Chapter 3: Microprocessor Types and Specifications

Microprocessors

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 easily the most expensive single component in the system, costing up to four or more times greater 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. This means that all PC-compatible systems use either Intel processors or Intel-compatible processors from a handful of competitors (such as AMD or Cyrix).

Intel's dominance in the processor market had not 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 PCs 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 today such as AMD, Cyrix, IDT, and Rise Technologies have cloned Intel's Pentium processors. In that case though, the clone had become more popular than the original.

Back then I had a system containing both of those processors, consisting of a 1MHz (yes, that's*1*, as in 1MHz!) 6502-based Apple main system with a Microsoft Softcard (Z-80 card) plugged into one of the slots. The Softcard contained a 2MHz Z-80 processor. This allowed me to run software for both types of 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, while the 6502 was best known for its use in the early Apple 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, PC-compatible systems have used a string of Intel or Intel-compatible processors, each new one capable of running the software of the processor before it, from the 8088 to the current Pentium III/Celeron and Athlon/Duron. The following sections cover the different 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.

Pre-PC Microprocessor History

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

The story of the development of the first microprocessor, the Intel 4004, can be read in Chapter 1, "Personal Computer Background." The 4004 processor 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 four 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.

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. Like 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, as 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 of 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 allowed 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 that was 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.

Radio Shack selected the Z-80 for the TRS-80 Model 1, its first PC. The chip was also 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 of 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 computer games, 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 chips. Intel introduced the 8086 in June 1978. The 8086 chip brought with it the original x86 instruction set that is still present on x86-compatible chips such as the Pentium III. 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, meaning it could directly address up to 1MB of memory. Although not directly backward compatible with the 8080, the 8086 instructions and language was very similar and allowed older programs to be ported over quickly 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 an expensive 16-bit support chip and board design. To help bring costs down, in 1979, Intel released 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 allowed support chips from the older 8-bit 8085 to be used, and far less expensive boards and systems could be made. It is for these reasons that IBM chose the crippled chip, the 8088, 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 allowed 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. In order 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.

In some ways the success of the PC, and the Intel architecture it contains, has limited the growth of the personal computer. In other ways, however, its success has caused a huge number of programs, peripherals, and accessories to be developed, and the PC to become a de facto standard in the industry. The original 8088 processor used in the first PC contained close to 30,000 transistors and ran at less than 5MHz. Intel recently introduced a version of the Pentium III Xeon with 2MB of on-die cache that has a whopping 140 million transistors, the largest ever in a single processor chip. Both AMD and Intel are manufacturing processors that run at 1GHz (AMD has some bragging rights there; it beat Intel to 1GHz by two days), and both have demonstrated processors running in the 2GHz range. And the progress doesn’t stop there, as according to Moore’s Law, processing speed and transistor counts are doubling every 1.5 to 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), which means millions of cycles per second—and faster is better! The width of a processor is a little more complicated to discuss because there are three main specifications in a processor that are expressed in width. They are

- Internal registers
- Data input and output bus
- Memory address bus

Systems below 16MHz usually had no cache memory at all. Starting with 16MHz systems, high-speed cache memory appeared on the motherboard because the main memory at the time could not run at 16MHz. Prior to the 486 processor, the cache on the motherboard was the only cache used in the system.

Starting with the 486 series, processors began including what was called L1 (Level 1) cache directly on the processor die. This meant that the L1 cache always ran at the full speed of the chip, especially important when the later 486 chips began to run at speeds higher than the motherboards they were plugged into. During this time the cache on the motherboard was called the second level or L2 cache, which ran at the slower motherboard speed.

Starting with the Pentium Pro and Pentium II, Intel began including L2 cache memory chips directly within the same package as the main processor. Originally this built-in L2 cache was implemented as physically separate chips contained within the processor package but not a part of the processor die. Since the speed of commercially available cache memory chips could not keep pace with the main processor, most of the L2 cache in these processors ran at one-half speed (Pentium II/III and AMD Athlon), while some ran the cache even slower, at two-fifths or even one-third the processor speed (AMD Athlon).

The original Pentium II, III, Celeron, and Athlon (Model 1 and 2) processors use 512KB of either one-half, two-fifths, or one-third speed L2 cache as Table 3.1 shows:

Table 3.1 L2 Cache Speeds

Processor	Speed	L2 Size	L2 Type	L2 Speed
Pentium III	450–600MHz	512KB	External	1/2 core (225–300MHz)
Athlon	550–700MHz	512KB	External	1/2 core (275–350MHz)
Athlon	750–	512KB	External	2/5 core (300–340MHz)

	850MHz			
Athlon	900–1000MHz	512KB	External	1/3 core (300–333MHz)

The Pentium Pro, Pentium II/III Xeon, newer Pentium III, Celeron, K6-3, Athlon (Model 4), and Duron processors include full-core speed L2 as shown in Table 3.2.

Table 3.2 Full-Core Speed Cache

Processor	Speed	L2 Size	L2 type	L2 Speed
Pentium Pro	150–200MHz	256KB–1MB	External	Full core
K6-3	350–450MHz	256KB	On-die	Full core
Duron	550–700+MHz	64KB	On-die	Full core
Celeron	300–600+MHz	128KB	On-die	Full core
Pentium II Xeon	400–450MHz	512KB–2MB	External	Full core
Athlon	650–1000+MHz	256KB	On-die	Full core
Pentium III	500–1000+MHz	256KB	On-die	Full core
Pentium III Xeon	500–1000+MHz	256KB–2MB	On-die	Full core

The problem originally forcing the L2 cache to run at less than the processor core speed was simple: The cache chips available on the market simply couldn't keep up. Intel built its own high-speed cache memory chips for the Xeon processors, but it also made them very expensive. A breakthrough occurred in the second-generation Celeron, where Intel built both the L1 and L2 caches directly on the processor die, where they both ran at the full-core speed of the chip. This type of design was then quickly adopted by the second generation Pentium III, as well as the AMD K6-3, Athlon, and Duron processors. In fact virtually all future processors from Intel and AMD have adopted or will adopt on-die L2 cache as it is the only cost-effective way to include the L2 and bring the speed up.

Table 3.3 lists the primary specifications for the Intel family of processors used in IBM and compatible PCs. Table 3.4 lists the Intel-compatible processors from AMD, Cyrix, NexGen, IDT, and Rise.

Note - Note in Table 3.3 that the Pentium Pro processor includes 256KB, 512KB, or 1MB of full-core speed L2 cache in a separate die within the chip. The Pentium II/III processors include 512KB of ½ core speed L2 cache on the processor card. The Celeron, Pentium II PE, and Pentium IIIE processors include full-core speed L2 cache integrated directly within the processor die. The Celeron III uses the same die as the Pentium IIIE, however half of the on-die cache is disabled, leaving 128KB functional.

The transistor count figures do not include the external (off-die) 256KB, 512KB, 1MB, or 2MB L2 cache built in to the Pentium Pro, Pentium II/III, Xeon, or AMD Athlon CPU packages. The external L2 cache in those processors contains an additional 15.5 (256KB), 31 (512KB), 62 million (1MB), or 124 million (2MB) transistors in separate chips!

Note in Table 3.4 that the Athlon includes either 512KB of L2 cache via separate chips, running at either one-half, two-fifths, or one-third the core speed, or 256KB of on-die L2 running at full-core speed, depending on which version you have.

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 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 of these cycles per second, so speed is measured in megahertz. (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 II 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 are normally 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.

Table 3.3 Intel Processor Specifications

Processor	**CPU Clock**	**Voltage**	**Internal Register Size**	**Data Bus Width**	**Max. Memory**	**Level 1 Cache**	**L1 Cache Type**	**Level 2 Cache**	**L2 Cache Speed**	**Integral FPU**	**M In**
8088	1x	5v	16-bit	8-bit	1MB	—	—	—	—	—	—
8086	1x	5v	16-bit	16-bit	1MB	—	—	—	—	—	—
286	1x	5v	16-bit	16-bit	16MB	—	—	—	—	—	—
386SX	1x	5v	32-bit	16-bit	16MB	—	—	—	Bus	—	—
386SL	1x	3.3v	32-bit	16-bit	16MB	0KB[1]	WT	—	Bus	—	—
386DX	1x	5v	32-bit	32-bit	4GB	—	—	—	Bus	—	—
486SX	1x	5v	32-bit	32-bit	4GB	8KB	WT	—	Bus	—	—
486SX2	2x	5v	32-bit	32-bit	4GB	8KB	WT	—	Bus	—	—
487SX	1x	5v	32-bit	32-bit	4GB	8KB	WT	—	Bus	Yes	—
486DX	1x	5v	32-bit	32-bit	4GB	8KB	WT	—	Bus	Yes	—
486SL[2]	1x	3.3v	32-bit	32-bit	4GB	8KB	WT	—	Bus	Opt.	—
486DX2	2x	5v	32-bit	32-bit	4GB	8KB	WT	—	Bus	Yes	—
486DX4	2–3x	3.3v	32-bit	32-bit	4GB	16KB	WT	—	Bus	Yes	—
486Pentium OD	2.5x	5v	32-bit	32-bit	4GB	2x16KB	WB	—	Bus	Yes	—
Pentium 60/66	1x	5v	32-bit	64-bit	4GB	2x8KB	WB	—	Bus	Yes	—
Pentium 75–200	1.5–3x	3.3–3.5v	32-bit	64-bit	4GB	2x8KB	WB	—	Bus	Yes	—
Pentium MMX	1.5–4.5x	1.8–2.8v	32-bit	64-bit	4GB	2x16KB	WB	—	Bus	Yes	MM
Pentium Pro	2–3x	3.3v	32-bit	64-bit	64GB	2x8KB	WB	256KB 512KB 1MB	Core	Yes	—
Pentium II	3.5–4.5x	1.8–2.8v	32-bit	64-bit	64GB	2x16KB	WB	512KB	? Core	Yes	MM
Pentium II PE	3.5–6x	1.6v	32-bit	64-bit	64GB	2x16KB	WB	256KB	Core[3]	Yes	MM
Celeron	3.5–4.5x	1.8–2.8v	32-bit	64-bit	64GB	2x16KB	WB	0KB	—	Yes	MM
Celeron A	3.5–8x	1.5–2v	32-bit	64-bit	64GB	2x16KB	WB	128KB	Core[3]	Yes	MM
Celeron III	4.5–9x	1.3–1.6v	32-bit	64-bit	64GB	2x16KB	WB	128KB	Core[3]	Yes	SS
Pentium III	4–6x	1.8–2v	32-bit	64-bit	64GB	2x16KB	WB	512KB	? Core	Yes	SS

Pentium IIIE	4–9x	1.3–1.7v	32-bit	64-bit	64GB	2x16KB	WB	256KB	Core[3]	Yes	SS
Pentium II Xeon	4–4.5x	1.8–2.8v	32-bit	64-bit	64GB	2x16KB	WB	512KB 1MB 2MB	Core	Yes	MM
Pentium III Xeon	5–6x	1.8–2.8v	32-bit	64-bit	64GB	2x16KB	WB	512KB 1MB 2MB	Core	Yes	SS
Pentium IIIE Xeon	4.5—6.5x	1.65v	32-bit	64-bit	64GB	2x16KB	WB	256KB 1MB 2MB	Core[3]	Yes	SS

Table 3.4 AMD, Cyrix, NexGen, IDT, and Rise Processors

Processor	CPU Clock	Voltage	Internal Register Size	Data Bus Width	Max. Memory	Level 1 Cache	L1 Cache Type	Level 2 Cache	L2 Cache Speed	Integral FPU	M I
AMD K5	1.5–1.75x	3.5v	32-bit	64-bit	4GB	16+8KB	WB	—	Bus	Yes	—
AMD K6	2.5–4.5x	2.2–3.2v	32-bit	64-bit	4GB	2x32KB	WB	—	Bus	Yes	M
AMD K6-2	2.5–6x	1.9–2.4v	32-bit	64-bit	4GB	2x32KB	WB	—	Bus	Yes	3
AMD K6-3	3.5–4.5x	1.8–2.4v	32-bit	64-bit	4GB	2x32KB	WB	256KB	Core[3]	Yes	3
AMD Athlon	5–10x	1.6–1.8v	32-bit	64-bit	8TB	2x64KB	WB	512KB	1/2–1/3 Core	Yes	E 3
AMD Duron	5–10x	1.5–1.8v	32-bit	64-bit	8TB	2x64KB	WB	64KB	Core[3]	Yes	E 3
AMD Athlon 4 (Thunderbird)	5–10x	1.5–1.8v	32-bit	64-bit	8TB	2x64KB	WB	256KB	Core[3]	Yes	E 3
Cyrix 6x86	2x	2.5–3.5v	32-bit	64-bit	4GB	16KB	WB	—	Bus	Yes	—
Cyrix 6x86MX/MII	2–3.5x	2.2–2.9v	32-bit	64-bit	4GB	64KB	WB	—	Bus	Yes	M
Cyrix III	2.5–	2.2v	32-bit	64-bit	4GB	64KB	WB	256KB	Core[3]	Yes	3

	7x										
NexGen Nx586	2x	4v	32-bit	64-bit	4GB	2x16KB	WB	—	Bus	Yes	—
IDT Winchip	3–4x	3.3–3.5v	32-bit	64-bit	4GB	2x32KB	WB	—	Bus	Yes	M
IDT Winchip2/2A	2.33–4x	3.3–3.5v	32-bit	64-bit	4GB	2x32KB	WB	—	Bus	Yes	3
Rise mP6	2–3.5x	2.8v	32-bit	64-bit	4GB	2x8KB	WB	—	Bus	Yes	M

FPU = Floating-Point Unit (internal math coprocessor)
WT = Write-Through cache (caches reads only)
WB = Write-Back cache (caches both reads and writes)
Bus = Processor external bus speed (motherboard speed)
Core = Processor internal core speed (CPU speed)
MMX = Multimedia extensions, 57 additional instructions for graphics and sound processing
3DNow = MMX plus 21 additional instructions for graphics and sound processing
Enh. 3DNow = 3DNow plus 24 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
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.
3. On-die integrated L2 cache—runs at full-core speed.
4. 128KB functional L2 cache (256KB total, 128KB disabled), uses same die as Pentium IIIE.

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 two 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/Celeron and Athlon/Duron.* These P6 class processors, as well as other sixth-generation processors such as those from AMD and Cyrix, can execute as many as three or more instructions per cycle.

Different instruction execution times (in cycles) make it difficult to compare systems based purely on clock speed or number of cycles per second. 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 why 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. This means that given the same clock speed, a Pentium will be 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 percent 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.

Comparing relative processor performance, you can see that a 1000MHz Pentium III is about equal to a (theoretical) 1,500MHz Pentium, which is about equal to an 3,000MHz 486, which is about equal to a 6,000MHz 386 or 286, which is about equal to a 12,000MHz 8088. The original PC’s 8088 ran at only 4.77MHz; today, we have systems that are comparatively about 2,500 times faster! As you can see, you have to 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 may be relatively faster at certain things and slower at others. To fairly compare different 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 3.5 shows the relative power, or iCOMP 2.0 index, for several processors.

Table 3.5 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	197	Pentium II	387

180		Overdrive 333	
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.

Recently Intel 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 compute-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 in to 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 3.6 shows the iCOMP Index 3.0 ratings for newer Intel processors.

Table 3.6 Intel iComp 3.0 Ratings

Processor	iCOMP3.0 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		

Considerations When Interpreting iCOMP Scores

Each processor's rating is calculated at the time the processor is introduced, using a particular, well-configured, commercially available system. Relative iCOMP Index 3.0 scores and actual system performance might be affected by future changes in software design and configuration. Relative scores and actual system performance also may be affected by differences in components or characteristics of microprocessors such as L2 cache, bus speed, extended multimedia or graphics instructions, or improvements in the microprocessor manufacturing process.

Differences in hardware components other than microprocessors used in the test systems also can affect how iCOMP scores relate to actual system performance. iCOMP 3.0 ratings cannot be compared with earlier versions of the iCOMP index because different benchmarks and weightings are used in calculating the result.

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 Celeron 600 runs at a multiple of nine times the motherboard speed of 66MHz, while a Pentium III 1GHz runs at 7½ times the motherboard speed of 133MHz. Up until early 1998, most motherboards ran at 66MHz or less because that is all Intel supported with its processors until then. Starting in April 1998, Intel released both processors and motherboard chipsets designed to run at 100MHz. Cyrix has a few processors designed to run on 75MHz motherboards, and many Pentium motherboards are capable of running that speed as well, although technically Intel never supported it. AMD also has versions of the K6-2 designed to run at motherboard speeds of 100MHz.

Starting in 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 at 100MHz but using a double transfer technique for an effective 200MHz data rate between the Athlon processor and the main chipset North Bridge chip.

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

Normally, you can set the motherboard speed and multiplier setting via jumpers or other 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 will 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 processors and motherboards run at the speeds shown in Table 3.7.

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

Table 3.7 Intel Processor and Motherboard Speeds

CPU Type	**CPU Speed (MHz)**	**CPU Clock Multiplier**	**Motherboard Speed (MHz)**
Pentium	60	1x	60
Pentium	66	1x	66
Pentium	75	1.5x	50
Pentium	90	1.5x	60
Pentium	100	1.5x	66
Pentium	120	2x	60
Pentium	133	2x	66
Pentium	150	2.5x	60
Pentium/Pentium Pro/MMX	166	2.5x	66
Pentium/Pentium Pro	180	3x	60
Pentium/Pentium Pro/MMX	200	3x	66
Pentium-MMX/Pentium II	233	3.5x	66
Pentium-MMX (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
Pentium II	350	3.5x	100
Pentium II/Xeon	400	4x	100
Pentium II/III/Xeon	450	4.5x	100
Pentium III/Xeon	500	5x	100

Pentium III/Xeon	550	5.5x	100
Pentium III/Xeon	600	6x	100
Pentium III/Xeon	650	6.5x	100
Pentium III/Xeon	700	7x	100
Pentium III/Xeon	750	7.5x	100
Pentium III/Xeon	800	8x	100
Pentium III/Xeon	850	8.5x	100
Pentium III/Xeon	533	4x	133
Pentium III/Xeon	600	4.5x	133
Pentium III/Xeon	667	5x	133
Pentium III/Xeon	733	5.5x	133
Pentium III/Xeon	800	6x	133
Pentium III/Xeon	866	6.5x	133
Pentium III/Xeon	933	7x	133
Pentium III/Xeon	1000	7.5x	133
Pentium III/Xeon	1066	8x	133
Pentium III/Xeon	1133	8.5x	133
Pentium III/Xeon	1200	9x	133
Pentium III/Xeon	1266	9.5x	133
Pentium III/Xeon	1333	10x	133

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 different 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.

It is possible in some systems to set the processor speed 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 since Intel, AMD, and others often build safety margins into their ratings. This means that a chip rated for, say, 800MHz may in fact run at 900MHz or more, but is instead 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 with playing with your system, and you can afford and are capable of dealing with any potential consequences, overclocking may allow you to get more performance from your system.

If you are intent on overclocking, there are several issues to consider. One is that most Intel processors since the Pentium II have been multiplier-locked before they are shipped out. This means that any changes to the multiplier setting on the motherboard will simply be ignored by the chip. Both Intel and AMD lock the multipliers on most of their newer processors. Although originally done to prevent remarkers from fraudulently relabeling processors, this has impacted the computing performance enthusiast, leaving tweaking the motherboard bus speed as the only way to achieve a clock speed higher than standard.

You can run into problems increasing motherboard bus speed as well. Intel motherboards, for example, simply don’t support clock speeds other than the standard 66MHz, 100MHz, or 133MHz settings. Also all of their boards with speed settings done via software (BIOS Setup) will read the proper settings from the installed processor and only allow those settings. In other words, you simply plug in the processor, and the Intel motherboard won’t allow any other settings other than what that processor is designed for.

Even if you could fool the processor into accepting a different setting, the jump from 66MHz to 100MHz, or from 100 to 133MHz, is a large one, and many processors would not make that much of a jump reliably. For example, a Pentium III 800E runs at a 100MHz bus speed with an 8x multiplier. Bumping the motherboard speed to 133MHz would cause the processor to try to run at 8x133 or 1066MHz. It is highly unlikely that the chip would run reliably at that speed. Likewise, a Celeron 600E runs at 9x66MHz. Raising the bus speed to 100MHz would cause the chip to try and run at 9x100MHz or 900MHz, likely an unsuccessful change.

What is needed is a board that supports intermediate speed settings and that allows the settings to be changed in smaller increments. For example, the Asus P3V4X motherboard supports front-side bus speed settings of 66, 75, 83, 90, 95, 100, 103, 105, 110, 112, 115, 120, 124, 133, 140, and 150MHz. By setting the 800MHz Pentium IIIE to increments above 100MHz, you’d have

Multiplier (fixed)	**Bus Speed**	**Processor Speed**
8x	100MHz	800MHz
8x	103MHz	824MHz
8x	105MHz	840MHz
8x	110MHz	880MHz
8x	112MHz	896MHz
8x	115MHz	920MHz
8x	120MHz	960MHz
8x	124MHz	992MHz
8x	133MHz	1066MHz

Likewise, using this motherboard with a Celeron 600, you could try settings above the standard 66MHz bus speed as follows:

Multiplier (fixed)	Bus Speed	Processor Speed
9x	66MHz	600MHz
9x	75MHz	675MHz
9x	83MHz	747MHz
9x	90MHz	810MHz
9x	95MHz	855MHz
9x	100MHz	900MHz

Normally a 10–20 percent increase will be successful, so with this motherboard, you are likely to get your processor running 100MHz or more faster than it was originally designed for.

Another trick used by overclockers is to play with the voltage settings for the CPU. All Slot 1, Slot A, Socket 8, Socket 370, and Socket A processors have automatic voltage detection, where the system will detect and set the correct voltage by reading certain pins on the processor. Some motherboards, such as those made by Intel, do not allow any changes to these settings manually. Other motherboards, such as the Asus P3V4X I mentioned earlier, allow you to tweak the voltage settings from the automatic setting up or down by tenths 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 running stable.

My recommendation is to be careful when playing with voltages. It is possible to 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. Especially when overclocking, it is essential that the system components and especially the CPU remain properly cooled. Going a little bit overkill on the processor heat sink and adding extra cooling fans to the case will never hurt and in many cases help a great deal when hotrodding a system in this manner.

Note - One good source of online overclocking information is located at http://www.tomshardware.com/. It includes, among other things, fairly thorough overclocking FAQs and an ongoing survey of users who have successfully (and sometimes unsuccessfully) overclocked their CPUs. Note that many of the newer Intel processors incorporate fixed bus multiplier ratios, which effectively prevent or certainly reduce the ability to overclock. Unfortunately this can be overridden with a simple hardware fix, and many counterfeit processor vendors are selling remarked (overclocked) chips.

The Processor Heat Sink Might Hide the Rating

Most processors have heat sinks on top of them, which can prevent you from reading the rating printed on the chip.

A *heat sink* is a metal device that draws heat away from an electronic device. Most processors running at 50MHz and faster should have a heat sink installed to prevent the processor from overheating.

Fortunately, most CPU manufacturers are placing marks on the top and bottom of the processor. If the heat sink is difficult to remove from the chip, you can take the heat sink and chip out of the socket together and read the markings on the bottom of the processor to determine what you have.

Cyrix P-Ratings

Cyrix/IBM 6x86 processors use a PR (Performance Rating) scale that is not equal to the true clock speed in megahertz. For example, the Cyrix 6x86MX/MII-PR366 actually runs at only 250MHz (2.5 x 100MHz). This is a little misleading—you must set up the motherboard as if a 250MHz processor were being installed, not the 366MHz you might suspect. Unfortunately this leads people to believe these systems are faster than they really are. Table 3.8 shows the relationship between the Cyrix 6x86, 6x86MX, and M-II P-Ratings versus the actual chip speeds in MHz.

Table 3.8 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 they are comparing to 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 does not compare well against the Celeron, Pentium II, or Pentium III processors. In that case these chips are more comparative at their true speed. In other words, the MII-PR366 really runs at only 250MHz and compares well against Intel processors running at closer to that speed. I consider calling a chip an MII-366 when it really runs at only 250MHz very misleading, to say the least.

AMD P-Ratings

Although both AMD and Cyrix concocted this misleading P-Rating system, AMD thankfully only used it for a short time and only on the older K5 processor. It still has the PR designation stamped on its newer chips, but all K6 and Athlon processors have PR numbers that match their actual CPU speed in MHz. Table 3.9 shows the P-Rating and actual speeds of the AMD K5, K6, and Athlon processors.

Table 3.9 AMD P-Ratings Versus Actual Chip Speeds in MHz

CPU Type	P-Rating	Actual CPU Speed (MHz)	Clock Multiplier	Motherboard Speed (MHz)
K5	PR75	75	1.5x	50
K5	PR90	90	1.5x	60
K5	PR100	100	1.5x	66
K5	PR120	90	1.5x	60
K5	PR133	100	1.5x	66
K5	PR166	116.7	1.75x	66
K6	PR166	166	2.5x	66
K6	PR200	200	3x	66
K6	PR233	233	3.5x	66
K6	PR266	266	4x	66
K6	PR300	300	4.5x	66
K6-2	PR233	233	3.5x	66
K6-2	PR266	266	4x	66
K6-2	PR300	300	4.5x	66
K6-2	PR300	300	3x	100
K6-2	PR333	333	5x	66
K6-2	PR333	333	3.5x	95
K6-2	PR350	350	3.5x	100
K6-2	PR366	366	5.5x	66
K6-2	PR380	380	4x	95
K6-2	PR400	400	6x	66
K6-2	PR400	400	4x	100
K6-2	PR450	450	4.5x	100
K6-2	PR475	475	5x	95
K6-2	PR500	500	5x	100
K6-2	PR533	533	5.5x	97
K6-2	PR550	550	5.5x	100
K6-3	PR400	400	4x	100
K6-3	PR450	450	4.5x	100
Athlon	PR500	500	5x	100[1]
Athlon	PR550	550	5.5x	100[1]
Athlon	PR600	600	6x	100[1]
Athlon	PR650	650	6.5x	100[1]

Athlon	PR700	700	7x	100[1]
Athlon	PR750	750	7.5x	100[1]
Athlon	PR800	800	8x	100[1]
Athlon	PR850	850	8.5x	100[1]
Athlon	PR900	900	9x	100[1]
Athlon	PR950	950	9.5x	100[1]
Athlon	PR1000	1000	10x	100[1]

1. Note the Athlon to North Bridge processor bus actually runs at a double transfer speed which is twice that of the motherboard clock speed (200MHz).

