is introducing new chips all the time, so visit its website and search for the Pentium processor “Quick Reference Guide” in the developer portion of its site. There you will find a complete listing of all current processor specifications by S-spec number.

One interesting item to note is that several subtly different voltages are required by different Pentium processors. Table 3.27 summarizes the various processors and their required voltages.

Table 3.27 Pentium Processor Voltages

Model	Stepping	Voltage Spec.	Voltage Range
1	—	Std.	4.75V–5.25V
1	—	5V1	4.90V–5.25V
1	—	5V2	4.90V–5.40V
1	—	5V3	5.15V–5.40V
2+	B1-B5	Std.	3.135V–3.465V
2+	C2+	Std.	3.135V–3.600V
2+	—	VR	3.300V–3.465V
2+	B1-B5	VRE	3.45V–3.60V
2+	C2+	VRE	3.40V–3.60V
4+	—	MMX	2.70V–2.90V
4	3	Mobile	2.285V–2.665V
4	3	Mobile	2.10V–2.34V
8	1	Mobile	1.850V–2.150V
8	1	Mobile	1.665V–1.935V

Many of the newer Pentium motherboards have jumpers that allow for adjustments to the different voltage ranges. If you are having problems with a particular processor, it might not be matched correctly to your motherboard voltage output.

If you are purchasing an older, used Pentium system today, I recommend using only Model 2 (second-generation) or later version processors that are available in 75MHz or faster speeds. You should definitely get stepping C2 or later. Virtually all the important bugs and problems were fixed in the C2 and later releases. The newer Pentium processors have no serious bugs to worry about.

AMD-K5

The AMD-K5 is a Pentium-compatible processor developed by AMD and available as the PR75, PR90, PR100, PR120, PR133, PR166, and PR200. Because it is designed to be physically and functionally compatible, any motherboard that properly supports the Intel Pentium should support the AMD-K5. However, a BIOS upgrade might be required to properly recognize the AMD-K5. The K5 has the following features:

- 16KB instruction cache, 8KB write-back data cache
- Dynamic execution—branch prediction with speculative execution
- Five-stage, RISC-like pipeline with six parallel functional units

- High-performance floating-point unit
- Pin-selectable clock multiples of 1.5x, 1.75x, and 2x

The K5 is sold under the P-Rating system, which means that the number on the chip does not indicate true clock speed, only apparent speed when running certain applications.

Note that the actual clock speeds of several of these processors are not the same as their apparent rated speeds. For example, the PR-166 version actually runs at only 117 true MHz. Sometimes this can confuse the system BIOS, which might report the true speed rather than the P-Rating, which compares the chip against an Intel Pentium of that speed. AMD's assertion is that because of architecture enhancements over the Pentium, they do not need to run the same clock frequency to achieve that same performance. Even with such improvements, AMD marketed the K5 as a fifth-generation processor, just like the Pentium.

The AMD-K5 operates at 3.52V (VRE setting). Some older motherboards default to 3.3V, which is below specification for the K5 and could cause erratic operation. Because of the relatively low clock speeds and compatibility issues some users experienced with the K5, AMD replaced it with the K6 family of processors.

Intel P6 (686) Sixth-Generation Processors

The P6 (686) processors represent a new generation with features not found in the previous generation units. The P6 processor family began when the Pentium Pro was released in November 1995. Since then, Intel has released many other P6 chips, all using the same basic P6 core processor as the Pentium Pro. Table 3.28 shows the variations in the P6 family of processors.

Table 3.28 Intel P6 Processor Variations

Pentium Pro	Original P6 processor, includes 256KB, 512KB, or 1MB of full-core speed L2 cache
Pentium II	P6 with 512KB of half-core speed L2 cache
Pentium II Xeon	P6 with 512KB, 1MB, or 2MB of full-core speed L2 cache
Celeron	P6 with no L2 cache
Celeron-A	P6 with 128KB of on-die full-core speed L2 cache
Pentium III	P6 with SSE (MMX2), 512KB of half-core speed L2 cache
Pentium IIPE	P6 with 256KB of full-core speed L2 cache
Pentium IIIE	P6 with SSE (MMX2) plus 256KB or 512KB of full-core speed L2 cache
Pentium III Xeon	P6 with SSE (MMX2), 512KB, 1MB, or 2MB of full-core speed L2 cache

The main new feature in the fifth-generation Pentium processors was the superscalar architecture, in which two instruction execution units could execute instructions simultaneously in parallel. Later fifth-generation chips also added MMX technology to the mix, as well. So then what did Intel add in the sixth generation to justify calling it a whole new generation of chip? Besides many minor improvements, the real key features of all sixth-generation processors are Dynamic Execution and the Dual Independent Bus (DIB) architecture, plus a greatly improved superscalar design.

Dynamic Execution

Dynamic execution enables the processor to execute more instructions on parallel, so tasks are completed more quickly. This technology innovation is comprised of three main elements:

- *Multiple branch prediction.* Predict the flow of the program through several branches
- *Dataflow analysis.* Schedules instructions to be executed when ready, independent of their order in the original program
- *Speculative execution.* Increases the rate of execution by looking ahead of the program counter and executing instructions that are likely to be necessary

Dual Independent Bus

The other main P6 architecture feature is known as the Dual Independent Bus. This refers to the fact that the processor has two data buses: one for the system (motherboard) and the other just for cache. This enables the cache memory to run at speeds previously not possible.

Other Sixth-Generation Improvements

Finally, the P6 architecture upgrades the superscalar architecture of the P5 processors by adding more instruction execution units and by breaking down the instructions into special micro-ops. This is where the CISC instructions are broken down into more RISC commands. The RISC-level commands are smaller and easier for the parallel instruction units to execute more efficiently. With this design, Intel has brought the benefits of a RISC processor—high-speed dedicated instruction execution—to the CISC world. Note that the P5 had only two instruction units, whereas the P6 has at least six separate dedicated instruction units. It is said to be three-way superscalar because the multiple instruction units can execute up to three instructions in one cycle.

Other improvements in efficiency also are included in the P6 architecture: built-in multiprocessor support, enhanced error detection and correction circuitry, and optimization for 32-bit software.

Rather than just being a faster Pentium, the Pentium Pro, Pentium II/III, and other sixth-generation processors have many feature and architectural improvements. The core of the chip is very RISC-like, whereas the external instruction interface is classic Intel CISC. By breaking down the CISC instructions into several RISC instructions and running them down parallel execution pipelines, the overall performance is increased.

Compared to a Pentium at the same clock speed, the P6 processors are faster—as long as you're running 32-bit software. The P6 Dynamic Execution is optimized for performance primarily when running 32-bit software, such as Windows NT. If you are using 16-bit software, such as Windows 95 or 98 (which still operate part time in a 16-bit environment) and most older applications, the P6 does not provide as marked a performance improvement over similarly speed-rated Pentium and Pentium-MMX processors. That's because the Dynamic Execution capability is not fully exploited. Because of this, Windows NT/2000/XP often are regarded as the most desirable operating systems for use with Pentium Pro/II/III/Celeron processors. Although this is not exactly true (a Pentium Pro/II/III/Celeron runs fine under Windows 95/98), Windows NT/2000/XP does take better advantage of the P6's capabilities.

Note that it is really not so much the operating system but which applications you use. Software developers can take steps to gain the full advantages of the sixth-generation processors. This includes using modern compilers that can improve performance for all current Intel processors, writing 32-bit code where possible, and making code as predictable as possible to take advantage of the processor's Dynamic Execution multiple branch prediction capabilities.

Pentium Pro Processors

Intel's successor to the Pentium is called the Pentium Pro. The Pentium Pro was the first chip in the P6 or sixth-generation processor family. It was introduced in November 1995 and became widely available in 1996. The chip is a 387-pin unit that resides in Socket 8, so it is not pin compatible with earlier Pentiums. The chip is unique among processors because it is constructed in a multichip module (MCM) physical format, which Intel calls a dual cavity PGA package. Inside the 387-pin chip carrier are two dies. One contains the actual Pentium Pro processor (shown in Figure 3.41), and the other contains a 256KB (the Pentium Pro with 256KB cache is shown in Figure 3.42), 512KB, or 1MB L2 cache. The processor die contains 5.5 million transistors, the 256KB cache die contains 15.5 million transistors, and the 512KB cache die(s) have 31 million transistors each—for a potential total of nearly 68 million transistors in a Pentium Pro with 1MB of internal cache! A Pentium Pro with 1MB cache has two 512KB cache die and a standard P6 processor die (see Figure 3.43).

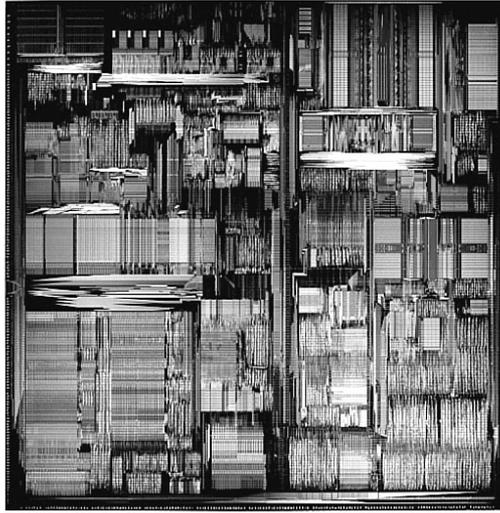


Figure 3.41 Pentium Pro processor die. *Photograph used by permission of Intel Corporation.*

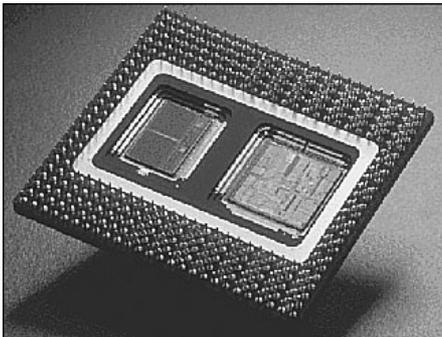


Figure 3.42 Pentium Pro processor with 256KB L2 cache (the cache is on the left side of the processor die). *Photograph used by permission of Intel Corporation.*

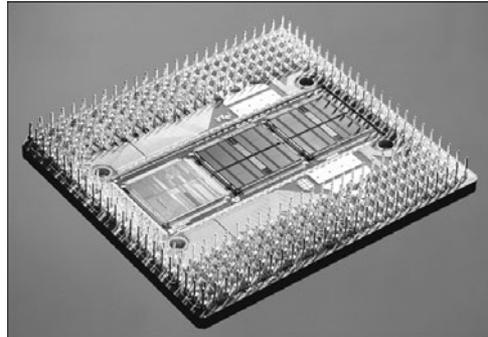


Figure 3.43 Pentium Pro processor with 1MB L2 cache (the cache is in the center and right portions of the die). *Photograph used by permission of Intel Corporation.*

The main processor die includes a 16KB split L1 cache with an 8KB two-way set associative cache for primary instructions and an 8KB four-way set associative cache for data.

Another sixth-generation processor feature found in the Pentium Pro is the DIB architecture, which addresses the memory bandwidth limitations of previous-generation processor architectures. Two buses make up the DIB architecture: the L2 cache bus (contained entirely within the processor package) and the processor-to-main memory system bus. The speed of the dedicated L2 cache bus on the Pentium Pro is equal to the full-core speed of the processor. This was accomplished by embedding the cache chips directly into the Pentium Pro package. The DIB processor bus architecture addresses processor-to-memory bus bandwidth limitations. It offers up to three times the performance bandwidth of the single-bus, “Socket 7” generation processors, such as the Pentium.

Table 3.29 shows Pentium Pro processor specifications, and Table 3.30 shows the specifications for each model within the Pentium Pro family because many variations exist from model to model.

Table 3.29 Pentium Pro Family Processor Specifications

Introduced	November 1995
Maximum rated speeds	150MHz, 166MHz, 180MHz, 200MHz
CPU clock	2x, 2.5x, 3x, 3.5x, 4x
Internal registers	32-bit
External data bus	64-bit
Memory address bus	36-bit
Addressable memory	64GB
Virtual memory	64TB
Integral L1 cache size	8KB code, 8KB data (16KB total)
Integrated L2 cache bus	64-bit, full-core speed
Socket/Slot	Socket 8
Physical package	387-pin dual cavity PGA
Package dimensions	2.46 (6.25cm)×2.66 (6.76cm)
Math coprocessor	Built-in FPU
Power management	SMM
Operating voltage	3.1V or 3.3V

Table 3.30 Pentium Pro Processor Specifications by Processor Model

<i>Pentium Pro Processor (200MHz) with 1MB Integrated Level 2 Cache</i>	
Introduction date	August 18, 1997
Clock speeds	200MHz (66MHz×3)
Number of transistors	5.5 million (0.35-micron process), plus 62 million in 1MB L2 cache (0.35-micron)
Cache memory	8K×2 (16KB) L1, 1MB core-speed L2
Die size	0.552" per side (14.0mm)
<i>Pentium Pro Processor (166/180/200MHz)</i>	
Introduction date	November 1, 1995
Clock speeds	200MHz (66MHz×3), 180MHz (60MHz×3), 166MHz (66MHz×2.5)
Number of transistors	5.5 million (0.35-micron process), plus 15.5 million in 256KB L2 cache (0.6-micron), or 31 million in 512KB L2 cache (0.35-micron)
Cache memory	8K×2 (16KB) L1, 256KB or 512KB core-speed L2
Die size	0.552" per side (14.0mm)
<i>Pentium Pro Processor (150MHz)</i>	
Introduction date	November 1, 1995
Clock speeds	150MHz (60MHz×2.5)
Number of transistors	5.5 million (0.6-micron process), plus 15.5 million in 256KB L2 cache (0.6-micron)
Cache memory	8K×2 (16KB) L1, 256KB core-speed L2
Die size	0.691" per side (17.6mm)

Performance comparisons on the iCOMP 2.0 Index rate a classic Pentium 200MHz at 142, whereas a Pentium Pro 200MHz scores 220. Just for comparison, note that a Pentium MMX 200MHz falls right about in the middle in regards to performance at 182. Keep in mind that using a Pentium Pro with any 16-bit software applications nullifies much of the performance gain shown by the iCOMP 2.0 rating.

Similar to the Pentium before it, the Pentium Pro runs clock multiplied on a 66MHz motherboard. The following table lists speeds for Pentium Pro processors and motherboards:

CPU Type/Speed	CPU Clock	Motherboard Speed
Pentium Pro 150	2.5x	60
Pentium Pro 166	2.5x	66
Pentium Pro 180	3x	60
Pentium Pro 200	3x	66

The integrated L2 cache is one of the really outstanding features of the Pentium Pro. By building the L2 cache into the CPU and getting it off the motherboard, the Pentium Pro can now run the cache at full processor speed rather than the slower 60MHz or 66MHz motherboard bus speed. In fact, the L2 cache features its own internal 64-bit back-side bus, which does not share time with the external 64-bit front-side bus used by the CPU. The internal registers and data paths are still 32-bit, as with the Pentium. By building the L2 cache into the system, motherboards can be cheaper because they no longer require separate cache memory. Some boards might still try to include cache memory in their designs, but the general consensus is that L3 cache (as it would be called) would offer less improvement with the Pentium Pro than with the Pentium. The incorporation of L2 cache is one of the most enduring legacies of the Pentium Pro because this feature has been incorporated into virtually every Intel and AMD processor built since, with the notable exception of the original Celeron.

One of the features of the built-in L2 cache is that multiprocessing is greatly improved. Rather than just SMP, as with the Pentium, the Pentium Pro supports a type of multiprocessor configuration called the Multiprocessor Specification (MPS 1.1). The Pentium Pro with MPS enables configurations of up to four processors running together. Unlike other multiprocessor configurations, the Pentium Pro avoids cache coherency problems because each chip maintains a separate L1 and L2 cache internally.

Pentium Pro-based motherboards were pretty much exclusively PCI and ISA bus-based, and Intel has produced its own chipsets for these motherboards. Because of the greater cooling and space requirements, Intel designed the new ATX motherboard form factor to better support the Pentium Pro and other future processors, such as the Pentium II/III/4. However, systems using the Pentium Pro use various types of motherboard form factors, including ATX, Baby-AT, and proprietary models.

- ▶▶ See "Motherboard Form Factors," p. 216, and "Sixth-Generation (P6 Pentium Pro/II/III Class) Chipsets," p. 264.

Four special VID pins are on the Pentium Pro processor. These pins can be used to support automatic selection of power supply voltage. Therefore, a Pentium Pro motherboard does not have voltage regulator jumper settings like most Pentium boards, which greatly eases the setup and integration of a Pentium Pro system. These pins are not actually signals, but are either an open circuit in the package or a short circuit to voltage. The sequence of opens and shorts defines the voltage the processor requires. In addition to allowing for automatic voltage settings, this feature was designed to support voltage specification variations on future Pentium Pro processors. The VID pins are named VID0 through VID3, and the definition of these pins is shown in Table 3.31. A 1 in this table refers to an open pin, and a 0 refers to a short to ground. The voltage regulators on the motherboard should supply the requested voltage or disable itself.

Table 3.31 Pentium Pro Voltage Identification Definition

VID [3:0]	Voltage Setting						
0000	3.5	1000	2.7	0100	3.1	1100	2.3
0001	3.4	1001	2.6	0101	3.0	1101	2.2
0010	3.3	1010	2.5	0110	2.9	1110	2.1
0011	3.2	1011	2.4	0111	2.8	1111	No CPU present

Most Pentium Pro processors run at 3.3V, but a few run at 3.1V. Note that the 1111 (or all opens) ID can be used to detect the absence of a processor in a given socket.

The Pentium Pro never did become very popular on the desktop, but it did find a niche in file-server applications primarily because of the full-core speed high-capacity internal L2 cache. For a time, Intel offered an OverDrive upgrade processor for the Pentium Pro, but it no longer offers any OverDrive processors. At one time, PowerLeap offered several upgrades for Pentium Pro that used 533MHz–700MHz-class Celeron PPGA processors in an adapter, but these products are no longer available.

Pentium II Processors

Intel revealed the Pentium II in May 1997. Prior to its official unveiling, the Pentium II processor was popularly referred to by its codename, Klamath, and was surrounded by much speculation throughout the industry. The Pentium II is essentially the same sixth-generation processor as the Pentium Pro, with MMX technology added (which included double the L1 cache and 57 new MMX instructions); however, there are a few twists to the design. The Pentium II processor die is shown in Figure 3.44.

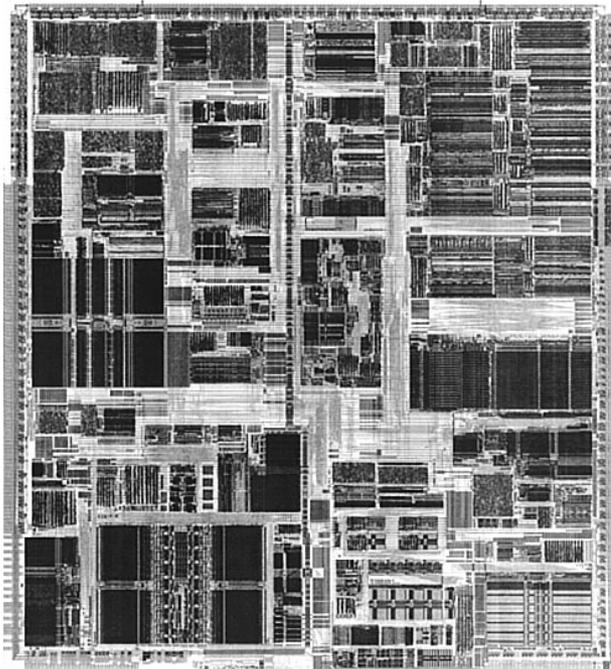


Figure 3.44 Pentium II Processor die. *Photograph used by permission of Intel Corporation.*

From a physical standpoint, it was a big departure from previous processors. Abandoning the chip in a socket approach used by virtually all processors up until this point, the Pentium II chip is characterized by its SEC cartridge design. The processor, along with several L2 cache chips, is mounted on a small circuit board (much like an oversized-memory SIMM) as shown in Figure 3.45, and the circuit board is then sealed in a metal and plastic cartridge. The cartridge is then plugged into the motherboard through an edge connector called Slot 1, which looks very much like an adapter card slot.

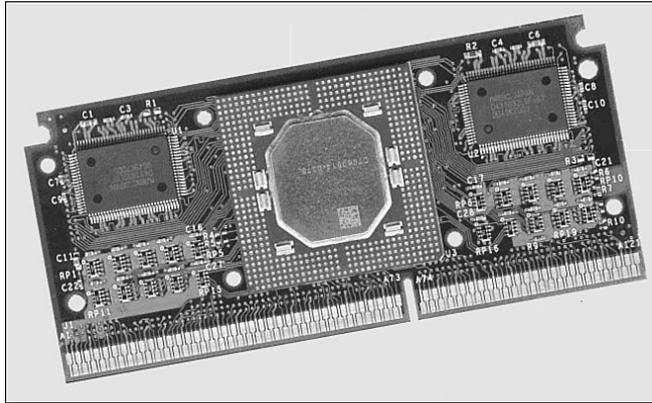


Figure 3.45 Pentium II processor board (normally found inside the SEC cartridge). *Photograph used by permission of Intel Corporation.*

The two variations on these cartridges are called SECC (single edge contact cartridge) and SECC2. Figure 3.46 shows a diagram of the SECC package; Figure 3.47 shows the SECC2 package.

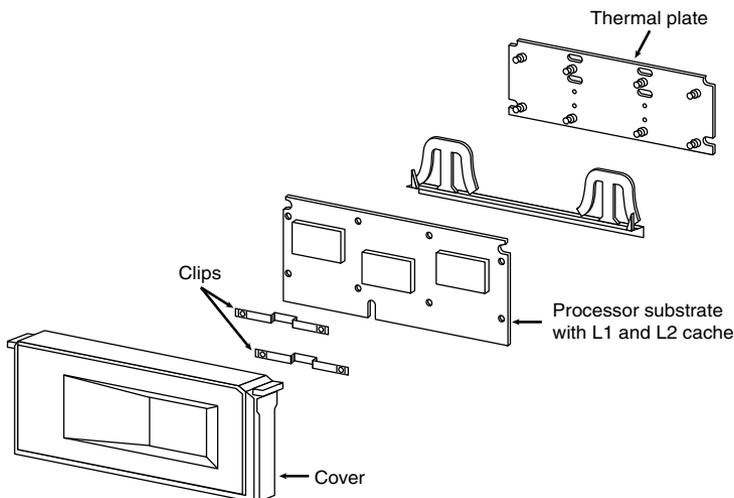


Figure 3.46 SECC components showing an enclosed processor board.

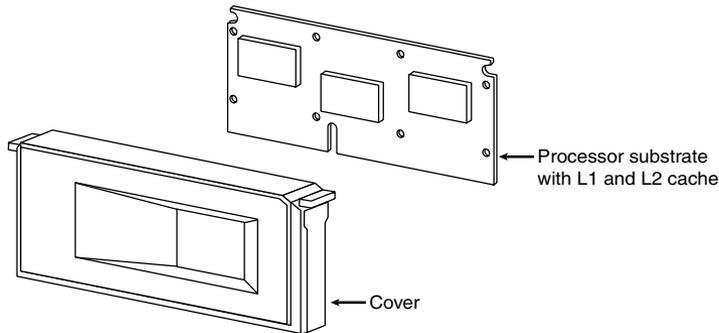


Figure 3.47 SECC, rev. 2 components showing a half-enclosed processor board.

As you can see from these figures, the SECC2 version was cheaper to make because it uses fewer over-all parts. It also allowed for a more direct heatsink attachment to the processor for better cooling. Intel transitioned from SECC to SECC2 in the beginning of 1999; all later PII chips, and the Slot 1 PIII chips that followed, use the improved SECC2 design.

By using separate chips mounted on a circuit board, Intel could build the Pentium II much less expensively than the multiple die within a package used in the Pentium Pro. Intel could also use cache chips from other manufacturers and more easily vary the amount of cache in future processors compared to the Pentium Pro design.

Intel offered Pentium II processors with the following speeds:

CPU Type/Speed	CPU Clock	Motherboard Speed
Pentium II 233MHz	3.5x	66MHz
Pentium II 266MHz	4x	66MHz
Pentium II 300MHz	4.5x	66MHz
Pentium II 333MHz	5x	66MHz
Pentium II 350MHz	3.5x	100MHz
Pentium II 400MHz	4x	100MHz
Pentium II 450MHz	4.5x	100MHz

The Pentium II processor core has 7.5 million transistors and is based on Intel's advanced P6 architecture. The Pentium II started out using a 0.35-micron process technology, although the 333MHz and faster Pentium IIs are based on 0.25-micron technology. This enables a smaller die, allowing increased core frequencies and reduced power consumption. At 333MHz, the Pentium II processor delivers a 75%–150% performance boost, compared to the 233MHz Pentium processor with MMX technology, and approximately 50% more performance on multimedia benchmarks. As shown earlier in Table 3.8, the iCOMP 2.0 Index rating for the Pentium II 266MHz chip is more than twice as fast as a classic Pentium 200MHz.

Aside from speed, the best way to think of the Pentium II is as a Pentium Pro with MMX technology instructions and a slightly modified cache design. It has the same multiprocessor scalability as the Pentium Pro, as well as the integrated L2 cache. The 57 new multimedia-related instructions carried over from the MMX processors and the capability to process repetitive loop commands more efficiently are included as well. Also included as a part of the MMX upgrade is double the internal L1 cache from the Pentium Pro (from 16KB total to 32KB total in the Pentium II).

Maximum power usage for the Pentium II is shown in the following table:

Core Speed	Power Draw	Process	Voltage
450MHz	27.1w	0.25-micron	2.0V
400MHz	24.3w	0.25-micron	2.0V
350MHz	21.5w	0.25-micron	2.0V
333MHz	23.7w	0.25-micron	2.0V
300MHz	43.0w	0.35-micron	2.8V
266MHz	38.2w	0.35-micron	2.8V
233MHz	34.8w	0.35-micron	2.8V

You can see that the highest speed 450MHz version of the Pentium II actually uses less power than the slowest original 233MHz version! This was accomplished by using the smaller 0.25-micron process and running the processor on a lower voltage of only 2.0V. Pentium III and subsequent processors used even smaller processes and lower voltages to continue this trend.

The Pentium II includes Dynamic Execution, which describes unique performance-enhancing developments by Intel and was first introduced in the Pentium Pro processor. Major features of Dynamic Execution include multiple branch prediction, which speeds execution by predicting the flow of the program through several branches; dataflow analysis, which analyzes and modifies the program order to execute instructions when ready; and speculative execution, which looks ahead of the program counter and executes instruction that are likely to be needed. The Pentium II processor expands on these capabilities in sophisticated and powerful new ways to deliver even greater performance gains.

Similar to the Pentium Pro, the Pentium II also includes DIB architecture. The term *Dual Independent Bus* comes from the existence of two independent buses on the Pentium II processor—the L2 cache bus and the processor-to-main-memory system bus. The Pentium II processor can use both buses simultaneously, thus getting as much as twice as much data in and out of the Pentium II processor as a single-bus architecture processor. The DIB architecture enables the L2 cache of the 333MHz Pentium II processor to run 2 1/2 times as fast as the L2 cache of Pentium processors. As the frequency of future Pentium II processors increases, so will the speed of the L2 cache. Also, the pipelined system bus enables simultaneous parallel transactions instead of singular sequential transactions. Together, these DIB architecture improvements offer up to three times the bandwidth performance over a single-bus architecture as with the regular Pentium.

Table 3.32 shows the general Pentium II processor specifications. Table 3.33 shows the specifications that vary by model.

