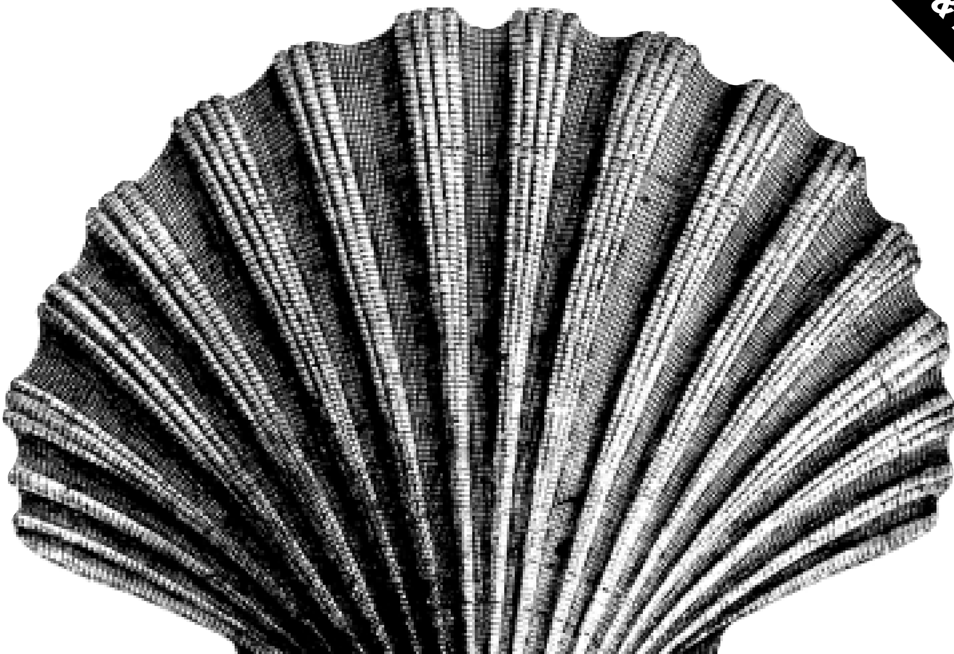
3rd Edition
Covers Windows & Linux



PC HARDWARE

IN A NUTSHELL

A Desktop Quick Reference

Robert Bruce Thompson
& Barbara Fritchman Thompson

O'REILLY®

PC HARDWARE

IN A NUTSHELL

Other resources from O'Reilly

oreilly.com *oreilly.com* is more than a complete catalog of O'Reilly books. You'll also find links to news, events, articles, weblogs, sample chapters, and code examples.

oreillynet.com is the essential portal for developers interested in open and emerging technologies, including new platforms, programming languages, and operating systems.

Conferences O'Reilly & Associates bring diverse innovators together to nurture the ideas that spark revolutionary industries. We specialize in documenting the latest tools and systems, translating the innovator's knowledge into useful skills for those in the trenches. Visit *conferences.oreilly.com* for our upcoming events.

Safari Bookshelf (*safari.oreilly.com*) is the premier online reference library for programmers and IT professionals. Conduct searches across more than 1,000 books. Subscribers can zero in on answers to time-critical questions in a matter of seconds. Read the books on your Bookshelf from cover to cover or simply flip to the page you need. Try it today with a free trial.

PC HARDWARE

IN A NUTSHELL

Third Edition

Robert Bruce Thompson and
Barbara Fritchman Thompson

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

4
Processors

The *processor*, also called the *microprocessor* or *CPU* (for *Central Processing Unit*), is the brain of the PC. It performs all general computing tasks and coordinates tasks done by memory, video, disk storage, and other system components. The CPU is a very complex chip that resides directly on the motherboard of most PCs, but may instead reside on a daughtercard that connects to the motherboard via a dedicated specialized slot.

Processor Design

A processor executes programs—including the operating system itself and user applications—all of which perform useful work. From the processor's point of view, a program is simply a group of low-level *instructions* that the processor executes more or less in sequence as it receives them. How efficiently and effectively the processor executes instructions is determined by its internal design, also called its *architecture*. The CPU architecture, in conjunction with CPU speed, determines how fast the CPU executes instructions of various types. The external design of the processor, specifically its external interfaces, determines how fast it communicates information back and forth with external cache, main memory, the chipset, and other system components.

Processor Components

Modern processors have the following internal components:

Execution unit
: The core of the CPU, the *execution unit* processes instructions.

Branch predictor
: The *branch predictor* attempts to guess where the program will jump (or branch) next, allowing the *prefetch and decode unit* to retrieve instructions

and data in advance so that they will already be available when the CPU requests them.

Floating-point unit
: The *floating-point unit* (*FPU*) is a specialized logic unit optimized to perform noninteger calculations much faster than the general-purpose logic unit can perform them.

Primary cache
: Also called *Level 1* or *L1 cache*, *primary cache* is a small amount of very fast memory that allows the CPU to retrieve data immediately, rather than waiting for slower main memory to respond. See Chapter 5 for more information about cache memory.

Bus interfaces
: *Bus interfaces* are the pathways that connect the processor to memory and other components. For example, modern processors connect to the chipset Northbridge via a dedicated bus called the *frontside bus* (*FSB)* or *host bus*.

Processor Speed

The processor *clock* coordinates all CPU and memory operations by periodically generating a time reference signal called a *clock cycle* or *tick*. Clock frequency is specified in *megahertz* (*MHz)*, which specifies millions of ticks per second, or *gigahertz* (*GHz*), which specifies billions of ticks per second. Clock speed determines how fast instructions execute. Some instructions require one tick, others multiple ticks, and some processors execute multiple instructions during one tick. The number of ticks per instruction varies according to processor architecture, its *instruction set*, and the specific instruction. *Complex Instruction Set Computer* (*CISC*) processors use complex instructions. Each requires many clock cycles to execute, but accomplishes a lot of work. *Reduced Instruction Set Computer* (*RISC*) processors use fewer, simpler instructions. Each takes few ticks but accomplishes relatively little work.

These differences in efficiency mean that one CPU cannot be directly compared to another purely on the basis of clock speed. For example, an AMD Athlon XP 3000+, which actually runs at 2.167 GHz, may be faster than an Intel Pentium 4 running at 3.06 GHz, depending on the application. The comparison is complicated because different CPUs have different strengths and weaknesses. For example, the Athlon is generally faster than the Pentium 4 clock for clock on both integer and floating-point operations (that is, it does more work per CPU tick), but the Pentium 4 has an extended instruction set that may allow it to run optimized software literally twice as fast as the Athlon. The only safe use of direct clock speed comparisons is within a single family. A 1.2 GHz Tualatin-core Pentium III, for example, is roughly 20% faster than a 1.0 GHz Tualatin-core Pentium III, but even there the relationship is not absolutely linear. And a 1.2 GHz Tualatin-core Pentium III is more than 20% faster than a 1.0 GHz Pentium III that uses the older Coppermine core. Also, even within a family, processors with similar names may differ substantially internally.

Processor Architecture

Clock speeds increase every year, but the laws of physics limit how fast CPUs can run. If designers depended only on faster clock speeds for better performance, CPU performance would have hit the wall years ago. Instead, designers have improved internal architectures while also increasing clock speeds. Recent CPUs run at more than 650 times the clock speed of the PC/XT's 8088 processor, but provide 6,500 or more times the performance. Here are some major architectural improvements that have allowed CPUs to continue to get faster every year:

Wider data busses and registers
: For a given clock speed, the amount of work done depends on the amount of data processed in one operation. Early CPUs processed data in 4-bit (*nibble*) or 8-bit (*byte*) chunks, whereas current CPUs process 32 or 64 bits per operation.

FPUs
: All CPUs work well with integers, but processing floating-point numbers to high precision on a general-purpose CPU requires a huge number of operations. All modern CPUs include a dedicated FPU that handles floating-point operations efficiently.

Pipelining
: Early CPUs took five ticks to process an instruction—one each to load the instruction, decode it, retrieve the data, execute the instruction, and write the result. Modern CPUs use *pipelining*, which dedicates a separate stage to each process and allows one full instruction to be executed per clock cycle.

Superscalar architecture
: If one pipeline is good, more are better. Using multiple pipelines allows multiple instructions to be processed in parallel, an architecture called *superscalar*. A superscalar processor processes multiple instructions per tick.

Intel Processors

Nearly all current PCs use either an Intel CPU or an Intel-compatible AMD Athlon CPU. The dominance of Intel in CPUs and Microsoft in operating systems gave rise to the hybrid term *Wintel*, which refers to systems that run Windows on an Intel or compatible CPU. Intel processors are referred to generically as *x86 processors*, based on Intel's early processor naming convention, 8086, 80186, 80286, etc. Intel has produced seven CPU generations, the first five of which are obsolete and the sixth obsolescent. They are as follows:

First generation
: The 8086 was Intel's first mainstream processor, and used 16 bits for both internal and external communications. The 8086 was first used in the late 1970s in dedicated word processors and minicomputers such as the DisplayWriter and the System/23 DataMaster. When IBM shipped its first PC in 1981, it used the 8088, an 8086 variant that used 16 bits internally but only 8 bits externally, because 8-bit peripherals were more readily available and less expensive then than were 16-bit components. The 8086 achieved prominence

much later when Compaq created the DeskPro as an improved clone of the IBM PC/XT. A few early PCs, notably Radio Shack models, were also built around the 80186 and 80188 CPUs, which were enhanced versions of the 8086 and 8088 respectively. The 8088 and 8086 CPUs did not include an FPU, although an *8087* FPU, called a *math coprocessor*, was available as an optional upgrade chip. First generation Intel CPUs (or their modern equivalents) are still used in some embedded applications, but they are long obsolete as general-purpose CPUs.

Second generation
: In 1982, Intel introduced the long-awaited follow-on to its first generation processors. The 80286, based on the iAPX-32 core, provided a quantum leap in processor performance, executing instructions as much as five times faster than an 808x processor running at the same clock speed. The 80286 processed instructions as fast as many mainframe processors of the time. The 80286 also increased addressable memory from 1 MB to 16 MB, and introduced *protected mode* operations. The IBM PC/AT was the first commercial implementation of the 80286. The optional *80287* FPU chip added floating-point acceleration to 80286 systems. Although long obsolete as a general-purpose CPU, the 80286 is still used in embedded controllers.

Third generation
: Intel's next generation debuted in 1985 as the 80386, later shortened to just 386. The 386 was Intel's first 32-bit CPU, which communicated internally and externally with a 32-bit data bus and 32-bit address bus. The 386 was available in 16, 20, 25, and 33 MHz versions. Although 386 clock speeds were only slightly faster than those of the 80286, improved architecture resulted in significant performance increases. The optional *80387* FPU added floating-point acceleration to 386 systems. Intel later renamed the 386 to the 386DX and released a cheaper version called the 386SX, which used 32 bits internally but only 16 bits externally. The 386SX was notable as the first Intel processor that included an internal (L1) cache, although it was only 8 KB and relatively inefficient. The 386 is long obsolete as a general-purpose CPU, but it is still commonly used in embedded controllers.

Fourth generation
: Intel's next generation debuted in 1989 as the 486 (there never was an 80486). The 486 was a full 32-bit CPU with 8 KB of L1 cache, included a built-in FPU, and was available in speeds from 20 MHz to 50 MHz. Intel released 486DX and 486SX versions. The 486SX was in fact a 486DX with the FPU disabled. Intel also sold the 487SX, which was actually a full-blown 486DX. Installing a 487SX in the coprocessor socket simply disabled the existing 486SX. The 486DX/2, introduced in 1992, was the first Intel processor that ran internally at a multiple of the memory bus speed. The 486DX/2 clock ran at twice bus speed, and was available in 25/50, 33/66, and 40/80 MHz versions. The 486DX/4, introduced in 1994, ran (despite its name) at thrice bus speed, doubled L1 cache to 16 KB, and was available in 25/75, 33/100, and 40/120 versions. The 486 is obsolete as a general-purpose CPU, although it is still popular in embedded applications.

Fifth generation

The Intel Pentium CPU defines the fifth generation. It provides much better performance than its 486 ancestors by incorporating several architectural improvements, most notably an increase in data bus width from 32 bits to 64 bits and an increase in CPU memory bus speed from 33 MHz to 60 and 66 MHz. Intel actually shipped several different versions of the Pentium, including:

- *Pentium P54*—the original Pentium shipped in 1993 in 50, 60, and 66 MHz versions using a 1X CPU multiplier, ran (hot) at 5.0 volts, contained a dual 8 KB + 8 KB L1 cache, and fit Socket 4 motherboards.
- *Pentium P54C*—the "Classic Pentium" first shipped in 1994, was available in speeds from 75 to 200 MHz using CPU multipliers from 1.5 to 3.0, used 3.3 volts, and contained the same dual L1 cache as the P54. P54C CPUs fit Socket 5 motherboards and most Socket 7 motherboards.
- *Pentium P55C*—the Pentium/MMX shipped in 1997, was available in speeds from 166 to 233 MHz, using CPU multipliers from 2.5 to 3.5, used 3.3 volts, and contained a dual 16 KB + 16 KB L1 cache, twice the size of earlier Pentiums. The other major change from the P54C was the addition of the MMX instruction set, a set of additional instructions that greatly improved graphics processing speed. P55C CPUs fit Socket 7 motherboards, and are still in limited distribution as of July 2003.

The Pentium and other fifth-generation processors are obsolete, although millions of Pentium systems remain in service. Any system that uses a fifth-generation processor is too old to upgrade economically.

Sixth generation

This generation began with the 1995 introduction of the Pentium Pro, and includes the Pentium II, Celeron, and Pentium III processors. Late sixth-generation Intel desktop processors had been relegated to entry-level systems by early 2002 and had been discontinued as mainstream products by mid-2002. By late 2002, only the Tualatin-core Celeron processors remained as representatives of this generation. Although it is still technically feasible to upgrade the processor in many sixth-generation systems, in practical terms it usually makes more sense to replace the motherboard and processor with seventh-generation products.

Seventh generation

This is the current generation of Intel processors, and includes Intel's flagship Pentium 4 as well as various Celeron processors based on the Pentium 4 architecture.

Intel currently manufactures several sixth-generation processors, including numerous variants and derivatives of the Celeron and Pentium III, and two seventh-generation processors, the Pentium 4 and the Celeron. The following sections describe current and recent Intel processors.

> There are times when it is essential to identify the processor a system uses. For information about identifying Intel processors, see *http://www.hardwareguys.com/supplement/cpu-id.html*.

Pentium, Pentium/MMX

Intel originally designated its processors by number rather than by name—Intel 8086, 8088, 80186, 80286, and so on. Intel dropped the "80" prefix early in the life cycle of the 80386, relabeling it as the 386. (Intel never made an "80486" processor despite what some people believe.) By the time Intel shipped its fourth-generation processors, it was tired of other makers using similar names for their compatible processors. Intel believed that these similar names could lead to confusion among customers, and so tried to trademark its X86 naming scheme. When Intel learned that part numbers cannot be trademarked, the company decided to drop the "86" naming scheme and create a made-up word to name its fifth generation processors. Intel came up with *Pentium*.

Intel has produced the following three major subgenerations of Pentium:

P54
: These earliest Pentium CPUs, first shipped in March 1993, fit Socket 4 motherboards, use a 3.1 million transistor core, have 16 KB L1 cache, and use 5.0 volts for both core and I/O components. P54-based systems use a 50, 60, or 66 MHz memory bus and a fixed 1.0 CPU multiplier to yield processor speeds of 50, 60, or 66 MHz.

P54C
: The so-called *Classic Pentium* CPUs, first shipped in October 1994, fit Socket 5 and most Socket 7 motherboards, use a 3.3 million transistor core, have 16 KB L1 cache, and generally use 3.3 volts for both core and I/O components. P54C-based systems use a 50, 60, or 66 MHz memory bus and CPU multipliers of 1.5, 2.0, 2.5, and 3.0x to yield processor speeds of 75, 90, 100, 120, 133, 150, 166, and 200 MHz.

P55C
: The *Pentium/MMX* CPUs (shown in Figure 4-1), first shipped in January 1997, fit Socket 7 motherboards, use a 4.1 million transistor core, have a 32 KB L1 cache, feature improved branch prediction logic, and generally use a 2.8 volt core and 3.3 volt I/O components. P55C-based systems use a 60 or 66 MHz memory bus and CPU multipliers of 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0x to yield processor speeds of 120, 133, 150, 166, 200, 233, 266, and 300 MHz.

The Pentium was a quantum leap from the 486 in complexity and architectural efficiency. It is a CISC processor, and was initially built on a 0.35 micron process (later 0.25 micron). Pentiums, like 486s, use 32-bit operations internally. Externally, however, the Pentium doubles the 32-bit 486 data bus to 64 bits, allowing it to access eight full bytes at a time from memory. With the Pentium, Intel also introduced new chipsets to support this wider data bus and other Pentium enhancements.

The Pentium uses a *dual-pipelined superscalar* design which, relative to the 486 and earlier CPUs, allows it to execute more instructions per clock cycle. The Pentium executes integer instructions using the same five stages as the 486—*Prefetch, Instruction Decode, Address Generate, Execute,* and *Write Back*—but the Pentium has two parallel integer pipelines versus the 486's one, which allows the Pentium to execute two integer operations simultaneously in parallel. This means

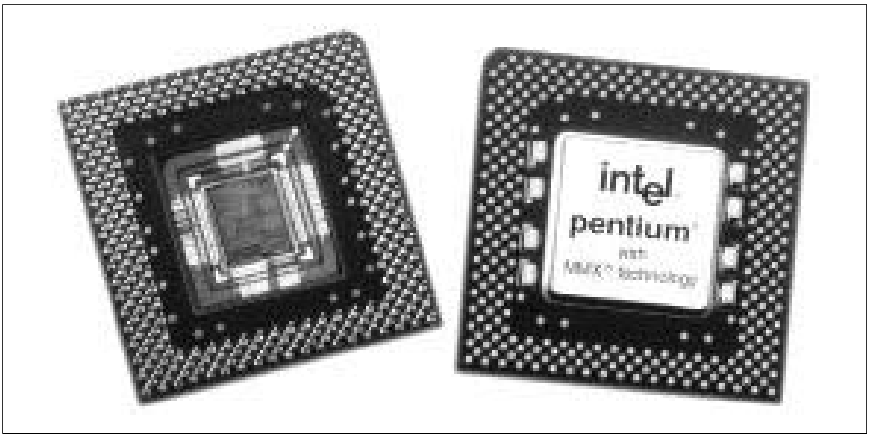


Figure 4-1. Intel Pentium/MMX processor (photo courtesy of Intel Corporation)

that, for equal clock speeds, the Pentium processes integer instructions about twice as fast as a 486.

The Pentium includes an improved 80-bit FPU that is much more efficient than the 486 FPU. The Pentium also includes a *Branch Target Buffer* to provide dynamic branch prediction, a process that greatly enhances instruction execution efficiency. Finally, the Pentium includes a *System Management Module* that can control power use by the processor and peripherals.

P54 Pentiums also improved upon 486 L1 caching. The 486 has one 8 KB L1 cache (16 KB for the 486DX/4) that uses the inefficient *write-through* algorithm. P54 and P54C Pentiums have dual 8 KB L1 caches—one for data and one for instructions—that use the much more efficient *two-way set associative write-back* algorithm. This doubling of L1 cache buffers and the improved caching algorithm combined to greatly enhance CPU performance. P55C Pentiums double L1 cache size to 16 KB, providing still more improvement.

The changes from the P54 to the P54C were relatively minor. Higher voltages and faster CPU speeds generate more heat, so Intel reduced the core and I/O voltages from 5.0/5.0V in the P54 to 3.3/3.3V in the P54C, allowing them to run the CPUs faster without excessive heating. Intel also introduced support for CPU multipliers, which allow the CPU to run internally at some multiple of the memory bus speed.

The changes from the P54C Classic to the P55C MMX were much more significant. In fact, had Intel not already introduced the Pentium Pro (its first sixth-generation CPU) before the P55C, the P55C might have been considered the first of a new CPU generation. In addition to doubling L1 cache size, the P55C incorporated two major architectural enhancements:

MMX
: Although sometimes described as *MultiMedia eXtensions* or *Matrix Math eXtensions*, Intel says officially that *MMX* stands for nothing. MMX is a set of 57 added instructions that are dedicated to manipulating audio, video, and graphics data more efficiently.

SIMD

Single Instruction Multiple Data (SIMD) is an architectural enhancement that allows one instruction to operate simultaneously on multiple sets of similar data.

In conjunction, MMX and SIMD greatly extend the Pentium's ability to perform parallel operations, processing 8 bytes of data per clock cycle rather than 1 byte. This is particularly important for heavily graphics-oriented operations such as video because it allows the P55C to retrieve and process eight 1-byte pixels in one operation rather than manipulating those 8 bytes as 8 separate operations. Intel estimates that MMX and SIMD used with nonoptimized software yield performance increases of as much as 20%, and can yield increases of 60% when used with MMX-aware applications.

Although the Pentium is technically obsolete, millions of Pentium systems remain in service as Linux firewalls or as dedicated appliance servers, and a significant number of them continue to be upgraded. As of July 2003 Intel still produced the Pentium/200 and /233 MMX processors in Socket 7, as well as several slower models for embedded applications. For additional information about Pentium processors, including detailed identification tables, visit *http://developer.intel.com/design/pentium/*.

Pentium Pro

Intel's first sixth-generation CPU, the Pentium Pro, was introduced in November 1995—along with the new 3.3 volt 387-pin Socket 8 motherboards required to accept it—and was discontinued in late 1998. Pentium Pro processors are no longer made, but remain available on the used market. Intel positioned the Pentium Pro for servers, a niche it never escaped, and where it continued to sell in shrinking numbers until its replacement, the Pentium II Xeon, shipped in mid-1998. The Pentium Pro predated the P55C Pentium/MMX, and never shipped in an MMX version. The Pentium Pro never sold in large numbers for two reasons:

Cost

The Pentium Pro was a very expensive processor to build. Its core logic comprised 5.5 million transistors (versus 4.1 million in the P55C), but the real problem was that the Pentium Pro also included a large L2 cache on the same substrate as the CPU. This L2 cache required millions of additional transistors, which in turn required a much larger die size and resulted in a much lower percentage yield of usable processors, both factors that kept Pentium Pro prices very high relative to other Intel CPUs.

32-bit optimization

The Pentium Pro was optimized to execute 32-bit operations efficiently at the expense of 16-bit performance. For servers, 32-bit optimization is ideal, but slow 16-bit operations meant that a Pentium Pro actually ran many Windows 95 client applications slower than a Pentium running at the same clock speed.