Data Bus

Perhaps the most common way to describe a processor is by the speed at which it runs and the width of the processor's external data bus. This defines the number of data bits that can be moved into or out of the processor in one cycle. A *bus* is a series of connections that carry common signals. Imagine running a pair of wires from one end of a building to another. If you connect a 110v AC power generator to the two wires at any point and place outlets at convenient locations along the wires, you have constructed a power bus. No matter which outlet you plug the wires into, you have access to the same signal, which in this example is 110v AC power. Any transmission medium that has more than one outlet at each end can be called a bus. A typical computer system has several internal and external buses.

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 allows for greater throughput.

Data in a computer is sent as digital information consisting of a time interval in which a single wire carries 5v to signal a 1 data bit, or 0v to signal a 0 data bit. The more wires you have, the more individual bits you can send in the same time interval. A chip such as the 286 or 386SX, which has 16 wires for transmitting and receiving such data, has a 16-bit data bus. A 32-bit chip, such as the 386DX and 486, has twice as many wires dedicated to simultaneous data transmission as a 16-bit chip; a 32-bit chip can send twice as much information in the same time interval as a 16-bit chip. Modern processors such as the Pentium series have 64-bit external data buses. This means that Pentium processors including the original Pentium, Pentium Pro, and Pentium II can all transfer 64 bits of data at a time to and from the 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 one byte flows through at a time. (One byte equals eight individual bits.) The 16-bit chip, with two bytes flowing at a time, resembles a two-lane highway. You may 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 four 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!

Just as you can describe a highway by its lane width, you can describe a chip by the width of its data bus. When you read an advertisement that describes a 32-bit or 64-bit computer system, the ad usually refers to the CPU's data bus. This number provides a rough idea of the chip's performance potential (and, therefore, the system).

Perhaps the most important 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. This means that a 32-bit processor, such as the 486 class chips, reads and writes memory 32 bits at a time. Pentium class processors, including the Pentium III and Celeron, read and write memory 64 bits at a time. Because standard 72-pin SIMMs (Single Inline Memory Modules) are only 32 bits wide, they must be installed one at a time in most 486 class systems; they're installed two at a time in most Pentium class systems. Newer DIMMs (Dual Inline Memory Modules) are 64 bits wide, so they are installed one at a time in Pentium class systems. Each DIMM is equal to a complete bank of memory in Pentium systems, which makes system configuration easy, because they can then be installed or removed one at a time.

Internal Registers (Internal Data Bus)

The size of the internal registers indicate 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. The register size is essentially the same as the internal data bus size. 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 the processor can operate on. 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 cannot. Most advanced processors today—chips from the 386 to the Pentium III—use 32-bit internal registers and can therefore run the same 32-bit operating systems and software.

Some 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 enables 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.

Internal registers often are larger than the data bus, which means that 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 has to "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.

More advanced sixth-generation processors such as the Pentium Pro and Pentium II/III have as many as six internal pipelines for executing instructions. Although some of these internal pipes are dedicated to special functions, these processors can still execute as many as three instructions in one clock cycle.

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 that a chip can address.

The highway analogy 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 3.10 describes the memory-addressing capabilities of processors.

Table 3.10 Processor Memory-Addressing Capabilities

Processor Family	Address Bus	Bytes	KB	MB	GB
8088/8086	20-bit	1,048,576	1,024	1	—
286/386SX	24-bit	16,777,216	16,384	16	—
386DX/486/P5 Class	32-bit	4,294,967,296	4,194,304	4,096	4
P6 Class	36-bit	68,719,476,736	67,108,864	65,536	64

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 Level 1 (L1) 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.

Since L1 cache is always built in to 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 in to 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 often is much slower than the CPU. This 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 frequently would be forced to wait until system memory caught up.

L1 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 they are plugged into. The Pentium III 1GHz, for example, runs at a multiple of 7½ times the true motherboard speed of 133MHz. Because the main memory is plugged in to the motherboard, it can also run at only 133MHz maximum. The only 1GHz memory in such a system is the L1 and L2 caches built into the processor core. In this example, the Pentium III 1GHz processor has 32KB of integrated L1 cache in two separate 16KB blocks and 256KB of L2, all running at the full speed of the processor core.

If the data that 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.

In order 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 is normally expressed in MHz (millions of cycles per second), while memory speeds are often expressed in nanoseconds (billionths of a second per cycle).

Both are really time- or frequency-based measurements, and a chart comparing them can be found in Chapter 6, "Memory," Table 6.3. 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 will normally run at 66MHz, which corresponds to a speed of 15ns per cycle, and require 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 in a typical Pentium 233 system, you have 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).

How Cache Works

To learn how the L1 and L2 cache work, 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 memory (SIMM/DIMM) RAM. The cache controller is the waiter, and the L1 cache is the table you are seated at. L2 cache will be introduced as a food cart, which is positioned between your table and the kitchen.

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. Once the waiter brings the food, you start eating at your normal rate. Pretty quickly you finish the hot dog, so you call the waiter and order a hamburger. Again you wait 60 seconds while the hamburger is being produced. When it arrives again you 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 again eat dessert 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 were 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 five 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 what 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 what 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, where 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 there is a cache miss. According to Intel, the L1 cache in most of its processors has approximately a 90 percent hit ratio. This means that the cache has the correct data 90 percent of the time and consequently the processor runs at full speed, 233MHz in this example, 90 percent of the time. However, 10 percent 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.

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 percent or greater, makes L1 cache very important for system performance.

Level 2 (L2) Cache

To mitigate the dramatic slowdown every time there is a L1 cache miss, a secondary or L2 cache can be 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 such that the waiter can retrieve food from it in 15 seconds. 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 instead. The effect is that the system slows down from 233MHz to 66MHz.

Just as with the L1 cache, most L2 caches have a hit ratio also in the 90 percent range, which means that if you look at the system as a whole, 90 percent 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 percent of that time you will have to go to the slow main memory to get the data due to an L2 cache miss. This means that by combining both caches, our sample system runs at full processor speed 90 percent of the time (233MHz in this case), motherboard speed nine percent (90 percent of 10 percent) of the time (66MHz in this case), and RAM speed about one percent (10 percent of 10 percent) of the time (16MHz in this case). You can clearly see the importance of both the L1 and L2 caches; without them the system will be using 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 only used directly about one percent of the time, if you doubled performance there, you would double the speed of your system only one percent 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 nine percent 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 devised methods of improving the performance of L2 cache. In Pentium (P5) class systems, the L2 cache was normally 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 like Sony, Toshiba, NEC, Samsung, and others. Since 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 since 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 there was a breakthrough, which first appeared in the Celeron processor 300A and above. These had 128KB of L2 cache, but there were no external chips used. Instead the L2 cache had been integrated directly into the processor core just like the L1. This meant that now both the L1 and L2 caches 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 the nine percent 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 there are now fewer parts involved.

Let’s revisit the restaurant analogy using a modern Pentium III 1GHz. You would now be taking a bite every one second (1GHz = 1ns 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 runs to the cart, and returns nine out of 10 times with the food you want in only one second (L2 speed = 1GHz or 1ns cycling). In this more modern system, you would run at 1GHz 99 percent of the time (L1 and L2 hit ratios combined), and only slow down to RAM speed (wait for the kitchen) 1 percent of the time as before. With faster memory running at 133MHz (7.5ns), you would only have to wait 7.5 seconds for the food to come from the kitchen. If only restaurant performance increased at the same rate processor performance has!

Cache Organization

The organization of the cache memory in the 486 and 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.

To understand how a four-way set associative cache works, consider a simple example. In the simplest cache design, the cache is set up as a single block into which you can load the contents of a corresponding block of main memory. This procedure is similar to using a bookmark to locate the current page of a book that you are reading. If main memory equates to all the pages in the book, the bookmark indicates which pages are held in cache memory. This procedure works if the required data is located within the pages marked with the bookmark, but it does not work if you need to refer to a previously read page. In that case, the bookmark is of no use.

An alternative approach is to maintain multiple bookmarks to mark several parts of the book simultaneously. Additional hardware overhead is associated with having multiple bookmarks, and you also have to take time to check all the bookmarks to see which one marks the pages of data you need. Each additional bookmark adds to the overhead, but also increases your chance of finding the desired pages.

If you settle on marking four areas in the book, you have essentially constructed a four-way set associative cache. This technique splits the available cache memory into four blocks, each of which stores different lines of main memory. 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 a four-way cache would improve performance greatly.

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 four 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 non-blocking. This is a technique for reducing or hiding memory delays by exploiting the overlap of processor operations with data accesses. A non-blocking cache allows 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 allow the processor to continue doing something non-dependent on the missing data.

The cache controller built into the processor also is responsible for watching the memory bus when alternative processors, known as busmasters, are in control of the system. This process of watching the bus is referred to as *bus snooping*. If a busmaster 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.

A secondary external L2 cache of extremely fast static RAM (SRAM) chips also is used in most 486 and Pentium-based systems. It further reduces the amount of time that the CPU must spend waiting for data from system memory. The function of the secondary processor cache is similar to that of the onboard cache. The secondary processor cache holds information that is moving to the CPU, thereby reducing the time that the CPU spends waiting and increasing the time that the CPU spends performing calculations. Fetching information from the secondary processor cache rather than from system memory is much faster because of the SRAM chips' extremely fast speed—15 nanoseconds (ns) or less.

Pentium systems incorporate the secondary cache on the motherboard, while Pentium Pro and Pentium II systems have the secondary cache inside the processor package. By moving the L2 cache into the processor, systems are capable of running at speeds higher than the motherboard—up to as fast as the processor core.

As clock speeds increase, cycle time decreases. Most SIMM memory used in Pentium and earlier systems was 60ns, which works out to be only about 16MHz! Standard motherboard speeds are now 66MHz, 100MHz, or 133MHz, and processors are available at 600MHz or more. Newer systems don't use cache on the motherboard any longer, as the faster SDRAM or RDRAM used in modern Pentium Celeron/II/III systems can keep up with the motherboard speed. The trend today is toward integrating the L2 cache into the processor die just like the L1 cache. This allows 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 3.11 illustrates the need for and function of L1 (internal) and L2 (external) caches in modern systems.

Table 3.11 CPU Speeds Relative to Cache, SIMM/DIMM, and Motherboard

CPU Type:	**Pentium**	**Pentium Pro**	**Pentium II 333**	**K6-2 500**
CPU speed	233MHz	200MHz	333MHz	500MHz
L1 cache speed	4ns (233MHz)	5ns (200MHz)	3ns (333MHz)	2ns (500MHz)
L2 cache speed	15ns (66MHz)	5ns (200MHz)	6ns (167MHz)	10ns (100MHz)
Motherboard speed	66MHz	66MHz	66MHz	100MHz
SIMM/DIMM speed	60ns (16MHz)	60ns (16MHz)	15ns (66MHz)	10ns (100MHz)
SIMM/DIMM type	FPM/EDO	FPM/EDO	SDRAM	SDRAM
CPU Type:	**Celeron 500**	**Pentium III 500**	**Athlon 1000**	**Pentium III 1000**
CPU speed	500MHz	500MHz	1000MHz	1000MHz
L1 cache speed	2ns (500MHz)	2ns (500MHz)	1ns (1000MHz)	1ns (1000MHz)
L2 cache speed	2ns (500MHz)	4ns (250MHz)	3ns (333MHz)	1ns (1000MHz)
Motherboard speed	66MHz	100MHz	200MHz	133MHz

SIMM/DIMM speed	15ns (66MHz)	10ns (100MHz)	10ns (100MHz)	5ns (200MHz)[1]
SIMM/DIMM type	SDRAM	SDRAM	SDRAM	RDRAM

1. Note RDRAM technically runs at 800MHz, but the channel is only 16 bits wide, resulting in a bandwidth of 1.6GB/sec, which is equivalent to running 200MHz at the 64-bit width of the processor data bus.

The Celeron processors at 300MHz and faster as well as the Pentium III processors at 600MHz and faster have on-die L2 cache which runs at the full-core speed of the processor. Newer Athlon processors and all Duron processors have full-core speed on-die cache as well. The older Pentium II and III processors, as well as the older Athlons, use external L2 and run the cache at either one-half, two-fifths, or one-third of the core processor speed. 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 allows the processor to keep working closer to its true speed.

Processor Modes

All Intel 32-bit and later 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.

Three different modes of operation possible are

- Real mode (16-bit software)
- Protected mode (32-bit software)
- Virtual Real mode (16-bit programs within a 32-bit environment)

Real Mode

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.

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 normally single-tasking, which means that only one program can run at a time. There is no built-in protection to keep one program from overwriting another program or even the operating system in memory, which means that if more than one program is running, it is possible for one of them to bring the entire system to a crashing halt.

Protected (32-bit) Mode

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, you needed a 32-bit operating system and a 32-bit application. 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, as an errant program cannot 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 in a backward-compatible real mode into the 386. That allowed it to run unmodified 16-bit operating systems and applications. It ran them quite well—much faster 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 faster. 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 III is running DOS (real mode), it acts like a "Turbo 8088." Turbo 8088 means that 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. This means if you have a 128MB Pentium III system running Windows 3.x or DOS, you are effectively using only the first megabyte of memory, leaving the other 127MB 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, 32-bit operating systems such as UNIX or variants (such as Linux), OS/2, and even Windows NT and Windows 2000 have had a very hard time getting any mainstream share in the PC marketplace. Out of those, Windows 2000 is the only one that will likely become a true mainstream product, and that is mainly because Microsoft has coerced us in that direction with Windows 95 through 98 and Me. Windows 3.x was the last full 16-bit operating system. In fact, it was not a complete operating system because it ran on top of DOS.

Microsoft realized how stubborn the installed base of PC users was so it developed Windows 95 through the current Windows Me as a bridge to a full 32-bit world. Windows 95, 98, Me are mostly 32-bit operating systems but retain enough 16-bit capability to fully run old 16-bit applications. Windows 95 came out in August 1995, a full 10 years later than the introduction of the first 32-bit PC processor! It has taken us only 10 years to migrate to software that can fully use the processors we have in front of us.

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 allows true multitasking, you can actually have several real mode sessions running, each with its own software running on a virtual PC. This 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 will automatically switch the processor into 32-bit mode and take control from there.

Some 16-bit (DOS and Windows 3.x) applications misbehave, which means they do things that even virtual real mode will not support. Diagnostics software is a perfect example of this. Such software will not run properly in a real mode (virtual real) window under Windows. In that case, you can still run your Pentium III in the original no-frills real mode by either booting to a DOS floppy or in the case of Windows 9x, 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 POST (power on self test) is completed. In the latter case it helps to hit the F8 key multiple times as it is hard to get the timing just right and Windows 9x will only look for the key 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, which 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.

Note that even though Windows Me (Millennium Edition) is based on Windows 98, Microsoft removed the Startup Menu option in an attempt to further wean us from any 16-bit operation. Windows NT and 2000 also lack the ability to interrupt the startup in this manner. For these operating systems, you’ll need a Startup Disk (floppy), which you can create and then use to boot the system in real mode. Normally you would do this to perform certain maintenance procedures, especially 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 were available that "extended" DOS and allow access to extended memory (over 1MB). These are sometimes called DOS extenders and are usually included as a 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 DPMI (DOS protected mode interface). DPMI was used by Windows 3.x to access extended memory for use with Windows 3.x applications. It allowed them to use more memory even though they were still 16-bit programs. DOS extenders are especially popular in DOS games, because they allow 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, or in the case of those that run under Windows, they use the DPMI interface built in to Windows, allowing 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).

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 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 allows 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 normally 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.

Superscalar architecture usually is associated with high-output RISC (Reduced Instruction Set Computer) chips. An 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 CISC (Complex Instruction Set Computer) 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 (see Figure 3.2) 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.

Figure 3.2
An Intel Pentium MMX chip shown from the top and bottom (exposing the die).
Photograph used by permission of Intel Corporation.

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 instructions 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 percent or less of the overall application code, can account for up to 90 percent 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 allows 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 often found in multimedia operations. Highly parallel refers to the fact that the same processing is done on many different data points, such as when modifying a graphic image. The main drawbacks to MMX were that it only worked on integer values and used the floating-point unit for processing, meaning that 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 (Streaming SIMD Extensions)

In February 1999, Intel introduced the Pentium III processor, and included in that processor was an update to MMX called Streaming SIMD Extensions (SSE). These were also called Katmai New Instructions (KNI) up until their debut, as 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 (Katmai New Instructions), SSE was also called MMX-2 by some before it was released. Besides 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.

The Streaming SIMD Extensions consist of 70 new instructions, including Single Instruction Multiple Data (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 SSE instructions are particularly useful with MPEG2 decoding, which is the standard scheme used on DVD video discs. This means that SSE equipped processors should be more capable of doing 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 (Single Instruction Multiple Data) 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, which means that 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 included support for SSE in its DirectX 6.1 and later video and sound drivers, which were included with Windows 98 Second Edition, Windows Me, Windows NT 4.0 (with service pack 5 or later), and Windows 2000.

Note that SSE is an extension to MMX, meaning that processors supporting SSE also support the original MMX instructions. This means that standard MMX-enabled applications run as they did on MMX-only processors.

3DNow and Enhanced 3DNow

3DNow technology is AMD’s alternative to the SSE instructions in the Intel processors. Actually 3DNow originally came out in the K6 series before Intel released SSE in the Pentium III, and then AMD added Enhanced 3DNow to the Athlon and Duron 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 later 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 allows greater performance for 3D graphics, multimedia, and other floating-point-intensive PC applications.

3DNow technology is a set of 21 instructions that uses SIMD (Single Instruction Multiple Data) techniques to operate on arrays of data rather than single elements. Enhanced 3DNow adds 24 more 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 (streaming SIMD extensions) 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 that software specifically written to support SSE will not support 3DNow, and vice versa.

Just as with SSE, 3DNow also supports single precision floating point SIMD (Single Instruction Multiple Data) 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 Microsoft Windows 9x, Windows NT 4.0, and all newer Microsoft operating systems. Application programming interfaces such as Microsoft's DirectX 6.x API and SGI's Open GL API have been optimized for 3DNow technology, as have the drivers for many leading 3D graphic accelerator suppliers, including 3Dfx, ATI, Matrox, and nVidia. While many games and video drivers support 3DNow, support is lacking from some of the major business graphics applications like Adobe Photoshop.

Dynamic Execution

First used in the P6 or sixth-generation processors, dynamic execution is an innovative combination of three processing techniques designed to help the processor manipulate data more efficiently. Those techniques are multiple branch prediction, data flow analysis, and speculative execution. Dynamic execution enables the processor to be more efficient by manipulating data in a more logically ordered fashion rather than simply processing a list of instructions, and it is one of the hallmarks of all sixth-generation processors.

The way software is written can dramatically influence a processor's performance. For example, performance will be adversely affected if the processor is frequently required to stop what it is doing and jump or branch to a point elsewhere in the program. Delays also occur when the processor cannot process a new instruction until the current instruction is completed. Dynamic execution allows the processor to not only dynamically predict the order of instructions, but execute them out of order internally, if necessary, for an improvement in speed.

Multiple Branch Prediction

Multiple branch prediction predicts the flow of the program through several branches. Using a special algorithm, the processor can anticipate jumps or branches in the instruction flow. It uses this to predict where the next instructions can be found in memory with an accuracy of 90 percent or greater. This is possible because while the processor is fetching instructions, it is also looking at instructions further ahead in the program.

Data Flow Analysis

Data flow analysis analyzes and schedules instructions to be executed in an optimal sequence, independent of the original program order. The processor looks at decoded software instructions and determines whether they are available for processing or are instead dependent on other instructions to be executed first. The processor then determines the optimal sequence for processing and executes the instructions in the most efficient manner.

Speculative Execution

Speculative execution increases performance by looking ahead of the program counter and executing instructions that are likely to be needed later. Because the software instructions being processed are based on predicted branches, the results are stored in a pool for later referral. If they are to be executed by the resultant program flow, the already completed instructions are retired and the results are committed to the processor’s main registers in the original program execution order. This technique essentially allows the processor to complete instructions in advance and then grab the already completed results when necessary.

Dual Independent Bus (DIB) 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 second or backside bus in a processor with DIB is used for the L2 cache, allowing it to run at much greater speeds than if it were to share the main processor bus.

> **Note -** The DIB architecture is explained more fully in Chapter 4, "Motherboards and Buses." To see the typical Pentium II/III system architecture, see Figure 4.34.

Two buses make up the DIB architecture: the L2 cache bus and the processor-to-main-memory, or system, bus. The P6 class processors from the Pentium Pro to the Celeron, Pentium II/III, and Athlon/Duron processors can use both buses simultaneously, eliminating a bottleneck there. The Dual Independent Bus architecture enables the L2 cache of the 1GHz Pentium III or Athlon, for example, to run 15 times faster than the L2 cache of older Pentium and K6 processors. Because the backside or L2 cache bus is coupled to the speed of the processor core, as the frequency of processors increases, so will the speed of the L2 cache.

The key to implementing DIB was to move the L2 cache memory off of the motherboard and into the processor package. L1 cache has always been directly a part of the processor die, but L2 was larger and 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. To move the L2 cache into the processor initially, modifications had to be made to the CPU socket or slot. There are two slot-based and three socket-based solutions that fully support DIB: Slot 1 (Pentium II/III/Celeron), Slot A (Athlon), Socket 8 (Pentium Pro), Socket 370 (Pentium III/Celeron), and Socket A (Athlon/Duron).

DIB also allows 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.

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 which 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 percent pure. These rods are then mechanically broken up into chunks and packed into quartz crucibles, which are loaded into the electric crystal pulling ovens. There the silicon chunks are melted at over 2,500° Fahrenheit. To prevent impurities, the ovens are normally mounted on very thick concrete cubes often on a suspension to prevent any vibration which would damage the crystal as it forms.

Once 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 to 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 approximately eight or 12 inches in diameter and over five feet long, weighing hundreds of pounds.

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

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

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.

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

Doping is the term used to describe chemical impurities added to silicon (which is naturally a non-conductor), 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. The Pentium III currently uses twenty or more masks to create six layers of metal and semiconductor interconnects.

As the light passes through a mask, the light is focused on the wafer surface, imprinting it 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 chips, a caustic solution washes away the areas where the light struck the photoresist, leaving the mask imprints of the individual chip vias (interconnections between layers) and circuit pathways. Then, another layer of semiconductor material is deposited on the wafer with more photoresist on top, and the next mask is used to produce 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.

The final masks add the *metallization* layers, which are the metal interconnects used to tie all the individual transistors and other components together. Most chips use aluminum interconnects today, although many will be moving to copper in the future. The first commercial PC chip using copper is the Athlon made in AMD’s Dresden fab. Copper is a better conductor than aluminum and will allow smaller interconnects with less resistance, meaning smaller and faster chips can be made. The reason copper hasn’t been used up until recently is that there were difficult corrosion problems to overcome during the manufacturing process that were not as much a problem with aluminum. As these problems have been solved, there will be more and more chips fabricated with copper interconnects.

A completed circular wafer will have as many chips imprinted on it as can possibly fit. Because each chip is normally 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 standard wafer size used in the industry today is 200mm in diameter, or just under eight inches. This results in a wafer of about 31,416 square millimeters. The Pentium II 300MHz processor, for example, was made up of 7.5 million transistors using a 0.35 micron (millionth of a meter) process. This process results in a die of exactly 14.2mm on each side, which is 202 square millimeters of area. This means that about 150 total Pentium II 300MHz chips on the .35 micron process could be made from a single 200mm-diameter wafer.

The trend in the industry is to go to both larger wafers and a smaller chip die process. Process refers to the size of the individual circuits and transistors on the chip. For example, the Pentium II 333MHz through 450MHz processors were made on a newer and smaller .25 micron process, which reduced the total chip die size to only 10.2mm on each side, or a total chip area of 104 square millimeters. On the same 200mm (8-inch) wafer as before, Intel can make about 300 Pentium II chips using this process, or double the amount over the larger .35 micron process 300MHz version.

The Pentium III in the 600MHz and faster speeds is built on a .18 micron process and has a die size of only 104 square millimeters, which is about 10.2mm on each side. This is the same size as the older Pentium II, even though the newer PIII has 28.1 million transistors (including the on-die L2 cache) compared to only 7.5 million for the Pentium II.

In the future, processes will move from .18 micron to .13 micron, and from 200mm (eight-inch) wafers to 300mm (12-inch) wafers. The larger 300mm wafers alone will allow for more than double the number of chips to be made, compared to the 200mm mostly used today. The smaller 0.13-micron process will allow more transistors to be incorporated into the die while maintaining a reasonable die size allowing for sufficient yield. This means the trend for incorporating L2 cache within the die will continue, and transistor counts will rise up to 200 million per chip or more in the future. The current king of transistors is the Intel Pentium III Xeon introduced in May 2000 with 2MB of on-die cache and a whopping 140 million transistors in a single die.

The trend in wafers is to move from the current 200mm (eight-inch) diameter to a bigger, 300mm (12-inch) diameter wafer. This will increase surface area dramatically over the smaller 200mm design and boost chip production to about 675 chips per wafer. Intel and other manufacturers expect to have 300mm wafer production in place during 2001. After that happens, chip prices should continue to drop dramatically as supply increases.