Table 3.32 Pentium II General Processor Specifications

Bus speeds	66MHz, 100MHz
CPU clock multiplier	3.5x, 4x, 4.5x, 5x
CPU speeds	233MHz, 266MHz, 300MHz, 333MHz, 350MHz, 400MHz, 450MHz
Cache memory	16K×2 (32KB) L1, 512KB 1/2-speed L2
Internal registers	32-bit
External data bus	64-bit system bus w/ ECC; 64-bit cache bus w/ optional ECC
Memory address bus	36-bit
Addressable memory	64GB
Virtual memory	64TB
Physical package	Single edge contact cartridge (S.E), 242 pins
Package dimensions	5.505" (13.98cm) × 2.473" (6.28cm) × 0.647" (1.64cm)
Math coprocessor	Built-in FPU
Power management	SMM

Table 3.33 Pentium II Specifications by Model

<i>Pentium II MMX Processor (350MHz, 400MHz, and 450MHz)</i>	
Introduction date	April 15, 1998
Clock speeds	350MHz (100MHz×3.5), 400MHz (100MHz×4), and 450MHz (100MHz×4.5)
iCOMP Index 2.0 rating	386 (350MHz), 440 (400MHz), and 483 (450MHz)
Number of transistors	7.5 million (0.25-micron process), plus 31 million in 512KB L2 cache
Cacheable RAM	4GB
Operating voltage	2.0V
Slot	Slot 1
Die size	0.400" per side (10.2mm)
<i>Mobile Pentium II Processor (266MHz, 300MHz, 333MHz, and 366MHz)</i>	
Introduction date	January 25, 1999
Clock speeds	266MHz, 300MHz, 333MHz, and 366MHz
Number of transistors	27.4 million (0.25-micron process), 256KB on-die L2 cache
Ball grid array (BGA)	Number of balls = 615
Dimensions	Width = 31mm; length = 35mm
Core voltage	1.6 volts
Thermal design power ranges by frequency	366MHz = 9.5 watts; 333MHz = 8.6 watts; 300MHz = 7.7 watts; 266MHz = 7.0 watts
<i>Pentium II MMX Processor (333MHz)</i>	
Introduction date	January 26, 1998
Clock speed	333MHz (66MHz×5)
iCOMP Index 2.0 rating	366
Number of transistors	7.5 million (0.25-micron process), plus 31 million in 512KB L2 cache
Cacheable RAM	512MB
Operating voltage	2.0V
Slot	Slot 1
Die size	0.400" per side (10.2mm)
<i>Pentium II MMX Processor (300MHz)</i>	
Introduction date	May 7, 1997
Clock speed	300MHz (66MHz×4.5)
iCOMP Index 2.0 rating	332
Number of transistors	7.5 million (0.35-micron process), plus 31 million in 512KB L2 cache
Cacheable RAM	512MB
Die size	0.560" per side (14.2mm)
<i>Pentium II MMX Processor (266MHz)</i>	
Introduction date	May 7, 1997
Clock speed	266MHz (66MHz×4)
iCOMP Index 2.0 rating	303
Number of transistors	7.5 million (0.35-micron process), plus 31 million in 512KB L2 cache

Table 3.33 Continued

<i>Pentium II MMX Processor (266MHz)</i>	
Cacheable RAM	512MB
Slot	Slot 1
Die size	0.560" per side (14.2mm)
<i>Pentium II MMX Processor (233MHz)</i>	
Introduction date	May 7, 1997
Clock speed	233MHz (66MHz×3.5)
iCOMP Index 2.0 rating	267
Number of transistors	7.5 million (0.35-micron process), plus 31 million in 512KB L2 cache
Cacheable RAM	512MB
Slot	Slot 1
Die size	0.560" per side (14.2mm)

The L1 cache always runs at full-core speeds because it is mounted directly on the processor die. The L2 cache in the Pentium II normally runs at half-core speed, which saves money and allows for less expensive cache chips to be used. For example, in a 333MHz Pentium II, the L1 cache runs at a full 333MHz, whereas the L2 cache runs at 167MHz. Even though the L2 cache is not at full-core speed as it was with the Pentium Pro, this is still far superior to having cache memory on the motherboard running at the 66MHz motherboard speed of most Socket 7 Pentium designs. Intel claims that the DIB architecture in the Pentium II enables up to three times the bandwidth of normal single-bus processors, such as the original Pentium.

By removing the cache from the processor's internal package and using external chips mounted on a substrate and encased in the cartridge design, Intel could use more cost-effective cache chips and more easily scale the processor up to higher speeds. The Pentium Pro was limited in speed to 200MHz, largely due to the inability to find affordable cache memory that ran any faster. By running the cache memory at half-core speed, the Pentium II can run up to 400MHz while still using 200MHz-rated cache chips. To offset the half-core speed cache used in the Pentium II, Intel doubled the basic amount of integrated L2 cache from 256KB standard in the Pro to 512KB standard in the Pentium II.

Note that the tag RAM included in the L2 cache enables up to 512MB of main memory to be cacheable in PII processors from 233MHz to 333MHz. The 350MHz, 400MHz, and faster versions include an enhanced tag-RAM that allows up to 4GB of main memory to be cacheable. If you support systems based on the Pentium II, be aware of the caching limitations in the slower processors before upgrading memory above 512MB. Uncached memory will slow down any system.

The system bus of the Pentium II provides "glueless" support for up to two processors. This enables low-cost, two-way multiprocessing on the L2 cache bus. These system buses are designed especially for servers or other mission-critical system use where reliability and data integrity are important. All Pentium IIs also include parity-protected address/request and response system bus signals with a retry mechanism for high data integrity and reliability. As a result, the Pentium II was used in many servers and workstations.

To install the Pentium II in a system, a special processor-retention mechanism is required. This consists of a mechanical support that attaches to the motherboard and secures the Pentium II processor in Slot 1 to prevent shock and vibration damage. Retention mechanisms should be provided by the motherboard manufacturer. (For example, the Intel Boxed AL440FX and DK440LX motherboards included a retention mechanism, plus other important system integration components.) The retention mechanism sometimes folds out of the way for easier storage of the motherboard component, or it might use a rigid design.

The Pentium II can generate a significant amount of heat that must be dissipated. This is accomplished by installing a heatsink on the processor. Many of the Pentium II processors use an active heatsink that incorporates a fan. Unlike heatsink fans for previous Intel boxed processors, the Pentium II fans draw power from a three-pin power header on the motherboard. Most motherboards provide several fan connectors to supply this power.

Special heatsink supports are necessary to furnish mechanical support between the fan heatsink and support holes on the motherboard. Normally, a plastic support is inserted into the heatsink holes in the motherboard next to the CPU, before installing the CPU/heatsink package. Most fan heatsinks have two components: a fan in a plastic shroud and a metal heatsink. The heatsink is attached to the processor's thermal plate and should not be removed. The fan can be removed and replaced if necessary—for example, if it has failed. Figure 3.48 shows the SEC assembly with fan, power connectors, mechanical supports, and the slot and support holes on the motherboard.

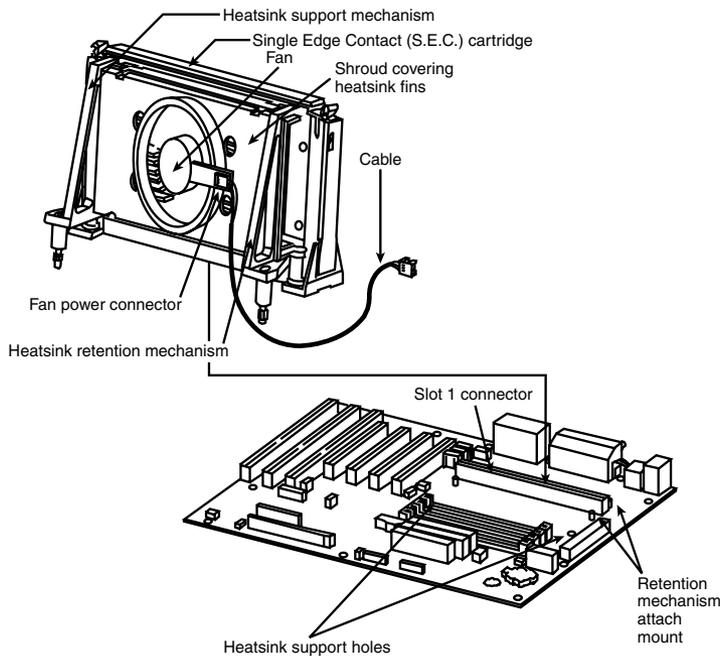


Figure 3.48 Pentium II/III processor and heatsink assembly.

The following tables show the specifications unique to certain versions of the Pentium II processor.

To identify exactly which Pentium II processor you have and what its capabilities are, look at the specification number printed on the SEC cartridge. You will find the specification number in the dynamic mark area on the top of the processor module. See Figure 3.49 to locate these markings.

After you have located the specification number (actually, it is an alphanumeric code), you can look it up in Table 3.34 to see exactly which processor you have.

For example, a specification number of SL2KA identifies the processor as a Pentium II 333MHz running on a 66MHz system bus, with an ECC L2 cache, and indicates that this processor runs on only 2.0V. The stepping is also identified, and by looking in the “Pentium II Specification Update Manual” published by Intel, you could figure out exactly which bugs were fixed in that revision.

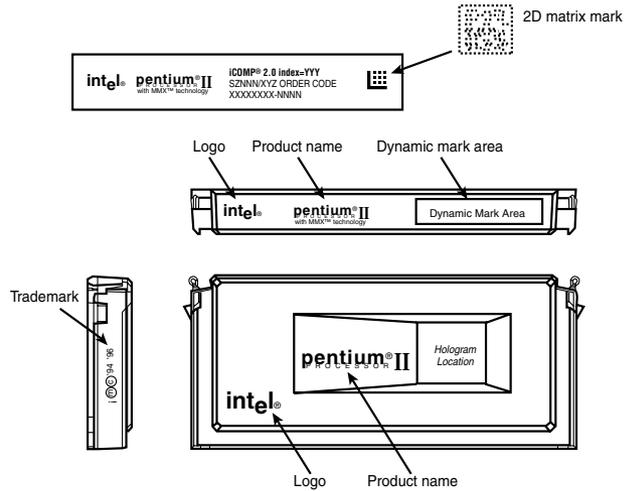


Figure 3.49 Pentium II/III SECC.

Table 3.34 Basic Pentium II Processor Identification Information

S-spec	Core Stepping	CPUID	Core/Bus Speed (MHz)	L2 Cache Size (KB)	L2 Cache Type	CPU Package	Notes (see footnotes)
SL264	C0	0633h	233/66	512	Non-ECC	SECC 3.00	5
SL265	C0	0633h	266/66	512	Non-ECC	SECC 3.00	5
SL268	C0	0633h	233/66	512	ECC	SECC 3.00	5
SL269	C0	0633h	266/66	512	ECC	SECC 3.00	5
SL28K	C0	0633h	233/66	512	Non-ECC	SECC 3.00	1, 3, 5
SL28L	C0	0633h	266/66	512	Non-ECC	SECC 3.00	1, 3, 5
SL28R	C0	0633h	300/66	512	ECC	SECC 3.00	5
SL2MZ	C0	0633h	300/66	512	ECC	SECC 3.00	1, 5
SL2HA	C1	0634h	300/66	512	ECC	SECC 3.00	5
SL2HC	C1	0634h	266/66	512	Non-ECC	SECC 3.00	5
SL2HD	C1	0634h	233/66	512	Non-ECC	SECC 3.00	5
SL2HE	C1	0634h	266/66	512	ECC	SECC 3.00	5
SL2HF	C1	0634h	233/66	512	ECC	SECC 3.00	5
SL2QA	C1	0634h	233/66	512	Non-ECC	SECC 3.00	1, 3, 5
SL2QB	C1	0634h	266/66	512	Non-ECC	SECC 3.00	1, 3, 5
SL2QC	C1	0634h	300/66	512	ECC	SECC 3.00	1, 5
SL2KA	dA0	0650h	333/66	512	ECC	SECC 3.00	5
SL2QF	dA0	0650h	333/66	512	ECC	SECC 3.00	1
SL2K9	dA0	0650h	266/66	512	ECC	SECC 3.00	
SL35V	dA1	0651h	300/66	512	ECC	SECC 3.00	1, 2
SL2QH	dA1	0651h	333/66	512	ECC	SECC 3.00	1, 2
SL2S5	dA1	0651h	333/66	512	ECC	SECC 3.00	2, 5
SL2ZP	dA1	0651h	333/66	512	ECC	SECC 3.00	2, 5
SL2ZQ	dA1	0651h	350/100	512	ECC	SECC 3.00	2, 5

Table 3.34 Continued

S-spec	Core Stepping	CPUID	Core/Bus Speed (MHz)	L2 Cache Size (KB)	L2 Cache Type	CPU Package	Notes (see footnotes)
SL2S6	dA1	0651h	350/100	512	ECC	SECC 3.00	2, 5
SL2S7	dA1	0651h	400/100	512	ECC	SECC 3.00	2, 5
SL2SF	dA1	0651h	350/100	512	ECC	SECC 3.00	1, 2
SL2SH	dA1	0651h	400/100	512	ECC	SECC 3.00	1, 2
SL2VY	dA1	0651h	300/66	512	ECC	SECC 3.00	1, 2
SL33D	dB0	0652h	266/66	512	ECC	SECC 3.00	1, 2, 5
SL2YK	dB0	0652h	300/66	512	ECC	SECC 3.00	1, 2, 5
SL2WZ	dB0	0652h	350/100	512	ECC	SECC 3.00	1, 2, 5
SL2YM	dB0	0652h	400/100	512	ECC	SECC 3.00	1, 2, 5
SL37G	dB0	0652h	400/100	512	ECC	SECC2 OLGA	1, 2, 4
SL2WB	dB0	0652h	450/100	512	ECC	SECC 3.00	1, 2, 5
SL37H	dB0	0652h	450/100	512	ECC	SECC2 OLGA	1, 2
SL2W7	dB0	0652h	266/66	512	ECC	SECC 2.00	2, 5
SL2W8	dB0	0652h	300/66	512	ECC	SECC 3.00	2, 5
SL2TV	dB0	0652h	333/66	512	ECC	SECC 3.00	2, 5
SL2U3	dB0	0652h	350/100	512	ECC	SECC 3.00	2, 5
SL2U4	dB0	0652h	350/100	512	ECC	SECC 3.00	2, 5
SL2U5	dB0	0652h	400/100	512	ECC	SECC 3.00	2, 5
SL2U6	dB0	0652h	400/100	512	ECC	SECC 3.00	2, 5
SL2U7	dB0	0652h	450/100	512	ECC	SECC 3.00	2, 5
SL356	dB0	0652h	350/100	512	ECC	SECC2 PLGA	2, 5
SL357	dB0	0652h	400/100	512	ECC	SECC2 OLGA	2, 5
SL358	dB0	0652h	450/100	512	ECC	SECC2 OLGA	2, 5
SL37F	dB0	0652h	350/100	512	ECC	SECC2 PLGA	1, 2, 5
SL3FN	dB0	0652h	350/100	512	ECC	SECC2 OLGA	2, 5
SL3EE	dB0	0652h	400/100	512	ECC	SECC2 PLGA	2, 5
SL3F9	dB0	0652h	400/100	512	ECC	SECC2 PLGA	1, 2
SL38M	dB1	0653h	350/100	512	ECC	SECC 3.00	1, 2, 5
SL38N	dB1	0653h	400/100	512	ECC	SECC 3.00	1, 2, 5
SL36U	dB1	0653h	350/100	512	ECC	SECC 3.00	2, 5
SL38Z	dB1	0653h	400/100	512	ECC	SECC 3.00	2, 5
SL3D5	dB1	0653h	400/100	512	ECC	SECC2 OLGA	1, 2

CPUID = The internal ID returned by the CPUID instruction

ECC = Error correcting code

OLGA = Organic land grid array

PLGA = Plastic land grid array

SECC = Single edge contact cartridge

SECC2 = Single edge contact cartridge revision 2

1. This is a boxed Pentium II processor with an attached fan heatsink.
2. These processors have an enhanced L2 cache, which can cache up to 4GB of main memory. Other standard PII processors can cache only up to 512MB of main memory.
3. These boxed processors might have packaging that incorrectly indicates ECC support in the L2 cache.
4. This is a boxed Pentium II OverDrive processor with an attached fan heatsink, designed for upgrading Pentium Pro (Socket 8) systems.
5. These parts operate only at the specified clock multiplier frequency ratio at which they were manufactured. They can be overclocked only by increasing the bus speed.

The two variations of the SECC2 cartridge vary by the type of processor core package on the board. The plastic land grid array (PLGA) is the older type of packaging used in previous SECC cartridges and was eventually phased out. A newer organic land grid array (OLGA), which is a processor core package that is smaller and easier to manufacture, took its place. It also enabled better thermal transfer between the processor die and the heatsink, which was attached directly to the top of the OLGA chip package. Figure 3.50 shows the open back side (where the heatsink would be attached) of SECC2 processors with PLGA and OLGA cores.

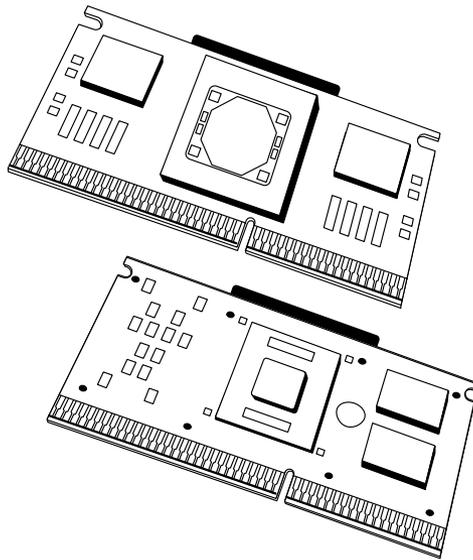


Figure 3.50 SECC2 processors with PLGA (top) and OLGA (bottom) cores.

Pentium II motherboards have an onboard voltage regulator circuit designed to power the CPU. Some Pentium II processors run at several different voltages, so the regulator must be set to supply the correct voltage for the specific processor you are installing. As with the Pentium Pro and unlike the older Pentium, no jumpers or switches must be set; the voltage setting is handled completely automatically through the VID pins on the processor cartridge. Table 3.35 shows the relationship between the pins and the selected voltage.

To ensure the system is ready for all Pentium II processor variations, the values in **bold** must be supported. Most Pentium II processors run at 2.8V, with some newer ones at 2.0V.

The Pentium II Mobile Module is a Pentium II for notebooks that includes the North Bridge of the high-performance 440BX chipset. This was the first chipset on the market that allowed 100MHz processor bus operation, although that feature was not supported in the mobile versions. The 440BX chipset was released at the same time as the 350MHz and 400MHz versions of the Pentium II.

Newer variations on the Pentium II include the Pentium IIPE, which is a mobile version that includes 256KB of L2 cache directly integrated into the die. Therefore, it runs at full-core speed, making it faster than the desktop Pentium II because the desktop chips use half-speed L2 cache.

Table 3.35 Pentium II/III/Celeron Voltage ID Pin Definitions

VID4	VID3	VID2	VID1	VID0	Voltage	VID4	VID3	VID2	VID1	VID0	Voltage
0	1	1	1	1	1.30	1	1	1	1	1	No Core
0	1	1	1	0	1.35	1	1	1	1	0	2.1
0	1	1	0	1	1.40	1	1	1	0	1	2.2
0	1	1	0	0	1.45	1	1	1	0	0	2.3
0	1	0	1	1	1.50	1	1	0	1	1	2.4
0	1	0	1	0	1.55	1	1	0	1	0	2.5
0	1	0	0	1	1.60	1	1	0	0	1	2.6
0	1	0	0	0	1.65	1	1	0	0	0	2.7
0	0	1	1	1	1.70	1	0	1	1	1	2.8
0	0	1	1	0	1.75	1	0	1	1	0	2.9
0	0	1	0	1	1.80	1	0	1	0	1	3.0
0	0	1	0	0	1.85	1	0	1	0	0	3.1
0	0	0	1	1	1.90	1	0	0	1	1	3.2
0	0	0	1	0	1.95	1	0	0	1	0	3.3
0	0	0	0	1	2.00	1	0	0	0	1	3.4
0	0	0	0	0	2.05	1	0	0	0	0	3.5

0 = Processor pin connected to Vss.

VID0–VID3 used on Socket 370.

VID0–VID4 used on Slot 1.

1 = Open on processor.

Socket 370 supports 1.30–2.05V settings only.

Slot 1 supports 1.30–3.5V settings.

Celeron

The Celeron processor is a chameleon. It was originally a P6 with the same processor core as the Pentium II in the original two versions; later it came with the same core as the PIII; and more recently it has been based on the various the Pentium 4 cores, including Prescott. It is designed mainly for lower-cost PCs.

Most of the features for the Celeron are the same as the Pentium II, III, or 4 because it uses the same internal processor cores. The main differences are in packaging, L2 cache amount, and CPU bus speed.

The first version of the Celeron was available in a package called the single edge processor package (SEPP or SEP package). The SEP package is basically the same Slot 1 design as the SECC used in the Pentium II/III, with the exception of the fancy plastic cartridge cover. This cover was deleted in the Celeron, making it cheaper to produce and sell. Essentially, the original Celeron used the same circuit board as is inside the Pentium II package.

◀◀ See “Single Edge Contact and Single Edge Processor Packaging,” p. 85.

Even without the plastic covers, the Slot 1 packaging was more expensive than it should have been. This was largely due to the processor retention mechanisms (stands) required to secure the processor into Slot 1 on the motherboard, as well as the larger and more complicated heatsinks required. This, plus competition from the lower-end Socket 7 systems using primarily AMD processors, led Intel to introduce the Celeron in a socketed form. The socket is called PGA-370 or Socket 370 because it has 370 pins. The processor package designed for this socket is called the plastic pin grid array (PPGA) package (see Figure 3.51) or flip chip PGA (FC-PGA). Both the PPGA and FC-PGA packages plug into the 370 pin socket and allow for lower-cost, lower-profile, and smaller systems because of the less expensive processor retention and cooling requirements of the socketed processor.

◀◀ See “Socket 370 (PGA-370),” p. 93.

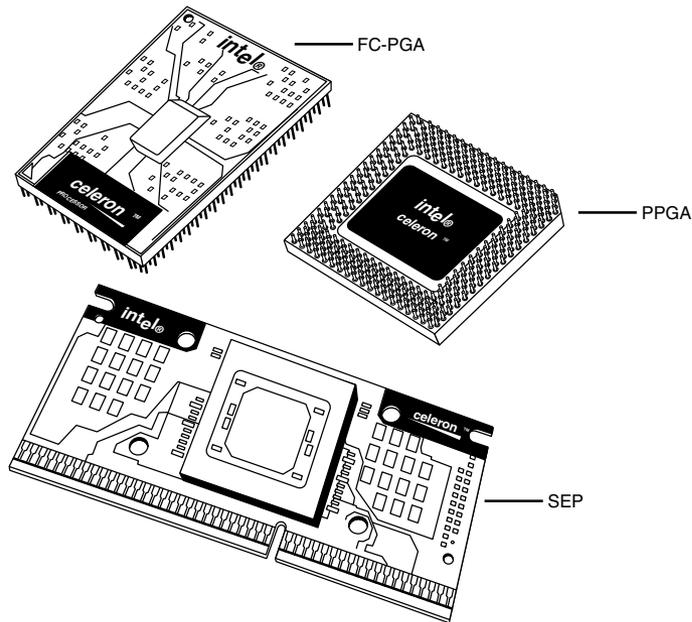


Figure 3.51 Celeron processors in the FC-PGA, PPGA, and SEP packages.

All Celeron processors at 433MHz and lower were available in the SEPP that plugs into the 242-contact slot connector (Slot 1). The 300MHz and higher versions were also made in the PPGA package. This means that the 300MHz to 433MHz have been available in both packages, whereas the 466MHz and higher-speed versions are available only in the PPGA. The fastest Celeron processor for Socket 370 runs at 1.4GHz; faster Celerons use Socket 478 and are based on the Pentium 4 design.

Motherboards that include Socket 370 can accept the PGA versions of both the Celeron and Pentium III in most cases. If you want to use a Socket 370 version of the Celeron in a Slot 1 motherboard, slot-to-socket adapters (usually called slot-kets) are available for about \$10–\$20 that plug into Slot 1 and incorporate a Socket 370 on the card. Figure 3.52 shows a typical slot-ket adapter.

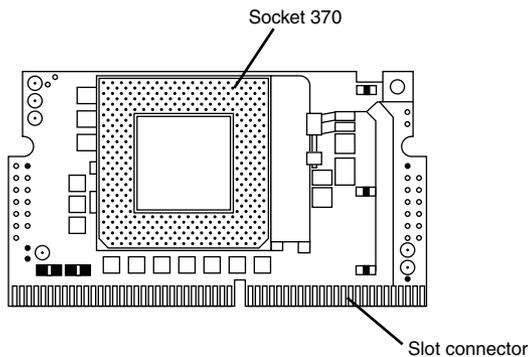


Figure 3.52 Slot-ket adapter for installing PPGA processors in Slot 1 motherboards.

Highlights of the Celeron include the following:

- Available at 300MHz (300A) and higher core frequencies with 128KB on-die L2 cache; 300MHz and 266MHz core frequencies without L2 cache
- L2 cache supports up to 4GB RAM address range and ECC
- Uses same P6 core processor as the Pentium II (266MHz through 533MHz), the Pentium III (533A MHz and higher), and the Pentium 4 (1.7GHz and higher)
- Dynamic execution microarchitecture
- Operates on a 66MHz, 100MHz, 400MHz, or 533MHz CPU bus depending on the version
- Specifically designed for lower-cost value PC systems
- Includes MMX technology; Celeron 533A and higher include SSE; Celeron 1.7GHz and higher include SSE2; Celeron D models include SSE3
- More cost-effective packaging technology, including SEP, PPGA, and FC-PGA or FC-PGA2 packages
- Integrated L1 and L2 cache on most models, with amount and type depending on the version; typically, the Celeron has half the L2 cache of the processor core it is patterned after
- Integrated thermal diode for temperature monitoring

The Intel Celeron processors from the 300A and higher include integrated 128KB L2 cache. The core for the 300A through 533MHz versions that are based on the Pentium II core include 19 million transistors because of the addition of the integrated 128KB L2 cache. The 533A and faster versions are based on the Pentium III core and incorporate 28.1 million transistors. The 1.7GHz and faster versions are based on the Pentium 4 core with 42 million transistors. The Pentium III and Pentium 4-based versions actually have 256KB of L2 cache on the die; however, 128KB is disabled, leaving 128KB of functional L2 cache. This was done because it was cheaper for Intel to simply make the Celeron using the same die as the Pentium III or 4 and just disable part of the cache on the Celeron versions, rather than coming up with a unique die for the newer Celerons. The Pentium III-based Celeron processors also support the SSE in addition to MMX instructions, whereas the Pentium 4-based versions support SSE2 instructions. The older Celerons based on the Pentium II core support only MMX.

All the Celerons in SEPP and PPGA form are manufactured using the 0.25-micron process, whereas those in FC-PGA and FC-PGA2 form are made using the better 0.18-micron and 0.13-micron processes. The smaller process reduces processor heat and enables higher speeds.

The latest Celeron processors for desktop computers use the Celeron D brand name, whereas the Celeron M brand name identifies Celeron-class processors designed for use in low-cost portable computers. Celeron D processors are manufactured using the 0.09-micron process.

A Brief Celeron History

The original Celerons were economy versions of the Intel Pentium II processor. Intel figured that by taking a Pentium II and deleting the separate L2 cache chips mounted inside the processor cartridge (and also deleting the cosmetic cover), it could create a “new” processor that was basically just a slower version of the Pentium II. As such, the first 266MHz and 300MHz Celeron models didn’t include any L2 cache. Unfortunately, this proved to have far too great a crippling effect on performance, so starting with the 300A versions, the Celeron received 128KB of on-die full-speed L2 cache, which was actually faster and more advanced than the 512KB of half-speed cache used in the Pentium II it was based on! In fact, the Celeron was the first PC processor to receive on-die L2 cache. It wasn’t until the Coppermine version of the Pentium III appeared that on-die L2 cache migrated to Intel’s main processors.

Needless to say, this caused a lot of confusion in the marketplace about the Celeron. Considering that the Celeron started out as a “crippled” Pentium II and then was revised so as to actually be superior in some ways to the Pentium II on which it was based (all the while selling for less), many didn’t know just where the Celeron stood in terms of performance. Fortunately, the crippling lack of L2 cache existed only in the earliest Celeron versions; all of those at speeds greater than 300MHz have on-die full-speed L2 cache.

The earliest Celerons from 266MHz up through 400MHz were produced in a SEPP design that physically looked like a circuit board and that was designed to fit into Slot 1. This is the same slot the Pentium II used, meaning the Celeron SEPP plugged into any Pentium II Slot-1 motherboard. As the Celeron continued to develop, the form factor was changed to correspond with changes in the Pentium II-, III-, and 4-class processors from which it was adapted. Starting with the 300A processor (300MHz Celeron with 128KB of on-die Level 2 cache), Celerons were produced in a PPGA package using the Socket 370 interface. This socket, with differences in voltage, was later used for most versions of the Pentium III. Celerons using Socket 370 range in speed from 300MHz all the way up to 1.4GHz. Along the way, the packaging changed from PPGA to FC-PGA and FC-PGA2. The latter added a metal heat spreader on top of the die offering better protection for the fragile die.

Celeron processors based on the Pentium 4 are produced in one of two package designs. Some use the FC-PGA2 package that fits into the same Socket 478 used by most Pentium 4 processors. However, the Celeron D is available in both the Socket 478 package and Socket T (LGA775) package used by the Prescott core version of the Pentium 4. The Celeron was never produced in the short-lived Socket 423 form factor the original Pentium 4 processors used.

As this very brief history shows, the name *Celeron* has never meant anything more specific than a reduced-performance version of Intel’s current mainstream processor. Before you can decide whether a particular Celeron processor is a suitable choice, you need to know what its features are and especially on which processor it is based. At least eight discrete variations of the Celeron processor exist, which are detailed in Table 3.36.