The Pentium Pro shipped in 133, 150, 166, 180, and 200 MHz versions with 256 KB, 512 KB, or 1 MB of L2 cache, and was never upgraded to a faster version. The Pentium Pro continued to sell long after the introduction of much faster Pentium II CPUs for only one reason: the first Pentium II chipsets supported only two-way

Symmetric Multiprocessing (SMP) while Pentium Pro chipsets supported four-way SMP. In some server environments, four 200 MHz Pentium Pro CPUs outperformed two 450 MHz Pentium II CPUs. The introduction of the 450NX chipset, which supports four-way SMP, and the mid-1998 introduction of the Pentium II Xeon processor, which supports eight-way SMP, removed the raison d'être for the Pentium Pro, and it died a quick death.

Pentium Pro processor architecture

Although the Pentium Pro is obsolete, it was the first Intel sixth-generation processor, and as such introduced many important architectural improvements. Understanding the Pentium Pro vis-à-vis the Pentium will help you understand current Intel CPU models. The two CPUs differ in the following major respects:

Secondary (L2) cache
: Pentium-based systems may optionally be equipped with an external L2 secondary cache of any size supported by the chipset. Typical Pentium systems have a 256 KB L2 cache, but high-performance motherboards may include a 512 KB, 1 MB, or larger L2 cache. But Pentium L2 caches use a narrow (32-bit), slow (60 or 66 MHz memory bus speed) link between the processor's L1 cache and the L2 cache. The Pentium Pro L2 cache is internal, located on the CPU itself, and the Pentium Pro uses a 64-bit data path running at full processor speed to link L1 cache to L2 cache. The dedicated high-speed bus used to connect to cache is called the *Backside Bus* (*BSB*), as opposed to the traditional CPU-to-chipset bus, which is now designated the *Frontside Bus* (*FSB*). In conjunction, the BSB and FSB are called the *Dual Independent Bus* (*DIB*) architecture. DIB architecture yields dramatically improved cache performance. In effect, 256 KB of Pentium Pro L2 cache provides about the same performance boost as 2 MB or more of Pentium L2 cache.

Dynamic execution
: The Pentium Pro uses a combination of techniques—including *branch prediction*, *data flow analysis*, and *speculative execution*—that collectively are referred to as *dynamic execution*. Using these techniques, the Pentium Pro productively uses clock cycles that would otherwise be wasted, as they are with the Pentium.

Super-pipelining
: *Super-pipelining* is a technique that allows the Pentium Pro to use *out-of-order instruction execution*, another method to avoid wasting clock cycles. The Pentium executes instructions on a first-come, first-served basis, which means that it waits for all required data to process an earlier instruction instead of processing a later instruction for which it already has all of the data. Because it uses *linear instruction sequencing*, or *standard pipelining*, the Pentium wastes what could otherwise be productive clock cycles executing no-op instructions. The Pentium Pro is the first Intel CPU to use super-pipelining. It has a 14-stage pipeline, divided into three sections. The first section, the *in-order front end*, comprises eight stages, and decodes and issues instructions. The second section, the *out-of-order core*, comprises three stages, and executes instructions in the most efficient order possible based on available

data, regardless of the order in which it received the instructions. The third and final section, the *in-order retirement section*, receives and forwards the results of the second section.

CISC versus RISC core

The most significant architectural difference between the Pentium and the sixth-generation processors is how they handle instructions internally. Pentiums use a *Complex Instruction Set Computer* (*CISC*) core. CISC means that the processor understands a large number of complicated instructions, each of which accomplishes a common task in just one instruction. The Pentium Pro was the first Intel CPU to use a *Reduced Instruction Set Computer* (*RISC*) core. RISC means that the processor understands only a few simple instructions. Complex operations are performed by stringing together multiple simple instructions. Although RISC CPUs must perform many simple instructions to accomplish the same task that CISC CPUs do with just one or a few complex instructions, the simple RISC instructions execute much faster than CISC instructions.

The Pentium Pro translates standard Intel x86 CISC instructions into RISC instructions that the Pentium Pro microcode uses internally, and then passes those RISC instructions to the internal out-of-order execution core. This translation helps avoid limitations of the standard x86 CISC instruction set, and supports the out-of-order execution that prevents pipeline stalls, but those benefits come at a price. Although the time required is measured in nanoseconds, converting from CISC to RISC does take time, and that slows program execution. Also, 16-bit instructions convert inefficiently and frequently result in pipeline stalls in the out-of-order execution unit, which commonly result in CPU wait states of as many as seven clock cycles. The upshot is that, for pure 32-bit operations, the benefit of RISC conversion greatly outweighs the drawbacks, but for 16-bit operations, the converse is true.

For additional information about Pentium Pro processors, including detailed identification tables, visit *http://developer.intel.com/design/pro/*.

Pentium II Family

Intel's first mainstream sixth-generation CPU, the Pentium II, shipped in May 1997. Intel subsequently shipped many variants of the Pentium II, which differ chiefly in packaging, the type and amount of L2 cache they include, the processor core they use, and the FSB speeds they support. All members of the Pentium II family use the Dynamic Execution Technology and DIB architecture introduced with the Pentium Pro. Intel reduced the core voltage from the 3.3 volts used by Pentium Pro to 2.8 volts or less in Pentium II processors, which allows them to run much faster while using less power and producing less heat. In effect, you're not far wrong if you think of Pentium II, sixth-generation Celeron, and Pentium III processors as faster versions of the Pentium Pro with MMX (or the enhanced *SSE* version of MMX) added, and the following major changes:

L2 cache

The Pentium Pro taught Intel the folly of embedding the L2 cache onto the CPU substrate itself, at least for the then-current state of the technology.

Early Pentium II family processors use discrete L2 cache *Static RAM* (*SRAM*) chips that reside within the CPU package but are not a part of the CPU substrate. Advances in fab technology have allowed Intel again to place L2 cache directly on the processor substrate on later Pentium II family processor models. Some Pentium II family processors run L2 cache at full processor speed, while others run it at half processor speed. The least-expensive Pentium II family processors have no L2 cache at all. The L2 cache in later members of the Pentium II family is improved, not just in size and/or speed, but also in functionality. The most recent Pentium III processors, for example, use an *eight-way set associative cache*, which is more efficient than the caching schemes used on earlier variants.

Packaging
: The Pentium Pro used the huge, complicated 387-pin *Dual Pattern-Staggered Pin Grid Array* (*DP-SPGA*) Socket 8. The extra pins provide data and power lines for the onboard L2 cache. Intel developed simplified alternative packaging methods for various members of the Pentium II family processors, which are described later in this chapter.

Improved 16-bit performance
: High cost aside, the major reason the Pentium Pro was never widely used other than in servers was its poor performance with 16-bit software. Although represented as a 32-bit operating system, Windows 95/98 still contains much 16-bit code. Users quickly discovered that Windows 95 actually ran slower on a Pentium Pro than on a Pentium of the same speed. Intel solved the 16-bit problem by using the Pentium segment descriptor cache in the Pentium II.

Members of the Pentium II family include the Pentium II, Pentium II Overdrive, Pentium II Xeon, sixth-generation Celeron, Pentium III, and Pentium III Xeon. Each of these processors is described in the following sections.

Pentium II

First-generation Pentium II processors shipped in 233, 266, 300, and 333 MHz versions with the Klamath core and a 66 MHz FSB. In mid-1998, Intel shipped second-generation Pentium II processors, based on the Deschutes core, that ran at 350, 400, and 450 MHz and used a 100 MHz FSB. Pentium II processors have 512 KB of L2 cache that runs at half internal CPU speed versus 256 KB to 1 MB of full CPU speed L2 cache in the Pentium Pro. Pentium II processors use a *Single Edge Contact connector* (*SECC*) or *SECC2* cartridge, which contains the CPU and L2 cache (see Figure 4-2). The SECC/SECC2 package mates with a *242-contact slot connector*, formerly known as *Slot 1*, which resembles a standard expansion slot. Klamath-based processors run at 2.8 volts and are built on a 0.35μ fab. Deschutes-based processors, including all 100 MHz FSB processors and recent 66 MHz FSB processors, run at 2.0 volts and are built on a 0.25μ fab. Excepting FSB speed and fab process, all Slot 1 Pentium II processors are functionally identical. As of July 2003, Pentium II processors remain in limited distribution, but they are obsolescent.

Figure 4-2. Intel Pentium II processor in the original SECC package (photo courtesy of Intel Corporation)

For additional information about Pentium II processors, including detailed identification tables, visit *http://developer.intel.com/design/pentiumii/*. For information about the Pentium II Xeon processor, see *http://www.intel.com/support/processors/pentiumii/xeon/*.

Celeron

The sixth-generation Celeron—we keep saying "sixth-generation" because Intel also makes a seventh-generation Celeron based on the Pentium 4—was initially an inexpensive variant of the Pentium II and, in later models, an inexpensive variant of the Pentium III. Klamath-based (Covington-core) Celerons shipped in April 1998 in 266 and 300 MHz versions without L2 cache. Performance was poor, so in fall 1998 Intel began shipping modified Deschutes-based (Mendocino-core) Celerons with 128 KB L2 cache. The smaller Celeron L2 cache runs at full CPU speed, and provides L2 cache performance similar to that of the larger but slower Pentium II L2 cache for most applications. Mendocino (0.25μ) Celerons have been manufactured in 300A (to differentiate it from the cacheless 300), 333, 366, 400, 433, 466, 500, and 533 MHz versions, all of which use the 66 MHz FSB.

With the introduction of the Coppermine-core Pentium III processor, Intel also introduced Celeron processors based on a variant of the Coppermine core called the *Coppermine128* core. Celerons based on this 0.18μ, 1.6v core began shipping in 533A, 566, and 600 MHz versions soon after their announcement in May 2000, and were eventually produced in speeds as high as 1.1 GHz, which approaches the limit of the Coppermine core itself.

Coppermine128-core Celerons have half of the 256 KB on-die L2 cache disabled to bring L2 cache size to the Celeron-standard 128 KB, and use a four-way set associate L2 cache rather than the eight-way version used by the Coppermine Pentium III. Coppermine128-core Celerons through the Celeron/766, shipped in November 2000, use the 66 MHz FSB speed. Coppermine128-core Celerons that use the 100 MHz FSB speed began shipping in March 2001, beginning with 800 MHz units and eventually reaching 1.1 GHz. Other than the differences in L2 cache size and type, processor bus speed differences, and official support for SMP, Coppermine128-core Celerons support the standard Coppermine-core Pentium III features, including SSE, described later in this chapter.

> Because Coppermine128 Celerons effectively *are* Pentium IIIs, some *may* be easy to overclock. For example, a Celeron/600 (66 MHz FSB) is effectively a down-rated Pentium III/900 (100 MHz FSB). During the ramp-up of the Coppermine128-core Celerons, we believe that Intel recycled Pentium III processors that tested as unreliable at 100 MHz or 133 MHz as 66 MHz Celerons, although Intel has never confirmed this. Many early Coppermine128-core Celerons were not good overclockers, although that changed as production ramped up. Note, however, that overclocking Coppermine128-core Celerons is viable only for the slower 66 MHz FSB models—the Celeron/566 and /600. Attempting to overclock a faster Celeron by running it with a 100 MHz FSB would cause it to run near or over 1.1 GHz, which appears to be the effective limit of the Coppermine core itself.

In November 2001, Intel began shipping Celerons based on the latest Pentium III core, code-named *Tualatin*. The first Tualatin-core Celerons ran at 1.2 GHz using the 100 MHz FSB. Intel later filled in the product line by shipping 100 MHz FSB Tualatin-core Celerons at 900 MHz, 1.0 GHz, 1.1 GHz, 1.3 GHz, and finally 1.4 GHz. Tualatin-core Celerons also differ from earlier Celeron models in that they include a full 256 KB eight-way set associative L2 cache, the same as Coppermine-core Pentium III models. Tualatin-core Celerons perform like full-blown Pentium IIIs because they effectively *are* full-blown Pentium IIIs.

So why did Intel suddenly decide to uncripple the Celeron? Basically, it had devoted a lot of resources to developing the Tualatin-core Pentium III only to find itself overtaken by events. Intel needed to ship the Pentium 4 to counter fast AMD Athlons, but there was no room in Intel's lineup for two premium processors. Accordingly, the Pentium III had to go, at least as mainstream product, giving way to the new-generation Pentium 4. But that left Intel with the perfectly good, new Tualatin core, which had been developed at great expense, with no way to sell it. Talk about being all dressed up with nowhere to go.

As a way of earning back the development costs of the Tualatin core while at the same time putting the screws to AMD's low-end Duron, Intel decided to ship Pentium III processors with the Celeron name on them. The new Celerons handily outperformed Durons running at the same clock speed, and in fact were surprisingly close to the performance level of the fastest Pentium 4 and Athlon processors then available. Selling for less than $100, the Tualatin-core Celerons provided incredibly high bang for the buck. In fact, they still do today. A Celeron/1.4G running in an 815-based motherboard is slower than a fast Pentium 4 and Athlon system, certainly, but is by no means a slow system.

Celerons have been produced in four form factors:

Single Edge Processor Package cartridge
: All Celerons through 433 MHz were produced in *Single Edge Processor Package* (*SEPP*) cartridge form, which resembles the Pentium II SECC and

SECC2 package, and is compatible with the Pentium II 242-contact slot. In mid-1999 Intel largely abandoned SEPP in favor of PPGA, and SEPP Celerons are no longer available new. Figure 4-3 shows an SEPP Celeron.

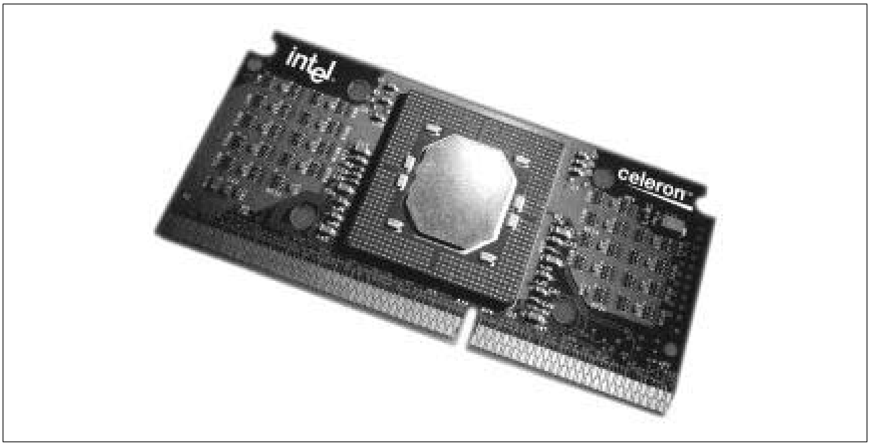


Figure 4-3. Intel Celeron processor in SEPP package (photo courtesy of Intel Corporation)

Plastic Pin Grid Array

As a cheaper alternative to SEPP, Intel developed the *Plastic Pin Grid Array* (*PPGA*). PPGA processors fit *Socket 370*, which resembles Socket 7 but accepts only PPGA Celeron and Pentium III processors. All Mendocino-core Celerons are manufactured in PPGA. The Celeron/466 was the first Celeron produced only in PPGA. PPGA processors can be used in most Socket 370 motherboards, although a few accept only Socket 370 Pentium III processors. PPGA Celerons are no longer available new. Figure 4-4 shows a PPGA Celeron.

Flip Chip Pin Grid Array

With the introduction of the Socket 370 version of the Pentium III, Intel introduced a modified version of PPGA called *Flip Chip PGA* (*FC-PGA*), which uses slightly different pinouts than PPGA. FC-PGA essentially reverses the position of the processor core from PPGA, placing the core on top (where it can make better contact with the heatsink) rather than on the bottom side with the pins. All Socket 370 Pentium III and Coppermine128-core Celerons (the 533A, 566, 600, and faster versions) require an FC-PGA compliant motherboard. FC-PGA processors physically fit older PPGA motherboards, but if you install an FC-PGA processor in a PPGA-only Socket 370 motherboard the processor doesn't work, although no harm is done. FC-PGA Celerons are no longer available new. Figure 4-5 shows an FC-PGA Celeron.

Flip Chip Pin Grid Array 2

Tualatin-core Celerons use the FC-PGA2 packaging, which is essentially FC-PGA with the addition of a flat metal plate, called an *Integrated Heat Spreader*, that covers the processor chip itself. Although these processors physically fit any Socket 370 motherboard, only very recent Socket 370 chipsets support the Tualatin core. Intel designates its own motherboard

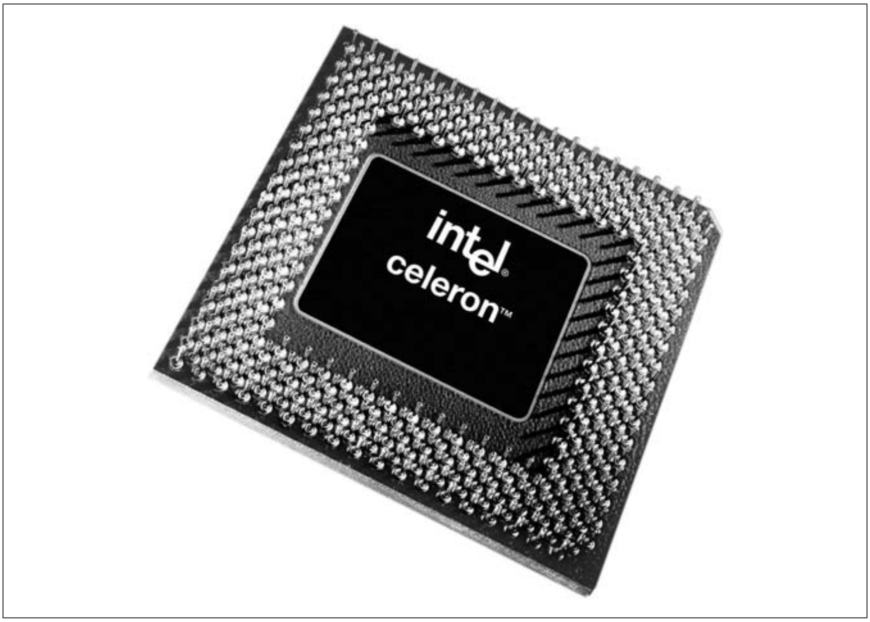


Figure 4-4. Intel Celeron processor in PPGA package (photo courtesy of Intel Corporation)

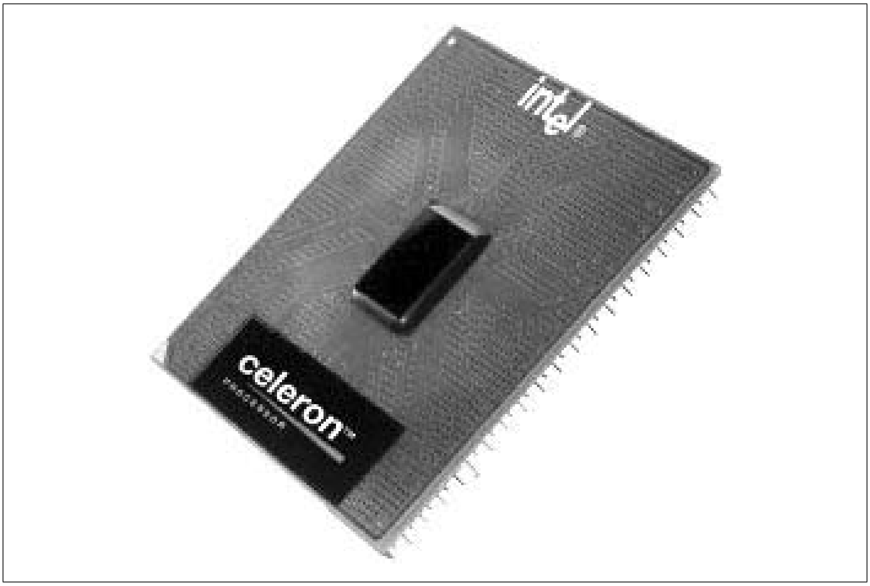


Figure 4-5. Intel Celeron processor in FC-PGA package (photo courtesy of Intel Corporation)

models that support Tualatin as "Universal" models. Other manufacturers use other terminology, but the important thing to remember is that the motherboard must explicitly support Tualatin if it is to run these processors. As of

July 2003, Intel still produces FC-PGA Celerons in 1.0, 1.1, 1.2, 1.3, and 1.4 GHz models. Figure 4-6 shows an FC-PGA2 Celeron.

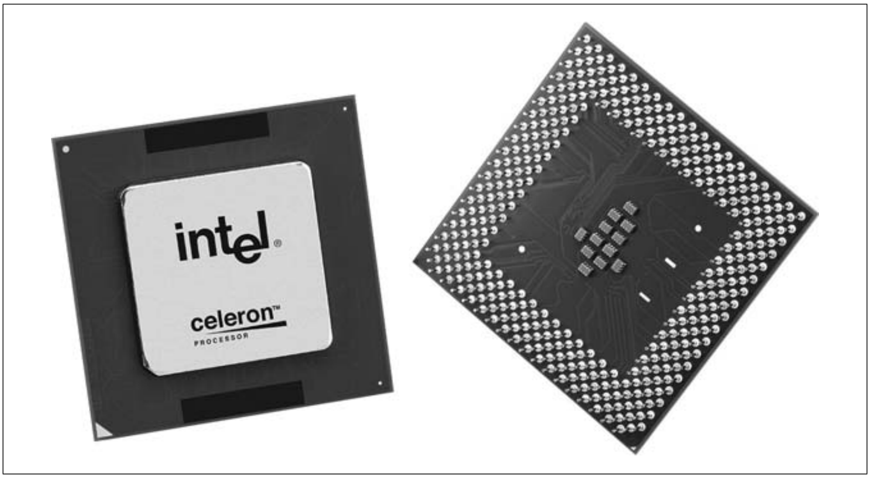


Figure 4-6. Intel Celeron processor in FC-PGA2 package (photo courtesy of Intel Corporation)

Intel has produced five major variants of the Celeron, using four packages, four cores, two bus speeds, four fab sizes, and more than 20 clock speeds. Table 4-1 summarizes the major differences between these variants.

