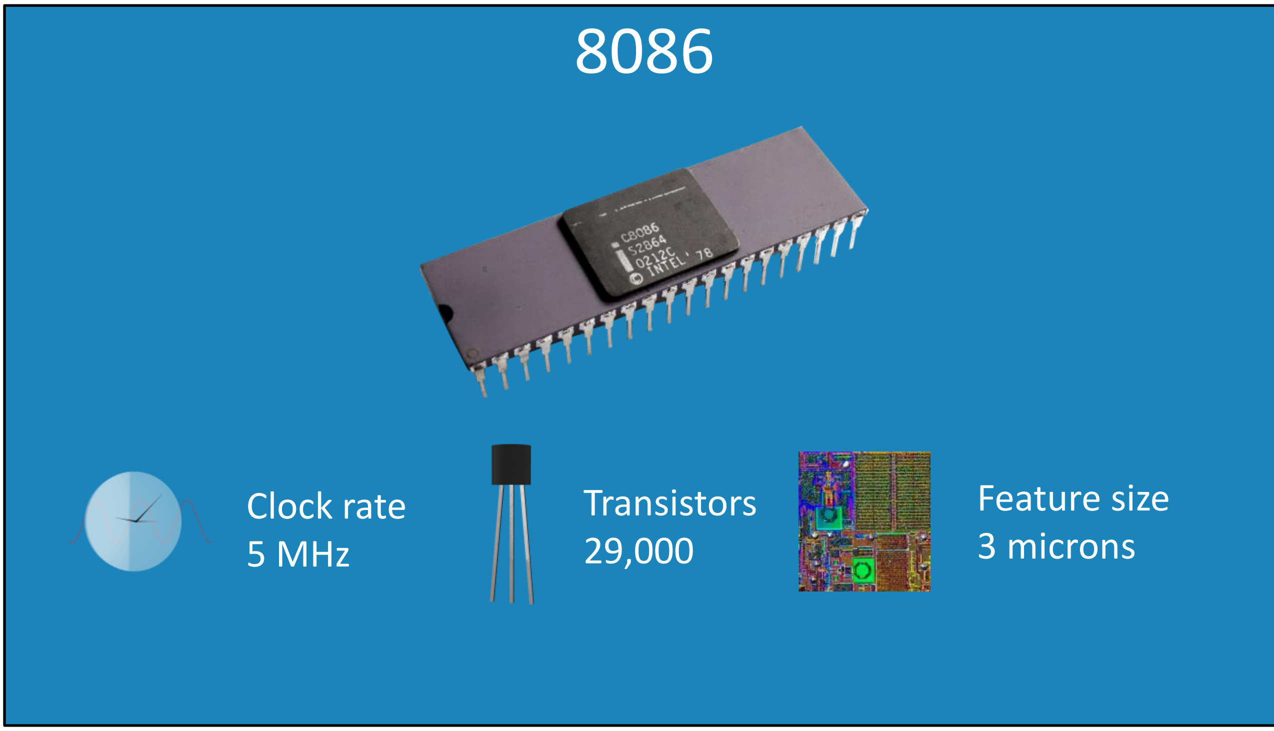
In this video from ITFreeTraining I will look at the different CPU sockets that are available. The CPU or Central Processing Unit has changed over the years and the sockets used by these CPUs have changed as technology has improved. Understanding what different sockets are available will give you an idea of which CPU you should purchase. Before I start, I will do a quick history of CPUs.



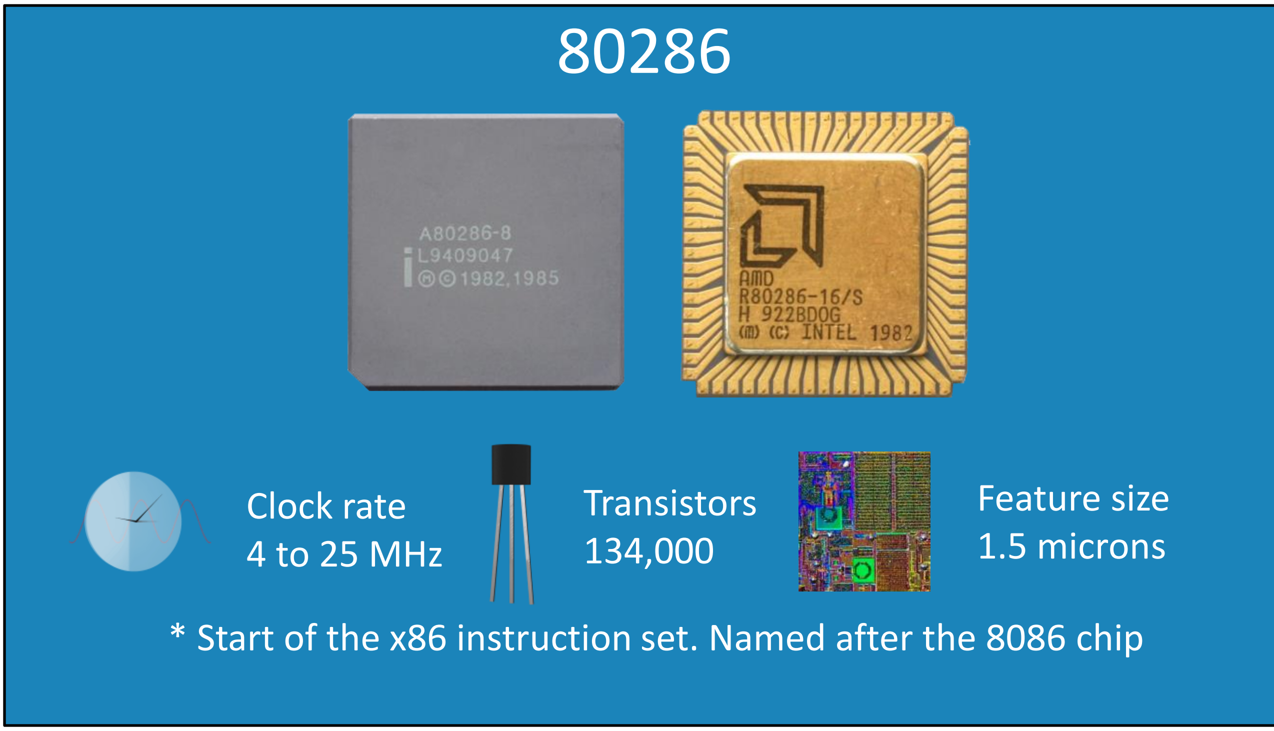
0:24 To start with, I will look at the Intel 8086. You won't need to know about the older CPUs for the exam, but having an understanding of how technology has changed will give you a better understanding of what can be achieved with modern CPUs.