Table 3.36 Celeron CPU Variations

Celeron Version	Based On	Codename	Process (Micron)	L2 Cache (KB)	Multimedia Support
Celeron	Pentium II Deschutes	Covington	0.25	0	MMX
Celeron A	Pentium II Deschutes	Mendocino	0.25	128	MMX
Celeron A-PGA	Pentium II Deschutes	Mendocino	0.25	128	MMX
Celeron III	Pentium III Coppermine	Coppermine-128	0.18	128	SSE
Celeron IIIA	Pentium III Tualatin	Tualatin-256	0.13	256	SSE
Celeron 4	Pentium 4 Willamette	Willamette-128	0.18	128	SSE2
Celeron 4A	Pentium 4 Northwood	Northwood-128	0.13	128	SSE2
Celeron D	Pentium 4 Prescott	Prescott-256	0.09	256	SSE3

1. All Celeron III below 800MHz use the 66MHz CPU bus; all Celeron III from 800MHz through 1.1GHz use the 100MHz bus.

2. Some Socket T processors support EM64T (64-bit extensions) and Execute Disable Bit antivirus protection. See Table 3.37 for details.

SEPP = Single edge processor package.

FC-PGA = Flip chip pin grid array.

FC-PGA2 = FC-PGA with added heat spreader.

Figure 3.53 shows most of the various Celeron package types.

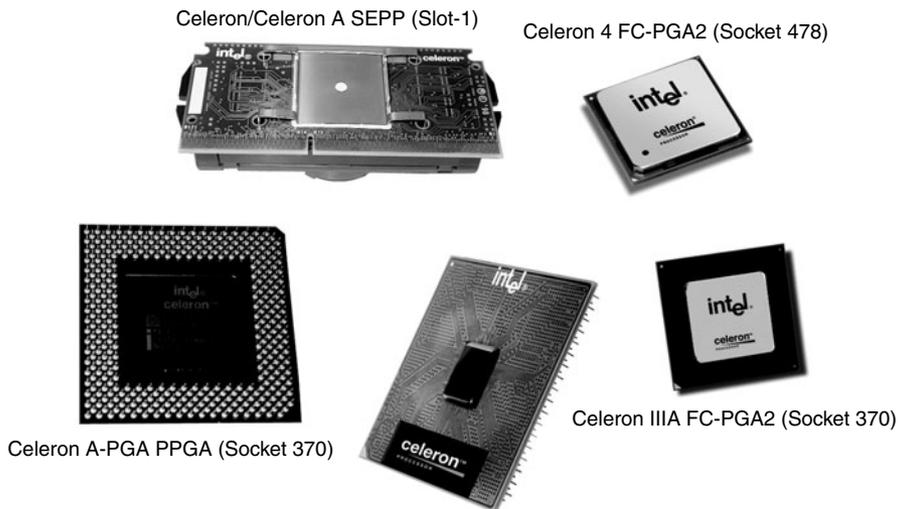


Figure 3.53 Processors released under the Celeron brand. Photos courtesy of Intel.

As you can see, there is a wide range of what is called a Celeron, and you could consider the Celeron as a family of different core processor models in several package variations.

The following sections discuss the differences between these Celeron processors.

Physical Interface	Package	CPU Bus Speed	Min. Speed	Max. Speed
Slot-1	SEPP	66MHz	266MHz	300MHz
Slot-1	SEPP	66MHz	300MHz	433MHz
Socket 370	PPGA	66MHz	300MHz	533MHz
Socket 370	FC-PGA	66/100MHz ¹	533MHz	1.1GHz
Socket 370	FC-PGA2	100MHz	900MHz	1.4GHz
Socket 478	FC-PGA2	400MHz	1.7GHz	1.8GHz
Socket 478	FC-PGA2	400MHz	2.0GHz	2.8GHz
Socket 478/Socket T (LGA775) ²	FC-PGA2	533MHz	2.13GHz	3.33GHz

MMX = Multimedia extensions; 57 additional instructions for graphics and sound processing.

SSE = Streaming SIMD (single instruction multiple data) extensions; MMX plus 70 additional instructions for graphics and sound processing.

SSE2 = Streaming SIMD extensions 2; SSE plus 144 additional instructions for graphics and sound processing.

The "Celeron Version" names listed here are not official; I made them up as a way to clearly identify the different Celeron processors.

Minimum and maximum speeds indicate the slowest and fastest rated speeds of each variation offered.

Socket 370 Celerons

Socket 370 Celerons are based on various versions of the Pentium II and Pentium III architecture.

Intel offered Celeron IIIA versions for Socket 370 motherboards in speeds from 900MHz to 1.4GHz. These processors have a CPU bus speed of 100MHz. Celeron IIIA versions based on the Pentium III Tualatin core have 256KB of L2 cache, whereas those based on the earlier Pentium III Coppermine core or Pentium II Deschutes core have 128KB of L2 cache. Compared to Celerons based on the previous Pentium III Coppermine core, Tualatin-based Celerons have the following differences:

- Larger L2 memory cache (256KB versus 128KB)
- Improved L2 cache design for better performance
- FC-PGA2 packaging, which includes a metal heat spreader over the fragile CPU core to protect it when attaching a heatsink

Like the Tualatin-core versions of the Pentium III, Celerons based on the Tualatin core don't work in motherboards designed for older Pentium III or Celeron chips. Socket 370 is physically the same, but the Tualatin core redefines 10 pins in the socket, which requires corresponding changes in the chipset and motherboard. So, if you're looking for a way to speed up an older Celeron by installing a Tualatin-core Celeron IIIA, make sure the motherboard is Tualatin-ready. Also note that Tualatin-core Celerons use the FC-PGA2 packaging, which includes a heat spreader on top of the CPU die. This requires a compatible heatsink.

Socket 478 Celeron and Celeron D Processors

Celeron processors in Socket 478 fall into three distinct camps, as Table 3.36 previously demonstrated:

- Celerons running at 1.7GHz and 1.8GHz are based on the original Pentium 4 Willamette core and have a 400MHz CPU bus, 128KB of L2 cache, and SSE2 support.

- Celerons running at 2GHz–2.8GHz with a 400MHz CPU bus are based on the Pentium 4 Northwood core, have 128KB of L2 cache, and have SSE2 support.
- Celeron D processors are based on the Prescott core used by the latest Pentium 4 processors; range in speed from 2.13GHz to 3.2GHz; and feature a 533MHz CPU bus, 256KB of L2 cache, and SSE3 support.

Socket T (LGA 775) Celeron D Processors

Celeron D processors in Socket T (LGA 775) range in speed from 2.53GHz to 3.2GHz and feature a 533MHz CPU bus, 256KB of L2 cache, and SSE3 support just like their Socket 478 Celeron D siblings. However, they also have two unique features compared to Celeron D processors in Socket 478:

- All-feature support for the Execute Disable Bit feature, which helps block buffer overrun virus attacks, when used with a compatible operating system such as Windows XP Service Pack 2.
- Some also feature support for EM64T, Intel's implementation of 64-bit extensions to the IA32 processor architecture. Thus, Celeron D processors with EM64T provide a low-cost way to use 64-bit operating systems such as Windows XP Professional x64 Edition or 64-bit Linux distributions.

Celeron D processors use the Intel processor numbering scheme introduced in 2004. Use Table 3.37 to determine the specific features supported by a particular Celeron D processor model number.

Table 3.37 Celeron D Model Numbers and Features

Celeron Model Number	Package	Clock Speed	EM64T (64-bit Extensions) Support	Execute Disable Bit Support
351	Socket T	3.2GHz	Yes	Yes
350	Socket 478	3.2GHz	No	No
346	Socket T	3.06GHz	Yes	Yes
345J	Socket T	3.06GHz	No	Yes
345	Socket 478	3.06GHz	No	No
341	Socket T	2.93GHz	Yes	Yes
340J	Socket T	2.93GHz	No	Yes
340	Socket 478	2.93GHz	No	No
336	Socket T	2.80GHz	Yes	Yes
335J	Socket T	2.80GHz	No	Yes
335	Socket 478	2.80GHz	No	No
331	Socket T	2.66GHz	Yes	Yes
330J	Socket T	2.66GHz	No	Yes
330	Socket 478	2.66GHz	No	No
326	Socket T	2.53GHz	Yes	Yes
325J	Socket T	2.53GHz	No	Yes
325	Socket 478	2.53GHz	No	No
320	Socket 478	2.40GHz	No	No
315	Socket 478	2.26GHz	No	No
310	Socket 478	2.13GHz	No	No

Because Intel has offered Celeron and Celeron D processors in many distinctive variations, it's easy to get confused as to which is which, or which is available at a specific speed. By reading the spec number

off a particular chip and looking up the number on the Intel developer website (developer.intel.com) or by using the reference charts in this book, you can find out the exact specification including socket type, voltage, stepping, cache size, and other information about the chip.

Pentium III

The Pentium III processor, shown in Figure 3.54, was first released in February 1999 and introduced several new features to the P6 family. It is essentially the same core as a Pentium II with the addition of SSE instructions and integrated on-die L2 cache in the later versions. SSE consists of 70 new instructions that dramatically enhance the performance and possibilities of advanced imaging, 3D, streaming audio, video, and speech-recognition applications.



Figure 3.54 Pentium III processor in SECC2 (Slot 1) and FC-PGA (Socket 370) packages.

Originally based on Intel's advanced 0.25-micron CMOS process technology, the PIII core started out with more than 9.5 million transistors. In late 1999, Intel shifted to a 0.18-micron process die (code-named Coppermine) and added 256KB of on-die L2 cache, which brought the transistor count to 28.1 million. The latest version of the Pentium III (codenamed Tualatin) uses a 0.13-micron process and has 44 million transistors; motherboards made before the Tualatin-core versions of the Pentium III generally do not support this processor because of logical pinout changes. The Pentium III was manufactured in speeds from 450MHz through 1.4GHz, as well as in server versions with larger or faster cache known as the Pentium Xeon. The Pentium III also incorporates advanced features such as a 32KB L1 cache and either half-core speed 512KB L2 cache or full-core speed on-die 256KB or 512KB L2 with cacheability for up to 4GB of addressable memory space. The PIII also can be used in dual-processing systems with up to 64GB of physical memory. A self-reportable processor serial number gives security, authentication, and system management applications a powerful new tool for identifying individual systems. Because of privacy concerns when the processor was released, you can disable this feature in the system BIOS on most systems that use the Pentium III or Celeron III processors.

Pentium III processors were first made available in Intel's SECC2 form factor, which replaced the more expensive older SEC packaging. The SECC2 package covers only one side of the chip and allows for better heatsink attachment and less overall weight. Architectural features of the Pentium III processor include

- *Streaming SIMD extensions (SSE)*. Seventy new instructions for dramatically faster processing and improved imaging, 3D streaming audio and video, web access, speech recognition, new user interfaces, and other graphics and sound-rich applications.
- *Intel processor serial number*. The processor serial number serves as an electronic serial number for the processor and, by extension, its system or user. This feature can be enabled or disabled as desired in the BIOS Setup. The serial number enables the system/user to be identified by company internal networks and applications. The processor serial number can be used in applications that benefit from stronger forms of system and user identification, such as:
 - *Applications using security capabilities*. Managed access to new Internet content and services; electronic document exchange.
 - *Manageability applications*. Asset management; remote system load and configuration.

Although the initial release of Pentium III processors was made in the improved SECC2 packaging, Intel later switched to the FC-PGA package, which is even less expensive to produce and enables a more direct attachment of the heatsink to the processor core for better cooling. The FC-PGA version plugs into Socket 370 but can be used in Slot 1 with a slot-let adapter.

All Pentium III processors have either 512KB or 256KB of L2 cache, which runs at either half-core or full-core speed. Pentium III Xeon versions have 512KB, 1MB, or 2MB of L2 cache that runs at full-core speed. The Pentium III Xeon is a more expensive version of the Pentium III designed for servers and workstations. All PIII processor L2 caches can cache up to 4GB of addressable memory space and include ECC capability.

Pentium III processors can be identified by their markings, which are found on the top edge of the processor cartridge. Figure 3.55 shows the format and meaning of the markings.

Table 3.38 shows variations of the Pentium III, indicated by the S-specification number.

Table 3.38 Intel Pentium III Processor Variations

Speed (MHz)	Bus Speed (MHz)	Multiplier	Boxed CPU S-spec	OEM CPU S-spec	Stepping	CPUID	L2 Cache
450	100	4.5x	SL3CC	SL364	kB0	0672	512K
450	100	4.5x	SL37C	SL35D	kC0	0673	512K
500	100	5x	SL3CD	SL365	kB0	0672	512K
500	100	5x	SL365	SL365	kB0	0672	512K
500	100	5x	SL37D	SL35E	kC0	0673	512K
500E	100	5x	SL3R2	SL3Q9	cA2	0681	256K
500E	100	5x	SL45R	SL444	cB0	0683	256K
533B	133	4x	SL3E9	SL3BN	kC0	0673	512K
533EB	133	4x	SL3SX	SL3N6	cA2	0681	256K
533EB	133	4x	SL3VA	SL3VF	cA2	0681	256K
533EB	133	4x	SL44W	SL3XG	cB0	0683	256K
533EB	133	4x	SL45S	SL3XS	cB0	0683	256K
550	100	5.5x	SL3FJ	SL3F7	kC0	0673	512K
550E	100	5.5x	SL3R3	SL3QA	cA2	0681	256K
550E	100	5.5x	SL3V5	SL3N7	cA2	0681	256K
550E	100	5.5x	SL44X	SL3XH	cB0	0683	256K
550E	100	5.5x	SL45T	N/A	cB0	0683	256K
600	100	6x	SL3JT	SL3JM	kC0	0673	512K
600E	100	6x	SL3NA	SL3H6	cA2	0681	256K
600E	100	6x	SL3NL	SL3VH	cA2	0681	256K
600E	100	6x	SL44Y	SL43E	cB0	0683	256K
600E	100	6x	SL45U	SL3XU	cB0	0683	256K
600E	100	6x	n/a	SL4CM	cC0	0686	256K
600E	100	6x	n/a	SL4C7	cC0	0686	256K
600B	133	4.5x	SL3JU	SL3JP	kC0	0673	512K
600EB	133	4.5x	SL3NB	SL3H7	cA2	0681	256K

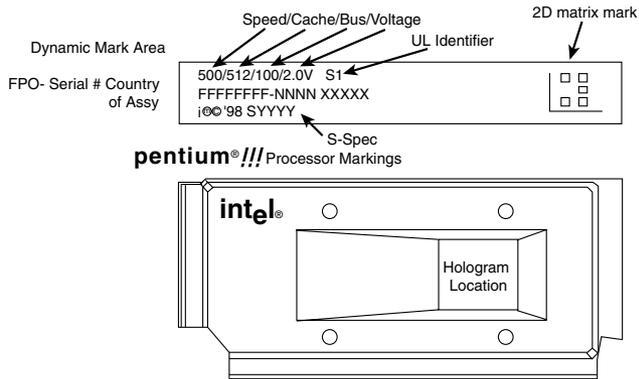


Figure 3.55 Pentium III processor markings.

L2 Speed	Max. Temp. (C)	Voltage	Max. Power (W)	Process (Microns)	Transistors	Package
225	90	2.00	25.3	0.25	9.5M	SECC2
225	90	2.00	25.3	0.25	9.5M	SECC2
250	90	2.00	28.0	0.25	9.5M	SECC2
250	90	2.00	28.0	0.25	9.5M	SECC2
250	90	2.00	28.0	0.25	9.5M	SECC2
500	85	1.60	13.2	0.18	28.1M	FC-PGA
500	85	1.60	13.2	0.18	28.1M	FC-PGA
267	90	2.05	29.7	0.25	9.5M	SECC2
533	85	1.65	14.0	0.18	28.1M	SECC2
533	85	1.65	14.0	0.18	28.1M	FC-PGA
533	85	1.65	14.0	0.18	28.1M	SECC2
533	85	1.65	14.0	0.18	28.1M	FC-PGA
275	80	2.00	30.8	0.25	9.5M	SECC2
550	85	1.60	14.5	0.18	28.1M	FC-PGA
550	85	1.60	14.5	0.18	28.1M	SECC2
550	85	1.60	14.5	0.18	28.1M	SECC2
550	85	1.60	14.5	0.18	28.1M	FC-PGA
300	85	2.00	34.5	0.25	9.5M	SECC2
600	82	1.65	15.8	0.18	28.1M	SECC2
600	82	1.65	15.8	0.18	28.1M	FC-PGA
600	82	1.65	15.8	0.18	28.1M	SECC2
600	82	1.65	15.8	0.18	28.1M	FC-PGA
600	82	1.7	15.8	0.18	28.1M	FC-PGA
600	82	1.7	15.8	0.18	28.1M	SECC2
300	85	2.05	34.5	0.25	9.5M	SECC2
600	82	1.65	15.8	0.18	28.1M	SECC2

Table 3.38 Continued

Speed (MHz)	Bus Speed (MHz)	Multiplier	Boxed CPU S-spec	OEM CPU S-spec	Stepping	CPUID	L2 Cache
600EB	133	4.5x	SL3VB	SL3VG	cA2	0681	256K
600EB	133	4.5x	SL44Z	SL3XJ	cB0	0683	256K
600EB	133	4.5x	SL45V	SL3XT	cB0	0683	256K
600EB	133	4.5x	SL4CL	SL4CL	cC0	0686	256K
600EB	133	4.5x	n/a	SL46C	cC0	0686	256K
650	100	6.5x	SL3NR	SL3KV	cA2	0681	256K
650	100	6.5x	SL3NM	SL3VJ	cA20	681	256K
650	100	6.5x	SL452	SL3XK	cB0	0683	256K
650	100	6.5x	SL45W	SL3XV	cB0	0683	256K
650	100	6.5x	n/a	SL4CK	cC0	0686	256K
650	100	6.5x	n/a	SL4C5	cC0	0686	256K
667	133	5x	SL3ND	SL3KW	cA2	0681	256K
667	133	5x	SL3T2	SL3VK	cA2	0681	256K
667	133	5x	SL453	SL3XL	cB0	0683	256K
667	133	5x	SL45X	SL3XW	cB0	0683	256K
667	133	5x	n/a	SL4CJ	cC0	0686	256K
667	133	5x	n/a	SL4C4	cC0	0686	256K
700	100	7x	SL3SY	SL3S9	cA2	0681	256K
700	100	7x	SL3T3	SL3VL	cA2	0681	256K
700	100	7x	SL454	SL453	cB0	0683	256K
700	100	7x	SL45Y	SL3XX	cB0	0683	256K
700	100	7x	SL4M7	SL4CH	cC0	0686	256K
700	100	7x	n/a	SL4C3	cC0	0686	256K
733	133	5.5x	SL3SZ	SL3SB	cA2	0681	256K
733	133	5.5x	SL3T4	SL3VM	cA2	0681	256K
733	133	5.5x	SL455	SL3XN	cB0	0683	256K
733	133	5.5x	SL45Z	SL3XY	cB0	0683	256K
733	133	5.5x	SL4M8	SL4CG	cC0	0686	256K
733	133	5.5x	SL4KD	SL4C2	cC0	0686	256K
733	133	5.5x	SL4FQ	SL4CX	cC0	0686	256K
750	100	7.5x	SL3V6	SL3WC	cA2	0681	256K
750	100	7.5x	SL3VC	SL3VN	cA2	0681	256K
750	100	7.5x	SL456	SL3XP	cB0	0683	256K
750	100	7.5x	SL462	SL3XZ	cB0	0683	256K
750	100	7.5x	SL4M9	SL4CF	cC0	0686	256K
750	100	7.5x	SL4KE	SL4BZ	cC0	0686	256K
800	100	8x	SL457	SL3XR	cB0	0683	256K
800	100	8x	SL463	SL3Y3	cB0	0683	256K
800	100	8x	SL4MA	SL4CE	cC0	0686	256K
800	100	8x	SL4KF	SL4BY	cC0	0686	256K
800EB	133	6x	SL458	SL3XQ	cB0	0683	256K
800EB	133	6x	SL464	SL3Y2	cB0	0683	256K

L2 Speed	Max. Temp. (C)	Voltage	Max. Power (W)	Process (Microns)	Transistors	Package
600	82	1.65	15.8	0.18	28.1M	FC-PGA
600	82	1.65	15.8	0.18	28.1M	SECC2
600	82	1.65	15.8	0.18	28.1M	FC-PGA
600	82	1.7	15.8	0.18	28.1M	FC-PGA
600	82	1.7	15.8	0.18	28.1M	SECC2
650	82	1.65	17.0	0.18	28.1M	SECC2
650	82	1.65	17.0	0.18	28.1M	FC-PGA
650	82	1.65	17.0	0.18	28.1M	SECC2
650	82	1.65	17.0	0.18	28.1M	FC-PGA
650	82	1.7	17.0	0.18	28.1M	FC-PGA
650	82	1.7	17.0	0.18	28.1M	SECC2
667	82	1.65	17.5	0.18	28.1M	SECC2
667	82	1.65	17.5	0.18	28.1M	FC-PGA
667	82	1.65	17.5	0.18	28.1M	SECC2
667	82	1.65	17.5	0.18	28.1M	FC-PGA
667	82	1.7	17.5	0.18	28.1M	FC-PGA
667	82	1.7	17.5	0.18	28.1M	SECC2
700	80	1.65	18.3	0.18	28.1M	SECC2
700	80	1.65	18.3	0.18	28.1M	FC-PGA
700	80	1.65	18.3	0.18	28.1M	SECC2
700	80	1.65	18.3	0.18	28.1M	FC-PGA
700	80	1.7	18.3	0.18	28.1M	FC-PGA
700	80	1.7	18.3	0.18	28.1M	SECC2
733	80	1.65	19.1	0.18	28.1M	SECC2
733	80	1.65	19.1	0.18	28.1M	FC-PGA
733	80	1.65	19.1	0.18	28.1M	SECC2
733	80	1.65	19.1	0.18	28.1M	FC-PGA
733	80	1.7	19.1	0.18	28.1M	FC-PGA
733	80	1.7	19.1	0.18	28.1M	SECC2
733	80	1.7	19.1	0.18	28.1M	SECC2
750	80	1.65	19.5	0.18	28.1M	SECC2
750	80	1.65	19.5	0.18	28.1M	FC-PGA
750	80	1.65	19.5	0.18	28.1M	SECC2
750	80	1.65	19.5	0.18	28.1M	FC-PGA
750	80	1.7	19.5	0.18	28.1M	FC-PGA
750	80	1.7	19.5	0.18	28.1M	SECC2
800	80	1.65	20.8	0.18	28.1M	SECC2
800	80	1.65	20.8	0.18	28.1M	FC-PGA
800	80	1.7	20.8	0.18	28.1M	FC-PGA
800	80	1.7	20.8	0.18	28.1M	SECC2
800	80	1.65	20.8	0.18	28.1M	SECC2
800	80	1.65	20.8	0.18	28.1M	FC-PGA

Table 3.38 Continued

Speed (MHz)	Bus Speed (MHz)	Multiplier	Boxed CPU S-spec	OEM CPU S-spec	Stepping	CPUID	L2 Cache
800EB	133	6x	SL4MB	SL4CD	cC0	0686	256K
800EB	133	6x	SL4G7	SL4XQ	cC0	0686	256K
800EB	133	6x	SL4KG	SL4BX	cC0	0686	256K
850	100	8.5x	SL47M	SL43F	cB0	0683	256K
850	100	8.5x	SL49G	SL43H	cB0	0683	256K
850	100	8.5x	SL4MC	SL4CC	cC0	0686	256K
850	100	8.5x	SL4KH	SL4BW	cC0	0686	256K
866	133	6.5x	SL47N	SL43G	cB0	0683	256K
866	133	6.5x	SL49H	SL43J	cB0	0683	256K
866	133	6.5x	SL4MD	SL4CB	cC0	0686	256K
866	133	6.5x	SL4KJ	SL4BV	cC0	0686	256K
866	133	6.5x	SL5B5	SL5QE	cD0	068A	256K
900	100	9x	n/a	SL4SD	cC0	0686	256K
933	133	7x	SL47Q	SL448	cB0	0683	256K
933	133	7x	SL49J	SL44J	cB0	0683	256K
933	133	7x	SL4ME	SL4C9	cC0	0686	256K
933	133	7x	SL4KK	SL4BT	cC0	0686	256K
933	133	7x	n/a	SL5QF	cD0	068A	256K
1000B	133	7.5x	SL4FP	SL48S	cB0	0683	256K
1000B	133	7.5x	SL4C8	SL4C8	cC0	0686	256K
1000B	133	7.5x	SL4MF	n/a	cC0	0686	256K
1000	100	10x	SL4BR	SL4BR	cC0	0686	256K
1000	100	10x	SL4KL	N/a	cC0	0686	256K
1000B	133	7.5x	SL4BS	SL4BS	cC0	0686	256K
1000B	100	10x	n/a	SL5QV	cD0	068A	256K
1000B	133	7.5x	SL5DV	N/a	cD0	068A	256K
1000B	133	7.5x	SL5B3	SL5B3	cD0	068A	256K
1000B	133	7.5x	SL52R	SL52R	cD0	068A	256K
1000B	133	7.5x	SL5FQ	n/a	cD0	068A	256K
1100	100	11x	n/a	SL5QW	cD0	068A	256K
1133	133	8.5x	SL5LT	n/a	tA1	06B1	256K
1133	133	8.5x	SL5GQ	SL5GQ	tA1	06B1	256K
1133-S	133	8.5x	SL5LV	n/a	tA1	06B1	512K
1133-S	133	8.5x	SL5PU	SL5PU	tA1	06B1	512K
1200	133	9x	SL5GN	SL5GN	tA1	06B1	256K
1200	133	9x	SL5PM	n/a	tA1	06B1	256K
1266-S	133	9.5x	SL5LW	SL5QL	tA1	06B1	512K
1333	133	10x	n/a	SL5VX	tA1	06B1	256K
1400-S	133	10.5x	SL657	SL5XL	tA1	06B1	512K

CPUID = The internal ID returned by the CPUID instruction
 ECC = Error correcting code

FC-PGA = Flip-chip pin grid array
 FC-PGA2 = Flip-chip pin grid array revision 2

L2 Speed	Max. Temp. (C)	Voltage	Max. Power (W)	Process (Microns)	Transistors	Package
800	80	1.7	20.8	0.18	28.1M	FC-PGA
800	80	1.7	20.8	0.18	28.1M	SECC2
800	80	1.7	20.8	0.18	28.1M	SECC2
850	80	1.65	22.5	0.18	28.1M	SECC2
850	80	1.65	22.5	0.18	28.1M	FC-PGA
850	80	1.7	22.5	0.18	28.1M	FC-PGA
850	80	1.7	22.5	0.18	28.1M	SECC2
866	80	1.65	22.9	0.18	28.1M	SECC2
866	80	1.65	22.9	0.18	28.1M	FC-PGA
866	80	1.7	22.5	0.18	28.1M	FC-PGA
866	80	1.7	22.5	0.18	28.1M	SECC2
866	80	1.75	26.1	0.18	28.1M	FC-PGA
900	75	1.7	23.2	0.18	28.1M	FC-PGA
933	75	1.7	25.5	0.18	28.1M	SECC2
933	75	1.7	24.5	0.18	28.1M	FC-PGA
933	75	1.7	24.5	0.18	28.1M	FC-PGA
933	75	1.7	25.5	0.18	28.1M	SECC2
933	77	1.75	27.3	0.18	28.1M	FC-PGA
1000	70	1.7	26.1	0.18	28.1M	SECC2
1000	70	1.7	26.1	0.18	28.1M	FC-PGA
1000	70	1.7	26.1	0.18	28.1M	FC-PGA
1000	70	1.7	26.1	0.18	28.1M	SECC2
1000	70	1.7	26.1	0.18	28.1M	SECC2
1000	70	1.7	26.1	0.18	28.1M	SECC2
1000	75	1.75	29.0	0.18	28.1M	FC-PGA
1000	75	1.75	29.0	0.18	28.1M	FC-PGA
1000	75	1.75	29.0	0.18	28.1M	FC-PGA
1000	75	1.75	29.0	0.18	28.1M	FC-PGA
1000	75	1.75	29.0	0.18	28.1M	FC-PGA
1100	77	1.75	33.0	0.18	28.1M	FC-PGA
1133	69	1.475	29.1	0.13	44M	FC-PGA2
1133	69	1.475	29.1	0.13	44M	FC-PGA2
1133	69	1.45	27.9	0.13	44M	FC-PGA2
1133	69	1.45	27.9	0.13	44M	FC-PGA2
1200	69	1.475	29.9	0.13	44M	FC-PGA2
1200	69	1.475	29.9	0.13	44M	FC-PGA2
1266	69	1.45	29.5	0.13	44M	FC-PGA2
1333	69	1.475	29.9	0.13	44M	FC-PGA2
1400	69	1.45	29.9	0.13	44M	FC-PGA2

SECC = Single edge contact cartridge

SECC2 = Single edge contact cartridge revision 2

Pentium III processors are all clock multiplier locked. This is a means to prevent processor fraud and over-clocking by making the processor work only at a given clock multiplier. Unfortunately, this feature can be bypassed by making modifications to the processor under the cartridge cover, and unscrupulous individuals have been selling lower-speed processors re-marked as higher speeds. It pays to purchase your systems or processors from direct Intel distributors or high-end dealers who do not engage in these practices.