Table 4-1. Comparison of sixth-generation Celeron variants

	Covington	Mendocino	Coppermine128	Coppermine128	Tualatin
Package	SECC	SECC-2 PPGA	FC-PGA	FC-PGA	FC-PGA2
Manufacturing dates	1998	1998 – 2000	2000 – 2002	2001 – 2002	2001 –
Clock speeds (MHz)	266, 300	300A, 333, 366, 400, 433, 466, 500, 533	500A, 533A, 566, 600, 633, 667, 700, 733, 766	800, 850, 900, 950, 1000, 1100	900, 1000, 1100, 1200, 1300, 1400
L2 cache size	none	128 KB	128 KB	128 KB	256 KB
L2 cache bus width	n/a	64 bits	256 bits	256 bits	256 bits
System bus speed	66 MHz	66 MHz	66 MHz	100 MHz	100 MHz
SSE instructions	❍	❍	●	●	●
Dual CPU capable	●	●	❍	❍	❍
Fabrication process	0.35μ	0.25μ	0.18μ	0.18μ	0.13μ

Dual-CPU capability deserves an explanation. Although Intel never officially supported Celerons for SMP operation, the two earliest Celeron variants did in fact support dual-CPU operation. For Covington-core and SECC-2 Mendocino-core Celerons, dual-CPU operation was impractical because enabling SMP required physical surgery on the processor package—literally drilling holes in the package and soldering wires. With PPGA Mendocino-core Celerons, dual-CPU operation was eminently practical because many dual Socket 370 motherboards were designed specifically to accept two Celerons, and no changes to the processors themselves were necessary. Beginning with the 66 MHz Coppermine128 Celerons, Intel physically disabled SMP operation in the core itself, so it is impossible to operate Coppermine- or Tualatin-core Celerons in SMP mode.

For additional information about Celeron processors, including detailed identification tables, visit *http://developer.intel.com/design/celeron/*.

Pentium III

The Pentium III, Intel's final sixth-generation processor, began shipping in February 1999. The Pentium III has been manufactured in numerous variants, including speeds from 450 MHz to 1.4 GHz (Intel defines 1 GHz as 1000 MHz), two bus speeds (100 MHz and 133 MHz), four packages (SECC, SECC2, FC-PGA, and FC-PGA2), and the following three cores:

Pentium III (Katmai core)

Initial Pentium III variants use the *Katmai core*, essentially an enhanced Deschutes with the addition of 70 new *Streaming SIMD instructions* (formerly called *Katmai New Instructions* or *KNI* and known colloquially as *MMX/2*) that improve 3D graphics rendering and speech processing. They use the 0. 25µ process, operate at 2.0V core voltage (with some versions requiring marginally higher voltage), use a 100 or 133 MHz FSB, incorporate 512 KB four-way set associative L2 cache running at half CPU speed, and have glueless support for two-way SMP. Katmai-core processors were made in SECC2 (Slot 1/SC242) at 450, 500, 550, and 600 MHz in 100 MHz FSB variants, and at 533 and 600 MHz in 133 MHz FSB variants.

Pentium III (Coppermine core)

Later Pentium III variants use the *Coppermine core*, which is essentially a refined version of the Katmai core. Later Coppermine processors use the updated *Coppermine-T core*. Coppermine processors use the 0.18µ process, which reduces die size, heat production, and cost. They operate at nominal 1.6V core voltage (with faster versions requiring marginally higher voltage), are available at either 100 MHz or 133 MHz FSB, and (in most variants) support SMP. Coppermine-core processors have been made in SECC2 (Slot 1/SC242) and FC-PGA (Socket 370) packaging in both 100 and 133 MHz FSB variants, running at speeds from 533 MHz to 1.13 GHz. Finally, Coppermine also incorporates the following significant improvements in L2 cache implementation and buffering:

Advanced Transfer Cache

Advanced Transfer Cache (*ATC*) is how Intel summarizes the several important improvements in L2 cache implementation from Katmai to Coppermine. Although L2 cache size is reduced from 512 KB to 256 KB, it is now on-die (rather than discrete SRAM chips) and, like the Celeron, operates at full CPU speed rather than half. Bandwidth is also quadrupled, from the 64-bit bus used on Katmai- and Mendocino-core Celeron processors to a 256-bit bus. Finally, Coppermine uses an eight-way set associative cache, rather than the four-way set associative cache used by earlier Pentium III and Celeron processors. Migrating L2 cache on-die increased transistor count from just under 10 million for the Katmai to nearly 30 million for Coppermine, which may account for the reported early yield problems with the Coppermine.

> When manufacturers begin producing a processor, a relatively high percentage of the processors made are unusable. In the initial phases, many of the processors on each wafer may be spoiled. As the manufacturer ramps up production and gains experience, the percentage of usable processors increases substantially, as does the percentage of processors that are usable at higher speeds. Marketing reasons aside, yield percentage is the major factor in the very high price of the fastest processors. During early production, only 1% to 10% of the processors produced may be able to run at the highest speed offered for that processor. As the yield percentage improves, manufacturers can cut processor prices. Yield percentages are one of the most closely guarded secrets in semiconductor manufacturing.

Advanced System Buffering

Advanced System Buffering (*ASB*) is how Intel describes the increase from Pentium III Katmai and earlier processors to the Coppermine from four to six fill buffers, four to eight queue entry buffers, and one to four write-back buffers. The increased number of buffers was primarily intended to prevent bottlenecks with 133 MHz FSB Coppermines, but also benefits those running at 100 MHz.

Pentium III (Tualatin core)

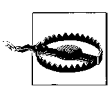
The latest Pentium III variants use the *Tualatin core*, which is the last Pentium III core Intel will ever produce. Tualatin processors use the 0.13μ process, which reduces die size, heat production, and cost, and allows considerably higher clock speeds than the Coppermine core. Had it not been for Intel's rapid transition to the Pentium 4, Tualatin-core Pentium IIIs could have been Intel's flagship processor through at least the end of 2002. Intel could have shipped Tualatins at ever-increasing clock speeds, beating the 0.18μ Palomino-core AMD Athlon on both clock speed and actual performance. Instead, Intel opted to compete using the Pentium 4. Intel has by its pricing mechanism effectively exiled Tualatin-core Pentium IIIs to niche status by selling fast Pentium 4 processors for less than Tualatin Pentium IIIs with comparable performance.

Tualatins use the 133 MHz FSB, and are available in two major variants, both of which use the FC-PGA2 packaging (with Integrated Heat Spreader). The first variant, intended for desktop systems, has the standard 256 KB L2 cache, uses the 133 MHz FSB, and was made in 1.0, 1.13, 1.2, 1.33, and 1.4 GHz models. The second variant, intended for entry-level servers and workstations, has 512 KB L2 cache, uses the 100 or 133 MHz FSB, and was made in models that run at 700, 800, 900, or 933 MHz, as well as models that run at 1.13, 1.26, and 1.4 GHz. Both variants are SMP-capable. Finally, Intel removed the much-hated Processor Serial Number from all Tualatin-core processors.

Table 4-2 summarizes the important differences between Pentium III variants as of July 2003. When necessary to differentiate processors of the same speed, Intel uses the *E* suffix to indicate support for ATC and ASB, the *B* suffix to indicate 133 MHz FSB, and the *EB* suffix to indicate both. An A suffix designates 0.13µ Tualatin-core processors. All processors faster than 600 MHz include both ATC and ASB. Note that A-step FC-PGA processors do not support SMP. B-step and higher FC-PGA and FC-PGA2 processors support SMP, except the 1B GHz processor, which is not SMP-capable in any stepping.

Table 4-2. Intel Pentium III variants

	1.40, 1.26, 1.13 GHz	1.33, 1.20, 1.13A, 1A GHz	1B GHz, 933, 866, 800EB, 733, 667, 600EB, 533EB	850, 800, 750, 700, 650, 600E, 550E	1.10G, 1G, 850, 800, 750, 700, 650, 600, 550E, 500E	1G, 933, 866, 800, 733, 667, 600EB, 533EB	600B, 533B	600, 550, 500, 450
Package	FC-PGA2	FC-PGA2	SECC2	SECC2	FC-PGA	FC-PGA	SECC2	SECC2
Process size	0.13µ	0.13µ	0.18µ	0.18µ	0.18µ	0.18µ	0.25µ	0.25µ
FSB speed (MHz)	133	133	133	100	100	133	133	100
L2 cache size (KB)	512	256	256	256	256	256	512	512
L2 cache speed	CPU	CPU	CPU	CPU	CPU	CPU	1/2 CPU	1/2 CPU
SMP support	●	●	●	●	●	●	●	●
Process or S/N	❍	❍	●	●	●	●	●	●



When Intel introduced the Pentium III in FC-PGA form, it changed Socket 370 pinouts. Those changes mean that, although an FC-PGA processor physically fits any Socket 370 motherboard, it will not run in motherboards designed for the Celeron/PPGA. Motherboards designed for FC-PGA processors are nearly all backward-compatible with PPGA Celeron processors. Similarly, as with Tualatin-core Celerons, Tualatin-core Pentium IIIs operate only in late-model Socket 370 motherboards that use chipsets with explicit Tualatin support. Most motherboards designed to use PPGA Celerons or FC-PGA Coppermine-core Pentium IIIs are not compatible with Tualatin-core Pentium IIIs.

Figure 4-7 shows a Pentium III processor in the SECC2 package. Some early Pentium III models were produced in the original SECC package, which closely resembles the Pentium II SECC package shown in Figure 4-2. Figure 4-8 shows a Pentium III processor in the FC-PGA package. Other than labeling, the Pentium III processor in the FC-PGA2 package closely resembles the FC-PGA2 Celeron processor shown in Figure 4-6.

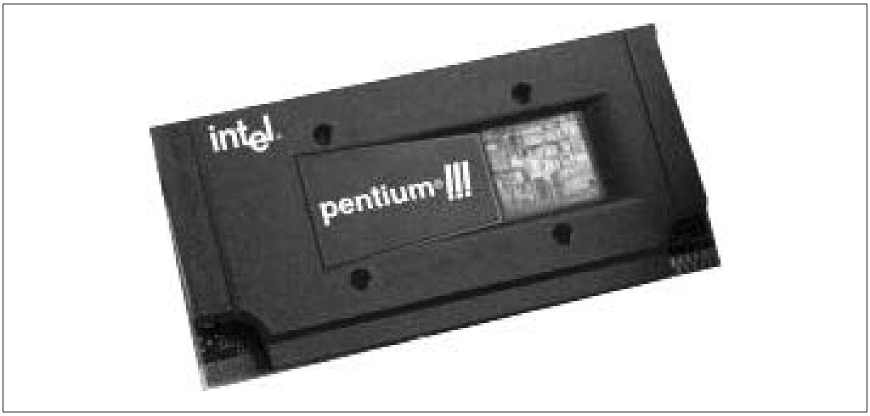


Figure 4-7. Intel Pentium III processor in SECC2 package (photo courtesy of Intel Corporation)

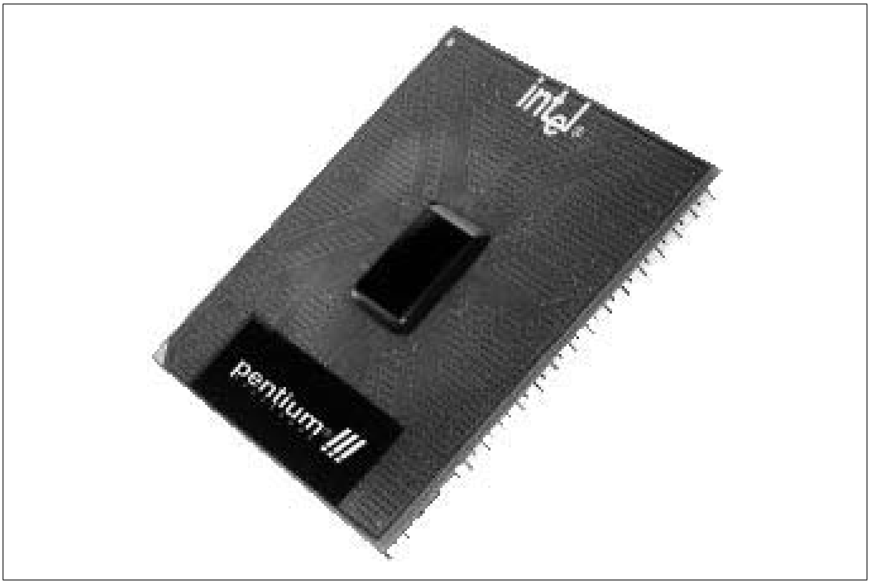


Figure 4-8. Intel Pentium III processor in FC-PGA package (photo courtesy of Intel Corporation)

For additional information about Pentium III processors, including detailed identification tables, visit *http://developer.intel.com/design/pentiumiii/*. For information about Pentium III Xeon processors, visit *http://developer.intel.com/design/pentiumiii/xeon/*.

Pentium 4

By late 2000, Intel found itself in a conundrum. In March of that year, AMD had forced Intel's hand by releasing an Athlon running at 1 GHz. Intel planned to release a 1.0 GHz version of its flagship processor, the Coppermine-core Pentium III, but not until much later. The Athlon/1.0G introduction was a wakeup call for Intel. It had to ship a Pentium III/1.0G immediately if it was to remain competitive on clock speed with the Athlon. One week after the Athlon/1.0G shipped, Intel shipped a Pentium III running at the magic 1.0 GHz.

The problem was that the Pentium III Coppermine core effectively topped out at about 1.0 GHz, while the Athlon Thunderbird core had plenty of headroom. For the next several months, AMD shipped faster and faster Athlons, while Intel remained stuck at 1.0 GHz. And to make matters worse, AMD could ship fast Athlons in volume, while Intel had very low yields on the fast Pentium III parts. Although 1.0 GHz Pentium IIIs were theoretically available, in reality even the 933 MHz parts were hard to come by. So Intel had to make the best of things, shipping mostly sub-900 MHz Pentium IIIs while AMD claimed the high end. Intel must have been gritting its collective teeth.

Adding insult to injury, Intel attempted unsuccessfully to ship a faster Pentium III, the ill-fated Pentium III/1.13G. These processors were available in such small volumes that many observers believed they must be almost handmade. Adding to Intel's embarrassment, popular enthusiast web sites including Tom's Hardware (*http://www.tomshardware.com*) and AnandTech (*http://www.anandtech.com*) reported that the 1.13 GHz parts did not function reliably. Intel was forced to admit this was true and withdrew the 1.13 GHz part, although it later reintroduced it successfully.

Intel had two possible responses to the growing clock speed gap. It could expedite the release of 0.13μ Tualatin-core Pentium IIIs, which have clock speed headroom at least equivalent to the Thunderbird-core and later Palomino-core Athlons, or it could introduce its seventh-generation Pentium 4 processor sooner than planned (see Figure 4-9). Intel wasn't anywhere near ready to convert its fabs to 0.13μ Tualatin-core Pentium III production, so its only real choice was to get the Pentium 4 to market quickly.

There were several problems with that course, not the least of which were that the 0.18μ Willamette-core Pentium 4 was not really ready for release and the only Pentium 4 chipsets Intel had available supported only Rambus RDRAM, which was hideously expensive at the time. But in November 2000, Intel was finally able, if only just, to ship the Pentium 4 processor running at 1.3, 1.4, and 1.5 GHz. Although many observers (including we) noted that that version of the Pentium 4 was a dead-end processor because it used Socket 423, which was due to be replaced by Socket 478 only months after the initial release, and that, despite its higher clock speed, the Pentium 4 had lower performance than Athlons or Pentium IIIs running at lower clock speeds, the Pentium 4 did at least allow Intel to regain the clock speed crown, an inestimable marketing advantage.

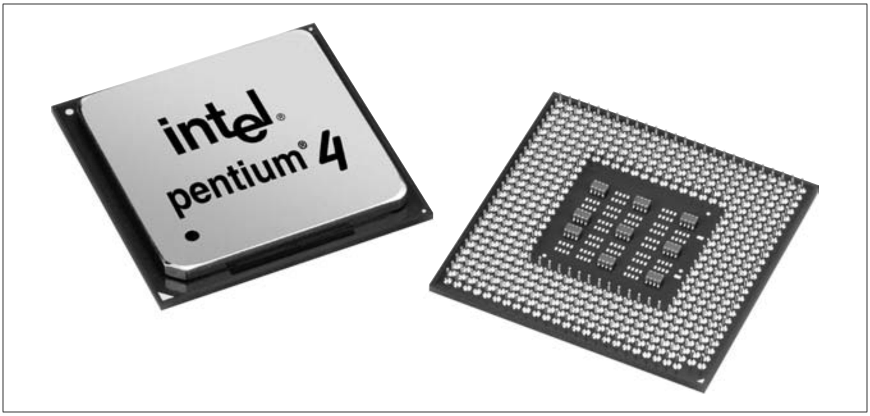


Figure 4-9. Intel Pentium 4 processor in mPGA478 package (photo courtesy of Intel Corporation)

AMD partisans gloated as the Athlon kicked sand in the face of the puny Socket 423 Pentium 4. But those who don't regard processors as a religious issue saw the writing on the wall. The Pentium 4 meant trouble for AMD, big trouble. The seventh-generation Pentium 4 is the most significant new Intel processor since the original Pentium Pro, which kicked off the sixth generation. The Pentium 4 has a lot of headroom, which the aging Athlon core did not.

That first Pentium 4 was significant, not so much for what it was as for what it would become. Just as Intel scaled the clock speeds of sixth-generation cores from the 120 MHz of the first Pentium Pro to the 1.4 GHz of the final Pentium III, we expect that it will scale the clock speed of the Pentium 4 by an order of magnitude or more—albeit using improved cores—eventually reaching 10 GHz to 15 GHz before introducing their next completely new core, which by that time may be named the Pentium 6, 7, or 8.

For the Pentium 4, Intel launched the fastest ramp-up in its history. In earlier generations, new processors coexisted with older processors for quite some time. Intel derived substantial revenues from the 386 long after the 486 shipped, from the 486 long after the Pentium shipped, and from the Pentium long after the Pentium II shipped. With the Pentium 4, it abandoned the idea of a staged introduction. Intel killed the market for sixth-generation processors quickly, leaving the Pentium 4 and its derivatives as the only mainstream Intel processors.

Pentium 4 processor features

Relative to sixth-generation processors, the Pentium 4 incorporates the following architectural improvements which together define the seventh generation and which Intel collectively calls NetBurst Micro-architecture.

Hyper pipelined technology

Hyper-pipelining doubles the pipeline depth compared to the Pentium III micro-architecture. The branch prediction/recovery pipeline, for example, is implemented in 20 stages in the Pentium 4, as compared to 10 stages in the Pentium III. Deep pipelines are a double-edged sword. Using a very deep pipeline makes it possible to achieve very high clock speeds, but a deep pipeline also means that fewer instructions can be completed per clock cycle. That means the Pentium 4 can run at much higher clock speeds than the Pentium III (or Athlon), but that it needs those higher clock speeds to do the same amount of work.

Early Pentium 4 processors were roundly condemned by many observers because they were outperformed by Pentium III and Athlon processors running at much lower clock speeds, which is solely attributable to the relative inefficiency of the Pentium 4 in terms of Instructions per Cycle (IPC). Ultimately, the low IPC efficiency of the Pentium 4 doesn't matter because Intel can easily boost the clock speed until the Pentium 4 greatly outperforms the fastest Pentium III or Athlon that can be produced. What superficially appears to be a weakness of the Pentium 4 is in fact its greatest strength.

Improved branch prediction

The deep pipeline of the Pentium 4 made it mandatory to use a superior Branch Prediction Unit (BPU) because a deep pipeline with anything less than excellent branch prediction would bring the processor to its knees. When the pipeline is very deep, a pipeline clog wastes massive numbers of clock ticks, and the function of a BPU is to prevent that from happening. The Pentium 4 BPU is the most advanced available, 33% more efficient at avoiding mispredictions than the Pentium III BPU or the comparable Athlon BPU. The Pentium 4 BPU uses a more effective branch-prediction algorithm and a dedicated 4 KB branch target buffer that stores detail about branching history to achieve these results. The improved BPU is one component of the Advance Dynamic Execution (ADE) engine, Intel's name for its very deep, out-of-order speculative execution engine.

Level 1 Execution Trace Cache

In addition to the standard Level 1 8 KB data cache, the Pentium 4 includes a 12 KB L1 Execution Trace Cache. This cache stores decoded micro-op instructions in the order they will be executed, optimizing storage efficiency and performance by removing the micro-op decoded from the main execution loop and storing only those micro-op instructions that will be needed. By caching micro-op instructions before they are needed, the Execution Trace Cache ensures that the processor execution units seldom have to wait for instructions, and that the effects of branch mispredictions are minimized.

Rapid Execution Engine

Even with an excellent BPU, integer code is more likely than floating-point code to be mispredicted, and such mispredictions have a catastrophic effect on throughput. To minimize their effect, the Pentium 4 includes two *Arithmetic Logic Units* (*ALUs*) that operate at twice the processor core frequency. For example, the Rapid Execution Engine on a 2 GHz Pentium 4 actually runs at 4 GHz. That allows a basic integer operation (e.g., Add, Subtract, AND, OR) to execute in half a clock cycle.

400, 533, or 800 MHz system bus

One Achilles' heel of the Pentium III (and, to a lesser extent, the Athlon) is the relatively slow link between the processor and memory. For example, using PC133 SDR-SDRAM, the Pentium III achieves peak data-transfer rates of only 1067 MB/s (133 MHz times 8 bytes/transfer). In practice, sustained data-transfer rates are lower still because SDRAM is not 100% efficient and the SDRAM interface uses only minimal buffering. Conversely, the Pentium 4 has the fastest system bus available on any desktop processor. Although the bus actually operates at only 100, 133, or 200 MHz, data transfers are quad-pumped for an effective bus speed of 400, 533, or 800 MHz. Also, Intel uses elaborate buffering that ensures sustained true 400/533/800 MHz data transfers when using Rambus RDRAM or dual-channel DDR-SDRAM memory. Sustained data-transfer rates using SDR-SDRAM or DDR-SDRAM are smaller than peak transfer rates, but are still much faster than the data-transfer rates of the Pentium III or Athlon using similar memory.

Hyper-Threading technology

Finally, with the November 2002 introduction of the Pentium 4/3.06G, Intel implemented *Hyper-Threading Technology* (*HTT*) on some of its Pentium 4 processors. To understand the potential benefit of HTT, it is necessary to understand a bit about how instructions are processed in a modern processor core.