This CPU was released in 1978 and you can see it looks nothing like what a modern CPU looks like. There were CPUs before this one, but I will only be talking about the main ones that had an impact on the CPUs that we use today.

This CPU is a 16-bit microprocessor. The minimum clock rate was five Megahertz. The CPU itself had 29000 transistors. The size of each transistor in a CPU is referred to as the feature size. For this CPU the feature size is three microns. A micron is one millionth of a meter. To put it in perspective, a human hair is 60 microns.

This CPU was in production until 1998. That is 20 years which is a long time in the world of computing. During this time there was a lot of demand for this CPU. The primary use of this CPU was for IBM personal computers. IBM also had a policy at the time that CPUs needed to be available from two different manufacturers.

To keep up with demand and also be compliant with IBM policy, Intel allowed other companies to manufacture this CPU. Many of these manufacturers no longer produce CPUs, however the big one that still manufactures CPUs is Advanced Micro Devices, otherwise known as AMD.



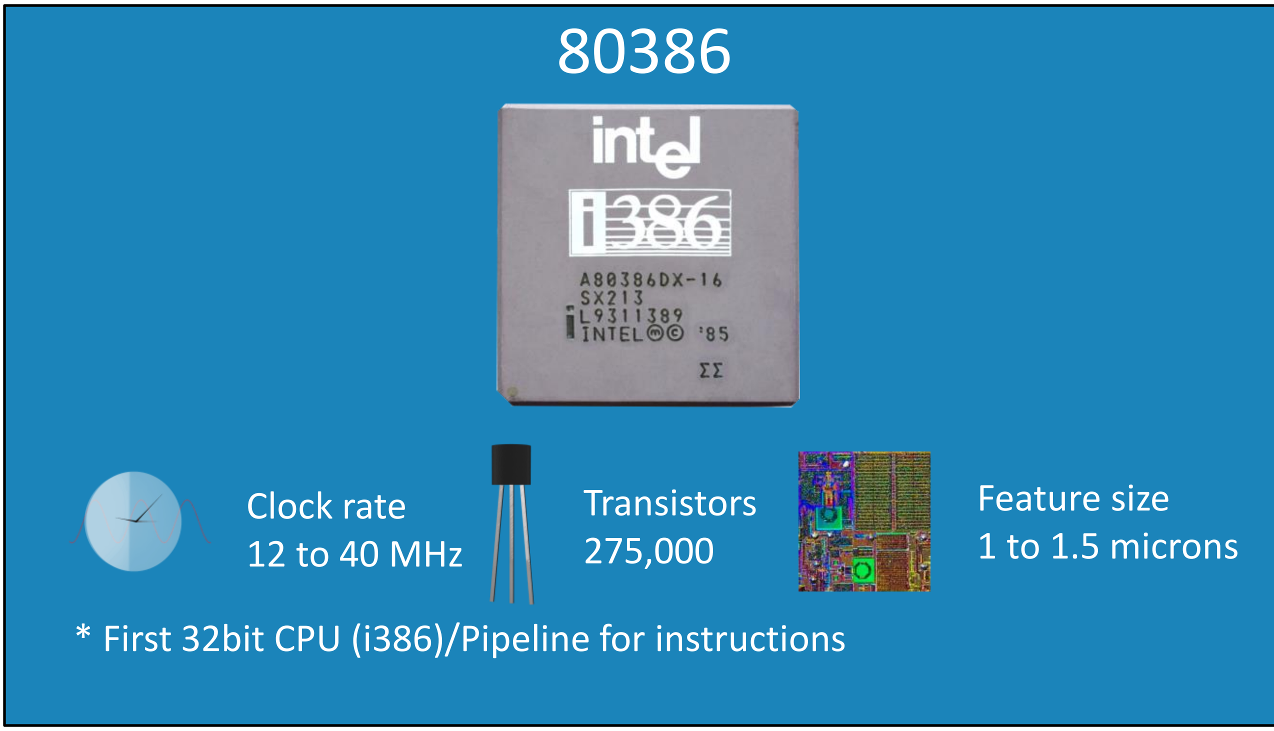
2:03 The next significant CPU is the 80286 chip, otherwise known as the 286. You can see the design of the chip has changed to something more like what we are used to seeing. This is an Intel chip; you will notice that the AMD chip, which is on the right, is identical in design.

