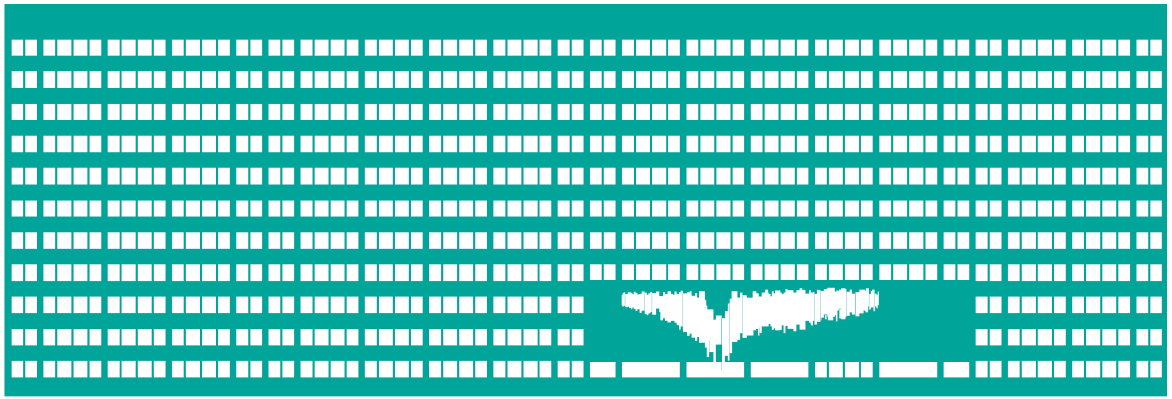
VŠB TECHNICKÁ UNIVERZITA OSTRAVA

VSB TECHNICAL UNIVERSITY OF OSTRAVA



www.vsb.cz

VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

APPS
Architektury počítačů a paralelních systémů / Architecture of Computers and Parallel Systems

Part 04: Intel x86 History

Ing. Petr Olivka, Ph.D.
Department of Computer Science,
FEECS, VSB-TUO
petr.olivka@vsb.cz
http://poli.cs.vsb.cz

CISC Processor - Intel x86

This chapter will introduce the CISC processors evolution. We will try to illustrate the history on one typical processor, because the comparison of multiple processors simultaneously would not be clear for readers. But the selection of one typical processor is complicated due to a variety of products and manufactures in the past 40 years.

We have decided to describe in this presentation one of the best-known and longest mass-produced processors in existence.

We definitely do not want to say that it is the best technology or that these are processors with the highest performance!

The Intel x86 processors are the selected product line.

Predecessor of x86 - Intel 8080

(Year-Technology-Transistors-Frequency-Data bus-Address Bus)

Y: 1974 T: NMOS 6µm Tr: 6000 F: 2MHz D: 8b A: 16b

This 8 bit processor is not directly the first member of x86 series, but it can not be skipped. It is one of the first commercially successful microprocessors. This microprocessor became the basis for a number of the first single-board computers and its instruction set inspired other manufacturers to develop 8-bit processors. It is compatible at the assembly level (not machine instructions) with its successor – a 16-bit version of the 8086.

In 1977 it was replaced by a newer version of the 8085 and manufactured until the mid-eighties.

Intel 8086

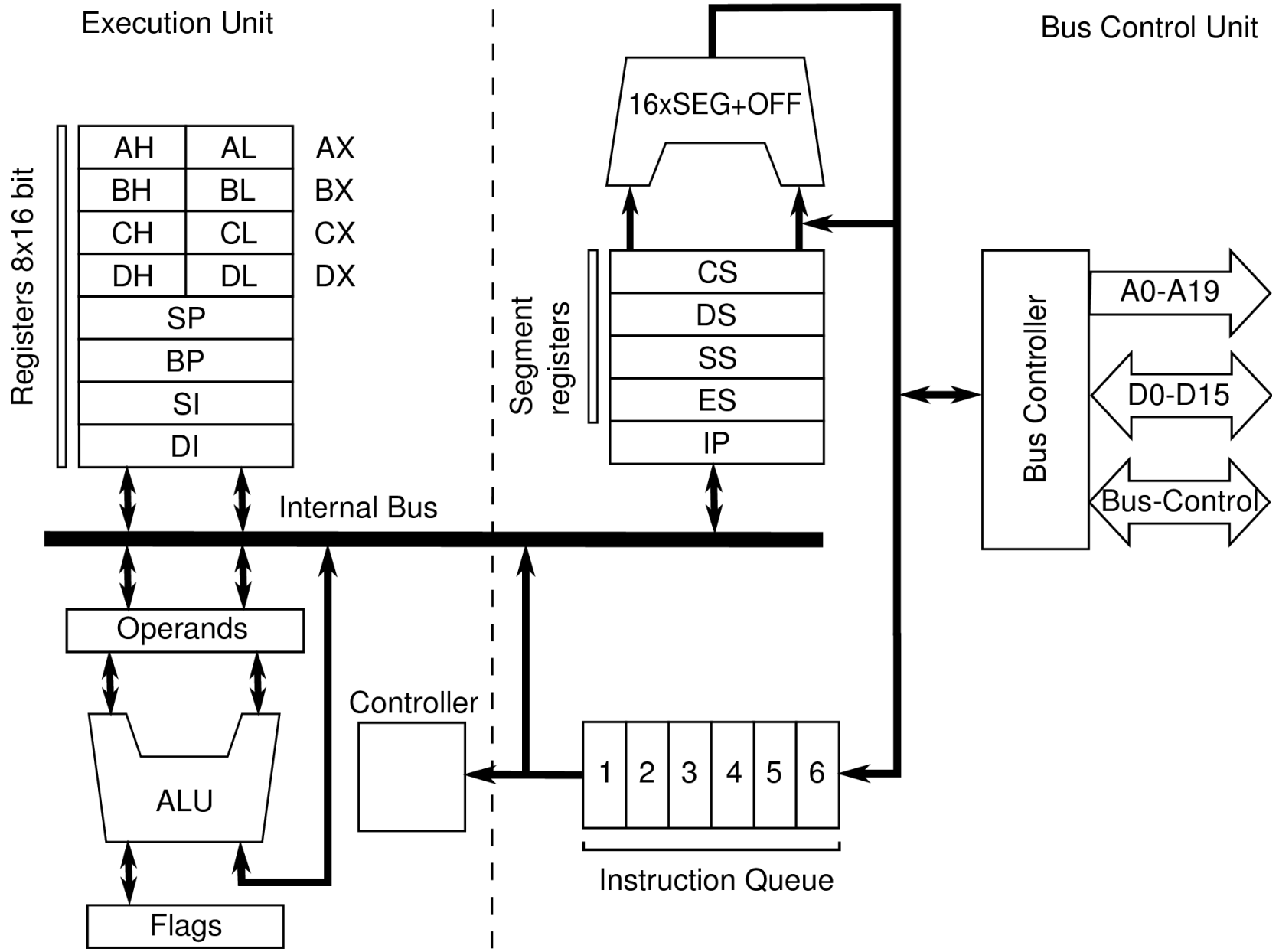
Y: 1976÷1990 T: HMOS 3.2µm Tr: 29000 F: 4.77÷10MHz D: 16b A: 20b