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 percent 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 percent 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. For example, AMD was plagued by low-yield problems during 1997 and 1998, which cost it significant market share. It has since solved the problems, and lately it seems Intel has had the harder time meeting production demands.

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 die 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 where 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 will 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 III 750, 866, and 1000 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. This means that out of a wafer of 150 total chips, perhaps more than 100 of them check out at 1000MHz, while only a few won't run at that speed. The paradox is that Intel often sells a lot more of the lower-priced 933 and 866MHz chips, so it will just dip into the bin of 1000MHz processors and label them as 933 or 866 chips and sell them that way. People began discovering that many of the lower-rated chips would actually run at speeds much higher than they were rated, and the business of overclocking was born. Overclocking describes the operation of a chip at a speed higher than it was rated for. In many cases, people have successfully accomplished this because, in essence, they had a higher-speed processor already—it was marked with a lower rating only because it was sold as the slower version.

An interesting problem then arose: Unscrupulous vendors began taking slower chips and remarking them 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 will normally run well past their rated speed—the remarked 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 remarked chips were just a case of rubbing off the original numbers and restamping with new official-looking numbers. These were normally easy to detect. Remarkers then resorted to manufacturing completely new processor housings, especially for the plastic-encased Slot 1 and Slot A processors from Intel and AMD. Although it may seem to be a huge bother to make a custom plastic case and swap it with the existing case, since the profits can be huge, criminals find it very lucrative. This type of remarking 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 remarking by building overclock protection in the form of a multiplier lock into most of its newer chips. 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. Normally this involves changing the bus frequency (BF) pins on the chip, which control the internal multipliers the chip uses. Even so, enterprising individuals have found ways to run their motherboards at bus speeds higher than normal, so even though the chip won't allow a higher multiplier, you can still run it at a speed higher than it was designed.

Be Wary of PII and PIII Overclocking Fraud

Also note that unscrupulous individuals have devised a small logic circuit that bypasses the multiplier lock, allowing the chip to run at higher multipliers. This small circuit can be hidden in the PII or PIII cartridge, and then the chip can be remarked or relabeled to falsely indicate it is a higher speed version. This type of chip remarketing fraud is far more common in the industry than people want to believe. In fact, if you purchase your system or processor from a local computer flea market show, you have an excellent chance of getting a remarked chip. I recommend purchasing processors only from more reputable direct distributors or dealers. Contact Intel, AMD, or Cyrix, for a list of their reputable distributors and dealers.

I recently installed a 200MHz Pentium processor in a system that is supposed to run at a 3x multiplier based off a 66MHz motherboard speed. I tried changing the multiplier to 3.5x but the chip refused to go any faster; in fact, it ran at the same or lower speed than before. This is a sure sign of overclock protection inside, which is to say that the chip won't support any higher level of multiplier than it was designed for. Today, all Intel Pentium II and III processors are multiplier locked, which means the multiplier can no longer be controlled by the motherboard. This means that overclocking can be accomplished only by running the motherboard at a higher bus speed than the processor was designed for. My motherboard at the time included a jumper setting for an unauthorized speed of 75MHz, which when multiplied by 3x resulted in an actual processor speed of 225MHz. This worked like a charm, and the system is now running fast and clean. Many new motherboards have BIOS or jumper settings which can be used to tweak the motherboard bus speeds a few MHz higher than normal, which is then internally multiplied by the processor to even higher speeds. Note that I am not necessarily recommending overclocking for everybody; in fact, I normally don't recommend it at all for any important systems. If you have a system you want to fool around with, it is interesting to try. Like my cars, I always seem to want to hotrod my computers.

The real problem with the overclock protection as implemented by Intel and AMD is that the professional counterfeiter can still override it by inserting some custom circuitry underneath the plastic case enclosing the processor. This again is particularly a problem with the slot-based processors, since they use a case cover that can hide this circuitry. Socketed processors are much more immune to these remarking attempts. 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 since it is so easy to defraud the purchaser. Also the traveling computer show/flea market arenas can be a hotbed of this type of activity.

Fraudulent computer components are not limited to processors; I have seen fake memory (SIMMs/DIMMs), fake mice, fake video cards, fake cache memory, counterfeit operating systems and applications, and even fake motherboards. The hardware that is faked normally 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 is the Microsoft mouse. They sell for $35 wholesale and yet I can 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 run off with a substantial profit.

PGA Chip Packaging

PGA packaging has been the most common chip package used until recently. It was used starting with the 286 processor in the 1980s and is still used today for Pentium and Pentium Pro processors. 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 ZIF (Zero Insertion Force) 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 SPGA (Staggered Pin Grid Array), where 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 3.5 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 3.5
PGA on Pentium 66 (left) and dual-pattern SPGA on Pentium Pro (right).

Single Edge Contact (SEC) and Single Edge Processor (SEP) Packaging

Abandoning the chip-in-a-socket approach used by virtually all processors until this point, the

Pentium II/III chips are characterized by their Single Edge Contact (SEC) cartridge design. The processor, along with several L2 cache chips, is mounted on a small circuit board (much like an oversized memory SIMM), which 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.

By placing the processor and L2 cache as separate chips inside a cartridge, Intel now has a CPU module that is easier and less expensive to make than the Pentium Pro that preceded it. The Single Edge Contact (SEC) cartridge is an innovative—if a bit unwieldy—package design that incorporates the backside bus and L2 cache internally. Using the SEC design, the core and L2 cache are fully enclosed in a plastic and metal cartridge. These subcomponents are surface mounted directly to a substrate (or base) inside the cartridge to enable high-frequency operation. The SEC cartridge technology allows the use of widely available, high-performance industry standard Burst Static RAMs (BSRAMs) for the dedicated L2 cache. This greatly reduces the cost compared to the proprietary cache chips used inside the CPU package in the Pentium Pro.

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 as the Pentium II, but without the fancy plastic cover. The SEP package plugs directly into the same Slot 1 connector used by the standard Pentium II. Four holes on the board allow for the heat sink to be installed.

Slot 1 is the connection to the motherboard and has 242 pins. The Slot 1 dimensions are shown in Figure 3.6. The SEC cartridge or SEP processor is plugged into Slot 1 and secured with a processor-retention mechanism, which is a bracket that holds it in place. There may also be a retention mechanism or support for the processor heat sink. Figure 3.7 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 3.8.

Figure 3.6
Pentium II Processor Slot 1 dimensions (metric/English).

Figure 3.7
Pentium II Processor SEC package parts.

Figure 3.8
Celeron Processor SEP package front-side view.

With the Pentium III, Intel introduced a variation on the SEC packaging called SECC2 (Single Edge Contact Cartridge version 2). This new package covers only one side of the processor board and allows the heat sink to directly attach to the chip on the other side. This direct thermal interface allows for better cooling, and the overall lighter package is cheaper to manufacture. Note that a new Universal Retention System, consisting of a new design plastic upright stand, is required to hold the SECC2 package chip in place on the board. The Universal Retention System will also work with the older SEC package as used on most Pentium II processors, as well as the SEP package used on the slot based Celeron processors, making it the ideal retention mechanism for all Slot 1-based processors. Figure 3.9 shows the SECC2 package.

Figure 3.9
SECC2 packaging used in newer Pentium II and III processors.

The main reason for going to the SEC and SEP packages in the first place was to be able to move the L2 cache memory off the motherboard and onto the processor in an economical and scalable way. Using the SEC/SEP design, Intel can easily offer Pentium II/III processors with more or less cache and faster or slower cache.

Processor Sockets and Slots

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 3.12 shows the specifications of these sockets.

Table 3.12 CPU Socket and Slot Types and Specifications

Socket Number	**Pins**	**Pin Layout**	**Voltage**	**Supported Processors**
Socket 1	169	17x17 PGA	5v	486 SX/SX2, DX/DX2[1], DX4 Overdrive
Socket 2	238	19x19 PGA	5v	486 SX/SX2, DX/DX2[1], DX4 Overdrive, 486 Pentium Overdrive
Socket 3	237	19x19 PGA	5v/3.3v	486 SX/SX2, DX/DX2, DX4, 486 Pentium Overdrive, AMD 5x86
Socket 4	273	21x21 PGA	5v	Pentium 60/66, Overdrive
Socket 5	320	37x37 SPGA	3.3/3.5v	Pentium 75-133, Overdrive
Socket 6[2]	235	19x19 PGA	3.3v	486 DX4, 486 Pentium Overdrive
Socket 7	321	37x37 SPGA	VRM	Pentium 75-233+, MMX, Overdrive, AMD K5/K6, Cyrix M1/II
Socket Number	Pins	Pin Layout	Voltage	Supported Processors
Socket 8	387	dual pattern SPGA	Auto VRM	Pentium Pro
Socket 370 (PGA370)	370	37x37 SPGA	Auto VRM	Celeron/Pentium III PPGA/FC-PGA
Slot A	242	Slot	Auto VRM	AMD Athlon PGA

Socket A (Socket 462)	462	PGA Socket	Auto VRM	AMD Athlon/Duron SECC
Slot 1 (SC242)	242	Slot	Auto VRM	Pentium II/III, Celeron SECC
Slot 2 (SC330)	330	Slot	Auto VRM	Pentium II/III Xeon

1. Non-overdrive DX4 or AMD 5x86 also can be supported with the addition of an aftermarket 3.3v voltage-regulator adapter.
2. Socket 6 was a paper standard only and was never actually implemented in any systems.
PGA = Pin Grid Array
PPGA = Plastic Pin Grid Array
FC-PGA = Flip Chip Pin Grid Array
SPGA = Staggered Pin Grid Array
SECC = Single Edge Contact Cartridge
VRM = Voltage Regulator Module

Sockets 1, 2, 3, and 6 are 486 processor sockets and are shown together in Figure 3.10 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 3.11 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 the thorough descriptions of the sockets.

Figure 3.10
486 processor sockets.

Figure 3.11
Pentium and Pentium Pro processor sockets.

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 3.12 shows the pinout of Socket 1.

Figure 3.12
Intel Socket 1 pinout.

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 heat sink consisting of aluminum fins that are glued to the processor with thermal transfer epoxy. Passive heat sinks don’t have any mechanical components like fans. Heat sinks with fans or other devices that use power are called active heat sinks. OverDrive processors rated at 40MHz or less do not have heat sinks.

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 will 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 is 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 3.13 shows the pinout configuration of the official Socket 2 design.

Figure 3.13
238-pin Intel Socket 2 configuration.

Notice that although the new 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 heat sink to the Pentium OverDrive processor. This active heat sink is a combination of a standard heat sink 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/heat sink assembly clips and plugs directly into the processor and provides for easy replacement if the fan fails.

Another requirement of the active heat sink is additional clearance—no obstructions for an area about 1.4 inches off the base of the existing socket to allow for heat-sink clearance. The Pentium OverDrive upgrade will be difficult or impossible in systems that were not designed with this feature.

Another problem with this particular upgrade is power consumption. The 5v Pentium OverDrive processor will draw up to 2.5 amps at 5v (including the fan) or 12.5 watts, which is more than double the 1.2 amps (6 watts) drawn by the DX2 66 processor.

Note - See Intel's Web site (http://www.intel.com) for a comprehensive list of certified OverDrive-compatible systems.

Socket 3

Because of problems with the original Socket 2 specification and the enormous heat the 5v version of the Pentium OverDrive processor generates, Intel came up with an improved design. The new processor is the same as the previous Pentium OverDrive processor, except that it runs on 3.3v and draws a maximum 3.0 amps of 3.3v (9.9 watts) and 0.2 amp of 5v (1 watt) to run the fan—a total 10.9 watts. This configuration provides a slight margin over the 5v version of this processor. The fan will be easy to remove from the OverDrive processor for replacement, should it ever fail.

Intel had to create a new socket to support both the DX4 processor, which runs on 3.3v, and the 3.3v Pentium OverDrive processor. In addition to the new 3.3v chips, this new socket supports the older 5v SX, DX, DX2, and even the 5v Pentium OverDrive chip. The design, called Socket 3, is the most flexible upgradable 486 design. Figure 3.14 shows the pinout specification of Socket 3.

Figure 3.14
237-pin Intel Socket 3 configuration.

Notice that Socket 3 has one additional pin and several others plugged compared with Socket 2. Socket 3 provides for better keying, which prevents an end user from accidentally installing the processor in an improper orientation. However, one serious problem exists: This socket cannot automatically determine the type of voltage that will be provided to it. A jumper is likely to be added on the motherboard near the socket to enable the user to select 5v or 3.3v operation.

> **Caution -** Because this jumper must be manually set, however, a user could install a 3.3v processor in this socket when it is configured for 5v operation. This installation will instantly destroy a very expensive chip when the system is powered on. So, it is up to the end user to make sure 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 that was designed for the original Pentium processors. The original Pentium 60MHz and 66MHz version processors had 273 pins and would plug into Socket 4—a 5v-only socket, because all the original Pentium processors run on 5v. This socket will accept the original Pentium 60MHz or 66MHz processor, and the OverDrive processor. Figure 3.15 shows the pinout specification of Socket 4.

Figure 3.15
273-pin Intel Socket 4 configuration.

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 heat sink (fan). The 66MHz OverDrive processor that replaced it consumes a maximum 2.7 amps (13.5 watts), including about one 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 will 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 newer 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 will draw less power than the original processor, additional clearance may have to be allowed for the active heat sink assembly that is mounted on top. As in other OverDrive processors with built-in fans, the power to run the fan will be drawn directly from the chip socket, so no separate power-supply connection is required. Also, the fan will be easy to replace should it ever fail.

Socket 5

When Intel redesigned the Pentium processor to run at 75, 90, 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 can use far less power than even the original 60MHz version. The newest 120 and higher Pentium, Pentium Pro, and Pentium II chips use an even smaller die 0.35 micron process. This results in lower power consumption and allows the extremely high clock rates without over- heating.

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 320 pins. The additional pins are used by the Pentium OverDrive for Pentium processors. This socket has the 320 pins configured in a staggered Pin Grid Array, in which the individual pins are staggered for tighter clearance.

Several OverDrive processors for existing Pentiums are currently available. If you have a first-generation Pentium 60 or 66 with a Socket 4, you can purchase a standard Pentium OverDrive chip that effectively doubles the speed of your old processor. An OverDrive chip with MMX technology is available for second-generation 75MHz, 90MHz, and 100MHz Pentiums using Socket 5 or Socket 7. Processor speeds after upgrade are 125MHz for the Pentium 75, 150MHz for the Pentium 90, and 166MHz for the Pentium 100. MMX greatly enhances processor performance, particularly under multimedia applications, and is discussed in the section "Pentium-MMX Processors," later in this chapter. Figure 3.16 shows the standard pinout for Socket 5.

The Pentium OverDrive for Pentium processors has an active heat sink (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 means 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!

Figure 3.16
320-pin Intel Socket 5 configuration.

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 two pins plugged for proper chip keying. Socket 6 has 235 pins and would accept 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 21x21 SPGA arrangement. The real difference with Socket 7 is not the socket but with the companion VRM (Voltage Regulator Module) or 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 has several different versions of the Pentium and Pentium-MMX processors that run on 3.3v (called VR), 3.465v (called VRE), or 2.8v, while AMD, Cyrix, and others use different variations of 3.3V, 3.2V, 2.9V, 2.4V, 2.3V, 2.2V, 2.1V, 2.0V, 1.9V, or 1.8V. Because of the variety of voltages that may be required to support different processors, most newer motherboard manufacturers are either including VRM sockets or building adaptable VRMs into the motherboard.

Figure 3.17 shows the Socket 7 pinout.

Figure 3.17
Socket 7 (Pentium) Pinout (top view).

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 allows for faster Socket 7–type systems to be made, which are nearly as fast as 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.

New chipsets are required for Super7 boards. Major third-party chipset suppliers, including Acer Laboratories Inc. (ALi), VIA Technologies, and SiS, are supporting the Super7 platform. ALi, VIA, and SiS all have chipsets for Super7 boards. Most of the major motherboard manufacturers are making Super7 boards in both Baby-AT and especially ATX form factors.

If you want to purchase a Pentium class board that can be upgraded to the next generation of even higher speed Socket 7 processors, look for a system with a Super7 socket and an integrated VRM that supports the different voltage selections required by your intended processor.

Socket 8

Socket 8 is a special SPGA (Staggered Pin Grid Array) socket featuring a whopping 387 pins! This was specifically designed for the Pentium Pro processor with the integrated L2 cache. The additional pins are to allow the chipset to control the L2 cache that is integrated in the same package as the processor. Figure 3.18 shows the Socket 8 pinout.

Figure 3.18
Socket 8 (Pentium Pro) pinout showing power pin locations.

Socket 370 (PGA-370)

In January 1999, Intel introduced a new socket for P6 class processors. The new socket is called Socket 370 or PGA-370, because it has 370 pins and was designed for lower-cost PGA (Pin Grid Array) versions of the Celeron and Pentium III processors. Socket 370 is designed to directly compete in the lower-end system market along with the Super7 platform supported by AMD and Cyrix. Socket 370 brings the low cost of a socketed design, with less expensive processors, mounting systems, heat sinks, and so on to the high-performance P6 line of processors.

Initially all the Celeron and Pentium III processors were made in SECC (Single Edge Contact Cartridge) or SEPP (Single Edge Processor Package) formats. These are essentially circuit boards containing the processor and separate L2 cache chips on a small board that plugs into the motherboard via Slot 1. This type of design was necessary when the L2 cache chips were made a part of the processor, but were not directly integrated into the processor die. Intel did make a multi-die chip package for the Pentium Pro, but this proved to be a very expensive way to package the chip, and a board with separate chips was cheaper, which is why the Pentium II looks different from the Pentium Pro.

Starting with the Celeron 300A processor introduced in August 1998, Intel began combining the L2 cache directly on the processor die; it was no longer in separate chips. With the cache fully integrated into the die, there was no longer a need for a board-mounted processor. Because it costs more to make a Slot 1 board or cartridge-type processor instead of a socketed type, Intel moved back to the socket design to reduce the manufacturing cost—especially with the Celeron, which competes on the low end with Socket 7 chips from AMD and Cyrix.

The Socket 370 (PGA-370) pinout is shown in Figure 3.19.

Figure 3.19
Socket 370 (PGA-370) Pentium III/Celeron pinout (top view).

The Celeron is gradually being shifted over to PGA-370, although for a time both were available. All Celeron processors at 333MHz and lower were only available 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 are only available in the Socket 370 version.

FC-PGA (Flip Chip Pin Grid Array)

Starting in October 1999, Intel also introduced Pentium III processors with integrated cache that plug into Socket 370. These use a packaging called FC-PGA (Flip Chip Pin Grid Array), which is a type of packaging where the raw die is mounted on the substrate upside down. It seems that all future Pentium III processors will continue in the socket version because the slot version is more expensive and no longer needed.

Note that due to some voltage changes and one pin change, many original Socket 370 motherboards will not accept the later FC-PGA (Flip Chip PGA) 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 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 if 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#.

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 on either 1.60 or 1.65 volts, whereas the Intel Celeron processors operate at 2.00 volts. There is a chance that 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, a low-cost adapter called a "slot-ket" has been made available by several manufacturers. This is essentially a Slot 1 board containing only a Socket 370, which allows you to use a PGA (Pin Grid Array) processor in any Slot 1 board. A typical slot-ket adapter is shown in the "Celeron" section, later in this chapter.

Socket A (Socket 462)

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

Socket A has 462 pins and 11 plugs oriented in a SPGA (Staggered Pin Grid Array) form (see Figure 3.20). 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 32 different 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 will normally be embedded on the motherboard.

Figure 3.20
Socket A (Socket 462) Athlon/Duron layout.

There are 11 total plugged holes, including two 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 3.21.

AMD has indicated that all future Athlon and Duron processors, at least for the time being, will be made in Socket A form, and the Slot A versions of the Athlon will be phased out.

Figure 3.21
Socket A (Socket 462) Athlon/Duron pinout (top view).

Zero Insertion Force (ZIF) Sockets

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 it typically takes 100 pounds of insertion force 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 socket during removal or reinstallation. Because of this, some motherboard manufacturers began using Low Insertion Force (LIF) sockets, which typically required only 60 pounds 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 pounds 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 going to ever replace his CPU.

Manufacturers began inserting special Zero Insertion Force (ZIF) sockets in their later Socket 1 motherboard designs. Since then, virtually all processor sockets have been of the ZIF design. Note, however, that a given Socket X specification has nothing to do with whether it is ZIF, LIF, or standard; the socket specification covers only the pin arrangement. These days, nearly all motherboard manufacturers are using ZIF sockets. These sockets almost eliminate the risk involved in upgrading because no insertion force is necessary to install the chip. Most ZIF sockets are handle-actuated; you lift the handle, drop the chip into the socket, and then close the handle. This design makes replacing the original processor with the upgrade processor an easy task.

Because of the number of pins involved, virtually all CPU sockets from Socket 2 through the present are implemented in ZIF form. This means that since the 486 era, removing the CPU from most motherboards does not require any tools.

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. In order to accept its new cartridges, Intel designed two different types of slots that could be used on motherboards.

Slot 1 is a 242-pin slot that is designed to accept Pentium II, Pentium III, and most Celeron processors. Slot 2 is a more sophisticated 330-pin slot that is designed for the Pentium II and 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 only allows single or dual processing functionality.

Note that Slot 2 is also called SC330, which stands for Slot Connector with 330 pins.

Slot 1 (SC242)

Slot 1, also called SC242 (Slot Connector 242 pins), is used by the SEC (Single Edge Cartridge) design used with the cartridge-type Pentium II/III and Celeron processors. Inside the cartridge is a substrate card that includes the processor and L2 cache. Unlike the Pentium Pro, the L2 cache was mounted on the circuit board and not within the same chip package as the processor. This allowed Intel to use aftermarket SRAM chips instead of making them internally and also allowed it to make processors with different amounts of cache easily. For example, the original Celeron was created by taking a Pentium II and simply leaving out the external cache chips. Figure 3.22 shows the Slot 1 connector dimensions and pin layout.

Figure 3.22
Slot 1 connector dimensions and pin layout.

Table 3.13 lists the names of each of the pins in the Slot 1 connector.

Table 3.13 Signal Listing in Order by Pin Number

Pin No.	Pin Name
A1	VCC_VTT
A2	GND
A3	VCC_VTT
A4	IERR#
A5	A20M#
A6	GND
A7	FERR#
A8	IGNNE#
A9	TDI
A10	GND
A11	TDO
A12	PWRGOOD
A13	TESTHI
A14	GND
A15	THERMTRIP#
A16	Reserved
A17	LINT[0]/INTR
A18	GND
A19	PICD[0]
A20	PREQ#
A21	BP#[3]
A22	GND
A23	BPM#[0]
A24	BINIT#
A25	DEP#[0]
A26	GND
A27	DEP#[1]
A28	DEP#[3]
A29	DEP#[5]

A30	GND
A31	DEP#[6]
A32	D#[61]
A33	D#[55]
A34	GND
A35	D#[60]
A36	D#[53]
A37	D#[57]
A38	GND
A39	D#[46]
A40	D#[49]
A41	D#[51]
A42	GND
A43	D#[42]
A44	D#[45]
A45	D#[39]
A46	GND
A47	Reserved
A48	D#[43]
A49	D#[37]
A50	GND
A51	D#[33]
A52	D#[35]
A53	D#[31]
A54	GND
A55	D#[30]
A56	D#[27]
A57	D#[24]
A58	GND
A59	D#[23]
A60	D#[21]
A61	D#[16]
A62	GND
A63	D#[13]
A64	D#[11]
A65	D#[10]

A66	GND
A67	D#[14]
A68	D#[9]
A69	D#[8]
A70	GND
A71	D#[5]
A72	D#[3]
A73	D#[1]
A74	GND
A75	BCLK
A76	BR0#
A77	BERR#
A78	GND
A79	A#[33]
A80	A#[34]
A81	A#[30]
A82	GND
A83	A#[31]
A84	A#[27]
A85	A#[22]
A86	GND
A87	A#[23]
A88	Reserved
A89	A#[19]
A90	GND
A91	A#[18]
A92	A#[16]
A93	A#[13]
A94	GND
A95	A#[14]
A96	A#[10]
A97	A#[5]
A98	GND
A99	A#[9]
A100	A#[4]
A101	BNR#

A102	GND
A103	BPRI#
A104	TRDY#
A105	DEFER#
A106	GND
A107	REQ#[2]
A108	REQ#[3]
A109	HITM#
A110	GND
A111	DBSY#
A112	RS#[1]
A113	Reserved
A114	GND
A115	ADS#
A116	Reserved
A117	AP#[0]
A118	GND
A119	VID[2]
A120	VID[1]
A121	VID[4]
B1	EMI
B2	FLUSH#
B3	SMI#
B4	INIT#
B5	VCC_VTT
B6	STPCLK#
B7	TCK
B8	SLP#
B9	VCC_VTT
B10	TMS
B11	TRST#
B12	Reserved
B13	VCC_CORE
B14	Reserved
B15	Reserved
B16	LINT[1]/NMI

B17	VCC_CORE
B18	PICCLK
B19	BP#[2]
B20	Reserved
B21	BSEL#
B22	PICD[1]
B23	PRDY#
B24	BPM#[1]
B25	VCC_CORE
B26	DEP#[2]
B27	DEP#[4]
B28	DEP#[7]
B29	VCC_CORE
B30	D#[62]
B31	D#[58]
B32	D#[63]
B33	VCC_CORE
B34	D#[56]
B35	D#[50]
B36	D#[54]
B37	VCC_CORE
B38	D#[59]
B39	D#[48]
B40	D#[52]
B41	EMI
B42	D#[41]
B43	D#[47]
B44	D#[44]
B45	VCC_CORE
B46	D#[36]
B47	D#[40]
B48	D#[34]
B49	VCC_CORE
B50	D#[38]
B51	D#[32]
B52	D#[28]

B53	VCC_CORE
B54	D#[29]
B55	D#[26]
B56	D#[25]
B57	VCC_CORE
B58	D#[22]
B59	D#[19]
B60	D#[18]
B61	EMI
B62	D#[20]
B63	D#[17]
B64	D#[15]
B65	VCC_CORE
B66	D#[12]
B67	D#[7]
B68	D#[6]
B69	VCC_CORE
B70	D#[4]
B71	D#[2]
B72	D#[0]
B73	VCC_CORE
B74	RESET#
B75	BR1#
B76	FRCERR
B77	VCC_CORE
B78	A#[35]
B79	A#[32]
B80	A#[29]
B81	EMI
B82	A#[26]
B83	A#[24]
B84	A#[28]
B85	VCC_CORE
B86	A#[20]
B87	A#[21]
B88	A#[25]

B89	VCC_CORE
B90	A#[15]
B91	A#[17]
B92	A#[11]
B93	VCC_CORE
B94	A#[12]
B95	A#[8]
B96	A#[7]
B97	VCC_CORE
B98	A#[3]
B99	A#[6]
B100	EMI
B101	SLOTOCC#
B102	REQ#[0]
B103	REQ#[1]
B104	REQ#[4]
B105	VCC_CORE
B106	LOCK#
B107	DRDY#
B108	RS#[0]
B109	VCC5
B110	HIT#
B111	RS#[2]
B112	Reserved
B113	VCC_L2
B114	RP#
B115	RSP#
B116	AP#[1]
B117	VCC_L2
B118	AERR#
B119	VID[3]
B120	VID[0]
B121	VCC_L2

Slot 2 (SC330)

Slot 2, otherwise called SC330 (Slot Connector 330 pins) is used on high-end motherboards that support the Pentium II and III Xeon processors. Figure 3.23 shows the Slot 2 connector.