To keep up with demand and also to meet IBM requirements to have two manufacturers for its CPUs, in 1981 Intel and AMD signed a technology exchange agreement. This agreement essentially meant that each company could manufacture the other company's CPUs. So essentially this meant (at the time) you could purchase any motherboard and install either an Intel or AMD CPU. Both would fit in the same socket. The customer did not have to make a choice when they purchased the motherboard which CPU they were going to use. Any 286 compatible CPU would work with any 286 motherboard.

The clock rate of the CPU was four to 25 Megahertz. At 25 Megahertz it was five times faster than the 8086. The transistors increased to 134,000, just over four times as many transistors as in the 8086. The feature size also decreased to 1.5 microns.

I will just make the point again, you don't need to know all these details for the exam, however having a bit of an understanding of where we have come from will give you an idea of why we do things the way we do now and why we name them the way we do.

The 286 CPU uses the same instruction set as the 8086. The instruction set essentially is the group of commands used to execute code on the CPU. Thus, this was the start of what is referred to as the x86 instruction set. This, and any other CPU that supports these instructions is referred to as an x86 compatible CPU.



4:03 The next CPU released was the 80386 in 1985, otherwise known as the 386. This CPU increased the clock speed to 12 to 40 Megahertz, the transistor count to 275,000 and the feature size, depending on the CPU, was either the same or reduced to one micron.

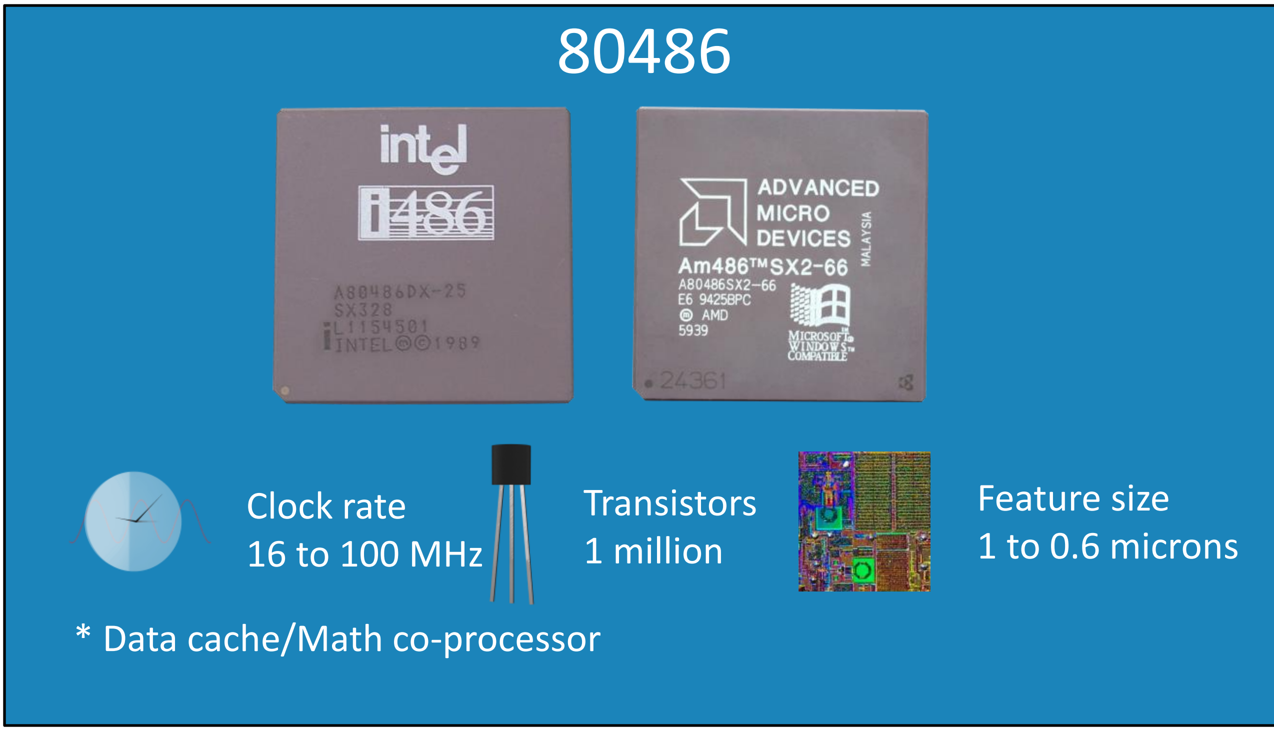
This CPU is significant because it was the first CPU released that was 32-bit and used a pipeline for instructions. 32-bit was made possible by creating extensions of the previous instruction set called the x86 instruction set. This instruction set is often referred to as the i386 instruction set. Since 32-bit is an extension of the x86 16-bit instruction, you may hear either referred to as x86. So, a download labeled as x86 will may be 16-bit or 32-bit. Now days, 16-bit code is very rare so it will most likely be 32-bit. If you see a download labeled as i386 this will be 32-bit code.