Consider a 24-hour supermarket with seven cash registers. On a Saturday afternoon, all seven of those cashiers may be busy, with customers backed up in each aisle waiting to complete their transactions. At 2:00 on a Wednesday morning, only one of the cash registers may be staffed because fewer customers are in the store. Even so, a flurry of activity may mean that a line forms at the one available cash register, leaving the remaining six unused.

The Pentium 4 has seven execution units, which are analogous to the cash registers. Two of those execution units, the double-pumped ALUs, process two operations per clock cycle. The other execution units, including the FPUs, process one operation per clock cycle. Because execution units operate independently, in theory the Pentium 4 could process a total of nine operations per clock cycle.

In practice, the Pentium 4 processes nowhere near nine operations per clock cycle because inefficiencies in matching the requirements of the running program code to the resources the processor has available mean that many of those resources go unused at any particular time. For example, typical desktop productivity software processes a lot of integer operations, loads, and stores, but leaves the floating-point execution units almost unused. Conversely, a scientific, CAD, or graphics program might use the FPUs almost exclusively, leaving the ALUs almost unused. Even programs that use integer operations almost exclusively will probably not saturate all of the ALUs. The upshot is that, during normal operations, most of the available execution units sit idle. According to Intel, the Pentium 4 typically uses only 35% of the available execution unit resources during normal operations. In effect, the CPU runs at only 35% of its potential performance.

With single-threaded programs, not much can be done to improve this situation. If, for example, the program has saturated the FPUs, all the ALUs in the world won't improve its performance. But in a multithreading environment, it's quite possible that resources not needed by one program thread might be usable by a different program thread. The problem is that a standard processor can execute only one program thread at a time. That means the second thread must wait its turn, even though the resources it needs are not being used by the currently active thread.

SMP is one solution to this problem. With multiple processors, each processor can be assigned a separate thread. These multiple threads are processed simultaneously, significantly increasing overall system performance. SMP does nothing to improve processor utilization, of course. Each of the multiple processors is still operating at only 35% or so of its potential throughput.

HTT is another solution to the problem. HTT splits each physical processor into virtual dual processors, allowing a single physical processor to process two threads simultaneously. To the extent that these two threads require different execution unit resources, they are not in conflict and can thus use a higher percentage of the available processor resources. Because each thread invariably requires resources that are also needed by the other thread, overall performance is not doubled. Performance may, however, increase by 20% or more in an HTT processor relative to a similar processor that does not support HTT.

HTT is not a panacea. If two program threads have similar resource requirements, a processor with HTT enabled may actually run those threads more slowly than the same processor with HTT disabled. For that reason, many vendors that ship HTT-capable systems turn HTT off by default. The only way to determine whether HTT will improve performance on your system is to run the system with HTT enabled and disabled and see which configuration runs faster for you. In our experience, HTT usually makes little difference either way if you are running only office applications, but if you run a mix of typical office applications and FPU-intensive applications, HTT can sometimes improve performance noticeably.

> Beware of enabling HTT if you run Windows 2000, which sees an HTT processor as two physical processors, and demands licenses for twice as many processors as you actually have. Even worse, Windows 2000 uses virtual processors and ignores "extra" physical processors. For example, if you run Windows 2000 Professional, which supports two processors, on a system with two physical HTT processors, Windows 2000 recognizes only the two virtual processors on the first physical processor, and ignores the second physical processor entirely. Duh. Microsoft's "solution" for this problem is to suggest that you buy an upgrade to Windows XP. Thanks, but no thanks. We'll upgrade to Linux instead.

At its introduction in November 2002, Intel supported HTT only in the Pentium 4/3.06G, the fastest and most expensive Pentium 4 at that time. In May 2003 Intel began shipping entry-level and midrange 800 MHz FSB Pentium 4 processors with HTT support, including the 2.40C, 2.60C, and

2.80C. In June 2003, Intel began shipping HTT-enabled Pentium 4 processors at 3.2 GHz, with faster versions due later in 2003 and throughout 2004.

> Enabling HTT requires that the processor, chipset, BIOS, and operating system all support HTT. The Intel 850E, 865-, and 875-series chipsets support HTT, as do most versions of the 845-series chipsets. The 845 chipset and the 845G chipset in steppings prior to B1 do not support HTT. Windows XP supports HTT, as does Linux with a 2.4.18 or higher kernel.

In addition to its new features, the Pentium 4 also has two features that have been significantly enhanced relative to the Pentium III:

Enhanced ATC
: Intel has enhanced the performance of the L2 ATC that first appeared in the Pentium III. The Pentium 4 uses a non-blocking, eight-way set associative, inclusive, full-CPU-speed, on-die, L2 cache with a 256-bit interface that transfers data during each clock cycle. Because the Pentium 4 clock is faster than that of the Pentium III, L2 cache transfers also support a much higher data rate. For example, a Pentium III operating at 1 GHz transfers L2 cache data at 16 GB/s, whereas a Pentium 4 at 1.5 GHz transfers L2 cache data at 48 GB/s (three times the transfer rate for a processor operating at 1.5 times the speed). The ATC also includes improved Data Prefetch Logic that anticipates what data will be needed by a program and loads it into cache before it is needed. Willamette-core Pentium 4 processors have a 256 KB L2 cache. Northwood-core Pentium 4 processors have a 512 KB L2 cache.

Enhanced floating-point and SSE functionality
: The Pentium 4 uses 128-bit floating-point registers and adds a dedicated register for data movement. These enhancements improve performance relative to the Pentium III on floating-point and multimedia applications. The Pentium 4 also includes SSE2, an updated version of the SSE that debuted with the Pentium III. SSE, which stands for Streaming SIMD Extensions, is an acronym within an acronym. SIMD, or Single Instruction Multiple Data, allows one instruction to be applied to a multiple data set (e.g., an array), which greatly speeds performance in such applications as video/image processing, encryption, speech recognition, and heavy-duty scientific number crunching. SSE2 adds 144 new instructions to the SSE instruction set, including 128-bit SIMD integer arithmetic operations and 128-bit SIMD double-precision floating-point operations. These new instructions can greatly reduce the number of steps needed to execute some tasks, but the catch is that the application software must explicitly support SSE2. For example, an application that is not designed to use SSE2 might run at the same speed on a Pentium 4 and an Athlon, while an SSE2-capable version of that application might run literally twice as fast on the Pentium 4.

Pentium 4 processor variants

Intel has produced Pentium 4 processors using two cores, the 0.18µ Willamette core and the 0.13µ Northwood core; two form factors, the 423-pin PGA-423

(Socket 423) and the smaller 478-pin mPGA-478 (Socket 478); and three FSB speeds, 400 MHz, 533 MHz, and 800 MHz:

Willamette-core processors
: Willamette-core Pentium 4 processors have 256 KB of eight-way set associative L2 cache and use the 400 MHz FSB. Intel has produced Willamette-core processors for Socket 423 and Socket 478 at core speeds of 1.30, 1.40, 1.50, 1.60, 1.70, 1.80, 1.90, and 2 GHz. Willamette-core processors have 42 million transistors and a die size of 217 square millimeters.

Northwood-core processors
: Northwood-core Pentium 4 processors have 512 KB of eight-way set associative L2 cache and use the 400, 533, or 800 MHz FSB. Intel has produced Northwood-core processors only for Socket 478 at core speeds of 1.6, 1.8, 2.0, 2.2., 2.26, 2.4, 2.5, 2.53, 2.6, 2.67, 2.8, 3.0, 3.06, and 3.2 GHz, with faster variants planned for release later in 2003. Northwood-core processors have 55 million transistors. The original Northwood core used a die size of 146 square millimeters, which in July 2002 was reduced to 131 square millimeters. Although Northwood-core processors dissipate less heat than Willamette-core processors running at the same speed, the smaller die size means the heat dissipated per unit surface area is actually higher. Northwood-core processors, particularly fast ones, accordingly require careful attention to proper cooling.

The Willamette core and Socket 423 were stopgap solutions, released solely to combat AMD's clock speed lead until the "real" Pentium 4—the Socket 478 Northwood-core processor—could be shipped. Intel intended to phase out Socket 423 as a mainstream technology by late 2001, relegating Socket 423 to upgrade status only, but the demand for Socket 478 motherboards and processors caused product shortages until mid-2002. When Intel had resolved those problems, it quickly discontinued Socket 423 motherboards and processors, which are now available only from overstock vendors and as used products.

For additional information about Pentium 4 processors, including detailed identification tables, visit *http://developer.intel.com/design/pentium4/*. For information about Xeon processors, visit *http://developer.intel.com/design/xeon/prodbref/*.

Celeron (Seventh-Generation)

In May 2002, Intel shipped a new series of seventh-generation Celeron processors. Just as the original Celerons were Pentium II and Pentium III variants with smaller L2 caches and slower FSB speeds, the new Celerons are Pentium 4 variants with, you guessed it, smaller caches and slower FSB speeds.

Confusingly, Intel uses the Celeron name for two entirely different series of processors. Like the sixth-generation Celerons, seventh-generation Celerons are positioned as entry-level processors with lower performance than Intel's mainstream processors. Intel walks a fine line with these processors because they must be fast enough to satisfy the price-sensitive entry-level market and compete successfully with low-end AMD processors, yet not be fast enough to cannibalize sales of the more profitable Pentium 4 processors.

Seventh-generation Celerons fit Socket 478 motherboards. Some Socket 478 motherboards do not support the Celeron, and those that do may require a BIOS upgrade. The first seventh-generation Celeron models used a modified 0.18µ Pentium 4 Willamette core called the Willamette-128 core, which has 128 KB of eight-way set associative L2 cache, half that of the Willamette-core Pentium 4. Willamette-128 Celerons were made in 1.7 and 1.8 GHz versions, which shipped in May and June 2002.

In September 2002, Intel began producing Celerons with a modified 0.13µ Pentium 4 Northwood core called the Northwood-128 core. Intel has produced Northwood-128 Celerons running at 2.0, 2.1, 2.2, 2.3, and 2.4 GHz. Like the Willamette-128 Celerons, these processors have 128 KB of eight-way set associative L2 cache, only one-quarter that of the Northwood-core Pentium 4.

One seldom-mentioned fact is that this tiny 128 KB L2 cache greatly impairs performance of a Northwood-128 Celeron relative to that of a Northwood Pentium 4 operating at the same speed. Whereas earlier sixth- and seventh-generation Celerons often had 85% or more the performance of the corresponding Pentium III or Pentium 4, with some benchmarks a Northwood-128 Celeron shows only 65% the performance of a Northwood Pentium 4 operating at the same clock speed. In effect, that means that the fastest available Northwood-128 Celeron is noticeably slower for some tasks, especially multimedia and gaming, than the slowest available Pentium 4, which sells for only a few dollars more. Intel really shot itself in the foot that time.

The days of the Celeron as a separate processor line may be numbered, although it's possible that Intel will take the same course it did by rebranding Tualatin-core Pentium IIIs as Celerons. That is, Intel may begin using the Pentium 4 brand only for its then-current midrange and faster processors. As faster processors are introduced, Intel may simply relabel older, slower Pentium 4 processors as Celerons, without making any actual changes to the processors.

The problem Intel faces with the Celeron is the same problem AMD faced with the Duron, which AMD recently discontinued. When processor prices ranged from $100 to $1,000, it made sense to have two separate lines of processors, economy lines such as the Celeron and Duron, and premium lines such as the Pentium III, Pentium 4, and Athlon. But processor prices have fallen dramatically, and average selling price (ASP) has plummeted even more. When the least-expensive Pentium 4 sold for $300, there was plenty of pricing room for a full series of Celeron processors. Now that entry-level Pentium 4 processors are routinely available for less than $150, there's not much room for a less-expensive, slower line of processors.

Our advice is to avoid seventh-generation Celeron processors except when low system price is the highest priority. In that case, use the least-expensive Northwood-128 Celeron you can find. Otherwise, you'll find that even the least-expensive Pentium 4 significantly outperforms the fastest Celeron and costs little more.

For additional information about Celeron processors, including detailed identification tables, visit *http://developer.intel.com/design/celeron/*.

> Intel has manufactured mobile variants of many of its processors, including the Pentium, Pentium II, Celeron, and Pentium III. These mobile versions are used in notebook computers and are not user-replaceable, so for all intents and purposes a notebook computer will always use the processor that was originally installed. For that reason, we have chosen to devote our available space to issues that are more likely to be important to more of our readers. For additional information about Intel mobile processors, visit *http://developer.intel.com/design/mobile/*.

AMD Processors

Until late 1999, Intel had the desktop processor market largely to itself. There were competing incompatible systems such as the Apple Mac, based on processors from Motorola, IBM, and others, but those systems sold in relatively small numbers. Some companies, including Cyrix, IDT, Harris, and AMD itself, made Intel-compatible processors, but those were invariably a step behind Intel's flagship processors. When those companies—which Intel calls "imitators"—were producing enhanced 286s, Intel was already shipping the 386 in volume. When the imitators began producing enhanced 386-compatible processors, Intel had already begun shipping the 486, and so on. Each time Cyrix, AMD, and the others got a step up, Intel would turn around and release its next-generation processor. As a result, these other companies' processors sold at low prices and were used largely in low-end systems. No one could compete with Intel in its core market.

All of that changed dramatically in late 1999, when AMD began shipping the Athlon processor. The Athlon didn't just match the best Intel processors. It was faster than the best Intel could produce, and was in many respects a more sophisticated processor. Intel had a fight on its hands, and it does to this day.

If you ever take a moment to appreciate how much processor you can get for so little money nowadays, give thanks to AMD. Without AMD, we'd all still be running sixth-generation Intel processors at 750 MHz or so. An entry-level Intel processor would cost $200 or $250, and a high-end one (that might run at 1 GHz) would probably cost $1,000 or more. The presence of AMD as a worthy competitor meant that Intel could no longer play the game of releasing faster processors in dribs and drabs at very high prices. Instead, Intel had to fight for its life by shipping faster and faster processors at lower and lower prices. We all have AMD to thank for that, and Intel should thank AMD as well. Although we're sure Intel wishes AMD would just disappear (and vice versa), the fact is that the competition has made both Intel and AMD better companies, as well as providing the obvious benefits to us, the users.

The following sections describe current and recent AMD processor models.

The AMD Athlon Family

The AMD Athlon, which was originally code-named the K7 and began shipping in August 1999, was the first Intel-compatible processor from any maker that could compete on an equal footing with mainstream Intel processors of the time.

First-generation Athlon processors matched or exceeded Katmai-core Pentium III processors in most respects, including (for the first time ever) floating-point performance. Intel finally had a real fight on its hands.

Although AMD represented the Athlon as the first seventh-generation processor, we regard the K7 Athlon as essentially an enhanced sixth-generation processor. Athlon has, in theory, several advantages relative to the aging Intel sixth-generation architecture, including the ability to perform nine operations per clock cycle (versus five for the Pentium III); more integer pipelines (three versus two); more floating-point pipelines (three versus one); a much larger L1 cache (128 KB versus 32 KB); more full x86 decoders (three versus one); and a faster FSB (100 MHz double-pumped to 200 MHz by transferring data on both the rising and falling edges of the clock cycle versus the single-pumped Intel 100/133 MHz bus, which transfers data only once during a clock cycle). While all that was very nice, tests showed that in practice the K7 Athlon and Pentium III were evenly matched at lower clock speeds, with the Pentium III sometimes showing a slight advantage in integer performance, and the Athlon a slight advantage in floating-point performance. At higher clock speeds, however, where the Pentium III L2 cache running at full CPU speed comes into play, the Coppermine Pentium III won most benchmarks handily.

AMD produced two variants of the first-generation Athlon, both in Slot A form. The earliest Athlons used the 0.25µ K7 core, but AMD transitioned within a few months to the improved 0.18µ K75 core, which was code-named Pluto for speeds lower than 1 GHz and Orion in the 1 GHz model. Although the K7 and K75 Athlons were good processors, they had the following drawbacks:

Poor chipset and motherboard support
: Initial acceptance of the Athlon was hampered because the only chipset available was the AMD-750, which was originally intended as a technology demonstrator rather than as a production chipset. The VIA KX133 chipset, originally planned to ship at the same time as the Athlon, was significantly delayed, and motherboards based on the KX133 began shipping in volume only in the second quarter of 2000. Many motherboard manufacturers delayed introducing Athlon motherboards, and their first products were crude compared to the elegant motherboards available for the Pentium III. In addition to indifferent quality, stability, compatibility, performance, and features, first-generation Athlon motherboards were in short supply and relatively expensive compared to comparable models for the Pentium III. In addition, KX133-based motherboards had problems of their own, including their inability to support Slot A Thunderbird-core Athlons. AMD soon made it clear that Slot A was an interim solution and that it would quickly transition to Socket A, so manufacturers devoted little effort to improving orphaned Slot A motherboards.

Fractional CPU-speed L2 cache
: Like the Deschutes-core Pentium II and the Katmai-core Pentium III, K7 Athlons run L2 cache at half CPU speed. Unlike the Coppermine Pentium III, which uses on-die L2 cache running at full CPU speed, the Athlon uses discrete L2 cache chips, which AMD had to buy from third parties. The Athlon architecture allows running L2 cache at anything from a small frac-

tion of CPU speed to full CPU speed. AMD took advantage of this as it introduced faster versions of the Athlon by reducing the speed of L2 cache relative to processor speed, allowing the company to use less expensive L2 cache chips. The Athlon/700 and slower run L2 cache at 1/2 CPU speed; The Athlon/750, /800, and /850 run L2 cache at 2/5 CPU speed. the Athlon/900 and faster run L2 cache at 1/3 CPU speed. Unfortunately, compared to the full-speed Pentium III Coppermine L2 cache, the slow L2 cache used on fast Athlons decreases performance substantially in many applications.

High power consumption

Early Athlon processors were power-hungry, with some 0.25μ models consuming nearly 60 watts. In comparison, typical Intel processors used one-half to one-third that amount. High power consumption and the resulting heat production had many implications, including the requirement for improved system cooling and larger power supplies. In fact, for the Athlon, AMD took the unprecedented step of certifying power supplies for use with its processor. If you built a system around a first-generation Athlon, you had to make sure that both cooling and power supply were adequate to meet the extraordinarily high current draw and heat dissipation of the processor.

Lack of SMP support

Until mid-2001, no multiprocessor Athlon systems existed. Although all Athlon processors from the earliest models have been SMP-capable (and in fact use the superior point-to-point SMP method rather than Intel's shared bus method), dual-processor Athlon systems had to wait for the release of the AMD-760MP chipset (originally designated the AMD-770) in mid-2001. This early absence of SMP support hurt Athlon acceptance in the critical corporate markets, not so much because there was a huge demand for SMP but because the lack of SMP support led buyers to consider the Athlon a less advanced processor than Intel's offerings.

With the exception of SMP support, which was never lacking in the processor, these faults were corrected in the second generation of Athlon CPUs, which are based on the enhanced K75 core code-named Thunderbird. All early Athlon models used Slot A, which is physically identical to Intel's SC242 (Slot 1), but uses EV-6 electrical signaling rather than the GTL signaling used by Intel. Figure 4-10 shows a Slot A Athlon processor.

Table 4-3 lists the important characteristics of first- and second-generation Slot A Athlon variants (Model 3 is missing because it was assigned to the Duron processor). All Slot A variants use the double-pumped 100 MHz FSB, for an effective 200 MHz FSB speed. First-generation (K7- and K75-core) Athlons are characterized by their use of 512 KB L2 cache running at a fraction of CPU speed and by their use of split core and I/O voltages. Second-generation (Thunderbird-core) Athlons are characterized by their use of a smaller 256 KB L2 cache that operates at full CPU speed and by the elimination of split voltages for core and I/O. Thunderbird processors were produced in very small numbers in Slot A for OEM use, and so are included in this table for completeness, but we've never actually seen a Slot A Thunderbird and don't know anyone who has.

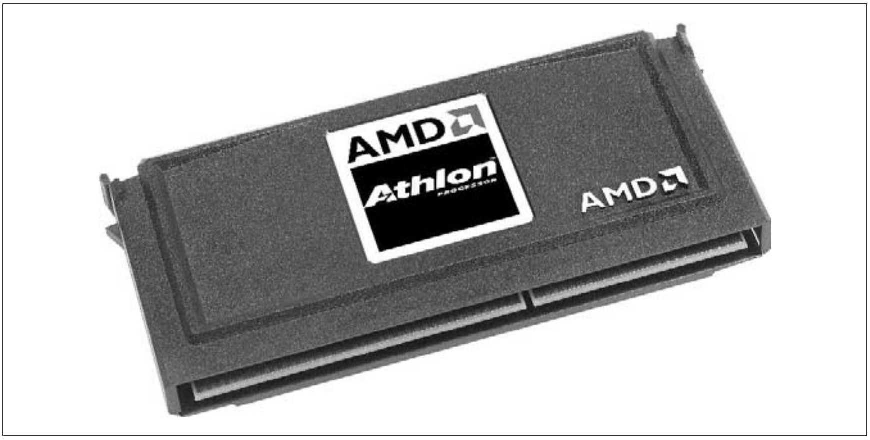


Figure 4-10. AMD Slot A Athlon processor

Table 4-3. Slot A Athlon variants

	Athlon	Athlon	Athlon	Athlon	Athlon	Athlon
Core	K7	K75	K75	K75	Thunderbird	Thunderbird
Model	1	2	2	2	4	4
Production dates	1999, 2000	2000	2000	2000	2000, 2001	2000, 2001
Clock speeds (MHz)	500, 550, 600, 650, 700	550, 600, 650, 700	750, 800, 850	900, 950, 1000	700, 750, 800, 850	900, 950, 1000
L2 cache size	512 KB	512 KB	512 KB	512 KB	256 KB	256 KB
L2 cache speed	1/2 CPU	1/2 CPU	2/5 CPU	1/3 CPU	CPU	CPU
L2 cache bus width	64 bits	64 bits	64 bits	64 bits	64 bits	64 bits
System bus speed	200 MHz	200 MHz	200 MHz	200 MHz	200 MHz	200 MHz
Core voltage	1.6	1.6	1.6 (750) 1.7 (800/850)	1.8	1.7	1.75
I/O voltage	3.3	3.3	3.3	3.3	1.7	1.75
Dual CPU-capable	m	m	m	m	m	m
Fabrication process	0.25µ	0.18µ	0.18µ	0.18µ	0.18µ	0.18µ
Interconnects	Al	Al	Al	Al	Al/Cu	Al/Cu
Die size (mm^2)	184	102	102	102	120	120
Transistors (million)	22	22	22	22	37	37

Like Intel, which shifted from Slot 1 to Socket 370 for low-end processors, AMD recognized that producing cartridge-based slotted processors was needlessly expensive for the low end, and made it more difficult to compete in the value segment. Also, improvements in fabrication made it possible to embed L2 cache directly on the processor die rather than using discrete cache chips. Accordingly, AMD developed a socket technology, analogous to Socket 370, which it called Socket A. AMD had never denied that Slot A was a stopgap technology, and that Socket A was its mainstream technology of the future. AMD rapidly phased out Slot A during 2000, and by late 2000 had fully transitioned to Socket A. AMD has to date produced four major Athlon variants in Socket A. From earliest to latest, these include:

Athlon (Thunderbird core)

The *Thunderbird Athlon* was originally designated *Athlon Professional* and targeted at the mainstream desktop and entry-level workstation market, in direct competition with the Intel Pentium III and Pentium 4. The first Thunderbird processors used an 0.18μ process with aluminum interconnects, but by late 2000 AMD had transitioned to a 0.18μ process with copper interconnects. During that transition, AMD phased out Slot A Thunderbird models, and shifted entirely to Socket A. Early Thunderbirds used the 100 MHz FSB (double-pumped to 200 MHz), with later models also available in 133 MHz FSB variants. Figure 4-11 shows a Socket A Athlon Thunderbird processor.