Figure 3.23
Slot 2 (SC330) connector dimensions and pin layout.

The Xeon processors are designed in a cartridge similar to, but larger than, that used for the standard Pentium II/III. Figure 3.24 shows the Xeon cartridge.

Motherboards featuring Slot 2 are primarily found in higher-end systems such as workstations or servers, which use the Xeon processors. These are Intel's high-end chips, which differ from the standard Pentium II/III mainly by virtue of having full-core speed L2 cache, and in some versions more of it.

Figure 3.24
Pentium II/III Xeon cartridge.

CPU Operating Voltages

One trend that is clear to anybody that 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 will be less expensive to run, but more importantly for portable or mobile systems, it will run 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, there will be less heat produced. Processors that run cooler can be packed into systems more tightly and will 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 allowing the clock rates of processors to go higher and higher.

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 5 volts, which was later was reduced to 3.5 or 3.3 volts 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 where the processor core ran off of a lower voltage while the I/O circuits remained at 3.3v. This was originally 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 A, Pentium Pro (Socket 8), or Pentium II (Slot 1 or Slot 2) processors, because these sockets and slots have special voltage ID (VID) pins that the processor uses to signal to the motherboard the exact voltage requirements. This allows the voltage regulators built in to the motherboard to be automatically set to the correct voltage levels by merely installing the processor.

Unfortunately, this automatic voltage setting feature is not available on Socket 7 and earlier motherboard and processor designs. This means you must normally 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 latest MMX versions are all 2.8v, except for mobile Pentium processors, which are as low as 1.8v. Table 3.11 lists the voltage settings used by Intel Pentium (non-MMX) processors that use a single power plane. This means that both the CPU core and the I/O pins run at the same voltage.

Table 3.14 shows voltages used by Socket 7 processors.

Table 3.14 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

Normally, the acceptable range is plus or minus five percent 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 different levels required by the processor core. Check the documentation for your motherboard and processor to find the appropriate settings.

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

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 different 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 only designed to supply 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 only supply 5v, plus they have a completely different pinout (fewer pins overall), so it is not possible to use a processor designed for Socket 7 or Socket 5 in Socket 4.

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 different levels required by the processor core. Check the documentation for your motherboard and processor to find the appropriate settings.

The Pentium Pro, Celeron, and Pentium II/III processors automatically determine their voltage settings by controlling the motherboard-based voltage regulator. That’s done through built-in voltage ID (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, it often helps to increase the voltage by a tenth of a volt or so. Of course this increases the heat output of the processor and must be accounted for with adequate heat sinking.

Heat and Cooling Problems

Heat can be a problem in any high-performance system. The higher-speed processors normally 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 might not be capable of handling the load without some help.

Heat Sinks

To cool a system in which processor heat is a problem, you can buy (for less than $5, in most cases) a special attachment for the CPU chip called a heat sink, which draws heat away from the CPU chip. Many applications may need only a larger standard heat sink with additional or longer fins for a larger cooling area. Several heat-sink manufacturers are listed in the Vendor List, on the CD.

A heat sink works like the radiator in your car, pulling heat away from the engine. In a similar fashion, the heat sink conducts heat away from the processor so that it can be vented out of the system. It does this by using a thermal conductor (usually metal) to carry heat away from the processor into fins that expose a high amount of surface area to moving air. This allows the air to be heated, thus cooling the heat sink and the processor as well. Just like the radiator in your car, the heat sink depends on airflow. With no moving air, a heat sink is incapable of radiating the heat away. To keep the engine in your car from overheating when the car is not moving, auto engineers incorporate a fan. Likewise, there is always a fan somewhere inside your PC helping to move air across the heat sink and vent it out of the system. Sometimes the fan included in the power supply is enough, other times an additional fan must be added to the case, or even directly over the processor to provide the necessary levels of cooling.

The heat sink is clipped or glued to the processor. A variety of heat sinks and attachment methods exist. Figure 3.25 shows various passive heat sinks and attachment methods.

Figure 3.25
Passive heat sinks for socketed processors showing various attachment methods.

Tip - According to data from Intel, heat sink clips are the number-two destroyer of motherboards (screwdrivers are number one). When installing or removing a heat sink that is clipped on, make sure you don’t scrape the surface of the motherboard. In most cases, the clips hook over protrusions in the socket, and when installing or removing the clips, it is very easy to scratch or scrape the surface of the board right below where the clip ends attach. I like to place a thin sheet of plastic underneath the edge of the clip while I work, especially if there are board traces that can be scratched in the vicinity.

Heat sinks are rated for their cooling performance. Typically the ratings are expressed as a resistance to heat transfer, in degrees centigrade per watt (°C/W), where lower is better. Note that the resistance will vary according to the airflow across the heat sink. To ensure a constant flow of air and more consistent performance, many heat sinks incorporate fans so they don't have to rely on the airflow within the system. Heat sinks with fans are referred to as active heat sinks (see Figure 3.26). Active heat sinks have a power connection, often using a spare disk drive power connector, although most newer motherboards now have dedicated heat sink power connections right on the board.

Figure 3.26
Active heat sinks for socketed processors.

Active heat sinks use a fan or other electric cooling device, which require power to run. The fan type is most common but some use a peltier cooling device, which is basically a solid-state refrigerator. Active heat sinks require power and normally plug into a disk drive power connector or special 12v fan power connectors on the motherboard. If you do get a fan-type heat sink, be aware that some on the market are very poor quality. The bad ones have motors that use sleeve bearings, which freeze up after a very short life. I only recommend fans with ball-bearing motors, which will last about 10 times longer than the sleeve-bearing types. Of course, they cost more, but only about twice as much, which means you'll save money in the long run.

Figure 3.27 shows an active heat sink arrangement on a Pentium II/III type processor. This is common on what Intel calls its "boxed processors," which are sold individually and through dealers.

Figure 3.27
An active (fan-powered) heat sink and supports used with Pentium II/III–type processors.

The passive heat sinks are 100 percent reliable, as they have no mechanical components to fail. Passive heat sinks (see Figure 3.28) are basically aluminum-finned radiators that dissipate heat through convection. Passive types don't work well unless there is some airflow across the fins, normally provided by the power supply fan or an extra fan in the case. If your case or power supply is properly designed, you can use a less-expensive passive heat sink instead of an active one.

Figure 3.28
A passive heat sink and supports used with Pentium II/III–type processors.

> **Tip -** To function effectively, a heat sink must be as directly attached to the processor as possible. To eliminate air gaps and ensure a good transfer of heat, in most cases, you should put a thin coating of thermal transfer grease on the surface of the processor where the heat sink attaches. This will dramatically decrease the thermal resistance properties and is required for maximum performance.

To have the best possible transfer of heat from the processor to the heat sink, most heat sink manufacturers specify some type of thermal interface material to be placed between the processor and the heat sink. This normally consists of a zinc-based white grease (similar to what skiers put on their noses to block the sun), but can also be a special pad or even a type of double-stick tape. Using a thermal interface aid such as thermal grease can improve heat sink performance dramatically. Figure 3.29 shows the thermal interface pad or grease positioned between the processor and heat sink.

Figure 3.29
Thermal interface material helps transfer heat from the processor die to the heat sink.

Most of the newer systems on the market use an improved motherboard form factor (shape) design called ATX. Systems made from this type of motherboard and case allow for improved cooling of the processor due to the processor being repositioned in the case near the power supply. Also, most of these cases now feature a secondary fan to further assist in cooling. Normally the larger case-mounted fans are more reliable than the smaller fans included in active heat sinks. A properly designed case can move sufficient air across the processor, allowing for a more reliable and less-expensive passive (no fan) heat sink to be used.

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, although the Intel ones are known for having the best performance.

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 3.15 summarizes the coprocessors available for the Intel family of processors.

Table 3.15 Math Coprocessor Summary

Processor	Coprocessor
8086	8087
8088	8087
286	287
386SX	387SX
386DX	387DX
486SX	487SX, DX2/OverDrive[1]
487SX[1]	Built-in FPU
486SX2	DX2/OverDrive[2]
486DX	Built-in FPU
486DX2	Built-in FPU
486DX4/5x86	Built-in FPU
Intel Pentium/Pentium MMX	Built-in FPU
Cyrix 6x86/MI/MII	Built-in FPU
AMD K5/K6/Athlon/Duron	Built-in FPU

Pentium II/III/Celeron/Xeon	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 3.16, indicates the maximum speed at which a system can run a math chip.

Table 3.16 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 rating 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 that they can run at different speeds compared with the rest of the system.

Tip - The performance increase in programs that use the math chip can be dramatic—usually, a geometric increase in speed occurs. If the primary applications that you use can take advantage of a math coprocessor, you should upgrade your system to include one.

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 don't even have a socket for the coprocessor because of cost and size considerations. These systems are 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 3.17 shows the specifications of the various math coprocessors.

Table 3.17 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 what CPU and math coprocessor are installed in a particular system by checking the markings on the chip.

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.

Companies such as Diagsoft, Symantec, Micro 2000, Trinitech, Data Depot, and others offer specialized diagnostics software that can test the system, including the processor. If you don't want to purchase this kind of software, you can perform a quick-and-dirty processor evaluation by using the diagnostics program supplied with your system.

Perhaps the most infamous of these 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.

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 manuals; this manual is available from Intel’s Web site. Most of the other processor manufacturers also have bulletins or tips on their Web sites 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. Now, a new feature built into the Intel P6 processors, including the Pentium Pro and Pentium II, can allow many bugs to 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 allows certain types of bugs to be worked around via microcode updates. The microcode updates reside in the system ROM BIOS and are loaded into the processor by the system BIOS during the power on self test (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 easiest method for checking the microcode update is to use the Pentium Pro and Pentium II processor update utility, which is developed and supplied by Intel. This utility can verify whether the correct update is present for all Pentium Pro processor-based motherboards. The utility displays the processor *stepping* and version of the microcode update. A stepping is the processor hardware equivalent of a new version. In software, we refer to minor version changes as 1.0, 1.1, 1.2, and so on, while in processors we call these minor revisions steppings.

To install a new microcode update, however, the motherboard BIOS must contain the routines to support the microcode update, which virtually all Pentium Pro and Pentium II BIOSes do have. The Intel processor update utility determines whether the code is present in the BIOS, compares the processor stepping with the microcode update currently loaded, and installs the new update, if needed. Use of this utility with motherboards containing the BIOS microcode update routine allows just the microcode update data to be changed; the rest of the BIOS is unchanged. A version of the update utility is provided with all Intel boxed processors. The term *boxed processors* refers to processors packaged for use by system integrators—that is, people who build systems. If you want the most current version of this utility, you have to contact an Intel processor dealer to download it, because Intel only supplies it to its dealers.

Note that if the BIOS in your motherboard does not include the processor microcode update routine, you should get a complete system BIOS upgrade from the motherboard vendor.

When you are building a system with a Pentium Pro, Celeron, or Pentium II/III processor, you must use the processor update utility to check that the system BIOS contains microcode updates that are specific to particular silicon stepping of the processor you are installing. In other words, you must ensure that the update matches the processor stepping being used.

Table 3.18 contains the current microcode update revision for each processor stepping. These update revisions are contained in the microcode update database file that comes with the Pentium Pro processor and Pentium II processor update utility. Processor steppings are listed in the sections on the Pentium, Pentium Pro, and Pentium II processors later in this chapter.

Table 3.18 Processor Steppings (Revisions) and Microcode Update Revisions Supported by the Update Database File PEP6.PDB

Processor	Stepping	Stepping Signature	Microcode Update Revision Required
Pentium Pro	C0	0x612	0xC6
Pentium Pro	sA0	0x616	0xC6
Pentium Pro	sA1	0x617	0xC6
Pentium Pro	sB1	0x619	0xD1
Pentium II	C0	0x633	0x32
Pentium II	C1	0x634	0x33
Pentium II	dA0	0x650	0x15

Using the processor update utility (CHECKUP3.EXE) available from Intel, a system builder can easily verify that the correct microcode update is present in all systems based on the P6 (Pentium Pro, Celeron, Pentium II/III, and Xeon) processors. For example, if a system contains a processor with stepping C1 and stepping signature 0x634, the BIOS should contain the microcode update revision 0x33. The processor update utility identifies the processor stepping, signature, and microcode update revision that is currently in use.

If a new microcode update needs to be installed in the system BIOS, the system BIOS must contain the Intel-defined processor update routines so the processor update utility can permanently install the latest update. Otherwise, a complete system BIOS upgrade is required from the motherboard manufacturer. It is recommended that the processor update utility be run after upgrading a motherboard BIOS and before installing the operating system when building a system based on any P6 processor. The utility is easy to use and executes in just a few seconds. Because the update utility may need to load new code into your BIOS, ensure that any jumper settings on the motherboard are placed in the "enable flash upgrade" position. This enables writing to the flash memory.

After running the utility, turn off power to the system and reboot—do not warm boot—to ensure that the new update is correctly initialized in the processor. Also ensure that all jumpers, such as any flash upgrade jumpers, and so on, are returned to their normal position.

Both Intel and AMD have a number of useful utilities available for downloading from their Web sites, utilities which can be very useful for somebody who is building a system. These include utilities which will identify exactly which processor you have, the speed the processor is running at, in some cases whether it has been overclocked, and even useful information like processor temperature, steppings, and more.

In addition to software utilities, both Intel and AMD have extensive technical documentation libraries online, all free for downloading. These documents range from the actual processor and chipset data books to motherboard schematics, design guides, and even information about cooling and thermal requirements.

Processor Codenames

Intel, AMD, and Cyrix have always used codenames when talking about future processors. The codenames are normally not supposed to become public, but often they do. They can often be found in 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 3.19 lists processor codenames for reference.

Table 3.19 Processors and Codenames

AMD Codename	**AMD Processor**
X5	5x86-133 [Socket 3]
SSA5	K5 (PR75-100) [Socket 5, 7]
5k86	K5 (PR120-200) [Socket 7]
K6	Original K6 core; killed after AMD acquired NexGen
NX686	NexGen core that became the K6 [Socket 7]
Little Foot	0.25µ K6 [Socket 7]
Chompers	K6-2 (formerly K6-3D) [Socket 7, Super 7]
Sharptooth	K6-3 (formerly K6 Plus-3D) [Super 7]
Argon	Original codename for K7
K7	Athlon [Slot A]
K75	0.18µ Athlon [Slot A]
Spitfire	Duron [Socket A]
Thunderbird	Athlon [Slot A, Socket A]
Mustang	Athlon w/copper interconnects [Slot A, Socket A]
Corvette	Mobile Athlon [Socket A]
SledgeHammer	K8 (64-bit CPU)
Cyrix Codename	**Cyrix Processor**
M6	486DX [Socket 1, 2, 3]
M7	486DX2/DX4 [Socket 3]
M9	5x86 [Socket 3]
M1sc	5x86 [Socket 3]
Chili	5x86 project
M1	6x86 (3.3v or 3.52v version) [Socket 7]
M1L	6x86L (2.8v/3.3v split version)

	[Socket 7]
M1R	Switch from 3M SGS process to 5M IBM process for 6x86
M2	6x86MX/M-II [Socket 7, Super 7]
Cayenne	MXi and Gobi core
Jedi	Original codename for Joshua (before Gobi)
Gobi	Former codename for Joshua
Joshua	VIA/Cyrix-III [Socket 370]
Jalapeno	Former codename for Mojave
Mojave	Cyrix/VIA M3 [Socket 370]
Serrano	Cyrix/VIA M4
C5	Samuel core (Winchip-4 plus on-Die L2 cache)
Samuel	Cyrix/VIA chip based on Winchip-4 [Socket 370]
Intel Codename	**Intel Processor**
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 enhanced version) [Socket 3]
P24C	486DX4 [Socket 3]
P23T	486DXODP (486 overdrive) [Socket 1, 2, 3]
P4T	486DXODPR (486 overdrive) [Socket 1, 2, 3]
P24T	PODP5V (Pentium OverDrive for 486) [Socket 2, 3]
P24CT	Pentium OverDrive for 486DX4 (3.3v core) [Socket 2, 3]

P5	Pentium (60/66MHz versions) [Socket 4]
P5T	Pentium OverDrive (120, 133) [Socket 4]
P54C	Pentium (75-120MHz versions) [Socket 5, 7]
P54CQS	Pentium (120-133MHz version) [Socket 5, 7]
P54CS	Pentium (120-200MHz versions) [Socket 7]
P54CTA	Pentium OverDrive (125, 150, 166) [Socket 5, 7]
Intel Codename	**Intel Processor**
P55C	Pentium MMX [Socket 7]
P54CTB	Pentium MMX OverDrive [Socket 5, 7]
Tillamook	mobile Pentium MMX
P6	Pentium Pro [Socket 8]
P6T	Pentium II OverDrive [Socket 8]
Klamath	Pentium II [Slot 1]
Drake	Pentium II Xeon [Slot 2]
Deschutes	0.25µ Pentium II [Slot 1 & 2]
Tonga	Mobile Pentium II
Covington	Celeron (cacheless Deschutes) [Slot 1]
Mendocino	Celeron (128KB on-die L2) [Slot 1, Socket 370]
Dixon	Mobile Pentium II (256KB on-die L2)
Katmai	Pentium III [Slot 1]
Tanner	Pentium III Xeon [Slot 2]
Coppermine	0.18µ PIII w/256KB on-die L2 [Slot 1, Socket 370]
Cascades	Coppermine Xeon (256KB on-die L2) [Slot 2]
Coppermine-128	Celeron III (128KB on-die L2) [Socket 370]
Timna	Celeron III w/integral chipset hub
P68	Former codename for Willamette

Willamette	Pentium IV [Socket 423]
Foster	Pentium IV server [Socket 603]
Gallatin	0.13µ successor to Foster [Socket 603]
Northwood	Mobile Pentium IV
P7	Former codename for Merced
Merced	Itanium (IA64) [Slot M]
McKinley	2nd generation Itanium [Slot M]
Madison	0.13µ McKinley [Slot M]
Deerfield	Low-cost Madison [Slot M]

Note that the codenames and information listed in these tables is not officially released, so by the time many of these future processors come out, names or specifications may change. Most companies who actually get this information from Intel are required to sign non-disclosure agreements which prevents them from sharing this information. This information has been gathered from a number of contacts and sources.

Intel-Compatible Processors (AMD and Cyrix)

Several companies—mainly AMD and Cyrix—have developed processors that are compatible with Intel processors. These chips are fully Intel-compatible, so they emulate every processor instruction in the Intel chips. Many of the chips are pin-compatible, which means that they can be used in any system designed to accept an Intel processor; others require a custom motherboard design. Any hardware or software that works on Intel-based PCs will work on PCs made with these third-party CPU chips. A number of companies currently offer Intel-compatible chips, and I will discuss some of the most popular ones here.

AMD Processors

Advanced Micro Devices (AMD) has become a major player in the Pentium-compatible chip market with its own line of Intel-compatible processors. AMD ran into trouble with Intel several years ago because its 486-clone chips used actual Intel microcode. These differences have been settled and AMD now has a five-year cross-license agreement with Intel. In 1996, AMD finalized a deal to absorb NexGen, another maker of Intel-compatible CPUs. NexGen had been working on a chip it called the Nx686, which was renamed the K6 and introduced by AMD. Since then, AMD has refined the design as the K6-2 and K6-3. Its newest chips, called the Athlon and Duron, are designed similarly to the Pentium II/III and Celeron and use a similar but not identical cartridge or slot design. AMD currently offers a wide variety of CPUs, from 486 upgrades to the K6 series and the Athlon/Duron.

Table 3.20 lists the basic processors offered by AMD and its Intel socket.

Table 3.20 AMD CPU Summary

CPU Type	**P-Rating**	**Actual CPU Speed (MHz)**	**Clock Multiplier**	**Motherboard Speed (MHz)**	**CPU Socket or Slot**
Am486DX4-100	n/a	100	3x	33	Socket 1,2,3
Am486DX4-120	n/a	120	3x	40	Socket 1,2,3
Am5x86-133	75	133	4x	33	Socket 1,2,3
K5	PR75	75	1.5x	50	Socket 5,7
K5	PR90	90	1.5x	60	Socket 5,7
K5	PR100	100	1.5x	66	Socket 5,7
K5	PR120	90	1.5x	60	Socket 5,7
K5	PR133	100	1.5x	66	Socket 5,7
K5	PR166	116.7	1.75x	66	Socket 5,7
K6	PR166	166	2.5x	66	Socket 7
K6	PR200	200	3x	66	Socket 7
K6	PR233	233	3.5x	66	Socket 7
K6	PR266	266	4x	66	Socket 7
K6	PR300	300	4.5x	66	Socket 7
K6-2	PR233	233	3.5x	66	Socket 7
K6-2	PR266	266	4x	66	Socket 7
K6-2	PR300	300	4.5x	66	Socket 7
K6-2	PR300	300	3x	100	Super7
K6-2	PR333	333	5x	66	Socket 7
K6-2	PR333	333	3.5x	95	Super7
K6-2	PR350	350	3.5x	100	Super7
K6-2	PR366	366	5.5x	66	Socket 7
K6-2	PR380	380	4x	95	Super7
K6-2	PR400	400	6x	66	Socket 7
K6-2	PR400	400	4x	100	Super7
K6-2	PR450	450	4.5x	100	Super7
K6-2	PR475	475	5x	95	Super7
K6-2	PR500	500	5x	100	Super7
K6-2	PR533	533	5.5x	97	Super7
K6-2	PR550	550	5.5x	100	Super7
K6-3	PR400	400	4x	100	Super7
K6-3	PR450	450	4.5x	100	Super7

Duron	PR550	550	5.5	100*	Socket A
Duron	PR600	600	6	100*	Socket A
Duron	PR650	650	6.5	100*	Socket A
Duron	PR700	700	7	100*	Socket A
Athlon	PR500	500	5x	100*	Slot A/Socket A
Athlon	PR550	550	5.5x	100*	Slot A/Socket A
Athlon	PR600	600	6x	100*	Slot A/Socket A
Athlon	PR650	650	6.5x	100*	Slot A/Socket A
Athlon	PR700	700	7x	100*	Slot A/Socket A
Athlon A	PR750	750	7.5x	100*	Slot A/Socket A
Athlon	PR800	800	8x	100*	Slot A/Socket A
Athlon	PR850	850	8.5x	100*	Slot A/Socket A
Athlon	PR900	900	9x	100*	Slot A/Socket A
Athlon	PR950	950	9.5x	100*	Slot A/Socket A
Athlon	PR1000	1000	10x	100*	Slot A/Socket A

Notice in the table that for the K5 PR120 through PR166 the model designation does not match the CPU clock speed. This is called a PR rating instead and is further described earlier in this chapter. Starting with the K6, the P-Rating equals the true MHz clock speed.

The model designations are meant to represent performance comparable with an equivalent Pentium-based system. AMD chips, particularly the new K6, have typically fared well in performance comparisons and usually have a much lower cost. There is more information on the respective AMD chips in the sections for each different type of processor.

As you can see from the table, most of AMD's newer K6 series processors are designed to use the Super7 interface it pioneered with Cyrix. Super7 is an extension to the standard Socket 7 design, allowing for increased board speeds of up to 100MHz. The AMD Athlon (K7) processors are designed to use Slot A, which is a 242-pin slot similar in appearance, but not in the pinout, to the Intel Slot 1.

Cyrix

Cyrix has become an even larger player in the market since being purchased by National Semiconductor in November 1997 and by VIA Technologies in 1999. Prior to that it had been a fabless company, meaning it had no chip-manufacturing capability. All the Cyrix chips were manufactured for Cyrix first by Texas Instruments and then mainly by IBM up through the end of 1998. Starting in 1999, National Semiconductor has taken over manufacturing of the Cyrix processors. More recently, National has been purchased by VIA technologies, who still uses National to manufacture the chips.

Like Intel, Cyrix has begun to limit its selection of available CPUs to only the latest technology. Cyrix is currently focusing on the Pentium market with the M1 (6x86) and M2 (6x86MX) processors. The 6x86 has dual internal pipelines and a single, unified 16KB internal cache. It offers speculative and out-of-order instruction execution, much like the Intel Pentium Pro processor. The 6x86MX adds MMX technology to the CPU. The chip is Socket 7 compatible, but some require modified chipsets and new motherboard designs. See Table 3.5 earlier in this chapter, which shows the Cyrix processors.