This CPU was the first to use a pipeline. A pipeline essentially means that rather than executing one instruction after the other, instructions can be executed out of order. The simplest way to understand this is to consider a conveyor belt. The conveyor belt contains all the instructions that need to be executed. A worker is at the end of the conveyor belt that processes the instructions.

Now consider that the conveyor belt has stopped. This generally occurs in a computer because it needs to retrieve something from memory. Before this CPU, when this occurred all operations on the CPU would stop. This is referred to as a CPU stall. With the introduction of the pipeline, when the CPU stops, the worker at the end of the conveyor belt simply moves down the stopped conveyor belt. The worker looks at each waiting instruction on the conveyor belt and does what it can. So essentially rather than stalling the CPU, the CPU is working on the upcoming instructions doing what it can.

This makes the CPU more efficient and faster as this is time that would normally be lost. Using this previously-lost time means that the CPU can process more instructions in the same time period than a CPU without a pipeline. You can start to understand why newer CPUs are faster than older CPUs. Even if they run at the same speed, the improvements in the CPU mean that it can execute more instructions faster. This is why you can find that a CPU with a slower clock rate can outperform a CPU with a faster clock rate.

The 386 is an important CPU, but this is when the legal problems started between Intel and AMD. Intel wanted to increase its market share and thus started to delay supplying information about its products. This caused problems with AMD selling 386 CPUs. Legal battles followed, and long story short, AMD started making clones of the 386. These cloned CPUs plugged into the same socket as the Intel 386, so the consumer could buy a motherboard and choose between buying an Intel or an AMD CPU. Essentially, they operated the same. However, sometimes in the early days, there were some compatibility problems. The Intel and AMD CPUs both had their advantages and disadvantages. This was the start of where Intel and AMD would start moving in different directions as CPU manufacturers.

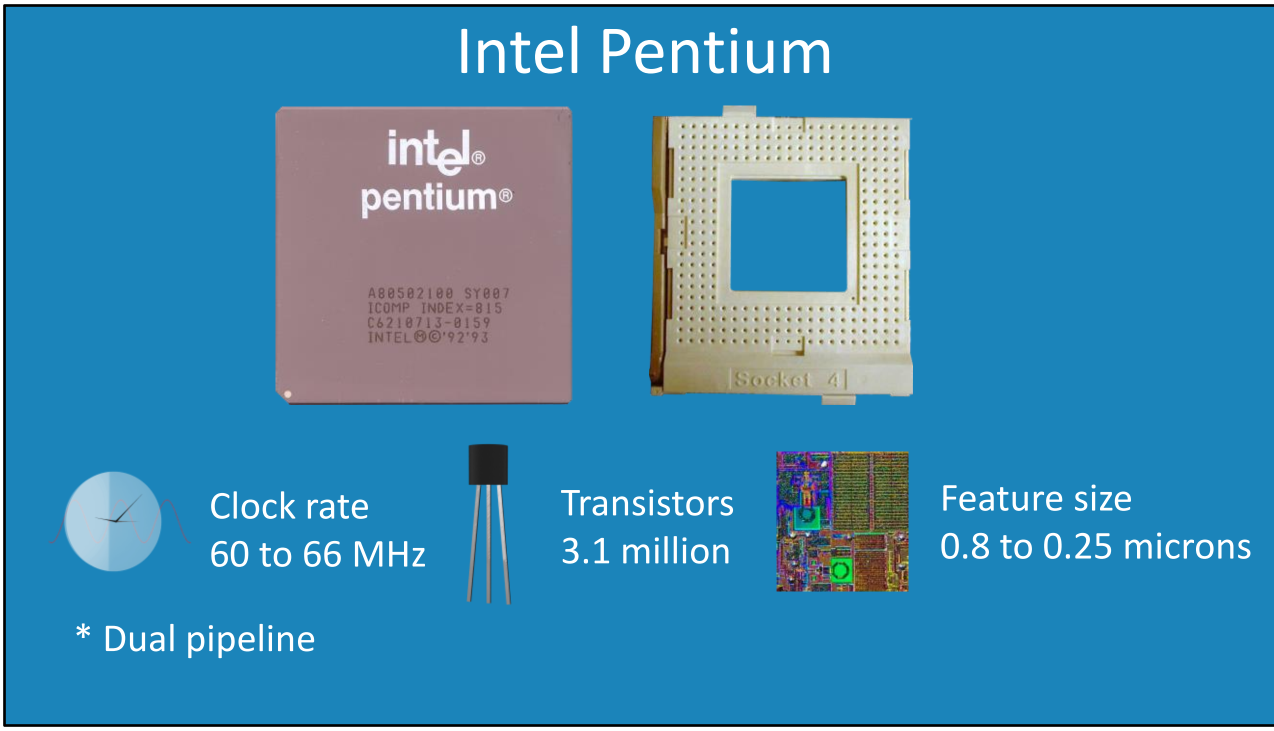


7:37 The next CPU that was released was the 80486 in 1989, otherwise known as the 486. Due to the legal issues between Intel and AMD, AMD was now creating CPUs that were a clone of the Intel CPU. Due to the AMD chip being a clone of the Intel CPU, it could still be used in the same motherboards as the Intel CPUs. However, the way Intel CPUs and AMD CPUs are designed inside is very different.

