Chapter 14 - Examples of CPU's

In this chapter I will briefly describe the important CPU's which have been on the market, starting from the PC's early childhood and up until today.

One could argue that the obsolete and discontinued models no longer have any practical significance. This is true to some extent; but the old processors form part of the "family tree", and there are still legacies from their architectures in our modern CPU's, because the development has been evolutionary. Each new processor extended and built "on top of" an existing architecture.
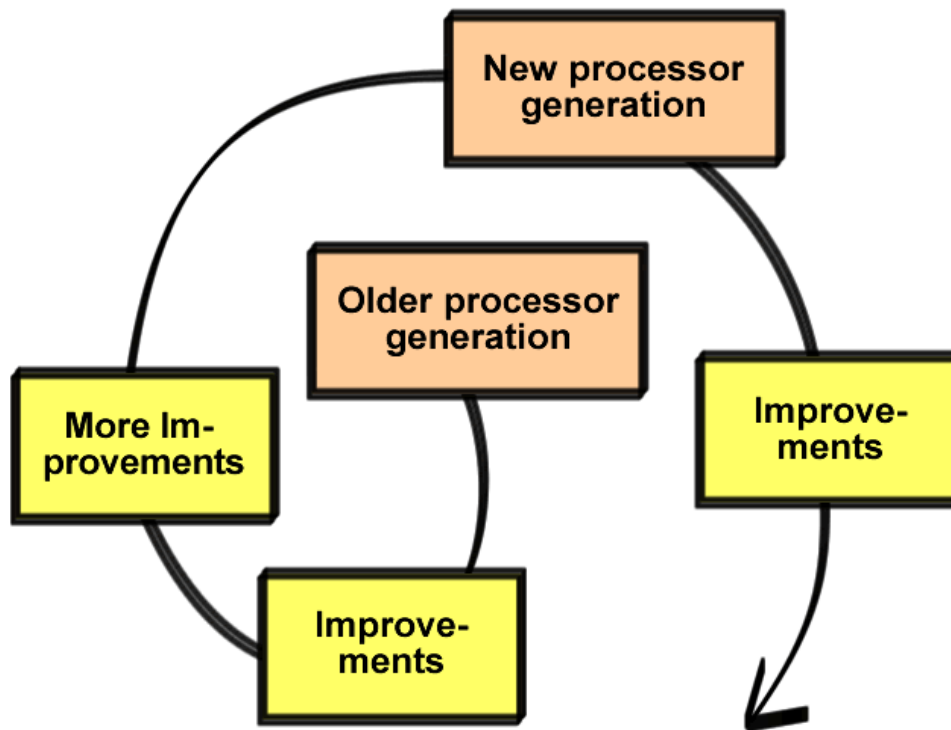

Fig. 98. The evolutionary development spirals ever outwards.

There is therefore value (one way or another) in knowing about the development from one generation of CPU's to the next. If nothing else, it may give us a feeling for what we can expect from the future.

16 bits – the 8086, 8088 and 80286

The first PC's were 16-bit machines. This meant that they could basically only work with text. They were tied to DOS, and could normally only manage one program at a time.

But the original 8086 processor was still "too good" to be used in standard office PC's. The Intel 8088 discount model was therefore introduced, in which the bus between the CPU and RAM was halved in width (to 8 bits), making production of the motherboard much cheaper. 8088 machines typically had 256 KB, 512 KB or 1 MB of RAM. But that was adequate for the programs at the time.

The Intel 80286 (from 1984) was the first step towards faster and more powerful CPU's. The 286 was much more efficient; it simply performed much more work per clock tick than the 8086/8088 did. A new feature was also the *32 bit protected mode* – a new way of working which made the processor much more efficient than under *real mode*, which the 8086/8088 processor forced programs to work in:

- Access to *all system memory* – even beyond the 1MB limit which applied to real mode.

Access to *multitasking*, which means that the operating system can run several programs at the same time.