The 6x86MX features 64KB of unified L1 cache and more than double the performance of the previous 6x86 CPUs. The 6x86MX is offered in clock speeds ranging from 180 to 266MHz, and like the 6x86, it is Socket 7 compatible. When running at speeds of 300MHz and higher, the 686MX was renamed the MII. Besides the higher speeds, all other functions are virtually identical. All Cyrix chips were manufactured by other companies such as IBM, which also markets the 6x86 chips under its own name. National began manufacturing Cyrix processors during 1998, but now that Cyrix is selling them off, the future is unclear.

Note that later versions of the 6x86MX chip have been renamed the MII to deliberately invoke comparisons with the Pentium II, instead of the regular Pentium processor. The MII chips are not redesigned; they are, in fact, the same 6x86MX chips as before, only running at higher clock rates. The first renamed 6x86MX chip is the MII 300, which actually runs at only 233MHz on a 66MHz Socket 7 motherboard. There is also an MII 333, which will run at a 250MHz clock speed on newer 100MHz Super7 motherboards.

Cyrix also has made an attempt at capturing even more of the low-end market than it already has by introducing a processor called the MediaGX. This is a low-performance cross between a 486 and a Pentium combined with a custom motherboard chipset in a two-chip package. These two chips contain everything necessary for a motherboard, except the Super I/O chip, and make very low-cost PCs possible. Expect to see the MediaGX processors on the lowest end, virtually disposable-type PCs. Later versions of these chips will include more multimedia and even network support.

IDT Winchip

Another offering in the chip market is from Integrated Device Technology (IDT). A longtime chip manufacturer that was better known for making SRAM (cache memory) chips, IDT acquired Centaur Technology, which had designed a chip called the C6 Winchip. Now with IDT’s manufacturing capability, the C6 processor became a reality.

Featuring a very simple design, the C6 Winchip is more like a 486 than a Pentium. It does not have the superscalar (multiple instruction pipelines) of a Pentium; it has a single high-speed pipeline instead. Internally, it seems the C6 has little in common with other fifth- and sixth- generation x86 processors. Even so, according to Centaur, it closely matches the performance of a Pentium MMX when running the Winstone 97 business benchmark, although that benchmark does not focus on multimedia performance. It also has a much smaller die (88 mm2) than a typical Pentium, which means it should cost significantly less to manufacture.

The C6 has two large internal caches (32KB each for instructions and data) and will run at 180, 200, 225, and 240MHz. The power consumption is very low—14W maximum at 200MHz for the desktop chip, and 7.1 to 10.6W for the mobile chips. This processor will likely have some success in the low-end market.

P-Ratings

To make it easier to understand processor performance, the P-Rating system was jointly developed by Cyrix, IBM Microelectronics, SGS-Thomson Microelectronics, and Advanced Micro Devices. This new rating, titled the (Performance) P-Rating, equates delivered performance of microprocessor to that of an Intel Pentium. To determine a specific P-Rating, Cyrix and AMD use benchmarks such as Winstone 9x. Winstone 9x is a widely used, industry-standard benchmark that runs a number of Windows-based software applications.

The idea is fine, but in some cases it can be misleading. A single benchmark or even a group of benchmarks cannot tell the whole story on system or processor performance. In most cases, the companies selling PR-rated processors have people believing that they are really running at the speed indicated on the chip. For example, a Cyrix/IBM 6x86MX-PR200 does not really run at 200MHz; instead, it runs at 166MHz. I guess the idea is that it "feels" like 200MHz or compares to some Intel processor running at 200MHz (which one?). I am not in favor of the P-Rating system and prefer to just report the processor's true speed in MHz. If it happens to be 166 but runs faster than most other 166 processors, so be it—but I don't like to number it based on some comparison like that.

> **Note -** See "Cyrix P-Ratings" and "AMD P-Ratings" earlier in this chapter to see how P-Ratings stack up against the actual processor speed in MHz.

The Ziff-Davis Winstone benchmark is used because it is a real-world, application-based benchmark that contains the most popular software applications (based on market share) that run on a Pentium processor. Winstone also is a widely used benchmark and is freely distributed and available.

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 a revolutionary new processor called the 8086 back in June of 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, which meant it could transfer data to memory 16 bits at a time.

The address bus was 20 bits wide, meaning that the 8086 could 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 would be 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 allowed 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, while a similarly configured Apple II cost $1,355.

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

The 8088 in the IBM PC ran at 4.77MHz, or 4,770,000 cycles (essentially computer heartbeats) per second. Each cycle represents a unit of time to the system, with different instructions or operations requiring one or more cycles to complete. 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 1,024KB (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 turned its sights toward producing a more powerful chip with an increased instruction set. The company's first efforts along this line—the 80186 and 80188—were unsuccessful. But incorporating system components into the CPU chip was an important idea for Intel because it led to faster, better chips, such as the 286.

The relationship between the 80186 and 80188 is the same as that of the 8086 and 8088; one is a slightly more advanced version of the other. Compared CPU to CPU, the 80186 is almost the same as the 8088 and has a full 16-bit design. The 80188 (like the 8088) is a hybrid chip that compromises the 16-bit design with an 8-bit external communications interface. The advantage of the 80186 and 80188 is that they combine on a single chip 15 to 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

Intel introduced the 8086 processor in 1976. The math coprocessor that was paired with the chip—the 8087—often was called the 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 1981, is the CPU behind the original IBM AT. Other computer makers manufactured what came to be known as IBM clones, 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. That means that 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 it offered a major performance boost to PCs used in businesses. The processing speed, or throughput, of the original AT (which ran at 6MHz) was five times greater than that of the PC running at 4.77MHz. The die for the 286 is shown in Figure 3.30.

Figure 3.30
286 Processor die.
Photograph used by permission of Intel Corporation.

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 an average 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.

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. The 286 was designed so that if one program locked up or failed, the entire system didn't need a warm boot (reset) or cold boot (power off, then back on). Theoretically, what happened in one area of memory didn't affect other programs. Before multitasked programs could be "safe" from one another, however, the 286 chip (and subsequent chips) needed an operating system that worked cooperatively with the chip to provide such protection.

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 were identical.

In most systems, the 80286 internally divides the system clock by two to derive the processor clock. The 80287 internally divides the system-clock frequency by three. 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.

In summary, the 80287 and the 8087 chips perform about the same at equal clock rates. The original 80287 is not better than the 8087 in any real way—unlike the 80286, which is superior to the 8086 and 8088. In most AT systems, the performance gain that you realize by adding the coprocessor is much less substantial than the same type of upgrade for PC- or XT-type systems or for the 80386.

286 Processor Problems

After you remove a math coprocessor from an AT-type system, you must rerun your computer's Setup program. Some AT-compatible Setup programs do not properly unset the math coprocessor bit. If you receive a POST error message because the computer cannot find the math chip, you might have to unplug the battery from the system board temporarily. All Setup information will be lost, so be sure to write down the hard drive type, floppy drive type, and memory and video configurations before unplugging the battery. This information is critical in reconfiguring your computer correctly.

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 (normally 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 chip rose in popularity for several years, which peaked around 1991. Obsolete 386 processor systems are mostly retired or scrapped, having been passed down the user chain. If they are in operating condition, they can be useful for running old DOS or Windows 3.x–based applications, which they can do quite nicely.

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, which means that 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. Many 286 system manufacturers were touting their 16MHz and 20MHz 286 systems as being just as fast as 16MHz and 20MHz 386 systems, and they were right! 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 3.31.

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 has 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 OS/2 and Windows NT, 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 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. Software commands could reboot the blown partition.

Numerous variations of the 386 chip exist, 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.

Figure 3.31
386 processor die.
Photograph used by permission of Intel Corporation.

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 VLSI (very large scale integration) 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 CMOS (complementary metal oxide semiconductor) materials. The CMOS design enables devices to consume extremely low levels of power.

The Intel 386 chip was available in clock speeds ranging from 16–33MHz; 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 1,000GB.)

386SX Processors

The 386SX was designed for systems designers who were looking for 386 capabilities at 286 system prices. Like 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 that the 386DX can address. Before it was discontinued, the 386SX was available in clock speeds ranging from 16 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.

Note - One common fallacy about the 386SX is that you can plug one into a 286 system and give the system 386 capabilities. This is not true; the 386SX chip is not pin-compatible with the 286 and does not plug into the same socket. Several upgrade products, however, have been designed to adapt the chip to a 286 system. In terms of raw speed, converting a 286 system to a 386 CPU chip results in little performance gain—

286 motherboards are built with a restricted 16-bit interface to memory and peripherals. A 16MHz 386SX is not markedly faster than a 16MHz 286, but it does offer improved memory management capabilities on a motherboard designed for it, and the capability to run 386-specific software.

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 needed. 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 16–64KB 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 runs 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 designed specifically 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 processors.

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 percent.

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 union, however, leave 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.

Weitek Coprocessors

In 1981, several Intel engineers formed the Weitek Corporation. Weitek developed math coprocessors for several systems, including those based on Motorola processor designs. Intel originally contracted Weitek to develop a math coprocessor for the Intel 386 CPU, because Intel was behind in its own development of the 387 math coprocessor. The result was the Weitek 1167, a custom math coprocessor that uses a proprietary Weitek instruction set, which is incompatible with the Intel 387.

To use the Weitek coprocessor, your system must have the required additional socket, which was different from the standard Intel coprocessor sockets.

80386 Bugs

Some early 16MHz Intel 386DX processors had a small bug that appeared as a software problem. The bug, which apparently was in the chip's 32-bit multiply routine, manifested itself only when you ran true 32-bit code in a program such as OS/2 2.x, UNIX/386, or Windows in enhanced mode. Some specialized 386 memory-management software systems also may invoke this subtle bug, but 16-bit operating systems (such as DOS and OS/2 1.x) probably will not.

The bug usually causes the system to lock up. Diagnosing this problem can be difficult because the problem generally is intermittent and software-related. Running tests to find the bug is difficult; only Intel, with proper test equipment, can determine whether your chip has a bug. Some programs can diagnose the problem and identify a defective chip, but they cannot identify all defective chips. If a program indicates a bad chip, you certainly have a defective one; if the program passes the chip, you still might have a defective one.

Intel requested that its 386 customers return possibly defective chips for screening, but many vendors did not return them. Intel tested returned chips and replaced defective ones. The defective chips later were sold to bargain liquidators or systems houses that wanted chips that would not run 32-bit code. The defective chips were stamped with a 16-bit SW Only logo, indicating that they were authorized to run only 16-bit software.

Chips that passed the test, and all subsequent chips produced as bug-free, were marked with a double-sigma code (SS). 386DX chips that are not marked with either of these designations have not been tested by Intel and might be defective.

The following marking indicates that a chip has not yet been screened for the defect; it might be either good or bad.

80386-16

The following marking indicates that the chip has been tested and has the 32-bit multiply bug. The chip works with 16-bit software (such as DOS) but not with 32-bit, 386-specific software (such as Windows or OS/2).

80386-16

16-bit SW Only

The following mark on a chip indicates that it has been tested as defect-free. This chip fulfills all the capabilities promised for the 80386.

80386-16

SS

This problem was discovered and corrected before Intel officially added DX to the part number. So, if you have a chip labeled as 80386DX or 386DX, it does not have this problem.

Another problem with the 386DX can be stated more specifically. When 386-based versions of XENIX or other UNIX implementations are run on a computer that contains a 387DX math coprocessor, the computer locks up under certain conditions. The problem does not occur in the DOS environment, however. For the lockup to occur, all the following conditions must be in effect:

- Demand page virtual memory must be active.
- A 387DX must be installed and in use.
- DMA (direct memory access) must occur.
- The 386 must be in a wait state.

When all these conditions are true at the same instant, the 386DX ends up waiting for the 387DX and vice versa. Both processors will continue to wait for each other indefinitely. The problem is in certain versions of the 386DX, not in the 387DX math coprocessor.

Intel published this problem (Errata 21) immediately after it was discovered to inform its OEM customers. At that point, it became the responsibility of each manufacturer to implement a fix in its hardware or software product. Some manufacturers, such as Compaq and IBM, responded by modifying their motherboards to prevent these lockups from occurring.

The Errata 21 problem occurs only in the B-stepping version of the 386DX and not in the later D-stepping version. You can identify the D-stepping version of the 386DX by the letters DX in the part number (for example, 386DX-20). If DX is part of the chip’s part number, the chip does not have this problem.

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 percent, which describes how often zero-wait-state read operations will 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, which means that 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.

Before the 486, many people avoided GUIs because they didn't have time to sit around waiting for the hourglass, which indicates that the system is performing behind-the-scenes operations that the user cannot interrupt. The 486 changed that situation. Many people believe that the 486 CPU chip spawned the widespread acceptance of GUIs.

With the release of its faster Pentium CPU chip, Intel began to cut the price of the 486 line to entice the industry to shift over to the 486 as the mainstream system. Intel later did the same thing with its Pentium chips, spelling the end of the 486 line. The 486 is now offered by Intel only for use in embedded microprocessor applications, used primarily in expansion cards.

Most of the 486 chips were offered in a variety of maximum speed ratings, varying from 16MHz up to 120MHz. 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 newer 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 3.21 shows the different speed combinations that can result from using the DX2 or DX4 processors with different motherboard clock speeds.

Table 3.21 Intel DX2 and DX4 Operating Speeds Versus Motherboard Clock Speeds

Processor speed	**DX2 (2 x mode) Speed**	**DX4 (2.5x mode) Speed**	**DX4 (3x mode) Speed**	**DX4**
16MHz Motherboard	32MHz	32MHz	40MHz	48MHz
40MHz Motherboard	80MHz	80MHz	100MHz	120MHz
20MHz Motherboard	40MHz	40MHz	50MHz	60MHz
50MHz Motherboard	n/a	100MHz	n/a	n/a
25MHz Motherboard	50MHz	50MHz	63MHz	75MHz
33MHz Motherboard	66MHz	66MHz	83MHz	100MHz

The internal core speed 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). The CLKMUL input is sampled only during a reset of the CPU and defines the ratio of the internal clock to the external bus frequency CLK signal at pin

C-3 (Socket 1) or D-4 (Socket 2, 3, or 6). If CLKMUL is sampled low, the internal core speed will be two times the external bus frequency. If driven high or left floating (most motherboards would leave it floating), triple speed mode is selected. If the CLKMUL signal is connected to the BREQ (Bus Request) output signal at pin Q-15 (Socket 1) or R-16 (Socket 2, 3, or 6), the CPU internal core speed will be two and a half times the CLK speed. To summarize, here is how the socket has to be wired for each DX4 speed selection:

CPU Speed	**CLKMUL (Sampled Only at CPU Reset)**
2x	Low
2.5x	Connected to BREQ
3x	High or Floating

You will have to determine how your particular motherboard is wired and whether it can be changed to alter the CPU core speed in relation to the CLK signal. In most cases, there would be one or two jumpers on the board near the processor socket. 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 would give 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.

Note - One caveat is that if your motherboard has VL-Bus slots, they will have to be slowed down to 33 or 40MHz to operate properly.

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 will not run properly at 50MHz. Consult your motherboard—or even better, your chipset documentation—to see how your board is set up.

Caution - If you are upgrading an existing system, be sure that your socket will support the chip that you are installing. In particular, if you are putting a DX4 processor in an older system, you need some type of adapter to regulate the voltage down to 3.3v. If you put the DX4 in a 5v socket, you will destroy the chip! See the earlier section on processor sockets for more information.

The 486-processor family is designed for greater performance than previous processors because it integrates formerly external devices, such as cache controllers, cache memory, and math coprocessors. Also, 486 systems were the first designed for true processor upgradability. Most 486 systems can be upgraded by simple processor additions or swaps that can effectively double the speed of the system.

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 now is also available in 5v, 196-pin PQFP (Plastic Quad Flat Pack), and 3.3v, 208-pin SQFP (Small Quad Flat Pack). These latter form factors are available in SL Enhanced versions, which are intended primarily for portable or laptop applications in which saving power is important.

Two main features separate the 486 processor from older processors:

- The 486DX integrates functions such as the math coprocessor, cache controller, and cache memory into the chip.

- The 486 also was designed with upgradability in mind; double-speed OverDrive are upgrades available for most systems.

The 486DX processor is fabricated with low-power CMOS (complementary metal oxide semiconductor) 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 inFigure 3.32.

The standard 486DX contains a processing unit, a floating-point unit (math coprocessor), a memory-management unit, and a 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 two 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).

Like 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.

- *In real mode*, the 486 (like the 386) runs unmodified 8086-type software.

- *In protected mode*, the 486 (like the 386) offers sophisticated memory paging and program switching.

- *In virtual real mode*, the 486 (like 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.

Figure 3.32
486 processor die.
Photograph used by permission of Intel Corporation.

The 486DX series has a built-in math coprocessor that sometimes is called an MCP (math coprocessor) or FPU (floating-point unit). 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 percent 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 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 reexecutes 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 one second from the suspend state (instant off) to exactly where it left off. You do not need to reboot, load the operating system, load the application program, and then load the application data. 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 is 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 chip is twice as fast as a 386DX with the same clock speed. Intel marketed the 486SX as being the ideal chip for new computer buyers, because fewer entry-level programs of that day used math-coprocessor functions.

The 486SX was normally available in 16, 20, 25, and 33MHz-rated speeds, and there was also a 486 SX/2 that ran at up to 50 or 66MHz. The 486SX normally comes in a 168-pin version, although other surface-mount versions are available 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 is a separate math coprocessor chip 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. If this situation sounds confusing, read on, because this topic brings you to the most important aspect of 486 design: upgradability.

487SX

The 487SX math coprocessor, as Intel calls 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.

Although in most cases you can upgrade a system by removing the 486SX CPU and replacing it with a 487SX (or even a DX or DX2/OverDrive), Intel originally discouraged this procedure. 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.) Now Intel recommends—or even insists on—the use of a single processor socket of a ZIF design, which makes upgrading an easy task physically.

Very few early 486 systems had a socket for the Weitek 4167 coprocessor chip for 486 systems that existed in November 1989.

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 different 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 will function 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 1.1 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. This means that 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 may 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 has to 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 percent, however, the CPU has to access system memory only 5–10 percent 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 only at 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 that the processor will have to 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 will still enjoy a near-doubling of CPU performance, but operations that involve a great deal of memory access will be slower.

This brings us to the DX4 processor. 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 may 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 heat sink 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 will run only in the tripled speed mode, and not the 2x or 2.5x modes of the standard DX4 processor.

> **Note -** As of this writing, Intel has discontinued all 486 and DX2/DX4/OverDrive processors, including the so-called Pentium OverDrive processor.

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 meant that it never saw the light of day. One thing to keep in mind about the 486 Pentium OverDrive chip is that although it is intended primarily for SX2 and DX2 systems, it should work in any upgradable 486SX or DX system that has a Socket 2 or Socket 3. If in doubt, check Intel’s online upgrade guide for compatibility.

The Pentium OverDrive processor is designed for systems that have a processor socket that follows the Intel Socket 2 specification. This processor also will work 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 will be realized. Unfortunately, most motherboards, especially older ones with the Socket 2 design, only support 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. Unfortunately, these are the only solutions still offered by Intel for upgrading the 486. 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. At the time, I did not recommend the 486 Pentium OverDrive chips as a viable solution for the future.

"Vacancy"—Secondary OverDrive Sockets

Perhaps you saw the Intel advertisements—both print and television—that featured a 486SX system with a neon Vacancy sign pointing to an empty socket next to the CPU chip. Unfortunately, these ads were not very informative, and they made it seem that only systems with the extra socket could be upgraded. I was worried when I first saw these ads because I had just purchased a 486DX system, and the advertisements implied that only 486SX systems with the empty OverDrive socket were upgradable. This, of course, was not true, but the Intel advertisements did not communicate that fact very well.

I later found that upgradability does not depend on having an extra OverDrive socket in the system and that virtually any 486SX or DX system can be upgraded. The secondary OverDrive socket was designed to make upgrading easier and more convenient. Even in systems that have the second socket, you can actually remove the primary SX or DX CPU and plug the OverDrive processor directly into the main CPU socket, rather than into the secondary OverDrive socket.

In that case, you would have an upgraded system with a single functioning CPU installed; you could remove the old CPU from the system and sell it or trade it in for a refund. Unfortunately, Intel does not offer a trade-in or core-charge policy; it does not want your old chip. For this reason, some people saw the OverDrive socket as being a way for Intel to sell more CPUs. Some valid reasons exist, however, to use the OverDrive socket and leave the original CPU installed.

One reason is that many PC manufacturers void the system warranty if the CPU has been removed from the system. Also, most manufacturers require that the system be returned with only the original parts when systems are serviced; you must remove all add-in cards, memory modules, upgrade chips, and similar items before sending the system in for servicing. If you replace the original CPU when you install the upgrade, returning the system to its original condition will be much more difficult.

Another reason for using the upgrade socket is that the system will not function if the main CPU socket is damaged when you remove the original CPU or install the upgrade processor. By contrast, if a secondary upgrade socket is damaged, the system still should work with the original CPU.

80487 Upgrade

The Intel 80486 processor was introduced in late 1989, and systems using this chip appeared during 1990. The 486DX integrated the math coprocessor into the chip.

The 486SX began life as a full-fledged 486DX chip, but Intel actually disabled the built-in math coprocessor before shipping the chip. As part of this marketing scheme, Intel marketed what it called a 487SX math coprocessor. Motherboard manufacturers installed an Intel-designed socket for this so-called 487 chip. In reality, however, the 487SX math chip was a special 486DX chip with the math coprocessor enabled. When you plugged this chip into your motherboard, it disabled the 486SX chip and gave you the functional equivalent of a full-fledged 486DX system.

AMD 486 (5x86)

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

The 5x85 offers high-performance features such as a unified 16KB write-back cache and 133MHz core clock speed; it is approximately comparable to a Pentium 75, which is why it is denoted with a P-75 in the part number. It is the ideal choice for cost-effective 486 upgrades, where changing the motherboard is difficult or impossible.

Not all 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 Web site.

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 may 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 3.45 volt processor will not operate in a 5-volt socket and may be damaged. To convert a 5-volt motherboard to 3.45 volts, adapters can be purchased from several vendors including Kingston, Evergreen, and AMP. In fact, Kingston and Evergreen sell 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.

- It is generally better to purchase a new motherboard with Socket 3 than to buy one of these adapters; however, 486 motherboards are hard to find these days, and your old board may be in a proprietary form factor for which it is impossible to find a replacement. Buying a new motherboard is also better than using an adapter because the older BIOS may not understand the requirements of the processor as far as speed is concerned. BIOS updates are often required with older boards.

- Most Socket 3 motherboards have jumpers, allowing 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 (± 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 there are problems running applications or the floppy drive doesn't work right (DMA conflicts).

- The 5x86 runs hot, so a heat sink is required; it normally must have a fan.

In addition to the 5x86, the AMD-enhanced 486 product line includes 80MHz; 100MHz; and 1,20MHz CPUs. These are the A80486DX2-80SV8B (40MHzx2), A80486DX4-100SV8B (33MHzx3), and the A80486DX4-120SV8B (40MHzx3).

Cyrix/TI 486

The Cyrix 486DX2/DX4 processors were available in 100MHz, 80MHz, 75MHz, 66MHz, and 50MHz versions. Like 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. Eventually, TI and Cyrix had a falling out, and then IBM made most of the Cyrix chips for a while. That changed in 1999 when first National Semiconductor bought Cyrix, and took over production from IBM. Then National sold Cyrix to VIA Technologies, although VIA still uses National to manufacture the chips.

P5 (586) Fifth-Generation Processors

After the fourth-generation chips like the 486, Intel and other chip manufacturers went back to the drawing board to come up with new architectures and features incorporated 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 (code-named P5) would be named the Pentium processor rather than the 586, as everybody had been assuming. 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 that they might develop. The actual Pentium chip shipped on March 22, 1993. Systems that use these chips were only a few months behind.

The Pentium is fully compatible with previous Intel processors, but it also 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.

The standard 486 chip can execute a single instruction in an average of two clock cycles—cut to an average of one clock cycle with the advent of internal clock multiplication used in the DX2 and DX4 processors. 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 (Reduced Instruction Set Computer) chips. The Pentium is one of the first CISC (Complex Instruction Set Computer) chips to be considered superscalar. The Pentium is almost like having two 486 chips under the hood. Table 3.22 shows the Pentium processor specifications.

Table 3.22 Pentium Processor Specifications

Introduced	March 22, 1993 (first generation); March 7, 1994 (second generation)
Maximum rated speeds	60, 66, 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
Integral-cache type	Two-way set associative, write-back Data
Burst-mode transfers	Yes
Number of transistors	3.1 million
Circuit size	0.8 micron (60/66MHz), 0.6 micron (75–100MHz), 0.35 micron (120MHz and up)
External package	273-pin PGA, 296-pin SPGA, tape carrier
Math coprocessor	Built-in FPU (floating-point unit)
Power management	SMM (system management mode), 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 allow 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 3.33 shows the internal architecture of the Pentium processor.

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 that 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, which means that each bank of memory is 64 bits.

On most motherboards, memory is installed via SIMMs (Single Inline Memory Modules) or DIMMs (Dual Inline Memory Modules). SIMMs are available in 8-bit-wide and 32-bit-wide versions, while DIMMs are 64 bits wide. There are also versions with additional bits for parity or ECC (error correcting code) 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. The newest Pentium systems and most Pentium II systems 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.

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.

Figure 3.33
Pentium processor internal architecture.

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) Static RAM (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 BiCMOS (bipolar complementary metal oxide semiconductor) process and superscalar architecture to achieve the high level of performance expected from the chip. BiCMOS adds about 10 percent to the complexity of the chip design, but adds about 30–35 percent better performance without a size or power penalty.