This is the first 16-bit processor. It is able to address up to 1MB of memory using the 64kB block segmentation.

The internal scheme is visible on the next slide. From the scheme, it is clear that CPU is divided into two parts: execution unit and bus control unit. It contains eight 16-bit wide registers and four segment registers. This set of registers is the basis for all subsequent generations.

The instruction queue is only 6 bytes long and after jump instruction it is erased and refilled.

... Intel 8086 Scheme



Intel 8088

The year after the introduction of 8086 processor, Intel presents modification of 8086 – the processor 8088. It was not a new type of processor because its all internal architecture remained unchanged, only the data bus width was reduced to 8-bit.

The performance of that processor was decreased, because a 16-bit wide data have to be read in two bus cycles. But at the time there were many available peripheral chips for 8-bit microprocessors and they could be used for the Intel 8088 as well. Computers with the narrowed data bus are cheaper and simpler.

Perhaps the best evidence of the suitability of this processor for the personal computer production, was the new standard PC-XT designed by IBM.

A legend is starting ...

Intel 80186/80188

Y: 1982÷2007 T: HMOS 1.3µm Tr: 55000
F: 6÷25MHz D: 16/8b A: 20b

In 1982 the Intel introduces the new processor 80186. It implements improved architecture and many newly modified and accelerated instructions.

The processor was primarily designed for embedded devices and it integrates many peripheral devices directly on the chip, especially DMA controller, clock, timers and ports. Thus it was incompatible with the PC-XT architecture. But some manufacturers used it in their personal computers.

The success of this processor is evident by the fact that it was produced for 25 years and over these years it was gradually modernized. It was also licensed to many other manufacturers.

Intel 80286

Y: 1982÷1993 T: HMOS 1.5µm Tr: 134000
F: 6÷25MHz D: 16b A: 24b

In 1982 the Intel introduced the second x86 generation – the processor 80286. Higher performance was achieved by two enhancements: most instructions need less machine cycles for execution and the processor works on the higher frequency.

The processor implements the new *Protected Mode* with possibility to address up to 16MB of RAM. For the backward compatibility the processor works in *Real Mode*. But programs written for *R-M* could not run in *P-M*. The processor uses in the *P-M* the MMU (Memory Management Unit) and it implements *Virtual Memory* in range 1GB.

P-M and *V-M* allow programmers to develop new safe, multitasking and multiuser OS. These technologies were at that time usually used only in mainframe computers.

The IBM uses this CPU for the new standard known as PC-AT.

Intel 80386DX

Y: 1985÷2007 T: CHMOS 1.5÷1µm Tr: 275000 F: 16÷40MHz D: 32b A: 32b

In 1985 the Intel presented the first full 32-bit processor with the full backward compatibility in *R-M*. All registers were extended to 32-bit. The address and data bus are 32-bit wide. In *P-M* the processor was able to address 4GB memory for each process and up to 64TB of the whole virtual memory. This addressing mode has been used until today. The processor implemented new *Virtual Mode* for compatibility with old *R-M* programs. Therefore old MS-DOS programs could be used in new operating systems.

The processor had cache memory controller for fast level 1 (L1) cache memory on the board. This memory was necessary for processors operating at a frequency over 30 MHz. Recommended size was 8÷32kB.

It is not without interest that the first computer with 80386DX processor was introduced by Compaq.

Intel 80386SX and FPU

In 1988 Intel launched the 80386SX processor. It was a backward step because at that time the 80286 was very popular and there were many peripherals available on the market for the 16-bit bus. Intel modified 80386DX for the 16-bit data bus and created the SX version. This processor allowed manufacturers to produce cheaper computers assembled from available chips.