- The possibility of *virtual memory*, which means that the hard disk can be used to emulate extra RAM, when necessary, via a *swap file*.
- 32 bit access to RAM and 32 bit drivers for I/O devices.

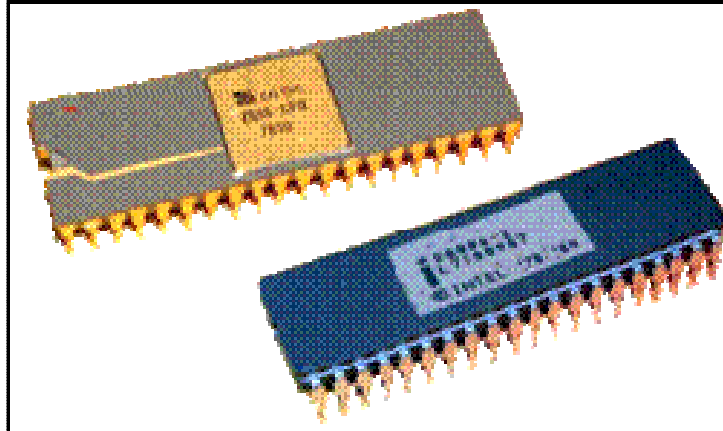Protected mode paved the way for the change from DOS to Windows, which only came in the 1990's.

Fig. 99. Bottom: an Intel 8086, the first 16-bit processor. Top: the incredibly popular 8-bit processor, the Zilog Z80, which the 8086 and its successors out competed.

32 bits – the 80386 and 486

The Intel 80386 was the first 32-bit CPU. The 386 has 32-bit long registers and a 32-bit data bus, both internally and externally. But for a traditional DOS based PC, it didn't bring about any great revolution. A good 286 ran nearly as fast as the first 386's – under DOS anyway, since it doesn't exploit the 32-bit architecture.

The 80386SX became the most popular chip – a discount edition of the 386DX. The SX had a 16-bit external data bus (as opposed to the DX's 32-bit bus), and that made it possible to build cheap PC's.

Fig. 100. Discount prices in October 1990 – but only with a b/w monitor.

The fourth generation

The fourth generation of Intel's CPU's was called the 80486. It featured a better implementation of the x86 instructions – which executed faster, in a more RISC-like manner. The 486 was also the first CPU with built-in L1 cache. The result was that the 486 worked roughly twice as fast as its predecessor – for the same clock frequency.

With the 80486 we gained a built-in FPU. Then Intel did a marketing trick of the type we would be better off without. In order to be able to market a cheap edition of the 486, they hit on the idea of *disabling* the FPU function in some of the chips. These were then sold under the name, 80486SX. It was ridiculous – the processors had a built-in FPU; it had just been switched off in order to be able *to segment the market*.

Fig. 101. Two 486's from two different manufacturers.

But the 486 was a good processor, and it had a long life under DOS, Windows 3.11 and Windows 95. New editions were released with higher clock frequencies, as they hit on the idea of *doubling* the *internal* clock frequency in relation to the external (see the discussion later in the guide). These double-clocked processors were given the name, 80486DX2.

A very popular model in this series had an external clock frequency of 33 MHz (in relation to RAM), while working at 66MHz internally. This principle (*double-clocking*) has been employed in one way or another in all later generations of CPU's. AMD, IBM, Texas Instruments and Cyrix also produced a number of 80486 compatible CPU's.

Pentium

In 1993 came the big change to a new architecture. Intel's Pentium was the first fifth-generation CPU. As with the earlier jumps to the next generation, the first versions weren't especially fast. This was particularly true of the very first Pentium 60 MHz, which ran on 5 volts. They got burning hot – people said you could fry an egg on them. But the Pentium quickly benefited from new process technology, and by using clock doubling, the clock frequencies soon skyrocketed.