The 486 had a clock rate of 16 to 100 Megahertz. The 486 was also the first CPU to have just over one million transistors in it. The feature size of the CPU was one to 0.6 microns. This is four times the transistor count of the 386. The main reason the transistor count is so high is that the 486 includes two things previously not included in the CPU. These were a data cache and a math co-processor.

The data cache allows the CPU to store a copy of a small amount of the computer's memory inside the CPU. This means, if the data is available in the cache, the CPU does not have to wait to access it from the memory, it can access it directly from the cache. This is faster than accessing it directly from memory.

The next added feature is a math co-processor. Before this, other than basic math functions, complex math functions were performed in an optional math co-processor that could be installed on the motherboard. Although not included in all 486s, the math co-processor sped up complex math functions that you would generally find in 3D software and computer design software.



9:26 The next CPU was the Intel Pentium released in 1993. Due to trademark issues – previously Intel had attempted to trademark the 486, but lost in court to AMD – rather than name it the 586, it was called the Pentium.

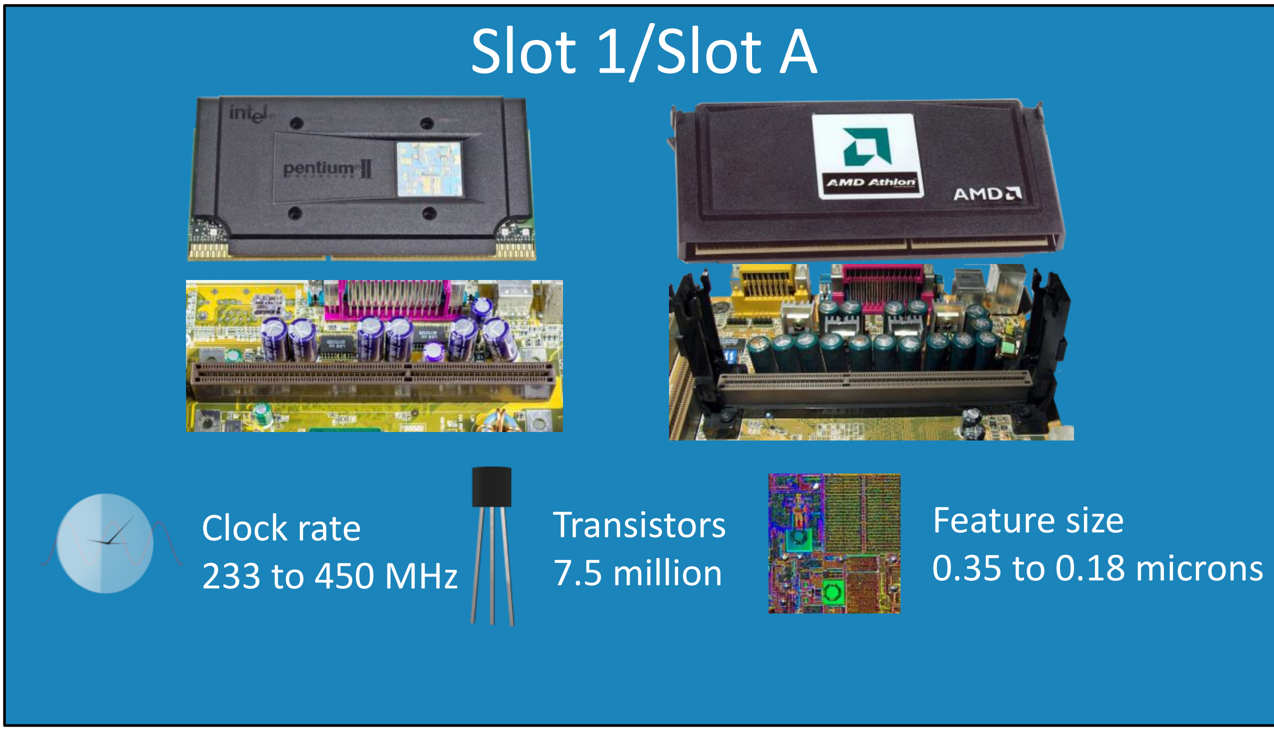
The first Pentium plugged into a socket called socket 4. Sockets are designed so essentially any CPU that plugs into that socket should work. This would mean that AMD could still create CPUs and as long as they were socket 4 compatible they would work in a socket 4 motherboard.

There were sockets before this, but this is when the socket concept and socket naming started to take off. Nowadays, most of the time if you purchase a particular CPU with a particular socket, it will work with the motherboard. However, in some cases certain motherboards will required a CPU of a particular generation, so you should always check that your CPU and motherboard are compatible.

The first Pentium ran at a clock rate of 60 to 66 Megahertz. You could get a 486 that ran at 100 Megahertz, but what made the Pentium faster was that it had a dual pipeline. Starting with the Pentium it could essentially execute two instructions at once. So even though it ran at a slower rate, it could still achieve more than the 486 could. This is why when looking at CPU performance, the clock rate is only one of the things that should be looked at.

The Pentium had 3.1 million transistors, an increase from the one million in the 486. The feature size was 0.8 to 0.25 microns. The Pentium became a brand name for Intel that still exists to this day.

There were some more sockets after this for Intel, with some of these supporting AMD. However, the biggest change would come in 1997.