All Pentium processors are SL enhanced, meaning that 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 allows 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 current 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. There is also a special 3.3v setting called VR (Voltage Reduced), which 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 are intended for mobile computer applications. They do not use a conventional chip package and are 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 are 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 Chapter 23, "Portable PCs."

The Pentium, like the 486, contains an internal math coprocessor or FPU. The FPU in the Pentium has been rewritten and performs significantly better than the FPU in the 486, yet it is fully compatible with the 486 and 387 math coprocessor. The Pentium FPU is estimated at two to as much as 10 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 Processor

The Pentium has been offered in three basic designs, each with several versions. The first-generation design, which is no longer available, came in 60 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 is used.

The first-generation Pentium was created through an 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 would solve 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 some cases.

If you do have one of these first-generation Pentiums, do not despair. As with previous 486 systems, Intel offers OverDrive upgrade chips that effectively double the processor speed of your Pentium 60 or 66. These are a single-chip upgrade, meaning they replace your existing CPU. Because subsequent Pentiums are incompatible with the Pentium 60/66 Socket 4 arrangement, these OverDrive chips were the only way to upgrade an existing first-generation Pentium without replacing the motherboard.

Rather than upgrading the processor with one only twice as fast, you should really consider a complete motherboard replacement, which would accept a newer design processor that would potentially be many times faster.

Second-Generation Pentium Processor

Intel announced the second-generation Pentium on March 7, 1994. This new processor was introduced in 90 and 100MHz versions, with a 75MHz version not far behind. Eventually, 120, 133, 150, 166, 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 3.34. Additionally, these new processors run on 3.3v power. The 100MHz version consumes a maximum 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).

The second-generation Pentium processors come in 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 is 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 interface 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 Multi-Processing (SMP) is being integrated into operating systems such as Windows NT and OS/2.

Figure 3.34
Pentium processor die. Photograph used by permission of Intel Corporation.

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.

Note - Some Pentium systems support 75MHz or even up to 100MHz with newer motherboard and chipset designs.

Virtually all Pentium motherboards have three speed settings: 50, 60, and 66MHz. Pentium chips are available with a variety of internal clock multipliers that cause the processor to operate at various multiples of these motherboard speeds. Table 3.23 lists the speeds of currently available Pentium processors and motherboards.

Table 3.23 Pentium CPU and Motherboard Speeds

CPU Type/Speed	CPU Clock	Motherboard Speed (MHz)
Pentium 75	1.5x	50
Pentium 90	1.5x	60
Pentium 100	1.5x	66
Pentium 120	2x	60
Pentium 133	2x	66
Pentium 150	2.5x	60
Pentium 166	2.5x	66
Pentium 200	3x	66
Pentium 233	3.5x	66
Pentium 266	4x	66

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 3.24 shows how the state of the BFx pins will affect the clock multiplication in the Pentium processor.

Table 3.24 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 233 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 will operate only at specific combinations of these settings or may even be fixed at one particular setting. Many of the newer motherboards have jumpers or switches that allow 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 "Processor Speed Ratings" section of 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.

A single-chip OverDrive upgrade is currently offered 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, as well. Simply stated, this means that a Pentium 100, 133, or 166 system equipped with the OverDrive chip will have a processor speed of 200MHz. Perhaps the best feature of these Pentium OverDrive chips is that they incorporate MMX technology. MMX provides greatly enhanced performance while running the multimedia applications that are so popular today.

If you have a Socket 7 motherboard, you might not need the special OverDrive versions of the Pentium processor that have built-in voltage regulators. Instead, you can purchase a standard Pentium or Pentium-compatible chip and replace the existing processor with it. You will have to be sure to set the multiplier and voltage settings so that they are correct for the new processor.

Pentium-MMX Processors

A third generation of Pentium processors (code-named P55C) was released in January 1997, which incorporates what Intel calls MMX technology into the second-generation Pentium design (see Figure 3.35). These Pentium-MMX processors are available in clock rates of 66/166MHz, 66/200MHz, and 66/233MHz, and 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 1.8 volts. With this newer technology, the 266 processor actually uses less power than the non-MMX 133.

Figure 3.35
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 allow a more universal motherboard solution with respect to these changing voltages, Intel has come up with 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, it is easy to reconfigure a motherboard to support any of the voltages required by the newer Pentium processors.

Of course, lower voltage is nice, but MMX is what this chip is really all about. MMX technology was developed by Intel as a direct response to the growing importance and increasing demands of multimedia and communication applications. Many such applications run repetitive loops of instructions that take vast amounts of time to execute. As a result, MMX incorporates a process Intel calls Single Instruction Multiple Data (SIMD), which allows one instruction to perform the same function on many pieces of data. Furthermore, 57 new instructions that are designed specifically to handle video, audio, and graphics data have been added to the chip.

If you want maximum future upgradability to the MMX Pentiums, make sure that your Pentium motherboard includes 321-pin processor sockets that fully meet the Intel Socket 7 specification. These would also include the VRM (Voltage Regulator Module) socket. If you have dual sockets, you can add a second Pentium processor to take advantage of SMP (Symmetric Multiprocessing) support in some newer operating systems.

Also make sure that any Pentium motherboard you buy can be jumpered or reconfigured for both 60 and 66MHz operation. This will enable you to take advantage of future Pentium OverDrive processors that will support the higher motherboard clock speeds. These simple recommendations will enable you to perform several dramatic upgrades without changing the entire motherboard.

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. There are tables listing all the different variations of Pentium processors and steppings and how to identify them later in this chapter.

This bug caused a tremendous fervor when it first was reported on the Internet by a mathematician 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 would incorrectly perform 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 that 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 later 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. This means that 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. Normally, all you have to do is call Intel and ask for the replacement. It will ship you a new part matching the ratings of the one you are replacing in an overnight shipping box. The replacement is free, including all shipping charges. You merely remove your old processor, replace it with the new one, and put the old one back in the box. Then you call the overnight service who will pick it up and send it back. Intel will take a credit card number when you first call for the replacement only to ensure that the original defective chip is returned. As long as it gets the original CPU back within a specified amount of time, there will be no charges to you. Intel has indicated that these defective processors will be 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 if 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), in the Microsoft Windows built-in calculator, or in any other calculating program that uses the FPU. Make sure that for the purposes of this test the FPU has not been disabled. That would normally require some special command or setting specific to the application, and would, of course, ensure that the test came out correct, no matter whether the chip is flawed or not.

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 (correct answer)

962,306,957,033 / 11,010,046 = 87,399.5805831329 (flawed Pentium)

Note - Note that your particular calculator program may 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 over 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.

Here is another division problem that will come out incorrectly on a Pentium with this flaw:

> 4,195,835 / 3,145,727 = 1.33382044913624100 (correct answer)
>
> 4,195,835 / 3,145,727 = 1.33373906890203759 (flawed Pentium)

This one shows an error in the fifth significant digit. A variation on the previous calculation can be performed as follows:

> x = 4,195,835
>
> y = 3,145,727
>
> z = x – (x/y)*y
>
> 4,195,835 – (4,195,835 / 3,145,727) x 3,145,727 = 0 (correct answer)
>
> 4,195,835 – (4,195,835 / 3,145,727) x 3,145,727 = 256 (flawed Pentium)

With an exact computation, the answer here should be zero. In fact, you will get zero on most machines, including those using Intel 286, 386, and 486 chips. But, on the Pentium, the answer is 256!

Here is one more calculation you can try:

> 5,505,001 / 294,911 = 18.66665197 (correct answer)
>
> 5,505,001 / 294,911 = 18.66600093 (flawed Pentium)

This one represents an error in the sixth significant digit.

There are several workarounds 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!

Power Management Bugs

Starting with the second-generation Pentium processors, Intel added functions that allow 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 will 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 may go through half a dozen or more such steppings.

See the previous editions of this book on the included CD-ROM for tables showing the Pentium processor steppings and revisions. This information is also available online from Intel via its Web site.

To determine the specifications of a given processor, you need to look up the S-spec number in the table of processor specifications. To find your S-spec number, you have to read it off of the chip directly. It can be found printed on both the top and bottom of the chip. If your heat sink is glued on, remove the chip and heat sink 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 that Intel publishes (via its Web site); it will tell you the specifications of that particular processor. Intel is introducing new chips all the time, so visit its Web site 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 there are several subtly different voltages required by different Pentium processors. Table 3.25 summarizes the different processors and their required voltages.

Table 3.25 Pentium Processor Voltages

Model	Stepping	Voltage Spec.	Voltage Range
1	—	Std.	4.75–5.25v
1	—	5v1	4.90–5.25v
1	—	5v2	4.90–5.40v
1	—	5v3	5.15–5.40v
2+	B1-B5	Std.	3.135–3.465v
2+	C2+	Std.	3.135–3.600v
2+	—	VR	3.300–3.465v
2+	B1-B5	VRE	3.45–3.60v
2+	C2+	VRE	3.40–3.60v
4+	—	MMX	2.70–2.90v
4	3	Mobile	2.285–2.665v
4	3	Mobile	2.10–2.34v
8	1	Mobile	1.850–2.150v
8	1	Mobile	1.665–1.935v

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 may 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. I would definitely want 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, and PR-166. 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. AMD keeps a list of motherboards that have been tested for compatibility.

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 (FPU)
- Pin-selectable clock multiples of 1.5x 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 several of these processors do not run at their apparent rated speed. For example, the PR-166 version actually runs at only 117 true MHz. Sometimes this can confuse the system BIOS, which may report the true speed rather than the P-Rating, which compares the chip against an Intel Pentium of that speed. AMD claims 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 markets the K5 as a fifth-generation processor, just like the Pentium.

The AMD-K5 operates at 3.52 volts (VRE Setting). Some older motherboards default to 3.3 volts, which is below specification for the K5 and could cause erratic operation.

Pseudo Fifth-Generation Processors

There is at least one processor that, while generally regarded as a fifth-generation processor, lacks many of the functions of that class of chip—the IDT Centaur C6 Winchip. True fifth-generation chips would have multiple internal pipelines, which is called superscalar architecture, allowing more than one instruction to be processed at one time. They would also feature branch prediction, another fifth-generation chip feature. As it lacks these features, the C6 is more closely related to a 486; however, the performance levels and the pinout put it firmly in the class with Pentium processors. It has turned out to be an ideal Pentium Socket 7–compatible processor for low-end systems.

IDT Centaur C6 Winchip

The C6 processor is a recent offering from Centaur, a wholly owned subsidiary of IDT (Integrated Device Technologies). It is Socket 7–compatible with Intel's Pentium, includes MMX extensions, and is available at clock speeds of 180, 200, 225, and 240MHz. Pricing is below Intel on the Pentium MMX.

Centaur is led by Glenn Henry, who spent more than two decades as a computer architect at IBM and six years as chief technology officer at Dell Computer Corp. The company is a well-established semiconductor manufacturer well-known for SRAM and other components.

As a manufacturer, IDT owns its own fabs (semiconductor manufacturing plants), which will help keep costs low on the C6 Winchip. Its expertise in SRAM manufacturing may be applied in new versions of the C6, which integrate onboard L2 cache in the same package as the core processor, similar to the Pentium Pro.

The C6 has 32KB each of instruction and data cache, just like AMD’s K6 and Cyrix’s 6x86MX, yet it has only 5.4 million transistors, compared with the AMD chip’s 8.8 million and the Cyrix chip’s 6.5 million. This allows for a very small processor die, which also reduces power consumption. Centaur achieved this small size with a streamlined design. Unlike competitor chips, the C6 is not superscalar—it issues only one instruction per clock cycle like the 486. However, with large caches, an efficient memory-management unit, and careful performance optimization of commonly used instructions, the C6 achieves performance that's comparable to a Pentium. Another benefit of the C6's simple design is low power consumption—low enough for notebook PCs. Neither AMD nor Cyrix has a processor with power consumption low enough for most laptop designs.

To keep the design simple, Centaur compromised on floating-point and MMX speed and focused instead on typical application performance. As a result, the chip's performance trails the other competitors' on some multimedia applications and games.

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, many other P6 chips have been released by Intel, all using the same basic P6 core processor as the Pentium Pro. Table 3.26 shows the variations in the P6 family of processors.

Table 3.26 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	P6 with 256KB of full-core speed L2

IIPE	cache
Pentium III Xeon	P6 with SSE (MMX2), 512KB, 1MB, or 2MB of full-core speed L2 cache

Even more are expected in this family, including versions of the Pentium III with on-die full-core speed L2 cache, and faster versions of the Celeron.

The main new feature in the fifth-generation Pentium processors was the superscalar architecture, where 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 enables the processor to execute more instructions on parallel, so that tasks are completed more quickly. This technology innovation is comprised of three main elements:

- *Multiple branch prediction*, to predict the flow of the program through several branches
- *Dataflow analysis*, which schedules instructions to be executed when ready, independent of their order in the original program
- *Speculative execution*, which increases the rate of execution by looking ahead of the program counter and executing instructions that are likely to be needed

Branch prediction is a feature formerly found only in high-end mainframe processors. It allows 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 like 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 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 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 allows the P6 to run existing Intel architecture software exactly as the P5 (Pentium) and previous processors did, just a whole lot more quickly!

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 allows the cache memory to run at speeds previously not possible.

Previous P5 generation processors have only a single motherboard host processor bus, and all data, including cache transfers, must flow through it. The main problem with that is the cache memory was restricted to running at motherboard bus speed, which was 66MHz until recently and has now moved to 100MHz. We have cache memory today that can run 500MHz or more, and main memory (SDRAM) that runs at 66 and 100MHz, so a method was needed to get faster memory closer to the processor. The solution was to essentially build in what is called a backside bus to the processor, otherwise known as a dedicated cache bus. The L2 cache would then be connected to this bus and could run at any speed. The first implementation of this was in the Pentium Pro, where the L2 cache was built right into the processor package and ran at the full-core processor speed. Later, that proved to be too costly, so the L2 cache was moved outside of the processor package and onto a cartridge module, which we now know as the Pentium II/III. With that design, the cache bus could run at any speed, with the first units running the cache at half-processor speed.

By having the cache on a backside bus directly connected to the processor, the speed of the cache is scalable to the processor. In current PC architecture—66MHz Pentiums all the way through the 333MHz Pentium IIs—the motherboard runs at a speed of 66MHz. Newer Pentium II systems run a 100MHz motherboard bus and have clock speeds of 350MHz and higher. If the cache were restricted to the motherboard as is the case with Socket 7 (P5 processor) designs, the cache memory would have to remain at 66MHz, even though the processor was running as fast as 333MHz. With newer boards, the cache would be stuck at 100MHz, while the processor ran as fast as 500MHz or more. With the Dual Independent Bus (DIB) design in the P6 processors, as the processor runs faster, at higher multiples of the motherboard speed, the cache would increase by the same amount that the processor speed increases. The cache on the DIB is coupled to processor speed, so that doubling the speed of the processor also doubles the speed of the cache.

The DIB architecture is necessary to have decent processor performance in the 300MHz and beyond range. Older Socket 7 (P5 processor) designs will not be capable of moving up to these higher speeds without suffering a tremendous performance penalty due to the slow motherboard-bound L2 cache. That is why Intel is not developing any Pentium (P5 class) processors beyond 266MHz; however, the P6 processors will be available in speeds of up to 500MHz or more.

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 (Complex Instruction Set Computer) instructions are broken down into more RISC (Reduced Instruction Set Computer) 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, while 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, while the external instruction interface is classic Intel CISC. By breaking down the CISC instructions into several different 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 operate part time in a 16-bit environment) and most older applications, the P6 will not provide as marked a performance improvement over similarly speed-rated Pentium and Pentium-MMX processors. That's because the Dynamic Execution capability will not be fully exploited. Because of this, Windows NT is often regarded as the most desirable operating system for use with Pentium Pro/II/III/Celeron processors. While this is not exactly true (a Pentium Pro/II/III/Celeron will run fine under Windows 95/98), Windows NT 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 new chip is unique among processors as it is constructed in a Multi-Chip Module (MCM) physical format, which Intel is calling a Dual Cavity PGA (Pin Grid Array) package. Inside the 387-pin chip carrier are two dies. One contains the actual Pentium Pro processor (shown in Figure 3.36), and the other a 256KB (the Pentium Pro with 256KB cache is shown in Figure 3.37), 512KB, or 1MB (the Pentium Pro with 1MB cache is shown in Figure 3.37) 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 3.38).

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 Dual Independent Bus (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.

Figure 3.36
Pentium Pro processor die.
Photograph used by permission of Intel Corporation.

Figure 3.37
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 3.38
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.

Table 3.27 shows Pentium Pro processor specifications. Table 3.28 shows the specifications for each model within the Pentium Pro family, as there are many variations from model to model.

Table 3.27 Pentium Pro Family Processor Specifications

Introduced	November 1995
Maximum rated speeds	150, 166, 180, 200MHz
CPU	2.5x, 3x
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	2.46 (6.25cm) x 2.66

dimensions	(6.76cm)
Math coprocessor	Built-in FPU
Power management	SMM (system management mode)
Operating voltage	3.1v or 3.3v

Table 3.28 Pentium Pro Processor Specifications by Processor Mode

Pentium Pro Processor (200MHz) with 1MB Integrated Level 2 Cache	
Introduction date	August 18, 1997
Clock speeds	200MHz (66MHz x 3)
Number of transistors	5.5 million (0.35 micron process), plus 62 million in 1MB L2 cache (0.35 micron)
Cache Memory	8Kx2 (16KB) L1, 1MB core-speed L2
Die size	0.552 (14.0mm)
Pentium Pro Processor (200MHz)	
Introduction date	November 1, 1995
Clock speeds	200MHz (66MHz x 3)
iCOMP Index 2.0 rating	220
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	8Kx2 (16KB) L1, 256KB or 512KB core-speed L2
Die size	0.552 inches per side (14.0mm)
Pentium Pro Processor (180MHz)	
Introduction date	November 1, 1995
Clock speeds	180MHz (60MHz x 3)
iCOMP Index 2.0 rating	197
Number of transistors	5.5 million (0.35 micron process), plus 15.5 million in 256KB L2 cache (0.6 micron)
Cache Memory	8Kx2 (16KB) L1, 256KB core-speed L2
Die size	0.552 inches per side (14.0mm)
Pentium Pro Processor (166MHz)	
Introduction date	November 1, 1995

Clock speeds	166MHz (66MHz x 2.5)
Number of transistors	5.5 million (0.35 micron process), plus 31 million in 512KB L2 cache (0.35 micron)
Cache Memory	8Kx2 L1, 512KB core-speed L2
Die size	0.552 inches per side (14.0mm)
Pentium Pro Processor (150MHz)	
Introduction date	November 1, 1995
Clock speeds	150MHz (60MHz x 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	8Kx2 speed L2
Die size	0.691 inches per side (17.6mm)

As you saw in Table 3.5, performance comparisons on the iCOMP 2.0 Index rate a classic Pentium 200MHz at 142, whereas a Pentium Pro 200MHz scores an impressive 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 will nullify much of the performance gain shown by the iCOMP 2.0 rating.

Like 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 60 or 66MHz motherboard bus speeds. In fact, the L2 cache features its own internal 64-bit backside bus, which does not share time with the external 64-bit frontside 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 may still try to include cache memory in their design, 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.

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 new type of multiprocessor configuration called the Multiprocessor Specification (MPS 1.1). The Pentium Pro with MPS allows 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 are pretty much exclusively PCI and ISA bus-based, and Intel is producing its own chipsets for these motherboards. The first chipset was the 450KX/GX (code-named Orion), while the most recent chipset for use with the Pentium Pro is the 440LX (Natoma). Due to 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. Even so, the Pentium Pro can be found in all types of motherboard designs; ATX is not mandatory.

Some Pentium Pro system manufacturers have been tempted to stick with the Baby-AT form factor. The big problem with the standard Baby-AT form factor is keeping the CPU properly cooled. The massive Pentium Pro processor consumes more than 25 watts and generates an appreciable amount of heat.

Four special Voltage Identification (VID) pins are on the Pentium Pro processor. These pins can be used to support automatic selection of power supply voltage. This means that 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 define the voltage required by the processor. In addition to allowing for automatic voltage settings, this feature has been 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 3.29. A 1 in this table refers to an open pin and 0 refers to a short to ground. The voltage regulators on the motherboard should supply the voltage that is requested or disable itself.

Table 3.29 Pentium Pro Voltage Identification Definition

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

Most Pentium Pro processors run at 3.3v, but a few run at 3.1v. Although those are the only versions available now, support for a wider range of VID settings will benefit the system in meeting the power requirements of future Pentium Pro processors. 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 has found a niche in file server applications due primarily to the full-core speed high-capacity internal L2 cache.

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 code name 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 3.39.

Figure 3.39
Pentium II Processor die.
Photograph used by permission of Intel Corporation.

From a physical standpoint, it is truly something new. Abandoning the chip in a socket approach used by virtually all processors up until this point, the Pentium II chip is characterized by its Single Edge Contact (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 3.40, which 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.

There are two variations on these cartridges, called SECC (Single Edge Contact Cartridge) and SECC2. Figure 3.41 shows a diagram of the SECC package. Figure 3.42 shows the SECC2 package.

Figure 3.40
Pentium II Processor Board (inside SEC cartridge).
Photograph used by permission of Intel Corporation.

Figure 3.41
SECC components showing enclosed processor board.

Figure 3.42
2 Single Edge Contact Cartridge, rev. 2 components showing half-enclosed processor board.

As you can see from these figures, the SECC2 version is cheaper to make because it uses fewer overall parts. It also allows for a more direct heat sink attachment to the processor for better cooling. Intel transitioned from SECC to SECC2 in the beginning of 1999; all newer PII/PIII cartridge processors use the improved SECC2 design.

By using separate chips mounted on a circuit board, Intel can build the Pentium II much less expensively than the multiple die within a package used in the Pentium Pro. Intel can 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 has 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 .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 percent performance boost, compared to the 233MHz Pentium processor with MMX technology, and approximately 50 percent more performance on multimedia benchmarks. These are very fast processors, at least for now. As shown in Table 3.3, 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 also included. 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).

The original Pentium II processors were manufactured using a 0.35 micron process. More recent models, starting with the 333MHz version, have been manufactured using a newer 0.25 micron process. Intel is considering going to a 0.18 micron process in the future. By going to the smaller process, power draw is greatly reduced.

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. Future Pentium III processors will use the 0.25- and 0.18 micron processes and even 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.

Like 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 than 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 3.30 shows the general Pentium II processor specifications. Table 3.31 shows the specifications that vary by model for the models that have been introduced to date.

Table 3.30 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	16Kx2 (32KB) L1, 512KB 1/2-speed L2
Internal registers	32-bit
Bus Speeds	**66MHz, 100MHz**
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 in. (12.82cm)x2.473 inches (6.28cm) x0.647 in. (1.64cm)
Math coprocessor	Built-in FPU (floating-point unit)
Power management	SMM (System Management Mode)

Table 3.31 Pentium II Specifications by Model

Pentium II MMX Processor (350, 400, and 450MHz)	
Introduction date	April 15, 1998
Clock speeds	350MHz (100MHzx3.5), 400MHz (100MHz x4), and 450MHz (100MHzx4.5)
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 2
Die size	0.400 inches per side (10.2mm)
Mobile Pentium II Processor (266, 300, 333, and 366MHz)	
Introduction date	January 25, 1999
Clock speeds	266, 300, 333, 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
Pentium II MMX Processor (333MHz)	
Introduction date	January 26, 1998
Clock speeds	333MHz (66MHzx5)
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 inches per side (10.2mm)
Pentium II MMX Processor (300MHz)	
Introduction date	May 7, 1997
Clock speeds	300MHz (66MHzx4.5)
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 inches per side (14.2mm)
Pentium II MMX Processor (266MHz)	
Introduction date	May 7, 1997
Clock speeds	266MHz (66MHzx4)
iCOMP Index 2.0 rating	303
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 inches per side (14.2mm)
Pentium II MMX Processor (233MHz)	
Introduction date	May 7, 1997

Clock speeds	233MHz (66MHzx3.5)
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 inches per side (14.2mm)

As you can see from the table, the Pentium II can handle up to 64GB of physical memory. Like the Pentium Pro, the CPU incorporates Dual Independent Bus architecture. This means the chip has two independent buses: one for accessing the L2 cache, the other for accessing main memory. These dual buses can operate simultaneously, greatly accelerating the flow of data within the system. 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, while 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 allows up to three times the bandwidth of normal single-bus processors like 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 can now 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 runs 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 will allow 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. This is very important if you ever plan on adding more than 512MB of memory. In that case, you would definitely want the 350MHz or faster version; otherwise, memory performance would suffer.

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.

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 include a retention mechanism, plus other important system integration components.)

The Pentium II can generate a significant amount of heat that must be dissipated. This is accomplished by installing a heat sink on the processor. Many of the Pentium II processors will use an active heat sink that incorporates a fan. Unlike heat sink 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 heat sink supports are needed to furnish mechanical support between the fan heat sink and support holes on the motherboard. Normally, a plastic support is inserted into the heat sink holes in the motherboard next to the CPU, before installing the CPU/heat sink package. Most fan heat sinks have two components: a fan in a plastic shroud and a metal heat sink. The heat sink 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 3.43 shows the SEC assembly with fan, power connectors, mechanical supports, and the slot and support holes on the motherboard.

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 3.44 to locate these markings.

After you have located the specification number (actually, it is an alphanumeric code), you can look it up in Table 3.32 to see exactly which processor you have.

Figure 3.43
Pentium II/III processor and heat sink assembly.

Figure 3.44
Pentium II/III Single Edge Contact Cartridge.

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 that this processor runs on only 2.0 volts. 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.