Basically, the major innovation was a *superscalar* architecture. This meant that the Pentium could process several instructions at the same time (using several pipelines). At the same time, the RAM bus width was increased from 32 to 64 bits.
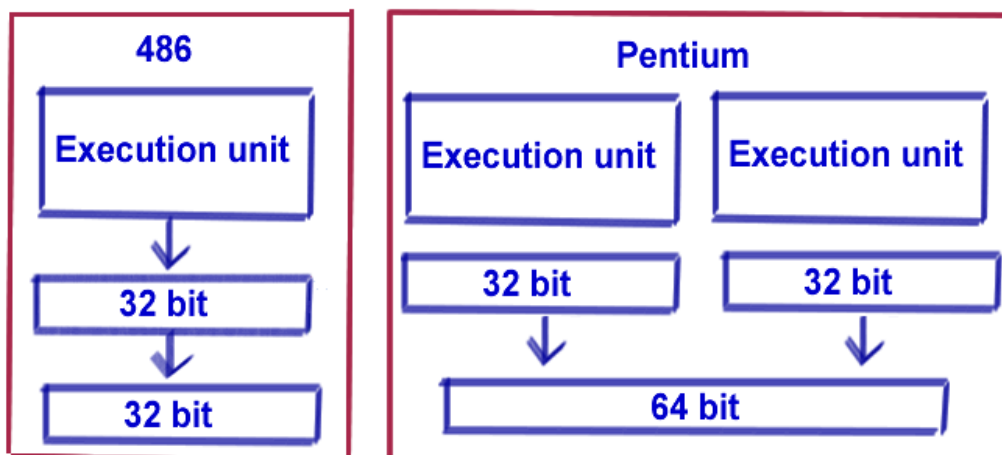

Fig. 102. The Pentium processor could be viewed as two 80486's built into one chip.

Throughout the 1990's, AMD gained attention with its *K5* and *K6* processors, which were basically cheap (and fairly poor) copies of the Pentium. It wasn't until the *K6-2* (which included the very

successful 3DNow! extensions), that AMD showed the signs of independence which have since led to excellent processors like the AthlonXP.
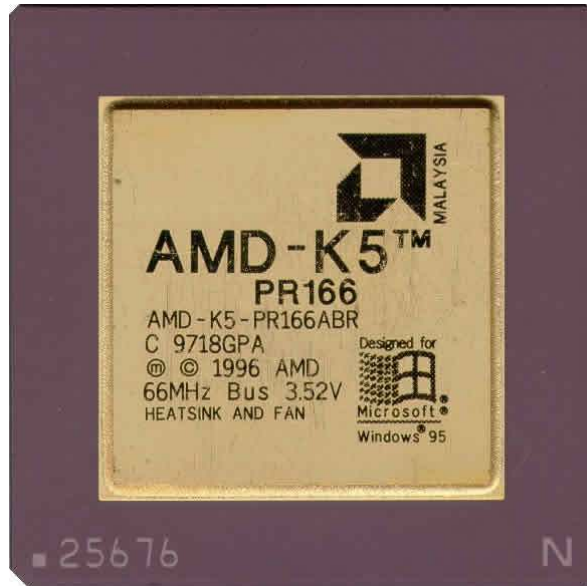

Fig. <![endif]-->103. One of the earlier AMD processors. Today you'd hesitate to trust it to run a coffee machine…

In 1997, the Pentium MMX followed (with the model name P55), introducing the MMX instructions already mentioned. At the same time, the L1 cache was doubled and the clock frequency was raised.

Fig. 104. The Pentium MMX. On the left, the die can be seen in the middle.

Pentium II with new cache

After the Pentium came the Pentium II. But Intel had already launched the Pentium Pro in 1995, which was the first CPU in the 6th generation. The Pentium Pro was primarily used in servers, but its architecture was re-used in the popular Pentium II, Celeron and Pentium III models, during 1997-2001.