Pentium II/III Xeon

The Pentium II and III processors were the basis for special high-end versions called Pentium II Xeon (introduced in June 1998) and Pentium III Xeon (introduced in March 1999). Intel now uses the term *Xeon* by itself to refer to Xeon processors based on the Pentium 4. These differ from the standard Pentium II and III in three ways: packaging, cache size, and cache speed.

Pentium II/III Xeon processors use a larger SEC cartridge than the standard PII/III processors, mainly to house a larger internal board with more cache memory.

Besides the larger package, the Xeon processors also include more L2 cache. They were produced in three variations, with 512KB, 1MB, or 2MB of L2 cache.

Even more significant than the size of the cache is its speed. All the cache in the Xeon processors run at the full-core speed. This is difficult to do considering that the cache chips were separate chips on the board in most versions. The original Pentium II Xeon processors had 7.5 million transistors in the main processor die, whereas the later Pentium III Xeon came with 9.5 million. When the Pentium III versions with on-die cache were released, the transistor count went up to 28.1 million transistors in the 256KB cache version, 84 million transistors in the 1MB cache version, and 140 million transistors in the 2MB cache version, which set an industry record at the time. The high transistor counts are due to the on-die L2 cache, which is very transistor intensive. The L2 cache in all Pentium II and III Xeon processors has a full 64GB RAM address range and supports ECC.

Table 3.39 provides an overview of the Pentium II and Pentium III Xeon processors.

Table 3.39 Intel Pentium II Xeon/Pentium III Xeon Processor Features

CPU	Processor Speed	FSB Speed	L2 Cache Sizes	Package
PII Xeon	400MHz	100MHz	512KB; 1024KB; 2048KB	SC330
	450MHz	100MHz	512KB; 1024KB; 2048KB	SC330
PIII Xeon	500MHz	100MHz	512KB; 1024KB; 2048KB	SC330
	550MHz	100MHz	512KB; 1024KB; 2048KB	SC330
	600MHz	133MHz	256KB	SC330.1
	667MHz	133MHz	256KB	SC330.1
	700MHz	100MHz	1024KB; 2048KB	SC330.1
	733MHz	133MHz	256KB	SC330.1 or 495-pin SECC
	800MHz	133MHz	256KB	SC330.1 or 495-pin SECC
	866MHz	133MHz	256KB	SC330.1 or 495-pin SECC
	900MHz	100MHz	2048KB	SC330.1
	933MHz	133MHz	256KB	SC330.1 or 495-pin SECC
	1000MHz	133MHz	256KB	495-pin SECC
Xeon	700MHz*	100MHz	1024KB	SC330.1

*Although this processor is listed as a Xeon by Intel, it's obvious from its specifications that it is really a Pentium III Xeon.

PII Xeon = Pentium II Xeon

PIII Xeon = Pentium III Xeon

For more details about Pentium II Xeon and Pentium III Xeon processors, see my book *Upgrading and Repairing Servers*.

Other Sixth-Generation Processors

Besides Intel, many other manufacturers have produced P6-type processors, but often with a difference. Most of them were designed to interface with P5 class motherboards for the lower-end markets. AMD later offered up the Athlon and Duron processors, which were true sixth-generation designs using their own proprietary connections to the system.

This section examines the various sixth-generation processors from manufacturers other than Intel.

NexGen Nx586

NexGen was founded by Thampy Thomas, who hired some of the people formerly involved with the 486 and Pentium processors at Intel. At NexGen, developers created the Nx586, a processor that was functionally the same as the Pentium but not pin compatible. As such, it was always supplied with a motherboard; in fact, it was usually soldered in. NexGen did not manufacture the chips or the motherboards they came in; for that it hired IBM Microelectronics. Later NexGen was bought by AMD, right before it was ready to introduce the Nx686—a greatly improved design by Greg Favor and a true competitor for the Pentium. AMD took the Nx686 design and combined it with a Pentium electrical interface to create a drop-in Pentium-compatible chip called the K6, which actually outperformed the original from Intel.

The Nx586 had all the standard fifth-generation processor features, such as superscalar execution with two internal pipelines and a high-performance integral L1 cache with separate code and data caches. One advantage is that the Nx586 includes separate 16KB instruction and 16KB data caches compared to 8KB each for the Pentium. These caches keep key instruction and data close to the processing engines to increase overall system performance.

The Nx586 also includes branch prediction capabilities, which are one of the hallmarks of a sixth-generation processor. Branch prediction means the processor has internal functions to predict program flow to optimize the instruction execution.

The Nx586 processor also featured a RISC core. A translation unit dynamically translates x86 instructions into RISC86 instructions. These RISC86 instructions were designed specifically with direct support for the x86 architecture while obeying RISC performance principles. They are thus simpler and easier to execute than the complex x86 instructions. This type of capability is another feature normally found only in P6 class processors.

The Nx586 was discontinued after the merger with AMD, which then took the design for the successor Nx686 and released it as the AMD-K6.

AMD-K6 Series

The AMD-K6 processor is a high-performance sixth-generation processor that is physically installable in a P5 (Pentium) motherboard. It essentially was designed for AMD by NexGen and was first known as the Nx686. The NexGen version never appeared because it was purchased by AMD before the chip was due to be released. The AMD-K6 delivers performance levels somewhere between the Pentium and Pentium II processor as a result of its unique hybrid design.

The K6 processor contains an industry-standard, high-performance implementation of the new multimedia instruction set, enabling a high level of multimedia performance for the time period. The K6-2 introduced an upgrade to MMX that AMD calls 3DNow!, which adds even more graphics and sound instructions. AMD designed the K6 processor to fit the low-cost, high-volume Socket 7 infrastructure. Initially, it used AMD's 0.35-micron, five-metal layer process technology; later the 0.25-micron process was used to increase production quantities because of reduced die size, as well as to decrease power consumption.

AMD-K6 processor technical features include

- Sixth-generation internal design, fifth-generation external interface
- Internal RISC core, translates x86 to RISC instructions
- Superscalar parallel execution units (seven)
- Dynamic execution
- Branch prediction
- Speculative execution
- Large 64KB L1 cache (32KB instruction cache plus 32KB write-back dual-ported data cache)
- Built-in floating-point unit
- Industry-standard MMX instruction support
- System Management Mode
- Ceramic pin grid array (CPGA) Socket 7 design
- Manufactured using 0.35-micron and 0.25-micron, five-layer designs

The K6-2 adds the following:

- Higher clock speeds
- Higher bus speeds of up to 100MHz (Super7 motherboards)
- 3DNow!; 21 new graphics and sound processing instructions

The K6-3 adds the following:

- 256KB of on-die full-core speed L2 cache

The addition of the full-speed L2 cache in the K6-3 was significant. It enabled the K6 series to fully compete with the Intel Pentium II processors and the Celeron processors based on the Pentium II. The 3DNow! capability added in the K6-2/3 was also exploited by newer graphics programs.

The AMD-K6 processor architecture is fully x86 binary code compatible, which means it runs all Intel software, including MMX instructions. To make up for the lower L2 cache performance of the Socket 7 design, AMD beefed up the internal L1 cache to 64KB total, twice the size of the Pentium II or III. This, plus the dynamic execution capability, enabled the K6 to outperform the Pentium and come close to the Pentium II and III in performance for a given clock rate. The K6-3 was even better with the addition of full-core speed L2 cache; however, this processor ran very hot and was discontinued after a relatively brief period.

Both the AMD-K5 and AMD-K6 processors are Socket 7 bus compatible. However, certain modifications might be necessary for proper voltage setting and BIOS revisions. To ensure reliable operation of the AMD-K6 processor, the motherboard must meet specific voltage requirements.

The AMD processors have specific voltage requirements. Most older split-voltage motherboards default to 2.8V Core/3.3V I/O, which is below specification for the AMD-K6 and could cause erratic operation. To work properly, the motherboard must have Socket 7 with a dual-plane voltage regulator supplying 2.9V or 3.2V (233MHz) to the CPU core voltage (Vcc2) and 3.3V for the I/O (Vcc3). The voltage regulator must be capable of supplying up to 7.5A (9.5A for the 233MHz) to the processor. When used with a 200MHz or slower processor, the voltage regulator must maintain the core voltage within 145mV of nominal (2.9V+/-145mV). When used with a 233MHz processor, the voltage regulator must maintain the core voltage within 100mV of nominal (3.2V+/-100mV).

If the motherboard has a poorly designed voltage regulator that cannot maintain this performance, unreliable operation can result. If the CPU voltage exceeds the absolute maximum voltage range, the processor can be permanently damaged. Also note that the K6 can run hot. Make sure your heatsink is securely fitted to the processor and that the thermally conductive grease or pad is properly applied.

The motherboard must have an AMD-K6 processor-ready BIOS with support for the K6 built in. Award has that support in its March 1, 1997 or later BIOS; AMI had K6 support in any of its BIOSs with CPU Module 3.31 or later; and Phoenix supports the K6 in version 4.0, release 6.0, or release 5.1 with build dates of 4/7/97 or later.

Because these specifications can be fairly complicated, AMD keeps a list of motherboards that have been verified to work with the AMD-K6 processor on its website.

The multiplier, bus speed, and voltage settings for the K6 are shown in Table 3.40. You can identify which AMD-K6 you have by looking at the markings on this chip, as shown in Figure 3.56.

Table 3.40 AMD-K6 Processor Speeds and Voltages

Processor	Core Speed	Clock Multiplier	Bus Speed	Core Voltage	I/O Voltage
K6-3	450MHz	4.5x	100MHz	2.4V	3.3V
K6-3	400MHz	4x	100MHz	2.4V	3.3V
K6-2	475MHz	5x	95MHz	2.4V	3.3V
K6-2	450MHz	4.5x	100MHz	2.4V	3.3V
K6-2	400MHz	4x	100MHz	2.2V	3.3V
K6-2	380MHz	4x	95MHz	2.2V	3.3V
K6-2	366MHz	5.5x	66MHz	2.2V	3.3V
K6-2	350MHz	3.5x	100MHz	2.2V	3.3V
K6-2	333MHz	3.5x	95MHz	2.2V	3.3V
K6-2	333MHz	5.0x	66MHz	2.2V	3.3V
K6-2	300MHz	3x	100MHz	2.2V	3.3V
K6-2	300MHz	4.5x	66MHz	2.2V	3.3V
K6-2	266MHz	4x	66MHz	2.2V	3.3V
K6	300MHz	4.5x	66MHz	2.2V	3.45V
K6	266MHz	4x	66MHz	2.2V	3.3V
K6	233MHz	3.5x	66MHz	3.2V	3.3V
K6	200MHz	3x	66MHz	2.9V	3.3V
K6	166MHz	2.5x	66MHz	2.9V	3.3V

**Figure 3.56** AMD Athlon processor for Slot A (cartridge form factor).

Older motherboards achieve the 3.5x setting by setting jumpers for 1.5x. The 1.5x setting for older motherboards equates to a 3.5x setting for the AMD-K6 and newer Intel parts. Getting the 4x and higher setting requires a motherboard that controls three BF pins, including BF2. Older motherboards can control only two BF pins. The settings for the multipliers are shown in Table 3.41.

Table 3.41 AMD-K6 Multiplier Settings

Multiplier Setting	BF0	BF1	BF2	Multiplier Setting	BF0	BF1	BF2
2.5x	Low	Low	High	4.5x	Low	Low	Low
3x	High	Low	High	5x	High	Low	Low
3.5x	High	High	High	5.5x	High	High	Low
4x	Low	High	Low				

These settings usually are controlled by jumpers on the motherboard. Consult your motherboard documentation to see where they are and how to set them for the proper multiplier and bus speed settings.

Unlike Cyrix and some of the other Intel competitors, AMD is a manufacturer and a designer. Therefore, it designs and builds its chips in its own fabs. Similar to Intel, AMD has migrated to 0.25-micron process technology and beyond (the AMD Athlon XP is built on a 0.13-micron process). The original K6 has 8.8 million transistors and is built on a 0.35-micron, five-layer process. The die is 12.7mm on each side, or about 162 square mm. The K6-3 uses a 0.25-micron process and incorporates 21.3 million transistors on a die only 10.9mm on each side, or about 118 square mm.

Because of its performance and compatibility with the Socket 7 interface, the K6 series is often looked at as an excellent processor upgrade for motherboards using older Pentium or Pentium MMX processors. Although they do work in Socket 7, the AMD-K6 processors have different voltage and bus speed requirements from the Intel processors. Before attempting any upgrades, you should check the board documentation or contact the manufacturer to see whether your board meets the necessary requirements. In some cases, a BIOS upgrade also is necessary.

AMD Athlon, Duron, and Athlon XP

The Athlon is AMD's successor to the K6 series (see Figure 3.57). The Athlon was designed as a new chip from the ground up and does not interface via the Socket 7 or Super7 sockets like its previous chips. In the initial Athlon versions, AMD used a cartridge design, called Slot A, almost exactly like that of the Intel Pentium II and III. This was due to the fact that the original Athlons used 512KB of external L2 cache, which was mounted on the processor cartridge board. The external cache ran at one-half core, two-fifths core, or one-third core depending on which speed processor you had. In June 2000, AMD introduced a revised version of the Athlon (codenamed Thunderbird) that incorporates 256KB of L2 cache directly on the processor die. This on-die cache runs at full-core speed and eliminates a bottleneck in the original Athlon systems. Along with the change to on-die L2 cache, the Athlon was also introduced in a version for AMD's own Socket A (Socket 462), which replaced the Slot A cartridge version. The most recent Athlon version, called the Athlon XP, has several enhancements such as 3DNow! Professional instructions, which also include the Intel SSE instructions. The latest Athlon XP models have also returned to the use of 512KB L2 cache, but this time at full processor speed.

Although the Slot A cartridge looks a lot like the Intel Slot 1, and the Socket A looks like Intel's Socket 370, the pinouts are completely different and the AMD chips do not work in the same motherboards as the Intel chips. This was by design because AMD was looking for ways to improve its chip architecture and distance itself from Intel. Special blocked pins in either socket or slot design prevent accidentally installing the chip in the wrong orientation or wrong slot. Figure 3.57 shows the Athlon in the Slot A cartridge. Socket A versions of the Athlon closely resemble the Duron.

Table 3.42 AMD Athlon Slot A Cartridge Processor Information

Part Number	Model	Speed (MHz)	Bus Speed (MHz)	Multiplier	L2 Cache
AMD-K7500MTR51B	Model 1	500	100x2	5x	512KB
AMD-K7550MTR51B	Model 1	550	100x2	5.5x	512KB
AMD-K7600MTR51B	Model 1	600	100x2	6x	512KB
AMD-K7650MTR51B	Model 1	650	100x2	6.5x	512KB
AMD-K7700MTR51B	Model 1	700	100x2	7x	512KB
AMD-K7550MTR51B	Model 2	550	100x2	5.5x	512KB
AMD-K7600MTR51B	Model 2	600	100x2	6x	512KB
AMD-K7650MTR51B	Model 2	650	100x2	6.5x	512KB
AMD-K7700MTR51B	Model 2	700	100x2	7x	512KB
AMD-K7750MTR52B	Model 2	750	100x2	7.5x	512KB
AMD-K7800MPR52B	Model 2	800	100x2	8x	512KB
AMD-K7850MPR52B	Model 2	850	100x2	8.5x	512KB

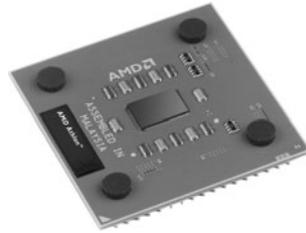


Figure 3.57 AMD Athlon XP 0.13-micron processor for Socket A (PGA form factor).

The Athlon was manufactured in speeds from 500MHz up to 1.4GHz and uses a 200MHz or 266MHz processor (front-side) bus called the EV6 to connect to the motherboard North Bridge chip as well as other processors. Licensed from Digital Equipment, the EV6 bus is the same as that used for the Alpha 21264 processor, later owned by Compaq. The EV6 bus uses a clock speed of 100MHz or 133MHz but double-clocks the data, transferring data twice per cycle, for a cycling speed of 200MHz or 266MHz. Because the bus is 8 bytes (64 bits) wide, this results in a throughput of 8 bytes times 200MHz/266MHz, which amounts to 1.6GBps or 2.1GBps. This bus is ideal for supporting PC1600 or PC2100 DDR memory, which also runs at those speeds. The AMD bus design eliminates a potential bottleneck between the chipset and processor and enables more efficient transfers compared to other processors. The use of the EV6 bus is one of the primary reasons the Athlon and Duron chips perform so well.

The Athlon has a very large 128KB of L1 cache on the processor die and one-half, two-fifths, or one-third core speed 512KB L2 cache in the cartridge in the older versions; 256KB of full-core speed cache in Socket A Athlon and most Athlon XP models; and 512KB of full-core speed cache in the latest Athlon XP models. All PGA socket A versions have the full-speed cache. The Athlon also has support for MMX and the Enhanced 3DNow! instructions, which are 45 new instructions designed to support graphics and sound processing. 3DNow! is very similar to Intel's SSE in design and intent, but the specific instructions are different and require software support. The Athlon XP adds the Intel SSE instructions, which it calls 3DNow! Professional. Fortunately, most companies producing graphics software have decided to support the 3DNow! instructions along with the Intel SSE instructions, with only a few exceptions.

The initial production of the Athlon used 0.25-micron technology, with newer and faster versions being made on 0.18-micron and 0.13-micron processes. The latest versions are even built using copper metal technology, a first in the PC processor business.

Table 3.42 shows detailed information on the Slot A version of the Athlon processor.

L2 Speed (MHz)	Voltage	Max. Power (W)	Process (Microns)	Transistors	Introduced
250	1.60V	42W	0.25	22M	Jun. 1999
275	1.60V	46W	0.25	22M	Jun. 1999
300	1.60V	50W	0.25	22M	Jun. 1999
325	1.60V	54W	0.25	22M	Aug. 1999
350	1.60V	50W	0.25	22M	Oct. 1999
275	1.60V	31W	0.18	22M	Nov. 1999
300	1.60V	34W	0.18	22M	Nov. 1999
325	1.60V	36W	0.18	22M	Nov. 1999
350	1.60V	39W	0.18	22M	Nov. 1999
300	1.60V	40W	0.18	22M	Nov. 1999
320	1.70V	48W	0.18	22M	Jan. 2000
340	1.70V	50W	0.18	22M	Feb. 2000

Table 3.42 Continued

Part Number	Model	Speed (MHz)	Bus Speed (MHz)	Multiplier	L2 Cache
AMD-K7900MNR53B	Model 2	900	100x2	9x	512KB
AMD-K7950MNR53B	Model 2	950	100x2	9.5x	512KB
AMD-K7100MNR53B	Model 2	1000	100x2	10x	512KB
AMD-A0650MPR24B	Model 4	650	100x2	6.5x	256KB
AMD-A0700MPR24B	Model 4	700	100x2	7x	256KB
AMD-A0750MPR24B	Model 4	750	100x2	7.5x	256KB
AMD-A0800MPR24B	Model 4	800	100x2	8x	256KB
AMD-A0850MPR24B	Model 4	850	100x2	8.5x	256KB
AMD-A0900MMR24B	Model 4	900	100x2	9x	256KB
AMD-A0950MMR24B	Model 4	950	100x2	9.5x	256KB
AMD-A1000MMR24B	Model 4	1000	100x2	10x	256KB

In most benchmarks the AMD Athlon compares as equal, if not superior, to the Intel Pentium III. AMD beat Intel to the 1GHz mark by introducing its 1GHz Athlon two days before Intel introduced the 1GHz Pentium III.

Table 3.43 shows information on the PGA or Socket A version of the AMD Athlon processor. All Socket A processors are Athlon Model 4.

Table 3.43 AMD Athlon PGA (Socket A) Processor Information

Speed (MHz) ¹	CPU Frequency Multiplier	Bus Speed (MHz) ²	CPU Frequency (MHz)	L2 Cache
650	6.5x	200	100	256KB
700	7x	200	100	256KB
750	6.5x	200	100	256KB
800	8x	200	100	256KB
850	8.5x	200	100	256KB
900	9x	200	100	256KB
950	9.5x	200	100	256KB
1000	10x	200	100	256KB
1000	7.5x	266	133	256KB
1100	11x	200	100	256KB
1133	8.5xx	266	133	256KB
1200	12x	200	100	256KB
1200	9x	266	133	256KB
1300	13x	200	100	256KB
1333	10x	266	133	256KB
1400	11x	266	133	256KB

1. Multiply the CPU frequency by the CPU frequency multiplier to obtain processor clock speed.

2. The CPU frequency is multiplied by 2 to obtain the bus speed. For best performance, use memory with a clock speed as fast as or faster than the bus speed.

L2 Speed (MHz)	Voltage	Max. Power (W)	Process (Microns)	Transistors	Introduced
300	1.80V	60W	0.18	22M	Mar. 2000
317	1.80V	62W	0.18	22M	Mar. 2000
333	1.80V	65W	0.18	22M	Mar. 2000
650	1.70V	36.1W	0.18	37M	Jun. 2000
700	1.70V	38.3W	0.18	37M	Jun. 2000
750	1.70V	40.4W	0.18	37M	Jun. 2000
800	1.70V	42.6W	0.18	37M	Jun. 2000
850	1.70V	44.8W	0.18	37M	Jun. 2000
900	1.75V	49.7W	0.18	37M	Jun. 2000
950	1.75V	52.0W	0.18	37M	Jun. 2000
1000	1.75V	54.3W	0.18	37M	Jun. 2000

Note

In Tables 3.43–46, the CPU frequency and CPU frequency multiplier are listed for each processor. These values are used in the system BIOS to configure your AMD processor if the BIOS is unable to manually configure the processor. The multiplier value listed in earlier editions of this book and in other sources is based on multiplying the bus speed to obtain the processor clock speed. However, the processor setup in the BIOS needs the actual values now shown in these tables.

L2 Speed (MHz)	Voltage	Max. Power (W)	Process (Microns)	Transistors
650	1.75V	38.5W	0.18	37M
700	1.75V	40.3W	0.18	37M
750	1.75V	43.8W	0.18	37M
800	1.75V	45.5W	0.18	37M
850	1.75V	47.3W	0.18	37M
900	1.75V	50.8W	0.18	37M
950	1.75V	52.5W	0.18	37M
1000	1.75V	54.3W	0.18	37M
1000	1.75V	54.3W	0.18	37M
1100	1.75V	59.5W	0.18	37M
1133	1.75V	63.0W	0.18	37M
1200	1.75V	66.5W	0.18	37M
1200	1.75V	66.5W	0.18	37M
1300	1.75V	68.3W	0.18	37M
1333	1.75V	70.0W	0.18	37M
1400	1.75V	72.0W	0.18	37M

Note

To configure an Athlon processor in the system BIOS, select the appropriate CPU frequency and CPU frequency multiplier from Table 3.43. The bus speed shown in Table 3.43 is twice that of the CPU frequency.

AMD Duron

The AMD Duron processor (originally code named Spitfire) was announced in June 2000 and is a derivative of the AMD Athlon processor in the same fashion as the Celeron is a derivative of the Pentium II and III. Basically, the Duron is an Athlon with less L2 cache; all other capabilities are essentially the same. It is designed to be a lower-cost version with less cache but only slightly less performance. In keeping with the low-cost theme, Duron contains 64KB on-die L2 cache and is designed for Socket A, a socket version of the Athlon Slot A (see Figure 3.58). Except for the Duron markings, the Duron is almost identical externally to the Socket A versions of the original Athlon.

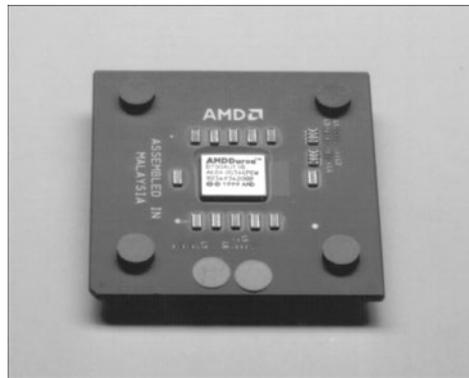


Figure 3.58 AMD Duron processor.

Essentially, the Duron was designed to compete against the Intel Celeron in the low-cost PC market, just as the Athlon was designed to compete in the higher-end Pentium III market. The Duron has since been discontinued, but most systems that use the Duron processor can use AMD Athlon or, in some cases Athlon XP or AMD Sempron processors using Socket A, as an upgrade.

Because the Duron processor is derived from the Athlon core, it includes the Athlon 200MHz front-side system bus (interface to the chipset) as well as enhanced 3DNow! instructions in Model 3. Model 7 processors include 3DNow! Professional instructions (which include a full implementation of SSE instructions).

Table 3.44 shows information on the PGA or Socket A version of the AMD Duron processor. Durons that require 1.6V are Model 3 processors, whereas those that require 1.75V are Model 7 processors. The Model 7 version was originally code named Morgan.

Table 3.44 AMD Duron Processor Information

Speed (MHz) ¹	CPU Frequency Multiplier	Bus Speed (MHz) ²	CPU Frequency (MHz)	L2 Cache	Voltage	Max. Power (W)	Process (Microns)	Transistors
550	5.5x	200	100	64KB	1.6V	25.3W	0.18	25M
600	6x	200	100	64KB	1.6V	27.4W	0.18	25M
650	6.5x	200	100	64KB	1.6V	29.4W	0.18	25M
700	7x	200	100	64KB	1.6V	31.4W	0.18	25M

Table 3.44 Continued

Speed (MHz) ¹	CPU Frequency Multiplier	Bus Speed (MHz) ²	CPU Frequency (MHz)	L2 Cache	Voltage	Max. Power (W)	Process (Microns)	Transistors
750	7.5x	200	100	64KB	1.6V	33.4W	0.18	25M
800	8x	200	100	64KB	1.6V	35.4W	0.18	25M
850	8.5x	200	100	64KB	1.6V	37.4W	0.18	25M
900	9x	200	100	64KB	1.6V	39.5W	0.18	25M
900	9x	200	100	64KB	1.75V	42.7W	0.18	25.2M
950	9.5x	200	100	64KB	1.6V	41.5W	0.18	25M
950	9.5x	200	100	64KB	1.75V	44.4W	0.18	25.2M
1000	10x	200	100	64KB	1.75V	46.1W	0.18	25.2M
1100	11x	200	100	64KB	1.75V	50.3W	0.18	25.2M
1200	12x	200	100	64KB	1.75V	54.7W	0.18	25.2M
1300	13x	200	100	64KB	1.75V	60.0W	0.18	25.2M
1400	11x	266	133	64KB	1.5V	45.5W	0.13	37.2M
1600	12x	266	133	64KB	1.5V	48.0W	0.13	37.2M
1800	13.5x	266	133	64KB	1.5V	53.0W	0.13	37.2M

1. Multiply the CPU frequency by the CPU frequency multiplier to obtain the processor clock speed.

2. The CPU frequency is multiplied by 2 to obtain the bus speed. For best performance, use memory with a clock speed as fast as or faster than the bus speed.

Note

To configure a Duron processor in the system BIOS, select the appropriate CPU frequency and CPU frequency multiplier from Table 3.44. The bus speed shown in Table 3.44 is twice that of the CPU frequency.

AMD Athlon XP

As mentioned earlier, the most recent version of the Athlon is called the Athlon XP. This is basically an improved version of the previous Athlon, with improvements in the instruction set so it can execute Intel SSE instructions and a new marketing scheme that directly competes with the Pentium 4. The latest Athlon XP models have also adopted a larger (512KB) full-speed on-die cache.

AMD uses the term "QuantiSpeed" (a marketing term, not a technical term) to refer to the architecture of the Athlon XP. AMD defines this as including the following:

- *A nine-issue superscalar, fully pipelined microarchitecture.* This provides more pathways for instructions to be sent into the execution sections of the CPU and includes three floating-point execution units, three integer units, and three address calculation units.
- *A superscalar, fully pipelined floating-point calculation unit.* This provides faster operations per clock cycle and cures a long-time deficiency of AMD processors versus Intel processors.
- *A hardware data prefetch.* This gathers the data needed from system memory and places it in the processor's Level 1 cache to save time.
- *Improved translation look-aside buffers (TLBs).* These enable the storage of data where the processor can access it more quickly without duplication or stalling for lack of fresh information.

These design improvements wring more work out of each clock cycle, enabling a "slower" Athlon XP to beat a "faster" Pentium 4 processor in doing actual work (and play).