Table 3.32 Basic Pentium II Processor Identification Information

S-spec	Core Stepping	CPUID	Core/Bus Speed (MHz)	L2 Cache Size (MB)	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	5

						3.00	
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	5
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	2, 5

						3.00	
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
SL2KE	TdB0	1632h	333/66	512	ECC	PGA	2, 4
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

SECC = Single Edge Contact Cartridge
SECC2 = Single Edge Contact Cartridge revision 2
PLGA = Plastic Land Grid Array
OLGA = Organic Land Grid Array
CPUID = The internal ID returned by the CPUID instruction
ECC = Error Correcting Code
1. This is a boxed Pentium II processor with an attached fan heat sink.
2. These processors have an enhanced L2 cache, which can cache up to 4GB of main memory. Other standard PII processors can only cache up to 512MB of main memory.
3. These boxed processors may have packaging which incorrectly indicates ECC support in the L2 cache.
4. This is a boxed Pentium II OverDrive processor with an attached fan heat sink, designed for upgrading Pentium Pro (Socket 8) systems.
5. These parts will only operate at the specified clock multiplier frequency ratio at which they were manufactured. They can only be overclocked by increasing bus speed.

The two variations of the SECC2 cartridge vary by the type of processor core package on the board. The PLGA (Plastic Land Grid Array) is the older type of packaging used in previous SECC cartridges as well and is being phased out. Taking its place is the newer OLGA (Organic Land Grid Array), which is a processor core package that is smaller and easier to manufacture. It also allows better thermal transfer between the processor die and the heat sink, which is attached directly to the top of the OLGA chip package. Figure 3.45 shows the open back side (where the heat sink would be attached) of SECC2 processors with PLGA and OLGA cores.

Figure 3.45
SECC2 processors with PLGA and OLGA cores.

Pentium II motherboards have an onboard voltage regulator circuit that is designed to power the CPU. Currently, there are Pentium II processors that 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, there are no jumpers or switches to set; the voltage setting is handled completely automatically through the Voltage ID (VID) pins on the processor cartridge. Table 3.33 shows the relationship between the pins and the selected voltage.

Table 3.33 Slot 1 and Socket 370 Voltage ID Pin Definitions

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

1	0	1	1	1	2.8
1	0	1	1	0	2.9
1	0	1	0	1	3.0
1	0	1	0	0	3.1
1	0	0	1	1	3.2
1	0	0	1	0	3.3
1	0	0	0	1	3.4
1	0	0	0	0	3.5

0 = Processor pin connected to Vss
1 = Open on processor
VID0-VID3 used on Socket 370
Socket 370 supports 1.30V through 2.05V settings only.
VID0-VID4 used on Slot 1
Slot 1 supports 1.30V through 3.5V settings.

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 is the first chipset on the market that allows 100MHz processor bus operation, although that is currently not supported in the mobile versions. The 440BX chipset was released at the same time as the 350 and 400MHz versions of the Pentium II; it is the recommended minimum chipset for any new Pentium II motherboard purchases.

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. This means that it runs at full-core speed, making it faster than the desktop Pentium II, because the desktop chips use half-speed L2 cache.

Celeron

The Celeron processor is a P6 with the same processor core as the Pentium II in the original two versions and now the same core as the PIII in the latest version. It is mainly designed for lower-cost PCs in the $1,000 or less price category. The best "feature" is that although the cost is low, the performance is not. In fact, due to the superior cache design, the Celeron outperforms the Pentium II at the same speed and at a lower cost.

Most of the features for the Celeron are the same as the Pentium II and III because it uses the same internal processor core. The main differences are in packaging and L2 cache design.

Up until recently, all Celeron processors were 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 (Single Edge Contact Cartridge) used in the Pentium II/III, with the exception of the fancy plastic cartridge cover. This cover is deleted in the Celeron, making it cheaper to produce and sell. Essentially the Celeron uses the same circuit board as is inside the Pentium II package.

Even without the plastic covers, the Slot 1 packaging was more expensive than it should be. 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 heat sinks 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 3.46) or FC-PGA (Flip Chip PGA). Both the PPGA and FC-PGA packages plug into the 370 pin socket and allows for lower-cost, lower-profile, and smaller systems because of the less expensive processor retention and cooling requirements of the socketed processor.

All Celeron processors at 433MHz and lower have been available in the SEPP that plugs into the 242-contact slot connector. The 300MHz and higher versions are also available in the PPGA package. This means that the 300MHz to 433MHz have been available in both packages, while the 466MHz and higher speed versions are only available in the PPGA.

Figure 3.46
Celeron processors in the FC-PGA, PPGA, and SEP packages.

Motherboards that include Socket 370 can accept the PGA versions of both the Celeron or Pentium III in most cases. If you want to use a Socket 370 version of the Celeron in a Slot 1 motherboard, there are slot-to-socket adapters (usually called slot-kets) available for about $10–$20 that plug into Slot 1 and incorporate a Socket 370 on the card. Figure 3.47 shows a typical slot-ket adapter.

Highlights of the Celeron include

- 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 (Error Correcting Code)
- Uses same P6 core processor as the Pentium II (266 through 533MHz) and now the Pentium III (533A MHz and higher)
- Dynamic execution microarchitecture
- Operates on a 66MHz CPU bus (future versions will likely also use the 100MHz bus)
- Specifically designed for lower-cost value PC systems
- Includes MMX technology; Celeron 533A and higher includes SSE
- More cost-effective packaging technology including Single Edge Processor (SEP), Plastic Pin Grid Array (PPGA), or Flip Chip Pin Grid Array (FCPGA) packages
- Integrated 32KB L1 cache, implemented as separate 16KB instruction and 16KB data caches
- Integrated thermal diode for temperature monitoring

Figure 3.47
Slot-ket adapter for installing PPGA processors in Slot 1 motherboards.

The Intel Celeron processors from the 300A and higher include integrated 128KB L2 cache. The core for the 300A through 533MHz versions which are based on the Pentium II core include 19 million transistors due to 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 Pentium III–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 Pentium III and Celeron using the same die and just disable part of the cache on the Celeron versions, rather than to come up with a unique die for the newer Celerons. The Pentium III–based Celeron processors also support the Streaming SIMD Extensions (SSE) in addition to MMX instructions. The older Celerons based on the Pentium II core only support MMX.

All the Celerons are manufactured using the .25 micron process, which reduces processor heat and enables the Intel Celeron processor to use a smaller heat sink compared to some of the Pentium II processors. Table 3.34 shows the power consumed by the various Celeron processors.

Table 3.34 Intel Celeron Processor Power Consumption

Speed (MHz)	L2 Cache	Max. Temp. (C)	Voltage Power	Max. (W)	Package
266	none	85	2.0V	16.59	SEPP
266	none	85	2.0V	16.59	SEPP
300	none	85	2.0V	18.48	SEPP
300	none	85	2.0V	18.48	SEPP
300A	128KB	85	2.0V	19.05	SEPP
300A	128KB	85	2.0V	19.05	SEPP
300A	128KB	85	2.0V	19.05	PPGA
333	128KB	85	2.0V	20.94	SEPP
333	128KB	85	2.0V	20.94	SEPP
333	128KB	85	2.0V	20.94	PPGA
366	128KB	85	2.0V	21.7	SEPP
366	128KB	85	2.0V	21.7	PPGA
400	128KB	85	2.0V	23.7	SEPP
400	128KB	85	2.0V	23.7	PPGA
433	128KB	85	2.0V	24.1	PPGA
466	128KB	70	2.0V	25.7	PPGA
500	128KB	70	2.0V	27.2	PPGA
533	128KB	70	2.0V	28.3	PPGA
533A	128KB	90	1.5V	11.2	FCPGA
566	128KB	90	1.5V	14.9	FCPGA
600	128KB	90	1.5V	15.8	FCPGA

Figure 3.48 shows the Intel Celeron processor identification information. Figure 3.49 shows the Celeron's PPGA processor markings.

Figure 3.48
Celeron SEPP (Single Edge Processor Package) processor markings.

Note - The markings on the processor identify the following information:

SYYYY = S-spec. number

FFFFFFFF = FPO # (test lot traceability #)

COA = Country of assembly

Note - The PPGA processor markings identify the following information:

AAAAAAA = Product code

ZZZ = Processor speed (MHz)

LLL = Integrated L2 cache size (in Kilobytes)

SYYYY = S-spec. number

FFFFFFFF-XXXX = Assembly lot tracking number

Figure 3.49
Celeron PPGA processor markings.

Table 3.35 shows all the available variations of the Celeron, indicated by the S-specification number.

Pentium III

The Pentium III processor, shown in Figure 3.50, was first released in February 1999 and introduced several new features to the P6 family. The most important advancements are the streaming SIMD extensions (SSE), consisting of 70 new instructions that dramatically enhance the performance and possibilities of advanced imaging, 3D, streaming audio, video, and speech-recognition applications.

Figure 3.50
Pentium III processor in SECC2 (Slot 1) and FC-PGA (Socket 370) packages.
Photograph used by permission of Intel Corporation.

Originally based on Intel’s advanced 0.25 micron CMOS process technology, the PIII core started out with over 9.5 million transistors. In late 1999 Intel shifted to a 0.18 micron process and added 256KB of on-die L2 cache, which brought the transistor count to a whopping 28.1 million! The Pentium III is available in speeds from 450MHz through 1000MHz and beyond, as well as server versions with larger or faster cache called 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 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.

Table 3.35 Intel Celeron Variations and Specifications

Speed (MHz)	**Bus Speed (MHz)**	**Multiplier**	**Boxed CPU S-spec**	**OEM CPU S-spec**	**Stepping**	**CPUID**	**L2 Cache**	**Graphics Extensions**	**Max. Temp. (C)**	**Voltag**
266	66	4x	SL2YN	SL2SY	dA0	0650	none	MMX	85	2.0V
266	66	4x	SL2QG	SL2TR	dA1	0651	none	MMX	85	2.0V
300	66	4.5x	SL2Z7	SL2YP	dA0	0650	none	MMX	85	2.0V
300	66	4.5x	SL2Y2	SL2X8	dA1	0651	none	MMX	85	2.0V
300A	66	4.5x	SL32A	SL2WM	mA0	0660	128KB	MMX	85	2.0V
300A	66	4.5x	SL2WM	SL2WM	mA0	0660	128KB	MMX	85	2.0V
300A	66	4.5x	SL35Q	SL36A	mB0	0665	128KB	MMX	85	2.0V
333	66	5x	SL32B	SL2WN	mA0	0660	128KB	MMX	85	2.0V
333	66	5x	SL2WN	SL2WN	mA0	0660	128KB	MMX	85	2.0V
333	66	5x	SL35R	SL36B	mB0	0665	128KB	MMX	85	2.0V
366	66	5.5x	SL37Q	SL376	mA0	0660	128KB	MMX	85	2.0V
366	66	5.5x	SL35S	SL36C	mB0	0665	128KB	MMX	85	2.0V
400	66	6x	SL37V	SL39Z	mA0	0660	128KB	MMX	85	2.0V
400	66	6x	SL37X	SL3A2	mB0	0665	128KB	MMX	85	2.0V
433	66	6.5x	SL3BS	SL3BA	mB0	0665	128KB	MMX	85	2.0V
466	66	7x	SL3FL	SL3EH	mB0	0665	128KB	MMX	70	2.0V
500	66	7.5x	SL3LQ	SL3FY	mB0	0665	128KB	MMX	70	2.0V
533	66	8x	SL3PZ	SL3FZ	mB0	0665	128KB	MMX	70	2.0V
533A	66	8x	n/a	SL46S	cBO	068x	128KB	SSE	90	1.5V
566	66	8.5x	SL3W7	SL46T	cB0	068x	128KB	SSE	90	1.5V
600	66	9x	SL3W8	SL46U	cB0	068x	128KB	SSE	90	1.5V

SEPP = Single Edge Processor Package (Card) SSE = MMX plus Streaming SIMD (Single Instruction Multiple Data) Extensions
PPGA = Plastic Pin Grid Array Boxed processors include a heat sink with fan

FCPGA = Flip Chip Pin Grid Array 266MHz through 533MHz are based on 0.25 micron Pentium II core
MMX = Multi Media Extensions 533A MHz and higher are based on 0.18 micron Pentium III core

Pentium III processors are available in Intel's Single Edge Contact Cartridge 2 (SECC2) form factor, which is replacing the more expensive older SEC packaging. The SECC2 package covers only one side of the chip, and allows for better heat sink attachment and less overall weight. It is also less expensive.

Architectural features of the Pentium III processor include

- *Streaming SIMD Extensions*. 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, the first of Intel's planned building blocks for PC security, serves as an electronic serial number for the processor and, by extension, its system or user. This enables the system/user to be identified by networks and applications. The processor serial number will be used in applications that benefit from stronger forms of system and user identification, such as the following:
  - *Applications using security capabilities*. Managed access to new Internet content and services; electronic document exchange.
  - *Manageability applications*. Asset management; remote system load and configuration.
  - *Intel MMX Technology.*
  - *Dynamic Execution Technology.*
  - *Incorporates an on-die diode.* This can be used to monitor the die temperature for thermal management purposes.

Most of the Pentium III processors will be made in the improved SECC2 packaging or, even better, the FC-PGA (Flip Chip PGA) package, which is much less expensive to produce and allows for a more direct attachment of the heat sink 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-ket adapter.

All Pentium III processors have either 512KB or 256KB of L2 cache, which runs at either half-core or full-core speed. Xeon versions have either 512KB, 1MB, or 2MB of L2 cache that runs at full-core speed. These are more expensive versions designed for servers and workstations.

All PIII processor L2 caches can cache up to 4GB of addressable memory space, and include Error Correction Code (ECC) capability.

Pentium III processors can be identified by their markings, which are found on the top edge of the processor cartridge. Figure 3.51 shows the format and meaning of the markings.

Figure 3.51

Pentium III processor markings.

Table 3.36 shows the available variations of the Pentium III, indicated by the S-specification number.

Pentium III processors are all clock multiplier locked. This is a means to prevent processor fraud and overclocking 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 remarked as higher speeds. It pays to purchase your systems or processors from direct Intel distributors or high-end dealers that do not engage in these practices.

Pentium II/III Xeon

The Pentium II and III processors are available in special high-end versions called Xeon processors. Originally introduced in June 1998 in Pentium II versions, later Pentium III versions were introduced in March 1999. These differ from the standard Pentium II and III in three ways: packaging, cache size, and cache speed.

Xeon processors use a larger SEC (Single Edge Contact) cartridge than the standard PII/III processors, mainly to house a larger internal board with more cache memory. The Xeon processor is shown in Figure 3.52; the Xeon's SEC is shown in Figure 3.53.

Figure 3.52
Pentium III Xeon processor.
Photograph used by permission of Intel Corporation.

Figure 3.53
Xeon processor internal components.

Table 3.36 Intel Pentium III Processor Variations

Speed (MHz)	Bus Speed (MHz)	Multiplier	Boxed CPU S-spec	OEM CPU S-spec	Stepping	CPUID	L2 Cache	L2 Speed	Temp. Max. (C)	Voltage	Max Pow (W)
450	100	4.5x	SL3CC	SL364	kB0	0672	512KB	225	90	2.00	25.3
450	100	4.5x	SL37C	SL35D	kC0	0673	512KB	225	90	2.00	25.3
500	100	5x	SL3CD	SL365	kB0	0672	512KB	250	90	2.00	28.0
500	100	5x	SL365	SL365	kB0	0672	512KB	250	90	2.00	28.0
500	100	5x	SL37D	SL35E	kC0	0673	512KB	250	90	2.00	28.0
500E	100	5x	SL3R2	SL3Q9	cA2	0681	256KB	500	85	1.60	13.2
500E	100	5x	SL45R	SL444	cB0	0683	256KB	500	85	1.60	13.2
533B	133	4x	SL3E9	SL3BN	kC0	0673	512KB	267	90	2.05	29.7
533EB	133	4x	SL3SX	SL3N6	cA2	0681	256KB	533	85	1.65	14.0
533EB	133	4x	SL3VA	SL3VF	cA2	0681	256KB	533	85	1.65	14.0
533EB	133	4x	SL44W	SL3XG	cB0	0683	256KB	533	85	1.65	14.0

533EB	133	4x	SL45S	SL3XS	cB0	0683	256KB	533	85	1.65	14.0
550	100	5.5x	SL3FJ	SL3F7	kC0	0673	512KB	275	80	2.00	30.8
550E	100	5.5x	SL3R3	SL3QA	cA2	0681	256KB	550	85	1.60	14.5
550E	100	5.5x	SL3V5	SL3N7	cA2	0681	256KB	550	85	1.60	14.5
550E	100	5.5x	SL44X	SL3XH	cB0	0683	256KB	550	85	1.60	14.5
550E	100	5.5x	SL45T	N/A	cB0	0683	256KB	550	85	1.60	14.5
600	100	6x	SL3JT	SL3JM	kC0	0673	512KB	300	85	2.00	34.5
600E	100	6x	SL3NA	SL3H6	cA2	0681	256KB	600	82	1.65	15.8
600E	100	6x	SL3NL	SL3VH	cA2	0681	256KB	600	82	1.65	15.8
600E	100	6x	SL44Y	SL43E	cB0	0683	256KB	600	82	1.65	15.8
600E	100	6x	SL45U	SL3XU	cB0	0683	256KB	600	82	1.65	15.8
600B	133	4.5x	SL3JU	SL3JP	kC0	0673	512KB	300	85	2.05	34.5
600EB	133	4.5x	SL3NB	SL3H7	cA2	0681	256KB	600	82	1.65	15.8
600EB	133	4.5x	SL3VB	SL3VG	cA2	0681	256KB	600	82	1.65	15.8
600EB	133	4.5x	SL44Z	SL3XJ	cB0	0683	256KB	600	82	1.65	15.8
600EB	133	4.5x	SL45V	SL3XT	cB0	0683	256KB	600	82	1.65	15.8
650	100	6.5x	SL3NR	SL3KV	cA2	0681	256KB	650	82	1.65	17.0
650	100	6.5x	SL3NM	SL3VJ	cA20	681	256KB	650	82	1.65	17.0
650	100	6.5x	SL452	SL3XK	cB0	0683	256KB	650	82	1.65	17.0
650	100	6.5x	SL45W	SL3XV	cB0	0683	256KB	650	82	1.65	17.0
667	133	5x	SL3ND	SL3KW	cA2	0681	256KB	667	82	1.65	17.5
667	133	5x	SL3T2	SL3VK	cA2	0681	256KB	667	82	1.65	17.5
667	133	5x	SL453	SL3XL	cB0	0683	256KB	667	82	1.65	17.5
667	133	5x	SL45X	SL3XW	cB0	0683	256KB	667	82	1.65	17.5
700	100	7x	SL3SY	SL3S9	cA2	0681	256KB	700	80	1.65	18.3
700	100	7x	SL3T3	SL3VL	cA2	0681	256KB	700	80	1.65	18.3
700	100	7x	SL454	SL453	cB0	0683	256KB	700	80	1.65	18.3
700	100	7x	SL45Y	SL3XX	cB0	0683	256KB	700	80	1.65	18.3
733	133	5.5x	SL3SZ	SL3SB	cA2	0681	256KB	733	80	1.65	19.1
733	133	5.5x	SL3T4	SL3VM	cA2	0681	256KB	733	80	1.65	19.1
733	133	5.5x	SL455	SL3XN	cB0	0683	256KB	733	80	1.65	19.1
733	133	5.5x	SL45Z	SL3XY	cB0	0683	256KB	733	80	1.65	19.1
750	100	7.5x	SL3V6	SL3WC	cA2	0681	256KB	750	80	1.65	19.5
750	100	7.5x	SL3VC	SL3VN	cA2	0681	256KB	750	80	1.65	19.5
750	100	7.5x	SL456	SL3XP	cB0	0683	256KB	750	80	1.65	19.5
750	100	7.5x	SL462	SL3XZ	cB0	0683	256KB	750	80	1.65	19.5

800	100	8x	SL457	SL3XR	cB0	0683	256KB	800	80	1.65	20.8
800	100	8x	SL463	SL3Y3	cB0	0683	256KB	800	80	1.65	20.8
800EB	133	6x	SL458	SL3XQ	cB0	0683	256KB	800	80	1.65	20.8
800EB	133	6x	SL464	SL3Y2	cB0	0683	256KB	800	80	1.65	20.8
850	100	8.5x	SL47M	SL43F	cB0	0683	256KB	850	80	1.65	22.5
850	100	8.5x	SL49G	SL43H	cB0	0683	256KB	850	80	1.65	22.5
866	133	6.5x	SL47N	SL43G	cB0	0683	256KB	866	80	1.65	22.9
866	133	6.5x	SL49H	SL43J	cB0	0683	256KB	866	80	1.65	22.5
933	133	7x	SL47Q	SL448	cB0	0683	256KB	933	75	1.65	25.5
933	133	7x	SL49J	SL44J	cB0	0683	256KB	933	75	1.65	24.5
1000	133	7.5x	n/a	SL48S	cB0	0683	256KB	1000	60	1.70	33.0

SECC = Single Edge Contact Cartridge ECC = Error Correcting Code
SECC2 = Single Edge Contact Cartridge revision 2 1. This is a boxed processor with an attached heat sink
CPUID = The internal ID returned by the CPUID instruction

Besides the larger package, the Xeon processors also include more L2 cache. They are available in three variations, with 512KB, 1MB, or 2MB of L2 cache. This cache is costly; the list price of the 2MB version is about $2,000!

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 are separate chips on the board; up until recently they were not integrated into the processor die. 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 a whopping 140 million transistors in the latest 2MB cache version, setting an industry record. The high transistor counts are due to the on-die L2 cache, which is very transistor intensive. The L2 cache in all Xeon processors has a full 64GB RAM address range and supports ECC (Error Correcting Code).

Table 3.37 shows the Xeon processor specifications for each model.

Table 3.37 Intel Pentium II and III Xeon Specifications

Pentium II Xeon:							
Speed (MHz)	**Bus Speed (MHz)**	**S-spec**	**Stepping**	**CPUID**	**L2 Cache**	**Transistors**	**Process (microns)**
400	100	SL2RH	B0	0652	512KB	7.5M	0.25
400	100	SL2NB	B0	0652	1024KB	7.5M	0.25
400	100	SL35N	B1	0653	512KB	7.5M	0.25
400	100	SL34H	B1	0653	512KB	7.5M	0.25
400	100	SL35P	B1	0653	1024KB	7.5M	0.25

400	100	SL34J	B1	0653	1024KB	7.5M	0.25
Pentium II Xeon:							
Speed (MHz)	**Bus Speed (MHz)**	**S-spec**	**Stepping**	**CPUID**	**L2 Cache**	**Transistors**	**Process (microns)**
450	100	SL33T	B1	0653	512KB	7.5M	0.25
450	100	SL354	B1	0653	512KB	7.5M	0.25
450	100	SL36W	B1	0653	512KB	7.5M	0.25
450	100	SL2XJ	B1	0653	512KB	7.5M	0.25
450	100	SL33U	B1	0653	1024KB	7.5M	0.25
450	100	SL2XK	B1	0653	1024KB	7.5M	0.25
450	100	SL33V	B1	0653	2048KB	7.5M	0.25
450	100	SL2XL	B1	0653	2048KB	7.5M	0.25
Pentium III Xeon:							
500	100	SL2XU	B0	0672h	512KB	9.5M	0.25
500	100	SL2XV	B0	0672h	1024KB	9.5M	0.25
500	100	SL2XW	B0	0672h	2048KB	9.5M	0.25
500	100	SL3C9	B0	0672h	512KB	9.5M	0.25
500	100	SL3CA	B0	0672h	1024KB	9.5M	0.25
500	100	SL3CB	B0	0672h	2048KB	9.5M	0.25
550	100	SL3FK	C0	0673h	512KB	9.5M	0.25
500	100	SL3D9	C0	0673h	512KB	9.5M	0.25
500	100	SL3DA	C0	0673h	1024KB	9.5M	0.25
500	100	SL3DB	C0	0673h	2048KB	9.5M	0.25
550	100	SL3AJ	C0	0673h	512KB	9.5M	0.25
550	100	SL3CE	C0	0673h	1024KB	9.5M	0.25
550	100	SL3CF	C0	0673h	2048KB	9.5M	0.25
550	100	SL3TW	C0	0673h	1024KB	9.5M	0.25
550	100	SL3Y4	C0	0673h	512KB	9.5M	0.25
550	100	SL3FR	C0	0673h	512KB	9.5M	0.25
500	100	SL385	C0	0673h	512KB	9.5M	0.25
500	100	SL386	C0	0673h	1024KB	9.5M	0.25
500	100	SL387	C0	0673h	2048KB	9.5M	0.25
550	100	SL3LM	C0	0673h	512KB	9.5M	0.25
550	100	SL3LN	C0	0673h	1024KB	9.5M	0.25
550	100	SL3LP	C0	0673h	2048KB	9.5M	0.25
600	133	SL3BJ	A2	0681h	256KB	28.1M	0.18
600	133	SL3BK	A2	0681h	256KB	28.1M	0.18

667	133	SL3BL	A2	0681h	256KB	28.1M	0.18
667	133	SL3DC	A2	0681h	256KB	28.1M	0.18
733	133	SL3SF	A2	0681h	256KB	28.1M	0.18
Pentium III Xeon:							
Speed (MHz)	**Bus Speed (MHz)**	**S-spec**	**Stepping**	**CPUID**	**L2 Cache**	**Transistors**	**Process (microns)**
733	133	SL3SG	A2	0681h	256KB	28.1M	0.18
800	133	SL3V2	A2	0681h	256KB	28.1M	0.18
800	133	SL3V3	A2	0681h	256KB	28.1M	0.18
600	133	SL3SS	A2	0681h	256KB	28.1M	0.18
667	133	SL3ST	A2	0681h	256KB	28.1M	0.18
733	133	SL3SU	A2	0681h	256KB	28.1M	0.18
800	133	SL3VU	A2	0681h	256KB	28.1M	0.18
600	133	SL3WM	B0	0683h	256KB	28.1M	0.18
600	133	SL3WN	B0	0683h	256KB	28.1M	0.18
667	133	SL3WP	B0	0683h	256KB	28.1M	0.18
667	133	SL3WQ	B0	0683h	256KB	28.1M	0.18
733	133	SL3WR	B0	0683h	256KB	28.1M	0.18
733	133	SL3WS	B0	0683h	256KB	28.1M	0.18
800	133	SL3WT	B0	0683h	256KB	28.1M	0.18
800	133	SL3WU	B0	0683h	256KB	28.1M	0.18
866	133	SL3WV	B0	0683h	256KB	28.1M	0.18
866	133	SL3WW	B0	0683h	256KB	28.1M	0.18
933	133	SL3WX	B0	683h	256KB	28.1M	0.18
933	133	SL3WY	B0	683h	256KB	28.1M	0.18
700	100	SL3U4	A0	6A0h	1024KB	84M	0.18
700	100	SL3U5	A0	6A0h	1024KB	84M	0.18
700	100	SL3WZ	A0	6A0h	2048KB	140M	0.18
700	100	SL3X2	A0	6A0h	2048KB	140M	0.18
700	100	SL4GD	A0	6A0h	1024KB	84M	0.18
700	100	SL4GE	A0	6A0h	1024KB	84M	0.18
700	100	SL4GF	A0	6A0h	2048KB	140M	0.18
700	100	SL4GG	A0	6A0h	2048KB	140M	0.18

Note that the Slot 2 Xeon processors do not replace the Slot 1 processors. Xeon processors for Slot 2 are targeted at the mid-range to high-end server and workstation market segments, offering larger, full-speed L2 caches and four-way multiprocessor support. Pentium III processors for Slot 1 will continue to be the processor used in the business and home desktop market segments, and for entry-level servers and workstations (single and dual processor systems).