The Pentium II initially represented a technological step backwards. The Pentium Pro used an integrated L2 cache. That was very advanced at the time, but Intel chose to place the cache outside the actual Pentium II chip, to make production cheaper.
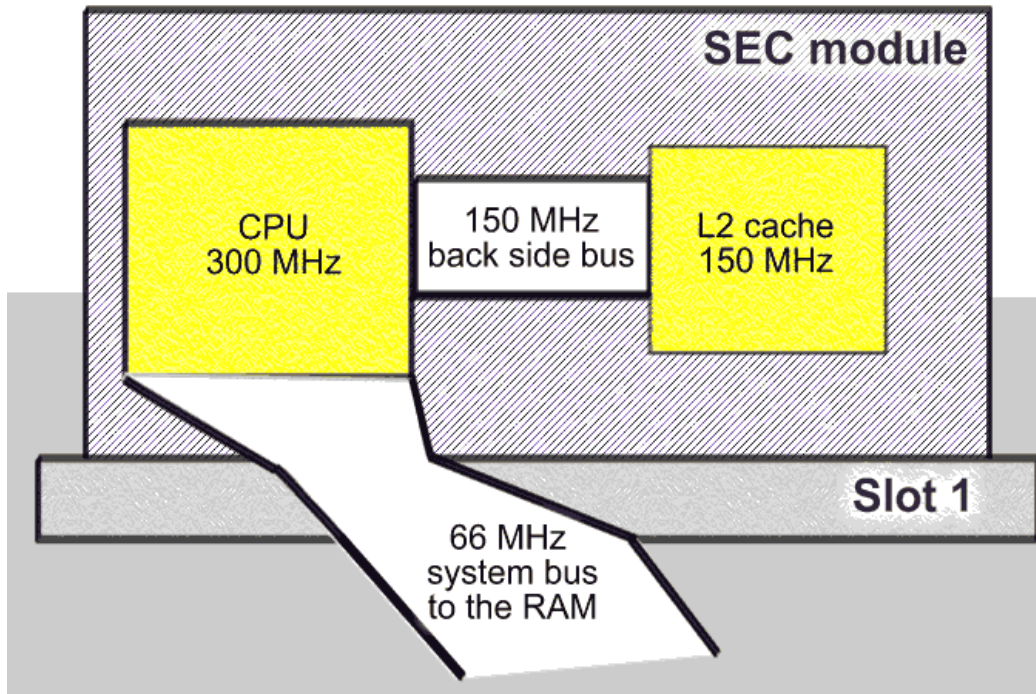
Fig. 105. L2 cache running at half CPU speed in the Pentium II.
The Level 2 cache was placed beside the CPU on a circuit board, an *SEC module* (e.g. see **Error! Reference source not found.**

Chapter 15. Evolution of the Pentium 4

As was mentioned earlier, the older P6 architecture was released back in 1995. Up to 2002, the Pentium III processors were sold alongside the Pentium 4. That means, in practise, that Intel's sixth CPU generation has lasted 7 years.

Similarly, we may expect this seventh generation Pentium 4 to dominate the market for a number of years. The processors may still be called Pentium 4, but it comes in al lot varietes.

A mayor modification comes with the version using 0.65 micron process technology. It will open for higher clock frequencies, but there will also be a number of other improvements.

Hyper-Threading Technology is a very exciting structure, which can be briefly outlined as follows: In order to exploit the powerful pipeline in the Pentium 4, it has been permitted to process *two threads at the same time*. Threads are series of software instructions. Normal processors can only process *one* thread at a time.

In servers, where several processors are installed in the same motherboard (MP systems), several threads can be processed at the same time. However, this requires that the programs be set up to exploit the MP system, as discussed on page 31.