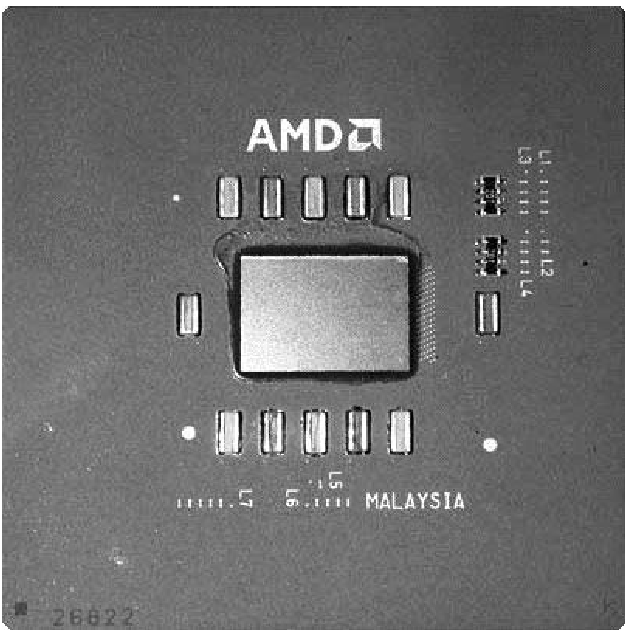


Figure 4-11. AMD Socket A Athlon Thunderbird processor

There was to have been another variant of the Thunderbird-core Athlon, code-named Mustang and formally named Athlon Ultra, but that processor shipped only as samples. Mustang was to be a Socket A part, targeted at servers and high-performance workstations and desktops. It was to be an enhanced version of Thunderbird, with reduced core size, lower power consumption, and large, full-speed, on-die L2 cache, probably 2 MB or more. Mustang was to have used a 133 MHz DDR FSB, yielding an effective FSB of 266 MHz. It was intended to use a 0.18µ process with copper interconnects from the start, and to require the AMD-760 chipset or later. Alas, the Mustang never shipped. It would have been a wonderful processor for its time.

Athlon XP (Palomino core)

AMD originally intended to name the Palomino-core Athlon the *Athlon 4*, for obvious reasons. In fact, the first Palomino-core Athlons that shipped were the Mobile Athlon 4 and the 1.0 GHz and 1.2 GHz versions of the Athlon MP. Instead, given Microsoft's schedule for introducing Windows XP, AMD decided its new processor might tag along on the coattails of the new Windows version. Accordingly, AMD finally named the Palomino-core Athlon the Athlon XP. Various architectural changes from the Thunderbird core, detailed later in this section, allow the Athlon XP to achieve considerably higher performance at a given clock speed than a comparable Thunderbird. The Athlon XP is also the first recent AMD processor to use a model designation unrelated to its actual clock speed. All Palomino-core Athlons use the 133/266 MHz FSB. Figure 4-12 shows a Palomino-core Athlon XP processor.

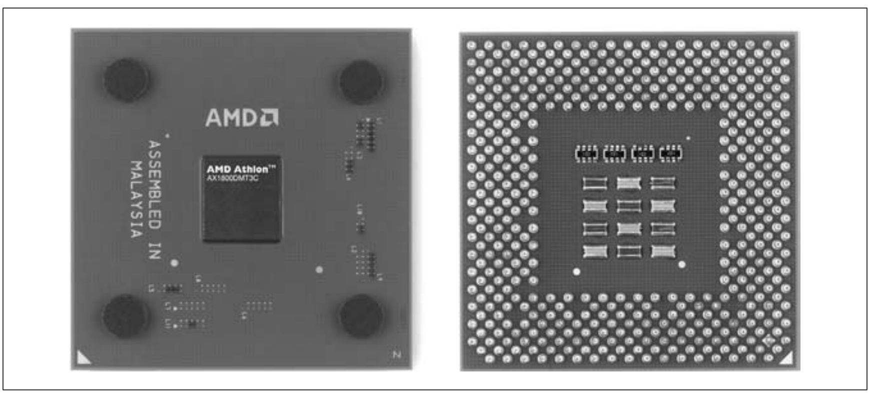


Figure 4-12. AMD Athlon XP processor (image courtesy of Advanced Micro Devices, Inc.)

Athlon XP (Thoroughbred core)

The Thoroughbred core, introduced in June 2002, is really just a die shrink of the Palomino core. In reducing the fabrication process size from 0.18µ to 0.13µ, AMD was able to shrink the die from 128 mm^2 to 81mm^2 (although that increased to 84mm^2 for the XP 2200+ and faster models).

There were no significant architectural changes from the Palomino core to the Thoroughbred core, so performance did not increase with the change to the new core. Transistor count did increase somewhat, from 37.2 million to 37.5 million. AMD also increased the number of metal layers from seven in the Palomino core to eight in the Thoroughbred core, which increases manufacturing complexity and cost, but allows improved routing by optimizing electrical paths within the processor, allowing closer placement of components and faster clock speeds. (For comparison, the Intel Northwood-core Pentium 4 uses only six layers.) The die shrink also allows using lower voltages, which reduces power consumption and heat output significantly. For example, the Palomino-core Athlon XP 2100+ dissipates 72.0W maximum, while the Thoroughbred-core Athlon XP 2100+ dissipates only 62.1W. All Thoroughbred-core Athlons use the 133/266 MHz FSB.

In August 2002, AMD introduced the Thoroughbred "B" core, which increased the number of metal layers to nine, again to allow faster clock speeds. From a functional standpoint, the major change is support for the 166/333 MHz FSB, which was first used with the Athlon XP 2400+ processor. Other than FSB, the only noticeable difference between the Thoroughbred and Thoroughbred "B" cores is that the former reports a CPUID string of 680, while the later reports 681.

Athlon XP (Barton core)

The Barton core, introduced in February 2003 with the Athlon XP 3000+, uses the same 0.13μ fab size as the Thoroughbred core, but the transistor count increases from 37.5 million to 54.3 million. That boost in transistor count increases die size from 84 mm^2 to 101 mm^2. Most of the increase in transistor count and die size is a result of L2 cache size being boosted from 256 KB to 512 KB. Other than the larger cache and larger die size, the Barton core is essentially the same as the Thoroughbred B core.

Despite the doubling of L2 cache size, the Barton core is a less significant upgrade to the Thoroughbred core than one might expect. Benchmarking a Willamette-core Pentium 4 with 256 KB of L2 cache against a Northwood-core Pentium 4 with 512 KB L2 cache running at the same clock speed typically shows performance increases in the 10% to 25% range, and often more. Those who expect a similar improvement going from a 256 KB Thoroughbred-core Athlon to a 512 KB Barton-core Athlon will be disappointed. Differences in processor bandwidth and caching technologies mean that the Athlon benefits much less from the larger L2 cache than does the Pentium 4. On most benchmarks, a Barton-core Athlon shows only a 1% to 5% performance improvement relative to a Thoroughbred-core Athlon running at the same clock speed.

Barton-core processors were initially available only with a 166/333 MHz FSB. Later Barton-core processors, including the Athlon XP 3200+, will ship with the 200/400 MHz FSB.

The really significant changes took place in the upgrade to the Thunderbird and Palomino cores. Other than the reduction from 0.18μ to 0.13μ and the substitution of copper interconnects for aluminum ones, the subsequent changes to the

Athlon core, particularly those to Thoroughbred and Barton, are largely minor tweaks that allow incrementally faster processor speeds. Faced with Intel's modern Pentium 4 core, AMD has been forced to squeeze as much as possible from its aging Athlon technology in order to remain competitive.

By updating the Athlon core and using such marketing gimmicks as naming its processors with model numbers higher than their actual clock speeds, AMD has generally remained competitive. But the Barton is almost certainly the last gasp for the Athlon. In order to counter faster Pentium 4 models from Intel, AMD has no choice. It must relegate the Athlon to the entry level and grab significant market share quickly for its forthcoming Hammer-series processors. The alternative doesn't bear thinking about.

AMD actually first shipped Palomino-core Athlon processors some months before the Athlon/XP desktop processor in the Athlon 4 mobile variant and the Athlon MP/1.0G and Athlon MP/1.2G variants, all of which were designated by their actual clock speeds. Subsequent Palomino-core Athlon processors are all designated using the QuantiSpeed performance rating rather than their actual clock speeds. For example, the Athlon XP/1500+, XP/1600+, XP/1700+, XP/1800+, and XP/1900+ actually run at clock speeds of 1333, 1400, 1466, 1533, and 1600 MHz, respectively, as do the similarly badged Athlon MP SMP-capable variants.

Although Palomino-core processors use the same 0.18μ fabrication process used for Thunderbird-core processors, AMD made several improvements in layout and architecture. Relative to the Thunderbird-core Athlon, Palomino-core Athlons (including the Athlon XP, the Athlon MP, and the Mobile Athlon 4) provide 3% to 7% faster performance clock for clock, and include the following enhancements:

Improved data prefetch mechanism

This allows the CPU, without being instructed to do so, to use otherwise unused FSB bandwidth to prefetch data that it thinks may be needed soon. This single feature accounts for most of the performance improvement in the Palomino core relative to the Thunderbird, and also increases the processor's dependence on a high-speed FSB/memory bus. Better data prefetch most benefits applications that require high memory bandwidth and have predictable memory access patterns, including video editing, 3D rendering, and database serving.

Enhanced Translation Look-aside Buffers

Translation Look-aside Buffers (*TLBs*) cache translated memory addresses. Translation is needed for the CPU to access data in main memory. Caching translated addresses makes finding data in main memory much faster. Palomino-core Athlons include the following three enhancements to the TLBs:

More L1 Data TLBs

Palomino-core Athlons increase the number of L1 Data TLBs from 32 to 40. The larger number of TLB entries increases the probability that the needed translated address will be cached, thereby improving performance. Even with 40 entries, though, the Palomino-core Athlon has fewer L1 TLB entries than the Intel Pentium III or Pentium 4, and the benefit of this small increase is minor.

L2 TLBs use exclusive architecture

In Thunderbird-core Athlons, the L1 and L2 TLBs are nonexclusive, which means that data cached in the L1 TLB is also cached in the L2 TLB. With the Palomino core, AMD uses an exclusive TLB architecture, which means that data cached in the L1 TLB is not replicated in the L2 TLB. The benefit of exclusive caching is that more entries can be cached in the L2 TLB. The drawback is that using exclusive caching results in additional latency when a necessary address is not cached in the L2 TLB. Overall, exclusive TLB caching again results in a minor performance increase.

TLB entries can be speculatively reloaded

Speculative reloading means that if an address is not present in the TLB, the processor can load the address into the TLB before the instruction that requested the address has finished executing, thereby making the cached address available without the latency incurred by earlier Athlon cores, which could load the TLB entry only after the requesting instruction had executed. Once again, speculative reloading provides a minor performance improvement.

SSE instruction set support

Palomino-core Athlons support the full Intel SSE instruction set, which AMD designates *3DNow! Professional*. Earlier Athlon processors supported only a subset of SSE and so could not set the processor flag to indicate full support. That meant that SSE-capable software could not use SSE on AMD processors, which in turn meant that AMD processors ran SSE-capable software much more slowly than did Intel SSE-capable processors. Palomino-core Athlons set the SSE flag to true, which allows software to use the full SSE instruction set (but *not* the SSE2 instruction set supported by Intel Pentium 4 processors). Also note that although Palomino-core Athlons support the full SSE instruction set, all that means is that they can run SSE-enabled software. It does not necessarily mean that they run SSE-enabled software as fast as a Pentium III or Pentium 4 processor does.

Reduced power consumption

Palomino-core Athlons have an improved design that reduces power consumption by 20% relative to Thunderbird, which reduces heat production and allows the Palomino core to achieve higher clock speeds than the Thunderbird core.

> Rather oddly, Morgan-core Durons (based on the Athlon Palomino core) actually draw more current than the older Spitfire-core Durons (based on the Athlon Thunderbird core). In fact, Morgan-core Durons draw the same current as Palomino-core Athlons operating at the same clock speed, which leads us to believe that Morgan-core Durons are literally simply Palomino-core Athlons with part of the L2 cache disabled.

Thermal diode

Palomino-core Athlons are the first AMD processors that include a thermal diode, which is designed to prevent damage to the processor from overheating by shutting down power to the processor if it exceeds the allowable design temperature. Intel processors have included a thermal diode for years. It is nearly impossible to damage an Intel Pentium III or Pentium 4 processor by overheating, even by so extreme a step as removing the heatsink/fan from the processor while it is running. Pentium III systems crash when they overheat badly, but the processor itself is protected from damage. Pentium 4 systems don't even crash, but simply keep running, albeit at a snail's pace. The AMD thermal diode, alas, is an inferior implementation. Although the thermal diode on an AMD processor can shut down the CPU safely when heat builds gradually (as with a failed CPU fan), it does not react quickly enough to protect the processor against a catastrophic overheating event, such as the heatsink falling off.

> The Godzilla-size heatsink/fan units used on modern high-speed processors cause catastrophic heatsink/fan unit failures more often than you might think. Whereas Pentium 4 processors use a heatsink/fan retention mechanism that clamps securely to the motherboard, AMD processors still depend on heatsink/fan units that clamp to the CPU socket itself, which isn't designed to support that much weight, particularly in a vertical configuration such as a minitower system. If the heatsink/fan unit comes loose, as it may do when the system is shipped or moved, an AMD processor will literally burn itself to a crisp within a fraction of a second of power being applied. We're talking smoke and flames here. This problem is one of the major causes of AMD systems arriving DOA, but may also occur anytime you move an AMD system. So, if you move an AMD system or if you've just received a new AMD system, *always* take the cover off and make sure the heatsink/fan unit is still firmly attached *before* you apply power to the system. You have been warned.

Although the Athlon XP included some significant technical enhancements over the Thunderbird-core Athlon, the change that received the most attention was AMD's decision to abandon clock speed labeling and instead designate Athlon XP models with a *Performance Rating* (*PR*) system

AMD K7-, K75-, and Thunderbird-core Athlon processors were labeled with their actual clock speeds. AMD Palomino-core and later Athlon XP processors use AMD's QuantiSpeed designations, which are simply a revival of the hoary PR system. Although AMD claims that these PR numbers refer to relative performance of Palomino-core processors versus Thunderbird-core processors, most observers believe that AMD hopes consumers will associate Athlon XP model numbers with Pentium 4 clock speeds. For example, although the AMD Athlon XP/2800+ processor actually runs at 2250 MHz, we think AMD believes buyers will at least subconsciously associate the 2800+ model number with the Pentium 4/2.8G, which does in fact run at a 2800 MHz clock speed.

AMD has gone to great pains to conceal the actual clock speed of Athlon MP processors from users. For example, it mandates that the actual clock speed not appear in advertisements, and has actually gone so far as to insist that system and motherboard makers modify the BIOS to ensure that it reports only the model number and not the actual clock speed. It's interesting that AMD trumpeted its faster clock speeds until Intel overtook AMD and left AMD in the dust in terms of clock speeds. Now that AMD can no longer match Intel's clock speeds, clock speeds are no longer important. Or so says AMD.

Table 4-4 lists the important characteristics of Socket A Athlon variants as of July 2003. Note that AMD has produced two Thoroughbred B processors using the same 2600+ designation. One runs at 2133 MHz on a 266 MHz FSB and the other at 2083 MHz on a 333 MHz FSB. All Socket A Athlon variants use a 64-bit back-side (L2 cache) bus running at full CPU speed and use a shared voltage rail for VCORE and VI/O. For more information about these processors, see *http://www.amd.com*.

Table 4-4. AMD Socket A Athlon variants

	Athlon	Athlon XP	Athlon XP	Athlon XP	Athlon XP
Core	Thunderbird	Palomino	Thoroughbred	Thoroughbred B	Barton
Model	4	6	8 (CPUID 680)	8 (CPUID 681)	10
Production dates	2000, 2001	2001 -	2002, 2003	2002, 2003	2003 -
Clock speeds (MHz)	750, 800, 850, 900, 950, 1000, 1100, 1133, 1200, 1300, 1333, 1400	1333, 1400, 1466, 1533, 1600, 1666, 1733	1467, 1533, 1600, 1667, 1733, 1800	1667, 1800, 2000, 2083 (333), 2133 (266), 2166, 2250	1833, 2083, 2166, 2200
Model designation	n/a	1500+, 1600+, 1700+, 1800+, 1900+, 2000+, 2100+	1700+, 1800+, 1900+, 2000+, 2100+, 2200+	2000+, 2200+, 2400+, 2600+ (333), 2600+ (266), 2700+, 2800+	2500+, 2800+, 3000+ 3200+
L2 cache size	256 KB	256 KB	256 KB	256 KB	512 KB
System bus speed (MHz)	200, 266	266	266	266, 333	333
Voltage (V)	1.7, 1.75	1.75	1.5, 1.6, 1.65	1.5, 1.6, 1.65	1.65
Fabrication process	0.18µ	0.18µ	0.13µ	0.13µ	0.13µ
Interconnects	Al/Cu	Cu	Cu	Cu	Cu
Die size (mm^2)	120	128	81, later 84	84	101
Transistors (million)	37	37.2	37.5	37.6	54.3

Other AMD processors

AMD has produced two special-purpose variants of the Athlon, the Duron and the SMP-certified Athlon MP:

Duron

The Duron was AMD's answer to the low-end Intel Celeron. Just as Intel introduced the Celeron in an attempt to maintain a high average selling price for its flagship Pentium III and Pentium 4 processors, AMD introduced the Duron as a "value" version of the Athlon. AMD has produced two models of the Duron:

Duron (Spitfire core)

The Duron, code-named Spitfire and for a short time designated Athlon Value, was targeted at the value desktop market, and was to be a Celeron-killer. With it AMD straddled a fine line between matching Celeron clock speeds and performance on the one hand, versus avoiding cannibalizing sales of Athlon processors on the other. Accordingly, AMD differentiated the Duron by limiting the clock speed of the fastest current Duron to one step below the clock speed of the slowest current Athlon, by using a smaller and less efficient L2 cache, and by making the Duron only in 100 MHz FSB versions (versus the 133 MHz or higher FSB available on some Athlon models). The Spitfire-core Duron was an excellent processor for its time. It unquestionably offered more bang for the buck than any other processor sold by AMD or Intel. Although it achieved reasonable sales volumes in Europe, the Duron never really took off in the U.S. because of the absence of high-quality integrated Duron motherboards.

Duron (Morgan core)

The Morgan-core Duron is simply a refresh of the Spitfire Duron to use the newer Palomino core. The advantages of the Morgan-core Duron over the Spitfire-core Duron are analogous to the advantages of the Palomino-core Athlon over the Thunderbird-core Athlon. The Morgan core is essentially a Palomino core with a smaller and less efficient L2 cache. As it did with the Spitfire, AMD carefully managed the Morgan to prevent cannibalizing sales of the Athlon XP. The fastest current Morgan was always at least one step slower than the slowest current Athlon XP. In terms of absolute performance clock for clock, the Morgan slightly outperforms the Coppermine-core Pentium III and the Tualatin-core Celeron.

The Appaloosa-core Duron, based on the Thoroughbred-core Athlon XP, was announced but later canceled. The Duron was a victim of AMD's success with the Athlon. As faster Athlons were introduced at lower prices, the Duron was simply squeezed out of its market niche. The Duron is still available as of July 2003, but is likely to disappear before year end. Figure 4-13 shows an AMD Duron processor.

Athlon MP

Even the first Athlon processors had the circuitry needed to support dual-processor operation. That feature was useless until the introduction of the

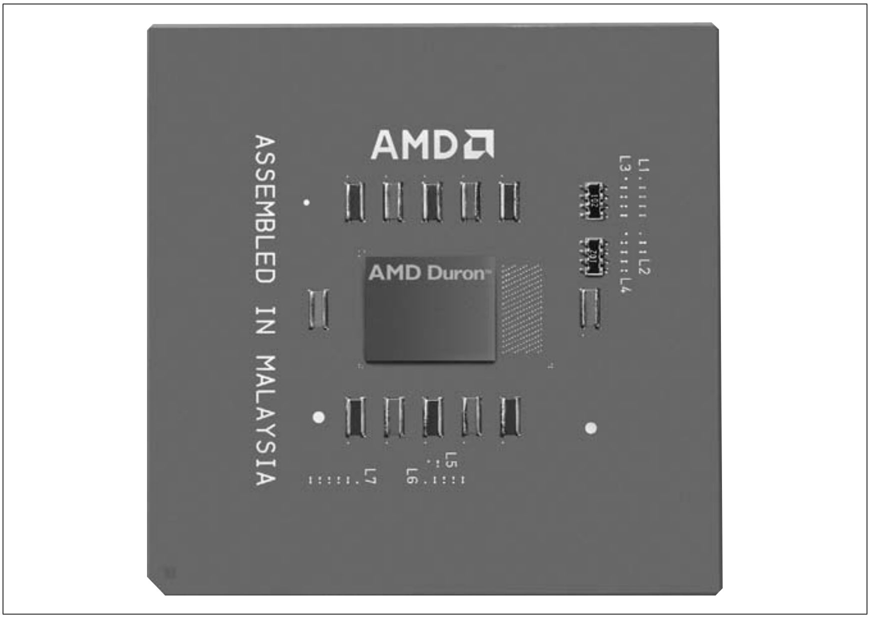


Figure 4-13. AMD Duron processor (image courtesy of Advanced Micro Devices, Inc.)

