The **Intel 8087**, announced in 1980, was the first x87 floating-point coprocessor for the 8086 line of microprocessors.[4][5][6]

The purpose of the 8087 was to speed up computations for floating-point arithmetic, such as addition, subtraction, multiplication, division, and square root. It also computed transcendental functions such as exponential, logarithmic or trigonometric calculations, and besides floating-point it could also operate on large binary and decimal integers. The performance enhancements were from approximately 20% to over 500%, depending on the specific application.

Intel Begins with The 4004

The first microprocessor sold by Intel was the four-bit 4004 in 1971. It was designed to work in conjunction with three other microchips, the 4001 ROM, 4002 RAM, and the 4003 Shift Register. Whereas the 4004 itself performed calculations, those other components were critical to making the processor function. The 4004 was mostly used inside of calculators and similar devices, and it was not meant for use inside of computers. Its max clock speed was 740 kHz.

The 4004 was followed by a similar processor known as the 4040, which was essentially an improved variation of the 4004 with an extended instruction set and higher performance.

8008 And 8080

The 4004 made a name for Intel in the microprocessor business, and to capitalize on the situation, Intel introduced a new line of eight-bit processors. The 8008 came first in 1972, followed by the 8080 in 1974 and the 8085 in 1975. Although the 8008 was the first eight-bit processor produced by Intel, it is not as notable as its predecessor or its successor, the 8080. It was faster than the 4004 thanks to its ability to process data in eight-bit chunks, but it was clocked rather conservatively between 200 and 800 kHz, and the 8008's performance simply didn't attract many system developers. The 8008 used 10-micrometer transistor technology.

Intel's 8080 was far more successful. It expanded on the design of the 8008 by adding new instructions and transitioning to six-micrometer transistors. This allowed Intel to more than double the clock rates, and the highest-performance 8080 chips in 1974 came running at 2 MHz. The 8080 was used in countless devices, which lead to several software developers, such as the recently formed Microsoft, to focus on software for Intel's processors.

Eventually when the 8086 was released, it was made source compatible with the 8080 to maintain backwards compatibility with this software. As a result, the 8080s and key hardware elements have been present inside of all x86-based processor ever produced, and 8080 software can technically still run on any x86 processor.

The 8085 was essentially a less expensive and higher-clocked variant of the 8080, which was highly successful as well though less influential.

8086: The Beginning Of x86

Intel's first 16-bit processor was the 8086, which helped to boost performance considerably compared to earlier designs. Not only was it clocked higher than the budget-oriented 8088, but it also employed a 16-bit external data bus and a longer six-byte prefetch queue. It was also able to run 16-bit tasks (though most software at this time was designed for eight-bit processors). The address bus was extended to 20-bit, which enabled the 8086 to access up to 1MB of memory and therefore increase performance.

The 8086 also became the first x86 processor, and it used the first revision of the x86 ISA, which nearly all of the processors created by AMD or Intel since the introduction of the 8086 have been based on.

Intel also produced the 8088 around the same time. This processor was based on the 8086, but with half as many data lines and a four-byte prefetch queue. This caused a loss of balance, as the narrower bus cut into instruction fetch rate, forcing Intel's execution unit to idle much of the time. It still had access to up to 1MB of RAM and ran at higher frequencies than previous processors; however, it was quite a bit slower than the 8086.

80186 And 80188

Intel followed up the 8086 with several other processors, all of which used a similar 16-bit architecture. The first was 80186, aimed at embedded applications. To facilitate this, Intel integrated several pieces of hardware typically found on the motherboard into the CPU, including the clock generator, interrupt controller, and timer. As a side effect, certain instructions ran notably faster on 80186 than 8086, even at the same clock rate. But of course, Intel naturally pushed the CPU's frequency up over time to further improve performance.

The budget-oriented 80188 similarly contained several pieces of hardware integrated into the processor. But like the 8088, its data bus was cut in half.