The first models of the Athlon XP used the Palomino core, which is also shared by the Athlon 4 mobile (laptop) processor. Later models have used the Thoroughbred core, which was later revised to improve thermal characteristics. The different Thoroughbred cores are sometimes referred to as Thoroughbred-A and Thoroughbred-B. The latest Athlon XP processors use a core with 512KB on-die full-speed L2 cache known as Barton. Additional features include

- 3DNow! Professional multimedia instructions (adding compatibility with the 70 additional SSE instructions in the Pentium III but not the 144 additional SSE2 instructions in the Pentium 4)
- 266MHz or 333MHz FSB
- 128KB Level 1 and 256KB or 512KB on-die Level 2 memory caches running at full CPU speed
- Copper interconnects (instead of aluminum) for more electrical efficiency and less heat

Also new to the Athlon XP is the use of a thinner, lighter organic chip packaging compound similar to that used by recent Intel processors. Figure 3.59 shows the latest Athlon XP processors that use the Barton core.

This packaging allows for a more efficient layout of electrical components. The latest versions of the Athlon XP are made using a new 0.13-micron die process that results in a chip with a smaller die that

Table 3.45 AMD Athlon XP Processor Information

P-Rating	Actual Speed (MHz) ¹	CPU Frequency Multiplier	CPU Frequency (MHz)	Bus Speed (MHz) ²	Multiplier
1500 ⁺³	1333	10x	133	266	5x
1600 ⁺³	1400	10.5x	133	266	5.25x
1700 ⁺³	1467	11x	133	266	5.5x
1800 ⁺³	1533	11.5x	133	266	5.75x
1900 ⁺³	1600	12x	133	266	6x
2000 ⁺³	1667	12.5x	133	266	6.25x
2100 ⁺³	1733	13x	133	266	6.5x
1700 ⁺⁴	1467	11x	133	266	5.5x
1700 ⁺⁵	1467	11x	133	266	5.5x
1800 ⁺⁴	1533	11.5x	133	266	5.75x
1800 ⁺⁵	1533	11.5x	133	266	5.75x
1900 ⁺⁴	1600	12x	133	266	6x
2000 ⁺⁴	1667	12.5x	133	266	6.25x
2000 ⁺⁵	1667	12.5x	133	266	6.25x
2100 ⁺⁴	1733	13x	133	266	6.5x
2100 ⁺⁵	1733	13x	133	266	6.5x
2200 ⁺⁴	1800	13.5x	133	266	6.75x
2200 ⁺⁵	1800	13.5x	133	266	6.75x
2400 ⁺⁵	2000	15x	133	266	7.5x
2500 ⁺⁵	1833	11x	166	333	5.5x
2600 ⁺⁵	2133	16x	133	266	8x
2600 ⁺⁶	2083	12.5x	166	333	6.25x
2700 ⁺⁶	2167	13x	166	333	6.5x
2800 ⁺⁷	2083	12.5x	166	333	6.25x
3000 ⁺⁷	2167	13x	166	333	6.5x
3000 ⁺⁷	2100	10.5x	200	400	5.25x
3200 ⁺⁷	2200	11x	200	400	5.5x

1. Multiply the CPU frequency by the CPU frequency multiplier to obtain the processor clock speed.

2. The CPU frequency is multiplied by 2 to obtain the bus speed. For best performance, use memory with a clock speed as fast as or faster than the bus speed.

3. Model 6 Athlon XP (Palomino).

uses less power, generates less heat, and is capable of running faster as compared to the previous models. The newest 0.13-micron versions of the Athlon XP run at actual clock speeds exceeding 2GHz. Table 3.45 provides detailed information about the Athlon XP.

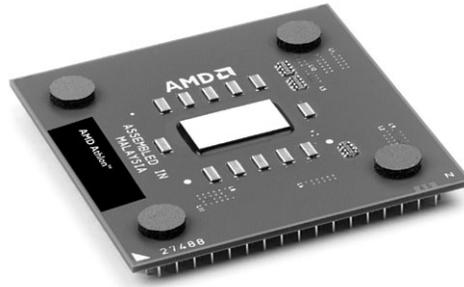


Figure 3.59 AMD Athlon XP 0.13-micron processor with 512KB of L2 cache for Socket A (PGA form factor). Photo courtesy of Advanced Micro Devices, Inc.

L2 Cache	Voltage	Max. Power (W)	Process (Microns)	Transistors
256KB	1.75V	60.0W	0.18	37.5
256KB	1.75V	62.8W	0.18	37.5
256KB	1.75V	64.0W	0.18	37.5
256KB	1.75V	66.0W	0.18	37.5
256KB	1.75V	68.0W	0.18	37.5
256KB	1.75V	70.0W	0.18	37.5
256KB	1.75V	72.0W	0.18	37.5
256KB	1.5V	49.4W	0.13	37.2
256KB	1.6V	59.8W	0.13	37.2
256KB	1.5V	51.0W	0.13	37.2
256KB	1.6V	59.8W	0.13	37.2
256KB	1.5V	52.5W	0.13	37.2
256KB	1.6V	60.3W	0.13	37.2
256KB	1.6V	61.3W	0.13	37.2
256KB	1.6V	62.1W	0.13	37.2
256KB	1.6V	62.1W	0.13	37.2
256KB	1.65V	67.9W	0.13	37.2
256KB	1.6V	62.8W	0.13	37.2
256KB	1.65V	68.3W	0.13	37.2
512KB	1.65V	68.3W	0.13	54.3
256KB	1.65V	68.3W	0.13	37.2
256KB	1.65V	68.3W	0.13	37.2
2167	1.65V	68.3W	0.13	37.2
2083	1.65V	68.3W	0.13	54.3
2167	1.65V	74.3W	0.13	54.3
512KB	1.65V	68.3W	0.13	54.3
512KB	1.65V	76.8W	0.13	54.3

4. Model 8 Athlon XP CPUID 680 (Thoroughbred).
5. Model 8 Athlon XP CPUID 681 (Thoroughbred).
6. Model 8 Athlon XP with 333MHz FSB (Thoroughbred).
7. Model 10 Athlon XP (Barton).

Note

To configure an Athlon XP processor in the system BIOS, select the appropriate CPU frequency and CPU frequency multiplier from Table 3.45. The bus speed shown in Table 3.45 is twice that of the CPU frequency.

The Athlon XP has been replaced by Socket A versions of the Sempron.

Athlon MP

The Athlon MP is AMD's first processor designed for multiprocessor support. Thus, it can be used in servers and workstations that demand multiprocessor support. The Athlon MP comes in the following three versions, which are similar to various Athlon and Athlon XP models:

- *Model 6 (1GHz, 1.2GHz)*. This model is similar to the Athlon Model 4.
- *Model 6 OPGA (1500+ through 2100+)*. This model is similar to the Athlon XP Model 6.
- *Model 8 (2000+, 2200+, 2400+, 2600+)*. This model is similar to the Athlon XP Model 8.
- *Model 10 (2500+, 2800+, 3000+)*. This model is similar to the Athlon XP Model 8, but with 512KB of L2 cache.

All Athlon MP processors use the same Socket A interface used by later models of the Athlon and all Duron and Athlon XP processors.

The Athlon MP has been replaced by the AMD Opteron. For more details about the Athlon MP, see the AMD website.

Sempron (Socket A)

AMD introduced the Sempron line of processors in 2004 to provide an economy line of processors designed to compete with the Intel Celeron D. As with the Celeron, the Sempron is a chameleon because the Sempron brand is used for both Socket A processors (based on and replacing the Athlon XP series) and Socket 754 processors (based on the Athlon 64). This section discusses Socket A versions of the Sempron. Socket 754 versions of the Sempron are discussed later in this chapter.

Table 3.46 AMD Sempron (Socket A) Processor Information

P-Rating	Actual Speed (MHz) ¹	CPU Frequency Multiplier	CPU Frequency (MHz)	Bus Speed (MHz) ²
2200+ ³	1500	166	9x	333
2200+ ⁴	1500	166	9x	333
2300+ ³	1583	166	9.5x	333
2400+ ³	1667	166	10x	333
2500+ ³	1750	166	10.5x	333
2600+ ³	1833	166	11x	333
2800+ ³	2000	166	12x	333
2800+ ⁴	2000	166	12x	333
3000+ ⁵	2000	166	12x	333

1. Multiply the CPU frequency by the CPU frequency multiplier to obtain the processor clock speed.

2. The CPU frequency is multiplied by 2 to obtain the bus speed. For best performance, use memory with a clock speed as fast as or faster than the bus speed.

►► See “AMD Sempron (Socket 754),” p. 201.

The Socket A version of the AMD Sempron is a replacement for, and is closely based on, the Athlon XP processor’s Thoroughbred (Model 8) and Barton (Model 10) versions. The major features of the Sempron are the same as the Athlon XP. Although the Sempron uses processor numbers that appear similar to those used by the Athlon XP, a Sempron with features similar to an Athlon XP does not use the same processor number. As with other AMD processors—and with Intel processors that use one of Intel’s new numbering schemes—you need to look up the specifics for a particular processor to determine its exact features.

Table 3.46 provides detailed information about Socket A versions of the Sempron.

Note

To configure a Socket A Sempron processor in the system BIOS, select the appropriate CPU frequency and CPU frequency multiplier from Table 3.46. The bus speed shown in Table 3.46 is twice that of the CPU frequency.

Cyrix/IBM 6x86 (M1) and 6x86MX (MII)

The Cyrix 6x86 processor family consists of the now-discontinued 6x86 and the newer 6x86MX processors. They are similar to the AMD-K5 and K6 in that they offer sixth-generation internal designs in a fifth-generation P5 Pentium-compatible Socket 7 exterior.

The Cyrix 6x86 and 6x86MX (renamed MII) processors incorporate two optimized superpipelined integer units and an on-chip floating-point unit. These processors include the dynamic execution capability that is the hallmark of a sixth-generation CPU design. This includes branch prediction and speculative execution.

The 6x86MX/MII processor is compatible with MMX technology to run MMX games and multimedia software. With its enhanced memory-management unit, a 64KB internal cache, and other advanced architectural features, the 6x86MX processor achieves higher performance and offers better value than competitive processors.

L2 Cache	Voltage	Max Power (W)	Process (Microns)	Transistors (Millions)
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	54.3
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	37.2
256KB	1.6V	62	.13	54.3
512KB	1.6V	62	.13	54.3

3. Model 8 Sempron (Thoroughbred core)

4. Model 10 Sempron with 256KB L2 Cache (Thorton core)

5. Model 10 Sempron (Barton Core)

Features and benefits of the 6x86 processors include

- *Superscalar architecture.* Two pipelines to execute multiple instructions in parallel
- *Branch prediction.* Predicts with high accuracy the next instructions needed
- *Speculative execution.* Enables the pipelines to continuously execute instructions following a branch without stalling the pipelines
- *Out-of-order completion.* Lets the faster instruction exit the pipeline out of order, saving processing time without disrupting program flow

The 6x86 incorporates two caches: a 16KB dual-ported unified cache and a 256-byte instruction line cache. The unified cache is supplemented with a small, quarter-K-size, high-speed, fully associative instruction line cache. The improved 6x86MX design quadruples the internal cache size to 64KB, which significantly improves performance.

The 6x86MX also includes the 57 MMX instructions that speed up the processing of certain computing-intensive loops found in multimedia and communication applications.

All 6x86 processors feature support for SMM. This provides an interrupt that can be used for system power management or software transparent emulation of I/O peripherals. Additionally, the 6x86 supports a hardware interface that enables the CPU to be placed into a low-power suspend mode.

The 6x86 is compatible with x86 software and all popular x86 operating systems, including Windows 95/98/Me, Windows NT/2000, OS/2, DOS, Solaris, and Unix. Additionally, the 6x86 processor has been certified Windows 95 compatible by Microsoft.

As with the AMD-K6, there are some unique motherboard and BIOS requirements for the 6x86 processors. The 6x86 processor has been discontinued since Cyrix was absorbed into VIA, but the 6x86MX (MII) design is still sold and supported by VIA. Check motherboard compatibility with the 6x86MX or MII processors before integrating one into an existing Socket 7/Super7 system. A BIOS update might be necessary in some cases. When installing or configuring a system with the 6x86 processors, you have to set the correct motherboard bus speed and multiplier settings. The Cyrix processors are numbered based on a P-Rating scale, which is not the same as the true megahertz clock speed of the processor.

See the section “Cyrix Processor Speeds,” earlier in this chapter, for the correct and true speed settings for the Cyrix 6x86 processors.

Note that because of the use of the P-Rating system, the actual speed of the chip is not the same number at which it is advertised. For example, the 6x86MX-PR300 is not a 300MHz chip; it actually runs at only 263MHz or 266MHz, depending on exactly how the motherboard bus speed and CPU clock multipliers are set. Cyrix says it runs as fast as a 300MHz Pentium, hence the P-Rating. Personally, I wish it would label the chip at the correct speed and then say that it runs faster than a Pentium at the same speed.

To install the 6x86 processors in a motherboard, you also must set the correct voltage. Normally, the markings on top of the chip indicate which voltage setting is appropriate. Various versions of the 6x86 run at 3.52V (use VRE setting), 3.3V (VR setting), or 2.8V (MMX) settings. The MMX versions use the standard split-plane 2.8V core 3.3V I/O settings.

The Cyrix MII is now sold by VIA Technologies.

VIA C3

The VIA C3 was originally known as the VIA Cyrix III and was designed to fit into the same Socket 370 used by the Pentium III and Celeron III. The initial versions of the C3, code named Joshua and Samuel, had 128KB L1 cache but didn't contain any L2 cache. As a consequence, they had much lower performance than similar 500MHz-class processors. The original Cyrix III/C3, code named Joshua, was developed by former Cyrix engineers after VIA bought Cyrix in late 1998, but the Samuel and subsequent versions are based on the Centaur Winchip (VIA purchased Centaur in 1999). The Samuel was

built with a .18-micron process, whereas the Samuel 2 is a development of the Samuel with 64KB of L2 cache on board and is built on a .15-micron process. The Ezra core was the first .13-micron process C3 processor, but it, like previous C3 processors, was not compatible with Tualatin (late Pentium III-compatible) motherboards. The Ezra-T core was the first C3 to reach 1GHz and the first to support Tualatin motherboards. The latest C3 uses the Nehemiah core and features clock speeds over 1GHz and built-in encryption. C3 models feature 100MHz FSB (750MHz and 900MHz models) or 133MHz FSB (733MHz, 800MHz, 866MHz, 933MHz, and higher).

The C3 is fully software compatible with other x86 processors, including Pentium III and Celeron, but its microarchitecture is designed to enhance the performance of most frequently used instructions while reducing the performance of seldom-used instructions. This design feature significantly reduces the die size needed for C3 processors, but it also reduces performance in multimedia and graphics operations. By reducing the die size, the C3 in its Nehemiah version offers typical power consumption of only 11.25 watts, making it the coolest running processor available for Socket 370 applications.

Because of its low power consumption, cool operation, and relatively low performance compared to the Intel Celeron, the C3 processor should be considered primarily for computing appliances, set-top boxes, and portable computers in which small size and low power/cooling requirements (rather than performance) are paramount.

The C3 is also available in an enhanced ball grid array (EBGA) package called the E-series. E-series C3 processors are used for permanent installation on motherboards such as the Mini-ITX ultra-compact form factor designs also produced by VIA.

For more details about various versions of the C3, refer to Table 3.2 or the VIA Technologies website.

Intel Pentium 4 (Seventh-Generation) Processors

The Pentium 4 was introduced in November 2000 and represented a new generation in processors (see Figure 3.60). If this one had a number instead of a name, it might be called the 786 because it represents a generation beyond the previous 686 class processors. Three main variations on the Pentium 4 have been released, based on the processor die and architecture. They are called the Willamette, Northwood, and Prescott. The processor dies are shown in Figure 3.61.



Figure 3.60 Pentium 4 FC-PGA2 processor.

The main technical details for the Pentium 4 include

- Speeds range from 1.3GHz to 3.8GHz.
- 42 million transistors, 0.18-micron process, 217 sq. mm die (Willamette).
- 55 million transistors, 0.13-micron process, 131 sq. mm die (Northwood).

- 125 million transistors, 0.09-micron process, 112 sq. mm die (Prescott).
- Software compatible with previous Intel 32-bit processors.
- Some Prescott versions support EM64T (64-bit extensions) and Execute Disable Bit (buffer overflow protection).
- Processor (front-side) bus runs at 400MHz, 533MHz, 800MHz, or 1066MHz.
- Arithmetic logic units (ALUs) run at twice the processor core frequency.
- Hyper-pipelined (20-stage or 31-stage) technology.
- Hyper-threading technology support in all 2.4GHz and faster processors running an 800MHz bus and all 3.06GHz and faster processors running a 533MHz bus.
- Very deep out-of-order instruction execution.
- Enhanced branch prediction.
- 8KB or 16KB L1 cache plus 12K micro-op execution trace cache.
- 256KB, 512KB, or 1MB of on-die, full-core speed 256-bit-wide L2 cache with eight-way associativity.
- L2 cache can handle up to 4GB RAM and supports ECC.
- 2MB of on-die, full-speed L3 cache (Extreme Edition).
- SSE2—SSE plus 144 new instructions for graphics and sound processing (Willamette and Northwood).
- SSE3—SSE2 plus 13 new instructions for graphics and sound processing (Prescott).
- Enhanced floating-point unit.
- Multiple low-power states.

◀◀ See "IA-32e 64-Bit Extension Mode (AMD64, x86-64, EM64T) ," p. 50.

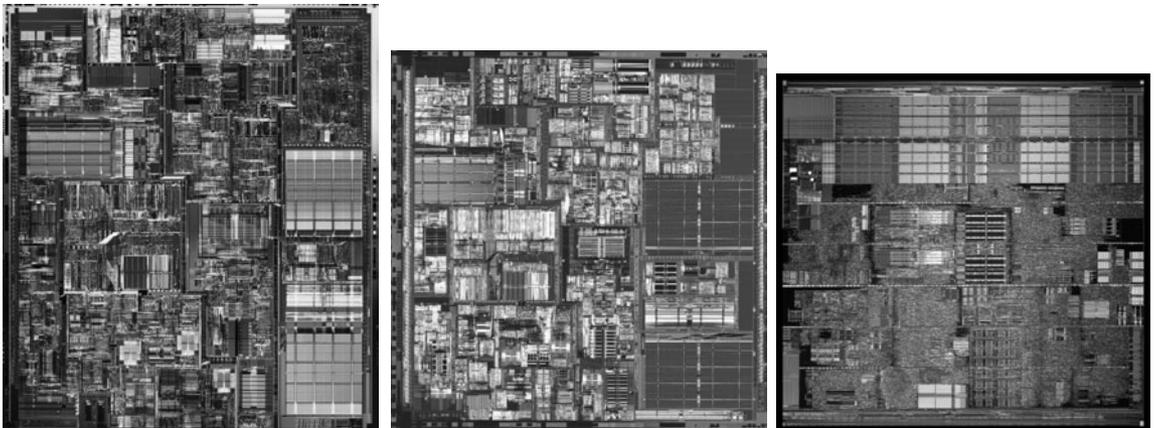


Figure 3.61 The CPU dies for the Pentium 4 CPU based on the Willamette, Northwood, and Prescott cores.

Intel abandoned Roman numerals for a standard Arabic numeral 4 designation to identify the Pentium 4. Internally, the Pentium 4 introduces a new architecture Intel calls NetBurst microarchitecture, which is a marketing term and not a technical term. Intel uses NetBurst to describe hyper-pipelined technology, a rapid execution engine, a high-speed (400MHz, 533MHz, 800MHz, or 1066MHz) system bus, and an execution trace cache. The hyper-pipelined technology doubles or triples the instruction pipeline depth as compared to the Pentium III (or Athlon/Athlon 64), meaning more and smaller steps are required to

execute instructions. Even though this might seem less efficient, it enables much higher clock speeds to be more easily attained. The rapid execution engine enables the two integer arithmetic logic units (ALUs) to run at twice the processor core frequency, which means instructions can execute in half a clock cycle. The 400MHz/533MHz/800MHz/1066MHz system bus is a quad-pumped bus running off a 100MHz/133MHz/200MHz/266MHz system clock transferring data four times per clock cycle. The execution trace cache is a high-performance Level 1 cache that stores approximately 12K decoded micro-operations. This removes the instruction decoder from the main execution pipeline, increasing performance.

Of these, the high-speed processor bus is most notable. Technically speaking, the processor bus is a 100MHz, 133MHz, 200MHz, or 266MHz quad-pumped bus that transfers four times per cycle (4x), for a 400MHz, 533MHz, 800MHz, or 1066MHz effective rate. Because the bus is 64 bits (8 bytes) wide, this results in a throughput rate of 3200MBps, 4266MBps, 6400MBps, or 8532MBps.

Table 3.44 shows how this transfer rate compares to various speeds of dual-channel RDRAM and DDR SDRAM.

Table 3.46 Pentium 4 Processor Bus and RDRAM Speed Comparison

Pentium 4 Processor Bus Speed	Throughput (Processor Bus×8)	Dual-channel RIMM Throughput	Dual-channel DDR DIMM Throughput
400MHz	3200MBps	3200MBps (PC800)	3200MBps (DDR266)
533MHz	4266MBps	4266MBps (PC1066)	4266MBps (DDR333)
800MHz	6400MBps	6400MBps (PC1200)	6400MBps (DDR400)
1066MHz	8532MBps	—	8600MBps (DDR533)

As you can see from Table 3.46, the throughput of the Pentium 4's processor bus is an exact match for the most common types of RDRAM and DDR SDRAM memory. The use of dual-channel memory means that modules must be added in matched pairs. Dual banks of PC1600 (DDR266), PC2100 (DDR333), or PC3200 (DDR400) DDR SDRAM are less expensive than equivalent RDRAM solutions, which is why virtually all newer Pentium 4 chipsets support DDR SDRAM or the newer DDR2 SDRAM.

In the Pentium 4's 20-stage or 31-stage pipelined internal architecture, individual instructions are broken down into many more substages than with previous processors such as the Pentium III, making this almost like a RISC processor. Unfortunately, this can add to the number of cycles taken to execute instructions if they are not optimized for this processor. Early benchmarks running existing software showed that existing Pentium III or AMD Athlon processors could easily keep pace with or even exceed the Pentium 4 in specific tasks; however, this is changing now that applications are being recompiled to work smoothly with the Pentium 4's deep pipelined architecture.

Another important architectural advantage is hyper-threading technology, which can be found in all Pentium 4 2.4GHz and faster processors running an 800MHz bus and all 3.06GHz and faster processors running a 533MHz bus. Hyper-threading enables a single processor to run two threads simultaneously, thereby acting as if it were two processors instead of one. For more information on hyper-threading technology, see the section "Hyper-Threading Technology," earlier in this chapter.

The Pentium 4 initially used Socket 423, which has 423 pins in a 39×39 SPGA arrangement. Later versions used Socket 478; recent versions use Socket T (LGA775), which has additional pins to support new features such as EM64T (64-bit extensions), Execute Disable Bit (protection against buffer overflow attacks), Intel Virtualization Technology, and other advanced features. The Celeron was never designed to work in Socket 423, but Celeron and Celeron D versions are available for Socket 478 and Socket T (LGA775), allowing for lower-cost systems compatible with the Pentium 4. Voltage selection is made via an automatic voltage regulator module installed on the motherboard and wired to the socket.

Use Table 3.47 as a comprehensive guide to Pentium 4 processor features. As you review the many Pentium 4 models listed in this table, you can easily see that there have actually been at least six distinct Pentium 4 generations, based on the most significant technology changes listed here:

- Socket 423
- Socket 478
- Socket 478 Hyper-Threading Technology
- Socket 478 Extreme Edition (L3 cache)
- Socket T (LGA775)
- Socket 775 EM64T (64-bit extensions)

For some time now, it has been obvious that “Pentium 4” has been far more of a brand than a single processor family, leading to endless confusion when users have considered processor upgrades or new system purchases. Because of the three form factors (Socket 423, Socket 478, and Socket 775) and the wide range of features available in the Pentium 4 family, it’s essential that you determine exactly what the features are of a particular processor before you purchase it as an upgrade to an existing processor or as part of a complete system.

Pentium 4 Extreme Edition

In November 2003, Intel introduced the Extreme Edition of the Pentium 4, which is notable for being the first desktop PC processor to incorporate L3 cache. The Extreme Edition (or Pentium 4EE) is basically

Table 3.47 Pentium 4 Processor Information

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
1.30	400	3.2	No	SL4QD	SL4SF	B2	0F07h
1.30	400	3.2	No	SL4SF	SL4SF	B2	0F07h
1.30	400	3.2	No	SL5GC	SL5FW	C1	0F0Ah
1.40	400	3.2	No	SL4SC	SL4SG	B2	0F07h
1.40	400	3.2	No	SL4SG	SL4SG	B2	0F07h
1.40	400	3.2	No	SL4X2	SL4WS	C1	0F0Ah
1.40	400	3.2	No	SL5N7	SL59U	C1	0F0Ah
1.40	400	3.2	No	SL59U	SL59U	C1	0F0Ah
1.40	400	3.2	No	SL5UE	SL5TG	D0	0F12h
1.40	400	3.2	No	SL5TG	SL5TG	D0	0F12h
1.50	400	3.2	No	SL4TY	SL4SH	B2	0F07h
1.50	400	3.2	No	SL4SH	SL4SH	B2	0F07h
1.50	400	3.2	No	SL4X3	SL4WT	C1	0F0Ah
1.50	400	3.2	No	SL4WT	SL4WT	C1	0F0Ah
1.50	400	3.2	No	SL5TN	SL5SX	D0	0F12h
1.50	400	3.2	No	SL5N8	SL59V	C1	0F0Ah
1.50	400	3.2	No	SL5UF	SL5TJ	D0	0F12h
1.50	400	3.2	No	SL5TJ	SL5TJ	D0	0F12h
1.50	400	3.2	No	SL62Y	SL62Y	D0	0F12h
1.60	400	3.2	No	SL4X4	SL4WU	C1	0F0Ah

a revamped version of the Prestonia core Xeon workstation/server processor, which has used L3 cache since November 2002. The Pentium 4EE has 512KB of L2 cache and 2MB of L3 cache, which increases the transistor count to 178 million transistors and makes the die significantly larger than the standard Pentium 4. Because of the large die based on the 130-nanometer process, this chip is expensive to produce and the extremely high selling price reflects that. The Extreme Edition is targeted toward the gaming market, where people are willing to spend extra money for additional performance. The additional cache doesn't help standard business applications as well as it helps power-hungry 3D games.

In 2004, revised versions of the Pentium 4 Extreme Edition were introduced. These processors are based on the 90-nanometer (0.09-micron) Pentium 4 Prescott core but with a larger 2MB L2 cache in place of the 512KB L2 cache design used by the standard Prescott-core Pentium 4. Pentium 4 Extreme Edition processors based on the Prescott core do not have L3 cache.

The Pentium 4 Extreme Edition is available in both Socket 478 and Socket T form factors, with clock speeds ranging from 3.2GHz to 3.4GHz (Socket 478) and from 3.4GHz to 3.73GHz (Socket T). For specific features of a particular Pentium 4 Extreme Edition processor, see Table 3.47.

The various Pentium 4 and Pentium 4 Extreme Edition versions, including thermal and power specifications, are shown in Table 3.47.