FPU coprocessor 8087/287/387

The first three generations of Intel processors were produced without units for floating point numbers computing. These Floating Point Units, called coprocessors, were produced as separate circuits. The coprocessor had to cooperate with the main processor and its independent activity was not possible. Computer manufacturers had to implement a separated slot for coprocessor on the boards.

Intel 80486DX

Y: 1989÷2007 T: CHMOS 1÷0.6µm Tr: 1.2 mil. F: 16÷100MHz D: 32b A: 32b

The forth generation of the Intel 80486DX processor brought the big step in the development. It was also the last pure CISC processor.

The great increase of number of transistors indicated a lot of modernizations. The processor was twice as powerful as the previous processor version running on the same frequency. Many improvements were done in ALU, in instruction queue and in throughput between internal parts. The processor contains the L1 cache with size of 8kB shared for data and instructions. MMU unit was improved too, especially for the higher performance in the protected mode. Another innovation was the FPU unit implementation directly on the chip. It did not improve performance of ALU, but FPU unit implemented directly in processor was much faster than standalone unit on the board.

... Intel 80486DX/DX2/SX

In 1991 Intel introduced cheaper version of the 486DX processor – 486SX version. This version did not implement the narrower data bus, as it was in the 80386SX version, but the 486SX did not implement the FPU.

The processor passed the next modernization in 1992. Intel introduced the 486DX2 version. This version doubled the internal clock frequency and it was fully pin-compatible with the 486DX. The first version was the DX2 50/25 MHz and 66/33 MHz followed.

In 1994 Intel introduced the next acceleration. The version 486DX4 multiplied internal frequency three times (not expected 4x). Two versions 75/25 MHz and 100/33 MHz were supplied due to high frequency only by 3.3V and the processor lost pin-compatibility with the version 486DX.

(Intel) Pentium

Y: 1993÷2001 T: BiCMOS 0.8÷0.25µm Tr: 3.1 mil.
F: 60÷300MHz D: 64b A: 32b

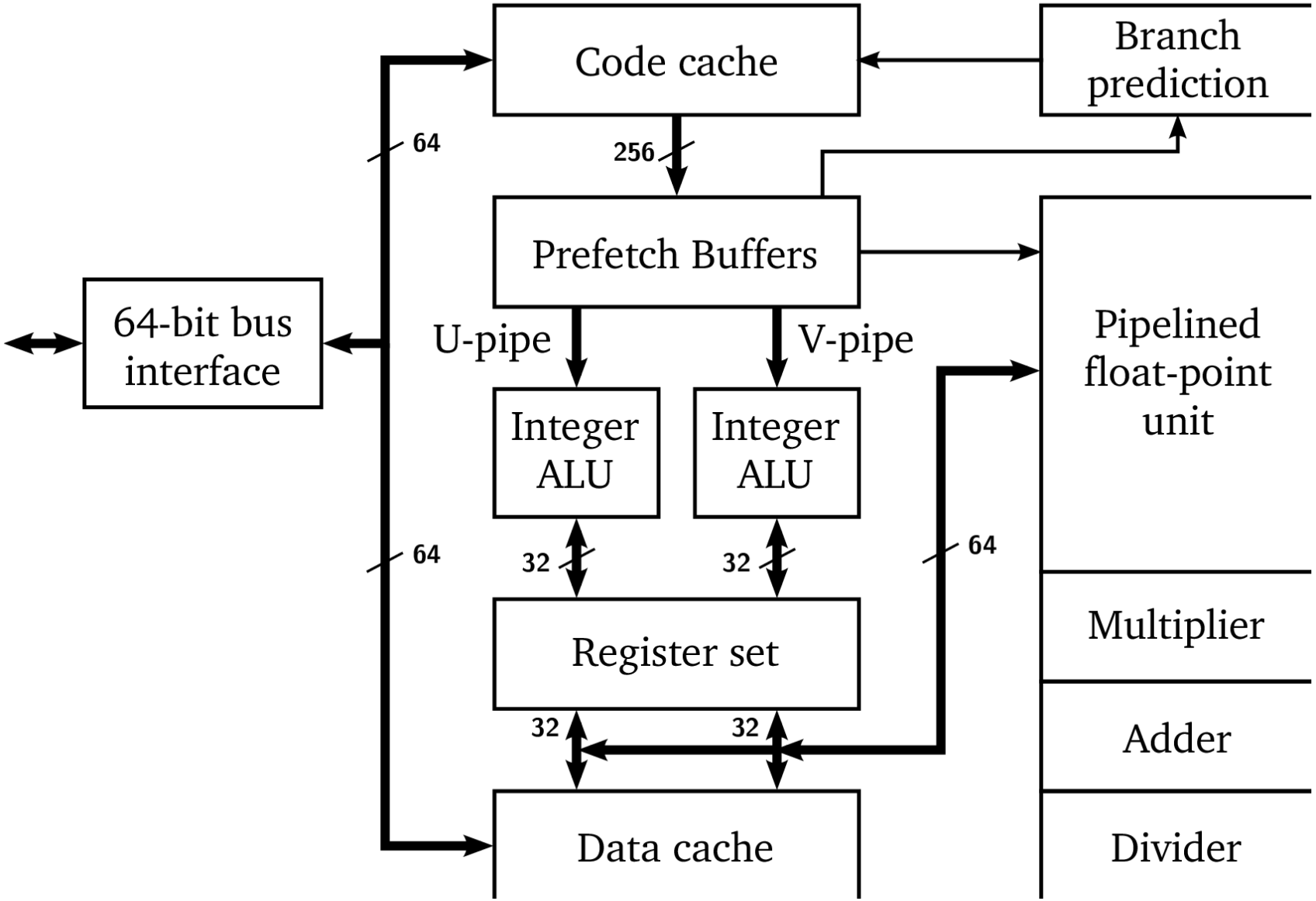
The fifth generation of processor called Pentium was introduced in 1994. It was the first x86 processor that implemented some RISC and superscalar features. The processor implemented two parallel ALU units. In ideal situation they could both work parallelly. When the processor executed some complex instruction both units had to cooperate. The branch prediction unit was implemented in the processor too. The processor contained separated L1 caches for data and code. The FPU was integrated there as well.