80286: More Memory, More Performance

The 80286 was released the same year as the 80186 and had nearly identical features, but it extended the address bus to 24-bit, which enabled the processor to access up to 16MB of memory.

iAPX 432

The iAPX 432 was an early attempt by Intel to diverge from its x86 portfolio in favor of an entirely different design. Intel expected iAPX 432 to be several times faster than its other offerings. The processor ultimately failed, however, due to some major design flaws. Although x86 processors are relatively complex, the iAPx 432 took CISC to a whole new level of complexity. The hardware design was rather large, which forced Intel to craft it out of two separate dies. The processor was also quite data hungry and failed to perform well without extremely high amounts of bandwidth. The iAPX 432 managed to outperform the 8080 and 8086, but it was quickly outpaced by newer x86 products, and eventually it was abandoned.

i960: Intel's First RISC

Intel created its first RISC processor in 1984. It was not designed as a direct competitor to the company's x86 processors because it was intended as a secure embedded solution. Internally, it was a 32-bit superscalar architecture that used Berkeley RISC design concepts. The first i960 processors were clocked relatively low, with the slowest model running at 10 MHz, but over the years it was improved and transitioned to smaller fabs that enabled it to hit up to 100 MHz. It also supported 4GB of protected memory.

The i960 was widely used inside of military systems as well as in business systems.

80386: x86 Turns 32-bit

Intel's first 32-bit x86 processor was the 80386, released in 1985. One key advantage that this processor had was its 32-bit address bus that allowed it to support up to 4GB of system memory. Although this was far more than anyone was using at the time, RAM limitations often hurt the performance of prior x86 and competing processors. Unlike modern CPUs, at the time the 80386 was released, more RAM almost always translated into a performance increase. Intel also implemented several architectural enhancements that helped push performance up above the 80286, even when both systems used the same amount of RAM. It also supported virtual mode processing, which increased multi-tasking support.

To segment its product line-up with a more budget-friendly offering, Intel also introduced the 80386SX. This processor was almost identical to the 80386; it still employed a 32-bit architecture, but half of its data bus was cut to 16 bits for cost-saving purposes.

i860

In 1989, Intel made another attempt to move away from its x86 processors. It created a new RISC CPU known as the i860. Unlike the earlier i960, this CPU was designed to be a high-performance model to compete in the desktop market, but the design proved problematic. Its most significant flaw was that the processor's performance relied entirely on the compiler to place instructions in the order they would need to be executed when the software was first created. This helped Intel keep the die size and overall complexity of the i860 down, but it was nearly impossible to correctly list every instruction from beginning to end when compiling the program. This caused the CPU to constantly stall while it attempted to work around the problem.

80486: Integrating The FPU

Intel's 80486 was another significant step up in terms of performance. The key to its success was tighter integration of components into the CPU. The 80486 was the first x86 CPU to contain L1 cache. Early 80486 models came with 8KB on-die, and were etched on a 1000nm process. But as the design transitioned to 600nm, the L1 cache size doubled to 16KB.

Intel also incorporated the FPU into the CPU, which up to that point had been a separate functional processing unit. By moving these pieces of hardware into the host processor, latency between them dropped sharply. The 80486 also used a faster FSB interface to increase bandwidth, and the core had various other tweaks to push up IPC. These changes increased the 80486's performance significantly, and high-end models were multiple times faster than the older 80386.

The first 80486 processors reached 50 MHz, and later models that used the improved 600nm process went as high as 100 MHz. To target budget-oriented users, Intel also released a version of the 80486 known as the 80486SX, which had the FPU disabled.

P5: The First Pentium

The Pentium emerged in 1993 as the first Intel x86 processor that didn't follow the 80x86 number system. Internally, the Pentium used the P5 architecture, which was Intel's first x86 superscalar design. Although the Pentium was generally faster than the 80486 in every way, its most prominent feature was a substantially improved FPU. The original Pentium's FPU was more than ten times faster than the 80486's aging unit. This became an even more significant feature in later years when Intel released the Pentium MMX. This processor was architecturally the same as the original Pentium, but featured support for Intel's new MMX SIMD instruction set that could drastically boost performance.

Intel also increased the L1 cache size on its Pentium processors relative to the 80486. Initial Pentiums contained 16KB, while the Pentium MMX moved up to 32KB. Naturally, these processsors also ran at higher clock rates. The first Pentium processors used 800nm transistors and could hit just 60 MHz, but subsequent revisions transitioned to Intel's 250nm process and pushed frequencies up to 300 MHz.