The new thing is that a single Pentium 4 logically can function as if there physically were two processors in the pc. The processor core (with its long pipelines) is simply so powerful that it can, in many cases, act as two processors. It's a bit like one person being able to carry on two independent telephone conversations at the same time.
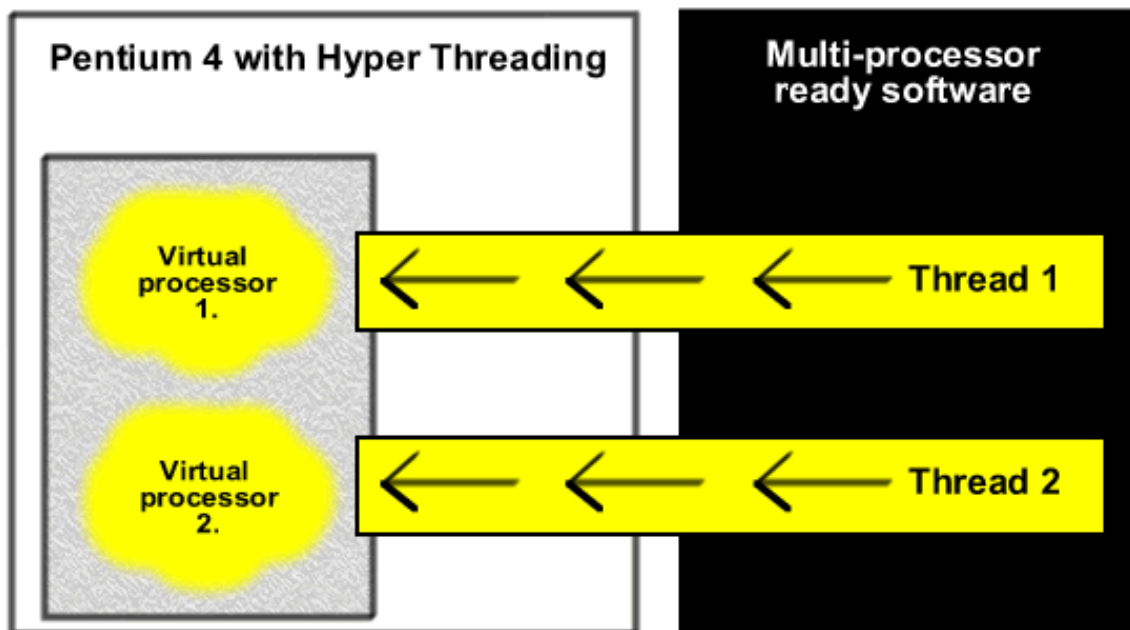
Figur 110. The Pentium 4 is ready for MP functions.

Hyper-Threading works very well in Intel's Prescott-versions of Pentium 4. You gain performance when you operate more than one task at the time. If you have two programs working simultaneously, both putting heavy pressure on the CPU, you will benefit from this technology. But you need a MP-compatible operating system (like Windows XP Professional) to benefit from it.