Pentium III Future

There are several new developments on target for the Pentium III processors. The primary trend seems to be the integration of L2 cache into the processor die, which also means it runs at full-core speed.

There will also be further reductions in the process size used to manufacture the processors. Pentium III processors first used the 0.25 micron Katmai core and later shifted to the 0.18 micron Coppermine core with on-die L2 cache. The future will see the migration to the Willamette core, which is an enhanced Pentium III, along with a migration to a 0.13-micron die with likely a larger integrated L2 cache. The shift to the 0.13 micron process will also include a shift from aluminum interconnects on the chip die to copper interconnects, as well as the use of larger 300mm (12") wafers.

Other Sixth-Generation Processors

Besides Intel, many other manufacturers are now making P6-type processors, but often with a difference. Most of them are designed to interface with P5 class motherboards and for the lower-end markets. AMD has recently offered up the Athlon and Duron processors, which are true sixth-generation designs using their own proprietary connection 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 normally 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 done 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 included 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 an RISC (Reduced Instruction Set Computer) core. A translation unit dynamically translates x86 instructions into RISC86 instructions. These RISC86 instructions were specifically designed 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 was essentially 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 due to its unique hybrid design. Because it is designed to install in Socket 7, which is a fifth-generation processor socket and motherboard design, it cannot perform quite as a true sixth-generation chip because the Socket 7 architecture severely limits cache and memory performance. However, with this processor, AMD is giving Intel a lot of competition in the low- to mid-range market, where the Pentium is still popular.

The K6 processor contains an industry-standard, high-performance implementation of the new multimedia instruction set (MMX), enabling a high level of multimedia performance. The K6-2 introduced an upgrade to MMX 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. This enables PC manufacturers and resellers to speed time to market and deliver systems with an easy upgrade path for the future. AMD’s state-of-the-art manufacturing facility in Austin, Texas (Fab 25) makes the AMD-K6 series processors. Initially it used AMD’s 0.35 micron, five-metal layer process technology; newer variations use the 0.25 micron processor 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 (FPU)
- Industry-standard MMX instruction support

- System Management Mode (SMM)
- Ceramic Pin Grid Array (CPGA) Socket 7 design
- Manufactured using a 0.35 micron and 0.25 micron, five-layer design

The K6-2 adds

- Higher clock speeds
- Higher bus speeds of up to 100MHz (Super7 motherboards)
- 3DNow; 21 new graphics and sound processing instructions

The K6-3 adds

- 256KB of on-die full-core speed L2 cache

The addition of the full speed L2 cache in the K6-3 is significant. It brings the K6 series to a level where it can fully compete with the Intel Celeron and Pentium II processors. The 3DNow capability added in the K6-2/3 is also being exploited by newer graphics programs, making these processors ideal for lower-cost gaming systems.

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 has beefed up the internal L1 cache to 64KB total, twice the size of the Pentium II or III. This, plus the dynamic execution capability, allows the K6 to outperform the Pentium and come close to the Pentium II in performance for a given clock rate. The K6-3 is even better with the addition of full-core speed L2 cache.

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 145 mV of nominal (2.9v+/-145 mV). When used with a 233MHz processor, the voltage regulator must maintain the core voltage within 100 mV of nominal (3.2v+/-100 mV).

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. Ensure your heat sink is securely fitted to the processor and 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 BIOS 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 Web site. All the motherboards on that list have been tested to work properly with the AMD-K6. So, unless these requirements can be verified elsewhere, it is recommended that you only use a motherboard from that list with the AMD-K6 processor.

The multiplier, bus speed, and voltage settings for the K6 are shown in Table 3.38. You can identify which AMD-K6 you have by looking at the markings on this chip, as shown in Figure 3.54.

Table 3.38 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 3.54
AMD-K6 processor markings.

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. To get the 4x and higher setting requires a motherboard that controls three BF (bus frequency) pins, including BF2. Older motherboards can only control two BF pins. The settings for the multipliers are shown in Table 3.39.

Table 3.39 AMD-K6 Multiplier Settings

Multiplier Setting	BF0	BF1	BF2
2.5x	Low	Low	High
3x	High	Low	High
3.5x	High	High	High
4x	Low	High	Low
4.5x	Low	Low	Low
5x	High	Low	Low
5.5x	High	High	Low

These settings are normally 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. This means it designs and builds its chips in its own fabs. Like Intel, AMD is migrating to 0.25 micron process technology and beyond. 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 now incorporates 21.3 million transistors on a die only 10.9mm on each side, or about 118 square mm. Further process improvements will enable even more transistors, smaller die, higher yields, and greater numbers of processors. AMD has recently won contracts with several high-end system suppliers, which gives it an edge on the other Intel competitors. AMD has delivered more than 50 million Windows-compatible CPUs in the last five years.

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 currently 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 than the Intel processors. Before attempting any upgrades, you should check the board documentation or contact the manufacturer to see if your board will meet the necessary requirements. In some cases, a BIOS upgrade will also be necessary.

AMD Athlon

The Athlon is AMD’s successor to the K6 series (seeFigure 3.55). The Athlon is a whole 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 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 either one-half core, two-fifths core, or one-third core depending on which speed processor you had. In June of 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 PGA (Pin Grid Array) or chip Socket A version, which is replacing the Slot A cartridge version.

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, as 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 in the wrong slot. Figure 3.55 shows the Athlon in the Slot A cartridge. Figure 3.56 shows the Athlon in the PGA package (Socket A).

Figure 3.55
AMD Athlon processor for Slot A (cartridge form factor).

Figure 3.56
AMD Athlon processor for Socket A (PGA form factor).

The Athlon is available in speeds from 550MHz up to 1GHz and beyond and uses a 200MHz 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, now owned by Compaq. The EV6 bus uses a clock speed of 100MHz but double-clocks the data, transferring data twice per cycle, for a cycling speed of 200MHz. Since the bus is eight bytes (64 bits) wide, this results in a throughput of eight bytes times 200MHz or 1.6GB/sec. This is superior to the Intel processors that use a front-side bus speed of only up to 133MHz, which results in 8 bytes times 133MHz or 1.07GB/sec. bandwidth. The AMD bus design eliminates a potential bottleneck between the chipset and processor and allows for 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, or 256KB of full-core speed cache in the later ones. 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 (Streaming SIMD Extensions) in design and intent, but the specific instructions are different and require software support. 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 a 0.18 micron process. The latest versions are even built using copper metal technology, a first in the PC processor business. Eventually all other processors will follow, as copper inconnects allow for lower power consumption and faster operation.

Table 3.40 shows detailed information on the Slot-A version of the Athlon processor.

Table 3.40 AMD Athlon Slot-A Cartridge Processor Information

Part Number	Model	Speed (MHz)	Bus Speed (MHz)	Multiplier	L2 Cache	(MHz)	L2 Speed Voltage	Current (A)	Max. Power (W)	Max. (micr
AMD-K7500MTR51B	Model 1	500	100x2	5x	512KB	250	1.60V	25A	42W	0.25
AMD-K7550MTR51B	Model 1	550	100x2	5.5x	512KB	275	1.60V	30A	46W	0.25
AMD-K7600MTR51B	Model 1	600	100x2	6x	512KB	300	1.60V	33A	50W	0.25
AMD-K7650MTR51B	Model 1	650	100x2	6.5x	512KB	325	1.60V	36A	54W	0.25
AMD-K7700MTR51B	Model 1	700	100x2	7x	512KB	350	1.60V	33A	50W	0.25
AMD-K7550MTR51B	Model 2	550	100x2	5.5x	512KB	275	1.60V	20A	31W	0.18
AMD-K7600MTR51B	Model 2	600	100x2	6x	512KB	300	1.60V	21A	34W	0.18
AMD-K7650MTR51B	Model 2	650	100x2	6.5x	512KB	325	1.60V	22A	36W	0.18
AMD-K7700MTR51B	Model 2	700	100x2	7x	512KB	350	1.60V	24A	39W	0.18
AMD-K7750MTR52B	Model 2	750	100x2	7.5x	512KB	300	1.60V	25A	40W	0.18
AMD-K7800MPR52B	Model 2	800	100x2	8x	512KB	320	1.70V	29A	48W	0.18
AMD-K7850MPR52B	Model 2	850	100x2	8.5x	512KB	340	1.70V	30A	50W	0.18
AMD-K7900MNR53B	Model 2	900	100x2	9x	512KB	300	1.80V	34A	60W	0.18
AMD-K7950MNR53B	Model 2	950	100x2	9.5x	512KB	317	1.80V	35A	62W	0.18
AMD-K7100MNR53B	Model 2	1000	100x2	10x	512KB	333	1.80V	37A	65W	0.18
AMD-A0650MPR24B	Model 4	650	100x2	6.5x	256KB	650	1.70V	23.8A	36.1W	0.18
AMD-A0700MPR24B	Model 4	700	100x2	7x	256KB	700	1.70V	25.2A	38.3W	0.18
AMD-A0750MPR24B	Model 4	750	100x2	7.5x	256KB	750	1.70V	26.6A	40.4W	0.18

AMD-A0800MPR24B	Model 4	800	100x2	8x	256KB	800	1.70V	28.0A	42.6W	0.18
AMD-A0850MPR24B	Model 4	850	100x2	8.5x	256KB	850	1.70V	29.4A	44.8W	0.18
AMD-A0900MMR24B	Model 4	900	100x2	9x	256KB	900	1.75V	31.7A	49.7W	0.18
AMD-A0950MMR24B	Model 4	950	100x2	9.5x	256KB	950	1.75V	33.2A	52.0W	0.18
AMD-A1000MMR24B	Model 4	1000	100x2	10x	256KB	1000	1.75V	34.6A	54.3W	0.18

Table 3.41 shows information on the PGA (Pin Grid Array) or Socket A version of the AMD Athlon processor.

Table 3.41 AMD Athlon PGA (Pin Grid Array) Processor Information

Part Number	Speed (MHz)	Bus Speed (MHz)	Multiplier	L2 Cache	L2 Speed (MHz)	Voltage	Max. Current (A)	Max. Power (W)	Process (microns)	Tra
A0650APT3B	650	100x2	6.5x	256KB	650	1.7V	23.8A	36.1W	0.18	37M
A0700APT3B	700	100x2	7x	256KB	700	1.7V	25.2A	38.3W	0.18	37M
A0750APT3B	750	100x2	7.5x	256KB	750	1.7V	26.6A	40.4W	0.18	37M
A0800APT3B	800	100x2	8x	256KB	800	1.7V	28.0A	42.6W	0.18	37M
A0850APT3B	850	100x2	8.5x	256KB	850	1.7V	29.4A	44.8W	0.18	37M
A0900AMT3B	900	100x2	9x	256KB	900	1.75V	31.7A	49.7W	0.18	37M
A0950AMT3B	950	100x2	9.5x	256KB	950	1.75V	33.2A	52.0W	0.18	37M
A1000AMT3B	1000	100x2	10x	256KB	1000	1.75V	34.6A	54.3W	0.18	37M

AMD is taking on Intel full force in the high-end market with the Athlon. It beat Intel to the 1GHz mark by introducing its 1GHz Athlon 2 days before Intel introduced the 1GHz Pentium III, and in most benchmarks the AMD Athlon compares as equal if not superior to the Intel Pentium III.

AMD Duron

The AMD Duron processor (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 (see Figure 3.57). 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, however 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. With the high-value design the Duron processor is expected to compete in the sub $1,000 PC market against the Celeron, just as the Athlon is designed to compete in the higher end Pentium III market.

Since 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.

Figure 3.57
AMD Duron processor.

Table 3.42 shows information on the PGA (Pin Grid Array) or Socket A version of the AMD Athlon processor.

Table 3.42 AMD Duron Processor Information

Part Number	Speed (MHz)	Bus Speed (MHz)	Multiplier	L2 Cache	L2 Speed (MHz)	Voltage	Max. Current (A)	Max. Power (W)	Process (microns)	Tran
D0550AST1B	550	100x2	5.5x	64KB	550	1.5V	15.8A	21.1W	0.18	25M
D0600AST1B	600	100x2	6x	64KB	600	1.5V	17.0A	22.7W	0.18	25M
D0650AST1B	650	100x2	6.5x	64KB	650	1.5V	18.2A	24.3W	0.18	25M
D0700AST1B	700	100x2	7x	64KB	700	1.5V	19.2A	25.5W	0.18	25M

Cyrix MediaGX

The Cyrix MediaGX is designed for low-end sub-$1,000 retail store systems that must be highly integrated and low priced. The MediaGX integrates the sound, graphics, and memory control by putting these functions directly within the processor. With all these functions pulled "on chip," MediaGX-based PCs are priced lower than other systems with similar features.

The MediaGX processor integrates the PCI interface, coupled with audio, graphics, and memory-control functions, right into the processor unit. As such, a system with the MediaGX doesn’t require a costly graphics or sound card. Not only that, but on the motherboard level, the MediaGX and its companion chip replace the processor, North and South Bridge chips, the memory control hardware, and L2 cache found on competitive Pentium boards. Finally, the simplified PC design of the MediaGX, along with its low-power and low-heat characteristics, allow the OEM PC manufacturer to design a system in a smaller form factor with a reduced power-supply requirement.

The MediaGX processor is not a Socket 7 processor; in fact, it does not go in a socket at all—it is permanently soldered into its motherboard. Because of the processor's high level of integration, motherboards supporting MediaGX processors and its companion chip (Cx5510) are of a different design than conventional Pentium boards. As such, a system with the MediaGX processor is more of a disposable system than an upgradable system. You will not be able to easily upgrade most components in the system, but that is often not important in the very low-end market. If upgradability is important, look elsewhere. On the other hand, if you need the lowest-priced system possible, one with the MediaGX might fill the bill.

The MediaGX is fully Windows-compatible and will run the same software as an equivalent Pentium. You can expect a MediaGX system to provide equivalent performance as a given Pentium system at the same megahertz. The difference with the MediaGX is that this performance level is achieved at a much lower cost. Because the MediaGX processor is soldered into the motherboard and requires a custom chipset, it is only sold in a complete motherboard form.

There is also an improved MMX-enhanced MediaGX processor that features MPEG1 support, Microsoft PC97 compliance for Plug-and-Play access, integrated game port control, and AC97 audio compliance. It supports Windows 95 and DOS-based games, and MMX software as well. Such systems will also include two universal serial bus (USB) ports, which will accommodate the new generation of USB peripherals such as printers, scanners, joysticks, cameras, and more.

The MediaGX processor is offered at 166 and 180MHz, while the MMX-enhanced MediaGX processor is available at 200MHz and 233MHz. Compaq is using the MMX-enhanced MediaGX processor in its Presario 1220 notebook PCs, which is a major contract win for Cyrix. Other retailers and resellers are offering low-end, low-cost systems in retail stores nationwide.

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 the latest 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.

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.* Allows 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 sized 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 System Management Mode (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 allows 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 requirements for the 6x86 processors. Cyrix maintains a list of recommended motherboards on its Web site that should be consulted if you are considering installing one of these chips in a board.

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 "Cyrix P-Ratings" earlier in this chapter to see 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 chips 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 have to 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.

Itanium (P7/Merced) Seventh-Generation Processors

What is coming after the Pentium III? The next-generation processor was code-named either P7 or Merced and will be called Itanium.

Intel has indicated that the new 64-bit Itanium processor will be available in late 2000. The Itanium processor will be the first processor in Intel’s IA-64 (Intel Architecture 64-bit) product family and will incorporate innovative performance-enhancing architecture techniques, such as prediction and speculation.

Itanium

The most current generation of processor is the P6, which was first seen in the Pentium Pro introduced in November of 1995 and most recently found in the latest Pentium II processors. Obviously, then, the next generation processor from Intel will be called the Itanium.

Intel’s IA-64 product family is expected to expand the capabilities of the Intel architecture to address the high-performance server and workstation market segments. A variety of industry players—among them leading workstation and server-system manufacturers, leading operating system vendors, and dozens of independent software vendors—have already publicly committed their support for the Itanium processor and the IA-64 product family.

As with previous new processor introductions, the P7 will not replace the P6 or P5, at least not at first. It will feature an all new design that will be initially expensive and found only in the highest end systems such as file servers or workstations. Intel expects the Itanium will become the mainstream processor by the year 2004 and that the P6 will likely be found in low-end systems only. Intel is already developing an even more advanced P7 processor, due to ship in 2001, which will be significantly faster than Itanium.

Intel and Hewlett-Packard began jointly working on the P7 processor in 1994. It was then that they began a collaboration on what will eventually become Intel's next-generation CPU. Although we don't know exactly what the new CPU will be like, Intel has begun slowly releasing information about the new processor to prepare the industry for its eventual release. In October of 1997, more than three years after they first disclosed their plan to work together on a new microprocessor architecture, Intel and HP officially announced some of the new processor's technical details.

The first chip to implement the P7 architecture won't ship until late 2000.

Itanium will be the first microprocessor that will be based on the 64-bit, next-generation Intel architecture-64 (IA-64) specification. IA-64 is a completely different processor design, which will use Very Long Instruction Words (VLIW), instruction prediction, branch elimination, speculative loading, and other advanced processes for enhancing parallelism from program code. The new chip will feature elements of both CISC and RISC design.

There is also a new architecture Intel calls Explicitly Parallel Instruction Computing (EPIC), which will let the processor execute *parallel instructions*—several instructions at the same time. In the Itanium, three instructions will be encoded in one 128-bit word, so that each instruction has a few more bits than today's 32-bit instructions. The extra bits let the chip address more registers and tell the processor which instructions to execute in parallel. This approach simplifies the design of processors with many parallel-execution units and should let them run at higher clock rates. In other words, besides being capable of executing several instructions in parallel within the chip, the Itanium will have the capability to be linked to other Itanium chips in a parallel processing environment.

Besides having new features and running a completely new 64-bit instruction set, Intel and HP promise full backward compatibility between the Itanium, the current 32-bit Intel x86 software, and even HP's own PA-RISC software. The P7 will incorporate three different kinds of processors in one and therefore be capable of running advanced IA-64 parallel processing software and IA-32 Windows and HP-RISC UNIX programs at the same time. In this way, Itanium will support 64-bit instructions while retaining compatibility with today's 32-bit applications. This backward compatibility will be a powerful selling point.

To use the IA-64 instructions, programs will have to be recompiled for the new instruction set. This is similar to what happened in 1985, when Intel introduced the 80386, the first 32-bit PC processor. The 386 was to give IBM and Microsoft a platform for an advanced 32-bit operating system that tapped this new power. To ensure immediate acceptance, the 386 and future 32-bit processors still ran 16-bit code. To take advantage of the 32-bit capability first found in the 386, new software would have to be written. Unfortunately, software evolves much more slowly than hardware. It took Microsoft a full 10 years after the 386 debuted to release Windows 95, the first mainstream 32-bit operating system for Intel processors.

Intel claims that won't happen with the P7. Despite that, it will likely take several years before the software market shifts to 64-bit operating systems and software. The installed base of 32-bit processors is simply too great, and the backward compatible 32-bit mode of the P7 will allow it to run 32-bit software very well, because it will be done in the hardware rather than through software emulation.

Itanium will use 0.18 micron technology for the initial Merced chips. This will allow Itanium to pack many more transistors in the same space. Early predictions have the Itanium sporting 100 million transistors!

Intel's initial goal with IA-64 is to dominate the workstation and server markets, competing with chips such as the Digital Alpha, Sun Sparc, and Motorola PowerPC. Microsoft will provide a version of Windows NT that runs on the P7, and Sun plans to provide a version of Solaris, its UNIX operating-system software, to support Itanium as well. NCR has already announced that it will build Itanium-powered systems that use Solaris.

Itanium will be available in a new package called the Pin Array Cartridge (PAC). This cartridge will include cache and will plug into a 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 3.58.) Itanium has clips on its sides, allowing four to be hung from a motherboard, both below and above.

Figure 3.58
Itanium processor.

Itanium will have three levels of cache. The L1 cache will be closely tied to the execution unit. It will be backed by on-die L2 cache. Finally the multimegabyte L3 cache will be housed in separate chips contained within the cartridge.

Itanium will be followed in late 2001 by a second IA-64 processor code-named McKinley. McKinley will have larger on-die L2 cache and target clock speeds of more than 1.5GHz, offering more than twice the performance of Itanium, according to Intel reps. Following McKinley will be Madison, based on 0.13 micron technology. Both Itanium and McKinley are based on 0.18 micron technology.

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.

To maximize your motherboard, you can almost always upgrade to the fastest processor your particular board will support. Normally, that can be determined by the type of socket on the motherboard. Table 3.43 lists the fastest processor upgrade solution for a given processor socket.

Table 3.43 Maximum Processor Speeds by Socket

Socket Type	Fastest Processor Supported
Socket 1	5x86–133MHz with 3.3v adapter
Socket 2	5x86–133MHz with 3.3v adapter
Socket 3	5x86–133MHz
Socket 4	Pentium OverDrive 133MHz
Socket 5	Pentium MMX 233MHz or AMD-K6 with 2.8v adapter
Socket 7	AMD-K6-2, K6-3, up to 550MHz
Socket 8	Pentium Pro OverDrive (333MHz Pentium II performance)
Socket 370	Celeron 600MHz (66MHz bus)
Socket 370	Pentium III 850MHz (100MHz bus)
Socket 370	Pentium III 1000MHz (133MHz bus)
Slot 1	Celeron 600MHz (66MHz bus)
Slot 1	Pentium III 850MHz (100MHz bus)
Slot 1	Pentium III 1000MHz (133MHz bus)
Slot 2	Pentium III Xeon 550MHz (100MHz bus)

For example, if your motherboard has a Pentium Socket 5, you can install a Pentium MMX 233MHz processor with a 2.8v voltage regulator adapter, or optionally an AMD-K6, also with a voltage regulator adapter. If you have Socket 7, your motherboard should be capable of supporting the lower voltage Pentium MMX or AMD-K6 series directly without any adapters. The K6-2 and K6-3 are the fastest and best processors for Socket 7 motherboards.

Rather than purchasing processors and adapters separately, I normally recommend you purchase them together in a module from companies such as Kingston or Evergreen (see the Vendor List on the CD).

Upgrading the processor can, in some cases, double the performance of a system, such as if you were going from a Pentium 100 to an MMX 233. However, if you already have a Pentium 233, you already have the fastest processor that goes in that socket. In that case, you really should look into a complete motherboard change, which would let you upgrade to a Pentium II processor at the same time. If your chassis design is not proprietary and your system uses an industry standard Baby-AT or 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 were normally 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 normally don’t recommend the OverDrive processors unless the deal is too good to pass up.

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 like 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. Though 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 normally be divided into two kinds: 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, while 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 that 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.

There are several companies that specialize in benchmark tests and software. The following table lists the company and the benchmarks they are known for. You can contact these companies via the information in the Vendor List on the CD.

Company	Benchmarks Published	Benchmark Type
Intel	iCOMP index 3.0	Processor
Intel	iCOMP index 3.0	System Intel Media Benchmark
Business Applications	SYSmark/NT	System
Performance Corporation	(BAPCo)	
Business Applications	SYSmark/NT, SYSmark95	System

Performance Corporation for Windows	(BAPCo)	
Standard Performance	SPECint95	Processor
Evaluation Corporation	(SPEC)	
Standard Performance	SPECint95,	Processor
Evaluation Corporation	SPECfp95	(SPEC)
Ziff-Davis Benchmark	CPUmark32	Processor Operation
Ziff-Davis Benchmark	Winstone 98	System Operation
Ziff-Davis Benchmark	WinBench 98	System Operation
Ziff-Davis Benchmark	CPUmark32, Winstone 98, WinBench 98, 3D WinBench 98	System Operation
Symantec Corporation	Norton SI32	Processor
Symantec Corporation	Norton SI32,	System
Norton Multimedia	Benchmark	

Processor Troubleshooting Techniques

Processors are normally very reliable. Most PC problems will be 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 3.44 provides a general troubleshooting checklist for processor-related PC problems.

Table 3.44 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.
Locks up during or shortly after POST	Poor Heat Dissipation	Check CPU heat sink/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 is not configured properly	Check manual and jumper board accordingly to proper bus and multiplier settings.
Operating system will not boot	Poor heat dissipation	Check CPU fan; replace if necessary, may need higher capacity heat sink.
	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	Monitor turned off or failed	Check monitor and power to monitor. Replace with

displayed		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 the processor that you have, and make sure that 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. Make 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 since there will normally be fewer problems.