The first processor version worked at 5V with frequency 60/66MHz and it was installed very seldom. The first commercially successful version came in 1995. It had 3.3V supply and it worked with frequencies of 75/90/120MHz.

In 1997 Intel added to the Pentium the multimedia unit MMX.

... Pentium Scheme

Pentium Block Diagram



... Pentium Scheme

It is clear that the execution unit consists of two ALUs. Both units are supplied from bottom by data and from top by instructions. The cache memory is strictly divided for data and code. Other circuits are internally connected together by the wide internal buses.

The instruction queue implemented in Prefetch Buffer uses Branch Prediction to decrease queue filling failures.

The Floating Point Unit is still in separated part of CPU and it cooperates with ALUs. Later it was extended by MMX unit.

Pentium Pro

Y: 1995÷1998 T: BiCMOS 0.5÷0.35µm Tr: 5.5 mil.+15 mil./256kB L2
F: 150÷200MHz D: 64b A: 36b

The sixth Intel processor generation brought a major technological breakthrough. The processor development started together with Pentium, but at that time the goal was very challenging. The new RISC processor should be developed with the backward CISC instruction compatibility.

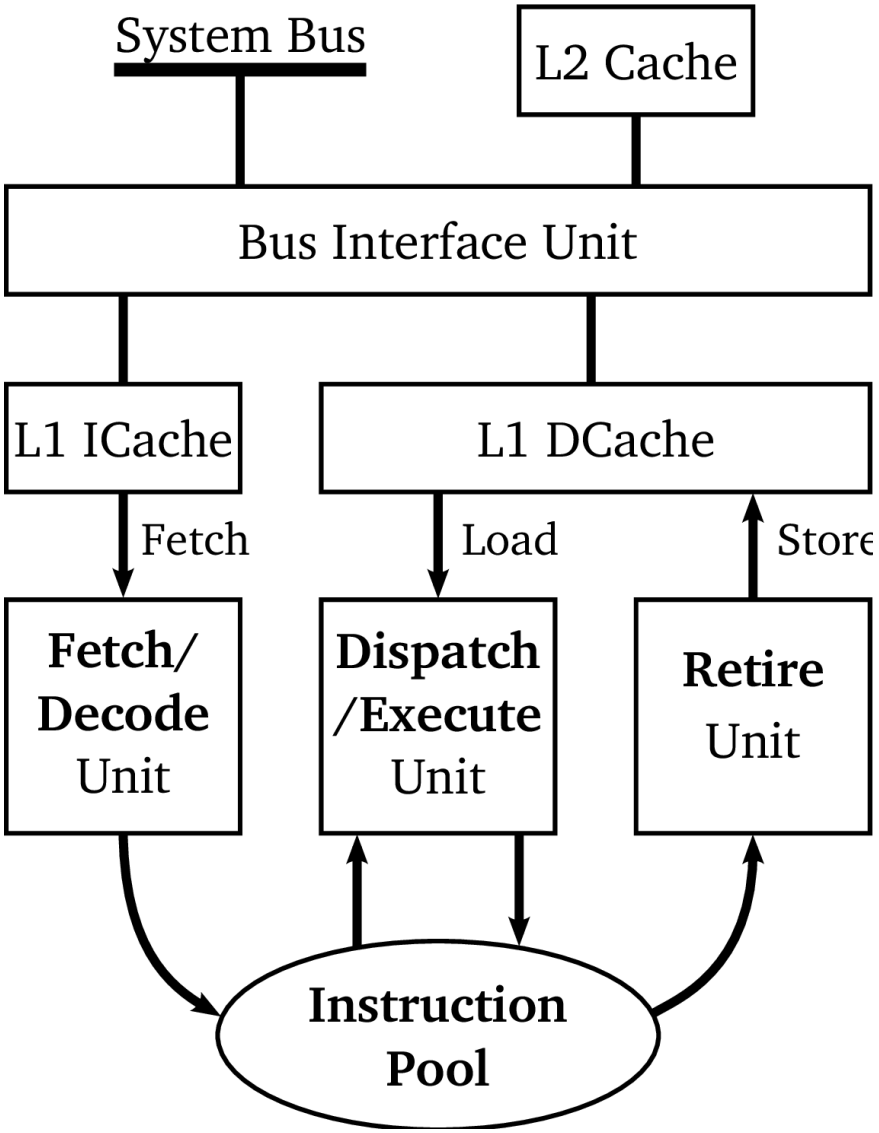
The result was introduced in 1995 – the processor Pentium Pro. It was primarily designed for servers thanks to its higher performance, but with higher price. The advantage was also wider Address Bus – 36 bits. It allows to address up to 64GB physical memory.

Because the processor introduced brand new architecture, which has been used until today, we will describe this processor in more detail.

It would be best to start with simplified internal scheme.