The next step in this evolution is the production of *dual-core processors.* AMD produces Opteron chips which hold two processors in one chip. Intel is working on dual core versions of the Pentium 4 (with the codename "Smithfield"). These chips will find use in servers and high performance pc's. A dual core Pentium 4 with Hyper-Threading enabled will in fact operate as a virtual *quad-core* processor.
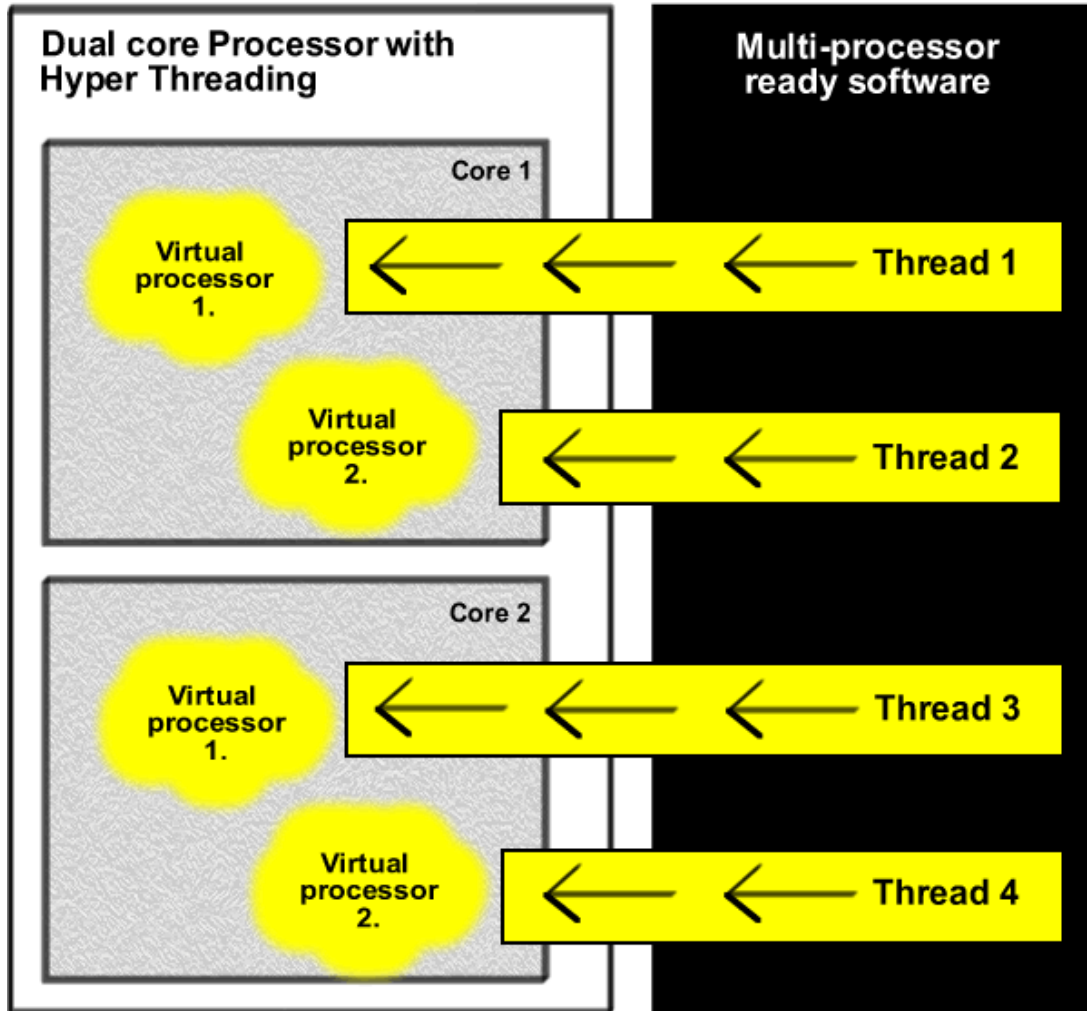
Figur 111. A dual core processor with Hyper Threading operates as virtual quad-processor.

Intel also produces EE-versions of the Pentium 4. EE is for *Extreme Edition*, and these processors are extremely speedy versions carrying 2 MB of L2 cache.

In late 2004 Intel changed the socket design of the Pentium 4. The new processors have no "pins"; they connect directly to the socket using little contacts in the processor surface.

Figur 112. The LGA 775 socket for Pentium 4.

Athlon

The last processor I will discuss is the popular Athlon and Athlon 64 processor series (or *K7* and *K8*).

It was a big effort on the part of the relatively small manufacturer, AMD, when they challenged the giant Intel with a complete new processor design.

The first models were released in 1999, at a time when Intel was the completely dominant supplier of PC processors. AMD set their sights high – they wanted to make a better processor than the Pentium II, and yet cheaper at the same time. There was a fierce battle between AMD and Intel between 1999 and 2001, and one would have to say that AMD was the victor. They certainly took a large part of the market from Intel.

The original 1999 Athlon was very powerfully equipped with pipelines and computing units:

- Three instruction decoders which translated X86 program CISC instructions into the more efficient RISC instructions (ROP's) – 9 of which could be executed at the same time.
- Could handle up to 72 instructions (*ROP out of order*) at the same time (the Pentium III could manage 40, the K6-2 only 24).
- Very strong FPU performance, with three simultaneous instructions.

All in all, the Athlon was in a class above the Pentium II and III in those years. Since Athlon processors were sold at competitive prices, they were incredibly successful. They also launched the *Duron* line of processors, as the counterpart to Intel's Celeron, and were just as successful with it.