P6: The Pentium Pro

Intel planned to quickly follow the Pentium up with the Pentium Pro based on its P6 architecture, but ran into technical difficulties. The Pentium Pro was considerably faster than the Pentium in 32-bit operations thanks to its Out-of-Order (OoO) design. It featured a significantly redesigned internal architecture that decoded instructions into micro-ops, which were then executed on general-purpose execution units. It also used a significantly extended 14-stage pipeline owing to the additional decoding hardware.

As the first Pentium Pro processors were targeted at the server market, Intel extended the address bus again to 36-bit and added its PAE technology that allowed it to support up to 64GB of RAM. This was far more than average users needed, but being able to support greater amounts of RAM was key to Intel's server customers.

The processor's cache system was reworked as well. The L1 cache was limited to two segmented 8KB caches, one for instructions and one for data. To make up for the 16KB deficit compared to the Pentium MMX, Intel placed between 256KB and 1MB of L2 cache on a separate chip attached to the CPU package. It connected to the CPU using a back-side-bus (BSB).

Intel initially planned to push the Pentium Pro out to consumers as well, but ultimately limited it as a server product. The Pentium Pro featured several revolutionary features, but it struggled against the Pentium and Pentium MMX in terms of performance. Both of the older Pentium parts were significantly faster at 16-bit operations, and 16-bit software was still heavily used back then. The processor also lacked support for the MMX instruction set, which resulted in the Pentium MMX outperforming the Pentium Pro in MMX-optimized software.

The Pentium Pro may have stood a chance in the consumer market, but it was also fairly expensive to produce due to the separate chip containing L2 cache. The fastest Pentium Pro processor ran at 200 MHz, and it was crafted with transistors ranging between 500 and 350nm.

P6: Pentium II

Intel didn't give up on the P6 architecture, but instead waited until 1997 when it released the Pentium II. The Pentium II managed to overcome nearly all of the negative aspects of the Pentium Pro. Its underlying architecture was similar to the Pentium Pro, and it continued to use a 14-stage pipeline with several enhancements to the core to improve IPC. The L1 grew to 16KB data + 16KB instruction caches.

Intel also moved to more affordable cache chips attached to a larger silicon package to reduce production costs. This was an effective way of making the Pentium II less expensive, but these memory modules were unable to operate at the CPU's full speed. Instead, the L2 cache ran at half-frequency, and on these early processors that was sufficient to increase performance.

Intel also added support for the MMX instruction set. The CPU cores used inside of the Pentium II, code-named "Klamath" and "Deschutes," were also sold as Xeon and Pentium II Overdrive products for servers. The highest-performance models contained 512KB of L2 cache and ran at 450 MHz.

P6: Pentium III And The Race To 1 GHz

Intel planned to follow up the Pentium II with a processor based on its Netburst architecture, but it wasn't quite ready. Instead, Intel pushed the P6 architecture out again as the Pentium III.

The first of these processors, code-named "Katmai," was rather similar to the Pentium II in that it used a slotted cartridge containing lower-quality L2 cache at half of the CPU's speed. The underlying architecture incorporated other significant changes, as several parts of the 14-stage pipeline were fused together, shortening it to 10 stages. Thanks to the updated pipeline and an increase in clock speed, the first of the Pentium III processors typically outperformed their Pentium II counterparts by a small margin.

Katmai was produced using 250nm transistors. However, following the move to a 180nm fabrication process, Intel was able to boost the Pentium III's performance significantly. This updated implementation, code-named "Coppermine," moved the L2 cache into the CPU and reduced its capacity by half (down to 256KB). But because it was able to run at the processor's frequency, performance still shot up.

Coppermine was Intel's competitor to AMD's Athlon in the race to break 1 GHz, which it succeeded in doing. Intel attempted to produce a 1.13 GHz model, but it was ultimately recalled after investigation from Dr. Tom Pabst of Tom's Hardware discovered that it was unstable. This left the 1 GHz model the fastest Coppermine-based Pentium III.
The last of Pentium III cores was named "Tualatin." It moved to a 130nm process that facilitated clock rates as high as 1.4GHz. It also increased the L2 cache back to 512KB, which helped improve performance somewhat.