... Pentium Pro

Pentium Pro Block Diagram



... Pentium Pro

The upper part of the scheme shows the first major change. The L2 cache memory is there implemented directly in CPU. But the cache memory is made in separate chip and it is connected directly to processor chip in package. It works on external bus frequency.

The Processor's Bus Interface Unit is directly connected to L1 Caches, separated for code and data.

The most significant changes follow.

The Fetch & Decode Unit reads old CISC x86 instructions from the memory and decodes them to one or more RISC instructions, called **micro-operations**. All micro-ops have the same length – 118 bits. The Decode unit is very complex and is composed of more internal parallelly working units to decode CISC instructions fast enough.

The decode unit is followed by the pure RISC processor.

... Pentium Pro

Decoded instructions are not progressing into the instruction queue as one may expect. Instructions are stored in the Instruction Pool. It is a bank of 40 instructions.

The Execution unit can select instructions from the pool in order to maximize the performance. This technology is called **Out of order execution**. The Execution unit is very complex and contains more parallelly working units: pair of ALUs and FPUs units.

Executed instructions are put back to the pool with results. After that the Retire Unit stores results back to registers and to the L1 cache. Results continue through the bus interface unit to the L2 cache and to the main memory.

The processor contains Branch Prediction Unit and in the previous scheme is hidden in Fetch and Decode Unit. It can store up to 512 predictions. Prediction success rate is usually about 90%.

Pentium II

Y: 1997 T: BiCMOS 0.35÷0.18μm Tr: 7.5 mil.+20 mil./256kB L2
F: 233÷533MHz D: 64b A: 36b

The Pentium II processor was introduced in 1997 and it directly succeeds Pentium Pro. Only the execution unit MMX was added. It was pretested in the Pentium processor. The L2 Cache is still in separated chip in package.

In 1998 Intel introduced the high-end version of Pentium II with 512kB of L2 cache and the cache was running on internal processor frequency. This processor was marked as **Xeon**.

The Low-end processor with small or none L2 cache was marked as Celeron and it was used in cheap personal computers.

Pentium III

Y: 1999÷2003 T: BiCMOS 0.25÷0.13µm Tr: 9.5 mil.+18 mil./256kB L2
F: 450MHz÷1.3GHz D: 64b A: 36b

The second successor of the Pentium Pro was introduced in 1999 and it had finally integrated the L2 cache on single chip together with the CPU core.

The execution unit got another helpers – the SSE unit and improved the branch prediction unit.

The power management was significantly improved too. The PIII was the best processor for laptops for the next few years.

The PIII processor was replaced by the next generation of the P4 processor, but later reappeared on the marked in upgraded form as Pentium M.

Pentium 4

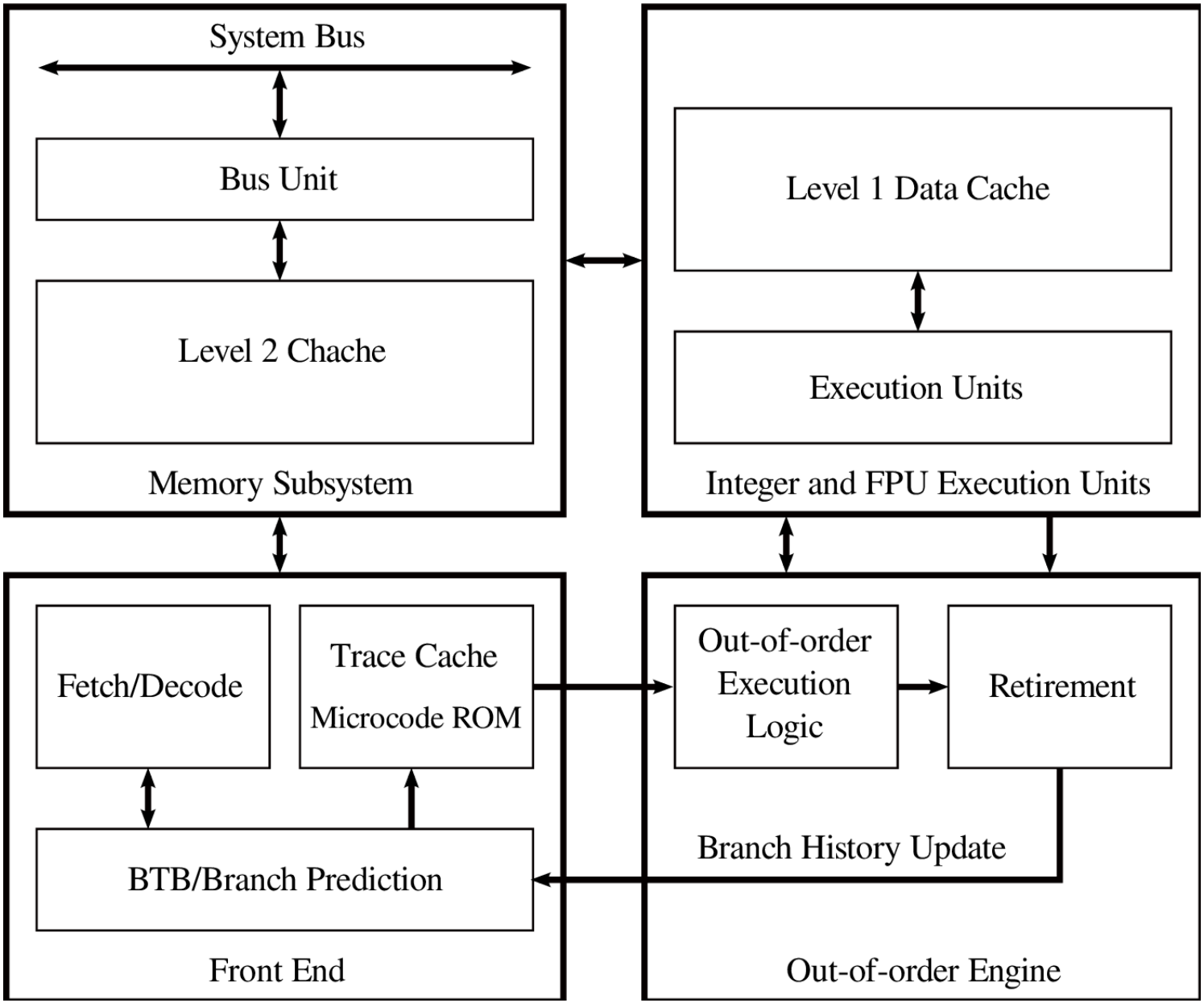
Y: 2000÷2008 T: BiCMOS 0.18÷0.065µm Tr: 42 mil. with 256kB L2
F: 1.3GHz÷3.8GHz D: 64b A: 36b