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
256K	0K	69°C	48.9W	423	180nm	42M	N/A
256K	0K	69°C	48.9W	423	180nm	42M	N/A
256K	0K	70°C	51.6W	423	180nm	42M	N/A
256K	0K	70°C	51.8W	423	180nm	42M	N/A
256K	0K	70°C	51.8W	423	180nm	42M	N/A
256K	0K	72°C	54.7W	423	180nm	42M	N/A
256K	0K	72°C	55.3W	478	180nm	42M	N/A
256K	0K	72°C	55.3W	478	180nm	42M	N/A
256K	0K	72°C	55.3W	478	180nm	42M	N/A
256K	0K	72°C	55.3W	478	180nm	42M	N/A
256K	0K	72°C	54.7W	423	180nm	42M	N/A
256K	0K	72°C	54.7W	423	180nm	42M	N/A
256K	0K	73°C	57.8W	423	180nm	42M	N/A
256K	0K	73°C	57.8W	423	180nm	42M	N/A
256K	0K	73°C	57.8W	423	180nm	42M	N/A
256K	0K	73°C	57.9W	478	180nm	42M	N/A
256K	0K	73°C	57.9W	478	180nm	42M	N/A
256K	0K	73°C	57.9W	478	180nm	42M	N/A
256K	0K	71°C	62.9W	478	180nm	42M	N/A
256K	0K	75°C	61.0W	423	180nm	42M	N/A

Table 3.47 Continued

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
1.60	400	3.2	No	SL5UL	SL5VL	D0	0F12h
1.60	400	3.2	No	SL5VL	SL5VL	D0	0F12h
1.60	400	3.2	No	SL5UW	SL5US	C1	0F0Ah
1.60	400	3.2	No	SL5UJ	SL5VH	D0	0F12h
1.60	400	3.2	No	SL5VH	SL5VH	D0	0F12h
1.60	400	3.2	No	SL6BC	SL679	E0	0F13h
1.60	400	3.2	No	SL679	SL679	E0	0F13h
1.60A	400	3.2	No	SL668	SL668	B0	0F24h
1.70	400	3.2	No	SL57V	SL57W	C1	0F0Ah
1.70	400	3.2	No	SL57W	SL57W	C1	0F0Ah
1.70	400	3.2	No	SL5TP	SL5SY	D0	0F12h
1.70	400	3.2	No	SL5N9	SL59X	C1	0F0Ah
1.70	400	3.2	No	SL5UG	SL5TK	D0	0F12h
1.70	400	3.2	No	SL5TK	SL5TK	D0	0F12h
1.70	400	3.2	No	SL62Z	SL62Z	D0	0F12h
1.70	400	3.2	No	SL6BD	SL67A	E0	0F13h
1.70	400	3.2	No	SL67A	SL67A	E0	0F13h
1.80	400	3.2	No	SL4X5	SL4WV	C1	0F0Ah
1.80	400	3.2	No	SL5UM	SL5VM	D0	0F12h
1.80	400	3.2	No	SL5VM	SL5VM	D0	0F12h
1.80	400	3.2	No	SL5UV	SL5UT	C1	0F0Ah
1.80	400	3.2	No	SL5UK	SL5VJ	D0	0F12h
1.80	400	3.2	No	SL5VJ	SL5VJ	D0	0F12h
1.80	400	3.2	No	SL6BE	SL67B	E0	0F13h
1.80	400	3.2	No	SL67B	SL67B	E0	0F13h
1.80A	400	3.2	No	SL63X	SL62P	B0	0F24h
1.80A	400	3.2	No	SL62P	SL62P	B0	0F24h
1.80A	400	3.2	No	SL68Q	SL66Q	B0	0F24h
1.80A	400	3.2	No	SL66Q	SL66Q	B0	0F24h
1.90	400	3.2	No	SL5WH	SL5VN	D0	0F12h
1.90	400	3.2	No	SL5VN	SL5VN	D0	0F12h
1.90	400	3.2	No	SL5WG	SL5VK	D0	0F12h
1.90	400	3.2	No	SL5VK	SL5VK	D0	0F12h
1.90	400	3.2	No	SL6BF	SL67C	E0	0F13h
1.90	400	3.2	No	SL67C	SL67C	E0	0F13h
2.0	400	3.2	No	SL5TQ	SL5SZ	D0	0F12h
2.0	400	3.2	No	SL5UH	SL5TL	D0	0F12h
2.0	400	3.2	No	SL5TL	SL5TL	D0	0F12h

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
256K	0K	75°C	61.0W	423	180nm	42M	N/A
256K	0K	75°C	61.0W	423	180nm	42M	N/A
256K	0K	75°C	60.8W	478	180nm	42M	N/A
256K	0K	75°C	60.8W	478	180nm	42M	N/A
256K	0K	75°C	60.8W	478	180nm	42M	N/A
256K	0K	75°C	60.8W	478	180nm	42M	N/A
256K	0K	75°C	60.8W	478	180nm	42M	N/A
512K	0K	66°C	46.8W	478	130nm	55M	N/A
256K	0K	76°C	64.0W	423	180nm	42M	N/A
256K	0K	76°C	64.0W	423	180nm	42M	N/A
256K	0K	76°C	64.0W	423	180nm	42M	N/A
256K	0K	76°C	63.5W	478	180nm	42M	N/A
256K	0K	76°C	63.5W	478	180nm	42M	N/A
256K	0K	76°C	63.5W	478	180nm	42M	N/A
256K	0K	73°C	67.7W	478	180nm	42M	N/A
256K	0K	73°C	67.7W	478	180nm	42M	N/A
256K	0K	73°C	67.7W	478	180nm	42M	N/A
256K	0K	78°C	66.7W	423	180nm	42M	N/A
256K	0K	78°C	66.7W	423	180nm	42M	N/A
256K	0K	78°C	66.7W	423	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
256K	0K	77°C	66.1W	478	180nm	42M	N/A
512K	0K	67°C	49.6W	478	130nm	55M	N/A
512K	0K	67°C	49.6W	478	130nm	55M	N/A
512K	0K	67°C	49.6W	478	130nm	55M	N/A
512K	0K	67°C	49.6W	478	130nm	55M	N/A
256K	0K	73°C	69.2W	423	180nm	42M	N/A
256K	0K	73°C	69.2W	423	180nm	42M	N/A
256K	0K	75°C	72.8W	478	180nm	42M	N/A
256K	0K	75°C	72.8W	478	180nm	42M	N/A
256K	0K	75°C	72.8W	478	180nm	42M	N/A
256K	0K	75°C	72.8W	478	180nm	42M	N/A
256K	0K	74°C	71.8W	423	180nm	42M	N/A
256K	0K	76°C	75.3W	478	180nm	42M	N/A
256K	0K	76°C	75.3W	478	180nm	42M	N/A

Table 3.47 Continued

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
2.0A	400	3.2	No	SL5ZT	SL5YR	B0	0F24h
2.0A	400	3.2	No	SL5YR	SL5YR	B0	0F24h
2.0A	400	3.2	No	SL68R	SL66R	B0	0F24h
2.0A	400	3.2	No	SL66R	SL66R	B0	0F24h
2.0A	400	3.2	No	SL6E7	SL6GQ	C1	0F27h
2.0A	400	3.2	No	SL6GQ	SL6GQ	C1	0F27h
2.0A	400	3.2	No	SL6QM	SL6PK	D1	0F29h
2.20	400	3.2	No	SL5ZU	SL5YS	B0	0F24h
2.20	400	3.2	No	SL5YS	SL5YS	B0	0F24h
2.20	400	3.2	No	SL68S	SL66S	B0	0F24h
2.20	400	3.2	No	SL66S	SL66S	B0	0F24h
2.20	400	3.2	No	SL6E8	SL6GR	C1	0F27h
2.20	400	3.2	No	SL6GR	SL6GR	C1	0F27h
2.20	400	3.2	No	SL6QN	SL6PL	D1	0F29h
2.26	533	4.3	No	SL683	SL67Y	B0	0F24h
2.26	533	4.3	No	SL67Y	SL67Y	B0	0F24h
2.26	533	4.3	No	SL6ET	SL6D6	B0	0F24h
2.26	533	4.3	No	SL6EE	SL6DU	C1	0F27h
2.26	533	4.3	No	SL6DU	SL6DU	C1	0F27h
2.26	533	4.3	No	SL6Q7	SL6PB	D1	0F29h
2.40	400	3.2	No	SL67R	SL65R	B0	0F24h
2.40	400	3.2	No	SL65R	SL65R	B0	0F24h
2.40	400	3.2	No	SL68T	SL66T	B0	0F24h
2.40	400	3.2	No	SL66T	SL66T	B0	0F24h
2.40	400	3.2	No	SL6E9	SL6GS	C1	0F27h
2.40	400	3.2	No	SL6GS	SL6GS	C1	0F27h
2.40A	533	4.3	No	SL7E8	SL7E8	C0	0F33h
2.40B	533	4.3	No	SL684	SL67Z	B0	0F24h
2.40B	533	4.3	No	SL67Z	SL67Z	B0	0F24h
2.40B	533	4.3	No	SL6EU	SL6D7	B0	0F24h
2.40B	533	4.3	No	SL6EF	SL6DV	C1	0F27h
2.40B	533	4.3	No	SL6DV	SL6DV	C1	0F27h
2.40B	533	4.3	No	SL6QP	SL6PM	D1	0F29h
2.40C	800	6.4	Yes	SL6WR	SL6WF	D1	0F29h
2.40C	800	6.4	Yes	SL6Z3	SL6Z3	M0	0F25h
2.50	400	3.2	No	SL6EB	SL6GT	C1	0F27h
2.50	400	3.2	No	SL6GT	SL6GT	C1	0F27h
2.50	400	3.2	No	SL6QQ	SL6QQ	D1	0F29h

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
512K	0K	68°C	52.4W	478	130nm	55M	N/A
512K	0K	68°C	52.4W	478	130nm	55M	N/A
512K	0K	68°C	52.4W	478	130nm	55M	N/A
512K	0K	68°C	52.4W	478	130nm	55M	N/A
512K	0K	69°C	54.3W	478	130nm	55M	N/A
512K	0K	69°C	54.3W	478	130nm	55M	N/A
512K	0K	74°C	54.3W	478	130nm	55M	N/A
512K	0K	69°C	55.1W	478	130nm	55M	N/A
512K	0K	69°C	55.1W	478	130nm	55M	N/A
512K	0K	69°C	55.1W	478	130nm	55M	N/A
512K	0K	69°C	55.1W	478	130nm	55M	N/A
512K	0K	70°C	57.1W	478	130nm	55M	N/A
512K	0K	70°C	57.1W	478	130nm	55M	N/A
512K	0K	70°C	57.1W	478	130nm	55M	N/A
512K	0K	70°C	56.0W	478	130nm	55M	N/A
512K	0K	70°C	56.0W	478	130nm	55M	N/A
512K	0K	70°C	56.0W	478	130nm	55M	N/A
512K	0K	70°C	58.0W	478	130nm	55M	N/A
512K	0K	70°C	58.0W	478	130nm	55M	N/A
512K	0K	70°C	58.0W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	71°C	59.8W	478	130nm	55M	N/A
512K	0K	71°C	59.8W	478	130nm	55M	N/A
1M	0K	69°C	89.0W	478	90nm	125M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	70°C	57.8W	478	130nm	55M	N/A
512K	0K	71°C	59.8W	478	130nm	55M	N/A
512K	0K	71°C	59.8W	478	130nm	55M	N/A
512K	0K	74°C	66.2W	478	130nm	55M	N/A
512K	0K	74°C	66.2W	478	130nm	55M	N/A
512K	0K	72°C	74.5W	478	130nm	55M	N/A
512K	0K	72°C	61.0W	478	130nm	55M	N/A
512K	0K	72°C	61.0W	478	130nm	55M	N/A
512K	0K	72°C	61.0W	478	130nm	55M	N/A
512K	0K	72°C	61.0W	478	130nm	55M	N/A

Table 3.47 Continued

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
2.53	533	4.3	No	SL685	SL682	B0	0F24h
2.53	533	4.3	No	SL682	SL682	B0	0F24h
2.53	533	4.3	No	SL6EV	SL6D8	B0	0F24h
2.53	533	4.3	No	SL6EG	SL6DW	C1	0F27h
2.53	533	4.3	No	SL6DW	SL6DW	C1	0F27h
2.53	533	4.3	No	SL6Q9	SL6PD	D1	0F29h
2.60	400	3.2	No	SL6HB	SL6GU	C1	0F27h
2.60	400	3.2	No	SL6GU	SL6GU	C1	0F27h
2.60	400	3.2	No	SL6QR	SL6QR	D1	0F29h
2.60B	533	4.3	No	SL6S3	SL6S3	C1	0F27h
2.60B	533	4.3	No	SL6QA	SL6PE	D1	0F29h
2.60C	800	6.4	Yes	SL6WS	SL6WH	D1	0F29h
2.60C	800	6.4	Yes	SL78X	N/A	D1	0F29h
2.66	533	4.3	No	SL6DX	SL6DX	C1	0F27h
2.66	533	4.3	No	SL6EH	N/A	C1	0F27h
2.66	533	4.3	No	SL6S3	SL6S3	C1	0F27h
2.66	533	4.3	No	SL6SK	N/A	C1	0F27h
2.66	533	4.3	No	SL6PE	SL6PE	D1	0F29h
2.66	533	4.3	No	SL6QA	N/A	D1	0F29h
2.66	533	4.3	No	SL7E9	N/A	C0	0F33h
2.66	533	4.3	No	SL7YU	N/A	D0	0f34h
2.66	533	4.3	No	SL85U	N/A	E0	0F41H
2.80	400	3.2	No	N/A	SL7EY	D1	0F29h
2.80	533	4.3	No	SL6K6	SL6HL	C1	0F27h
2.80	533	4.3	No	SL6HL	SL6HL	C1	0F27h
2.80	533	4.3	No	SL6SL	SL6S4	C1	0F27h
2.80	533	4.3	No	SL6S4	SL6S4	C1	0F27h
2.80	533	4.3	No	SL6QB	SL6PF	D1	0F29h
2.80	533	4.3	No		SL7PK	E0	0F41h
2.80	533	4.3	No	SL88G	SL88G	E0	0F41h
2.80	800	6.4	Yes	SL6WJ	SL6WJ	D1	0F29h
2.80	800	6.4	Yes	SL6WT	SL6WT	D1	0F29h
2.80	800	6.4	Yes	SL6Z5	N/A	M0	0F25h
2.80	800	6.4	Yes	SL7E2	SL7E2	D0	0f34h
2.80	800	6.4	Yes	SL7E3	SL7E3	D0	0f34h
2.80	800	6.4	Yes	N/A	SL7J5	D0	0f34h
2.80	800	6.4	Yes	SL7KA	SL7KA	D0	0f34h
2.80	800	6.4	Yes	N/A	SL7J5	D0	0f34h

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
512K	0K	71°C	59.3W	478	130nm	55M	N/A
512K	0K	71°C	59.3W	478	130nm	55M	N/A
512K	0K	71°C	59.3W	478	130nm	55M	N/A
512K	0K	72°C	61.5W	478	130nm	55M	N/A
512K	0K	72°C	61.5W	478	130nm	55M	N/A
512K	0K	72°C	61.5W	478	130nm	55M	N/A
512K	0K	72°C	62.6W	478	130nm	55M	N/A
512K	0K	72°C	62.6W	478	130nm	55M	N/A
512K	0K	75°C	69.0W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	73°C	66.1W	478	130nm	55M	N/A
512K	0K	73°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
512K	0K	74°C	66.1W	478	130nm	55M	N/A
1M	0K	73.1°C	103W	478	90nm	125M	N/A
1M	0K	69.1°C	84W	775	90nm	125M	505
1M	0K	67.7°C	84W	775	90nm	125M	505
512K	0K	75°C	68.4W	478	130nm	55M	N/A
512K	0K	75°C	68.4W	478	130nm	55M	N/A
512K	0K	75°C	68.4W	478	130nm	55M	N/A
512K	0K	75°C	68.4W	478	130nm	55M	N/A
512K	0K	75°C	68.4W	478	130nm	55M	N/A
512K	0K	75°C	69.7W	478	130nm	55M	N/A
1M	0K	69.1°C	89.0W	478	90nm	125M	N/A
1M	0K	69.1°C	89.0W	478	90nm	125M	N/A
512K	0K	75°C	69.7W	478	130nm	55M	N/A
512K	0K	75°C	69.7W	478	130nm	55M	N/A
512K	0K	73°C	76.0W	478	130nm	55M	N/A
1M	0K	69.1°C	89.0W	478	90nm	125M	N/A
1M	0K	69.1°C	89.0W	478	90nm	125M	N/A
1M	0K	67.7°C	84.0W	775	90nm	125M	520
1M	0K	69.1°C	89.0W	478	90nm	125M	N/A
1M	0K	67.7°C	84.0W	775	90nm	125M	520

Table 3.47 Continued

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
2.80	800	6.4	Yes	N/A	SL7PL	E0	0F41h
2.80	800	6.4	No	N/A	SL7PT	E0	0F41h
2.80	800	6.4	Yes	N/A	SL88H	E0	0F41h
2.80	800	6.4	Yes	SL8HX	SL8HX	E0	0F41h
2.80A	533	4.3	No	SL7K9	SL7K9	D0	0f34h
2.80A	533	4.3	No	SL7D8	SL7D8	C0	0F33h
2.80C	800	6.4	Yes	SL78Y	N/A	D1	0F29h
2.80E	800	6.4	Yes	SL79K	SL79K	C0	0F33h
2.93	533	6.4	No	N/A	SL7YV	D0	0f34h
2.93	533	6.4	No	N/A	SL85V	E0	0F41h
3.0	800	6.4	Yes	SL6WU	SL6WK	D1	0F29h
3.0	800	6.4	Yes	SL78Z	N/A	D1	0F29h
3.0	800	6.4	Yes	SL7BK	N/A	M0	0F25h
3.0	800	6.4	Yes	SL7E4	SL7E4	D0	0f34h
3.0	800	6.4	Yes	SL7KB	SL7KB	D0	0f34h
3.0	800	6.4	Yes	SL7PM	SL7PM	E0	0F41h
3.0	800	6.4	Yes	SL7PU	SL7PU	E0	0F41h
3.0	800	6.4	Yes	SL7Z9	SL7Z9	N0	0F43h
3.0	800	6.4	Yes	SL88J	N/A	E0	0F41h
3.00	800	6.4	Yes	SL7KK	SL7KK	D0	0f34h
3.00	800	6.4	Yes	SL7J6	SL7J6	D0	0f34h
3.0E	800	6.4	Yes	SL79L	SL79L	C0	0F33h
3.06	533	4.3	Yes	SL6K7	SL6JJ	C1	0F27h
3.06	533	4.3	Yes	SL6JJ	SL6JJ	C1	0F27h
3.06	533	4.3	Yes	SL6SM	SL6S5	C1	0F27h
3.06	533	4.3	Yes	SL6S5	SL6S5	C1	0F27h
3.06	533	4.3	Yes	SL6QC	SL6PG	D1	0F29h
3.06	533	4.3	No	N/A	SL87L	E0	0F41h
3.20	800	6.4	Yes	SL6WE	SL6WG	D1	0F29h
3.20	800	6.4	Yes	SL792	N/A	D1	0F29h
3.20	800	6.4	Yes	SL79M	SL79M	C0	0F33h
3.20	800	6.4	Yes	SL7B8	SL7B8	C0	0F33h
3.20	800	6.4	Yes	SL7E5	SL7E5	D0	0f34h
3.20	800	6.4	Yes	SL7J7	SL7J7	D0	0f34h
3.20	800	6.4	Yes	SL7KC	SL7KC	D0	0f34h
3.20	800	6.4	Yes	SL7KL	SL7KL	D0	0f34h
3.20	800	6.4	Yes	SL7LA	SL7LA	D0	0f34h
3.20	800	6.4	Yes	SL7PN	SL7PN	E0	0F41h

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
1M	OK	69.1°C	89.0W	775	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	505
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	521 ²
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	69°C	89.0W	478	90nm	125M	N/A
512K	OK	75°C	69.7W	478	130nm	55M	N/A
1M	OK	69°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	515
1M	OK	67.7°C	84.0W	775	90nm	125M	515
512K	OK	70°C	81.9W	478	130nm	55M	N/A
512K	OK	70°C	81.9W	478	130nm	55M	N/A
512k	OK	66°C	82.0W	478	130nm	55M	N/A
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	530J
2M	OK	67.7°C	84.0W	775	90nm	169M	630 ²
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	530
1M	OK	67.7°C	84.0W	775	90nm	125M	530
1M	OK	69°C	89.0W	478	90nm	125M	N/A
512K	OK	69°C	81.8W	478	130nm	55M	N/A
512K	OK	69°C	81.8W	478	130nm	55M	N/A
512K	OK	69°C	81.8W	478	130nm	55M	N/A
512K	OK	69°C	81.8W	478	130nm	55M	N/A
512K	OK	69°C	81.8W	478	130nm	55M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	519
512K	OK	70°C	82.0W	478	130nm	55M	N/A
512K	OK	70°C	82.0W	478	130nm	55M	N/A
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	540
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	540
1M	OK	67.7°C	103.0W	775	90nm	125M	N/A ³
1M	OK	73.2°C	103.0W	775	90nm	125M	N/A

Table 3.47 Continued

CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping	CPUID
3.20	800	6.4	Yes	SL7PW	SL7PW	E0	0F41h
3.20	800	6.4	Yes	N/A	SL7PX	E0	0F41h
3.20	800	6.4	Yes	SL7Z8	SL7Z8	E0	0F43h
3.20	800	6.4	Yes	SL88K	N/A	E0	0F41h
3.2EE	800	6.4	Yes	SL7AA	SL7AA	M0	0F25h
3.40	800	6.4	Yes	SL793	SL793	D1	0F29h
3.40	800	6.4	Yes	SL7AJ	SL7AJ	C0	0F33h
3.40	800	6.4	Yes	SL7B9	N/A	C0	0F33h
3.40	800	6.4	Yes	SL7E6	SL7E6	D0	0f34h
3.40	800	6.4	Yes	SL7J8	SL7J8	D0	0f34h
3.40	800	6.4	Yes	SL7KD	SL7KD	E0	0F41h
3.40	800	6.4	Yes	SL7KM	SL7KM	D0	0f34h
3.40	800	6.4	Yes	SL7LH	SL7LH	D0	0f34h
3.40	800	6.4	Yes	N/A	SL7PP	E0	0F41h
3.40	800	6.4	Yes	SL7PY	SL7PY	E0	0F41h
3.40	800	6.4	Yes	N/A	SL7PZ	E0	0F41h
3.40	800	6.4	Yes	SL7RR	SL7RR	M0	0F25h
3.40	800	6.4	Yes	SL7Z7	SL7Z7	N0	0F43h
3.4EE	800	6.4	Yes	SL7CH	SL7CH	M0	0F25h
3.4EE	800	6.4	Yes	SL7GD	SL7GD	M0	0F25h
3.46EE	1066	8.5	Yes	SL7NF	SL7NF	M0	0F25h
3.46EE	1066	8.5	Yes	N/A	SL7RT	M0	0F25h
3.60	800	6.4	Yes	SL7J9	SL7J9	D0	0f34h
3.60	800	6.4	Yes	N/A	SL7KN	D0	0f34h
3.60	800	6.4	Yes	SL7L9	SL7L9	D0	0f34h
3.60	800	6.4	Yes	N/A	SL7NZ	E0	0F41h
3.60	800	6.4	Yes	SL7Q2	SL7Q2	E0	0F41h
3.60	800	6.4	Yes	SL8J6	SL8J6	E0	0F41h
3.60	800	6.4	Yes	SL7Z5	SL7Z5	N0	0F43h
3.80	800	6.4	Yes	SL7P2	SL7P2	E0	0F41h
3.80	800	6.4	Yes	SL82U	SL82U	E0	0F41h
3.80	800	6.4	Yes	N/A	SL8J7	E0	0F41h
3.80	800	6.4	Yes	N/A	SL7Z3	N0	0F43h

180nm = Willamette core

130nm = Northwood core

90nm = Prescott core

L2 Cache	L3 Cache	Max. Temp	Max. Power	Socket	Process	Transistors	Processor Model Number, Notes
1M	OK	67.7°C	84.0W	775	90nm	125M	540J
1M	OK	67.7°C	84.0W	775	90nm	125M	540 ³
2M	OK	67.7°C	84.0W	775	90nm	169M	640 ²
1M	OK	69.1°C	89.0W	478	90nm	125M	N/A
512K	2M	64°C	92.1W	478	130nm	178M	N/A
512K	OK	70°C	89.0W	478	130nm	55M	N/A
1M	OK	73°C	103.0W	478	90nm	125M	N/A
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	72.8°C	115.0W	775	90nm	125M	550
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	72.8°C	115.0W	775	90nm	125M	550
1M	OK	72.8°C	115.0W	775	90nm	125M	N/A ³
1M	OK	73.2°C	103.0W	478	90nm	125M	N/A
1M	OK	67.7°C	84.0W	775	90nm	125M	550J
1M	OK	67.7°C	84.0W	775	90nm	125M	550 ³
512K	2M	66°C	109.6W	775	130nm	169M	N/A
2MB	OK	67.7°C	84.0W	775	90nm	125M	650 ²
512K	2M	68°C	102.9W	478	130nm	178M	N/A
512K	2M	66°C	109.6W	775	130nm	178M	N/A
512K	2M	66°C	110.7W	775	130nm	178M	N/A
512K	2M	66°C	110.7W	775	130nm	178M	N/A
1M	OK	72.8°C	115.0W	775	90nm	125M	560
1M	OK	72.8°C	115.0W	775	90nm	125M	560
1M	OK	72.8°C	115.0W	775	90nm	125M	560 ³
1M	OK	72.8°C	115.0W	775	90nm	125M	560 ³
1M	OK	72.8°C	115.0W	775	90nm	125M	560J
1M	OK	72.8°C	115.0W	775	90nm	125M	N/A ²
2MB	OK	72.8°C	115.0W	775	90nm	125M	660 ²
1M	OK	72.8°C	115.0W	775	90nm	125M	N/A ²
1M	OK	72.8°C	115.0W	775	90nm	125M	570J
1M	OK	72.8°C	115.0W	775	90nm	125M	N/A ²
2MB	OK	72.8°C	115.0W	775	90nm	125M	670 ²

1. HT = Hyper-Threading Technology; EE = Extreme Edition.

2. This processor supports Intel Extended Memory 64 Technology (EM64T) and Execute Disable Bit (NX).

3. This processor supports EM64T.

Memory Requirements

Pentium 4–based motherboards use RDRAM, SDRAM, DDR SDRAM, or DDR2 SDRAM memory, depending on the chipset; however, most Pentium 4 systems use DDR or DDR2 SDRAM. Since Intel's contract with RAMBUS expired in 2001, DDR SDRAM and DDR2 SDRAM have become Intel's preferred memory type for mainstream systems.

Note

The early Pentium 4 motherboards that used RDRAM used the same RAMBUS RDRAM RIMM modules introduced for use with some of the chipsets used in Pentium III motherboards. However, the dual RDRAM channels the Pentium 4 uses require you to install pairs of identical modules (called *RIMMs*). Pentium 4 motherboards that use RDRAM accept either one or two pairs of RIMMs. Both pairs of memory must be the same speed, but need not be the same size.

Pentium 4 Power Supply and Cooling Issues

Compared to older processors, the Pentium 4 requires a lot of electrical power, and because of this, most Pentium 4 motherboards use a new design voltage regulator module powered from 12V instead of 3.3V or 5V, as with previous designs. By using the 12V power, more 3.3V and 5V power is available to run the rest of the system and the overall current draw is greatly reduced with the higher voltage as a source. PC power supplies generate more than enough 12V power, but the ATX motherboard and power supply design originally allotted only one pin for 12V power (each pin is rated for only 6 amps), so additional 12V lines were necessary to carry this power to the motherboard.

The fix appears in the form of a third power connector, called the ATX12V connector. This new connector is used in addition to the standard 20-pin ATX power supply connector and 6-pin auxiliary (3.3V/5V) connector. Fortunately, the power supply itself doesn't require a redesigned power supply; more than enough 12V power is available from the drive connectors. To utilize this, companies such as PC Power and Cooling sell an inexpensive (\$8) adapter that converts a standard Molex-type drive power connector to the ATX12V connector. Typically, a 300-watt (the minimum recommended) or larger power supply has more than adequate levels of 12V power for both the drives and the ATX12V connector.

If your power supply is less than the 300-watt minimum recommended, you need to purchase a replacement. Because the ATX12V power supply connector is required for most Intel-based systems from the past few years, virtually all vendors sell an off-the-shelf ATX12V-ready model or one that uses the adapter mentioned previously.

►► See "Motherboard Power Connectors," p. 1167.

Cooling a high-wattage unit such as the Pentium 4 requires a large active heatsink. These heavy (sometimes more than 1 lb.) heatsinks can damage a CPU or destroy a motherboard when subjected to vibration or shock, especially during shipping. To solve this problem with Pentium 4 motherboards, various methods have been used. Intel's specifications for Socket 423 added four standoffs to the ATX chassis design flanking the Socket 423 to support the heatsink retention brackets. These standoffs enabled the chassis to support the weight of the heatsink instead of depending on the motherboard, as with older designs. Vendors also used other means to reinforce the CPU location without requiring a direct chassis attachment. For example, Asus's P4T motherboard was supplied with a metal reinforcing plate to enable off-the-shelf ATX cases to work with the motherboard.

Socket 478 systems do not require any special standoffs or reinforcement plates; instead they use a unique scheme in which the CPU heatsink attaches directly to the motherboard rather than to the CPU socket or chassis. Motherboards with Socket 478 can be installed into any ATX chassis—no special standoffs are required.

Socket T (LGA775) systems use a unique clamping mechanism that holds the processor in place. The heatsink is attached over the processor and clamping mechanism and attaches to the motherboard.

Because the Pentium 4 processor family has been manufactured in three socket types with a wide variation in clock speed and power dissipation, it's essential that you choose a heatsink made specifically for the processor form factor and speed you have purchased (or intend to purchase). This is just one more reason I think it's worth getting a boxed processor instead of an OEM version when building or upgrading a system. If you purchase the shrink-wrapped or "boxed" processor, you get an Intel-specified high-quality heatsink in the box with the process. In addition, you get a 3-year warranty with Intel, making the boxed version ideal for upgraders and system builders.