P5 And P6: Celeron And Xeon

Around the release of the Pentium II, Intel also introduced its Celeron and Xeon product lines. These products used the same core as the Pentium II or Pentium III, but with varying amounts of cache. The first Celeron-branded processors based on the Pentium II had no L2 cache at all, which resulted in horrible performance. Later models based on the Pentium III had half of the L2 cache disabled compared to their Pentium III counterparts. This resulted in Celeron processors that used the Coppermine core containing just 128KB of L2 cache; later models based on Tualatin increased this to 256KB.

These half-cache derivatives became known as the Coppermine-128 and the Tualatin-256. Intel sold them at clock speeds comparable to the Pentium III, which allowed them to perform well and made them highly competitive against AMD's Duron processors. Microsoft used one of the Coppermine-128 Celeron processors clocked at 733 MHz as the CPU inside of its Xbox gaming console.

The first Xeon processors were similar, but they contained more L2 cache. The Pentium II-based Xeon processors contained at least 512KB, the same as Pentium II CPUs, whereas higher-end models could have up to 2MB.

MORE: Best CPUs
MORE: Best CPU Cooling
MORE: Intel & AMD Processor Hierarchy
MORE: All CPU Content

Netburst: Introduction

Before discussing Intel's Netburst architecture and the Pentium 4, it is important to examine the idea behind its deep pipeline, which describes the process whereby instructions move through a core. Pipeline stages often perform multiple tasks, but sometimes they're devoted to single functions. By either adding new hardware or splitting one stage into multiple stages, the execution pipeline can be extended. The processor pipeline can also be shrunk by removing hardware or by combining the components in multiple stages down into a single stage.

The length or depth of the pipeline has a direct impact on latency, IPC, clock speed and the architecture's throughput requirements. Longer pipelines typically require higher amounts of bandwidth, but if the pipeline is kept adequately fed with data, then each stage in the pipeline stays busy. Processors that have longer pipelines typically are able to run at higher clock rates as well.

The trade-off is significantly higher latency inside of the processor, as data flowing through it must stop at each stage for a certain number of clock cycles. Processors using a long pipeline tend to have lower IPC as well, which is why they rely on significantly higher frequencies to increase performance. Over the years, processors implementing both philosophies have proven successful. Neither approach is necessarily flawed.

Netburst: Pentium 4 Willamette And Northwood

In 2000, Intel's Netburst architecture was finally ready, and it was pushed into production as the Pentium 4. The combination would carry Intel's top-end CPUs for the next six years. The first implementation was named "Willamette," which carried Netburst and the Pentium 4 through the first two years of its life. This was a troubled time for Intel, however, and the chip struggled to outperform the Pentium III. Netburst enabled significantly higher frequencies, and Willamette managed to hit 2 GHz, but the Pentium III at 1.4 GHz was still faster in some tasks. AMD's Athlon processors enjoyed a healthy performance lead during this period.

The problem with Willamette was that Intel stretched the pipeline out to 20 stages and planned to hit even higher clock rates beyond 2 GHz, but due to power consumption and heat issues, it was unable to reach those goals. The situation improved with Intel's 130nm design known as "Northwood," which scaled up to 3.2 GHz and doubled the L2 cache from 256KB to 512KB. Netburst's power consumption and heat issues persisted. However, Northwood nevertheless performed significantly better and was highly competitive against AMD.

On high-end models, Intel also introduced its Hyper-Threading technology to improve resource utilization in environments that emphasized multitasking. Hyper-Threading wasn't as beneficial on Northwood as it is on present-day Core i7 processors, but it did push performance up by a few percentage points.

Willamette and Northwood were released inside of Celeron- and Xeon-branded CPUs as well. As with the previous generation of Celeron- and Xeon-based products, Intel raised or lowered the L2 cache size in order to distinguish their performance.

P6: Pentium-M

As Netburst was designed as a high-performance architecture that was fairly power hungry, it didn't translate well to mobile systems. Instead, in 2003 Intel created its first architecture designed exclusively for notebooks. The Pentium-M was based on the P6 architecture, but with a longer 12-14 stage pipeline. This was also Intel's first variable-length pipeline, which meant that instructions could be executed after moving through

just 12 stages if the information required for the instruction was already loaded into cache. If not, it had to go through two additional stages to load the data.