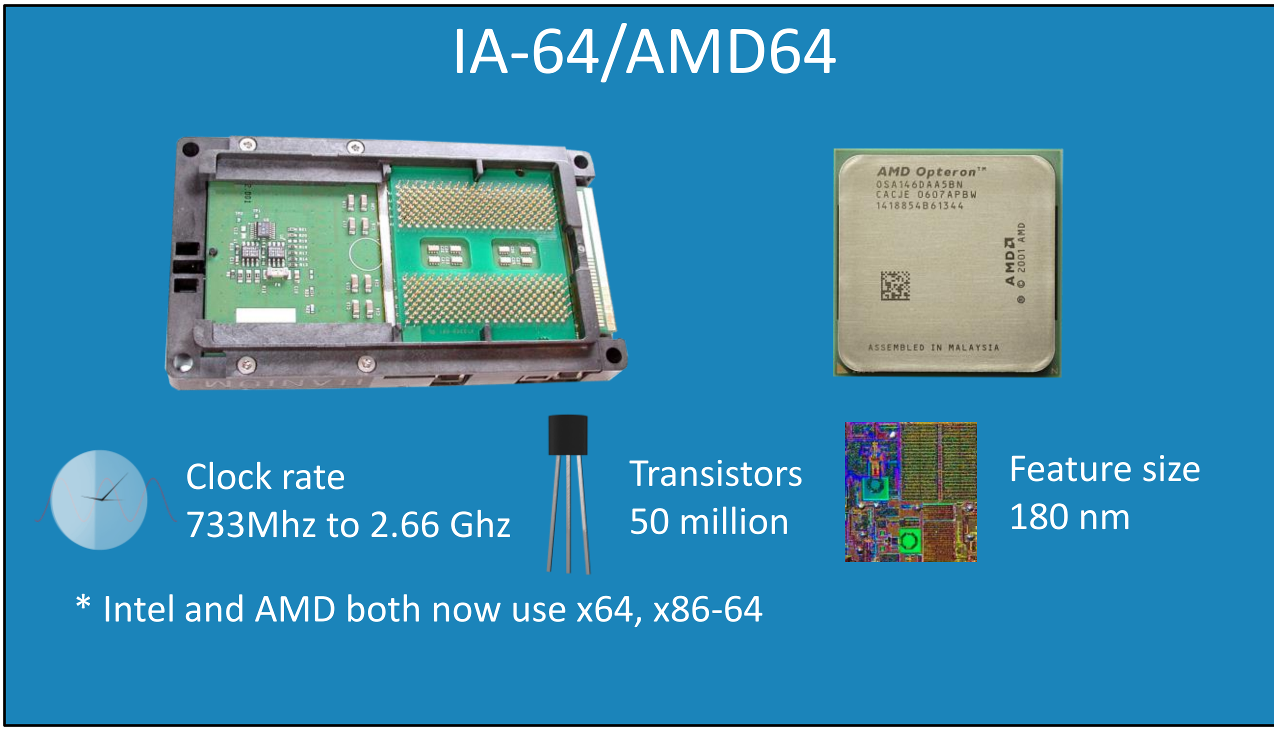


11:22 In 1997, Intel changed the look and design of the CPU completely; this was the Slot 1. Slot 1 uses a single edge cartridge style connector. The reason behind this was that, with the cache and the CPU being combined together, the failure rate for CPUs was too high. Using a circuit board, the CPU and cache could be manufactured separately and put on a circuit board. This allowed each to be tested separately. If one failed testing, they could simply use another one, thus decreasing the cost of manufacturing.

AMD released Slot A in response to Intel. It had the same number of pins as Slot 1, however it was not electrically compatible. This essentially meant that an Intel motherboard could not be used with an AMD CPU. This change had already started a year before this with the release of the AMD Super Socket 7. However, with the release of Slot 1 and Slot A, this really highlighted that both companies were going in different directions in relation to motherboard and CPU compatibility.

There were a few CPUs released for these slots, so I will use the first one released, the Pentium II as an example. The Pentium II had a clock rate of 233 to 450 Megahertz. The transistors increased to seven and a half million and the feature size decreased to 0.35 to 0.18 microns.

Slot 1 was short lived and was discontinued in 1999 to return back to the socket approach. With improvements in CPU manufacturing, there was no need to separate the cache and the CPU and they were once again put inside the same chip. Before I look at that, something else important happened in that year.



13:11 In 1999 Intel and AMD both released their 64-bit architecture, and about two years later both companies released CPUs for their 64-bit architectures. In the case of Intel, this would take the form of the Itanium CPU. Don't be fooled by the design, it is not like the Slot 1 design. The CPU and cache are inside the same chip. The board the CPU is attached to only contains some electronic components. Later releases of the Itanium would go back to having the single chip design.

The Itanium CPU was a complete redesign of the old x86 instruction set. The Itanium CPU required instructions to be optimized and to be in a certain order to execute. This required new compilers to be created to generate code for these CPUs. This turned out to be difficult to achieve and thus limited the amount of software that could be run on the Itanium CPU. The Itanium CPU was not very popular in the market place. Intel announced that Itanium CPUs would be discontinued and that sales of this CPU would end in 2021. Itanium was only used in the server market. Itanium servers are pretty rare, so you most likely won't come across any. If you see something labeled like IA-64 this is software for Itanium.

AMD released their 64-bit CPU which was the AMD Opteron. This CPU, rather than completely re-designing the instruction set, simply created extensions to the 32-bit instruction set to make it support 64-bit. This turned out to be very popular.