Xeon Processors

Xeon processors are based on the Pentium 4 and are designed for Socket 603 and Socket 604. Xeon DP processors (often referred to simply as *Xeon*) are designed for single- and dual-processor workstations:

- Xeon DP processors with a 400MHz CPU bus feature clock speeds from 1.4GHz to 3GHz.
- Xeon DP processors with a 533MHz CPU bus feature clock speeds from 2GHz to 3.2GHz.
- Xeon DP processors with a 667MHz CPU bus (a speed never used by the Pentium 4, by the way) feature clock speeds from 3.33GHz to 3.66GHz.
- Xeon DP processors with an 800MHz CPU bus feature clock speeds from 2.8GHz to 3.8GHz.

Xeon MP processors are designed for four-way and larger servers. They are available in speeds ranging from 1.4GHz to 3GHz, and all support the 400MHz CPU bus.

For more information about Xeon DP and Xeon MP processors, see my book *Upgrading and Repairing Servers*.

Eighth-Generation (64-Bit Register) Processors

As of 2001, it had been about 15 years since PCs had begun to support 32-bit processors (all processors from the 80386 up through the Intel Pentium 4 and AMD Athlon XP). However, in 2001, Intel introduced the first 64-bit processor for servers—the Itanium—followed in 2002 by the improved Itanium 2. In 2003, AMD introduced the first 64-bit processor for x86-compatible desktop computers—the Athlon 64—followed by its first 64-bit server processor, the Opteron. In 2004, Intel introduced a series of 64-bit-enabled versions of its Pentium 4 desktop processor. Then in 2005, Intel introduced 64-bit versions of its Xeon workstation and server processors and new 64-bit desktop processors—the Pentium Extreme Edition and dual-core Pentium D.

The following sections discuss the major features of these processors and the different approaches taken by Intel and AMD to bring 64-bit computing to the PC server and desktop.

Intel Itanium and Itanium 2

Introduced on May 29, 2001, the Itanium was the first processor in Intel's IA-64 (Intel Architecture 64-bit) product family, and it incorporated innovative performance-enhancing architecture techniques, such as prediction and speculation. It and its newer sibling, the Itanium 2 (introduced in June 2002), are the highest-end processors from Intel and are designed mainly for the server market.

If Intel was still using numbers to designate its processors, the Itanium family might be called the 886 because the Itanium and Itanium 2 are the eighth-generation processors in the Intel family, and they represent the most significant processor architecture advancement since the 386.

Intel's IA-64 product family is designed to expand the capabilities of the Intel architecture to address the high-performance server and workstation market segments.

The Itanium and Itanium 2 were never designed to replace the Pentium 4. They feature an all-new design that is initially expensive and is found only in the highest-end systems such as file servers or advanced workstations.

The Itanium's technical details are listed in Table 3.48.

Table 3.48 Intel Itanium and Itanium 2 Technical Details

Processor	Processor Speed	L2 Cache	L3 Cache Size	FSB Speed	Memory Bus Width	Bandwidth	Number of Transistors
Itanium	733MHz 800MHz	96KB	2MB ¹ or 4MB ¹	266MHz	64-bit	2.1GBps	25 million (core) 150 or 300 million (cache)
Itanium 2	900MHz	256KB	1.5MB ²	400MHz	128-bit	6.4GBps	221 million
Itanium 2	1GHz ⁴ 1.3GHz ⁴	256KB	3MB ²	400MHz	128-bit	6.4GBps	221 million
Itanium 2	1.4GHz ³	256KB	1.5MB	400MHz	128-bit	6.4GBps	221 million
Itanium 2	1.6GHz ³	256KB	3MB	400MHz 533MHz	128-bit	6.4GBps 8.5GBps	500 million
Itanium 2	1.4GHz 1.5GHz ⁵ 1.6GHz ³	256KB	4MB ²	400MHz	128-bit	6.4GBps	410 million
Itanium 2	1.5GHz 1.6GHz ⁵	256KB	6MB ²	400MHz	128-bit	6.4GBps	500 million
Itanium 2	1.66GHz ⁵	256KB	6MB ²	667MHz	128-bit	10.6GBps	500 million
Itanium 2	1.6GHz ⁵	256KB	9MB ²	400MHz	128-bit	6.4GBps	592 million
Itanium 2	1.66GHz ⁵	256KB	9MB ²	667MHz	128-bit	10.6GBps	592 million

1. On-cartridge, full-speed unified 128-bit wide

2. On-die, full-speed unified 128-bit wide

3. Optimized for dual-processor operation (DP Optimized)

4. Also available in a low-voltage version

5. Optimized for multiple-processor operation

As noted in Table 3.48, the Itanium and Itanium 2 are the first Intel processors with three levels of integrated cache. Even though a few previous system designs featured L3 cache, the L3 cache was located on the motherboard and was therefore much slower. By building L3 cache in to the cartridge (Itanium) or on the processor die (Itanium 2), all three cache levels run at the full processor speed.

The following features apply to both Itanium and Itanium 2 processors:

- 16TB (terabytes) physical memory addressing (44-bit address bus).
- Full 32-bit instruction compatibility in hardware.
- EPIC (explicitly parallel instruction computing) technology, which enables up to 20 operations per cycle.
- Two integer and two memory units that can execute four instructions per clock.
- Two FMAC (floating-point multiply accumulate) units with 82-bit operands.
- Each FMAC unit is capable of executing two floating-point operations per clock.
- Two additional MMX units are capable of executing two single-precision FP operations each.
- A total of eight single-precision FP operations can be executed every cycle.
- 128 integer registers, 128 floating-point registers, 8 branch registers, 64 predicate registers.

The Itanium 2 also features

- 400MHz, 533MHz, or 667MHz CPU Bus (versus 266MHz for Itanium)
- 128-bit-wide CPU Bus (versus 64-bit-wide for Itanium)

Itanium and Itanium 2 were initially based on 0.18-micron technology; however, current versions of the Itanium 2 are based on 0.13-micron, allowing for higher speeds and larger caches.

The original Itanium used a cartridge known as the pin array cartridge (PAC). This cartridge includes L3 cache and plugs into a PAC418 (418-pin for Itanium) or PAC611 (611-pin for Itanium 2) socket on the motherboard and not a slot. The package is about the size of a standard index card, weighs about 6oz. (170g), and has an alloy metal on its base to dissipate the heat (see Figure 3.62). Itanium has clips on its sides, enabling four of them to be hung from a motherboard, both below and above.

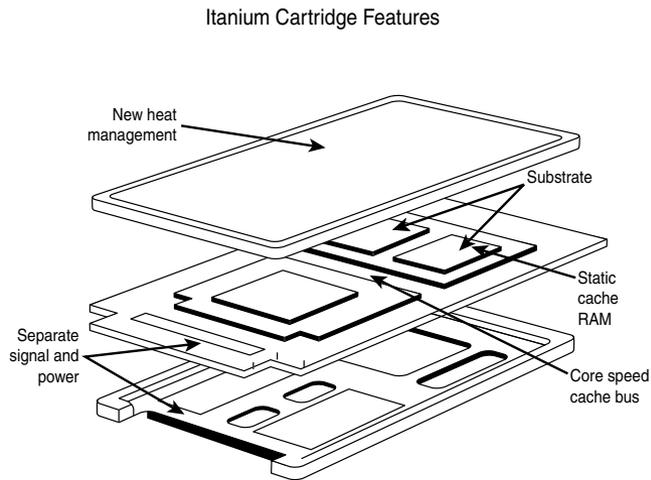


Figure 3.62 The Itanium's pin array cartridge.

The first Itanium 2 was codenamed McKinley and officially introduced in June 2002. The current version uses the 0.13-micron Madison core, which has up to a whopping 592 million transistors in its 9MB on-die L3 cache version. Because the Itanium 2 has a significantly higher CPU bus bandwidth (up to 10.6GBps), higher clock speeds, larger caches, and a processor FSB twice as wide (128 bits) as the original Itanium, the Itanium 2 is significantly faster in overall processing. The Itanium 2 integrates all three levels of cache inside the processor die, so a cartridge is unnecessary (see Figure 3.63). The Itanium and Itanium 2 are not interchangeable and are supported by different sockets and chipsets.

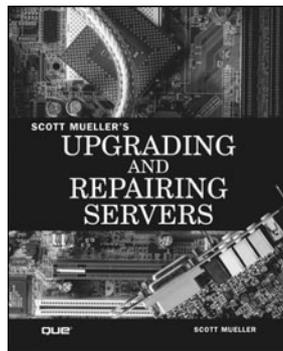


Figure 3.63 The Itanium 2 is a more compact design than the original Itanium. *Photograph used by permission of Intel Corporation.*

The Itanium and Itanium 2 are supported by a variety of operating systems, including Microsoft Windows (XP 64-bit Edition and 64-bit Windows Advanced Server Limited Edition 2002), Linux (from four distributor companies: Red Hat, SuSE, Caldera, and Turbo Linux), and two Unix versions (Hewlett-Packard's HP-UX and IBM's AIX).

Tip

If you need to run 32-bit x86 software on an Itanium 2 processor, be sure your OS supports the IA-32 Execution Layer (IA-32 EL) technology. IA-32 EL improves the performance of 32-bit software on the Itanium 2 processor. Operating systems that include or support IA-32 EL include Windows Server 2003 Enterprise Edition, Windows Server 2003 Data Center, Windows XP 64-bit Edition, and most current Linux distros that support Itanium 2. You can download IA-32 EL for Red Hat Enterprise Linux 4; Red Hat Enterprise Linux 3 UP5; Red Hat Enterprise 3 UP4; SUSE Enterprise Server 9 SP1; SUSE Enterprise Server Linux SP1, Kernel 2.6 from the "IA-32 Execution Layer" page at <http://www.intel.com/cd/software/products/asmo-na/eng/219773.htm>.

Although Itanium 2 has broad OS support, it has not proven to be as popular as Intel and hardware vendors initially hoped. Although Intel continues to develop new versions of the Itanium 2 platform, it's more likely that your first foray into 64-bit computing will use one of the AMD or Intel processors discussed in the following sections. That's because other 64-bit processors natively use extensions of the existing IA-32 architecture for full-speed 32-bit and 64-bit computing and cost little more (if anything) than comparable 32-bit-only processors.

AMD Athlon 64 and 64 FX

The AMD Athlon 64 and 64 FX, introduced in September 2003, are the first 64-bit processors for desktop (and not server) computers. Originally code named ClawHammer, the Athlon 64 and 64 FX are the desktop element of AMD's 64-bit processor family, which also includes the Opteron (code named SledgeHammer) server processor. The Athlon 64 and 64 FX (shown in Figure 3.64) are essentially Opteron chips but are designed for single-processor systems, and in some cases have decreased cache or memory bandwidth capabilities.



Figure 3.64 AMD Athlon 64 FX (Socket 939 version). *Photo courtesy of AMD.*

Besides support for 64-bit instructions, the biggest difference between the Athlon 64 and 64 FX and other processors is the fact that their memory controller is built in. The memory controller is normally part of the motherboard chipset North Bridge or memory controller hub (MCH) chip, but with the Athlon 64 and 64 FX, the memory controller is now built in to the processor. This means that the typical CPU bus architecture is different with these chips. In a conventional design, the processor talks to the chipset North Bridge, which then talks to the memory and all other components in the

system. Because the Athlon 64 and 64 FX have integrated memory controllers, they talk to memory directly, and also talk to the North Bridge for other system communications. Separating the memory traffic from the CPU bus allows for greatly improved performance not only in memory transfers, but also in CPU bus transfers. The main difference in the Athlon 64 and 64 FX is in the different configurations of cache sizes and memory bus widths.

The major features of the Athlon 64 design include

- Speeds ranging from 1.8GHz to 2.4GHz.
- 68.5 million transistors (512KB L2 cache versions) or 114 million transistors (1MB L2 cache versions).
- 12-stage pipeline.
- DDR memory controller with ECC support integrated into the processor (instead of the North Bridge or MCP, as in other recent chipsets).
- Socket 754 features single-channel memory controller; Socket 939 features dual-channel memory controller.
- 128KB L1 cache (some Athlon 64s include up to 1MB).
- 512KB or 1MB of on-die full-speed L2 cache.
- Support for AMD64 (also called IA-32e or x86-64) 64-bit extension technology (extends 32-bit x86 architecture).
- 3.2GBps (Socket 754) or 4GBps (Socket 939) Hypertransport link to chipset North Bridge.
- Addressable memory size up to 1TB, greatly exceeding the 4GB or 64GB limit imposed by 32-bit processors.
- SSE2 (SSE plus 144 new instructions for graphics and sound processing).
- Multiple low-power states.
- 130-nanometer (ClawHammer, Newcastle cores) or 90-nanometer (Winchester, Venice, San Diego cores).

The Athlon 64 FX differs from the standard Athlon 64 in the following ways:

- Supports only Socket 939 or Socket 940 (initial versions).
- Has dual-channel DDR memory controller with ECC support.
- Socket 940 versions require registered memory.
- Features speeds from 2.2GHz to 2.8GHz.
- 1MB L2 cache (standard).

Although Socket 939 versions of the Athlon 64 have closed the performance gap, the Athlon 64 FX is still the fastest single-core Athlon 64 processor.

Although AMD has been criticized by many, including me, for its confusing performance-rating processor names in the Athlon XP series, AMD also uses this naming scheme with the Athlon 64. As I suggest with the Athlon XP, you should look at the actual performance of the processor with the applications you use most to determine whether the Athlon 64 is right for you and which model is best suited to your needs. The integrated memory bus in the Athlon 64 means that the Athlon 64 connects to memory more directly than any 32-bit chip and makes North Bridge design simpler. AMD offers its own chipsets for the Athlon 64, but most Athlon 64 motherboards and systems use third-party chipsets from the same vendors that now produce Athlon XP chipsets. See Chapter 4 for details.

The various models and features of the Athlon 64 and 64 FX are summed up in Tables 3.49 and 3.50.

Table 3.49 Athlon 64 Processor Information

Part Number*	Model Number	CPU Speed	Bus Speed (GBps)	Stepping
ADA2800AEP4AX	2800+	1.8GHz	3.2	CG
ADA2800AEP4AP	2800+	1.8GHz	3.2	C0
ADA2800AEP4AR	2800+	1.8GHz	3.2	CG
ADA3000AIK4BX	3000+	2.0GHz	3.2	E6
ADA3000AEP4AP	3000+	2.0GHz	3.2	C0
ADA3000AEP4AR	3000+	2.0GHz	3.2	CG
ADA3000AEP4AX	3000+	2.0GHz	3.2	CG
ADA3000DAA4BW	3000+	1.8GHz	4.0	E6
ADA3000DAA4BP	3000+	1.8GHz	4.0	E3
ADA3000DIK4BI	3000+	1.8GHz	4.0	D0
ADA3000DEP4AW	3000+	1.8GHz	4.0	CG
ADA3200DKA4CG	3200+	2.0GHz	4.0	E4
ADA3200AI04BX	3200+	2.0GHz	3.2	E6
ADA3200DAA4BW	3200+	2.0GHz	4.0	E6
ADA3200DAA4BP	3200+	2.0GHz	4.0	E3
ADA3200DIK4BI	3200+	2.0GHz	4.0	D0
ADA3200DEP4AW	3200+	2.0GHz	4.0	CG
ADA3200AEP5AP	3200+	2.0GHz	3.2	C0
ADA3200AEP5AR	3200+	2.0GHz	3.2	CG
ADA3200AEP4AX	3200+	2.2GHz	3.2	CG
ADA3400AIK4BO	3400+	2.2GHz	3.2	E3
ADA3400AEP5AP	3400+	2.2GHz	3.2	C0
ADA3400AEP4AX	3400+	2.2GHz	3.2	CG
ADA3400AEP4AR	3400+	2.2GHz	3.2	CG
ADA3400AEP5AR	3400+	2.2GHz	3.2	CG
ADA3500DAA4BN	3500+	2.2GHz	4.0	E4
ADA3500DEP4AS	3500+	2.2GHz	4.0	CG
ADA3500DKA4CG	3500+	2.2GHz	4.0	E4
ADA3500DAA4BW	3500+	2.2GHz	4.0	E6
ADA3500DAA4BP	3500+	2.2GHz	4.0	E3
ADA3500DIK4BI	3500+	2.2GHz	4.0	D0
ADA3500DEP4AW	3500+	2.2GHz	4.0	CG
ADA3700AEP5AR	3700+	2.4GHz	3.2	CG
ADA3700DAA5BN	3700+	2.2GHz	4.0	E4
ADA3700AEP5AR	3700+	2.4GHz	3.2	CG
ADA3800DEP4AW	3800+	2.4GHz	4.0	CG
ADA3800DAA4BW	3800+	2.4GHz	4.0	E6
ADA3800DEP4AS	3800+	2.4GHz	4.0	CG
ADA3800DAA4BP	3800+	2.4GHz	4.0	E3
ADA4000DEP5AS	4000+	2.4GHz	4.0	CG
ADA4000DKA5CF	4000+	2.4GHz	4.0	E6
ADA4000DAA5BN	4000+	2.4GHz	4.0	E4

*Part numbers for tray processors; PIB (processor-in-box) numbers are different.

L2 Cache	Max. Temp	Voltage	Power	Socket	Process
512K	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
512K	65°C	1.4V	51W	754	90nm
512K	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
512K	65°C	1.35V	67W	939	90nm
512K	70°C	1.35V	67W	939	90nm
512K	70°C	1.4V	67W	939	90nm
512K	70°C	1.5V	89W	939	130nm
512K	65°C	1.35V	67W	939	90nm
512K	69°C	1.4V	59W	754	90nm
512K	65°C	1.35V–1.4V	67W	939	90nm
512K	70°C	1.35V–1.4V	67W	939	90nm
512K	70°C	1.4V	67W	939	90nm
512K	70°C	1.5V	89W	939	130nm
1M	70°C	1.5V	89W	754	130nm
1M	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
512K	65°C	1.4V	67W	754	90nm
1M	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	754	130nm
1M	70°C	1.5V	89W	754	130nm
1M	70°C	1.5V	89W	754	130nm
512K	70°C	1.35V–1.4V	67W	939	90nm
512K	70°C	1.5V	89W	939	130nm
512K	65°C	1.35V	67W	939	90nm
512K	65°C	1.35V–1.4V	67W	939	90nm
512K	65°C	1.35V–1.4V	67W	939	90nm
512K	70°C	1.4V	67W	939	90nm
512K	70°C	1.5V	89W	939	130nm
1M	70°C	1.5V	89W	754	130nm
1M	70°C	1.35V–1.4V	89W	939	90nm
1M	70°C	1.5V	89W	754	130nm
512K	70°C	1.5V	89W	939	130nm
512K	70°C	1.35V–1.4V	89W	939	90nm
512K	70°C	1.5V	89W	939	130nm
512K	70°C	1.35V–1.4V	89W	939	90nm
1M	70°C	1.5V	89W	939	130nm
1M	65°C	1.35V	89W	939	90nm
1M	65°C	1.35V	89W	939	90nm

Table 3.50 Athlon 64 FX Processor Information

Part Number*	Model Number	CPU Speed	Bus Speed (GBps)	Stepping
ADAFX51CEP5AK	FX-51	2.2GHz	3.2	C0
ADAFX51CEP5AT	FX-51	2.2GHz	3.2	CG
ADAFX53CEP5AT	FX-53	2.4GHz	3.2	CG
ADAFX53DEP5AS	FX-53	2.4GHz	4.0	CG
ADAFX55DAA5BN	FX-55	2.6GHz	4.0	E4
ADAFX55DEI5AS	FX-55	2.6GHz	4.0	CG
ADAFX57DAA5BN	FX-57	2.8GHz	4.0	E4

*Part numbers for tray processors; PIB (processor-in-box) numbers are different.

The Athlon 64 and 64 FX are available in three socket versions (see Table 3.51). The Athlon 64 is available in Socket 754 and Socket 939 versions, whereas the 64 FX is available in Socket 939 and Socket 940 versions. Socket 754 supports only a single-channel memory bus, whereas Sockets 939 and 940 both support dual-channel memory for double the memory bandwidth. Socket 939 also supports faster and cheaper unbuffered DDR SDRAM DIMMs; Socket 940 supports slower and more expensive registered DIMMs. Because of this, you should avoid any Socket 940 processors or motherboards because they require registered modules that are both slower and more expensive than unbuffered types. Socket 754 versions of the Athlon 64 are also designed to use more affordable unbuffered modules, but only in single-channel mode.

Table 3.51 AMD Athlon 64 and 64 FX Socket and Memory Types

Socket	Processor	Channels	Type
754	Athlon 64	Single-channel	Unbuffered
939	Athlon 64 Athlon 64 FX	Dual-channel	Unbuffered
940	Athlon 64 FX	Dual-channel	Registered

The Athlon 64 essentially comes in two versions: a Socket 754 version that has only a single-channel memory bus and an improved Socket 939 version that has a dual-channel memory bus. The Athlon 64 FX is also available in two versions: a Socket 940 version that uses expensive (and slower) registered memory and an improved Socket 939 version that uses unbuffered memory. The Socket 939 versions of the Athlon 64 and 64 FX are essentially the same chip, differing only in the amount of L2 cache included. For example, the Athlon 64 3800+ and Athlon 64 FX-53 both run at 2.4GHz and run dual-channel memory. The only difference is that the 3800+ has only 512KB of L2 cache whereas the FX-53 has 1MB of L2. Because the 64 and 64 FX chips are essentially the same, you need to read the fine print to determine the minor differences in configuration.

The Athlon 64 and 64 FX can draw up to 104W or more of power, which is high but still somewhat less than the more power-hungry Pentium 4 processors. As with the Pentium 4, motherboards for the Athlon 64 and 64 FX require the ATX12V connector to provide adequate 12V power to run the processor voltage regulator module.

The initial version of the Athlon 64 is built on a 0.13-micron (130-nanometer) process (see Figure 3.65). Subsequent versions use a 0.09-micron (90-nanometer) process.

L2 Cache	Max. Temp	Voltage	Power	Socket	Process
1M	70°C	1.5V	89W	940	130nm
1M	70°C	1.5V	89W	940	130nm
1M	70°C	1.5V	89W	940	130nm
1M	70°C	1.5V	89W	939	130nm
1M	65°C	1.35V–1.4V	104W	939	90nm
1M	63°C	1.5V	104W	939	130nm
1M	65°C	1.35V–1.4V	104W	939	90nm

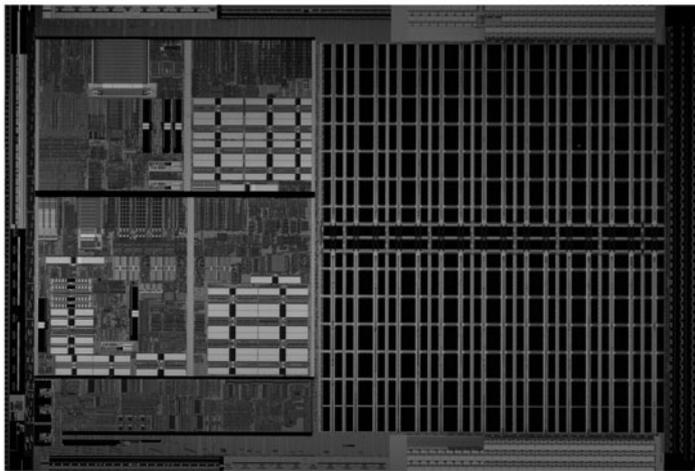


Figure 3.65 AMD Athlon 64 die (130-nanometer process, 106 million transistors, 193 sq. mm). *Photo courtesy of AMD.*

AMD Sempron (Socket 754)

Just as the Intel Celeron name long ago ceased to identify a particular processor and instead is a brand used by Intel to identify various types of low-cost, reduced-performance processors, AMD's Sempron brand follows a similar course. Sempron is used to identify both Socket A processors that have replaced the Athlon XP and Socket 754 processors that provide a low-cost alternative to the Athlon 64.

◀ See "Sempron (Socket A)," p. 174, for more information on the Socket A version of the Sempron.

The Socket 754 Sempron is based on the Socket 754 version of the Athlon 64 processor. However, some versions of the Sempron operate only in a 32-bit mode. The major features of the Socket 754 Sempron include

- 90-nanometer manufacturing process (except as noted in Table 3.52)
- 128KB or 256KB of L2 cache
- 3.2GBps HyperTransport connection to chipset
- 32-bit only or 32/64-bit operation supporting AMD64 (IA-32e or x86-64) applications
- 63.5–68.5 million transistors
- SSE3 instructions (90-nm process only)

A system using a Socket 754 Sempron processor can be easily upgraded to a Socket 754 Athlon 64 processor. Table 3.52 provides detailed information about Socket 754 Sempron processors.

Table 3.52 Sempron (Socket 754) Processors

Part Number ¹	Model Number	CPU Speed	Bus Speed (GBps)	Stepping
SDA2500AIO3BX	2500+	1.4GHz	3.2	E6
SDA2600AIO2BO	2600+	1.6GHz	3.2	E3
SDA2600AIO2BX	2600+	1.6GHz	3.2	E6
SDA2600AIO2BA	2600+	1.6GHz	3.2	D0
SDA2800AIO3BX	2800+	1.6GHz	3.2	D0
SDA2800AIO3BO	2800+	1.6GHz	3.2	E3
SDA2800AIO3BX	2800+	1.6GHz	3.2	E6
SDA3000AIO2BA	3000+	1.8GHz	3.2	D0
SDA3000AIO2BX	3000+	1.8GHz	3.2	E6
SDA3000AIO2BO	3000+	1.8GHz	3.2	E3
SDA3000AIO2BA	3000+	1.8GHz	3.2	D0
SDA3100AIP3AX	3100+	1.8GHz	3.2	CG
SDA3100AIO3BX	3100+	1.8GHz	3.2	E6
SDA3100AIO3BO	3100+	1.8GHz	3.2	E3
SDA3100AIO3BA	3100+	1.8GHz	3.2	D0
SDA3300AIO2BA	3300+	2.0GHz	3.2	D0
SDA3300AIO2BX	3300+	2.0GHz	3.2	E6
SDA3300AIO2BO	3300+	2.0GHz	3.2	E3
SDA3400AIO3BX	3400+	2.0GHz	3.2	E6

1. Part numbers for tray processors; PIB (processor-in-box) numbers are different.

2. PIB SDA2600CVBOX supports AMD64; PIB SDA2600BABOX does not.

3. PIB SDA2800CVBOX supports AMD64; PIB SDA2800BABOX does not.

As Table 3.52 indicates, most Socket 754 Sempron models support AMD64 64-bit computing. With both Intel and AMD offering entry-level 64-bit processors, it's easier than ever to move into 64-bit computing.

AMD Opteron

The AMD Opteron is the workstation and server counterpart to the AMD Athlon 64, supporting the same AMD64 (x86-64) architecture as the Athlon 64. The Opteron was introduced in the spring of 2003.

The following are the major features of the Opteron:

- 128KB L1 cache
- 1MB L2 cache
- Clock speeds of 1.8GHz–2.8GHz
- Three 3.2MBps Hypertransport links to chipset
- Socket 939 or 940
- Integrated dual-channel memory controller with ECC
- Maximum addressable memory of 1 terabyte (40-bit physical) and 256 terabytes (48-bit virtual)
- AMD64 (x86-64) architecture
- 130-nanometer or 90-nanometer production process
- Single-core or dual-core design

The Opteron is available in three series: 100 (single-processor workstations), 200 (dual-processor workstations and servers), and 800 (up to eight-way servers). Dual-core versions of Opteron processors are available in all three of these series.

L2 Cache	Max. Temp	Voltage	Power	AMD64 Support	Process
256K	69°C	1.4V	62W	Yes	90nm
128K	69°C	1.4V	62W	Yes	90nm
128K	69°C	1.4V	62W	Yes	90nm
256K	70°C	1.4V	62W	— ²	90nm
256K	70°C	1.4V	62W	— ³	90nm
256K	69°C	1.4V	62W	Yes	90nm
256K	69°C	1.4V	62W	Yes	90nm
128K	70°C	1.4V	62W	No	90nm
128K	69°C	1.4V	62W	Yes	90nm
128K	69°C	1.4V	62W	Yes	90nm
128K	70°C	1.4V	62W	Yes	90nm
256K	70°C	1.4V	62W	No	130nm
256K	69°C	1.4V	62W	Yes	90nm
256K	69°C	1.4V	62W	Yes	90nm
256K	70°C	1.4V	62W	— ⁴	90nm
128K	70°C	1.4V	62W	— ⁵	90nm
128K	69°C	1.4V	62W	Yes	90nm
128K	69°C	1.4V	62W	Yes	90nm
256K	69°C	1.4V	62W	Yes	90nm

4. PIB SDA3100CVBOX supports AMD64; PIB SDA3100BABOX does not.

5. PIB SDA3300CVBOX supports AMD64; PIB SDA3300BABOX does not.

Unlike the Itanium series, which has been supported primarily by Intel chipsets, the Opteron has broad third-party chipset support from companies such as VIA, SiS, ULI, NVIDIA, and ATI (just like the Athlon 64 does).

For more information on Opteron configurations and features, see the book *Upgrading and Repairing Servers*.