The first of these processors was crafted with 130nm transistors and contained a 1MB L2 cache. It managed to hit 1.8 GHz while consuming just 24.5W of power. A later revision known as "Dothan" was released in 2004 and transitioned to 90nm transistors. This enabled Intel to increase the L2 cache to 2MB and, combined with a number of core enhancements, provide a decent IPC throughput improvement. The CPU also scaled up to 2.27 GHz with a slight increase of power to 27W.

The Pentium-M architecture would eventually be used inside of the Stealey A100 mobile CPUs before being replaced by Intel's line of Atom processors.

Netburst: Prescott

Northwood carried the Netburst architecture from 2002 until 2004, after which Intel launched Prescott with numerous enhancements. It used a 90nm fabrication process that enabled Intel to increase the L2 cache to 1MB. Intel also introduced the new LGA 775 interface that featured support for DDR2 memory and a faster quad-pumped FSB than the first Northwood-based CPUs. These changes resulted in Prescott having significantly more bandwidth than Northwood, which was vital to increasing Netburst's performance. Prescott was also Intel's first 64-bit x86 processor, allowing it to access more RAM and operate on 64 bits at a time.

Prescott was supposed to be the crown jewel in Intel's family of Netburst-based processors, but instead it was a fiasco. Intel again extended its execution pipeline, this time to 31 stages. The company hoped to increase clock rates enough to offset the longer pipe, but it was only able to hit 3.8 GHz. Prescott simply ran too hot and consumed too much power. Intel expected the move to 90nm to alleviate this issue, but the increased transistor density made cooling more difficult. As it was not able to hit higher frequencies, Prescott's evolutionary changes hurt overall performance.

Even with all of the enhancements and extra cache, Prescott was, at best, on par with Northwood at any given clock rate. Around the same time, AMD's K8 processors were also moving to smaller transistors that enabled them to hit higher frequencies. For this brief time period, AMD dominated the desktop CPU market.

Netburst: Pentium D

In 2005, the race was on to produce the first consumer-oriented dual-core processor. AMD had already announced its dual-core Athlon 64, but it wasn't available yet. Intel rushed to beat AMD by using a multi-core module (MCM) that contained two Prescott

dies. The company christened its dual-core processor the Pentium D, and the first model was code-named "Smithfield."

The Pentium D launched to criticism, however, as it faced the same issues that plagued Prescott. The heat and power of two Netburst-based dies limited clock rates to 3.2 GHz at most. And because the architecture was bandwidth-limited, Smithfield's IPC suffered as throughput was split between two cores. The implementation wasn't particularly elegant either; AMD's dual-core CPU constructed from one die was considered superior.

Smithfield was followed by Presler, which moved to 65nm transistor technology. It contained two Ceder Mill dies on an MCM. This helped reduce the processor's heat and power consumption, and let Intel raise its clock rate to 3.8 GHz.

There are two key steppings of Presler. The first one had a higher 125W TDP, whereas the later model dropped down to 95W. Thanks to the smaller die size, Intel was able to double the L2 cache as well, so each die had 2MB. A few enthusiast models also featured Hyper-Threading, allowing the CPU to address four threads simultaneously.

All Pentium D processors supported 64-bit software and could take advantage of more than 4GB of RAM.

Core: Core 2 Duo

Intel eventually gave up on its Netburst architecture and instead put its support behind the P6 and Pentium-M design. The company realized that P6 was still viable, and capable of being both efficient and providing excellent performance. It reworked the architecture into its Core design. Like the Pentium-M, it used a 12 to 14 stage pipeline that was significantly shorter than Prescott's 31-stage implementation.

Core proved to be highly scalable, and Intel was able to push it into service on mobile systems with TDPs as low as 5W and high-end servers with 130W ceilings. Intel mostly sold it as "Core 2 Duo" or "Core 2 Quad" products, but Core was also used inside of Core Solo-, Celeron-, Pentium- and Xeon-branded CPUs. The dies used were built using two CPU cores, and quad-core designs used two dual-core dies on an MCM. Single-core versions, meanwhile, had one core disabled. L2 cache size ranged from 512KB up to 12MB.

With the improvements made to the Core architecture, Intel could again compete against AMD. The PC market entered a golden age filled with extremely competitive high-performance processors that are still viable to this day.