Processors

AMD-760MP chipset because no prior Athlon chipset supported dual processors. In mid-2001, Tyan shipped its 760MP-based Thunder motherboard. It supported dual Athlons, but was expensive and required a special power supply. In late 2001, Tyan shipped the inexpensive Tiger MP dual Athlon board, which used a standard power supply. Suddenly, dual Athlon systems were affordable, and many enthusiasts set out to build them.

AMD capitalized on this new market by introducing Athlon XP processors certified for dual-processor operation, which they named the Athlon MP. Athlon MP processors are binned (hand-picked and individually tested) for reliable SMP operation, or so the rumor has it. We have our doubts. We and many of our readers have run dual Athlon XPs successfully. Alas, AMD has disabled SMP operation on recent Athlon XP processors. If you want a dual Athlon system using current products, the only option is to use SMP-certified (and more expensive) Athlon MP processors. AMD has made Athlon MP processors using two cores:

Athlon MP (Palomino core)

The first Athlon MP models used the Palomino core. They shipped in June 2001, four months before AMD introduced the first Palomino-core Athlon XP models. At that time, AMD had not yet decided to use model numbers rather than clock speeds to designate its processors, so the first two Athlon MP models were designated the Athlon MP/1.0G and the Athlon MP/1.2G. Those numbers accurately reflect their true clock speeds of 1000 MHz and 1200 MHz, respectively. By October 2001, when AMD began rolling out the new Palomino-core Athlon XPs, it had decided to designate the first model the Athlon XP/1500+, even though its actual clock speed was only 1333 MHz. All subsequent Athlon MP processors are designated by model number rather than clock speed. Functionally, the Palomino-core Athlon MP is identical to the Palomino-core Athlon XP.

Athlon MP (Thoroughbred core)

Functionally, the Thoroughbred-core Athlon MP is identical to the Thoroughbred-core Athlon XP. When AMD transitioned to Thoroughbred-core Athlon XPs, it did not immediately introduce Athlon MP processors based on the Thoroughbred core. Instead, AMD began the staged introduction of Athlon MP processors that continues today. For example, in June 2002, AMD introduced Thoroughbred-core Athlon XP models 1700+ through 2200+. It was not until late August that AMD introduced Thoroughbred-core Athlon MP models at 2000+ and 2200+, just days after it introduced the Athlon XP 2400+ and 2600+. AMD says the delay is needed to certify faster models for SMP operation, which seems to us a reasonable explanation.

Athlon MP (Barton core)

In May 2003 AMD shipped the Athlon MP 2800+, the first Athlon MP based on the Barton core. The 2800+ may also be the final Athlon MP model, because AMD now devotes all of its attention to the Opteron. Functionally, the Barton-core Athlon MP is identical to the Barton-core Athlon XP, including the increase from 256 KB to 512 KB of L2 cache. Interestingly, a few examples of the Athlon MP 2800+ with 333 MHz FSB have surfaced. We don't understand why AMD would produce such a processor. The 760MPX (the only Athlon chipset that supports SMP) supports a maximum FSB speed of 266 MHz, which seems to render a 333 MHz FSB Athlon MP pointless. We can only speculate that AMD plans a refresh of the 760MPX to add support for the 333 MHz FSB.

Table 4-5 lists the important characteristics of Socket A Duron and Athlon MP variants as of July 2003. For more information about these processors, see *http://www.amd.com*.

Table 4-5. Socket A Duron and Athlon MP variants

	Duron	Duron	Athlon MP	Athlon MP	Athlon MP	Athlon MP
Core	Spitfire	Morgan	Palomino	Palomino	Thorough-bred	Barton
Model	3	7	6	6	8	10
Production dates	2000 - 2001	2001 - 2003	2001 - 2002	2001 - 2002	2002 -	2003-
Clock speeds (MHz)	600, 650, 700, 750, 800, 850, 900, 950	1000, 1100, 1200, 1300	1000, 1200	1333, 1400, 1533, 1600, 1667, 1733	1667, 1800, 2000, 2133	2133
Model designation	n/a	n/a	n/a	1500+, 1600+, 1800+, 1900+, 2000+, 2100+	2000+, 2200+, 2400+, 2600+	2800+

Table 4-5. Socket A Duron and Athlon MP variants (continued)

	Duron	Duron	Athlon MP	Athlon MP	Athlon MP	Athlon MP
L2 cache size	64 KB	64 KB	256 KB	256 KB	256 KB	512 KB
System bus speed	200 MHz	200 MHz	266 MHz	266 MHz	266 MHz	266 MHz
Voltage (V)	1.5V, 1.6V	1.75V	1.75V	1.75V	1.6V, 1.65V	1.6V
Dual CPU-capable	❍	❍	●	●	●	●
Fabrication process	0.18µ	0.18µ	0.18µ	0.18µ	0.13µ	0.13µ
Interconnects	Cu	Cu	Cu	Cu	Cu	Cu
Die size (mm^2)	100	106	128	128	85	101
Transistors (million)	25.00	25.18	37.2	37.2	37.5	54.3

Choosing a Processor

The processor you choose determines how fast the system runs, and how long it will provide subjectively adequate performance before you need to replace the processor or the system itself. Buying a processor just fast enough to meet current needs means that you'll have to upgrade in a few months. But processor pricing has a built-in law of diminishing returns. Spending twice as much on a processor doesn't buy twice the performance. In fact, you'll be lucky to get 25% more performance for twice the money. So although it's a mistake to buy too slow a processor, it's also a mistake to buy one that's too fast. Consider the following issues when choosing a processor:

Horizon
: What kind of applications do you run and how long do you want the system to be usable without requiring an upgrade? If you run mostly standard productivity applications and don't upgrade them frequently, a low-end processor may still be fast enough a year or more after you buy it. If you run cutting-edge games or other demanding applications, buy a midrange or faster processor initially, and expect to replace it every six months to a year. But expect to pay a price for remaining on the bleeding edge.

Hassle
: Do you mind upgrading your system frequently? If you don't mind replacing the processor every six to 12 months, you can get most of the performance of a high-end system at minimal cost by replacing the processor frequently with the then-current midrange processor. In the past, this was easier with AMD processors because AMD has used Socket A for years and had standardized on 100/200 MHz and 133/266 MHz FSBs. It was sometimes possible to install a current processor in a two-year-old motherboard with only a BIOS upgrade.

Intel made things much more difficult, replacing Socket 370 with Socket 423 and then Socket 478, and introducing faster FSB speeds frequently. Although many considered these changes as cynical planned obsolescence, in fact these changes resulted simply from Intel's much faster product development cycle. The situation is different now. Intel has stabilized around Socket 478 and the 800 MHz FSB (although the forthcoming Prescott processors will use a different socket), and AMD is in a state of flux. AMD recently introduced the 166/333 MHz and 200/400 MHz FSBs for the Athlon, which will rapidly render older motherboards obsolete. Also, AMD has deemphasized Athlon product development in favor of its forthcoming Hammer-series processors, which are entirely incompatible with the Athlon series. On balance, Intel actually offers a better upgrade path for now, although that may change depending on the decisions AMD makes with regard to Hammer-series processors.

Trade-offs

If you're working on a fixed budget, don't spend too much on the processor to the detriment of the rest of the system. Instead of spending $300 on a fast processor and compromise on the other components, you're better off spending $150 on a midrange processor and using the other $150 to buy more memory, a faster hard disk, and better video. A low-end Pentium 4 with lots of memory, a fast hard drive, and a good video adapter blows the doors off the fastest Pentium 4 with inadequate memory, a slow hard drive, and a cheesy video card every day of the week. Don't make yourself "processor-poor."

Form factor

Keep form factor in mind when you're shopping for a processor, particularly if you're also buying a motherboard:

Socket 7

Don't consider buying a Socket 7 processor, even as an inexpensive upgrade to a working system. Any money spent on Socket 7 is wasted. Retire the old system to less-demanding duties, and build or buy a new system instead.

Slot 1

Slot 1 was obsolete by the end of 2001. Although new Slot 1 processors remain in limited distribution, new Slot 1 motherboards are now almost impossible to find. An existing Slot 1 system may or may not be a good upgrade candidate depending on the motherboard characteristics. Some Slot 1 motherboards support fast Pentium III processors, and can be upgraded at reasonable expense. For example, we recently upgraded an older Pentium II server to a Pentium III using a salvaged processor. Because we used a relatively slow Pentium III processor, even if we had to buy the processor, the total upgrade cost would have been about $75. Performance more than doubled, which gives that server another two years or more of useful life.

Other Slot 1 motherboards have neither BIOS support nor adequate VRMs to support faster processors. Although it's possible to upgrade those systems with marginally faster Slot 1 processors, doing so makes no economic sense. Before you upgrade any Slot 1 system, check prices carefully. Some Slot 1 processors are very expensive relative to the performance boost they provide. You may be able to replace the motherboard, processor, *and* memory with Socket 478 Pentium 4 or Socket A Athlon components for little more than the cost of the Slot 1 processor alone.

Slot A

Like Intel processors, AMD Athlon processors were originally produced in slotted versions, which were subsequently replaced by socketed versions. Slot A motherboards and processors are now almost impossible to find, and any Slot A motherboard is now so old that it is a poor upgrade candidate. If for some reason you must replace the processor in a Slot A system, pay careful attention to the chipset it uses. Motherboards based on the AMD-750 chipset can use Slot A processors based on the K7, K75, and Thunderbird cores (although Slot A Thunderbirds are difficult to find). Motherboards based on the VIA KX133 chipset are incompatible with Slot A Thunderbird Athlon processors, but can use Athlons based on the K7 and K75 cores. As of July 2003, Slot A processors are still in limited distribution, but soon the only alternative will be the used market.

Socket 370

As of July 2003 Socket 370 is moribund. Intel pulled out all the stops to push the Pentium 4 at the expense of its sixth-generation Celeron and Pentium III processors, and by mid-2002 Socket 370 was no longer a mainstream technology. Intel still offers a limited selection of Socket 370 Celeron and Pentium III processors. Alas, Intel no longer makes Socket 370 motherboards, so third-party motherboard makers are now the only source for new Socket 370 motherboards.

Although it no longer makes economic sense to build a new Socket 370 system, existing Socket 370 systems may be economically upgradeable. When upgrading an older Socket 370 system, verify compatibility between your motherboard and the Socket 370 processor you propose to buy. There are many incompatibilities between older motherboards and newer processors. Some problems can be solved with a simple BIOS update, but many are unsolvable because the older motherboard's chipset or VRMs do not support newer Socket 370 processors.

Socket A

In the past, AMD did a much better job than Intel at maintaining backward compatibility. Intel changed sockets and FSB speeds frequently, but AMD just kept using Socket A and the standard 100/200 and 133/266 MHz FSB speeds. The Hammer-series processors, due later in 2003, will change that, but Socket A motherboards and processors will remain available for at least the next year or two. As long as you don't mind buying into an obsolescent technology, Socket A remains a good choice for a new system until Hammer-series processors and motherboards become inexpensive mainstream products.

Older Socket A systems may or may not be good upgrade candidates. In general, older-model Socket A motherboards can use newer Socket A processors, although perhaps not the fastest models. A Socket A system that supports only the 200 MHz FSB is probably too old to be economically upgradeable. For such systems, replace the motherboard, processor, and memory with current products. Most Socket A systems that support the 266 MHz FSB or higher and that support at least PC2100 DDR-SDRAM are excellent upgrade candidates. By replacing an older Duron or Athlon processor with a current low-end Athlon, you may be able to double system performance for much less than $100. Before you make such an upgrade, verify that your motherboard supports the specific processor model and speed that you plan to install. You will probably need to upgrade the BIOS as well.

If your goal is to build a dual-processor system, your best option is a pair of Socket A Athlon MP processors running in an AMD-760MPX based motherboard. As always, an older motherboard may have BIOS or VRM issues with newer processors, so you still need to verify compatibility.

> Always verify the cooling requirements of a replacement processor. The existing CPU heatsink/fan unit may fit the new processor, but that's no guarantee that it is adequate to cool the new processor adequately. We almost learned that the hard way. In late 2002, AMD sent us a preproduction sample of its new 333 MHz FSB Athlon 2600+, including just the bare CPU. We verified that the ASUS A7N8X Deluxe motherboard supported the 2600+, but we didn't think about the heatsink. We'd already squirted thermal goop onto the processor and were about to install an off-the-shelf heatsink when we remembered that we'd gotten in some sample heatsinks from DynaTron, and decided to try one of those. That was fortunate because as we were reading the DynaTron literature we realized that the heatsink we were about to use was rated only for XP 2000+ and slower Athlons. If we'd installed that heatsink and powered up the system, our shiny new 2600+ processor might have burnt itself to a crisp in seconds. Processors aren't much good if you let the smoke out.

Socket 423

Socket 423 was Intel's first socket for the Pentium 4, and was simply a stopgap solution that allowed Intel to bring Pentium 4 processors to market quickly to compete with the AMD Athlon on clock speed. Socket 423 processors and motherboards are obsolete. Socket 423 motherboards are nearly impossible to find, although Socket 423 processors remain in limited distribution. A Socket 423 system is a poor upgrade candidate because the fastest available Socket 423 processor will be little or no faster than the processor already installed. Replacing the motherboard, processor, and memory is a far better solution.

Socket 478

A Socket 478 processor is the best choice if you are building a new mainstream system. An existing Socket 478 system can easily be upgraded simply by dropping in a faster Socket 478 processor, a condition that is likely to remain true for some time. As always, it's possible that BIOS, chipset, and VRM issues may restrict the speed of the fastest Socket 478 processor that can be installed in a particular motherboard, but Socket 478 currently offers the best options for future upgradability.

When upgrading a system, the existing motherboard determines upgradability, as follows:

Socket 7 and earlier motherboards
: These motherboards are simply too old to upgrade economically. We recommend retiring such ancient systems, or discarding them entirely.

Slot 1 motherboards
: Slot 1 Pentium II and Celeron processors remain in limited distribution, although we expect them to disappear entirely by the end of 2003 or early 2004. Fortunately, some Slot 1 motherboards can be upgraded by using a slocket adapter, which accepts a Socket 370 processor and plugs into the motherboard Slot 1. The best candidates for such upgrades are motherboards designed for the Pentium III that support the 100 MHz or 133 MHz FSB. Even if a particular motherboard can be upgraded via slocket, it may be limited by BIOS, chipset, or VRM issues as to which particular Socket 370 processors are usable. In general, FC-PGA Celerons are the most likely to work, assuming that the motherboard supports the Celeron L2 caching method. An FC-PGA Coppermine-core Pentium III may or may not work, depending on the particular slocket/processor combination and the chipset and BIOS configuration of the motherboard. We know of no slocket that allows FC-PGA2 Celerons and Pentium IIIs to be used in Slot 1 motherboards. Before you attempt to upgrade a Slot 1 motherboard with a slocket, verify with the slocket maker that the slocket, processor, and motherboard you plan to use are compatible.

Slot A motherboards
: Slot A processors are now almost impossible to find new. Slot A motherboards are now so old that it makes no sense to spend money upgrading them. Instead, replace the processor, motherboard, and memory with current products. You can buy a decent Socket A processor, motherboard, *and* memory for less than $200, which makes messing around with an obsolete processor and motherboard a complete waste of time.

Socket 370 motherboards
: Upgrading a Socket 370 system *should* be easy. Unfortunately, it often isn't. The problem with upgrading Socket 370 motherboards is that there have been so many variants of the socket itself and processors intended to fit it that determining compatibility can be difficult. Any Socket 370 processor physically fits any Socket 370 socket, but there are actual pinout differences between early Socket 370 sockets and processors and later versions. Late-model Socket 370 processors—Coppermine- and Tualatin-core Celerons and Pentium IIIs—will not operate in early-model Socket 370 motherboards, and early-model Socket 370 processors—Mendocino-core Celerons and Katmai-core Pentium IIIs—may or may not operate in later-model Socket 370 motherboards. In addition, chipset issues are important with Socket 370 because

early Socket 370 chipset revisions do not support later Socket 370 processors, even though the processor is otherwise compatible electrically and physically with the socket. Intel rationalized this situation in late 2001 by introducing its so-called "Universal" Socket 370 motherboards, which can accept any Socket 370 processor. If you intend to upgrade the processor in a Socket 370 system, the best advice is first to determine exactly what motherboard you have (including revision level). Once you've done that, visit the motherboard maker's web site and read the technical documentation to determine which currently available Socket 370 processors can be used in that motherboard.

Socket A, Socket 423, and Socket 478

Motherboards that use any of these sockets can be upgraded using current processors. Socket 423 is a poor upgrade candidate because only relatively slow processors are available for it. Socket A and Socket 478 motherboards are generally good upgrade candidates because there are numerous models of fast, inexpensive processors available for both of them. As always, check the documentation for the motherboard to ensure that it supports the type, FSB speed, and clock speed of the processor you plan to install. Ordinarily, such upgrades are relatively straightforward, requiring a BIOS upgrade at most.

Forthcoming AMD and Intel Processors

Intel and AMD constantly strive to out-do each other in bringing faster and more capable processors to market. In late 2003 and into 2004, each company will be ramping up its new-generation desktop processors. Although the current Athlon XP and Pentium 4 processors will continue to sell in large numbers throughout 2003 and into 2004, the future definitely belongs to these new processor lines. AMD hopes to get a foothold in the corporate market and to increase their general market share with their new desktop processors, but Intel has some plans of its own to protect its 80%+ general market share and its nearly 100% corporate market share.

As we write this in July 2003, only the Opteron processor is shipping, and only in limited numbers. The Athlon 64 and the Prescott/Pentium 5 are not yet shipping and we have been unable to get preproduction samples from AMD and Intel. Accordingly, much of this section is speculative, based on published information that is subject to change, industry rumors, and informed speculation. However, we thought it worthwhile to include the best information we had available as we went to press, because even imperfect or incomplete information may be useful to our readers.

AMD Opteron and Athlon 64

By mid-2002, AMD was struggling to produce Athlons that could match Pentium 4 performance. By July 2003, it was obvious to nearly everyone that the Athlon XP had reached the end of the line and that the 3200+ would almost certainly be the final Athlon XP processor. AMD was able to push the Athlon core further than anyone expected, eventually reaching a core clock speed of 2.2 GHz in the Barton-core Athlon XP 3200+ model. AMD also expanded L2 cache from 256 KB on earlier cores to 512 KB on the Barton core, and increased FSB speeds from 266 MHz to 333 MHz and eventually to 400 MHz on the final Athlon XP models.

But all of these enhancements yielded only marginal performance improvements over earlier Athlon models. The real problem was that the Athlon core itself had reached its limits, while Intel's Pentium 4 core wasn't even breathing hard. AMD badly needed an entirely new processor core if they were to compete with Intel on anything like a level playing field.

In April 2003, AMD shipped their new-generation processor, code-named K8 or Sledgehammer, officially named Opteron, and ironically dubbed "Lateron" by pundits because of the repeated and lengthy delays AMD suffered in bringing this processor to market. (Nor is AMD alone in having evil nicknames applied to its processors. Some wags called the original Itanium 1 the "Itanic" because, like its namesake, it sank without a trace.)

AMD will produce two processor lines based on the K8 core. The Opteron is intended for servers, and began shipping in April 2003. The Athlon 64 is a cut-down version of the Opteron intended for desktop systems, and is to begin shipping in September 2003. The key feature of both processors is that they support both 32-bit and 64-bit instructions, and can dynamically alternate 32- and 64-bit threads.

In contrast to the 64-bit Intel Itanium, which executes 64-bit code natively but 32-bit IA-32 code only via slow translation, the Opteron and Athlon 64 are 64-bit processors that can execute 64-bit code using the AMD64 instruction set—called "long" mode—and can also execute standard 32-bit code natively, called "legacy" mode. To support 32- and 64-bit operations in one processor, AMD modified the Athlon XP core to add eight 64-bit general-purpose registers and eight 64-bit versions of the original eight 32-bit general purpose registers. These 64-bit registers are accessible only when the processor is operating in long mode. In legacy mode, the Opteron and Athlon 64 processors appear to 32-bit software as a standard 32-bit Athlon processor.

The Opteron and Athlon 64 are incompatible with current chipsets and motherboards, so using either requires buying or building a new system. As of July 2003, Opteron systems and motherboards are in limited distribution. We expect Athlon 64 products to become available in September 2003.

Opteron

The Opteron is based on the variant of the K8 core codenamed Sledgehammer. Various Opteron models support 1-, 2-, 4-, and 8-way operation and are targeted at servers. AMD plans to produce at least three Opteron series. Opteron 100-series processors support only 1-way processing, and are due in September 2003. Opteron 200-series processors support 1- and 2-way processing, and shipped in April 2003. Opteron 400-series processors support 1-, 2-, 4-, and 8-way processing, and are to ship in September 2003 and into 2004.

Rather than the clock speed designations or QuantiSpeed model numbers AMD used for earlier processors, AMD assigns each Opteron model an arbitrary number to indicate relative performance. For example, the Opteron processor roadmap includes the 140, 240, and 840 models, which operate at 1.4 GHz; the 1.6 GHz 142, 242, and 842 models; and the 1.8 GHz 144, 244, and 844 models. AMD plans to release later Opteron models operating at 2.0 GHz (presumably the 146, 246, and 846 models), as well as models operating at 2.2 GHz (148, 248, and 848).

Opteron processors use 6.4 GB/s HyperTransport Technology (HTT) channels to provide a high-speed link between the processor components themselves and to the outside world. The Opteron has three HTT channels, which may be either of two types. Coherent HTT channels link the processor to other Opteron processors. Opteron 100-series, 200-series, and 800-series processors have zero, one, or three coherent HTT channels, respectively. Standard HTT channels link the processor to I/O interfaces such as a Southbridge or PCI Express bridge.

> Do not confuse AMD HTT (HyperTransport Technology) with Intel HTT (Hyper-Threading Technology). You'd think they could come up with different TLAs. It isn't like there aren't lots of letters to choose from.

The Opteron features a 1024 KB L2 cache and a dual-channel DDR333 memory controller, which uses a 144-bit interface that requires 72-bit ECC memory. Relocating the memory controller from the chipset, where it has traditionally resided, directly onto the processor core allows memory to be more tightly integrated with the processor for higher performance. The downside is that the Opteron is limited to using memory no faster than DDR333 unless AMD changes the processor core itself, or unless a chipset maker adds an external memory controller.