In 2000 Intel introduced with great expectations the new Pentium 4. This processor implements new architecture NetBurst. It was to be the processor for the new multimedia world. But the first demonstration for experts was very unconvincing. The processor had the same performance as PIII on the same frequency, together with significantly higher current consumption.

When we are observing this processor from today's perspective, we can't help thinking that Intel deliberately made its life difficult by this processor.

Let's look at the internal schema.

... Pentium 4


System Bus
Bus Unit
Level 2 Chache
Memory Subsystem
Level 1 Data Cache
Execution Units
Integer and FPU Execution Units
Fetch/Decode
Trace Cache
Microcode ROM
BTB/Branch Prediction
Front End
Out-of-order Execution Logic
Retirement
Branch History Update
Out-of-order Engine

... Pentium 4

The NetBurst architecture is on the scheme. The scheme looks different than Pentium Pro. But at a closer look we can see that the principle of operation is very similar.

The Fetch and Decode unit translates CISC instructions into RISC ones and sends them to the Out-of-order Execution Unit. The Retirement unit stores then results back to memories. The attentive reader may notice one major change. The L1 code cache is placed after Fetch/Decode unit. Thus it contains decoded micro-operations.

What is not clear from the diagram is that this execution unit has some of its parts working at twice frequency than the rest of the processor.

There is a pipeline with 20 stages in Pentium 4. It is two times longer than in Pentium Pro and it places greater demands on the branch prediction.

Pentium 4 EM64T

Y: 2004 T: BiCMOS 90nm Tr: 125 mil. with 1MB L2
F: 2.8GHz D: 64b A: 40b

In 2004 Intel introduced its first processor, where they had to accept third-party standard. AMD at that time had a successful 64-bit technology, which was created by extending the old 32-bit architecture. Because Intel did not want to lose market position, they had to adapt.

All registers in the processor were extended to 64-bit, 8 new registers were added and address bus was extended to 40 bits. The new technology was marked as EM64T.

The processor had very long pipeline (30 stages), it had to be clocked by high frequency and it was overheating. Power losses of processors were in range from 85 to 115W!

Pentium M

Y: 2003÷2008 T: BiCMOS 130÷65nm Tr: 77 mil. with 1MB L2
F: 900MHz÷2.2GHz D: 64b A: 32b

The new Pentium M processor was introduced in 2003. It was designed primarily for notebooks. Intel took the best from PIII architecture and used latest experiences with bus communication and branch prediction from P4. Intel proposed this new processor with 1MB L2 cache on the chip. The result was surprisingly good.

The Pentium M with 1.5GHz clock had nearly the same performance as P4 with 2.5GHz clock. And the power consumption was only 30% in comparison with P4!

But this processor was strictly sold only for notebook as part of Centrino technology. Intel still produced the P4 for desktops.

Intel Core, Core Duo, Core Solo

Y: 2006 T: BiCMOS 65nm Cores: 1 or 2
F: 1.5GHz ÷ 2.2GHz D: 64b A: 36b

In 2006 the Pentium M continues in the new series known as the Intel Core. This processor was not only designed for notebooks. It had 36-bit address bus and it was sold for desktops and servers too.

The wider bus was not the only new feature. Intel started to implement two cores on the single chip. The power consumption was so improved that it was possible to use multiple cores even in battery-powered computers.

The L2 cache implemented on the chip had capacity usually from 2 to 4MB.

Intel Core 2

Y: 2006 T: BiCMOS 65÷45nm Cores: 1, 2 or 4
F: 1GHz÷3.3GHz D: 64b A: 36b or 40b

In 2006 the new 64-bit processor was introduced – Intel Core 2. It implements the EM64T technology and it was designed for all computer platforms – notebooks, desktops and servers. Manufactured variants differed only by the bus width, number of cores and L2 cache size.