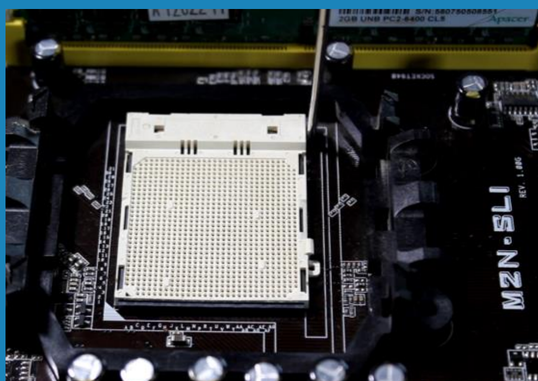
These CPUs started with having speeds of 733 Megahertz up to 2.66 Gigahertz. The transistor count was now around 50 million. Feature size was also reduced to around 180 nanometers.

With the success of AMD 64-bit instruction set, Intel had to follow. Now both Intel and AMD use the same 64-bit instruction set. Although it may be referred to by many different names, it is commonly called x64 or x86-64.

Essentially nowadays both the Intel and AMD CPUs use the same basic instruction set. Therefore, software compatibility should not be a problem. Where the instructions differ is for extensions like those used to run virtual machines. Since both manufacturer's CPUs are implemented differently, you may find that in running certain software you get different performance depending on which CPU you are using.

Earlier Sockets

- PGA-ZIF (Pin Grid Array-Zero Insertion Force)
- Intel generally name after the number of pins



15:54 With the change from slot back to sockets, let's now have a closer look at the socket. The socket we will look at is the PGA-ZIF socket. This socket dominated the market until 2004. PGA stands for Pin Grid Array. This essentially means that the pins are mounted on the CPU and the socket contains pin holes.

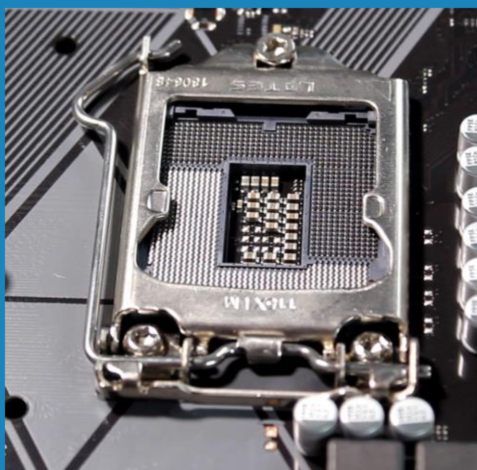
The second part of the name, ZIF, stands for Zero Insertion Force. This, as the name suggests, means zero force is used to insert the CPU. You can see that when I put a CPU in the socket, gravity does the work of it going into the socket. The sockets are keyed so the CPU will only go in one way. If you find the CPU will not go into the socket, it may be that it is the wrong way around. The socket and CPU will have a triangle in one corner to indicate which way it should be put in. Before putting the CPU into the socket, make sure these triangles line up.

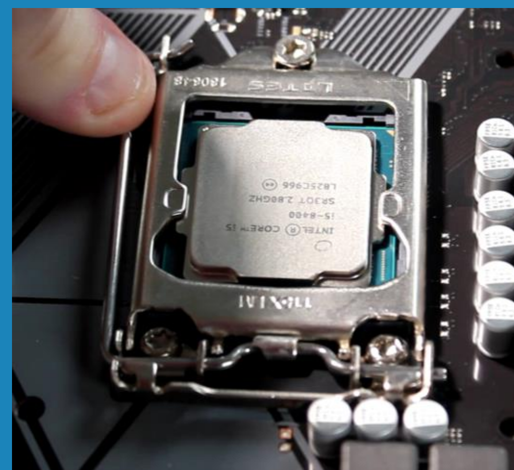
Once the CPU is in place, there is generally a lever or similar device that locks the CPU into place. This prevents the CPU from moving and gives it a good connection to the socket.

Intel generally names the sockets after the number of pins in the socket. For example, a socket 495 would have 495 pins. More pins allows for more data to be transferred in and out of the CPU, which generally means better performance. However, there are a lot of factors to consider when purchasing a CPU which effect the performance that you will receive.

Current Sockets

- LGA (Land Grid Array)





17:31 Next, I will look at the current sockets being used on the market. The current sockets being used are LGA or Land Grid Array. The big difference is that the pins are not in the CPU but on the motherboard. Having the pins on the chip increases the chance they may be bent or damaged when handling the CPU. This change was made to reduce the risk of damaging the CPU when installing or handling it.

LGA has a triangle in one corner like the ZIF socket to indicate which way the CPU should be installed. Make sure you match the triangles up when installing the CPU. Since the CPU and socket are keyed, if the triangles do not line up the CPU will not go into the socket. If you try to force it, this will damage the motherboard or the CPU.

To install the CPU, there will generally be a retention lever which, when released, will also release the hold down bracket. It is just a matter of placing the CPU into the socket. If installed correctly, the CPU should go gently into the socket. Once installed, put the hold down bracket back into place and secure the retention lever.

As before, Intel named the socket after the number of pins in the socket. For example, the LGA 2011 socket has 2011 pins. AMD however continues to follow their trend of not naming the socket after the number of pins in the socket.