> Informed sources speculate that AMD may tweak the shipping K8 core to add support for DDR400 and perhaps DDR533. Support for DDR-II will come no earlier than mid-2004, pending JEDEC approval of a final DDR-II specification.

The Opteron uses Socket 940, newly introduced by AMD for this processor. Relative to Socket 462, those extra contacts are used primarily to support the three HTT channels.

Athlon 64

The Athlon 64 processor is based on the variant of the K8 core codenamed Clawhammer. The Athlon 64 supports 1- and 2-way operation, is due in September 2003, and is targeted at desktop systems. The Athlon 64 differs from the Opteron in the following important respects:

HyperTransport Technology channels
: Rather than the three HTT channels used by the Opteron, the Athlon 64 has only one HTT channel.

Memory controller
: Rather than the 144-bit dual-channel DDR333 ECC memory controller used by the Opteron, the Athlon 64 has a 64-bit single-channel DDR333 non-ECC

memory controller. (Shipping models may include DDR400 support.) The narrower memory interface of the Athlon 64 means its memory bandwidth is half that of the Opteron. Like the Opteron, the Athlon 64 integrates the memory controller onto the processor.

Cache size
: The Athlon 64 and Opteron both have the AMD-standard 128 KB L1 cache, with 64 KB allocated to instructions and 64 KB to data. Opteron processors provide 1 MB of L2 cache. Athlon 64 processors are available with either 256 KB or 1 MB L2 cache. Our moles tell us that for performance reasons, AMD may decide to ship the "small" Athlon 64 with 512 KB L2 cache rather than 256 KB.

Chipset support
: Most Opteron systems will be built around the server-class AMD 8000-series chipset. Most Athlon 64 systems will use desktop-class chipsets such as the *n*VIDIA *n*Force3, the VIA K8T800/K8M800, and others. Based on our experiences with the *n*Force and *n*Force2 Athlon chipsets, we expect the *n*Force3 to be the best Athlon 64 chipset.

The Athlon 64 uses Socket 754, another new AMD socket. As with Socket 940, the additional contacts are necessary to support the single HTT channel supported by the Athlon 64. Because the Athlon 64 has only one HTT channel, it can use the smaller socket.

Table 4-6 details the important characteristics of the Opteron and Athlon 64 processors, with the Barton-core Athlon XP shown for comparison. Most of the items are self-explanatory, but a couple deserve comment.

Generation
: AMD regards the Athlon XP as seventh-generation and the Opteron/Athlon 64 as eighth-generation. We consider both of those processor families to be hybrids, straddling the generational boundaries defined by Intel processors. In particular, the 64-bitness of the Opteron and Athlon 64 give them a definite claim to eighth-generation status, but architecturally they remain near relatives of the hybrid sixth/seventh-generation Athlon XP.

Fabrication process
: With the Opteron and Athlon 64, AMD uses the Silicon-on-Insulator (SOI) process rather than the traditional CMOS process. SOI offers potentially huge benefits, but at a correspondingly high risk. During the first half of 2003, AMD's problems with SOI in getting high yields at fast clock speeds were widely reported in the industry press. We think the most important issue for the new AMD processors is how well and how quickly the AMD Dresden fab will be able to master SOI production. If they succeed, they will produce high yields of the new processors and be able to scale clock speeds up quickly. If they fail, the Opteron and Athlon 64 will be expensive to produce and will languish at lower clock speeds. The phrase "bet the company" is often used in the high technology field, but in this case we think AMD is indeed betting the company on the success of their SOI process.

Table 4-6. Characteristics of Opteron and Athlon 64 versus Athlon XP

	Opteron	Athlon 64	Athlon XP
Core	Sledgehammer	Clawhammer	Barton
Generation	7th/8th	7th/8th	6th/7th
CPU Socket	940	754	462
Production dates	April 2003 –	September 2003 –	February 2003 –
Clock speeds (MHz)	1400, 1600, 1800	1600, 1800, 2000	1833, 2083, 2133, 2200
Model designation	240, 242, 244	3400+, 3600+, 3800+	2500+, 2800+, 3000+, 3200+
L2 cache size	1024 KB	256, 512 (?), or 1024 KB	512 KB
External bus speed	333 MHz DDR-SDRAM 19.2 GB/s HTT (triple)	333 MHz DDR-SDRAM 6.4 GB/s HTT (single)	333, 400 MHz DDR-SDRAM EV-6
Instruction set	IA-32/AMD64	IA-32/AMD64	IA-32
Multimedia support	MMX, 3DNow!, SSE, SSE2	MMX, 3DNow!, SSE, SSE2	MMX, 3DNow!, SSE
Voltage (V)	1.55	1.55	1.65
Fabrication process	0.13 (CMOS, SOI)	0.13 (CMOS, SOI)	0.13 (CMOS)
Interconnects	Cu	Cu	Cu
Die size	193 mm^2	104 mm^2	101 mm^2
Transistors (million)	105.9+	67	54.3

Intel Pentium 5?

Intel and AMD play a constant game of leapfrog. The introduction of the Opteron/Athlon 64 almost demanded that Intel introduce a new processor of its own. That processor is the Prescott-core Pentium, due in the fourth quarter of 2003, which Intel may or may not call the Pentium 5.

On balance, we think Intel will decide to name their new processor the Pentium 5, both for marketing reasons and for technical reasons. From a marketing standpoint, Intel would clearly like to counter the Opteron and Athlon processors with a newly-named processor of their own. From a technical standpoint, the improvements in architecture and instruction set are sufficient to justify the Pentium 5 name for the Prescott-core processor.

No matter what Intel chooses to call this processor, it is a significant improvement on the current Northwood-core Pentium 4. Relative to current Northwood-core processors, the Prescott-core processors increase L1 cache size, boost L2 cache from 512 KB to 1024 KB (matching the new AMD processors), and increase pipeline depth to enable higher core frequencies.

Just those enhancements would have made life difficult for the new AMD processors. But a more significant enhancement lurks within Prescott. The Prescott New Instructions (PNI) are 13 new instructions that extend the SSE and SSE2 multimedia instruction sets used by earlier Intel processors. In particular, three of the new PNI instructions are worth noting. One adds support for AV encoding—as opposed to AV decoding, which was supported by earlier Intel processors—and two improve thread control for Hyper-Threading Technology (HTT) operations.

The new HTT thread control instructions are likely to boost performance substantially, with less sensitivity to application mix. In the past, the benefit of HTT depended largely on the specific applications being run. Some applications showed major performance improvements with HTT, most applications showed no change, and some actually ran slower with HTT enabled. The improved HTT threading support available with PNI means that HTT will become more generally useful. For more information about PNI, visit *http://cedar.intel.com/media/pdf/PNI_LEGAL3.pdf*.

Prescott-core processors may also have a major hidden feature. We admit that this is pure speculation on our part, but we do have some historical evidence for our beliefs. Intel built Hyper-Threading Technology into Northwood-core processors, where it remained hidden until Intel chose to reveal it. We think history may repeat itself. Intel may have embedded their Yamhill technology into Prescott as a hidden feature.

Intel's world view is that 32-bit processors are sufficient for desktop systems, that only datacenters require 64-bit processors, and that 64-bit processors should operate natively in 64-bit mode rather than as 32/64-bit hybrids. But Intel always has a Plan B, and in this case Plan B is Yamhill. Yamhill is, in effect, Intel's version of AMD's hybrid AMD64 architecture. Intel would prefer to drive people to its native 64-bit Itanium architecture. But if that fails and AMD64 catches on, Intel can spring Yamhill as a nasty surprise to AMD. Don't be surprised if that happens.

Table 4-7 shows the important characteristics of the Prescott-core "Pentium 5", with the Northwood-core Pentium 4 shown for comparison.

Table 4-7. Characteristics of Prescott "Pentium 5" versus Pentium 4

	"Pentium 5"	Pentium 4
Core	Prescott	Northwood "A"
Generation	7th/8th	7th
CPU Socket	478, 775	478
Production dates	October 2003 (?) –	November 2002 –
Clock speeds (MHz)	3200, 3400, and higher	2400, 2600, 2800, 3000, 3060, 3200
L2 cache size	1024 KB	512 KB
External bus speed	800, 1066, 1200 MHz	400, 533, 800 MHz
Instruction set	IA-32/Yamhill-64 (?)	IA-32
Multimedia support	MMX, SSE, SSE2, PNI	MMX, SSE, SSE2
Voltage (V)	1.25	1.500, 1.525, 1.550
Fabrication process	0.09 (CMOS)	0.13 (CMOS)
Interconnects	Cu	Cu
Die size	109 mm^2	131 mm^2
Transistors (million)	~ 100	55

Our Thoughts

We won't comment in detail on server processors, because we don't understand that market well enough. We note, however, that IT managers are notoriously conservative in adopting new platforms, and the perception of Intel as the tried-and-true 64-bit solution, particularly with regard to chipsets, probably militates against the broad acceptance of the Opteron in the datacenter. We're sure that the Opteron will have some "wins", but overall we think that 32-bit Intel processors will continue to dominate PC-server space. Those who need the additional memory addressability and other features of 64-bit processors will probably continue using heavy iron, at least in the short term.

On the desktop side, the picture isn't much better for AMD. We think the Intel Pentium 5 (or whatever Intel chooses to call it) will walk all over the Athlon 64. Although the Athlon 64 runs 32-bit code competently—something Intel has never been able to achieve with its 64-bit processors—its forte is 64-bit operations, and for now 32-bit operations are sufficient for the desktop. The only 64-bit operating system available is Linux, although Microsoft promises a 64-bit Windows Real Soon Now. Even if that comes to pass, the dearth of 64-bit applications programs means that the Athlon 64 will be operating in 32-bit mode nearly all the time.

Considered as a 32-bit processor, the Athlon 64 is in effect a slightly enhanced Athlon XP. It operates at a severe disadvantage relative to the Prescott-core Pentium. AMD had severe teething pains getting the K8 core running faster than 1.8 GHz, and we do not expect the K8 core to scale nearly as well as the new 0.09µ Intel core. We think it likely that when the new Intel core debuts at 3.4 GHz, it will match or exceed the fastest Athlon 64 model in most 32-bit operations. And, while AMD has to work very hard for each increment in Athlon 64 clock speed, we expect the new Intel core to scale effortlessly to 5 GHz or faster.

Although we admire AMD and appreciate the results of their competition with Intel, we're forced to conclude that AMD is likely to be an also-ran in the desktop processor race throughout 2003 and well into 2004. The arrival of 64-bit Windows and 64-bit applications may help somewhat, but we think it will be insufficient to turn the tide. Certainly, 64-bit processing (and memory addressability) will be a blessing for some people. Those who work with huge databases or do serious image processing and video work can use every bit of horsepower and memory they can get. But for the most part we think 64-bit processing for the desktop is a technology of the future, and is unlikely in the short term to create a large demand for the new 64-bit AMD processors.

Installing a Processor

The following sections describe the steps required to install and configure standard slotted and socketed processors. The steps we describe are generally applicable to any modern processor of a given type, but the details may vary slightly between different processors, particularly with regard to such things as configuring the motherboard and installing heatsink/fan units. If this is the first time you've installed a processor, or if you are in doubt about any step, refer to the documentation provided by the manufacturer of your specific processor and motherboard.

Before you install any processor, make sure that you have identified exactly both the processor itself and the motherboard you plan to install it in. If the processor is not new, you can identify it using the steps described earlier in this chapter. All high-quality motherboards have information printed on the board itself that identifies the manufacturer, model, and revision number. If the board does not contain such information, you may be able to identify the board by writing down the full BIOS string displayed by the BIOS boot screen and checking that string against one of the BIOS sites listed in Chapter 3. However, such "anonymous" boards are generally of very low quality, so it's usually better to replace such a board rather than attempt to use it.

> Before you install a processor, make absolutely sure the processor is compatible with the motherboard. It is not safe to assume merely because the processor fits the socket or slot that that processor will function properly in that motherboard. In some cases, the processor simply will not work. For example, there are many incompatibilities between Socket 370 processors and motherboards. Not all Socket 370 processors can be used in all Socket 370 motherboards, even if the processor and the motherboard were both made by Intel. In that situation, no damage is done. The processor simply doesn't work.
>
> There are, however, two common situations in which installing an incorrect processor may damage the processor and/or the motherboard:
>
> - Installing a fast processor in a motherboard designed to use only slower versions of that processor. For example, a Slot 1 Pentium II/III motherboard may be rated to accept processors no faster than 450 MHz. Installing a 550 MHz Slot 1 Pentium III may damage the processor or motherboard because the faster processor draws more current than the VRM on the motherboard is designed to supply.
> - Installing a processor that requires low voltage in a motherboard that can supply only higher voltage. This problem arises only with Socket 7 and earlier motherboards. Slot 1 and later motherboards and processors automatically negotiate the proper voltage. If the motherboard cannot supply the voltage required by the processor, it simply does not power the processor at all. But if you are installing a late-model Socket 7 processor in an older motherboard, be very certain that that motherboard can supply the proper lower voltages required by the new processor (and that it is configured to do so). Otherwise, your new processor may literally go up in smoke the first time you apply power.

The exact sequence of steps required to install a processor depends on its packaging (slotted versus socketed) and whether it comes with a heatsink and fan installed. Regardless of processor type, always begin by laying the motherboard flat on a firm surface, padding it with the antistatic foam or bag supplied with it. Inserting the CPU (and memory) may require substantial force, so it's important to ensure that the motherboard is fully supported to avoid cracking it.

Before you install any processor, obtain and read the installation documentation for both the processor and the motherboard. Spending a few minutes doing that may well save you hours of frustration.

Installing a Socketed Processor

All modern mainstream processors are socketed rather than slotted. These include the Intel Pentium III and Celeron (Socket 370), the Intel Pentium 4 (Socket 423 and Socket 478), and the AMD Athlon and Duron (Socket A). Fifth-generation processors such as the Intel Pentium and AMD K5 use Socket 5 or Socket 7, and hybrid fifth/sixth-generation processors such as the AMD K6 series and Cyrix 686 series use Socket 7.

Installing any socketed processor is a straightforward operation if you do things by the numbers. The most important thing to remember is that processors are particularly sensitive to static shock. Take great care to observe antistatic procedures while you are handling the processor. It's a good rule of thumb to always keep one hand in contact with the PC power supply while you handle the processor.

All recent socketed motherboards have a *Zero Insertion Force* (*ZIF*) socket. As its name implies, the ZIF socket allows a chip with hundreds of pins to be seated easily. Older friction-fit sockets made it nearly impossible to seat a complex chip with hundreds of pins properly. If you encounter a motherboard without a ZIF socket, that in itself is a good reason to replace the motherboard before installing the new processor.

Regardless of the type of socketed processor you are installing, take the following preliminary steps:

1. If you are installing a new processor in an older system, before you begin work check to see if an updated BIOS is available for the system. The new processor may require a BIOS update to function at full capacity, or indeed to function at all. If a new BIOS is available, download it and update your PC as described in Chapter 3.
2. Move the PC or motherboard to a well-lit work area, preferably one with all-around access. Collect all of the tools, software, manuals, and upgrade components you need. Read through the processor documentation before proceeding.
3. To install a processor in a new motherboard, ground yourself, remove the motherboard from its packaging, and place it flat on its accompanying anti-static bag. If you are installing a new processor in an existing PC, you can probably do so without removing the motherboard, although you may have to reroute or temporarily disconnect cables to gain unobstructed access to the socket.

If a heatsink and/or fan are not already installed on the processor, check the instructions or examine the components to determine whether the cooling devices need to be installed before or after you install the processor in the socket. Some cooling devices are easy to install regardless of whether the processor is already in its socket. Most are designed to be installed with the processor already seated in its socket, but a few are easier to install on a loose processor. If your cooling device appears to be easy to install either way, install it after the processor is in the socket. That makes it much easier to get the processor aligned and seated correctly. When you install the cooling device, don't forget to apply thermal compound if the documentation recommends it.

Installing Socket 5 and Socket 7 processors

Socket 5 and Socket 7 motherboards must be configured properly to support the particular processor you are installing. If you are installing a Slot 1 or later processor, skip to the following section. If you are installing a Socket 5 or Socket 7 processor, take the steps described in the preceding section, and then continue as follows:

1. Use the processor and motherboard documentation to verify that the processor and motherboard are compatible, and to determine the proper settings for bus speed, CPU multiplier, core voltage, and I/O voltage. Use the motherboard manual or manufacturer's web site to locate the configuration jumpers and to determine the jumper settings that match those required by the new processor. On some systems, settings are made by a combination of jumper settings and entries in BIOS setup. There are four settings you may have to make, all of which may not be present on a given motherboard:

   Bus speed

   All Socket 5 and Socket 7 motherboards provide settings at least for 60 and 66 MHz. Some motherboards provide higher bus speeds, often including 75 and 83 MHz. These higher bus speeds are used to overclock a 60 or 66 MHz processor—running it faster than its rated speed. Don't use these settings unless you are sure you want to overclock the processor. More recent Socket 7 motherboards, called Super7 motherboards, also provide 95 and 100 MHz bus settings, which are the standard speeds for newer Socket 7 processors. These motherboards may also include various overclocking settings, including 103, 112, and 124 MHz. Again, avoid using overclocking unless you are making an informed decision to do so.

   CPU multiplier

   The product of the bus speed and CPU multiplier determines how fast the processor runs. For example, using a 60 MHz bus speed with a 2.5X multiplier runs the processor at 150 MHz. Note that some processors convert the chosen CPU multiplier internally to a different multiplier. For example, some processors convert a 1.5X CPU multiplier motherboard setting to an internal 4.0X multiplier. Note also that some CPUs are named with a "performance rating" rather than their actual speed. For example, the WinChip2-300 actually runs at 250 MHz (100 MHz x 2.5), but uses the "300" name to indicate its supposed performance relative to other processors. When setting the bus speed and CPU multiplier, it is important to choose settings that run the processor at its actual rated speed rather than the labeled performance equivalent.

   You can sometimes choose between two combinations of bus speed and CPU multiplier that have the same product. In this case, choose the

Processors

combination of the higher bus speed and lower multiplier, so long as the higher bus speed is supported. For example, when installing a 300 MHz processor, you can choose 66MHz/4.5X or 100MHz/3.0X. Either setting runs the processor at 300 MHz, but the latter setting provides marginally faster performance by allowing data to be communicated faster between the CPU and the external L2 cache memory.

Voltage
: Different processors require different voltages. Some processors operate on a single voltage, and others (called *split rail processors*) require different values for *Core Voltage* and *I/O Voltage*. Old motherboards may support only one fixed voltage, and so may not be usable with recent low-voltage or dual-voltage CPUs. Pay close attention to voltage because installing a low-voltage CPU in a high-voltage motherboard may destroy the processor. Adapters are available to allow installing newer low-voltage processors in older motherboards, but in that situation it is better in every respect simply to replace the motherboard.

Asynchronous PCI
: Systems with a 60 or 66 MHz FSB run the PCI bus at half speed—30 MHz and 33 MHz, respectively. Systems with a 100 MHz FSB run the PCI bus at one third speed—33 MHz. This process of using these fixed divisors is called synchronous PCI. But PCI devices are unreliable much above 33 MHz, and overclocking the system by using a 75, 83, or 95 MHz FSB would cause the PCI bus to run at 37.5 MHz (marginal), 41.5 MHz (unusable), or 47.5 MHz (ridiculous). So many motherboards designed to support overclocking include a jumper that allows setting the PCI bus to 33 MHz regardless of the FSB speed.

2. Once you have set and verified all jumpers, lift the ZIF lever, which is located on one side of the socket, as far as it will go. If there is a processor in the socket, grasp it firmly and lift it free. It should come away without resistance.
3. Locate Pin 1 on the new processor. Pin 1 is usually indicated by a dot or beveled edge on one corner of the processor, or by a missing pin on that corner. Locate Pin 1 on the ZIF socket, which is usually indicated by a dot or beveled edge, and sometimes by a numeral 1 silk-screened onto the motherboard itself. Orient Pin 1 on the new processor to Pin 1 on the socket and then gently press the processor into the socket, as shown in Figure 4-14. The processor should seat fully with little or no resistance, dropping into place because of its own weight. If the processor does not seat easily, remove it, verify that the pins align correctly, and try to seat it again. Avoid excessive force when seating the processor. It's easy to bend pins, and straightening them is next to impossible.

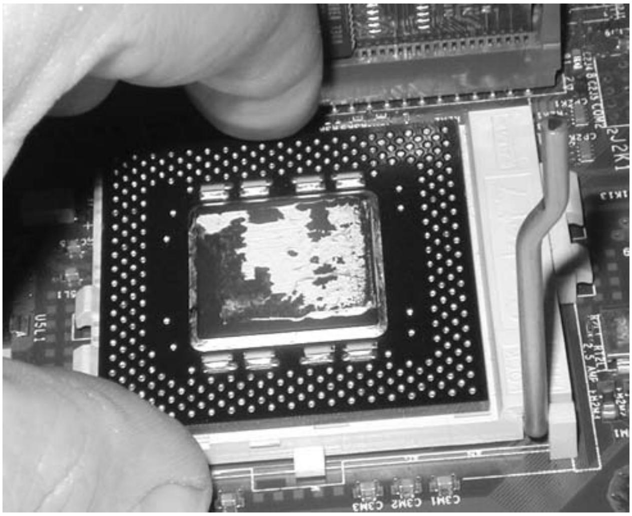


Figure 4-14. Dropping the processor into the socket, where it should seat fully by its own weight (Be sure to align the pins first)

> The processor shown in Figure 4-14 is an Intel Pentium/200, which we were relocating from a system with a failed motherboard to replace a slower processor in another system. The mottling visible on the processor is the remnants of the thermal pad from the old heatsink. Good practice would have been to clean the leftover parts of the thermal pad from the processor and heatsink before proceeding, but we simply added a dollop of thermal goop, which worked fine. We certainly wouldn't do this on any processor we cared about.

4. Once the processor is fully seated, press the ZIF lever down until it is parallel to the edge of the socket, as shown in Figure 4-15. This locks the processor into the socket and makes electrical contact on all pins.
5. If you did not previously install the cooling device, do so now. Don't forget to use thermal compound to improve heat transfer between the processor and the cooling device. Most heatsinks and heatsink/fan units clip directly to the processor or to the socket. Once you have the heatsink aligned properly with the processor (most fit properly in only one orientation) align the clip and press down until it locks into place, as shown in Figure 4-16. If your cooling unit includes a fan, attach the fan power cable to a motherboard fan power header or to an available power supply connector, as appropriate.