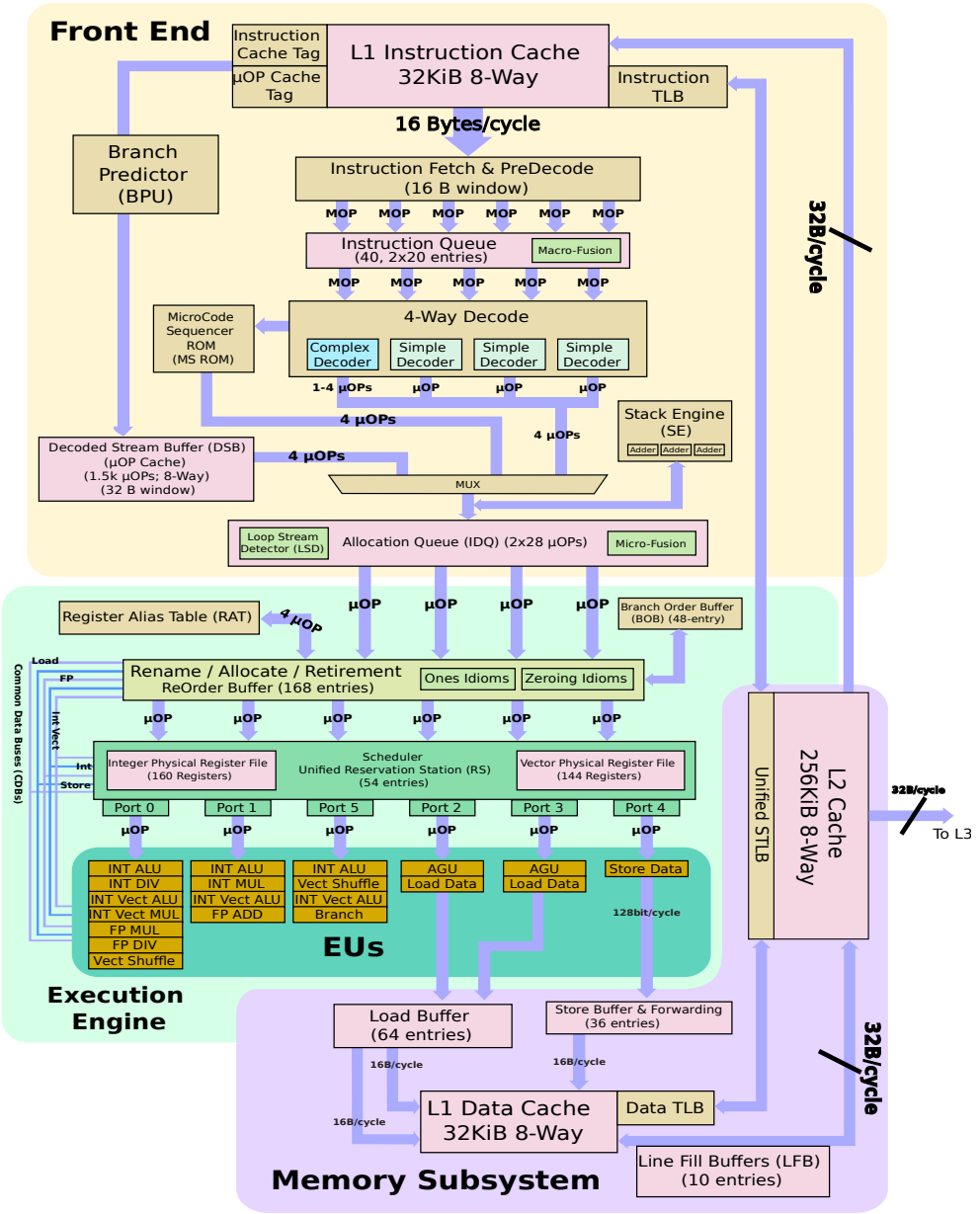
The introduction of this processor virtually stopped the development of NetBurst architecture.

Intel i7 - 9.5th generation

Coffee Lake CPUs are built using the second refinement of Intel's 14 nm process (14 nm++).[6] It features increased transistor gate pitch for a lower current density and higher leakage transistors that allows higher peak power and higher frequency at the expense of die area and idle power.

Coffee Lake marks a shift in the number of cores for Intel's mainstream desktop processors, the first such update for the previous ten-year history of Intel Core CPUs. In the 8th generation, mainstream desktop i7 CPUs feature six cores and 12 threads, i5 CPUs feature six single-threaded cores and i3 CPUs feature four single-threaded cores.

… Intel i7 - 9.5th generation – single core schema



... Intel i7 - 9.5th generation

Features specific to Coffee Lake include:

- Increased core count to six cores on Core i5 and 8th generation i7 parts; Core i3 is now a quad-core brand. 9th generation i7 and i9 parts feature eight cores.
- Increased L3 cache in accordance to the number of threads
- Increased turbo clock speeds across i5 and i7 CPUs models (increased by up to 400 MHz)
- Increased iGPU clock speeds by 50 MHz and rebranded it UHD (Ultra High Definition)
- DDR4 memory support updated for 2666 MHz (for i5, i7 and i9 parts) and 2400 MHz (for i3 parts); DDR3 memory is no longer supported on LGA1151 parts, unless using with H310C chipset
- 300 series chipset on the second revision of socket LGA 1151
- Support for CNVi