Dual-Core Processors

No matter how fast a conventional single-core processor operates or how much RAM is installed in a system, it must ensure that each program and process that is running is properly serviced. As more and more programs are opened, the amount of time the processor can devote to each program is reduced. The result is that system performance declines. Workstations and servers have long enjoyed the benefits of multiple processors, including better responsiveness when multitasking, faster performance in single multithreaded applications, and better overall throughput for both business and creativity applications (in terms of instructions processed per clock cycle).

However, the high cost of multiprocessor motherboards and multiple processors has kept most desktop computer users from enjoying the same benefits.

Note

A multithreaded application can run different parts of the program, known as *threads*, at the same time in the same address space. They can share code and data. Currently, relatively few applications other than video editing programs are multithreaded. A multithreaded program runs faster on a dual-core processor or an Intel processor with HT Technology enabled than on a single-core processor.

If you use multiple applications at the same time, such as email, web browsers, office suite components such as word processors and spreadsheets, graphics editors, and so forth, you should consider the latest development in processor technology: a dual-core processor. The dual-core processors introduced by Intel and AMD are designed to bring the benefits of multiprocessor operation to desktop systems by placing two processor cores in a single physical processor.

Dual-core processors include two processor cores in the same physical package, providing virtually all the advantages of a multiple-processor computer at a cost lower than that of two matched processors. Unlike Intel's HT Technology—which simulates two processors in a single physical unit—dual-core processors do not need specific application support to improve performance. Dual processor cores provide more time to service each running application or application thread, providing faster performance in a multitasking environment.

Intel introduced the first dual-core processors (the Pentium D and Pentium Extreme Edition) in early 2005, and AMD introduced its dual-core Opteron and Athlon 64 X2 processors shortly thereafter. Although both vendors offer dual-core processors, their designs are quite different in some ways, as are the systems that support them. Before looking at the specifics of these new processors, though, it's useful to determine whether you need a dual-core processor.

Who Needs a Dual-Core Processor?

A dual-core processor is designed for users who frequently multitask (run multiple programs at the same time) or who use multithreaded applications. Figure 3.66 illustrates how a dual-core processor handles multiple applications for faster performance.

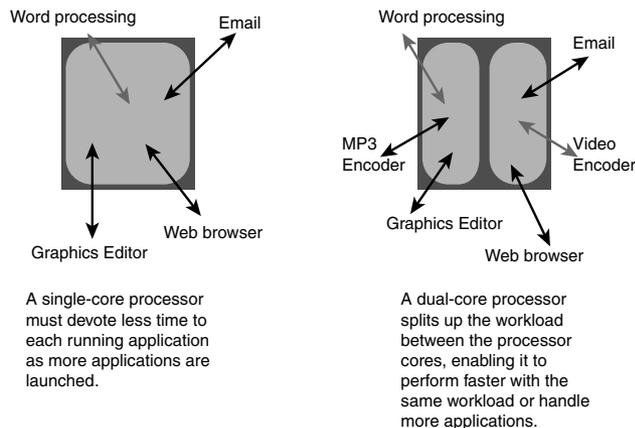


Figure 3.66 How a single-core processor (left) and a dual-core processor (right) handle multitasking.

It's important to realize that a dual-core processor does not improve single-task performance. If you play 3D games on your PC, for example, it's very likely that's all you're doing at the time so no multi-tasking is taking place that would take advantage of a dual-core CPU. Until such time as games are designed to be multithreaded, gamers might prefer to choose a high-performance single-core processor instead of a dual-core processor.

However, if you want to play 3D games at the same time as you perform other processor-intensive tasks, such as video or audio encoding, a dual-core processor might be a worthwhile investment. Benchmark tests indicate that some dual-core processors experience only slight slowdowns when playing a 3D game such as Doom 3 and performing other entertainment-oriented tasks such as audio or video encoding. Whether at work or play, a dual-core processor can help you get more done at once, if you use multiple applications.

Intel Pentium D and Pentium Extreme Edition

Intel introduced its first-dual core processors, the Pentium Extreme Edition and Pentium D, in April 2005. Although these processors used the code name Smithfield before their introductions, they are based on the Pentium 4 Prescott core. In fact, to bring dual-core processors to market as quickly as possible, Intel used two Prescott cores in each Pentium D or Pentium Extreme Edition processor. Each core communicates with the other via the MCH (North Bridge) chip on the motherboard (see Figure 3.67).

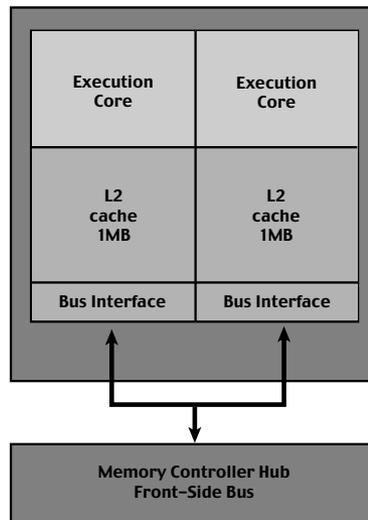


Figure 3.67 The Pentium D and Pentium Extreme Edition's processor cores communicate with each other via the chipset's MCH (North Bridge) chip.

For this reason, Intel 915 and 925 chipsets and some third-party chipsets made for the Pentium 4 cannot be used with the Pentium D or Pentium Extreme Edition. Intel's 945 series, 955X and 975X desktop chipsets, and E7230 workstation chipset are the first Intel chipsets to support these processors. The nForce 4 series from NVIDIA also works with these processors.

- ▶▶ See "Intel 945 Express Family," p. 290, and "Intel 955X and 975X Family," p. 290, for more information on these chipsets.

Because the Prescott core is the highest-wattage core Intel has produced for desktop computers and because each chip contains two cores, Intel has limited the speed of these processors to a maximum of 3.2GHz—compared to 3.8GHz for Pentium 4 processors. Even at a 3.2GHz top speed, however, the thermal design power of the Pentium Extreme Edition 840 and the Pentium D 840 is 130W, compared to 115W for Pentium 4 Prescott processors.

The major features of the Pentium D include

- Clock speeds of 2.8GHz–3.2GHz
- 800MHz processor bus
- EM64T 64-bit extensions
- Execute Disable Bit support
- 90-nanometer manufacturing process
- 2MB L2 cache (1MB per core)
- Socket T (LGA775)

The 830 and 840 models also include Enhanced Intel Speed Step Technology, which results in cooler and quieter PC operation by providing a wide range of processor speeds in response to workload and thermal issues.

The Pentium Extreme Edition 840 is similar to the Pentium D 840, but with the following differences:

- HT Technology is supported, enabling each core to simulate two processor cores for even better performance with multithreaded applications.
- Enhanced Intel Speed Step Technology is not supported.
- It includes unlocked clock multipliers, enabling easy overclocking.

Table 3.53 compares the features of the various Pentium D and Pentium Extreme Edition processors.

Although a motherboard upgrade is necessary for most users of Pentium 4 processors to move to the Pentium D or Pentium Extreme Edition, the advent of dual-core processing is an exciting one, especially for those of us who are constantly running multiple programs at the same time.

Table 3.53 Pentium D and Pentium Extreme Edition Processors

Processor	Model Number	CPU Speed (GHz)	Bus Speed (MHz)	Bus Speed (GBps)	HT Support	Boxed S-spec	OEM S-spec	Stepping
Pentium D	820	2.8	800	6.4	No	SL88T	SL88T	A0
Pentium D	820	2.8	800	6.4	No	—	SL8CP	B0
Pentium D	830	3.0	800	6.4	No	—	SL88S	A0
Pentium D	830	3.0	800	6.4	No	SL8CM	SL8CM	B0
Pentium D	840	3.2	800	6.4	No	SL8CM	SL8CM	B0
Pentium D	840	3.2	800	6.4	No	—	SL88R	A0
Pentium Extreme Edition	840	3.2	800	6.4	Yes	—	SL8FK	A0

In 2006, look for new dual-core designs that will take advantage of the forthcoming 65-nanometer production process. These processors will run cooler than the processors shown in Table 3.52, which should allow for faster clock speeds.

Intel Processor Model Numbers

Most people associate clock speed with the processor, and Intel has always used the raw clock speed of its processors to market them. This has led many people to believe that faster-speed processors always result in faster or better systems, but that is not always the case. Processor architectures have a major effect on the performance of a processor, and it is entirely possible that a slower clock speed processor can handily outperform a faster one when running actual programs or doing real work. Unfortunately, this message is hard to convey when the main attribute used to market a chip is its raw clock speed.

AMD has long been marketing its chips with model numbers, which in this case do relate to speed—but not directly. Starting in 2004, Intel also began to use model numbers, but its model numbering scheme is distinctly different from AMD's. Intel has decided to use a BMW-esque numbering scheme across its various processor families. Currently, it uses 8xx designations for its top-of-the-line desktop processors (Pentium Extreme Edition and Pentium D), 7xx for its Pentium M mobile processors, 6xx for advanced Pentium 4 processors, 5xx for mainstream Pentium 4 and mobile Pentium 4 processors, and 3xx for economy Celeron D desktop and Celeron M mobile processors. Dual-core Intel Xeon processors are numbered in the 7xxx series.

Intel is not extending the numbering system to processor models already released. Thus, it will be useful for some time to come to use comprehensive references such as Table 3.47 for Pentium 4 processors because this table incorporates both processors with the numbering system and those that were introduced before the numbering system was developed.

When creating the specific model number for a chip, Intel takes into account not only the raw clock speed of the chip, but also the internal architecture, cache sizes, bus speeds, and other features. In general, the higher the number, the more feature-rich the processor. In addition, within each series, the higher numbers are generally faster chips.

Examples of the model numbers currently assigned to Pentium Extreme Edition, Pentium D, Pentium 4, and Celeron D processors are shown in Table 3.54.

CPUID	L2 Cache	L3 Cache	Max. Temp.	Max. Power	Socket	Process	Transistors
0F44h	2MB	—	64.1°C	95W	775	90nm	230 million
0F47h	2MB	—	64.1°C	95W	775	90nm	230 million
0F44h	2MB	—	69.8°C	130W	775	90nm	230 million
0F47h	2MB	—	69.8°C	130W	775	90nm	230 million
0F47h	2MB	—	69.8°C	130W	775	90nm	230 million
0F44h	2MB	—	69.8°C	130W	775	90nm	230 million
0F44h	2MB	—	69.8°C	130W	775	90nm	230 million

Table 3.54 Intel Desktop Processor Model Numbers and Meanings

Processor	Model No.	Clock Speed	Bus Speed	L2 Cache	Hyper-Threading	EM64T Support	Dual-Core	Other Features
Pentium Extreme Edition	840	3.2GHz	800MHz	2MB	Yes	Yes	Yes	XDB
Pentium D	840	3.2GHz	800MHz	2MB	No	Yes	Yes	XDB, EISS
	830	3.0GHz	800MHz	2MB	No	Yes	Yes	XDB, EISS
	820	2.8GHz	800MHz	2MB	No	Yes	Yes	XDB
Pentium 4	672	3.8GHz	800MHz	2MB	Yes	Yes	No	EISS, IVT, XDB
	670	3.8GHz	800MHz	2MB	Yes	Yes	No	EISS, XDB
	662	3.6GHz	800MHz	2MB	Yes	Yes	No	EISS, IVT, XDB
	660	3.6GHz	800MHz	2MB	Yes	Yes	No	EISS, XDB
	650	3.4GHz	800MHz	2MB	Yes	Yes	No	EISS, XDB
	640	3.2GHz	800MHz	2MB	Yes	Yes	No	EISS, XDB
	630	3GHz	800MHz	2MB	Yes	Yes	No	EISS, XDB
	571	3.8GHz	800MHz	1MB	Yes	Yes	No	XDB
	570J	3.8GHz	800MHz	1MB	Yes	No	No	XDB
	561	3.6GHz	800MHz	1MB	Yes	Yes	No	XDB
	560J	3.6GHz	800MHz	1MB	Yes	No	No	XDB
	560	3.6GHz	800MHz	1MB	Yes	No	No	—
	551	3.4GHz	800MHz	1MB	Yes	Yes	No	XDB
	550J	3.4GHz	800MHz	1MB	Yes	No	No	XDB
	550	3.4GHz	800MHz	1MB	Yes	No	No	—
	541	3.2GHz	800MHz	1MB	Yes	Yes	No	XDB
	540J	3.2GHz	800MHz	1MB	Yes	No	No	XDB
	540	3.2GHz	800MHz	1MB	Yes	No	No	—
	531	3.0GHz	800MHz	1MB	Yes	Yes	No	XDB
	530J	3.0GHz	800MHz	1MB	Yes	No	No	XDB
	530	3.0GHz	800MHz	1MB	Yes	No	No	—
	521	2.8GHz	800MHz	1MB	Yes	Yes	No	XDB
	520J	2.8GHz	800MHz	1MB	Yes	No	No	XDB
520	2.8GHz	800MHz	1MB	Yes	No	No	—	
506	2.66GHz	533MHz	1MB	No	Yes	No	XDB	
505J	2.66GHz	533MHz	1MB	Yes	No	No	XDB	
505	2.66GHz	533MHz	1MB	Yes	No	No	—	
Celeron D	351	3.2GHz	533MHz	256KB	No	Yes	No	XDB
	350	3.2GHz	533MHz	256KB	No	No	No	—
	346	3.06GHz	533MHz	256KB	No	Yes	No	XDB
	345J	3.06GHz	533MHz	256KB	No	No	No	XDB
	345	3.06GHz	533MHz	256KB	No	No	No	—
	341	2.93GHz	533MHz	256KB	No	Yes	No	XDB

Table 3.54 Continued

Processor	Model No.	Clock Speed	Bus Speed	L2 Cache	Hyper-Threading	EM64T Support	Dual-Core	Other Features
	340J	2.93GHz	533MHz	256KB	No	No	No	XDB
	340	2.93GHz	533MHz	256KB	No	No	No	—
	336	2.80GHz	533MHz	256KB	No	Yes	No	XDB
	335J	2.80GHz	533MHz	256KB	No	No	No	XDB
	335	2.80GHz	533MHz	256KB	No	No	No	—
	331	2.66GHz	533MHz	256KB	No	Yes	No	XDB
	330J	2.66GHz	533MHz	256KB	No	No	No	XDB
	330	2.66GHz	533MHz	256KB	No	No	No	—
	326	2.53GHz	533MHz	256KB	No	Yes	No	XDB
	325J	2.53GHz	533MHz	256KB	No	No	No	XDB
	325	2.53GHz	533MHz	256KB	No	No	No	—
	320	2.40GHz	533MHz	256KB	No	No	No	—
	315	2.26GHz	533MHz	256KB	No	No	No	—
	310	2.13GHz	533MHz	256KB	No	No	No	—

EISS = Enhanced Intel SpeedStep Technology for power and thermal management.

XDB = Execute Disable Bit protection against buffer overflow virus attacks.

IVT = Intel Virtualization Technology; it enables the system to run multiple virtual systems for maintenance; testing; or customized environments for gaming, business, and other applications.

Not all 8xx chips are faster than 6xx chips, and not all 5xx chips are faster than 3xx chips. The model numbers are not strictly comparisons of speed and certainly don't pertain to speed comparisons outside the model line. For example, using the BMW automobile analogy from which these numbers seem to be derived, some 3-series cars are faster than some 5-series cars, and some 5-series cars are faster than some 7-series cars. However, as you go up in the series numbers, the higher-numbered series generally have more features or are premium models. Within a particular series, the model numbers do give somewhat of an indication of speed, in that a Pentium 4 660 is faster than a Pentium 4 650, and so on.

It will be interesting to see how these model numbers play out in the marketplace. There are indications that Intel might change its processor numbering system again in 2006. Whatever Intel, or AMD for that matter, decides to do with processor naming, I wouldn't purchase either an Intel or an AMD chip for an upgrade or as part of a new computer without knowing what the true clock speeds are, as well as knowing the cache sizes and other features in the chip. As we have seen, the model numbers don't strictly tell that and are useful only for a rough comparison.

AMD Athlon 64 X2 and Dual-Core Opteron Processors

One of the ironies of the processor business is that AMD, whose 64-bit Athlon 64 and Opteron processors were designed with dual-core updates in mind from the very beginning, was actually the second x86 chip vendor to introduce dual-core chips. AMD's first dual-core Opterons were introduced just after Intel's Pentium Extreme Edition and Pentium D in April 2005, and the desktop Athlon 64 X2 was introduced in May 2005. The Athlon 64 X2 uses two core designs:

- Systems with 1MB of total L2 cache (512KB per core) use the Manchester core.
- Systems with 2MB of total L2 cache (1MB per core) use the Toledo core.

Other major features of the Athlon 64 X2 include

- 90nm manufacturing process
- Actual clock speeds of 2.2GHz–2.4GHz
- Socket 939 form factor
- 1GHz HyperTransport interconnect

The dual-core Opteron processors are available in all three series at speeds ranging from 1.8GHz (x65) to 2.4GHz (x80):

- 100-Series dual-core models for single processor configurations include 165, 170, 175, and 180.
- 200-Series dual-core models for dual-processor configurations include 265, 270, 275, and 280.
- 800-Series dual-core models for up to eight-way processor configurations include 865, 870, 875, and 880.

Although AMD was not the first to introduce dual-core chips, there are several advantages—especially for existing Socket 939 Athlon 64 and all Opteron users—to the AMD approach. The design of these processors has always included room for the second processor core along with a crossbar memory controller to enable the processor cores to directly communicate with each other without using the North Bridge, as with Intel's initial dual-core processors. Figure 3.68 illustrates the internal design of the Athlon 64 X2.

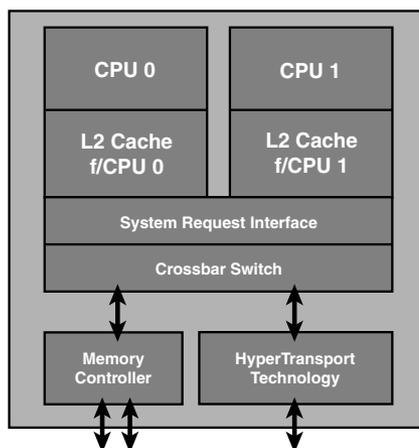


Figure 3.68 The Athlon 64 X2 use the integrated crossbar memory controller present from the beginning of the Athlon 64 processor design to enable the processor cores to communicate with each other.

Table 3.55 Athlon 64 X2 Processor Information

Part Number*	Model Number	CPU Speed	Bus Speed (GBps)	Stepping
ADA3800DAA5BV	3800+	2.0GHz	4.0	E4
ADA3800DAA5CD	3800+	2.0GHz	4.0	E6
ADA4200DAA5BV	4200+	2.2GHz	4.0	E4
ADA4400DAA5CD	4200+	2.2GHz	4.0	E6
ADA4400DAA6CD	4400+	2.2GHz	4.0	E6
ADV4400DAA6CD	4400+	2.2GHz	4.0	E6
ADA4600DAA5BV	4600+	2.4GHz	4.0	E4
ADA4600DAA5CD	4600+	2.4GHz	4.0	E6
ADA4800DAA6CD	4800+	2.4GHz	4.0	E6

*Part numbers for tray processors; PIB (processor-in-box) numbers are different.

The result is that most existing systems based on Socket 939 Athlon 64 and Socket 940 Opterons can be upgraded to a dual-core processor without a motherboard swap. As long as the motherboard supports the 90-nanometer production process versions of these processors and a dual-core BIOS upgrade is available from the motherboard or system vendor, the upgrade is possible.

Another benefit of AMD's approach is the lack of a performance or thermal penalty in moving to a dual-core design. Because the Athlon 64/Opteron design included provisions for a dual-core upgrade from the beginning, the thermal impact of the second core is minimal, even though the dual-core processors run at the same speeds as their predecessors. For example, the hottest Athlon 64 X2 models (running at 2.4GHz or 2.2GHz) dissipate only 110W of heat, compared to 130W for the Pentium Extreme Edition and Pentium D. Most 2.2GHz Athlon 64 X2 models dissipate only 89W, which is the same wattage as the 2.4GHz versions of the Athlon 64 single-core processors.

Although the clock speeds of the Athlon 64 X2 and the Opteron are slower than Intel Pentium D or Pentium Extreme Edition processors, the increased efficiency of AMD's design provides performance that's comparable to or better than Intel's processors, depending on the benchmark. Table 3.55 provides a detailed comparison of the various Athlon 64 X2 processors.

The ability to upgrade most existing Socket 939 Athlon 64 and all Opteron systems with a dual-core processor opens the way for many users to move into dual-core computing with minimal difficulty. As with Intel's dual-core processors, AMD's dual-core processors are best suited to users who multitask or run multithreaded single applications. Gamers are still advised to use the fastest single-core processor, which in AMD's case is the fastest Athlon 64 FX series currently available.

Processor Upgrades

Since the 486, processor upgrades have been relatively easy for most systems. With the 486 and later processors, Intel designed in the capability to upgrade by designing standard sockets that would take a variety of processors. Thus, if you have a motherboard with Socket 3, you can put virtually any 486 processor in it; if you have a Socket 7 motherboard, it should be capable of accepting virtually any Pentium processor (or Socket 7-based third-party processor). This trend has continued to the present, with most motherboards being designed to handle a range of processors in the same family (Pentium III/Celeron III, Athlon/Duron/Athlon XP, Pentium 4/Celeron 4, and so forth).

To maximize your motherboard, you can almost always upgrade to the fastest processor your particular board will support. Because of the varieties of processor sockets and slots—not to mention voltages, speeds, and other potential areas of incompatibility—you should consult with your motherboard manufacturer to see whether a higher-speed processor will work in your board. Usually, that can be determined by the type of socket or slot on the motherboard, but other things such as the voltage regulator and BIOS can be deciding factors as well.

L2 Cache	Max. Temp.	Voltage	Power	Socket	Process
1M	71°C	1.35V–1.4V	89W	939	90nm
1M	71°C	1.35V–1.4V	89W	939	90nm
1M	65°C	1.35V–1.4V	89W	939	90nm
1M	65°C	1.35V–1.4V	89W	939	90nm
2M	65°C	1.35V–1.4V	110W	939	90nm
2M	71°C	1.35V–1.4V	89W	939	90nm
1M	65°C	1.35V–1.4V	110W	939	90nm
1M	65°C	1.35V–1.4V	110W	939	90nm
2M	65°C	1.35V–1.4V	110W	939	90nm

For example, if your motherboard supports Socket 478, you might be able to upgrade to the fastest 3.8GHz version of the Pentium 4. Before purchasing a new CPU, you should verify that the motherboard has proper bus speed, voltage settings, and ROM BIOS support for the new chip. Visit the motherboard or system manufacturer's website to obtain the most up-to-date processor compatibility information and to download BIOS updates that might be necessary.

Caution

If you are trying to upgrade the processor in a low-cost micro-ATX system from a company such as HP, you might have very few processor upgrade options. This is because many of the motherboards in low-cost computers don't provide much in the way of adjustments for clock speed or voltage.

If you are unable to install a faster processor directly into your system, a variety of third-party solutions are available, including adapters that can help first-generation Socket 423 Pentium 4 motherboards use Socket 478 processors, faster Socket 370 processors for older Slot 1 motherboards, and so on. Rather than purchasing processors and adapters separately, I usually recommend you purchase them together in a module from companies such as PowerLeap (see the Vendor List on the disc).

Upgrading the processor can, in some cases, double the performance of a system. However, if you already have the fastest processor that will fit a particular socket, you need to consider other alternatives. In that case, you really should look into a complete motherboard change, which would let you upgrade to a Pentium 4, Athlon XP, or Athlon 64 processor at the same time. If your chassis design is not proprietary and your system uses an industry-standard ATX motherboard design, I normally recommend changing the motherboard and processor rather than trying to find an upgrade processor that will work with your existing board.

OverDrive Processors

Intel at one time offered special OverDrive processors for upgrading systems. Often these were repackaged versions of the standard processors, sometimes including necessary voltage regulators and fans. Unfortunately, they frequently were overpriced, even when compared against purchasing a complete new motherboard and processor. They have all been withdrawn, and Intel has not announced any new versions. I don't recommend the OverDrive processors or third-party upgrades unless the deal is too good to pass up and you need to keep a very old system operating.

Processor Benchmarks

People love to know how fast (or slow) their computers are. We have always been interested in speed; it is human nature. To help us with this quest, various benchmark test programs can be used to measure different aspects of processor and system performance. Although no single numerical measurement can completely describe the performance of a complex device such as a processor or a complete PC, benchmarks can be useful tools for comparing different components and systems.

However, the only truly accurate way to measure your system's performance is to test the system using the actual software applications you use. Although you think you might be testing one component of a system, often other parts of the system can have an effect. It is inaccurate to compare systems with different processors, for example, if they also have different amounts or types of memory, different hard disks, video cards, and so on. All these things and more will skew the test results.

Benchmarks can typically be divided into two types: component or system tests. *Component* benchmarks measure the performance of specific parts of a computer system, such as a processor, hard disk, video card, or CD-ROM drive, whereas system benchmarks typically measure the performance of the entire computer system running a given application or test suite.

Benchmarks are, at most, only one kind of information you can use during the upgrading or purchasing process. You are best served by testing the system using your own set of software operating systems and applications and in the configuration you will be running.

Several companies specialize in benchmark tests and software. The following table lists the companies and the benchmarks they are known for. You can contact these companies via the information in the Vendor List on the disc.

Company	Benchmarks Published	Benchmark Type
Futuremark (formerly MadOnion.com)	SysMark PCMark Pro 3DMark	System System 3D graphics
Business Applications Performance Corporation (BAPCo)	MobileMark	Notebook battery life
Standard Performance Evaluation Corporation	SPECint	Processor Integer
SiSoftware	SPECfp Sandra	Processor Floating-Point System, memory, processor, multimedia

Processor Troubleshooting Techniques

Processors are normally very reliable. Most PC problems are with other devices, but if you suspect the processor, there are some steps you can take to troubleshoot it. The easiest thing to do is to replace the microprocessor with a known-good spare. If the problem goes away, the original processor is defective. If the problem persists, the problem is likely elsewhere.

Table 3.56 provides a general troubleshooting checklist for processor-related PC problems.

Table 3.56 Troubleshooting Processor-Related Problems

Problem Identification	Possible Cause	Resolution
System is dead, no cursor, no beeps, no fan.	Power cord failure.	Plug in or replace power cord. Power cords can fail even though they look fine.
	Power supply failure.	Replace the power supply. Use a known-good spare for testing.
	Motherboard failure.	Replace motherboard. Use a known-good spare for testing.
	Memory failure.	Remove all memory except 1 bank and retest. If the system still won't boot replace bank 1.
System is dead, no beeps, or locks up before POST begins.	All components either not installed or incorrectly installed.	Check all peripherals, especially memory and graphics adapter. Reseat all boards and socketed components.
System beeps on startup, fan is running, no cursor on screen.	Improperly seated or failing graphics adapter.	Reseat or replace graphics adapter. Use known-good spare for testing.
System powers up, fan is running, no beep or cursor.	Processor not properly installed.	Reseat or remove/reinstall processor and heatsink.

Table 3.56 Continued

Problem Identification	Possible Cause	Resolution
Locks up during or shortly after POST.	Poor heat dissipation.	Check CPU heatsink/fan; replace if necessary, use one with higher capacity.
	Improper voltage settings.	Set motherboard for proper core processor voltage.
	Wrong motherboard bus speed.	Set motherboard for proper speed.
	Wrong CPU clock multiplier.	Jumper motherboard for proper clock multiplier.
Improper CPU identification during POST.	Old BIOS.	Update BIOS from manufacturer.
	Board not configured properly.	Check manual and jumper board accordingly to proper bus and multiplier settings.
System won't start after new processor is installed.	Processor not properly installed.	Reseat or remove/reinstall processor and heatsink.
	BIOS doesn't support new processor.	Update BIOS from system or motherboard manufacturer.
	Motherboard can't use new processor.	Verify motherboard support.
Operating system will not boot.	Poor heat dissipation.	Check CPU fan; replace if necessary; it might need a higher-capacity heatsink or heatsink/fan on the North Bridge chip.
	Improper voltage settings.	Jumper motherboard for proper core voltage.
	Wrong motherboard bus speed.	Jumper motherboard for proper speed.
	Wrong CPU clock multiplier.	Jumper motherboard for proper clock multiplier.
	Applications will not install or run.	Improper drivers or incompatible hardware; update drivers and check for compatibility issues.
System appears to work, but no video is displayed.	Monitor turned off or failed.	Check monitor and power to monitor. Replace with known-good spare for testing.

If during the POST the processor is not identified correctly, your motherboard settings might be incorrect or your BIOS might need to be updated. Check that the motherboard is jumpered or configured correctly for your processor, and make sure you have the latest BIOS for your motherboard.

If the system seems to run erratically after it warms up, try setting the processor to a lower speed setting. If the problem goes away, the processor might be defective or overclocked.

Many hardware problems are really software problems in disguise. Be sure you have the latest BIOS for your motherboard, as well as the latest drivers for all your peripherals. Also, it helps to use the latest version of your given operating system because there usually will be fewer problems.