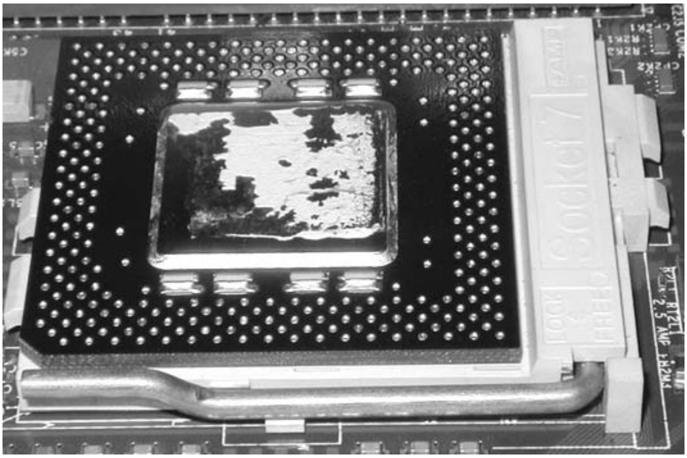


Figure 4-15. The ZIF lever pressed down and locked into place

If no thermal compound or pad was supplied with the heatsink/fan, buy a tube of thermal goop at Radio Shack (it costs $2 or so) and use it. A processor installed without thermal compound may run 20C or more hotter than one with thermal compound, which at best may shorten the life of the processor and at worst may cause frequent system hangs or physical damage to the processor. Thermal compound is frequently omitted, sometimes even on name-brand commercial PCs, so it's worth checking any processor that you didn't install yourself. If you are installing a recent AMD processor, pay close attention to AMD's published requirements for cooling. Using anything other than a brand of phase-change media specifically approved by AMD may void your warranty.

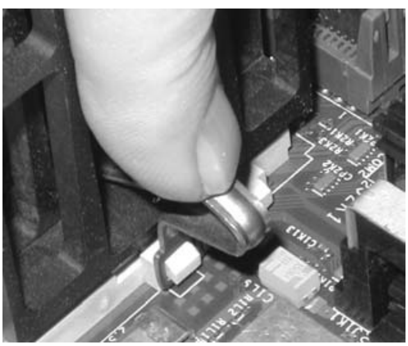


Figure 4-16. Attaching the heatsink securely using the locking clips

If you are upgrading to a faster processor, do not assume that you can use the heatsink/fan unit from the original processor. Faster processors may generate more heat, and may require a more capable HSF unit. Running a newer, hotter processor with the old HSF may at best result in sporadic lockups and at worst in damaging the processor. We said this earlier in the chapter, but it bears repeating.

6. Install the motherboard, if necessary, connect or reroute any cables you moved, do a quick visual once-over, reconnect the monitor, keyboard, and mouse, and then apply power to the system. The system should begin a normal boot sequence. If nothing (or something strange) happens, immediately turn the power off and reverify all connections and settings.
7. Once the system boots normally, enter CMOS Setup and make whatever changes, if any, the processor documentation recommends. Once the system is working normally, turn off the power, reinstall the chassis cover, return the PC to its working location, reconnect all cables, and restart the system.

Installing modern socketed processors

Installing recent socketed processors—the Intel Pentium III/4/Celeron or the AMD Athlon/Duron—requires essentially the same steps described in the preceding section, except that recent processors do not require the motherboard be configured manually.

Most Socket 370, Socket 423, Socket 478, and Socket A motherboards are self-configuring. They detect the type and speed of processor installed and properly configure FSB speed, CPU multiplier, voltage, and other settings automatically. However, some motherboards intended for overclockers allow overriding information supplied by the processor—for example, by setting a 66 MHz FSB Celeron to run at 100 MHz FSB. Depending on the motherboard, changing such settings may require setting jumpers or altering the default BIOS settings. All such motherboards we have seen default to "Auto," which uses the settings supplied by the processor.

There are, however, several issues to be aware of when installing a modern socketed processor:

Compatibility

As we explained in some detail earlier in this chapter, compatibility between motherboard and processor is a major issue. That a processor physically fits the motherboard socket is no guarantee that it will work at all, or even that attempting to use it will not damage the processor and/or motherboard.

With Socket A, AMD has done a much better job of maintaining forward- and backward-compatibility than Intel has done with Socket 370. Even so, with either AMD or Intel processors, it's important to check that the motherboard supports the exact processor you plan to install.

In particular, make sure that the motherboard is rated for processors at least as fast as the processor you plan to install. If the motherboard documentation mentions only slower processors, don't give up hope. High-quality motherboards are often over engineered, using larger VRMs than necessary to support the processors they were designed for. It's quite possible that the motherboard maker issued updated specifications for your motherboard that include support for faster processors. Check the motherboard manufacturer's web site to make sure.

Also verify that the motherboard supports the FSB speed of the processor. If it doesn't, the processor will still operate, but at a much reduced speed. For example, installing a 133 MHz FSB Pentium III/933 in a motherboard that supports only a 100 MHz FSB causes that processor to run at only 700 MHz. Similarly, installing a 266 MHz FSB Athlon in a motherboard that supports only the 200 MHz FSB means that processor runs at only 75% of its rated speed.

BIOS revision level

The BIOS revision level can determine which processors your motherboard supports. A later BIOS may add support for faster versions of a given processor, and may also add support for an entirely new processor. For example, we have an early Slot 1 board that was designed for the cacheless Slot 1 Celeron and did not support later Slot 1 Celerons, which included embedded L2 cache. A BIOS update for that board added support for cached Celerons, and a subsequent BIOS update added support for the new features and changed caching scheme of the Pentium III. Don't assume that because you just purchased a motherboard that it necessarily has the latest BIOS. Some makers, notably Intel, issue BIOS revisions very frequently, and the motherboard you receive may have been in the pipeline for weeks or even months. Before you install a processor in any motherboard, new or old, the first thing you should do is identify the motherboard precisely, check the manufacturer's web site for the most recent BIOS update, and download that update. Once you have the system up and running, install the updated BIOS before you do anything else.

> A motherboard with an early BIOS revision may create a "can't get there from here" situation. That is, the processor you want to install may refuse to boot without a later BIOS revision than is currently installed on the motherboard. In that case, the best alternative is to install temporarily a processor that the earlier BIOS supports. That's why when we upgrade older systems, we install the latest BIOS version on the old system before we remove the original processor. That's also why we keep a stack of old processors around.

Chipset revision level

Many motherboard manufacturers, including top-notch ones such as Intel, have a nasty habit of slipstreaming revisions. Even two motherboards with identical model numbers may be significantly different. In some cases, that difference is as trivial as different BIOS versions, which is easily fixed. Other times, though, there are very real hardware differences between the boards, and those differences may determine which processors a particular board supports. For example, Intel has produced the popular D815EEA2 "Easton 2" motherboard in two distinct forms. Both versions use the 815E chipset, but the version with an early chipset revision level does not support Tualatin-core Pentium III and Celeron processors. If you have the earlier version, you're out of luck. The newer processors simply won't run in it.

If you're buying a new motherboard, check the manufacturer's web site to determine the current rev level and ask the vendor whether the motherboard he wants to sell you is the latest rev level. If not, buy your motherboard elsewhere. If you're using an older motherboard, check the manufacturer's web site to determine what variants exist and what implication those variants have for processor support.

Heatsink compatibility

Socket 370 processors are a particular problem in this respect. There are three different physical forms of Socket 370 processors you are likely to encounter. Early Socket 370 processors use PPGA packaging. These processors have a flat top, with the processor chip itself on the bottom (pin) side of the package. Pentium III and Celeron FC-PGA processors also have a flat top, but with the processor chip protruding above the surface of the processor, on the side opposite the pins, where it comes into direct contact with the heatsink. The most recent Pentium III and Celeron processors use FC-PGA2 packaging, which is similar to FC-PGA but includes a flat metal integrated heat spreader that shrouds the processor chip itself.

Each of these styles requires a physically different heatsink. Using an incorrect heatsink may damage the processor, either physically or by allowing it to overheat. For example, clamping a PPGA heatsink (which has a flat contact surface) onto an FC-PGA processor (which has a raised processor chip) may literally crush the processor. Conversely, installing an FC-PGA heatsink on a PPGA processor may allow the processor to overheat because a portion of it is not in contact with the heatsink.

Heatsink rating is another issue. Faster processors generate more heat, and require larger or more efficient heatsinks. Don't assume that just because a heatsink is designed to be used with a particular type of processor it is usable with that processor running at any arbitrary speed. For example, a particular heatsink may be designed to cool an AMD Duron running at 850 MHz or less. Using that heatsink on a 1.2 GHz Duron will likely allow the processor to overheat and perhaps damage itself.

> Don't assume that all heatsink/fan units will necessarily fit your motherboard and case. Some heatsink/fan units are physically quite large and may not fit. In particular, the portion of the heatsink that overhangs the processor may come into contact with capacitors and other components that protrude above the motherboard. It's not uncommon to find that clamping the heatsink/fan unit into place crushes components that immediately surround the processor socket, so be very careful. Some case/motherboard combinations are also incompatible with some heatsink/fan units because the heatsink/fan is so tall that it cannot be installed because the power supply or portions of the chassis block the space needed by the heatsink fan. If in doubt, measure the available clearances before you order a heatsink/fan unit, and make sure you can return a unit that is incompatible with your motherboard and/or case.

Whichever processor you install, make absolutely certain that the heatsink you plan to use both fits that processor properly and is rated for the processor speed. If you buy a retail-boxed processor, it will come with a heatsink/fan unit appropriate for the processor. If you buy an OEM processor or are reinstalling a processor pulled from another system, make sure the heatsink you use is rated for that particular processor.

Power supply compatibility

Most people don't think about the power supply when they're building or upgrading a system, but the power supply can be a critical issue. Many systems, particularly mass-market systems and consumer-grade systems from major OEMs such as Gateway and Dell, have power supplies that are marginal at best, both in terms of quality and output rating. For example, we have a full-tower Gateway system that arrived with a 150W power supply, and that's *after* we paid for an upgraded power supply. How small must the standard power supply have been?

Modern fast processors have high current draws, and you cannot safely assume that the existing power supply has enough reserve capacity to power them adequately. If you're building a system or upgrading the processor speed significantly in an existing system, make sure that your power supply is up to the job. Otherwise, you may find that the system will not even boot. If the power supply is barely adequate, you may find that the system crashes frequently. We often hear from people who've upgraded their systems with first-rate motherboards and processors, only to find that the new system crashes at the drop of a hat. When that happens, it usually turns out that they've used generic memory or that they just assumed the original power supply would be good enough. Often, it wasn't.

Installing a Slotted Processor

Although mainstream slotted processors are now obsolescent, they remain in limited distribution. A faster slotted processor may be a worthwhile upgrade for an older system. Installing a faster slotted processor can greatly improve system performance and extend the useful life of an otherwise obsolescent system.

For example, until late 2001 our Internet gateway system was an older Celeron. We'd been having some problems with it locking up, which we suspected were caused by the commodity memory installed in it or by the undersized power supply. One day, after three lockups in as many hours, Robert (who is a procrastinator) finally decided to do something about it. We tore down that system and replaced the power supply with an Antec unit and the 64 MB of generic memory with a 128 MB Crucial stick.

While we had the case open for a cleaning and general upgrading, we noticed that the system still had its original Celeron/333 installed, so we decided to replace it with a Pentium II/450 that we'd pulled from another system. The faster clock speed and larger L2 cache of the Pentium II yield performance nearly twice that of the original processor, which takes that system from marginal to more than sufficient for the gateway and mail server tasks to which it is devoted. For a cost of less than $100 (even if we'd had to buy the processor), we now have a reliable Internet gateway system that we expect to continue using for several years to come.

Installing a slotted processor is in some ways easier than installing a socketed processor and in some ways harder. Intel manufactures processors for two similar but incompatible slots. The 242-pin connector, formerly called Slot 1, accepts slotted Celeron, Pentium II, and Pentium III processors. The 330-pin connector, formerly called Slot 2, accepts Pentium II/III Xeon-class processors. These various processors come in different physical packaging (SEC, SEC2, SEPP, etc.), each of which uses a different *retention mechanism*. For example, an SEC Pentium II and an SEPP Celeron both fit the same Slot 1, but use different and incompatible retention mechanisms.

To further complicate matters, Intel ships the same processor in different variants. For example, the retail-boxed version of the Pentium II processor comes with an attached fan, while the OEM version of that processor does not. If you purchase an OEM processor with an attached fan, that package may or may not fit the standard retention mechanism (although it usually does fit). So, the first rule is to make sure that the retention mechanism accepts the processor. If you purchase a cooling device that does not fit the standard retention mechanism, it should be supplied with a mechanism that fits it. Thankfully, all retention mechanisms mount to the standard set of holes in Slot 1 motherboards. Fortunately, AMD Slot A processors are a much simpler matter. All of them use the same physical mounting mechanism, and all Slot A motherboards can accept any Slot A processor. To install a Slot 1 Intel Celeron/Pentium II/Pentium III or a Slot A AMD Athlon processor, take the following steps:

1. When installing a new processor in an older system, determine if a BIOS update is available because the processor may require a later BIOS to support its new features. For example, the Intel SE440BX2-V motherboard accepts various Slot 1 processors, including some Pentium IIIs. But you must upgrade the BIOS to take advantage of the new Pentium III SIMD instructions. Installing the Pentium III without upgrading the BIOS simply makes the

Pentium III run like a faster Pentium II. If a new BIOS is available, download it and update your PC as described in the preceding chapter.

2. Move the PC or motherboard to a well-lit work area, preferably one with all-around access. Collect all of the tools, software, manuals, and upgrade components you need. Read through the processor documentation before proceeding.
3. To install a processor in a new motherboard, ground yourself, remove the motherboard from its packaging, and place it flat on its accompanying anti-static bag. If you are installing a new processor in an existing PC that uses a compatible retention mechanism, you can probably do so without removing the motherboard, although you may have to reroute or disconnect cables to gain unobstructed access to the slot. If the retention mechanism needs to be replaced—e.g., when upgrading a Celeron system to a Pentium III—you may or may not have to remove the system board to replace the retention mechanism.
4. If it is not already installed, install the retention mechanism by following the instructions supplied with it or with the motherboard. Standard retention mechanisms are notched at one end to match the notch in the Slot 1 connector on the motherboard. Align the retention mechanism and seat the four posts into the matching holes on the motherboard. Press down firmly until the retention mechanism seats. Each post has a sliding internal pin topped by a flat, circular piece of white plastic. Forcing that pin down into the post expands the bottom of the post on the far side of the motherboard, securing the post to the motherboard. Press down each of the pins until it snaps into place. Some newer Slot 1 motherboards come with the retention mechanism already installed, but with the vertical supports folded flat. If your motherboard is like this, lift the vertical supports until they snap into place.
5. If the cooling device is not already installed on the processor, install it now. Some processor packages also contain a supplementary support mechanism designed to secure the processor against the additional weight and vibration of the cooling fan. If your package contains such a supplemental support, install it on the processor according to the instructions provided with it.
6. Refer to the processor documentation to determine the proper settings for bus speed and CPU multiplier. Refer to the motherboard manual or manufacturer web site to locate configuration jumpers and to determine the jumper settings that match those required by the new processor. Some boards have separate jumpers for FSB speed and CPU multiplier, others have jumpers for CPU speed only (which implicitly sets both FSB speed and CPU multiplier), and still others use “jumperless setup” which sets FSB and CPU multiplier options in CMOS Setup. Slot 1 processors do not require voltages to be set manually. All current Slot 1 processors use 3.3 volts for external I/O. Klamath-based processors use 2.8 volts internally, and Deschutes-based processors use 2.0 volts. Voltage setting is handled completely automatically via the Voltage ID (VID) pins on the processor itself.
7. Once you have made necessary jumper changes, if any, install the processor, first removing the existing processor if necessary. Note that the card-edge

connector on the processor has a key notch, as does the slot. Slide the processor into the support bracket, making sure that the key is oriented properly, as shown in Figure 4-17.

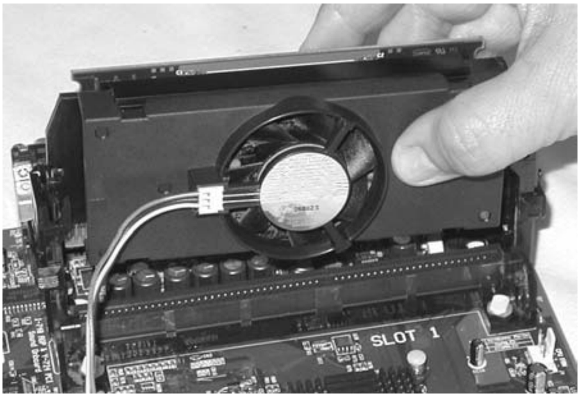


Figure 4-17. Guiding the processor into place, while aligning the keying tab in the slot with that in the processor's card-edge connector

8. Using both thumbs, press down firmly on the processor until it seats fully, as shown in Figure 4-18. This may require applying significant pressure, but you should feel and hear the processor seat. Most support brackets have locking tabs at the top that will snap into place to secure the processor once it is fully seated.

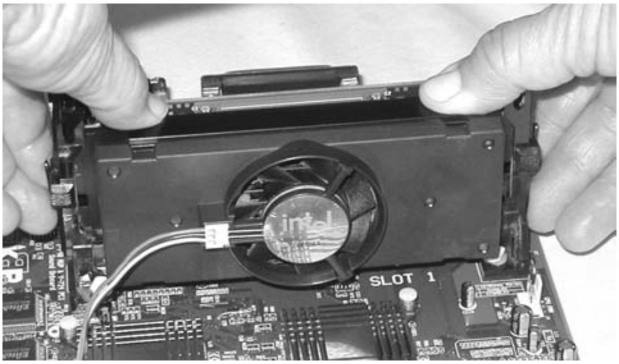


Figure 4-18. Using both thumbs to press firmly until the processor seats fully

9. If the fan power lead is designed to connect to a motherboard power header, connect it now. If the fan power lead instead is designed to connect to a power supply power connector, you'll make that connection after the motherboard is installed in the case.

Completing the Installation

Once you have physically installed the processor and memory (as described in Chapter 5), installed the motherboard in the case, and connected all cables, you're ready to test the system. Verify that everything is connected properly and that you haven't left any tools in the patient. Connect a monitor, keyboard, and mouse to the system. With the case cover still off, apply power. Everything should spin up properly, including the processor fan if one is installed. If it doesn't, immediately turn the power off and recheck your work.

Once the processor is functioning, restart the system, enter BIOS Setup, and set the processor speed if necessary. Setting processor speed is unnecessary with modern processors. In fact, it's usually impossible to do so because the processor reports its speed to the motherboard, which automatically configures itself for that speed. Older processors may or may not require setting processor speed manually, depending on the particular processor and motherboard. Note that with some motherboards, you must move a jumper from "Normal" to "Configure" mode before you can change some settings, including processor speed. Once you have configured BIOS settings appropriately, save the changes and turn off the system.

Our Picks

Although processor makers probably hate us for saying so, the processor actually plays a relatively minor role in overall system performance. The difference in absolute processor performance between a $50 processor and a $500 processor may be a factor of two or less. Nor does buying a $500 processor make your system run twice as fast because processor speed is only one element of system performance. Before you plunk down $500 for a processor, consider instead spending that extra money on more memory, a faster video card, a SCSI hard drive, or all of those.

Inexpensive system ($750 or less)

AMD Athlon XP. In this price range, spend $75 or so on the processor. We recommend choosing the least-expensive Athlon XP you can find in retail-boxed form. Low-end Athlon processors provide incredible bang for the-buck. Your system won't be quite as fast as one that uses a midrange or faster Athlon XP or Pentium 4, but it won't be all that much slower, either.

> Be very careful when buying inexpensive processors. A price much lower than $75 may mean that you're being quoted on discontinued inventory. That's fine if you know what you're getting and are willing to accept the older processor. A lowball price may also mean you're being quoted on an OEM version, which does not include the heatsink and fan, and may have only a short guarantee rather than the three-year guarantee on the retail-boxed model. By the time you buy a separate heatsink and fan, you'll probably spend more than you would have by buying the retail-boxed model in the first place. Caveat emptor definitely applies when you buy any processor. Ask specifically what you're getting or you may get a nasty surprise.

Mainstream system ($750 to $1,200)

Intel Pentium 4 or AMD Athlon XP. In this price range, you have a bit more room to play, and it makes sense to allocate some of that extra money to a faster processor. At the lower end of this price range, choose the fastest retail-boxed AMD Athlon XP you can find for $100 or so (the Intel Pentium 4 doesn't compete in the $100 price range). At the higher end, choose the fastest retail-boxed Intel Northwood-core Pentium 4 or AMD Athlon XP you can find for $150.

Performance system ($1,200+)

Intel Pentium 4 or AMD Athlon XP. This budget level provides considerably more options. For a system in this price range, choose the fastest retail-boxed AMD Athlon XP or Intel Pentium 4 (533 or 800 MHz FSB) you can find in the sub-$200 range.

Dual-processor system

AMD Athlon MP. If you run Windows 2000/XP, Linux, or another SMP-capable operating system, we recommend using a dual-processor system. In our experience, responsiveness in a multitasking environment is better with two midrange processors than with one fast processor. If you choose components carefully, you can build a dual-processor Athlon system for only $250 or so more than the cost of a comparable mainstream system. Your system won't run any one task as fast as it would with a faster single processor, but it won't bog down when you're running many tasks, as the fast single-processor system will.

Processor cooling solutions

TaiSol or DynaTron. A retail-boxed processor includes a heatsink/fan unit that is perfectly adequate for routine use. If you buy an OEM processor, which does not include a heatsink/fan unit, or if you overclock or otherwise push your system beyond its design limits, you'll need a high-quality third-party heatsink/fan unit. Such units vary widely in price, cooling efficiency, and noise level, but the best units overall in our opinion are those manufactured by TaiSol and DynaTron. We have used various TaiSol and DynaTron heatsink/fan models on various processors—including Celerons, Pentium IIIs, Pentium 4s, Athlons, and Durons running at various speeds—and have found them to be effective, quiet, and reasonably priced.

We constantly test and review processors. For information about which specific processors we recommend by brand and model, visit *http://www.hardwareguys.com/picks/processors.html*. We also maintain a set of system guides that detail our currently recommended system configurations for various purposes and in various price ranges. You can view the latest system guides at *http://www.hardwareguys.com/guides/guides.html*.