Figur 113. Athlon was a huge success for AMD. During 2001-2002, the Athlon XP was in strong competition with the Pentium 4.

Athlon XP versus Pentium 4

The Athlon processor came in various versions. It started as a Slot A module (see Fig. 107 on page 42). It was then moved to Socket A, when the L2 cache was integrated.

In 2001, a new Athlon XP version was released, which included improvements like a new *Hardware Auto Data Prefetch Unit* and a bigger *Translation Look-aside Buffer*. The Athlon XP was much less advanced than the Pentium 4 but quite superior at clock frequencies less than 2000 MHz. A 1667 MHz version of AthlonXP was sold as 2000+. This indicates, that the processor as a minimum performs like a 2000 MHz Pentium 4.

Later we saw Athlons in other versions. The latest was based on a new kernel called "Barton". It was introduced in 2003 with a L2-cachen of 512 KB. AMD tried to sell the 2166 MHz version under the brand 3000+. It did not work. A Pentium 4 running at 3000 MHz had no problems outperforming the Athlon.

Opteron/ Athlon64

AMD's 8th generation CPU was released in 2003. It is based on a completely new core called *Hammer.*

A new series of 64-bits processors is called Athlon 64, Athlon 64 FX and Opteron. These CPU's has a new design in two areas:

• The memory controller is integrated in the CPU. Traditionally this function has been housed in the north bridge, but now it is placed inside the processor.
• AMD introduces a completely new 64-bit set of instructions.

Moving the memory controller into the CPU is a great innovation. It gives a much more efficient communication between CPU and RAM (which has to be ECC DDR SDRAM – 72 bit modules with error correction).)

Every time the CPU has to fetch data from normal RAM, it has to first send a request to the chipset's controller. It has to then wait for the controller to fetch the desired data – and that can take a long time, resulting in wasted clock ticks and reduced CPU efficiency. By building the memory controller directly into the CPU, this waste is reduced. The CPU is given much more direct access to RAM. And that should reduce latency time and increase the effective bandwidth.