Intel Atom

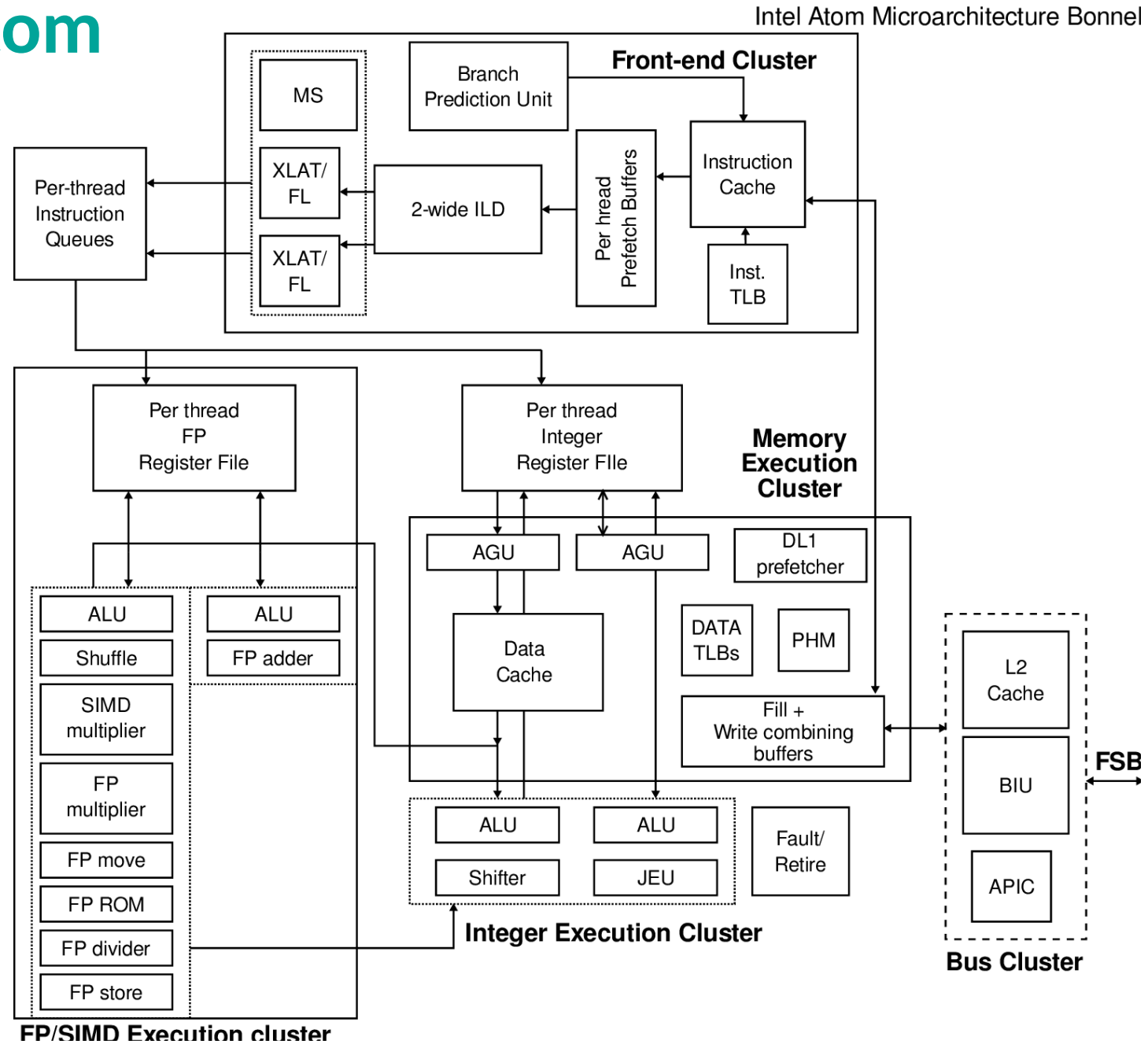
Y: 2008 T: BiCMOS 45nm Cores: 1 or 2
F: 800MHz ÷ 2GHz D: 64b A: 32b

In 2008 Intel presented the processor Atom. Only few processors with the ultra low power consumption were available at that time on the market. These processors were produced by AMD – processors Geode, and by VIA – C7 and Eden. Intel at that time did not produce small processors and they needed to fill the gap on the market.

Atom is not a poor relation of Pentium M as one may think. It is completely new architecture named Bonnell.

The architecture will be shown on the next slide. Principles are clear. The Front-end cluster fetches and decodes instructions and passes them to Instruction queue. The Integer Execution Unit or FP Unit get them from the queue. The Front Side Bus controller and L2 cache are implemented in Bus Unit.

... Intel Atom


Intel Atom Microarchitecture Bonnel
Front-end Cluster
MS
Branch Prediction Unit
XLAT/ FL
XLAT/ FL
2-wide ILD
Per thread Prefetch Buffers
Instruction Cache
Inst. TLB
Per-thread Instruction Queues
Per thread FP Register File
Per thread Integer Register FIle
Memory Execution Cluster
AGU
AGU
DL1 prefetcher
DATA TLBs
PHM
Data Cache
Fill + Write combining buffers
ALU
Shuffle
SIMD multiplier
FP multiplier
FP move
FP ROM
FP divider
FP store
ALU
FP adder
ALU
ALU
Shifter
JEU
Fault/ Retire
Integer Execution Cluster
FP/SIMD Execution cluster
L2 Cache
BIU
APIC
FSB
Bus Cluster

Itanium, Itanium 2

Y: 2001,2002 T: BiCMOS 180÷65nm Tr: 220 mil÷2 bil.
F: 733Hz÷1.7GHz D: 128b A: 50b

Intel introduced Itanium in 2001 and it was supposed to be the 64-bit successor of Pentium family. It was completely designed as the new pure RISC processor.

Its main weakness was the poor backward compatibility with the 32-bit predecessors. It had implemented slow 32-bit decoding unit and users did not accept it.

The first version had the same performance in 32-bit mode as Pentium 100MHz. The second version was produced without that slow 32-bit simulator and the backward compatibility was solved by the software.

The Itanium is designed now for high performance servers and implements up to 24 MB of L3 cache directly on the chip.