Current Market

- Technology changes very quickly
  - Microarchitecture changes about every 3 years
  - Motherboards don't support all CPUs
- Different CPUs support different features
  - E.g. power saving/performance

Clock rate
3 to 5 Ghz

Transistors
5 Billion

Feature size
10-20nm

18:59 This covers all the CPU sockets and their development over time. I will now have a look at the current market. Technology changes very quickly. Approximately every three years, the microarchitecture changes. Microarchitecture is the CPU design and the chipset that supports it. Since everything changes so quickly, not all motherboards support all CPUs of a particular socket. So, don't assume that if you purchase a particular socket of that CPU it will work in a motherboard of the same socket.

Different CPUs have different features. For example, some CPUs are designed with power saving in mind and others are designed with performance in mind. Some CPUs may also have features like built-in graphics. Other CPUs are designed to be used in servers. Nowadays, there are so many different CPUs available, all designed with a different purpose in mind, your choice of CPU will be determined by what you need – things like performance, amount of memory you want to be able to address, will all be factors in your decision making.

Nowadays, the clock rate of a CPU can be as high as three Gigahertz to five Gigahertz. Currently, the higher end speeds are found in server CPUs. CPUs have billions of transistors. With so many CPUs on the market it is hard to get an exact number, but let's say the average is around the five billion mark.

The feature size has decreased to 10-20 nanometers. Ten nanometers has only happened recently. However, the feature size is starting to get so small, manufacturers are starting to meet some hard engineering limitations. If you look hard enough, you will find that a CPU of ten nanometers is probably not all ten nanometers. Parts of the CPU will be ten nanometers and other parts will not. But the point to take away is that we started at three microns and are now down to ten nanometers. You will find that even if we can't get any smaller, the trend is to start making the CPU larger and also technology is starting to be used to stack multiple layers on top of each other. Who knows what the future may bring, but we can say for sure we have come a long way since the 8086.

Summary

- i386 refers to 32-bit Intel instruction set
  - Clones of Intel also used this instruction set
- x64 or x86-64 refers to 64-bit instruction set
  - Both Intel and AMD use this
  - Extensions to instruction set are vendor specific
- Current socket is LGA (Land Grid Array)

21:18 Before ending this video, I will do a quick summary of the main points. i386 refers to the 32-bit instruction set. This was originally developed by Intel. However, clones of the Intel CPU also used this instruction set.

x64 or x86-64 refers to the 64-bit instruction set. This is the instruction set used both by Intel and AMD in CPUs currently on the market. In the old days there may have been some software that would not run correctly on both Intel and AMD due to small differences in the instruction sets. Nowadays, this should not occur. However, both manufacturers are always adding extensions to their CPUs. Extensions include support for things like virtual machines and improvements to graphical processing. Your software will query the CPU to see if it supports what it needs. In specialized cases like virtual machines, there may be two different versions of the software, one for Intel and one for AMD.

The current socket used by both Intel and AMD is LGA or Land Grid Array. Essentially this means the pins for the socket are on the motherboard reducing the chance of damage to the CPU when installing or handling.

This concludes this video on CPUs and sockets. I hope you have found this video useful and I look forward to seeing you in the next video from us. Until the next video, I would like to thank you for watching.

References

"The Official CompTIA A+ Core Study Guide (Exam 220-1001)" Chapter 3 location 5509 – 5661

“CompTIA A+ Certification exam guide. Tenth edition” Pages 96 – 101
“CPU die” https://commons.wikimedia.org/wiki/File:PXE.jpg
“Advanced Micro Devices” https://en.wikipedia.org/wiki/Advanced_Micro_Devices
“Intel 8086” https://en.wikipedia.org/wiki/Intel_8086
“File:Intel C8086.jpg” https://commons.wikimedia.org/wiki/File:Intel_C8086.jpg
“Intel 80186” https://en.wikipedia.org/wiki/Intel_80186
“80286 picture” https://commons.wikimedia.org/wiki/File:KL_Intel_i286.jpg
“AMD bottom of chip”
https://en.wikipedia.org/wiki/File:KL_AMD_80286_CLCC_Bottom.jpg
“Picture of 386” https://en.wikipedia.org/wiki/File:KL_Intel_i386DX.jpg
“Picture of 486”
https://commons.wikimedia.org/wiki/File:Intel_i486_DX_25MHz_SX328.jpg
“Picture of am486” https://commons.wikimedia.org/wiki/File:KL_AMD_486SX2.jpg
“Picture Intel Pentium”
https://commons.wikimedia.org/wiki/File:A80502100_sy007_pentium_observe.png
“Picture Socket 4” https://en.wikipedia.org/wiki/File:Socket_4.jpg
“Picture Slot 1” https://en.wikipedia.org/wiki/File:Asus_P3C2000_-_Slot_1-8626.jpg
“Picture Slot A” https://en.wikipedia.org/wiki/File:Slot_A.jpg
“Picture AMD Opteron”
https://en.wikipedia.org/wiki/File:AMD_Opteron_146_Venus,_2005.jpg
“Transistor count” https://en.wikipedia.org/wiki/Transistor_count
“Pin grid array” https://en.wikipedia.org/wiki/Pin_grid_array

Credits
Trainer: Austin Mason http://ITFreeTraining.com
Voice Talent: HP Lewis http://hplewis.com