The Athlon 64 processors are designed for 64 bits applications. This should be more powerful than the existing 32 bit software. We will probably see plenty of new 64 bit software in the future, since Intel is releasing 64 bit processors compatible with the Athlon 64 series.
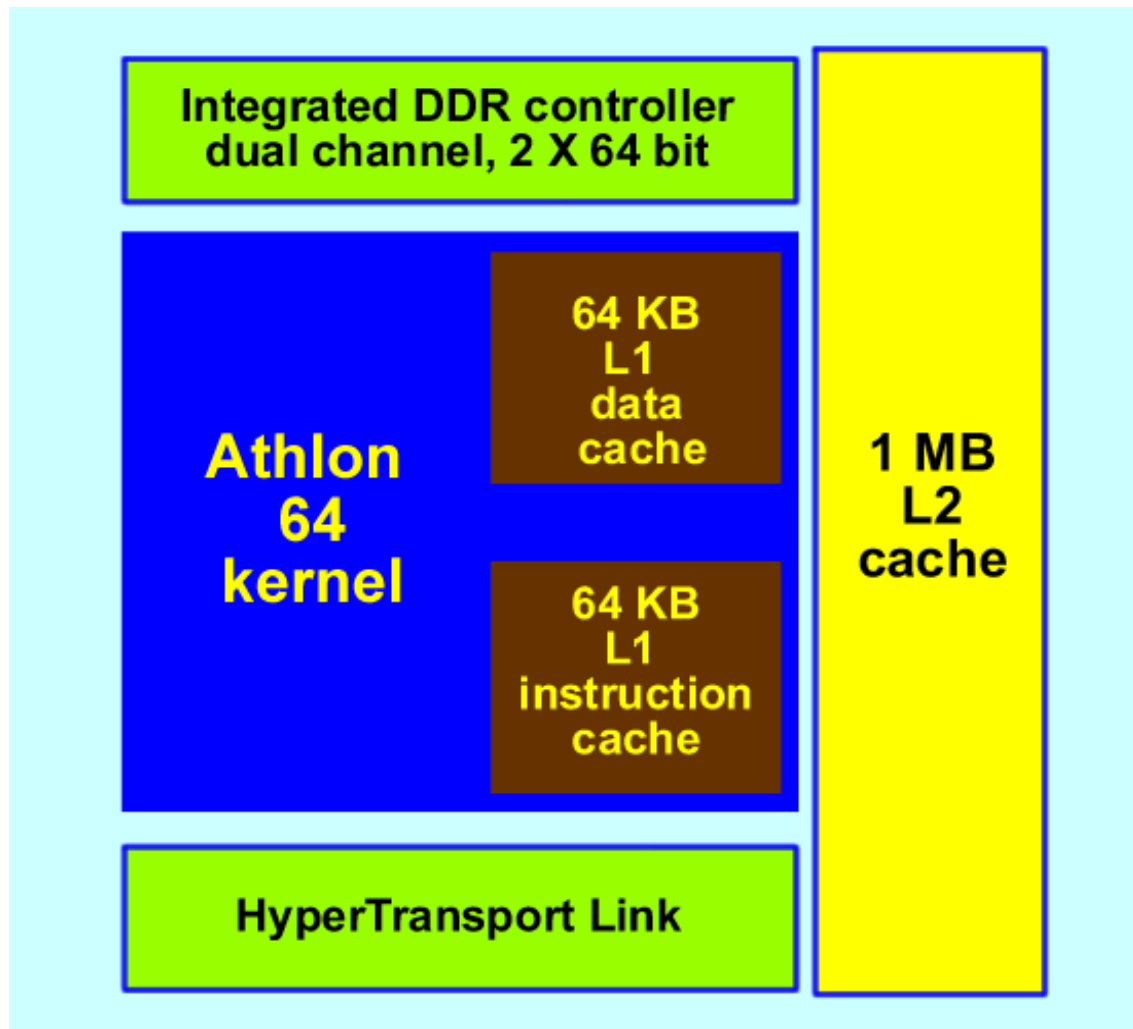
Figur 114. In the Athlon 64 the memory controller is located inside the processor. Hence, the RAM modules are interfacing directly with the CPU.

Overall the Athlon 64 is an updated Athlon-processor with integrated north bridge and 64 bits instructions. Other news are:

- Support for SSE2 instructions and 16 registers for this.
- Dual channel interface to DDR RAM giving a 128 bit memory bus, although the discount version Athlon 64 keeps the 64 bit bus.
- Communikationen to and from the south bridge via a new HyperTransport bus, operating with high-speed serial transfer.
- New sockets of 754 and 940 pins.

A complete line of chips

AMD expects to use the K8 kernel in all types of processors:

	The Opteron is the most expensive and advanced version to be used in multi-processor servers. The models are called 200, 400 and 800, and they use 2, 4 or 8 CPUs on the same motherboard – without use of a north bridge. All processors share a common memory of up to 64 GB. Each Opteron has three HyperTransport I/O channels, which each can move 6,4 GB/secund.
	The Athlon FX is a Opteron to be used in single processor configurations, high-end pc's and workstations. There is dual RAM interface, but only one channel of Hyper Transport Link.
	This is the discount version with reduced performance and lower prices. Only 64 bit RAM interface and smaller L2-cache.

Figur 115. Three versions of the latest AMD processor.

Historical overview

I will close off this review with a graphical summary of a number of different CPU's from the last 25 years. The division into generations is not always crystal clear, but I have tried to present things in a straightforward and reasonably accurate way:

Figur 116. There are scores of different processors. A selection of them is shown here, divided into generations.

But what is the most powerful CPU in the world? IBM's Power4 must be a strong contender. It is a monster made up of 8 integrated 64-bit processor cores. It has to be installed in a 5,200 pin socket, uses 500 watts of power (there are 680 million transistors), and connects to a 32 MB L3 cache, which it controls itself. Good night to Pentium.